Modeling continuous processes from data

Michael Small*
Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong

Kevin Judd and Alistair Mees
Centre for Applied Dynamics and Optimization, Department of Mathematics and Statistics, University of Western Australia, Perth, Australia

(Received 31 July 2001; published 11 April 2002)

Experimental and simulated time series are necessarily discretized in time. However, many real and artificial systems are more naturally modeled as continuous-time systems. This paper reviews the major techniques employed to estimate a continuous vector field from a finite discrete time series. We compare the performance of various methods on experimental and artificial time series and explore the connection between continuous (differential) and discrete (difference equation) systems. As part of this process we propose improvements to existing techniques. Our results demonstrate that the continuous-time dynamics of many noisy data sets can be simulated more accurately by modeling the one-step prediction map than by modeling the vector field. We also show that radial basis models provide superior results to global polynomial models.

DOI: 10.1103/PhysRevE.65.046704 PACS number(s): 95.75.Wx, 02.60.−x, 05.45.Tp, 05.45.Pq

I. INTRODUCTION

It is natural to choose to model many dynamical systems as flows, or as the vector fields that generate these flows. However, observed data are necessarily discretized in time and it may be more appropriate to model them using a map. Various function fitting algorithms are widely employed to estimate a map from an observed data set. Estimation techniques include neural networks [1]; radial basis networks [2,3]; global polynomials [4–6]; local polynomials [1]; global polynomials and standard functions [7,8]; local linear fits [1]; triangulations and tesselations [9]; and Volterra functional expansions [10,11]. Since the underlying dynamical system is often continuous, the quantity of interest is not the ''one-step'' prediction but rather the vector field. There are techniques for estimating the vector field directly from data and we shall consider these in more detail later.

The modeling methods that we examine are (i) radial basis models of the one-step predictions (map), (ii) estimates of vector fields from radial basis map models (Euler derivative of the map), (iii) modeling of the vector field directly using implicit Adams integration schemes and global polynomial models [12,13], and (iv) a technique that involves Adams integration modeling with radial basis models. We do not test any other approaches in this paper. In particular, we only consider the usual polynomial approach and do not consider either rational approximation or the standard function and global polynomial approach of Gouesbet and co-workers [7,8].

To compare models we examined root-mean-square (one-step) prediction errors and qualitative features of the free run dynamics. We found that modeling regimes based on estimating the map [(i) and (ii)] perform better than vector field [(iii) and (iv)] modeling methods. Vector field methods require a large amount of clean, rapidly sampled data. However, even in these situations map models perform better. Overall, simulations from iterated one-step predictions made on the map (i) performed better than the alternatives.

The key to the recovery of the underlying dynamics is to extract an evolution operator (either a map or a vector field). In the remainder of this introduction we review some mathematical techniques employed in this paper. In Sec. I A we introduce the Adams integration technique to estimate a vector field, Sec. I B describes the radial basis modeling techniques we use to estimate a map, and Sec. I C described the minimum description length (MDL) model selection criterion.

Section II describes various methods that we will use to estimate vector fields. Section III compares numerical results obtained by applying these techniques to artificial and experimental data from several dynamical systems.

A. Adams integration

Modeling a vector field usually proceeds by fitting a function of the data values (the embedded time series) to an estimate of the vector derivative. This estimate will usually be based on the numerical Euler derivative $\dot{f}(x_{n+1})=(x_{n+1}-x_n)/\tau$. Euler integration utilizes this formula to numerically integrate ordinary differential equations [14,15]. Brown and co-workers [12,13,16] propose an extension to this method based on the implicit Adams integration scheme [14]. The implicit Adams integration formula is given by

$$x_{n+1}=x_n+\tau\sum_{i=0}^{m} a_i\dot{f}(x_{n+1-i}), \tag{1}$$

where m is the order of the Adams integrator and a_i are the Adams coefficients (see Ref. [14]). For $m=0$ Adams integration is equivalent to Euler integration, and the Adams/

*Formerly at Center for Applied Dynamics and Optimization, Department of Mathematics and Statistics, University of Western Australia, Perth, Australia; electronic address: ensmall@polyu.edu.hk

MDL technique proposed by Brown *et al.* [13] is equivalent to fitting embedded data to the usual first difference.

Brown and co-workers have implemented this method to build global polynomial models of vector fields from data. They select the order of the global polynomial and the order of Adams integration according to minimum description length [17] (see Sec. I C). The global polynomial model is built from an orthonormal (on the data/attractor) set of polynomials and they recommend the selection of coefficients (the weights of the orthonormal polynomials) not according to least mean squares but by utilizing the orthonormality [13,18]. They demonstrate these techniques with artificial noise-free data from the Lorenz system and from integration of chemical passivation equations, and experimental data from an electronic circuit [19], a chemical reaction and an experiment involving a vibrating string [16].

B. Radial basis modeling techniques

Utilizing a time-delay embedding one may construct a scalar function of a vector (embedded) variable, such that the vector map in state space

$$x_{n+1}=F(x_n)+e_n, \tag{2}$$

[where e_n are independent and identically distributed (i.i.d.) random variates corresponding to model prediction error and system noise] is given by

$$F(x_n)=(g(x_n),(x_n)_1,(x_n)_2,\ldots,(x_n)_{d_e-1}),$$

where $(x_n)_i \in \mathbf{R}$ is the ith component of $x_n \in \mathbf{R}^{\mathbf{d}_e}$ and

$$g(x_n)=\lambda_0+\sum_{i=1}^{a}\lambda_i(x_n)_{\ell_i}+\sum_{j=1}^{b}\lambda_{j+a+1}\phi\left(\frac{\|x_n-c_j\|}{r_j}\right). \tag{3}$$

Here λ_i, λ_j, and r_j are scalar constants, $1\leqslant \ell_i<\ell_{i+1}\leqslant de\tau$ are integers and c_j are arbitrary points in $\mathbf{R}^{\mathbf{d}}$. The integer parameters a and b are selected to minimize the description length (to be described in Sec. I C). The scalar function $\phi(\cdot)$ represents the class of radial basis function from which the model will be built. We choose to use Gaussian basis functions because they appear to be capable of modeling a wide variety of phenomena. In this case, the basis functions are described by

$$\phi(x)=\exp\left[\frac{-x^2}{2}\right].$$

C. Minimum description length

The minimum description length criterion is based on information theoretic ideas [17] and measures the compression achieved by describing the data in terms of a model and the model prediction error, compared to simply describing the data. The minimum description length criterion is a generalization of both the Akaike [20] and Schwarz [21] information criteria.

Rissanen [17] shows that the description length of a parameter λ_i specified to some accuracy δ_i is $\ln(\gamma/\delta_i)$ [2,17]. The constant γ is not critical and is related to the binary representation of floating point numbers [2]. Therefore, the description length of k model parameters $\Lambda=\{\lambda_i\}_{i=1}^{k}$ is given by

$$L(\Lambda)=\sum_{i=1}^{k}\ln\left(\frac{\gamma}{\delta_i}\right).$$

The description length of the data $\{x_t\}_{t=1}^{N}$, and the model with parameters Λ, is given by

$$L(x,\Lambda)=L(x|\Lambda)+L(\Lambda), \tag{4}$$

where the description length of the data given the model (i.e., the description length of the model prediction errors) $L(x|\Lambda)$ is the negative logarithm of the likelihood of the data under the assumed distribution, $-\ln[P(x|\Lambda)]$.

If a model provides a ''good'' description of the data then the description length of the model parameters and model prediction error will be small. If a model provides a ''poor'' description, or is excessively large (i.e., overfits the data) then the description length of the model parameters and model prediction error will be large. Detailed discussion of the minimum description length principle may be found in Ref. [17], and application of this technique to radial basis modeling is described in more detail in Ref. [2].

D. Outline

The remainder of this paper describes our alterations to these techniques and compares the modeling results with experimental and artificial data. Specifically, we compare (i) radial basis models of the ''one-step'' map (Sec. I B); (ii) an Euler approximation to the derivative of the map estimated in (i); (iii) the implicit Adams integration/global polynomial techniques described in Sec. I A; and (iv) radial basis techniques utilizing the implicit Adams integration schemes.

When faced with the problem of reconstructing a vector field from experimental (possibly scalar) data one may take many different approaches. In general, we will assume that Takens' embedding theorem [22], or some equivalent technique, has been applied, and we will consider only vector time series, and the problem of vector prediction. Section II describes methods to estimate the vector field from experimental data and Sec. III presents the results of some numerical work.

II. MODELING VECTOR FIELDS

In this section we explore alternative approaches to modeling vector fields from data. In a separate paper we have suggested deriving the vector field by examining the limiting values ($\tau_0 \to 0$) of a map [23]. The methods in this section

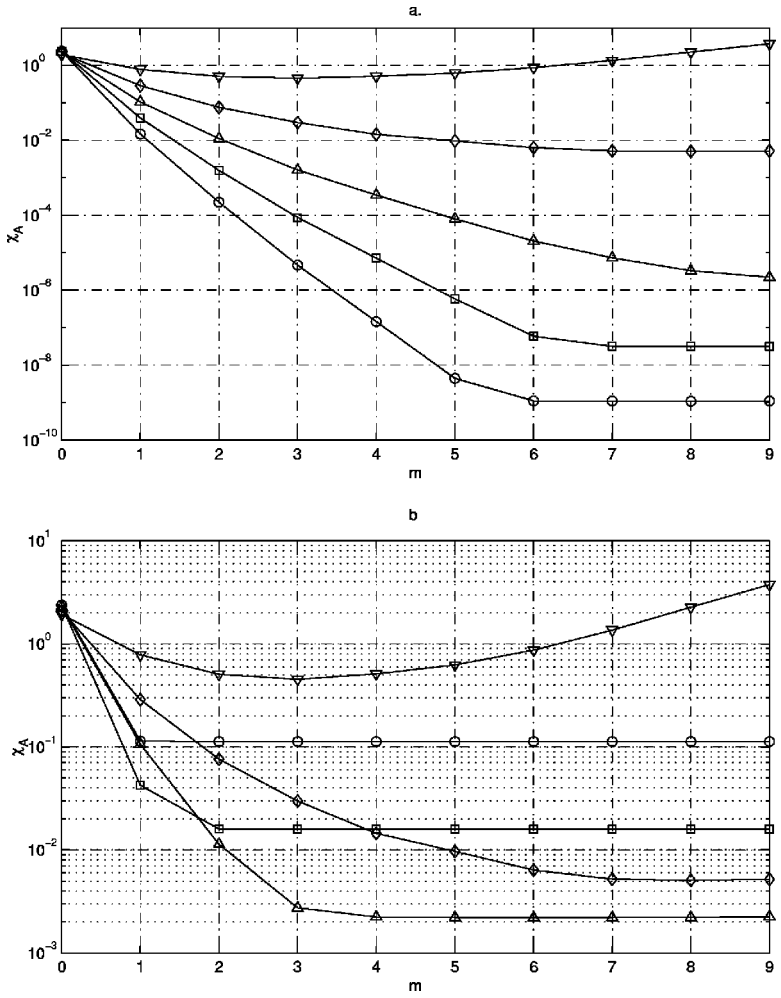

FIG. 1. Relative prediction error for the Lorenz system (1). The relative prediction error is given by $\chi_A=(1/N\Sigma_{n=1}^{N}\|x_{n+1}-[x_n+\tau\Sigma_{i=0}^{m}a_iF(x_{n+1-i})]\|^2/\|x_{n+1}-n_n\|^{2\,1/2}$ for a set of N embedded data points. The calculations in panel (a) and (b) are the same calculations for an 18 sec segment of the Lorenz system, with a sampling interval of $\tau=0.001$ sec (○), 0.002 659 sec (□), 0.007 071 sec (△), 0.01880 sec (◇), and 0.05 sec (▽). The value of χ_A is calculated for each data set with the given sampling rates for the Adams integration (m) of order 0 to 9. Panel (b) shows the results when the Lorenz data is calculated to five decimal figures, and panel (a) are the same calculations for Lorenz data calculated using double precision arithmetic. A similar calculation was presented in Refs. [12,13]. One can see that in the presence of minimal noise (one part in 2^{16}) and small to medium step size, high order Adams integration performs best. Note also, that the variation between panel (a) and panel (b) may be attributed to additional factors besides the truncation at the fifth decimal place.

attempt to determine the vector field directly, either by applying a variant of Adams/MDL techniques described by Brown and co-workers [12,13], or by estimating the Euler derivative from a map model of the form (2). These methods are based on finite approximations to the derivative and, therefore, require data sampled ''sufficiently'' often to be able to approximate the vector field accurately.

Estimation of the Euler derivative from a map is trivial. In Secs. II A and II B we consider methods based on the implicit Adam integration scheme and minimum description length. Section II A describes global polynomial based modeling methods, and Sec. II B describes the application of radial basis modeling techniques to the same scheme.

A. Implicit Adams integration

In Sec. I A we described the application of an implicit Adams integration scheme, minimum description length, and global polynomial modeling as described in Ref. [12]. We have coded the algorithm suggested by Brown and co-workers [12] with consistent results. We have been able to accurately reproduce the calculations of Fig. 3 of Ref. [12]. However, our results are only comparable to those in Ref. [12] if the data of the Lorenz system are truncated (as they do) to five decimal places. Allowing for double precision arithmetic we achieve substantially different results. Figure 1 demonstrates this comparison. These calculations support the conclusions made by Brown and co-workers [12] that these methods are suitable in the presence of only small amounts of noise and moderately highly sampled systems. For larger amounts of noise (i.e., one part in 10^5) there appears to be an optimal, intermediate, step size.

The results of Fig. 1 are calculated from a second-order global polynomial model, and the Lorenz system can be adequately described by a second-order polynomial model. However, we suggest that higher-order terms are required to adequately account for the approximation associated with finite sampling of a continuous system. To test this it is necessary to modify the algorithm described by Ref. [12] and apply the minimum description length criterion in a stricter sense.

For a fixed maximum order polynomial N_p and fixed order of Adams integration m, Brown and co-workers build a global polynomial model of the vector field using *all* the polynomial basis functions and calculate the description length of *that* model. However, using a subset selection algorithm described elsewhere [2] we suggest an alteration to

TABLE I. Global polynomial model of the Lorenz system (1). Model built from 5000 data points (sampled every 0.05 sec) of the original Lorenz system using an Adams integration scheme and strict minimum description length. The data was contaminated with observational noise on all three components (normal with standard deviation 5% of the standard deviation of the data values). This model has polynomial terms up to fourth order and was built using $m=0$ Adams integration (i.e., Euler integration). In Fig. 1 we demonstrated that high-order Adams integration works well for systems with no noise; for systems with more noise, a much lower-order Adams integration scheme is selected. In this case Euler integration is selected as optimal according to minimum description length. Note that, many more terms are selected in this model than are present in the original Lorenz equations. The higher-order terms in this model are possibly related to the higher-order terms in an evaluation of Eq. (6)—this expansion is necessary because of the finite set size and finite sampling rate.

x component		y component		z component	
Coefficient	Term	Coefficient	Term	Coefficient	Term
0.0005189113	xy^3	-0.00075278227	xz^3	0.31095388	y^2
-14.809195	y	15.952365	y	0.0012658103	x^2z^2
37.857094	x	-0.0074572455	x^2yz	-0.00038309294	z^3
0.63649195	x^3	-0.0063938793	yz^2	-1.6396998	1
0.01607577	x^2yz	0.00079663814	y^3z	-0.00026448879	y^4
-0.07983783	y^3	0.00048246071	yz^3	0.052881899	xyz
-7.4128861×10^{-5}	yz^3	-12.144337	x	0.0036643532	x^3y
-2.3880938	xz	0.0094646419	x^3z	-2.6404106	z
-4.2530913×10^{-5}	xyz^2	-0.35127338	x^3	-0.0004175926	y^2z^2
-0.0082001822	xy^2z	-0.02775479	y^3	-0.0060493741	x^4
-0.037756234	y^2	-0.03153251	x^2	2.3421078×10^{-6}	yz^3
1.0451196	yz	0.28592214	x^2y	-0.59313148	x^2
-0.8351578	x^2y	0.065101065	z	-0.00024747432	y^3
0.0015206632	y^3z	-4.3400433×10^{-6}	z^4	-0.00066517359	xyz^2
-0.012620472	x^3z	5.9716446×10^{-5}	x^2z^2		
0.092927835	x^2	-0.0001289389	x^3y		
3.8935607×10^{-5}	y^2z^2	0.97795022	xz		
-5.1375924×10^{-5}	z^3	3.2744309×10^{-5}	y^4		
0.42323624	xy^2	-0.69856633	yz		
-3.3274152×10^{-5}	x^2z^2				

this algorithm. A strict application of the minimum description length principle would involve the following; for fixed values of N_p and m one builds *the* model with minimum description length by using *only* those basis functions that are necessary. That is, only those terms that contribute an overall decrease in the description length. One then selects the values of m and N_p, which yield the minimum description length model.

Our calculations indicate that an implementation of this algorithm will generally produce superior results to the method described by Brown and co-workers. However, the difference is usually marginal. Using this strict application of minimum description length we are able to determine the presence of higher-order terms in the minimum description length best model of the Lorenz system (see Table I). These terms correspond to additional approximation associated with the finite sampling of a continuous system.

The expected values of these additional terms may be calculated analytically. Given $\dot{x}=f(x)$, then

$$x(t+\tau_0)\approx x(t)+\tau_0 f(x(t)), \tag{5}$$

so define

$$F(x)=x+\tau_0 f(x). \tag{6}$$

For example, for a small step size τ_0 the evolution over time τ_0 of the Lorenz system may be (approximately) represented in this form. One may take the nth iterate $F^n([x,y,z]^T)$ of the map (6) as an approximation to the difference equations of the Lorenz system with step size $n\tau_0$. Alternatively, one may take the Taylor series (in powers of τ_0) and using the identity provided by the Lorenz equations to obtain a similar approximations. Calculations show that the remainder term of the Taylor series expansion diminishes slowly, and therefore, may not provide a good approximation to the map.

Furthermore, an evaluation of Eq. (6) yields substantially different results to those shown in Table I. Calculations show that the coefficients estimated in Table I are sensitive to noise

TABLE II. Global polynomial model of the Lorenz system (1). The same calculation as Table I, except the data was sampled every 0.001 sec. For this higher sampling rate the improved MDL criteria described in this paper has been able to accurately determine the exact equations of the underlying systems (to one part in 10^{10}).

x component		y component		z component	
Coefficient	Term	Coefficient	Term	Coefficient	Term
1.5034525×10^{-13}	xz	-1	y	-6.5077874×10^{-14}	z^2
4.3673276×10^{-13}	yz	1.3248492×10^{-12}	yz	2.3392705×10^{-12}	x^2
10	y	-1	xz	3.1620093×10^{-11}	1
-10	x	28	x	1	xy
8.5557938×10^{-14}	xy	9.474207×10^{-14}	y^2	-6.2292401×10^{-13}	y^2
				-1.4828462×10^{-12}	y
				9.5964037×10^{-14}	yz
				-2.6666667	z

level and the length of the time series (also see Table II). Improved results are also obtained by fitting the model to many short transients in the neighborhood of the attractor (instead of a single trajectory *on* the attractor). A large number of short transients in the neighborhood of the attractor provide superior coverage of phase space compared to a single trajectory. The improved modeling results obtained with many transients provides further evidence of the numerical sensitivity of global polynomial modeling techniques. Similar observations have been expressed by other authors [24,25].

B. Adams/MDL/radial basis methods

It appears that global polynomial models are sufficient if the underlying system is known to be described by polynomial nonlinearities and the time step is sufficiently small. In fact, the Weiestrass convergence theorem [26] guarantees that for an arbitrary continuous function there exist some sequence of polynomials converging to it. However, for certain functions this convergence can be very slow or the sequence of polynomials may be nonobvious. For example, polynomials interpolated at equally spaced points to the function $|\cdot|$ (defined in a symmetric interval about zero) converge only at zero and the interval end points [26]. It has also been noted that like other unbounded basis functions, extrapolation using polynomials can be hazardous [15].

We have found that better results are obtained by the use of radial basis function networks [3], and in particular, Gaussian basis functions [2] or variants [27]. In a method directly analogous to that described in Sec. II A we can apply strict minimum description length to implicit Adams integration modeling (1) of a vector field. We apply radial basis modeling techniques to fit Eq. (1) as a sum of Gaussian basis functions.

That is, we fit the function $F(x_n)$ to expressions of the form

$$x_{n+1}=x_n+\tau\sum_{i=0}^{m} a_i F(x_{n+1-i}), \tag{7}$$

where

$$F(x_n)=(g_1(x_n),g_2(x_n),\ldots,g_{d_e}(x_n))$$

and each g_i is a function of the form

$$g_i(x_n)=\lambda_0+\sum_{i=1}^{n_i}\lambda_i(x_n)_{\not\subset_i}+\sum_{j=1}^{n_j}\lambda_{j+n+1}\phi\left(\frac{\|x_n-c_j\|}{r_j}\right).$$

The model that minimizes description length is then selected as the best. In this formulation we fit d_e scalar functions and do not utilize additional information available from a time-delay reconstruction. (See Sec. I B and [2]). A weakness of this approach is that there is no guarantee that the predicted values $g_i(x_n)$ will be appropriately correlated [38].

For a set of d candidate pseudolinear basis functions, let X be an $N\times d$ matrix such that the ith column of X is the evaluation of the ith basis function over the data, and the jth row of X is the evaluation of all the candidate basis functions at the jth (vector) datum. The subset selection algorithm discussed in Ref. [2] will select columns $\mathcal{I}=\{i_1,i_2,\ldots,i_n\}$ of the matrix X and a weight vector $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ so that the description length of $\{x_n\}_{n=1}^N$ is minimized by describing the model prediction error and the model itself, namely,

$$\sum_{k=1}^{n}\lambda_k X_{(i_k,:)}. \tag{8}$$

Here $X_{(i_k,:)}$ denotes the i_kth column of X. That is, this algorithm selects a set of basis functions from a larger group of candidates, based on the evaluation of these functions over the data. However, if we are to build a function F to minimize the description length of the modeling errors of Eq. (7) we must generalize Eq. (8) and the associated subset selection algorithm. Let

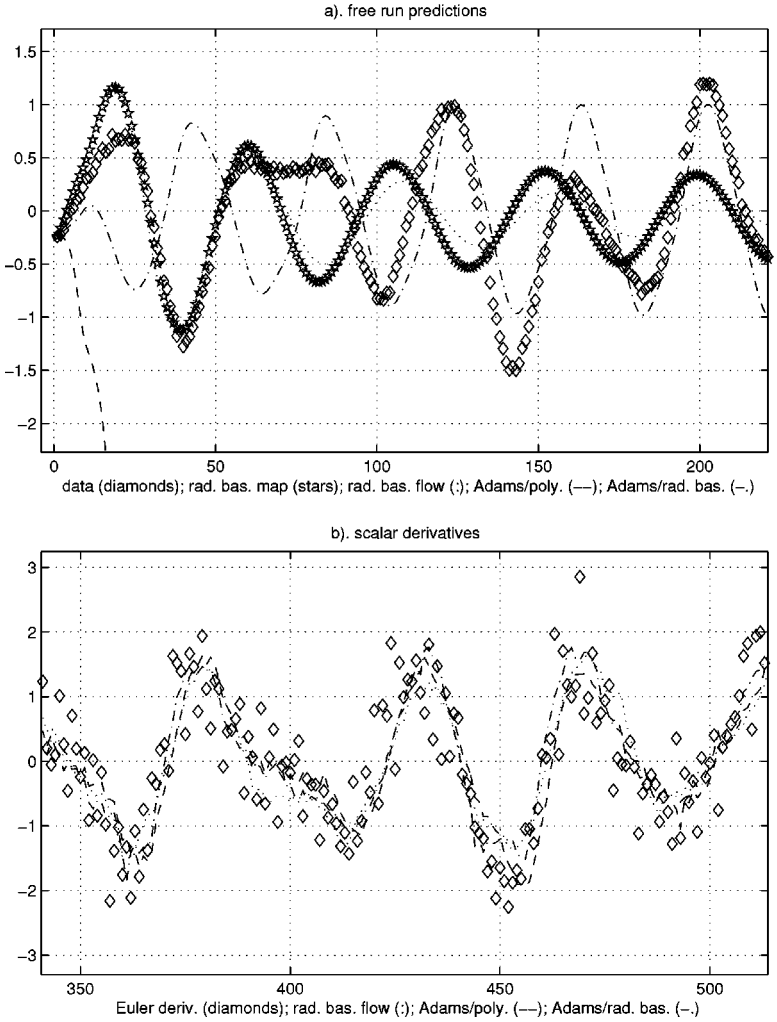

FIG. 2. Simulated behavior from reconstructed vector field of circuit equations (3). Reconstruction of the vector field from a time-delay embedding of one coordinate of the circuit equation described by Ref. [19] in the presence of additive noise [$N(0,\sigma^2)$, $\sigma^2 \approx 0.05$(standard deviation data)]. The data shown in this figure is different from that used to build these models (''honest'' predictions). Panel (a) shows trajectories integrated on the reconstructed vector field, and panel (b) shows estimates of the (scalar) derivatives of the original trajectory from the vector fields. The data are shown as diamonds ($\diamond$), radial basis map predictions as stars ($\star$), radial basis flow (Euler derivative of the map and integrated) as a dotted line ($\cdots$), the Adams integration/global polynomial scheme as a dashed line ($--$), and Adams integration/radial basis model as a dot-dashed line ($-\cdot-$). Note that the Adams/global polynomial approach is quickly divergent, and the other simulations tend to periodic orbits or stable foci. The original system is chaotic.

$$A_m = \begin{bmatrix} a_m & a_{m-1} & \cdots & a_0 & 0 & 0 & \cdots & 0 \\ 0 & a_m & a_{m-1} & \cdots & a_0 & 0 & \cdots & 0 \\ 0 & 0 & a_m & \cdots & a_1 & a_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_m & a_{m-1} & \cdots & a_0 \end{bmatrix}, \tag{9}$$

$$y = \frac{1}{\tau}\begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_N - x_{N-1} \end{bmatrix}. \tag{10}$$

The parameters $a_0, a_1, a_2, \ldots, a_m$ are the mth order Adams integration coefficients as in Eq. (1).

The minimum description length best model of Eq. (7) may now be obtained by applying the subset selection algorithm of Ref. [2] to the matrix $A_m X$ to fit y. Because the basis function weights appear only linearly this is identical to fitting a function of the form (2).

III. RESULTS

Numerical experiments were conducted with data from the following four simulated and three experimental systems:

(1) The ubiquitous Lorenz system [(x,y,z) coordinates, $\tau = 0.05$, $N = 4000$] in the chaotic regime ($s = 10$, $r = 28$, $b = 8/3$). The Lorenz equations are

$$\dot{x} = s(y - x),$$

$$\dot{y} = rx - y - xz,$$

$$\dot{z} = xy - bz.$$

(2) Reconstructed chaotic ($s = 10$, $r = 28$, $b = 8/3$) Lorenz system (x component, $\tau = 0.05$, $N = 4000$, $d_e = 3,4$, lag$= 7$)with and without observational noise [$N(0,\sigma^2)$, $\sigma^2 \approx 0.05$(standard deviation data)].

(3) Circuit equations described by Rulkov and Volkovskii [19] (y component, $\tau = 0.1$, $N = 4000$, $d_e = 4$, lag$= 9$) with and without observational noise [$N(0,\sigma^2)$, $\sigma^2 \approx 0.05$(standard deviation data)], see Fig. 2. The circuit

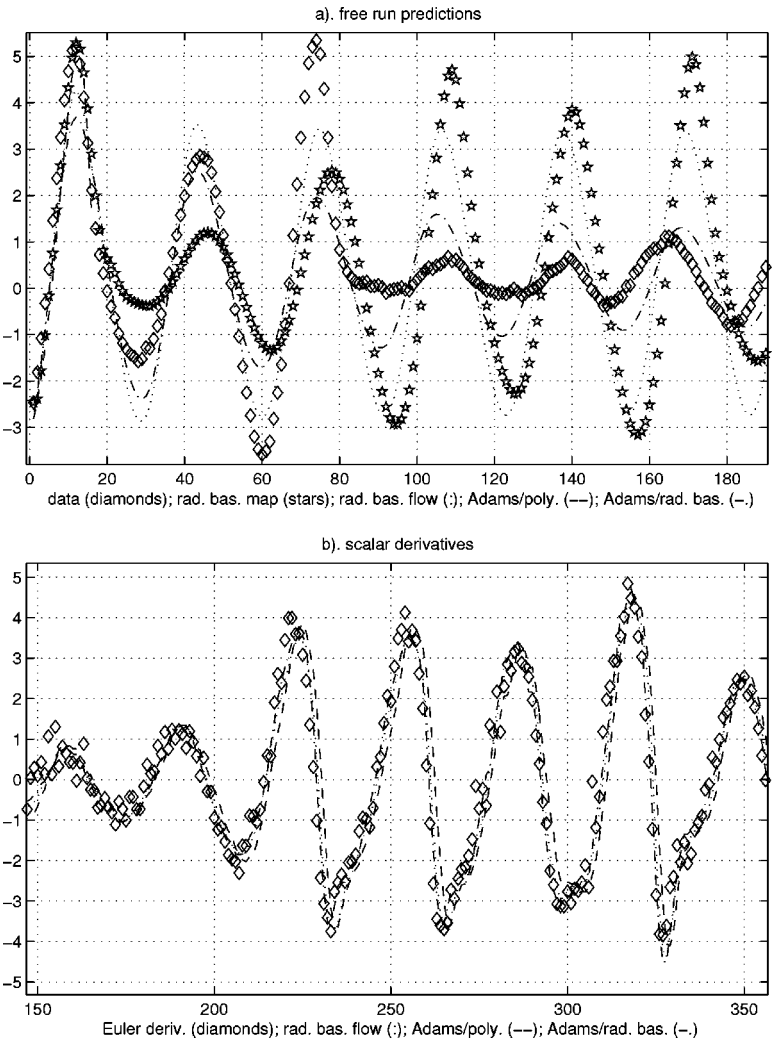

FIG. 3. Simulated behavior from reconstructed vector field of Rössler equations (4). Reconstruction of the vector field from a time-delay embedding of one coordinate of the Rössler equations in the presence of dynamic [$N(0,\sigma^2)$, $\sigma^2 \approx 0.025$ (standard deviation data), dynamic noise on each component of the system] and observational [$N(0,\sigma^2)$, $\sigma^2 \approx 0.05$ (standard deviation data)] noise. The data shown in these figures is different from that used to build these models. Panel (a) shows trajectories integrated on the reconstructed vector field, and panel (b) shows estimates of the (scalar) derivatives of the original trajectory from the vector fields. The data are shown as diamonds ($\diamond$), radial basis map predictions as stars ($\star$), radial basis flow (Euler derivative of the map) as a dotted line ($\cdots$), the Adams integration/global polynomial scheme as a dashed line ($--$), and Adams integration/radial basis model as a dot-dashed line ($-\cdot-$). Note that the Adams/global polynomial approach quickly becomes singular (at about the 18th datum with a value of approximately 1.7), the Adams/radial basis model slowly converges to a stable foci. The other models exhibit periodic orbits. The original system exhibits a noise driven periodic orbit.

equations in Ref. [19] are

$$\dot{x} = y,$$

$$\dot{y} = -x - \delta y + z,$$

$$\dot{z} = \gamma[h(x) - z] - \sigma y,$$

$$h(x) = \begin{cases} 0.528\alpha, & x < -1.2, \\ x(1-x^2)\alpha, & -1.2 \leqslant x < 1.2, \\ 0.528\alpha, & x \geqslant 1.2, \end{cases}$$

$$\alpha = 22.3, \quad \delta = r\sqrt{\frac{C_2}{L}}, \quad \gamma = \frac{\sqrt{LC_2}}{RC_1}, \quad \sigma = \frac{C_2}{C_1}.$$

For the simulations described in this paper we used the following parameter values: $C_1 = C_2 = 375$ nF, $L = 233.7$ mH, $r = 0.407$ kΩ, and $R = 6$ kΩ. With these parameter values the attractor is vaguely ''Lorenz-like''—it has two separate ''wings'' and a central separatrix.

(4) Reconstruction of Rössler equations with period 3 behavior ($a = 0.411$, $b = 2$, $c = 4$, y component, $\tau = 0.1$, $N = 4000$, $d_e = 4$, lag$= 9$) with and without noise (normal dynamic and/or observational noise with a standard deviation of 2.5% and 5% of the standard deviation of the data, respectively), see Fig. 3. The Rössler equations are

$$\dot{x} = -y - z,$$

$$\dot{y} = x + ay,$$

$$\dot{z} = b + z(x - c).$$

(5) Experimental data from an apparently chaotic laser [28] ($\tau = 800$ ns, $N = 1150$, $d_e = 5$, lag$= 5$), see Fig. 4.

(6) Experimental data from a vibrating string [16,29] ($\tau = 3.73$ ms, $N = 3200$, $d_e = 4$, lag$= 12$).

(7) Experimental data from Japanese vowel (u) sounds [30] ($\tau = 104.167$ μs, $N = 3200$, $d_e = 4$, lag$= 10$).

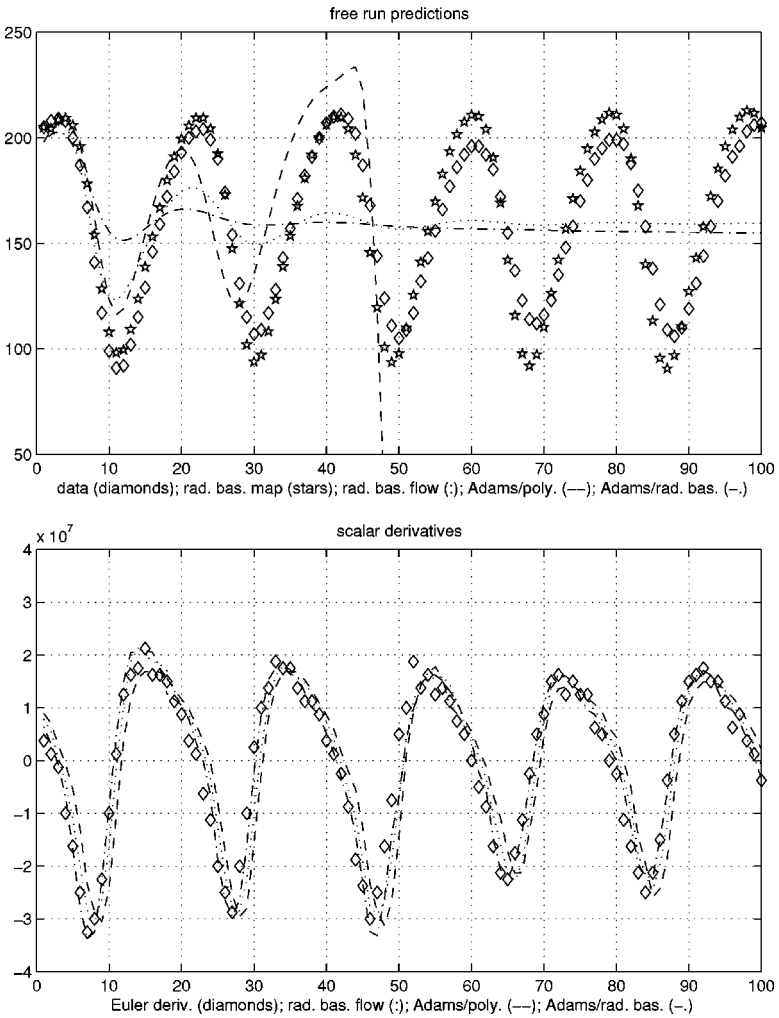

FIG. 4. Simulated behavior from reconstructed vector field of experimental laser data (5). Reconstruction of the vector field from a time-delay embedding of one coordinate of experimental laser data described by Ref. [28]. The data shown in these figures is different from that used to build these models (the model was built on 1150 data points and tested on the next 350). Panel (a) shows trajectories integrated on the reconstructed vector field, and panel (b) shows estimates of the (scalar) derivatives of the original trajectory from the vector fields. The data are shown as diamonds ($\diamond$), radial basis map predictions as stars ($\star$), radial basis flow (Euler derivative of the map) as a dotted line ($\cdots$), the Adams integration/global polynomial scheme as a dashed line ($--$), and Adams integration/ radial basis model as a dot-dashed line ($-\cdot-$). Note that the Adams/global polynomial approach is quickly divergent, and the other vector field based methods exhibit stable foci. The data is (apparently) chaotic [28], and so is the radial basis model map.

Root-mean-square one-step prediction error and minimum description length of all models of each system were roughly comparable (see Table III). We tested the performance of the various modeling schemes described in Secs. II A and II B by comparing the trajectories of numerical integration of these models to the original data. By comparing the behavior of the numerical integration on the reconstructed vector fields we found that the global polynomial implementation of an Adams integration scheme did not produce adequate results for any of these data sets. In each case, this method produced a vector field that soon became singular when tested with a variety of integration schemes [39]. That is, the required precision for numerical integration along a trajectory exceeded the available machine precision. The implementation of Adams integration using radial basis modeling produced superior results. However, for noisy or reconstructed systems the vector field generally exhibited a stable focus. In all cases the radial basis modeling method described in Refs. [2,27] and applied to model the map performed best. For a typical example of the results of these calculations see Figs. 2, 3, and 4.

We have assessed the performance of each of these models using purely qualitative comparisons. In situations where this is insufficient it would be necessary to seek qualitative techniques such as those utilized in Ref. [4] or statistical techniques suggested by Ref. [31]. To adopt these approaches one would apply some significant measure (such as correlation dimension) to the data and model simulations and compare the deviation [4,31]. A good model would produce simulations that are typical of the data (with respect to the chosen measure). A poor model would be obviously distinct from the data. However, for the data and models presented here this is unnecessary. In each case a valid comparison can be made based on the asymptotic dynamics of the systems. Figure 5 compares the attractor reconstructed from the laser data (5), and the attractor provided by a simulation from a radial basis map model.

In each data set and every trial, the vector field estimated with global polynomials was either rapidly divergent or became singular. Vector fields estimated from radial basis modeling and fitted using the implicit Adams integration formula produce superior results. The vector field produced by this method was more stable (it could be integrated numerically), and the results were more consistent with the data. Long term behavior was generally still poor. For the data considered here only one of these methods accurately and consistently modeled the long term dynamics—the radial basis model of the map (see Fig. 5). Methods aimed at modeling the vector field from data appear to perform poorly when faced with small to moderate noise level, reconstructed sys-

TABLE III. Prediction error and description length. This table summarizes representative results of the modeling algorithms discussed in this paper. These algorithms were applied repeatedly to these and other data sets. The table columns, from left to right, are (i) the system/time series studied, (ii) observational noise, (iii) dynamic noise, (iv) sampling rate in seconds, (v) reconstruction embedding dimension, and (vi) reconstruction embedding lag expressed as a number of data points. Observational and dynamic noise levels are standard deviations of Gaussian random processes, both given as a fraction of the standard deviation of the data and as an absolute number (in parentheses). The remaining columns of the table give values for root-mean-square modeling prediction error and minimum description length for radial basis map model, Adams integration/global polynomial model, and Adams integration/radial basis model. For the map radial basis model, these values are calculated in terms of the prediction error $\|x_{n+1}-x_n\|$, while for the Adams integration scheme methods the error is in terms of $\|x_{n+1}-(x_n+\tau F(x_n))\|$. An approximate comparison may be made between the modeling error calculated in either way. (However, they are not identical.) An equivalence between values of description length is not as straightforward because of the difficulty in comparing different model types in an unbiased way. The other map model prediction errors are for scalar functions (predicting the first component).

System	Observational noise	Dynamic noise	τ (sec)	d_e	Lag	Modeling error			Minimum description length		
						Map	Polynomial	Radial basis	Map	Polynomial	Radial basis
(1) Lorenz	0.1 (1.606) [a]	0	0.05	NA	NA	2.0766 [a]	2.0450	2.1859	24051	59 941	59644
(2) Lorenz	0	0	0.05	3	7	0.56	1.05	1.01	4261	54710	54405
(2) Lorenz	0.05 (0.40)	0	0.05	3	7	1.015	1.22	1.155	6331	75202	75127
(3) circuit	0	0	0.1	4	9	0.01	0.01144	0.01035	−1231	−1151	−988
(3) circuit	0.05 (0.03)	0	0.1	4	9	0.0556	0.04592	0.04662	−5843	10623	10900
(4) Rössler	0	0.025 (0.055)	0.2	4	9	0.0687	0.07084	0.06448	−4582	6621	6057
(4) Rössler	0.05 (0.11)	0.025 (0.055)	0.2	4	9	0.188	0.1737	0.17504	−995	21238	21187
(5) Laser	NA	NA	8×10^{-7}	5	5	1.619	1.3426	1.6374	4304	127440	129150
(6) String	NA	NA	3.73×10^{-3}	4	12	108	30.929	28.556	20869	136247	138085
(7) Vowel	NA	NA	1.04×10^{-4}	4	10	581	474.683	478.671	25028	215098	215228

[a]The root-mean-square prediction errors are vector predictions of the vector variable (this system is in the original coordinates, not a time delay-reconstruction).

tems, and/or moderately sparsely sampled systems. These results indicate that estimating the dynamics as a map produce superior results to estimating the vector field.

IV. CONCLUSIONS

Our calculations corroborate the results presented by Brown and co-workers [13]. Adams integration based methods applied to estimate vector fields outperform Euler integration techniques. When applying these methods with a strict minimum description length model selection criterion it is clear that this must be the case. The difference between the results obtained with strict MDL and those described by Brown and co-workers appear to be minimal. However, we contend that global polynomial modeling may not be the most appropriate method, in general. While this class of functions may be well suited to some situations (in particular, polynomial nonlinear ordinary differential equations with a small integration time step and minimal noise) they may not work so well in general. As we have noted, the Weiestrass convergence theorem guarantees the existence of a good polynomial model of a smooth function [26]. However, it is often observed that polynomials may perform poorly in practice, especially when extrapolating [15]. In this paper we have considered polynomial functions, but not rational functions. Menard and co-workers [32] have observed that for map models, a better choice is rational approximation. Whereas ordinary polynomials do not have poles, a ratio of two polynomials allows for a finite number and is therefore better equipped to fit the dynamics observed in many maps [32].

Furthermore, Sec. II A demonstrated that global polynomial models are sensitive to noise, sampling rate, and the ''coverage'' of phase space provided by the data (these models do not extrapolate well). The issue of ''coverage'' of the attractor, and some concerns on applications of global polynomial methods are discussed in Ref. [33]. Furthermore, Aguirre and Billings have observed that global polynomial models are particularly prone to the problem of over parameterization [24,34]. Simulations from models built from several transient trajectories, gave better simulations compared to models built from a single long trajectory. The recent work of Bezruchko and co-workers [35] corroborates this observation. Conversely, Letellier and co-workers [36] have shown that polynomial nonlinearities may be reconstructed with a polynomial model using a single unstable periodic orbit, or the laminar phase of intermittent chaos. Of course, for nonpolynomial nonlinearities the effectiveness of extrapolating in this way may be less [15]. Among the current polynomial methods to estimate vector fields from data it appears that global polynomial modeling with an Adams integration scheme is one of the more effective model. Of course, these results do not necessarily extend to rational functions. From the results presented here it is unclear whether rational function approximations would outperform radial basis models—especially for modeling maps.

We compared numerical results for several modeling techniques for many data sets. Statistics such as minimum de-

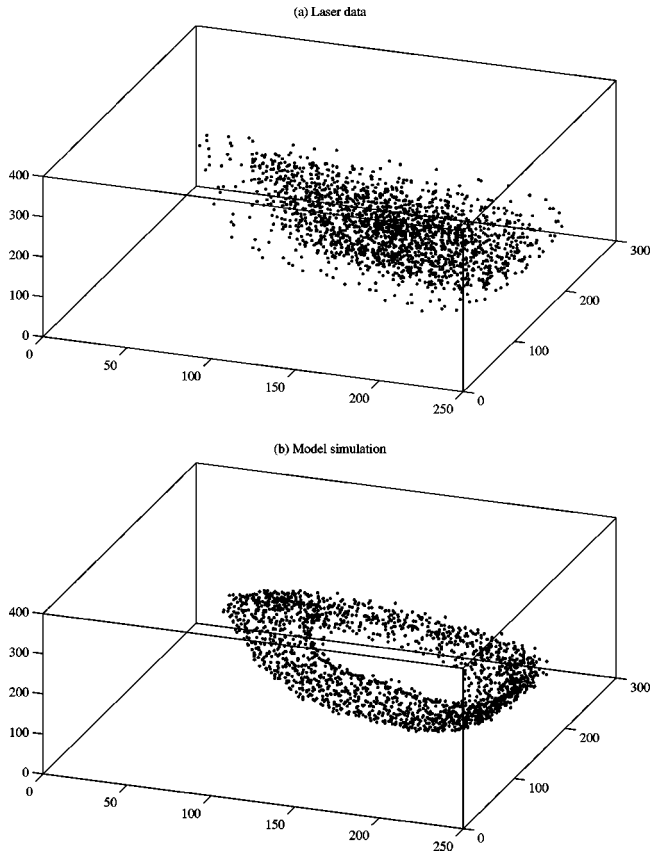

FIG. 5. Attractor reconstructed from experimental data and model simulations for experimental laser data (5). Panel (a) is the time-delay reconstructed attractor (lag=5) from the experimental laser data (2000 points) and panel (b) shows the attractor reconstructed (lag=5) from a simulation (2000 points, initial transient of 2000 points removed) of a radial basis map model of that data. The deterministic trajectory [panel (b)] is clearly bounded and nonperiodic. One can also observe that both attractors occupy the same region of phase space and that the attractor shown in (b) appears to be smoother than that in (a). The model used to compute the trajectory for (b) was built from only 1150 observations of (5). During this 1150 observations the laser underwent only one ''collapse'' [trajectory moving radially from the outside to the center of the attractor in (a)] and this feature is not modeled well. Models built from longer time series capture this feature exactly.

scription length and model fit (root-mean-square prediction error) of all methods proved to be comparable. Even for very small amounts of noise all methods provided smoother estimates of the vector field than by differencing successive (embedded) data values directly. We contend that a more appropriate test of the ''goodness'' of a model is the asymptotic dynamics. For these data this proved to be true. In all cases, simulation from global polynomial models of experimental and artificial data produced equations that, when integrated, were either divergent or singular. This provides further support for our argument that global polynomial models do not extrapolate well. In some further calculations we built global polynomial models from many transients (as opposed to a single trajectory) and have found the dynamic behavior to improve. These simulations exhibit stable foci.

Adams integration/radial basis modeling techniques performed slightly better. Generally these systems also exhibited stable foci; in all cases the data was either chaotic (with some pseudoperiod) or a noisy periodic orbit. Simulations produced from radial basis models of the dynamics as a map produced simulations that appeared most like the data and were most stable. Estimating the vector field by differencing this map gave simulations that appeared to be more similar to the data than Adams integration based methods.

A more rigorous comparison of the results of these different modeling techniques may be obtained by applying nonlinear surrogates data techniques [31] as described in Refs. [27,37]. However, in the cases we considered this proved to be unnecessary. The Adams integration techniques produced simulations that consistently exhibited divergent or singular simulations (in the case of global polynomial methods) and stable foci (in the case of radial basis model methods). This behavior is clearly distinct from the data, and clearly inferior to the results of models based on estimating the map.

The main result of this paper is that a map model produces superior results to models of the vector field in systems with moderate noise, time-delay reconstructions or medium to slow sampling rates. Thus, if one intends to estimate vector fields it may be best to calculate them from a map model of the dynamics. It may be more practical to examine the equivalence between a continuous system and the discrete model of data sampled from it, than to attempt to reconstruct the vector field directly.

ACKNOWLEDGMENT

M.S. is currently funded by the Hong Kong Polytechnic University.

[1] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge Nonlinear Science Series Vol. 7 (Cambridge University Press, Cambridge, UK, 1997).
[2] K. Judd and A. Mees, Physica D **82**, 426 (1995).
[3] M. J. D. Powell, in *Advances in Numerical Analysis, Wavelets, Subdivision Algorithms and Radial Basis Functions*, Vol. II edited by W. Light, (Oxford Science Publications, Oxford 1992), Chap. 3, pp. 105–210.
[4] G. Gouesbet, Phys. Rev. A **43**, 5321 (1991).
[5] G. Gouesbet and J. Maquet, Physica D **58**, 202 (1992).
[6] G. Rowlands and J. Sprott, Physica D **58**, 251 (1992).
[7] G. Gouesbet and C. Letellier, Phys. Rev. E **49**, 4955 (1994).
[8] C. Letellier, L.L. Sceller, E. Maréchal, P. Dutertre, B. Maheu, G. Gouesbet, Z. Fei, and J. Hudson, Phys. Rev. E **51**, 4262 (1995).
[9] A.I. Mees, Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 777 (1991).
[10] A. Irving and T. Dewson, Physica D **102**, 15 (1997).

[11] X. Zhao and V.Z. Marmarelis, Automatica **33**, 81 (1997).
[12] R. Brown, N.F. Rulkov, and E.R. Tracy, Phys. Lett. A **194**, 71 (1994).
[13] R. Brown, N.F. Rulkov, and E.R. Tracy, Phys. Rev. E **49**, 3784 (1994).
[14] L. Fox and F. Mayers, *Computing Methods For Scientists and Engineers*, Monographs on numerical analysis (Clarendon Press, Oxford, 1968).
[15] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, UK, 1988).
[16] N.B. Tufillaro, P. Wyckoff, R. Brown, T. Schreiber, and T. Molteno, Phys. Rev. E **51**, 164 (1995).
[17] J. Rissanen, *Stochastic Complexity in Statistical Inquiry* (World Scientific, Singapore, 1989).
[18] M. Giona, F. Lentini, and V. Cimagalli, Phys. Rev. A **44**, 3496 (1991).
[19] N. Rul'kovl and A. Volkovskii, Int. J. Bifurcation Chaos Appl. Sci. Eng. **2**, 669 (1992).
[20] H. Akaike, IEEE Trans. Autom. Control **19**, 716 (1974).
[21] G. Schwarz, Ann. Stat. **6**, 461 (1978).
[22] F. Takens, Lect. Notes Math. **898**, 366 (1981).
[23] K. Judd and M. Small, Physica D **136**, 31 (2000).
[24] L. Aguirre and S.A. Billings, Physica D **80**, 26 (1995).
[25] S. Chen, S.A. Billings, C.F.N. Cowan, and P.M. Grant, Int. J. Control **52**, 1327 (1990).
[26] E. Cheney, *Introduction to Approximation Theory*, 2nd ed. (American Mathematical Society, Providence, RI, 1982).
[27] M. Small and K. Judd, Physica D **117**, 283 (1998).
[28] H.D. Abarbanel, Z. Gills, C. Liu, and R. Roy, Phys. Rev. A **53**, 440 (1996).
[29] K. Judd and A. Mees, Physica D **92**, 221 (1996).
[30] T. Ikeguchiand K. Aihara, J. Intell. Fuzzy Syst. **5**, 33 (1997).
[31] M. Small and K. Judd, Physica D **120**, 386 (1998).
[32] O. Ménard, C. Letellier, J. Maquet, and G. Gouesbet, Phys. Rev. E **62**, 6325 (2000).
[33] J.P. Crutchfield and B.S. McNamara, Complex Syst. **1**, 417 (1987).
[34] L. Aguirre and S.A. Billings, Int. J. Control **62**, 569 (1995).
[35] B.P. Bezruchko, T.V. Dikanev, and D.A. Smirnov, Phys. Rev. E **64**, 036210 (2001).
[36] C. Letellier, E. Ringuet, J. Maquet, B. Maheu, and G. Gouesbet, Entropie **202/203**, 147 (1997).
[37] M. Small and K. Judd, Int. J. Bifurcation Chaos Appl. Sci. Eng. **8**, 1231 (1998).
[38] In general, the predicted values $g_i(x_n)$ will not be correlated. Hence, the free run predictions will rapidly diverge from the expected system dynamics.
[39] MATLAB's ODE23, ODE45, ODE113, ODE15S, and ODE23S functions.