ELSEVIER

Physica D 132 (1999) 133–149

PHYSICA D

www.elsevier.com/locate/physd

# Models of knowing and the investigation of dynamical systems

P.E. Rapp [a,b,*], T.I. Schmah [c], A.I. Mees [d]

[a] *Department of Physiology, The Medical College of Pennsylvania Hahnemann University, 2900 Queen Lane, Philadelphia, PA 19121, USA*
[b] *The Arthur P. Noyes Clinical Research Center, Norristown State Hospital, Norristown, PA 19401, USA*
[c] *Département de Mathematiques, Ecole Polytechnique, Federale de Lausanne, Switzerland*
[d] *Centre for Applied Dynamics and Optimization, The University of Western Australia, Nedlands, Perth, WA 6907, Australia*

Received 15 August 1998; received in revised form 10 February 1999; accepted 12 February 1999
Communicated by A.M. Albano

**Abstract**

We present three distinct concepts of what constitutes a scientific understanding of a dynamical system. The development of each of these paradigms has resulted in a significant expansion in the kind of system that can be investigated. In particular, the recently-developed 'algorithmic modelling paradigm' has allowed us to enlarge the domain of discourse to include complex real-world processes that cannot necessarily be described by conventional differential equations. 

*Keywords:* Dynamical systems; Algorithmic modelling; Embedding

## 1. Introduction

What do we mean when we say that we understand a dynamical system? In this essay, we identify three distinct paradigms for scientific understanding of dynamical systems. These paradigms are the *models of knowing* of the title. The introduction of new models of knowing has resulted in a significant expansion in the kinds of systems that can be investigated scientifically.

The first paradigm, which we shall refer to as the Newtonian [1], was established in the seventeenth century. According to this approach, a dynamical system is understood by modeling it with a differential equation and then solving that equation.

The second paradigm, the geometric theory of differential equations (also known as the qualitative theory), emerged at the beginning of the twentieth century, largely as a result of the work of Poincaré. It is firmly grounded in the Newtonian paradigm, in that it begins with differential equations, but it endeavors to provide qualitative information about system behavior even when formulas to solve the differential equations are not available.

* Corresponding author. Tel.: +1-610-3131156; fax: +1-610-3135753; e-mail: rapp@mcphu.edu.

[1] We call this the Newtonian model without prejudice as to what Newton's world view may actually have been. It might be argued that 'Laplacian' is a more appropriate term.

0167-2789/99/$ – see front matter 
PII: S0167-2789(99)00035-4

The third and most recent is what we call the algorithmic modeling paradigm. In this approach, there is a strong emphasis on understanding and automating the process of modeling, usually with the aid of computers. The dynamical models used are usually maps (discrete-time dynamical systems) rather than differential equations, and often are best described by algorithms rather than conventional formulas. This paradigm is described in greater detail later. The main purpose of this essay is to describe the algorithmic modeling paradigm and put it in historical context.

In our historical discussions, we will restrict attention to ordinary differential equations, ignoring, for example, partial differential equations, difference equations, and stochastic models. We do this for clarity, and because of the historical centrality of ordinary differential equations, and also because this is the case with which we are most familiar. Our main points apply to other kinds of dynamical systems as well [2].

## 2. The Newtonian paradigm

In the Newtonian paradigm, change is characterized by a differential equation. The two basic steps to understanding a dynamical system are writing down the differential equation and solving it. The Newtonian approach concentrates more on solving differential equations than on how they are obtained in the first place; we shall return to this matter in Section 4, but for now, we assume the equation is given.

A simple example of the Newtonian approach is the analysis of harmonic motion [3]. Suppose that a weight of mass $m$ is connected to an ideal spring with a modulus of elasticity $K$. We let $x$ denote the position of the weight at time $t$. Position and time are related by the differential equation

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -Kx.$$

We can solve this equation for any initial conditions. For example, if the weight is motionless and at position $x_0$ at time $t = 0$, then

$$x(t) = x_0 \cos\sqrt{\frac{K}{m}}t.$$

The position of the weight, $x(t)$, is completely specified for any value of time $t$. The Newtonian ideal for dynamical analysis is achieved because we know both the differential equation and a formula giving its solution.

During the century following Newton's death in 1727, analytical mechanics and the theory of differential equations (which remained very closely linked fields) advanced on two fronts. First, more general procedures for constructing differential equations that describe dynamical systems were established. For example, Joseph-Louis Lagrange (1737–1813), who was particularly proud of the fact that his treatise on mechanics did not contain a single diagram, showed how the calculus of variations could be used to determine the differential equations describing mechanical systems. Second, procedures for solving differential equations were extended. There were many successes; but by the early 19th century, it was becoming clear that there were serious difficulties, both theoretical and practical,

[2] The most general mathematical definition of a deterministic dynamical system is that it is a flow on some space, in other words a function from $\Phi : T \times M \to M$ for some 'time' space $T$ (an interval in $\mathbb{R}$ or $\mathbb{Z}$) and 'phase' space $M$. For example, for the flow of an ordinary differential equation, $T$ is an interval in $\mathbb{R}$ and $M$ is usually a smooth manifold; for the solution to a difference equation on $\mathbb{R}$ of order $k$, we have $T = \mathbb{Z}$ and $M = \mathbb{R}^k$. In *stochastic* dynamical systems, $\Phi(t, x)$ is a probability distribution on $M$.

[3] The fact that our example deals with a linear system is not particularly relevant here; we have chosen it merely for simplicity.

in solving differential equations [4]. To appreciate these difficulties, we need to be clearer about what we mean by 'solving'. We now outline different kinds of solutions that can be found. We will only consider ordinary (finite-dimensional) differential equations in this essay. Some of what we say is applicable to partial (infinite-dimensional) differential equations, but the situation there is even more complex.

### 2.1. Closed-form solutions

Traditionally, solving a differential equation has meant writing down a formula which satisfies the equation. But what exactly *is* a formula, or if you prefer, what *kind* of formulas are we interested in?

A common answer is: a finite (preferably short) expression built out of rational functions, radicals ($n$th roots), exponentials, logarithmic and trigonometric functions. The solution in the above example is of this form, as are almost all functions in introductory calculus books. The list of 'building-block' functions is commonly expanded to include all algebraic functions, $f(t)$, which are those that are solutions to algebraic equations like

$$a_n(t)x^n + a_{n-1}(t)x^{n-1} + \cdots + a_0(t) = 0$$

where the $a_i$ are polynomials. Functions built out of rational functions and radicals are algebraic, but there are other algebraic functions that can only be defined implicitly by an equation like that above. A function that can be built from algebraic, exponential, logarithmic and trigonometric functions, in a finite number of steps, is called an *elementary* function. An elementary function that is a solution to a differential equation is called an *elementary solution.*

Unfortunately, some elementary functions have integrals which are not elementary [5]. This was proved by Neils Henrik Abel (1802–1829) and Joseph Liouville (1809–1882), and it had been suspected earlier, for example, by Marie-Jean Marquis de Condorcet (1743–1794) and Pierre-Simon de Laplace (1749-1827) [3]. A simple example (and the one Liouville seems to have started with) is the elliptic integral

$$\int \frac{\mathrm{d}t}{\sqrt{1+t^4}}.$$

The integral exists, but is not elementary. It is the solution to the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{1}{\sqrt{1+t^4}}.$$

This differential equation is one of the simplest possible, so it is clearly hopeless to restrict attention to equations with elementary solutions. It has long been known [4] that functions whose integrals can be expressed in terms of elementary functions are extremely rare.

Even if an integral of an elementary function is not elementary, it can be manipulated mathematically, and its values approximated, using well-understood techniques. Thus it is advantageous to be able to express solutions to differential equations as formulas containing integrals. For example, the solution to

$$\frac{\mathrm{d}x}{\mathrm{d}t} + \mathrm{e}^{t^2}x = 0$$

[4] There is also the important theoretical question, do solutions even exist? Powerful existence and uniqueness theorems have been proved for large classes of differential equations [1], but these are beyond the scope of this essay. We focus here on the process of finding solutions where they are known, or assumed, to exist.

[5] Relatively recently, a class of functions slightly larger than the traditional elementary functions has been defined which is closed under integration [2]. This is of use in computer algebra but does not significantly affect the themes of this paper.

can be written [5] as

$$x(t) = A\,\mathrm{e}^{(-\int \mathrm{e}^{s^2}\,\mathrm{d}s)}.$$

Expressing a solution as a formula containing integrals of elementary functions is called *integration by quadrature*, or *integration in finite terms*, and is considered to 'solve' the equation. The result is called a *solution in finite terms*, or more commonly, a *closed-form* solution. More precisely, we may define the set of closed form functions to be the closure of the set of algebraic functions under the operation of integration (thus allowing integrals of integrals, etc.).

Liouville showed that some differential equations cannot be solved even using this expanded definition of what a solution is. For example,

$$\frac{\mathrm{d}x}{\mathrm{d}t} + x^2 = t$$

is not solvable by quadrature; in fact, no particular solution to this equation can be expressed in closed form [6]. In other words, though this equation has a solution, that solution cannot be expressed as a formula involving elementary functions and integration.

Of greater practical significance is the fact that at present we have no satisfactorily general procedure for finding closed-form solutions when they exist. Texts on differential equations are filled with a variety of special-purpose methods [5]. These methods have been quite successful in solving the simple differential equations of classical physics, which examples have largely driven the development of the theory. But even classical physics quickly gives rise to differential equations whose general solutions are beyond the reach of these methods, for example, the famous three-body problem [7]; and problems from other fields, such as biology and economics, are generally much more complicated.

In summary, the traditional goal of finding a closed-form solution to a differential equation is usually very difficult to achieve, and sometimes impossible.

### *2.2. Series solutions*

When a closed-form solution is not available, a solution in the form of a power series can often be found. For example, to solve

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} - 2t\frac{\mathrm{d}x}{\mathrm{d}t} - 2x = 0,$$

we start by writing

$$x(t) = a_0 + a_1t + a_2t^2 + \cdots = \sum_{n=0}^{\infty} a_n t^n.$$

We substitute this expression into the original equation and simplify, to get the recursive relationships

$$a_{n+2} = \frac{2a_n}{n+2}.$$

The coefficients $a_0$ and $a_1$ are determined by the initial conditions. For example, for $a_0 = 1$ and $a_1 = 0$, one can show that

$$x(t) = 1 + t^2 + \frac{t^4}{2!} + \frac{t^6}{3!} + \cdots = \mathrm{e}^{t^2}$$

is a solution to the original equation [5]. Newton was aware of this kind of solution. Unfortunately, sometimes the series constructed by such a method does not converge. Sometimes it converges but one cannot find a formula for the coefficients, though one may be able to find a recursive algorithm for determining them. If the power series converges and one can find a formula for the coefficients of the series, the solution is called an analytic solution or a series solution. These are considered a good second-best to closed-form solutions.

Even if a series solution can be found, it may not prove very helpful. For example, contrary to what is often supposed, the three-body problem of classical mechanics, which motivated the research of Poincaré that we shall discuss later, has in fact been solved analytically. Karl Sundman published a series solution for almost all initial values in a series of papers beginning in 1907 [8–10]. In 1991, Quidong Wang accomplished the same feat for the general $n$-body problem [11]. The solution consists of a very slowly converging infinite series which has not been very helpful for generating insights and understanding, and furthermore, is essentially useless for numerical approximation of solutions. The history of the series solution of the $n$-body problem is reviewed briefly in [12].

### *2.3. Numerical methods*

There are efficient numerical methods for approximating the solution to nearly any differential equation, given the initial conditions. One popular such algorithm is the Runge–Kutta method [13,14]. Numerical methods of solution were at one time considered much less satisfying than formulas because they provide less insight into the nature of the general solution. This view has changed, however, with the advent of powerful computers, and particularly of graphics workstations. Often a good numerical approximation can show qualitative features of the solution that cannot be seen by other methods, or can be seen only after a great deal of effort.

### *2.4. Summary*

We have described a variety of approaches used to solve differential equations. For a more detailed historical outline, see [15] or [16]. Much progress has been made in this field, and much interesting mathematics has been created in the process. Nevertheless, there remain great practical difficulties in solving differential equations, and in some cases, the conventionally most satisfying types of solution can be shown not to exist. Other less restrictive definitions of 'solution' have been invented, which enlarge the domain of solvable equations, but these kinds of solution, even if they can be found, are not always as informative.

It is clear that the Newtonian paradigm, while of immense historical, mathematical and scientific importance, has serious limitations, both theoretical and practical. A quite different approach is outlined in the next section.

## 3. The qualitative theory of differential equations

Much can be learned about the solutions to a differential equation without finding formulas for them. For example, consider the equation $(\mathrm{d}x/\mathrm{d}t) = f(x)$. We can find equilibrium points by finding zeros $\hat{x}$ of $f$. We can often determine their stability by examining the eigenvalues of the derivative $\mathrm{D}f(\hat{x})$; when this fails, we can still often determine stability by looking at the effects of nonlinear terms. We can ask other qualitative questions about the solutions. For example, are there periodic solutions? Are they stable? How does the system respond to parameter changes? In the Newtonian paradigm, questions like these are addressed by attempting to solve the equation completely and then analyze the solution. However, as we have seen, closed-form solutions may not exist, and series solutions, when they can be found, may be difficult to analyze and understand. So it is useful to be able to answer qualitative questions by using differential equations directly, without requiring a formula for the solution.

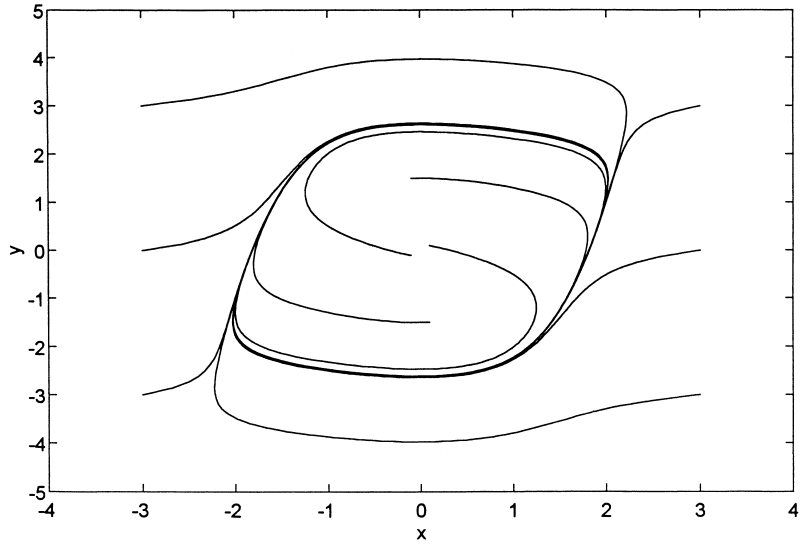


Fig. 1. Numerical solution of the Van der Pol equation for the specific case of $\mu = 2$ and several initial conditions.

The first person to carry out a major investigation along these lines was Henri Poincaré (1854–1912). The study of dynamical systems expanded greatly with his publication of four major papers from 1881 to 1886 in which he emphasized the geometrical and topological nature of differential equations [17–20] (reviewed in [21]). This emphasis was so successful that the qualitative theory of differential equations is often referred to as the geometrical theory.

### *3.1. The geometrical viewpoint*

Any ordinary differential equation of order greater than one can be reduced to a system of first order equations on a higher-dimensional space. As an example of this *reduction of order*, consider the Van der Pol equation, a differential equation that describes an electronic circuit. The original form of the equation is

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} - \mu(1 - y^2)\frac{\mathrm{d}y}{\mathrm{d}t} + y = 0.$$

By defining the new state to be $\boldsymbol{x} = (y, \dot{y})$, we see that original equation is equivalent to

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = x_2, \qquad \frac{\mathrm{d}x_2}{\mathrm{d}t} = \mu(1 - x_1^2)x_2 - x_1.$$

The two-dimensional $\boldsymbol{x}$ plane is called the *phase space*. The system of first-order equations defines a 'velocity' vector at each point in phase space, the collection of which is called a *vector field*. This is just another way of thinking of the differential equations, but in a geometric setting. A particular solution to the differential equation is a parameterized curve in phase space, called a *trajectory*, that is always a tangent to the vector field given by the equation. Fig. 1 shows a few numerically-calculated trajectories for the Van der Pol example.

The set of all trajectories is called the *flow*; it corresponds to the general solution of the equation. Poincaré's program was to study the general properties of the flow so that it was possible to describe the behavior of infinitely many trajectories in qualitative terms. In some cases, as in Fig. 1, trajectories converge onto periodic solutions. In other cases, trajectories originating from a large region of phase space converge onto a single fixed point. Convergence onto more complex structures that are neither fixed points nor periodic solutions can also occur, and

the realization that such structures exist and may be common is the motivation for the study of chaotic dynamics. During the past century many new methods have been developed to extend our understanding of the qualitative properties of differential equations, including the classification of fixed points, the method of isoclines, stability analysis, bifurcation theory, various indices, and existence theorems for structures including periodic orbits and strange attractors [22–27].

### 3.2. The goals of dynamical systems analysis

Instead of deriving a formula for the solution, the Poincaré program directly addresses questions that a formula could be hoped to answer. Having a formula for the solution mainly allows us to do two things: accurately compute a particular solution for a given initial condition; and make qualitative statements about the system's behavior (We might call these goals prediction and insight, though the two cannot be separated quite so neatly.). When a formula for the solution is not available, the first goal can usually still be met by computing a particular solution with a numerical algorithm, as mentioned in the previous section. The second goal is addressed by the qualitative geometrical theory. Thus the development of qualitative methods allows the goals of dynamical analysis to be met without ever giving a formula for the solution to the differential equation.

The development of the qualitative theory resulted in an enormous expansion of the range of systems that could be investigated. This is particularly true for the analysis of biological systems: much of the present activity in mathematical biology followed from the introduction of geometrical methods in the 1970s [28]. Most differential equations studied today by scientists do not have known analytical solutions, so they are studied with qualitative methods.

### 3.3. Difficulties in finding closed-form models

Though the qualitative theory does not require a formula for the *solution* to an equation, it does, as originally conceived, require a formula for the differential equation itself; that is, the *vector field* in phase space must be expressed in closed form. There are some theoretical and practical difficulties with finding systems of closed-form differential equations to model phenomena of interest.

To start with, there exist smooth vector fields that cannot be described explicitly in closed-form. For example, we know from Section 3.1, that there are smooth functions that cannot be expressed in closed form and we can take such a function as a vector field (i.e. the right hand side of a differential equation).

We could always expand our notion of closed-form functions to include functions particularly appropriate to the problems in which we are interested; this is a classical approach, which has lead to a variety of useful 'special functions'. But this is clearly an endless task, and an approach that quickly loses usefulness as the bag of 'special functions' grows.

It might be thought that, from a practical viewpoint, one is unlikely to encounter a vector field that cannot be described in closed form. But we will suggest a way in which one might appear in the investigation of a system defined in closed form. Suppose that, contrary to what is expected, we have somehow obtained an 'equation of the universe', or at least a differential equation that describes a significantly large system as accurately as we could wish. In general, the system's phase space will have very large dimension, and we have little hope of solving the equations in any of the senses described earlier: analytically, numerically, or qualitatively. To have any hope of gaining insight or performing calculations, we must simplify. This means extracting some dynamical or physical subsystem. By a dynamical subsystem, we mean an attractor of lower dimension than that of the phase space, preferably much lower. A biological example is provided by Michaelis–Menten enzyme kinetics [29] where the dimension of the original dynamical model is reduced by assuming that the concentration of the enzyme-substrate complex is always at its

equilibrium value. As a simpler example that occurs even in linear systems, when analyzing an electronic circuit we can usually ignore stray capacitances and inductances: it is easy to show that they correspond to fast motions onto a lower-dimensional manifold on which the conventional circuit equations are defined. This concept of a so-called inertial manifold [6] can be seen as central to application of the theory of differential equations to modeling the world. It would be surprising for all inertial manifolds to have closed-form descriptions, and even more so, for the restriction of the differential equations to the inertial manifold to have a closed-form description [7].

The existence of smooth vector fields that cannot be expressed in closed form need not worry us too much, as long as we are content with *approximations*. For example, the Stone–Weierstrass Theorem [31] tells us that, on a compact subset of $\mathbb{R}^n$, we can uniformly approximate any continuous vector field to any desired accuracy with polynomials. Unfortunately, unless our vector field is already a polynomial, we will have to accept that, the better the desired accuracy, the larger the formula.

The same kind of problem arises in approximation by *any* class of functions. Consider some 'nice' class of functions, for example, polynomials, elementary functions, or those that can be expressed in closed form, or maybe a larger group, including some special functions; and assume that we have some agreed system for writing them down. If we define a 'manageable formula' to be one which is at most, say, 6 ft long, then nice functions with manageable formulas comprise a negligible subset of the set of all continuous (or smooth) vector fields. In fact, it is easy to show that vector fields without manageable formulas are open dense in the set of, say, all continuous vector fields on $R^k$. This remains true even if we allow infinite-precision parameters [8]. This means that, not only will we have to be content with approximations, but there exist systems that we cannot approximate as accurately as we would want with formulas small enough to be useful to us.

Recognition of the problem of vector fields without manageable formulas is by no means new. We can do no better than quote one of the masters:

> *Insofar as the weather may be due to an unlimited assemblage of local instabilities, it may not be amenable to a finite scheme of law at all.*
>
> (James Clerk Maxwell, Matter and Motion, 1877.)

There are also practical difficulties in finding a good formula to model a system of interest, even supposing that such a formula exists. In the early days of the mathematical study of dynamical systems, the systems under study were generally from physics, and they were comparatively simple: it may have taken much insight and cleverness to model a situation, but the model could usually be developed from simple physical laws. Today, we are more often in the position of modeling an incompletely understood and very complex system from, for example, biology, economics, meteorology or the social sciences. While we may know, or have theories about, some of the underlying laws, we must depend in part on observing features of the system that do not yet have a theoretical basis. We search for models that approximate experimental data well, while being consistent with available theories. This is a much different and more complex kind of modeling process than Newton and his immediate followers faced.

---

[6] An invariant and exponentially attracting manifold is called an inertial manifold.

[7] To investigate this further, we would need to expand our definition of *closed-form* to manifolds. One obvious way to do this is to require that the manifold be covered by local coordinate charts, in each of which that vector field can be expressed in closed-form, using our previous definition. But this definition is too flexible. Indeed, at any non-equilibrium point, it is possible to find some coordinate system in which the vector field is trivial (see the 'straightening out theorem' in [30]). But that coordinate system may be very hard to find. It may also be difficult to work with, for example it may not be expressible in closed form with respect to the original coordinate axes. It would be more interesting to define some class of 'natural' coordinate systems for a submanifold with respect to the space in which it is embedded; but the choice of an appropriate class might depend strongly on the scientific context.

[8] Assuming a finite set of typographical symbols, there are only a finite number of formulas shorter than a given maximum length. Even if we allow arbitrary real-valued parameters, the set of manageable formulas is still finite-dimensional. The set of continuous vector fields on $R^k$ is an infinite-dimensional vector space, so the complement of the set of functions with manageable formulas in the set of continuous vector fields is open dense.

Nonetheless, most mathematical modeling today still uses conventional differential equations, derived more from theories of the underlying system than directly from experimental data. It is accepted that it is very rare to be able to do system *identification*, that is, to get a good fit between the model and the real-world system, but it is argued that the model in some way captures the essence of the system's behavior. In the hands of gifted practitioners, this does indeed seem to be the case, as long as the domain of discourse is suitably limited, but even in physics and engineering there are notoriously many systems which have not been satisfactorily modeled in this way, and outside these areas the majority of systems do not seem to yield to this approach. There has been some work on systematizing the process of differential equation modeling [32], but much remains to be done.

When faced with an extremely complex system, a common approach is to focus attention on isolated subsystems or otherwise simplified systems. This allows the introduction of rigor into the analysis, and brings the hope that models of simplified systems can shed light on the behavior of the original system, and perhaps provide enough insight to allow one to build more comprehensive models. But there is also the risk of becoming entangled in the details of the reduced systems and losing sight of the original compelling problem. While studying simplified systems it is often very useful, and would clearly also be desirable to study the original dynamical system in some way that does not require one to first describe it with a differential equation.

To summarize, there are serious theoretical and practical difficulties in modeling the world with conventional differential equations. We must at least think more carefully about the process of modeling complex systems from experimental data. The remainder of this essay discusses the modeling process, and also explores the use of models other than conventional differential equations. We claim that the new approach gives us the ability to model a far larger class of systems.

## 4. Algorithmic modeling

This section describes a new paradigm for understanding dynamical systems, developed largely in the last two decades and still immature, with roots in statistics, signal processing and computer science as well as traditional dynamical systems theory. The focus is on understanding the process of producing models from experimental data. We accept the arguments made above and do not necessarily assume that there *exist* simple formulas to accurately describe our data. Our response is two-fold: allow more complex and non-traditional models, and accept that we are looking for approximations. The types of models described here are quite varied, ranging from conventional formulas to complex models most easily described by algorithms such as computer programs. We will call models of the latter kind *algorithmic* models, without defining this term precisely. The approach we describe here is often called *reconstruction*. We are calling it the *algorithmic modeling* paradigm, emphasizing the importance of modeling process, and the algorithmic nature of many of the models produced. A number of investigators have contributed to the development of the concepts presented in this section. Notable in this regard is the work of Abarbanel [33]; the account on pages 112–114 of his monograph is particularly pertinent.

In general, algorithmic modeling begins with experimental data, possibly augmented by existing knowledge or theories (for example, symmetries, boundaries or conservation laws). We will concentrate on the experimental data. We will touch on the important problem of incorporating theoretical knowledge, but much less has been done in this area.

We will assume that the experimental data consists of, say, $k$ real variables sampled at regular time intervals, in other words, $x_1, \ldots, x_N$, where each $x_i$ is a $k$-vector. Discrete variables are often encountered, but we will concentrate on real variables for simplicity. Sometimes data is gathered at irregular time intervals, which poses more difficulties, but we will not consider this here. A *model* for this data consists of a dynamical system and a relationship between its phase space and the observed variables.

We will concentrate mostly on deterministic models with finite-dimensional manifolds as phase spaces. Note that it is possible that our system is inherently stochastic or infinite-dimensional, or too complex to be accurately represented by any model small enough for our computers to handle [9]. Nonetheless, we will begin by *assuming* our system is deterministic, finite-dimensional, and relatively simple, remembering that it will be important to question these assumptions if our resulting model is not very good.

In the real world, most dynamical systems change state continuously (at least, as far as we can observe), so it makes sense to consider continuous-time models, such as conventional differential equations. But since the data are sampled at regular time intervals, it is also reasonable to consider *discrete-time* models. These are simply functions $\psi : A \to A$ from the phase space to itself [10]; for any initial state $x$, the state of the system after $t$ time steps is $\psi^t(x)$ (the $t$th iterate of $\psi$). This is the approach that has received most attention in the development of algorithmic modeling.

One of the first problems we encounter is the choice of phase space for our model.

### 4.1. Constructing a phase space

We want to model the data with a map (a continuous function) $\psi : M \to M$ for some phase space $M$. If we are beginning with experimental data and little theoretical understanding, we probably do not know a priori what the phase space of our model should be, and how it relates to the observable variables. If there are many observable variables and we have reason to suppose that the system is quite simple, we might assume that the values of the observables completely determine the state of the system. In that case, if we have $N$ real variables, our phase space can be $\mathbb{R}^N$, or some subset of $\mathbb{R}^N$, and the observables can be just the standard coordinates of $\mathbb{R}^N$. If the system is complex and we have only one observable variable or only a few, the situation is less clear.

Here is a commonly used procedure for constructing a phase space from observations of a single real variable (A similar procedure can be used with multiple variables.). Beginning with data $x_1, \ldots, x_N$, choose a phase space dimension $m$ (more about this later) and construct the following points in $\mathbb{R}^m$,

$$z_j = (x_j, x_{j+1}, \ldots, x_{j+m-1}),$$

for all applicable $j$, that is $1 \leq j \leq N-m+1$. We take as our phase space some subset of $\mathbb{R}^m$ containing the trajectory $z_1, z_2, \ldots, z_{N-m+1}$. This construction is called an *$m$-dimensional embedding*. Since the word 'embedding' already has a standard definition in topology, we suggest Sauer's term *delay embedding* [34].

This simple procedure is justified by the Takens–Mañé embedding theorem [34–37], which we will now state. Assume that the true dynamical system consists of iterations of a diffeomorphism [11] $\psi : M \to M$ on a compact phase space $M$ of dimension $d$, and that the observations $x_i$ are related to the true states $y_i$ of the system by a smooth map $c : M \to R$ such that $c(y_i) = x_i$ for all $i$. For any integer $m > 2d$, define the *delay coordinate map* $\Phi : M \to R^m$ by

$$\Phi(y) = (c(y), c(\psi(y)), c(\psi^2(y)), \ldots, c(\psi^{m-1}(y)).$$

[9] Distinguishing between stochastic systems and extremely complex deterministic ones is impossible in practice. Whether there really exist two different categories, stochastic and deterministic, is an interesting philosophical question, but not very relevant to the present discussion.

[10] This is the definition of an *autonomous* discrete-time dynamical system. A non-autonomous system is a function $\psi : \mathbb{Z} \times A \to A$, in other words the dynamics are time-dependent. But the distinction is academic, as we can always consider the new phase space $\mathbb{Z} \times A$, with autonomous dynamic $\Psi(t, x) = (t+1, \psi(t, x))$.

[11] A *diffeomorphism* is a differentiable function with a differentiable inverse.

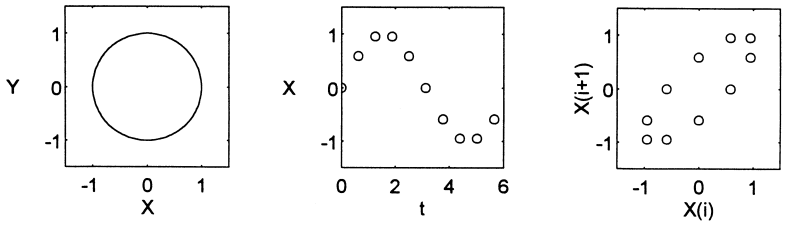


Fig. 2. Reconstruction of a circular phase space from data. Consider a constant-speed motion around a circle. (a) The true phase space; (b) observations at regular time intervals; (c) a reconstruction of the phase space.

The Takens–Mañé theorem says that, for almost every $\psi$ and $c$, the map $\Phi$ is an embedding, in other words, $M$ is diffeomorphic to its image [12]. This means that if we had an infinite amount of data, from states that formed a dense subset of the phase space, and we chose the dimension of the delay embedding to be greater than $2d$, then the closure of the set of embedded points $\{z_i\}$ would (with probability one) be diffeomorphic to the phase space. (We call this the *reconstructed phase space*.) Further, the continuous extension of the map $z_i \mapsto z_{i+1}$ would correspond (under the diffeomorphism) to the original map $\psi$. Fig. 2 shows how the delay embedding method could be used in practice to reconstruct a circular phase space from observations of a single variable [13,14].

While the ideal of infinite amounts of noise-free data from a set of states dense in the phase space can never be met in practice, the procedure can give useful information about the topological dynamics of the system that generated the data.

There are many practical details to be considered in applying the delay embedding method. One important issue is that the embedding theorem requires that the dimension $m$ of the delay embedding satisfy $m > 2d$, where $d$ is the dimension of the true phase space, but of course, in the experimental setting, we do not know $d$ a priori. Simply choosing huge $m$ values is unsatisfactory, because that leads to a very sparsely populated phase space. Another issue is the choice of the sampling interval, that is, the time interval between observations. (For example, note that the reconstructed phase space in Fig. 2(c) is elliptical; for some sampling intervals, the reconstruction gives an ellipse so elongated that the data points appear to be on a line segment along the diagonal.) Criteria for making these selections are reviewed in several papers [39,40] and in Abarbanel's monograph [33].

With appropriate selection of embedding dimension and sampling interval, the delay embedding method can be a very useful starting point in the modeling process. It is used in most applications of the methods we will be discussing. Indeed, the delay embedding by itself, perhaps followed by a singular value decomposition to identify the dimensions with highest variance, can give useful pictures of the dynamics. Some researchers have used the delay-embedded trajectory to directly assess evidence for a low-dimensional model, without making an explicit model [41].

---

[12] Our statement of the theorem incorporates extensions given in [34,37]. In particular, the original theorem proved that $\Phi$ is an embedding for all $\psi$ and $c$ in an open dense set. The proof for *almost every* $\psi$ and $c$, for an appropriate extension of the idea of Lebesgue measure to spaces of smooth maps, is due to Sauer, Yorke and Casdagli [34].

[13] The central idea of the delay embedding method is that the state of the system is largely determined by the previous $m$ observations, an idea which appears, for example, in the AR method in statistics. But AR models are always linear: one could describe them in the present context as delay embedding followed by linear regression. In algorithmic modeling, the delay embedding can be followed by *any* modeling technique. Moreover, recent work in embedding [38] has tended to move away from the idea of a global embedding, using notions which only make sense in the nonlinear context.

[14] Note the similarity between the delay embedding technique and the reduction of order of a differential equation, discussed in Section 3. Given an $n$th order differential equation on $\mathbb{R}$, we can reduce the system to a first order equation on $\mathbb{R}^n$, which we call phase space. The coordinates of $\mathbb{R}^n$ correspond to the first $n$ derivatives of the single coordinate, call it $x$, of the original space $\mathbb{R}$. The first $n$ derivatives of $x$ can be estimated from $n+1$ data points, $x_i, \ldots, x_{i+n}$, which are the coordinates of the delay embedding space for $m = n+1$.

### 4.2. Modeling the dynamics

Once we have chosen a phase space, and have a sequence of data points in this space, we want to find a map on the phase space that agrees well with the data (For any given point in the phase space, this map will give an approximation of the point's new position one time step into the future.). One basic question which arises is, how is this approximation to be described? A simple formula in closed form would be helpful, but is usually impossible to find. The arguments made in the previous section about vector fields apply equally well to maps: most maps can not be expressed by simple formulas in closed form, and even when it is theoretically possible, there is no general method known for finding such a formula. One response to this problem is to allow more complex models, so complex in some cases as be impractical to manipulate without a computer.

A related response is to allow non-traditional maps that are best expressed as algorithms rather than formulas. In particular, we allow computer programs that take as input a point in the phase space, and output the value of the map at that point. Any mathematical formula can be encoded in such an algorithm, but not all algorithms correspond to conventional formulas [15]. Even if an algorithm does correspond to an acceptable mathematical formula, the algorithm may be just as useful and natural a description as the formula. Of course, there is a great body of mathematical knowledge available to us if we work with formulas, some of which we lose access to if we work with general algorithms instead [16]. But we believe that the benefits outweigh the losses.

We now mention some specific methods for finding a model $f$ for the dynamics on the reconstructed phase space, corresponding to the true dynamical map $\psi : M \to M$.

The simplest algorithmic model is naive proximity: for any test point $z$ in phase space, find the closest data point $z_i$, and set $f(z) = z_{i+1}$. Another simple model is linear interpolation, using the nearest $m + 1$ data points (in an $m$-dimensional phase space). Due to experimental error, it is often preferable to use more than $m + 1$ data points to fit a local linear or local polynomial model [43] (the entire model $f$ is piecewise linear or polynomial).

The methods mentioned so far are *local*, in that data points far from our test point $z$ do not affect our prediction $f(z)$. By contrast, *global* methods use all data points. The simplest global model is linear regression, or AR modeling. As an extension of this idea, we can fit global polynomials, or linear combinations of other basis functions. Global nonlinear approximations which have been successfully applied to the reconstruction problem include radial basis functions [44–47] and neural nets [48,49]. Methods based on tesselations and triangulations have some global and some local character [50,51].

One argument for global models is that using more data points can lessen the effects of noise. On the other hand, one needs a larger class of potential models in order to model global behaviour accurately; for example, a piecewise linear model can often succeed for a non-linear system where a global linear model is almost useless.

These approximation methods all have their advantages and disadvantages: even for a single dataset, there may be more than one 'best' model, depending on the criteria employed. The principal disadvantage that all of these methods share is that at present there is no really satisfactory theory of functional approximation in higher dimensional spaces. We regard this as a temporary limitation and a challenge to future research rather than a fundamental flaw in the

[15] This is true if 'conventional' means closed form. But there is a long tradition of encoding logic in mathematical formulas, for example,

$$\phi(z) = \begin{cases} z, & \text{if } f(z) = 0, \\ 2z, & \text{if } g(z) = h(z) \text{ and } f(z) = 4, \\ -4z, & \text{otherwise.} \end{cases}$$

If we added some sort of looping construct (based on mathematical recursion), and allowed nested logic, we could conceivably encode any algorithm in a formula. But we would very soon approach a level of complexity at which the algorithm is the clearer description.

[16] But we lose less than we might expect: for example, it is possible [42] to differentiate many *algorithms* in the sense that application of simple rules to successive lines of a program that returns $f(x)$ yields a program that returns the derivative $\mathrm{D}f(x)$.

approach. The beginnings of such a theory can be seen in the work by Jones [52], Barron [49] and others, who in effect divide a broad class of functional approximation methods into what we shall call weak methods, which require exponentially rapidly increasing model size as dimension increases, and strong methods, for which the model size grows much less rapidly. Weak methods include global linear and global polynomial fits, while strong methods include neural nets, radial basis functions, wavelets, and triangulations. Work such as this is essential in overcoming the limitations and in reducing the risks of applying an inadequately understood approximation method.

We conclude by outlining features that a good model should have, and indicating how these considerations fit into the context of algorithmic modeling. We hope that these issues will receive more theoretical attention in the future.

In general terms, a 'good' model for a certain dataset should satisfy the following criteria:

- it fits the data well,
- it is consistent with prior knowledge about the system from which the data was produced,
- it is simple, and
- it is in a form that we can use to answer our questions about the system.

The problem, of course, is how to define these criteria precisely. We will discuss each of these in the following sections.

### *4.3. Fitting the data well*

The question of how well a model fits the data is addressed by standard statistical methods. In general, one wants to do something like choosing parameters of the model to maximize the likelihood of the given data. Note that, from this point of view, the assumed error or noise distribution of the model is as important as the deterministic part, something that is not always made clear in the dynamical systems literature.

For example, for every embedded data point $z_k$ (except the last one), we can compare our model's predicted $f(z_k)$ with the actual data value $z_{k+1}$. We can compute the sum of the squared distances between these predicted and actual values, and consider that the model fits the data best when this is smallest. This is equivalent to assuming that the model errors are independent identically distributed realizations of a Gaussian distribution.

More broadly, we might consider that the model fits the data well if model and data share certain dynamical properties of interest to us. But it is often difficult to describe properties of the data without first making a model of it. For example, assuming the dataset is non-constant (which has to be for meaningful analysis), we know that none of the data points is an equilibrium point of the system. So any statement regarding the position and nature of equilibrium points must be based on some model of the dynamics of the system at points other than the data points.

### *4.4. Incorporating prior knowledge*

We can incorporate prior knowledge or theories about the system in our choice of approximation method. For example, we usually believe that the time-step map is continuous. In this case, we should prefer continuous models. Of the models mentioned above, naive proximity models or local linear models are not usually continuous, whereas triangulations, radial basis functions and neural nets are continuous. Another example is the symmetry properties that many physical systems have; these can be incorporated explicitly into the approximation process. A more interesting kind of prior knowledge would involve qualitative or geometrical features of the sort mentioned in the previous section. For example, we might believe that the system has a stable periodic orbit in or near a certain position. We are not aware of modeling methods that explicitly try to incorporate this kind of prior knowledge, though often a model is evaluated post-facto on these grounds. It should be understood that using algorithmic

modeling methods does not mean we cannot *in principle* exploit our physical knowledge, even if the current state of the art is somewhat lacking in this area.

### 4.5. Simplicity

Any data sequence in $\mathbb{R}$ can be fitted *exactly* with a polynomial with high enough order, by Lagrange interpolation. Yet this model is practically useless to us, because our data are noisy: fitting the noise exactly distorts our model severely. Ideally, we would like to fit the 'true' dynamics and ignore the noise, but in practice, we can not tell the two apart very well. There are many useful techniques for filtering noise [53,54], using fairly mild assumptions which are all variations on the idea that the true system is simpler (eg, smoother, or lower order) than accurate modeling of the noisy data would suggest. But many of these methods have, as yet, incomplete theoretical justification.

The simplicity principle, known as Occam's razor, asserts that when two models seem equally good, the simpler one is more likely to be correct. This can be seen as just another form of prior knowledge or theory about a dynamical system, but it deserves special attention. It may seem paradoxical that we emphasize simplicity at this stage, since we began our consideration of algorithmic modeling by explicitly allowing more complex models than has been conventional. But what we are advocating is an understanding of the trade-offs between simplicity and accuracy, and a systematic incorporation of simplicity into modeling methods.

Bayesian statistics provides one starting point [55] [17]. We will explain the basic idea by contrasting it with maximum likelihood estimation. Suppose we have a class of possible models $\{M_i\}$ (each of which include error distributions), and experimental data $D$. To apply the maximum likelihood method, we choose the model $M_i$ that gives the greatest conditional probability density $p(D|M_i)$ (called the *likelihood*); in other words, we choose the model that would make the observed data most likely. In Bayesian statistics, we modify the method by assigning a priori probability distribution to the models, $p(M_i)$, and then choosing the model which gives the greatest combined probability density $p(M_i \text{ and } D) = p(M_i)p(D|M_i)$.

The interesting problem is then to come up with the a priori probability distribution on the class of models. Two similar, independently developed answers, with theoretical justifications, are the Schwarz Information Criterion [56] and the Akaike Information Criterion [57]. In each method, we assume each model in the class is determined by a finite number of parameters (for example, polynomials of order $n$ are determined by $n$ coefficients). We assign the probability density $\alpha^n$ to models with $n$ parameters [18]. The choice of $\alpha$ is where the methods differ; in all cases, it depends on the number of data points.

These can be seen as attempts to quantify Occam's razor. The most thorough of such an attempt is Rissanen's *minimum description length* [58], which is in effect a way to try to approximate the (incomputable) Kolmogorov complexity of the data [59]. The minimum description length criterion looks at the total number of bits required to describe the model, the parameters and the errors, and considers such matters as the accuracy required of individual parameters. The number of bits to describe the errors turns out to be proportional to the log likelihood, and this is why there is an asymptotic equivalence between the description length and the Schwarz criterion.

These methods have been successfully applied to reconstruction of dynamics using several model classes [47].

[17] Bayesian statistics could incorporate any type of theoretical knowledge about the dynamical system to be modeled, if that theory could be specified as a probability distribution on the model space. But we have not seen this approach used for prior theoretical knowledge other than simplicity.

[18] Actually, this is only a *relative* probability density, since we have not normalized it, and we have not even specified a measure on the space of models; but relative probability is all that is required to apply the Bayesian method.

### *4.6. How the models can be used*

Our models should be in a form that allow us to answer our questions about the underlying systems. It is in this area that most criticisms of the reconstruction approach have been made. We accept this concern as partially valid. After all, there are centuries of mathematics available to us if we work with conventional formulas, so it is unsurprising if newer methods are less sophisticated. Here, we discuss the extent to which algorithmic models can be used to understand dynamical systems. As we did in the previous section, we will roughly categorize the goals of dynamical analysis as prediction and insight.

In many instances effective prediction can be based on the algorithmic models we have discussed above [43,44–48,50,51]. In fact, while the predictions are not always successful, our experience shows that algorithmic models tend to be better than traditional models for this purpose. In some cases, accurate prediction is precluded by the complexity of the system, the quality of the data, or defects in our ability to build good models. When this occurs, algorithmic methods may still be useful in assessing the degree of predictability, and hence whether there appears to be modelable deterministic dynamics even if the models we have available are clearly inadequate.

Perhaps the most misunderstood point is how insight is obtained. Insight is obtained in essentially the same ways as for more traditional system representations: recall from our earlier discussion that algorithmic models are not fundamentally different from formulas, though they are often non-traditional and more complex. To understand either a formula or an algorithm, one does various standard things such as finding attractors, such as fixed points, and tries to estimate their basins of attraction; we may choose to calculate various averages such as Lyapunov exponents, and we may conduct a mixture of numerical simulation and geometrical study to try to provide a qualitative description. In practice, algorithmic models are about as well suited to such studies as are formulas: we can find, for example, fixed points and derivatives with equal ease (or difficulty) in the two cases. For a case where an algorithmic model may be superior to a formulaic description, see [60] which shows how to find *all* the fixed points up to any order, and their stability properties, for a triangulation reconstruction. One apparent advantage of conventional formulas is that we can more often find *exact* answers to qualitative questions. But, when modeling the real world, our formulas will be approximations anyway, so the value of having exact answers is unclear.

We remark that given either a conventional or an algorithmic model, one would like to be able to say something like ‘the dynamics takes this volume of phase space, folds it up thus, and maps it back onto itself’. Such a *geometric* description is seldom achieved using any method; for an approach that works in low dimensions with either old or new fashioned models, see [61].

## 5. Conclusions

We have outlined three distinct ways of mathematically analyzing a dynamical system: the Newtonian paradigm (find formulas for both the vector field and the solution); an approach based on the geometrical theory of differential equations (find a formula for the vector field, but if solutions are required, approximate them using numerical algorithms); and algorithmic modeling (both the dynamical system and its solution are approximated algorithmically).

While there is still much to learn, we are confident in predicting that algorithmic modeling of dynamical systems, based on observational data and conducted without the requirement of a conventional formula for the map or differential equation, will be recognized as an essential and inevitable step in the maturation of dynamical analysis. Just as the introduction of geometrical methods greatly broadened the range of problems that could be treated, investigations based on algorithmic reconstruction will have a similarly liberating effect.

## Acknowledgements

We thank Lenny Smith for interesting discussions, and for directing us to the quotation from James Clerk Maxwell. We would like to thank Sir Peter Swinnerton-Dyer for directing us to Hardy's monograph. Discussions with Henry Abarbanel and Celso Grebogi were particularly important in developing the ideas presented in this essay. This research was partially conducted at the Institute for Nonlinear Science, University of California at San Diego, and we thank Henry Abarbanel for hospitality. The research was supported by NIH Grants NS19716 and NS32983 and by a grant from the Australian Research Council. The continuing support of the Arthur P. Noyes Research Foundation, Norristown State Hospital, is gratefully acknowledged.

## References

[1] P. Hartman, Ordinary Differential Equations, Birkhäuser, Basel, 1982.
[2] M. Bronstein, J.H. Davenport, B.M. Trager, Symbolic integration is algorithmic! Computers and Mathematics (1989).
[3] J. Lützen, Joseph Liouville 1809–1882: Master of Pure and Applied Mathematics, Springer, Berlin, 1990.
[4] G.H. Hardy, The integration of functions of a single variable, in: J.G. Leathem, E.T. Whitaker (Eds.), Cambridge Tracts in Mathematics and Mathematical Physics, vol. 2, Cambridge University Press, Cambridge.
[5] M. Braun, Differential Equations and Their Applications, Springer, New York, 1993.
[6] J. Ritt, Integration in Finite Terms, Columbia University Press, New York, 1948.
[7] C. Marchal, The Three-Body Problem, Elsevier, Amsterdam, 1990.
[8] K. Sundman, Recherches sur le problème des trois corps, Acta Societatis Scientiarum Fennicae 34 (1908).
[9] K. Sundman, Nouvelle recherches sur le problème des trois corps, Acta Societatis Scientiarum Fennicae 35 (1909).
[10] K. Sundman, Mèmoire sur le problème des trois corps, Acta Mathematica 36 (1912) 105–179.
[11] Q. Wang, The global solution of the $n$-body problem, Celestial Mech. 50 (1991) 73–88.
[12] F. Diacu, The solution of the $n$-body problem, Math. Intelligencer 18 (1996) 66–70.
[13] J.D. Lambert, Computational Methods for Ordinary Differential Equations, Wiley, New York, 1973.
[14] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C, Cambridge University Press, Cambridge, 1988.
[15] E.L. Ince, Ordinary Differential Equations, Longmans & Green, London, 1927.
[16] M. Kline, Mathematical Thought from Ancient to Modern Times, Oxford University Press, New York, 1972.
[17] H. Poincaré, Mémoire sur les courbes définies par une équation différentielle, Journal de Mathématiques 3(7) (1881) 375–422.
[18] H. Poincaré, Mémoire sur les courbes définies par une équation différentielle, Journal de Mathématiques 3(8) (1882) 251–296.
[19] H. Poincaré, Sur les courbes définies par les équations différentielles, Journal de Mathématiques 4(1) (1885) 167–244.
[20] H. Poincaré, Sur les courbes définies par les équations différentielles, Journal de Mathématiques 4(2) (1886) 151–217.
[21] J. Barrow-Green, Poincaré and the three body problem, in: History of Mathematics, vol. 11, American Mathematical Society and London Mathematical Society, 1997.
[22] A.A. Andronov, A.A. Vitt, S.E. Khaikin, Theory of Oscillators, Pergamon, London, 1966.
[23] N. Minorsky, Nonlinear Oscillations, Van Nostrand, New York, 1962.
[24] C. Hayashi, Nonlinear Oscillations in Physical Systems, Princeton University Press, Princeton, NJ, 1964.
[25] M.W. Hirsch, S. Smale, Differential Equations, Dynamical Systems and Linear Algebra, Academic Press, New York, 1974.
[26] J. Guckenheimer, P. Holmes, Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, vol. 42, Springer, New York, 1983.
[27] S.H. Strogatz, Nonlinear Dynamics and Chaos, Addison-Wesley, New York, 1994.
[28] C.H. Waddington, Towards a Theoretical Biology, Edinburgh University Press, Edinburgh, 1972.
[29] L. Michaelis, M.L. Menten, Die Kinetic der Invertinwirkung, Biochem. Z. 49 (1913) 333–369.
[30] R. Abraham, J.E. Marsden, Foundations of Mechanics, Addison-Wesley, New York, 1978.
[31] R.G. Bartle, The Elements of Real Analysis, Wiley, New York, 1964.
[32] B.P. Zeigler, Simplification of biochemical reaction systems, in: L.A. Segel (Ed.), Mathematical Models in Molecular and Cellular Biology, Cambridge University Press, Cambridge, 1980, pp. 112–143.
[33] H.D.I. Abarbanel, Analysis of Observed Chaotic Data, Springer, New York, 1996.
[34] T. Sauer, J.A. Yorke, M. Casdagli, Embedology, J. Stat. Phys. 65 (1992) 579–616.
[35] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.S. Young (Eds.), Dynamical Systems and Turbulence, vol. 898, Springer, Berlin, 1981, pp. 365–381.
[36] R. Mañé, On the dimension of the compact invariant sets of certain nonlinear maps, in: D.A. Rand, L.-S. Young (Eds.), Dynamical Systems and Turbulence, vol. 898, Springer, New York, 1980, pp. 230–242.

[37] L. Noakes, The Takens embedding theorem, Int. J. Bifurcation and Chaos 1(4) (1991) 867–872.
[38] K. Judd, A.I. Mees, Embedding as a modeling problem, Physica D 120 (1998) 273–286.
[39] J.-P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. 57 (1985) 617–656.
[40] C.J. Cellucci, A.M. Albano, P.E. Rapp, A.D. Krystal, A comparative study of embedding methods, Phys. Rev. (1998), submitted for publication.
[41] D.T. Kaplan, L. Glass, Direct test for determinism in a time series, Phys. Rev. Lett. 68(4) (1992) 427–430.
[42] A. Griewank, G.F. Corliss, Automatic differentiation of algorithms: theory, implementation and application, in: SIAM Proc. on Applied Mathematics, vol. 53, 1991.
[43] J.D. Farmer, J.J. Sidorowich, Predicting chaotic time series, Phys. Rev. Lett. 59 (1987) 845–848.
[44] M. Casdagli, Nonlinear prediction of chaotic time series, Physica D 35 (1989) 335–356.
[45] L.A. Smith, Identification and prediction of low-dimensional dynamics, Physica D 58 (1992) 50–76.
[46] A.I. Mees, Parsimonious dynamical reconstruction, Int. J. Bifurcation and Chaos 3 (1993) 669–675.
[47] K. Judd, A.I. Mees, On selecting models for nonlinear time series, Physica D 82 (1995) 426–444.
[48] A.S. Weigend, B.A. Huberman, D.E. Rumelhart, Predicting the future: a connectionist approach, Int. J. Neural Systems 1 (1990) 193–209.
[49] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. on Information Theory 39 (1993) 930–945.
[50] A.I. Mees, Dynamical systems and tesselations: Detecting determinism in data, Int. J. Bifurcation and Chaos 1 (1991) 777–794.
[51] S. Allie, A.I. Mees, Reconstructing noisy dynamical systems by triangulation, Phys. Rev. E 55 (1997) 87–93.
[52] L.K. Jones, A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training, The Annals of Statistics 20 (1992) 608–613.
[53] R.W. Hamming, Digital Filters, 3rd ed. Prentice-Hall, NJ, 1989.
[54] J.B. Kadtke, J. Brush, J. Holzfuss, Global dynamical equations and Lyapunov exponents from noisy chaotic time series, Intern. J. Bifurcation and Chaos 3(3) (1993) 607–616.
[55] H.J. Larson, Introduction to Probability Theory and Statistical Inference, Wiley, New York, 1982.
[56] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[57] H. Akaike, A New Look at the Statistical Model Identification Problem, IEEE Trans. on Automatic Control 19(6) (1974) 716–723.
[58] J. Rissanen, Stochastic Complexity in Statistical Inquiry, Series in Computer Science, vol. 15. World Scientific, Singapore, 1989.
[59] M. Li, P. Vitanyi, An Introduction to Kolmogorov Complexity Theory and its Applications, Springer, Berlin, 1993.
[60] S. Allie, A. Mees, Finding periodic orbits from short time series, Phys. Rev. Lett., 1997, submitted for publication.
[61] K. Judd, A.I. Mees, K. Aihara, M. Toyoda, Grid imaging for a two-dimensional map, Int. J. Bifurcation and Chaos in Appl. Sci. Eng. 1 (1991) 197–210.