estimation the data values obtained by applying these techniques to artificial and experimental data from several dynamical systems. Section II compares numerical results of various methods on experimental and artificial time series and explores 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

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

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/
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

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

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

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

where $(x_n)_i \in \mathbb{R}$ is the $i$th component of $x_n \in \mathbb{R}^d$ and

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

Here $\lambda_i$, $\lambda_j$, and $r_j$ are scalar constants, $1 \leq i \leq a$ and $1 \leq j \leq b$ are integers and $c_j$ are arbitrary points in $\mathbb{R}^d$. 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 $\lambda_i$ specified to some accuracy $\delta_i$ is $\ln(\gamma/\delta_i)$ [2,17]. The constant $\gamma$ 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 $(x_n)_{i=1}^N$, and the model with parameters $\Lambda$, is given by

$$L(x,\Lambda) = L(x|\Lambda) + L(\Lambda),$$

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 ($\tau_0 \rightarrow 0$) of a map [23]. The methods in this section
attempt to determine the vector field directly, either by applying a variant of Adams/MDL techniques described by Brown and co-workers \cite{12,13}, or by estimating the Euler derivative from a map model of the form \( \varphi \). 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 Adams 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. \cite{12}. We have coded the algorithm suggested by Brown and co-workers \cite{12} with consistent results. We have been able to accurately reproduce the calculations of Fig. 3 of Ref. \cite{12}. However, our results are only comparable to those in Ref. \cite{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 \cite{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 \( 2^{16} \)) 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.

FIG. 1. Relative prediction error for the Lorenz system (1). The relative prediction error is given by \( x_A = (1/N \sum_{n=1}^{N} (x_{n+1} - x_n + \tau \sum_{i=0}^{m} a_i F(x_{n+1-i})))^2 \) 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 \( x_A \) 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. \cite{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.
the values of $m$ overall decrease in the description length. One then selects are necessary. That is, only those terms that contribute an description length by using $N_p$.

A strict application of the minimum description length principle would involve the following; for fixed this algorithm. A strict application of the minimum description length model 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 $x = f(x)$, then

$$x(t + \tau_0) = x(t) + \tau_0 f(x(t)),$$

so define

$$F(x) = x + \tau_0 f(x).$$

For example, for a small step size $\tau_0$ the evolution over time $\tau_0$ of the Lorenz system may be (approximately) represented in this form. One may take the $n$th 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 $\tau_0$) and use 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 I. Global polynomial model of the Lorenz system (1).

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$ component</td>
<td>$y$ component</td>
<td>$z$ component</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>$xy$</td>
<td>$x^2$</td>
<td>$y$</td>
</tr>
<tr>
<td>$0.0005189113$</td>
<td>$0.00075278227$</td>
<td>$0.310935388$</td>
<td>$0.092927835$</td>
</tr>
<tr>
<td>$-14.809195$</td>
<td>$15.952365$</td>
<td>$0.0012658103$</td>
<td>$x^2z^2$</td>
</tr>
<tr>
<td>$37.857904$</td>
<td>$-0.0074572455$</td>
<td>$-0.0003809924$</td>
<td>$x^3$</td>
</tr>
<tr>
<td>$0.63649195$</td>
<td>$-0.0063938793$</td>
<td>$-1.6396998$</td>
<td>$y$</td>
</tr>
<tr>
<td>$0.01607577$</td>
<td>$0.00079663814$</td>
<td>$-0.0002644879$</td>
<td>$y^4$</td>
</tr>
<tr>
<td>$-0.07983783$</td>
<td>$0.00048246071$</td>
<td>$0.052881899$</td>
<td>$xyz$</td>
</tr>
<tr>
<td>$-7.4128861 \times 10^{-5}$</td>
<td>$-12.144337$</td>
<td>$0.0036643532$</td>
<td>$x^2y$</td>
</tr>
<tr>
<td>$-2.3880938$</td>
<td>$0.0004646419$</td>
<td>$-2.6404106$</td>
<td>$z$</td>
</tr>
<tr>
<td>$-4.2530913 \times 10^{-5}$</td>
<td>$-0.35127338$</td>
<td>$-0.0001475926$</td>
<td>$y^2z^2$</td>
</tr>
<tr>
<td>$-0.0082001822$</td>
<td>$-0.02775479$</td>
<td>$-0.0060493741$</td>
<td>$x^4$</td>
</tr>
<tr>
<td>$-0.037756234$</td>
<td>$-0.03135251$</td>
<td>$2.3421078 \times 10^{-6}$</td>
<td>$z^2x^3$</td>
</tr>
<tr>
<td>$1.0451196$</td>
<td>$0.28592214$</td>
<td>$-0.59313148$</td>
<td>$z^2$</td>
</tr>
<tr>
<td>$-0.8351578$</td>
<td>$0.065101065$</td>
<td>$-0.00024747432$</td>
<td>$y^3$</td>
</tr>
<tr>
<td>$0.0015206632$</td>
<td>$-4.3400433 \times 10^{-6}$</td>
<td>$-0.00066517359$</td>
<td>$xyz^2$</td>
</tr>
<tr>
<td>$-0.012620472$</td>
<td>$5.9716446 \times 10^{-5}$</td>
<td>$x^2z^2$</td>
<td>$x^2y$</td>
</tr>
<tr>
<td>$0.092972835$</td>
<td>$-0.0001289389$</td>
<td>$x^3y$</td>
<td>$z$</td>
</tr>
<tr>
<td>$3.8935607 \times 10^{-5}$</td>
<td>$0.97795022$</td>
<td>$xz$</td>
<td>$y^4$</td>
</tr>
<tr>
<td>$-5.1375924 \times 10^{-5}$</td>
<td>$3.2744309 \times 10^{-5}$</td>
<td>$y^4$</td>
<td>$xyz$</td>
</tr>
<tr>
<td>$0.42323624$</td>
<td>$-0.69856633$</td>
<td>$yz$</td>
<td>$x^2z^2$</td>
</tr>
<tr>
<td>$-3.3274152 \times 10^{-5}$</td>
<td>$x^2z^2$</td>
<td>$xy$</td>
<td>$y^2z^2$</td>
</tr>
</tbody>
</table>
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).

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Term</th>
<th>Coefficient</th>
<th>Term</th>
<th>Coefficient</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5034525 \times 10^{-13}</td>
<td>x z</td>
<td>-1</td>
<td>y</td>
<td>-6.5077874 \times 10^{-14}</td>
<td>z^2</td>
</tr>
<tr>
<td>4.3673276 \times 10^{-13}</td>
<td>y z</td>
<td>1.3248492 \times 10^{-12}</td>
<td>y z</td>
<td>2.3392705 \times 10^{-12}</td>
<td>x^2</td>
</tr>
<tr>
<td>10</td>
<td>y</td>
<td>-1</td>
<td>x z</td>
<td>3.1621093 \times 10^{-11}</td>
<td>1</td>
</tr>
<tr>
<td>-10</td>
<td>x</td>
<td>28</td>
<td>x</td>
<td>1</td>
<td>xy</td>
</tr>
<tr>
<td>8.5557938 \times 10^{-14}</td>
<td>x y</td>
<td>9.474207 \times 10^{-14}</td>
<td>y^2</td>
<td>-6.2292401 \times 10^{-13}</td>
<td>y^2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.4828462 \times 10^{-12}</td>
<td>y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.5964037 \times 10^{-14}</td>
<td>y z</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.6666667</td>
<td>z</td>
</tr>
</tbody>
</table>

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 Weierstrass 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 \( f(x) \) (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

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

where

\[
F(x_n) = (g_1(x_n), g_2(x_n), \ldots, g_d(x_n))
\]

and each \( g_i \) is a function of the form

\[
g_i(x_n) = \lambda_0 + \sum_{j=1}^{n_i} \lambda_j x_{n_i} + \sum_{j=1}^{n_j} \lambda_{j+n+1} \phi \left( \frac{\|x_n - c_j\|}{r_j} \right).
\]

The model that minimizes description length is then selected as the best. In this formulation we fit \( d \), 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 \( i \)th column of \( X \) is the evaluation of the \( i \)th basis function over the data, and the \( j \)th row of \( X \) is the evaluation of all the candidate basis functions at the \( j \)th (vector) datum. The subset selection algorithm discussed in Ref. [2] will select columns \( \mathcal{S} = \{i_1, i_2, \ldots, i_n\} \) 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,

\[
\sum_{k=1}^{n} \lambda_k X_{(i_k \cdot \cdot)}.
\]

Here \( X_{(i_k \cdot \cdot)} \) denotes the \( i_k \)th 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
The parameters $a_0, a_1, a_2, \ldots, a_m$ are the $m$th 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).

\[ A_m = \begin{bmatrix} a_m & a_{m-1} & \cdots & a_0 & 0 & 0 & \cdots & 0 \\ 0 & a_m & a_{m-1} & \cdots & a_0 & 0 & \cdots & 0 \\ 0 & 0 & a_m & \cdots & a_1 & a_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_m & a_{m-1} & \cdots & a_0 \end{bmatrix} \]

(9)

\[ y = \frac{1}{\tau} \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_N - x_{N-1} \end{bmatrix} \]

(10)

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, \( \tau = 0.05, N = 4000, d_e = 3.4, \text{lag} = 7\)) with and without observational noise \([N(0, \sigma^2), \sigma^2 \approx 0.05(\text{standard deviation data})]\).

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

\[
\dot{x} = y,
\]
\[
\dot{y} = -x - \delta y + z,
\]
\[
\dot{z} = \gamma [h(x) - z] - \sigma y,
\]
\[
h(x) = \begin{cases} 
0.528 \alpha, & x < -1.2, \\
 x(1 - x^2) \alpha, & -1.2 \leq x < 1.2, \\
0.528 \alpha, & x \geq 1.2,
\end{cases}
\]

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

For the simulations described in this paper we used the following parameter values: \(C_1 = C_2 = 375 \text{ nF}, \ L = 233.7 \text{ mH}, \ r = 0.407 \text{ k\Omega}, \) and \(R = 6 \text{ k\Omega}.\) 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 \((a = 0.411, \ b = 2, \ c = 4, \ y \text{ component}, \ \tau = 0.1, \ N = 4000, \ d_e = 4, \ \text{lag} = 9)\) 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 \([28] \) \((\tau = 800 \text{ ns}, \ N = 1150, \ d_e = 5, \ \text{lag} = 5),\) see Fig. 4.

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

(7) Experimental data from Japanese vowel \((u)\) sounds \([30] \) \((\tau = 104.167 \mu\text{s}, \ N = 3200, \ d_e = 4, \ \text{lag} = 10).\)
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. 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, and the attractor provided by a simulation from a radial basis model map.

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 (◇), radial basis map predictions as stars (●), radial basis flow (Euler derivative of the map) as a dotted line (···), the Adams integration/global polynomial scheme as a dashed line (---), and Adams integration/radial basis model as a dot-dashed line (----). 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.
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 the error is in terms of \( \| x_{n+1} - (x_n + \Delta F(x_n)) \| \). 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).

<table>
<thead>
<tr>
<th>System</th>
<th>Observational noise</th>
<th>Dynamic</th>
<th>( \tau )</th>
<th>( d_r )</th>
<th>Lag</th>
<th>Modeling error</th>
<th>Minimum description length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Lorenz</td>
<td>0.1 (1.606) (^a)</td>
<td>0</td>
<td>0.05</td>
<td>NA</td>
<td>NA</td>
<td>2.0766 (^a)</td>
<td>20.450 2.1859 24051 59941 59644</td>
</tr>
<tr>
<td>(2) Lorenz</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>3</td>
<td>7</td>
<td>0.56</td>
<td>1.05 1.01 4261 54710 54045</td>
</tr>
<tr>
<td>(3) circuit</td>
<td>0.05 (0.40)</td>
<td>0</td>
<td>0.05</td>
<td>3</td>
<td>7</td>
<td>1.015</td>
<td>1.22 1.155 6331 75202 75127</td>
</tr>
<tr>
<td>(4) circuit</td>
<td>0.05 (0.03)</td>
<td>0</td>
<td>0.1</td>
<td>4</td>
<td>9</td>
<td>0.01</td>
<td>0.01144 0.01035 (-1231) (-1151) (-988)</td>
</tr>
<tr>
<td>(4) Rössler</td>
<td>0</td>
<td>0.025 (0.055)</td>
<td>0.2</td>
<td>4</td>
<td>9</td>
<td>0.0556</td>
<td>0.04592 0.04662 (-5843)</td>
</tr>
<tr>
<td>(5) Laser</td>
<td>NA</td>
<td>0</td>
<td>8 \times 10^{-7}</td>
<td>5</td>
<td>5</td>
<td>1.619</td>
<td>1.3426 1.6374 4304 127440 129150</td>
</tr>
<tr>
<td>(6) String</td>
<td>NA</td>
<td>NA</td>
<td>3.73 \times 10^{-3}</td>
<td>4</td>
<td>12</td>
<td>108</td>
<td>30.929 28.556 20869 136247 138085</td>
</tr>
<tr>
<td>(7) Vowel</td>
<td>NA</td>
<td>NA</td>
<td>1.04 \times 10^{-4}</td>
<td>4</td>
<td>10</td>
<td>581</td>
<td>474.683 478.671 25028 215098 215228</td>
</tr>
</tbody>
</table>

\(^a\)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).

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 Weierstrass 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-
description 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.

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.

MATLAB’s ODE23, ODE45, ODE113, ODE15S, and ODE23S functions.