PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Observing complexity, seeing simplicity

Jaroslav Stark

Phil. Trans. R. Soc. Lond. A 2000 **358**, 41-61
doi: 10.1098/rsta.2000.0518

THE ROYAL SOCIETY

Observing complexity, seeing simplicity

By Jaroslav Stark

Centre for Nonlinear Dynamics and its Applications, University College London, Gower Street, London WC1E 6BT, UK

This century has seen the formulation of a number of novel mathematical and computational frameworks for the study, characterization and control of various classes of complex phenomena. Most of these involve some non-trivial dynamics. In order to be of genuine use in the real world, it is essential that such theoretical developments are related to observed data. This paper is concerned with the question of how this might be achieved. In particular, it investigates how much information about a complex unknown system one can hope to recover from observations. The vast majority of theoretical analysis assumes that we have an accurate model of a system and that we know the variables that uniquely determine its state. In principle, the application of such a theory to real problems requires the simultaneous measurement of all these variables. This is rarely feasible in practice, where often we will not even know what the important variables are. All that we may be able to achieve is to make a sequence of repeated measurements of one or more observables. The relationship between such observations and the state of the system is often uncertain. It is therefore unclear how much information about the behaviour of the system we can deduce from such measurements. It turns out that for a certain class of mathematically idealized systems it is, in principle, possible to reconstruct the whole system from a sequence of measurements of just a single observable. As a consequence, we may be able to build remarkably simple models of apparently complex looking behaviour. We shall outline the theoretical framework behind this remarkable result, and discuss its limitations and its generalizations to more realistic systems. Finally, we shall speculate that the complexity of theoretical models may sometimes outstrip our ability to detect them in real data.

Keywords: time-series; Takens' theorem; embedding; delay reconstruction; prediction; spatio-temporal

1. Introduction

In a certain sense, the problem of complexity has been the fundamental issue of all the sciences throughout history. Of course, as our understanding of the world has improved, our notion of what we mean by complex has changed. In particular, systems whose properties initially seem to lack any rhyme or reason gradually come to be seen as simpler and simpler as we discover the mechanisms underlying their behaviour and devise scientific tools to analyse and ultimately control their operation. The development of such understanding, and the apparent simplification of complex observations, has been the driving force of most scientific enquiry.

A good example of this is the field of celestial mechanics. Right from the dawn of the earliest human societies the night sky has inspired scientific enquiry. Indeed, ancient astronomers could probably claim to be the first scientists. Through painstaking observation they sought to organize and make sense of the vast number of celestial objects that at first must have appeared beyond all comprehension. In this they were surprisingly successful, and even 5000 years ago could predict many events, such as solar and lunar eclipses, and the appearance of some meteor showers and comets with a significant degree of accuracy. Such forecasts were entirely based on the recognition of specific patterns in observed data, rather than on any appreciation of fundamental mechanisms.

As time went on, the requirements for better and better predictions (to a large extent driven by the practical needs of marine navigation) necessitated the development of theoretical frameworks, ultimately leading to the development of both Newton's laws of motion and of differential calculus. While in principle these provided a complete description of the behaviour of, for instance, the Solar System (as well as supplying the tools and motivation for a substantial part of applied mathematics), they by no means afforded a total solution to the problem.

In particular, at the end of the 19th century, Poincaré realized that the behaviour of even just three bodies subject to Newton's laws could be so complex as to defy complete understanding. This realization led to the development of the modern theory of nonlinear dynamics and the discovery of chaotic behaviour in simple low-dimensional systems. Indeed today, the Solar System still manages to provide surprises, such as the collision of comet Shoemaker–Levy 9 with Jupiter and the sudden appearance of comet Hale–Bopp after 4000 years. Furthermore, even with our powerful computational facilities we still cannot give a detailed description of the dynamics of the entire asteroid belt, or entirely agree on the mechanisms that created the rings of Saturn.

This example illustrates the fundamental role that complexity plays in science, and the importance of the interplay between observation and theory. It highlights how an appropriate theoretical framework can lead to a simple description of apparently complex observations, but also how even simple and well-understood mechanisms can lead to behaviour that even after centuries of study is not completely understood. However, this example also runs the danger of being deceptively over-simplistic. The interactions underlying many of the most interesting and practically relevant problems today are often an order of magnitude more complicated, and several orders less well understood, than those governing the motion of the Solar System. Thus, when attempting to study the nervous and immune systems, analyse the stability of ecosystems or the climate, or seek to efficiently control transport, telecommunication or energy supply networks, we are perhaps not in much better a position than the first human staring up with awe and bewilderment at the night sky.

This century has seen the formulation of a number of novel mathematical and computational frameworks for the study, characterization and control of such complex phenomena. In order to be of any relevance to practical problems, it is important that such theoretical advances are closely related to observed data. In particular, it is essential to understand how much information about an unknown system one can hope to recover from observations.

This paper is concerned with two aspects of this fundamental problem. Firstly, note that the vast majority of theoretical analysis of complex behaviour assumes that we have an accurate model of a system and that we know the variables that uniquely determine its state. In principle, the application of such a theory to real problems requires the simultaneous measurement of all these variables. Unfortunately, this is rarely feasible in practice, particularly in the real world (as opposed to the laboratory bench). Even in the simple example of the Solar System, introduced above, it is problematical. Thus we know that a proper description of the dynamics requires all three components of both the position and the velocity of each body in the system. Throughout much of history we had no direct method of observing the instantaneous velocity of a planet, or our distance from it, and had to be content with the two components of its position projected onto the sky. Only recently, with the advent of laser and radar techniques, have we been able to measure all the required variables directly. Furthermore, even today, should we wish to incorporate all the asteroids in our model, we would have great difficulty in observing all their positions and velocities simultaneously! In many other problems, one will not even know what the important variables are.

Often, therefore, we are limited to making a sequence of measurements of one or more observables, such as components of the position of a few planets, the population of a number of different species in a given ecosystem, the air temperature and pressure at a finite number of geographical locations, or the density of traffic at several positions on a motorway. The relationship between such observations and the state of the system is often uncertain. It is therefore not immediately obvious how relevant theoretical models of complex behaviour are to the analysis of real systems.

Remarkably, over the last two decades it has come to be realized that in many cases one can reconstruct an unknown system from a sequence of measurements of just a single observable. This can result in relatively simple descriptions of apparently complex datasets and has stimulated a variety of applications in fields ranging from fluid dynamics, through electrical engineering to biology, medicine and economics. It has led to the development of novel algorithms that can characterize, predict and manipulate a system on the basis of observed data. Similar concepts have appeared in control theory, and, more recently, in computer science, with applications as diverse as speech recognition and data compression. The first half of the paper will describe these ideas and some of the ways in which they can be applied.

Unfortunately, despite the obvious success of these techniques, the theoretical framework on which they are based is rather limited and, strictly speaking, fails to encompass any system in the real world. This is rather unsatisfactory, since it means that all the algorithms motivated by this framework are operating in circumstances beyond its scope. As a result, we have little understanding of what such methods are actually doing, what their fundamental limits are, and how much information they actually preserve. These considerations have led to a number of recent extensions of this framework to systems closer to those encountered in the real world. These are outlined in the latter part of the paper.

However, even these generalizations are confined to relatively simple systems whose behaviour is governed by only a few important variables. As far as we are aware, there is still no systematic approach to the observation of genuinely complex behaviour. We shall conclude the paper by arguing that this may be due to fundamental limits imposed by the process of observation, and that for many complex phenomena there is just no way of validating a proposed model against measured data. This suggests that the best we may be able to do in such situations is to explain as much as possible of the observed phenomenon by a relatively simple model, and attribute the remainder to random effects, which we do not attempt to describe in detail. At least in some idealized examples, such an approach of observing complexity but seeing only simplicity can work surprisingly well.

2. Complex systems and deterministic dynamics

Today, in many scientific circles the term 'complexity' has acquired a specific technical meaning. In this paper, however, we want to use the term 'complex' in a manner consistent with its everyday usage, rather than adopt any precise technical definition. By a complex system, we shall thus simply mean one that exhibits either structure or behaviour that is complicated to describe, to predict or to control. Within this context, we shall be led to focus largely on systems with some non-trivial dynamical evolution. Most complex systems of interest fall into this category. Furthermore, even when one encounters complex static patterns, these are almost always generated by some dynamical process that one has to understand in order to analyse their properties.

We shall also use a very broad notion of a dynamical system, and use this term to simply mean any system whose state changes (or potentially could change) with time. Initially, however, we shall concentrate on systems that are deterministic and autonomous (that is independent of any external events), since, until recently, this has been the setting for the ideas that we explore in this paper. This will allow us to present these ideas in their simplest possible form. Of course, all real systems are subject to at least some noise, and many are affected by external perturbations or inputs. We shall return to these later in the paper.

We shall denote the state of our system by the symbol x and the space of all such possible states as X (called, for obvious reasons, the *state space*). Often the state will be a vector $x = (v_1, v_2, \dots, v_m)$, with each component v_i representing the value of some property of the system, such as the position or velocity of a particular planet, the voltage or current at a particular point in an electronic circuit, the population of some species in an ecosystem, the price of a particular commodity, etc.

The dynamical evolution of our system is defined by a rule that given the current state of the system determines the state some specified time in the future. Mathematically, this is described by a function f from the state space to itself, such that if the system is in state x now, it will be in state $f(x)$ a time τ later. It will then be in state $f(f(x))$ after a time 2τ, state $f(f(f(x)))$ after a time 3τ, and so on. If we use the notation x_n to mean the state of the system at time $n\tau$, we get a sequence of states $x_0, x_1, \dots, x_n, \dots$, mapping out the evolution of our system through time, with $x_{n+1} = f(x_n)$.

Many systems, of course, do not evolve in discrete time-steps, so this may be considered a rather restricted framework. However, we shall be largely concerned with observation, and any practical data-gathering mechanism can only operate at discrete times. Hence, even if the system has continuous time evolution, we can only observe a sequence of discrete snapshots, much as a video camera actually records a sequence of still images. Thus, in the example of the Solar System described above, we would typically only be interested in the position of the planets at a fixed time on successive nights, and the function f would take the positions and velocities on one night to those on the next.

3. Observing deterministic dynamical systems

Historically, most of the theory of such dynamical systems has been carried out under the assumption that the state space, X, and the function f are known. While this is the case in the above example of the Solar System, there are many other problems, particularly in biology and the social sciences, where we only have the sketchiest idea of the variables that determine the behaviour of the system (i.e. $v_1, v_2, \ldots, v_m$) and the interactions between them (i.e. f). Furthermore, even when we do possess such knowledge, the application of most nonlinear dynamics techniques requires the simultaneous observation of all the variables $v_1, v_2, \ldots, v_m$. As already indicated in § 1, this is difficult even for a system as well understood as the motion of the planets. When we turn to problems in biology or the social sciences, there is little hope of carrying out the required measurements. A typical example is given by an ecosystem. We rarely know which species need to be incorporated in an accurate description of a particular ecological system, and even more rarely can we quantify the interactions between them. However, even if we could build a realistic model, we are very unlikely to be able to measure the populations of all the relevant species, and any other necessary variables (such as nutrient levels or carbon dioxide concentrations). There is thus little chance of ever knowing, let alone measuring, all the state variables in this system.

Such examples illustrate that usually all that we can expect is to make a sequence of repeated measurements of one or more observables. These may be some subset of the state variables, but more generally will be some arbitrary function $\varphi(x)$ of the state x. In the simplest case of a single observable, $\varphi(x)$ will just be a number, and the sequence of observed quantities $\varphi(x_0), \varphi(x_1), \ldots, \varphi(x_n), \ldots$, will form a so-called *scalar time-series*. In the case of the above examples, this might represent a sequence of measurements of the angle above the horizon of Jupiter, or the number of Canadian lynx pelts offered for sale by trappers in the Mackenzie River district in successive months. The latter demonstrates why we want to consider an arbitrary measurement function φ, rather than just assuming that we are observing one of the state variables: we cannot hope to directly observe the total population of lynx, but can reasonably assume that the numbers trapped are some function of this.

Increasingly, we are in the position of being able to measure several observables simultaneously. We shall discuss this in a subsequent section. For the moment, however, we shall restrict ourselves to the case of a scalar observable, since this allows us to present the fundamental ideas in their simplest possible setting. Graphically, it is easiest to illustrate the observation procedure in a simulated system on a computer. We do this here with the well-known Lorenz equations. The state space of these consists of three variables (v_1, v_2, v_3) whose evolution is governed by the differential equations

$$\begin{aligned}
\dot{v}_1 &= 10(v_2 - v_1),\\
\dot{v}_2 &= -v_1 v_3 + 28 v_1 - v_2,\\
\dot{v}_3 &= v_1 v_2 - \tfrac{8}{3} v_3.
\end{aligned}$$

Here, a dot over a variable represents its rate of change, so these equations relate the direction and speed of the variables (left-hand side) to functions of their present values (right-hand side). By following a trajectory generated by this prescription,

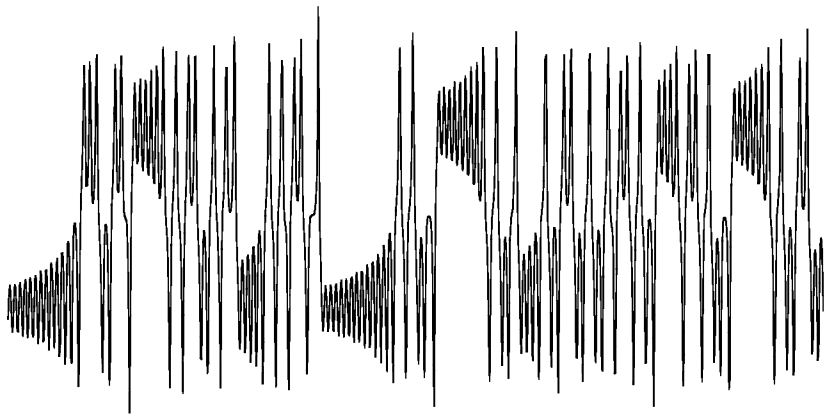

Figure 1. Time-series from Lorenz equations.

and sampling at equal time intervals, we obtain a dynamical system of the kind described above.

The Lorenz system is one of the classic examples of chaotic dynamics, derived by the meteorologist E. Lorenz (1963) as a simple caricature of fluid convection, and, hence, in some sense, of large weather systems. Despite their apparent simplicity, these equations have helped to provide the motivation for some of the most fundamental developments in the subject over the past four decades.

These equations illustrate the difficulties we encounter in trying to characterize complexity, and the way in which the process of observation can affect such characterization. Thus, the equations written as above look very innocuous: only three variables v_1, v_2 and v_3, and only two simple nonlinear terms ($-v_1v_3$ and v_1v_2 in the second and third equations, respectively). If we were told that a real system was governed by such equations, we would be tempted to say that it must be a simple system! This would certainly have been the accepted opinion of scientists and engineers until at least the 1960s. Indeed, in writing these equations down, Lorenz was not looking for anything complex, and believed that at most they would exhibit simple periodic oscillations.

Much to his surprise, when Lorenz integrated them on a computer and plotted one of the variables against time, he obtained a time-series like that in figure 1: far from periodic, and far from simple! However, after careful checking, he became convinced that he was seeing a genuine phenomenon, and one of the major steps in the development of chaotic dynamics was made.

It is now well known that even the simplest and most innocent-looking nonlinear systems can lead to complex-looking and apparently unpredictable behaviour, and this phenomenon is called *chaos*. Many of the mechanisms underlying this are now well understood, though it has to be said that even after 35 years of concerted effort there are fundamental properties of the Lorenz equations that still defy complete analysis.

The time-series in figure 1 was obtained by sampling the v_1 variable at regular time-intervals. In the context of our framework above, this corresponds to observing the system using the observable $\varphi(v_1, v_2, v_3) = v_1$. This produces a complex and irregular pattern, and, as we shall see, is not the ideal viewpoint from which to

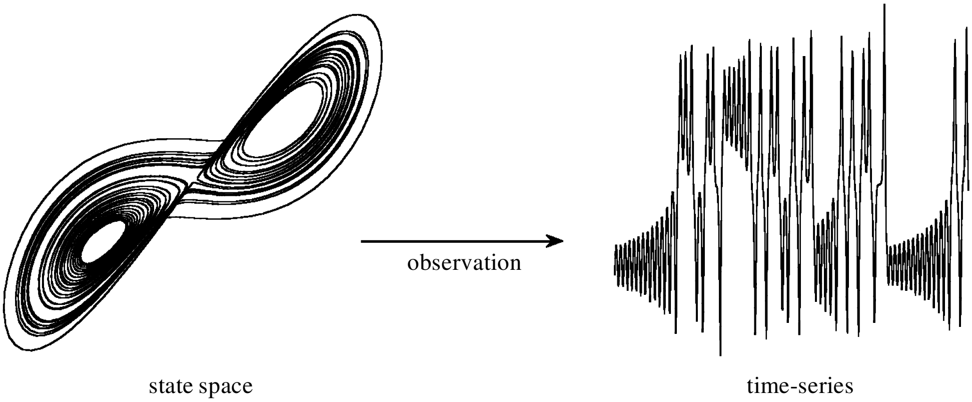

Figure 2. State space and time-series views of the Lorenz system.

study these equations. If, instead, we plot all three variables (v_1, v_2, v_3) as vectors in a three-dimensional state space, we find that the system quickly approaches and settles on a complicated set shown on the left in figure 2, called the *Lorenz attractor* (actually, this is a projection of this object onto the two-dimensional page). While still quite intricate, it displays far more apparent structure than the time-series plot. Objects such as this can be investigated using the modern tools of topology and geometry, and are the focus of much of modern nonlinear dynamics.

The remainder of figure 2 illustrates the process of observation, using the observable $\varphi(v_1, v_2, v_3) = v_1$. As we can see, this seems to lead to a loss of structure and information and an apparent increase in complexity. Certainly, in moving from the equations, through the state space picture, to the time-series, we have gone from something apparently simple to something apparently complex. Yet, as already discussed above, it is precisely the right-hand side of figure 2 that represents the situation that we frequently face in the real world. The crucial issue that we therefore wish to address in this paper is to what extent the apparent complexity of the time-series picture is real, to what extent does the process of observation lose information, and to what extent can we reconstruct the left-hand side of figure 2 from the right-hand side?

In the context of the Solar System, this would be like trying to study the dynamics of the whole system by observing just the position of Jupiter. Even in the simpler case of the Lorenz equations, the state of the system at any given time is specified by three variables v_1, v_2 and v_3, while the time-series consists of only a single variable. The time-series thus appears to contain far less information than the state space representation, and the fact that it originates in a deterministic system appears to be of little use.

4. Reconstructing the state space

Remarkably enough, however, it turns out that these difficulties can be overcome, and, at least in a certain sense, it is possible to 'reconstruct' the state space picture just from the observed time-series, using the so called *method of delays*. The systematic use of this technique was first suggested by Packard *et al.* (1980), who attributed the basic idea to Ruelle, though a number of other authors around that

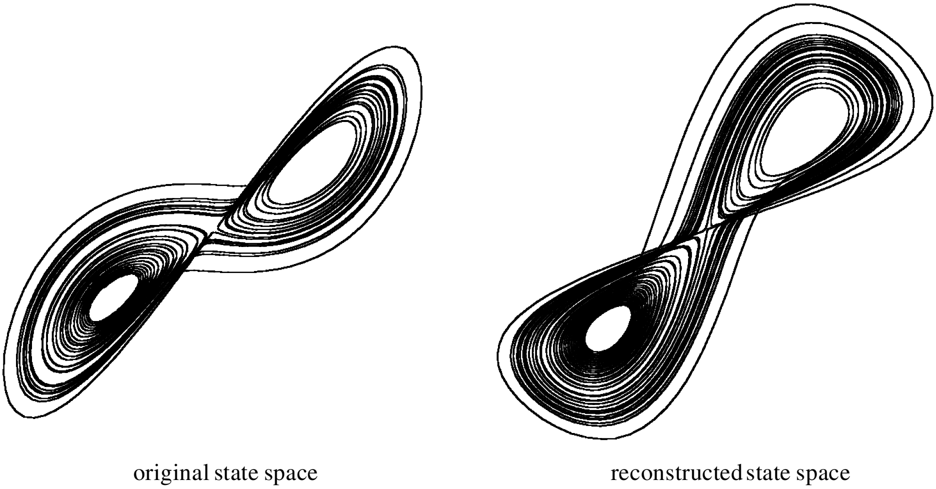

Figure 3. Original and reconstructed Lorenz attractors.

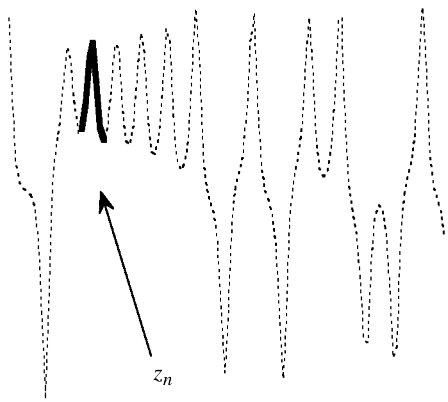

Figure 4. The state z_n corresponds to a pattern of successive values of the time-series.

time were beginning to experiment with it for specific systems (see, for example, Ott *et al*. 1994).

We illustrate this method in figure 3, which shows both the original attractor, drawn with the full knowledge of v_1, v_2 and v_3, and a reconstructed attractor, drawn using just the v_1 time-series. While the two are not identical, they are astonishingly similar. In fact, the two pictures look like two different views of the same object, which, as it turns out, is exactly what they are. It is therefore clear that by passing to the time-series we have not really lost anything fundamental.

To explain this apparent piece of black magic, let us use the notation $\varphi_n = \varphi(x_n)$ to denote the time-series. We thus assume that our only knowledge of a dynamical system is given by the single sequence of numbers $\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots$. In order to reconstruct the state space and the dynamics of our system, we need to fabricate something consisting of many variables out of the single variable φ_n. The only conceivable way of doing this is to group together a number of successive φ_n to create

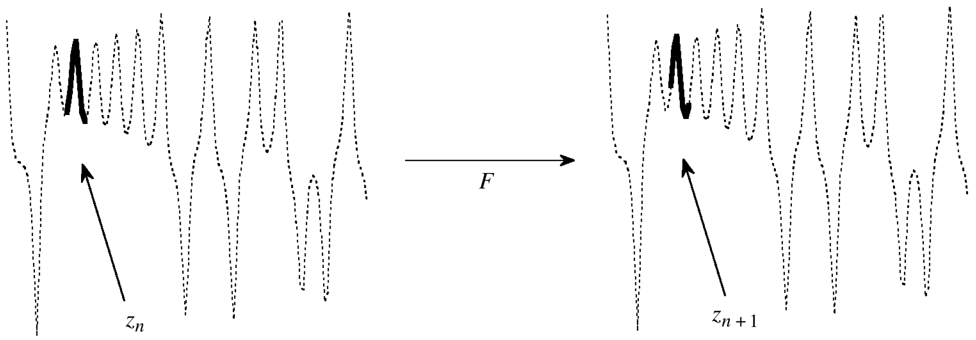

Figure 5. The reconstructed dynamics F, in terms of the time-series.

a vector z_n. Thus, let us define

$$z_n = (\varphi_{n-d}, \varphi_{n-d+1}, \ldots, \varphi_{n-1}).$$

Graphically, the construction of z_n is shown in figure 4, for the case of the Lorenz system.

Here, d is simply the number of successive values that we choose to take, and is called the *embedding dimension*. A variety of algorithms exists for estimating an appropriate choice (see, for example, Ott *et al.* 1994; Abarbanel 1995; Kantz & Schreiber 1998), and, hence, for the moment, we shall simply assume that such a choice has been made. If we now plot all the vectors derived in this way from the time-series, we obtain the reconstructed attractor on the right of figure 3.

In order to understand how and why this procedure works and in what sense the two sides of figure 3 are the same, we need to turn to a fundamental theorem proved nearly 20 years ago by Takens (1980). This provides a proper mathematical justification for the method of delays. It also shows that not only can we recover the geometric properties of the attractor, as a static object in state space, but we can also reconstruct the mapping f that governs the dynamics.

5. Reconstructing the dynamics

The state space of our reconstructed dynamical system consists of all the delay vectors constructed as above. We can thus think of the method of delays as 'synthesizing' this multivariable space out of the single-variable time-series by using successive observations. We now attempt to define a dynamical evolution, which we shall call F, on this 'synthesized' state space by sending z_n to z_{n+1}, as shown in figure 5.

The map F thus simply represents a 'sliding window' moving through the time-series. In doing this, it might appear that we are simply playing mathematical games, and just defining objects arbitrarily. In particular, why should F have anything to do with our original unknown dynamical system f? Before we address that, though, we need to ask whether F is well-defined. At first, this might seem a nonsensical question, after all, have we not just defined F by setting $F(z_n) = z_{n+1}$? However, what if the time-series at two different times yields the same reconstructed vector, so that $z_n = z_p$ for some p different from n. Then, if F is to be a bona fide function, we need $F(z_n) = F(z_p)$, that is $z_{n+1} = z_{p+1}$. This is far from guaranteed, and is closely related to the predictability of the time-series, which we shall discuss below.

It also suggests that it would be sensible to enquire about the regularity of F, i.e. is it continuous, differentiable and so on? In other words, suppose that z_n is not exactly the same as z_p, but is very close. Does it then follow that z_{n+1} must lie close to z_{p+1}?

It turns out that the answer to all of these questions is yes, thanks to the *Takens' embedding theorem* (Takens 1980). This guarantees that for typical systems and observations, and for d sufficiently large, the reconstructed dynamics F is well-defined and as regular as the original dynamics f. Furthermore, f and F really represent the same dynamics, in different 'disguises', or, more precisely, in different coordinate systems. Thus, all those properties that are independent of coordinates, and this includes the vast majority of those considered by modern nonlinear dynamics, will be the same for both the original and the reconstructed systems.

To explain this better, consider the example of cartographic projections employed in drawing maps. The surface of the Earth is curved and cannot be represented on a two-dimensional page without some inevitable distortion. To see this, just try cutting a tennis ball in half and pressing it flat. Some stretching or compression of the rubber is inescapable. Throughout the ages, map makers have, therefore, evolved a variety of coordinate systems, or projections (see, for example, *The Times Concise Atlas*, 5th edn (1991), London, Harper Collins), each designed to minimize the deformation in some particular way. Thus, conformal projections preserve shape, equal-area projections preserve area, and equidistant projections preserve distances from a reference point. However, the more we try to preserve one of these features, the more we find we have to distort the others.

This can be seen, for instance, in the apparent shape and size of, say, Greenland in the Mercator and Gall projections. Even when drawn on the same scale, straight-line distances on the page between corresponding points will be drastically different in these two projections. On the other hand, the underlying geographical reality is the same in both cases, and the fundamental properties of Greenland remain independent of our choice of map. The two maps, therefore, simply provide different descriptions for different purposes.

In exactly the same way, the two halves of figure 3 are pictures of the same phenomenon, seen through two different 'projections'. From this perspective, Takens' theorem thus informally says that the method of delays recovers the original dynamical system, but viewed in a new coordinate system. What use is this in practice? If the coordinate system were known, as it is in cartography, we could, in principle, transform back into the original coordinates, and reconstruct the original system exactly. Unfortunately, in the method of delays, no explicit formula for the coordinate transformation is usually available. Nevertheless, the knowledge that such a transformation exists is sufficient in many applications. It gives us a firm mathematical justification for practical algorithms that use observed time-series to characterize the complexity and chaotic nature of a dynamical system. Furthermore, it guarantees that the time-series is predictable and allows us to manipulate it in a variety of sophisticated ways. These techniques are described further in the following sections. It must, however, be stressed that all such applications depend on the availability of adequate amounts of good-quality data. Takens' theorem only guarantees reconstruction in a mathematically idealized sense. It cannot possibly rescue us if we have poor or insufficient data. The actual amount of data, and their quality, are very dependent on the specific applications, though some general guide-lines exist (see, for example, Ott *et al*. 1994; Abarbanel 1995; Kantz & Schreiber 1998).

We comment next on two technical aspects of Takens' theorem. Firstly, it is crucial that the term 'typical' is included in its statement: the result is certainly not true for all systems and all observables. Thus, suppose we take φ to be a constant function. This might, for instance, correspond to a broken measuring apparatus that just gives a fixed output. Then the time-series will consist of a repeated sequence of the same value. No amount of mathematical sophistication can recover any information about the original system from this. In the actual mathematical formulation of the theorem, the word 'typical' has a precise technical definition, which need not concern us here (see also Sauer *et al*. 1991).

The second issue is the size of d required to ensure a valid reconstruction. The theorem in fact states that any d greater than or equal to $2m + 1$ will do, where m is the number of state variables in the original state space X. Unfortunately, in any real system, m will be unknown, so this condition is of little use. In practice, there exist many algorithms that allow us to make a reasonable choice (see, for example, Ott *et al*. 1994; Abarbanel 1995; Kantz & Schreiber 1998).

Finally, we remark that similar concepts appear in control theory, and, indeed, a weaker version of Takens' theorem was proved independently by the control theorist Aeyels (1981). Related ideas are also beginning to appear in computer science, with applications as diverse as speech recognition and data compression (A. Mees, personal communication). Furthermore, the whole question of how much information can be deduced about a system from observed data is also a central one to statistics. There is, therefore, an urgent need to relate these various approaches to each other.

6. Characterizing complexity

Let us now consider the first of the applications mentioned in § 5. Nonlinear dynamics has developed a variety of measures designed to characterize various aspects of chaotic systems, including fractal dimensions, Liapunov exponents and entropies. All of these attempt to quantify the complexity of a system in one way or another. In particular, they measure, respectively, the number of independent variables governing the asymptotic dynamics, the loss of predictability with time, and the rate of production of information (see, for example, Ott *et al*. 1994). The problem is that all such measures are defined in the context of the state space of a system, i.e. on the left-hand side of figure 3. *A priori*, it is therefore not clear how they can be used when the only information that we possess is an observed time-series.

Fortunately, all of these characteristic quantities are independent of the coordinate system in which they are computed. Takens' theorem thus guarantees that if we calculate them in the reconstructed state space, we will get the same answer as in the original state space. Thus, for example, the fractal dimension of the two different pictures of the Lorenz attractor in figure 3 is the same. Recall, however, that the one on the right-hand side was drawn using only the observed time-series data. Hence, provided that this time-series is sufficiently long, and of sufficiently high quality, we can use it to compute the fractal dimension of the original attractor. The same holds for the other invariants mentioned above, and, thus, Takens' theorem provides a powerful framework for the characterization of chaotic dynamical systems from observed data.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

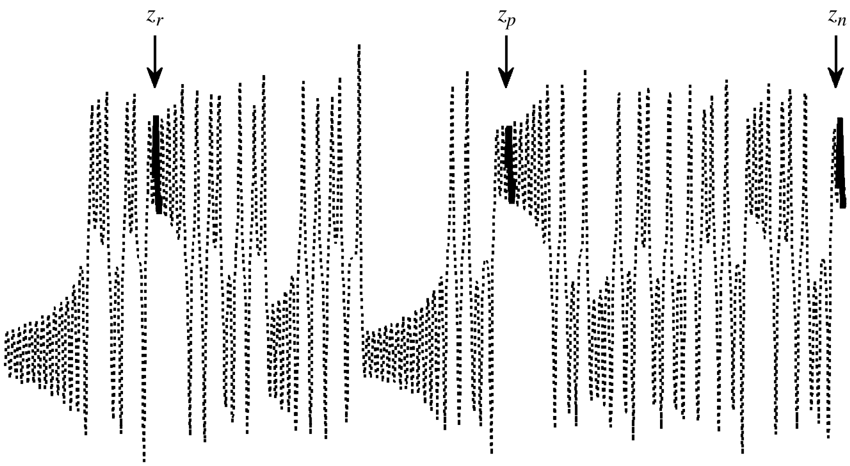

Figure 6. Forecasting by analogy: patterns z_r and z_p are a good match to z_n, and the values following them are a good estimate of the next value after z_n.

7. Prediction

Although significant, the ability to characterize complex systems is often only a small part of our goal. In many practical applications, we need to go much further and predict, control and otherwise manipulate such systems. Once again, Takens' theorem comes to the rescue. Recall that this gives us a function F such that $F(z_n) = z_{n+1}$. Using the definition of z_n and z_{n+1} in terms of the time-series gives

$$(\varphi_{n-d}, \varphi_{n-d+1}, \dots, \varphi_{n-1}) \xrightarrow{F} (\varphi_{n-d+1}, \varphi_{n-d+2}, \dots, \varphi_n).$$

This is, in fact, just a representation of figure 5 in mathematical symbols! Now, observe that F consists of d components, that is $F(z_n) = (F_1(z_n), \dots, F_d(z_n))$, with $F_i(z_n) = \varphi_{n-d+i}$. Writing this as $F_i(\varphi_{n-d}, \varphi_{n-d+1}, \dots, \varphi_{n-1}) = \varphi_{n-d+i}$, we see that the first $d-1$ components $F_1, \dots, F_{d-1}$ are trivial and simply consist of copying one of the arguments of F_i. The only non-trivial component of F is, thus, the last one, F_d, which, for convenience, we will denote by G, so that

$$\varphi_n = G(\varphi_{n-d}, \varphi_{n-d+1}, \dots, \varphi_{n-1}). \tag{7.1}$$

In other words, any given element of the time-series is determined by the preceding d values. Thus, in principle, the time-series is entirely predictable. In practice, there is a minor stumbling block: we will generally not know G explicitly. However, if we have a sufficiently large sample of the time-series in the past, we can use one of a plethora of available nonlinear function-fitting methods to estimate G. We can then use such an estimate from that point on for forecasting purposes. Possible fitting techniques include neural networks, radial basis functions, Gaussian processes, kernel estimators, local polynomial fits, and many others (see, for example, Ott *et al*. 1994; Kantz & Schreiber 1998).

It may help to understand equation (7.1), and this approach to prediction, by considering the intuitive notion of 'forecasting by analogy'. It is quite natural to attempt to predict a time-series by searching for common patterns in its past history. This is a common technique for everyone from a casino gambler to a stock-market analyst, and was, essentially, the method used by astronomers until Newton's time. Thus, suppose that we currently see the pattern $z_n = (\varphi_{n-d}, \varphi_{n-d+1}, \ldots, \varphi_{n-1})$ of d successive values. We search through the past history of the time-series and find a similar pattern in the past, such as $z_p = (\varphi_{p-d}, \varphi_{p-d+1}, \ldots, \varphi_{p-1})$ in figure 6. A reasonable prediction for the next value, φ_n, is then the value, φ_p, that followed $(\varphi_{p-d}, \varphi_{p-d+1}, \ldots, \varphi_{p-1})$. If we have more matching patterns, such as z_r in figure 6, we can use any of these, or take a suitable average. Takens' theorem justifies this intuitively appealing idea, and gives it a precise mathematical formulation.

If we want to predict further into the future than one time-step, we can either iterate (7.1) for however many steps we want, or attempt a direct fit of φ_{n+i} as a function of $\varphi_{n-d}, \varphi_{n-d+1}, \ldots, \varphi_{n-1}$. However, we must realize that if the underlying dynamics is chaotic, we will be limited in how far ahead we can predict. This is because any real data will be subject to at least some noise and any real function fit will not be exact. The resulting errors will be amplified by a chaotic process leading to a rapid increase in forecasting errors the further ahead we attempt to predict.

8. Other applications

The ability to construct a predictive model from observed data opens up a range of other possibilities. Some of the most exciting are novel algorithms for noise reduction and signal detection or separation. In real life, we rarely encounter a perfect time-series obtained from a completely deterministic system. Instead, we are often presented with a mixture ${}_n = \varphi_n + \varepsilon_n$ of a dynamical time-series φ_n and some other signal ε_n. The latter may represent noise, in which case we want to remove it from ${}_n$, or it may be a signal that we wish to detect, in which case we want to extract it from ${}_n$ and discard φ_n. An example of the latter might be a foetal electrocardiogram (ECG) signal ε_n masked by a much stronger maternal signal φ_n (see, for example, Schreiber & Kaplan (1996); in this particular problem, ε_n is also deterministic, but with significantly different dynamics, which is all that matters). In both cases, the mathematical issue that we face amounts to separating ${}_n$ into its two components φ_n and ε_n.

Starting with the pioneering work of Kostelich & Yorke (1988), a number of algorithms have appeared that can, under appropriate circumstances, perform such a separation (see, for example, Ott *et al*. 1994). Ultimately, these all rely on the ability to make predictions, and, hence, rely on Takens' theorem. Thus if ε_n is absent, equation (7.1) will give good predictions, while if ε_n is large, then we expect our prediction errors also to be large. These errors can be further processed to recover the actual form of ε_n.

The observant reader might point out that this is all very well if we know G, or have a 'clean' sample of φ_n from which we can estimate it. But what if the only data available are the contaminated time-series ${}_n$, as in the foetal ECG example? Fortunately, it turns out that as long as ε_n is not too large relative to φ_n, we can often still estimate G from ${}_n$ (see, for example, Davies & Stark 1994). This allows these ideas to be applied increasingly to practical signal-processing problems, potentially achieving results far superior to those possible using conventional signal-processing approaches.

A closely related problem, which has received somewhat less attention, yet, perhaps, is of far greater practical significance, is that of detecting shifts in the under-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

lying dynamical system. Thus, suppose that this system is slowly undergoing some change, or perhaps is subject to some sudden minor variation that presages a more dramatic and perhaps even catastrophic later event. Can we recognize this simply by observing the time-series? In other words, in terms of real applications, what can we deduce about fundamental changes to the global climate by observing a time-series of the temperature? What can we learn about the possible damage to an ecosystem by observing a few species? In entirely different contexts, can we identify subtle dynamical changes in an ECG that herald a heart attack? Can we predict the imminent collapse of an electricity supply network by monitoring subtle changes in voltage variations?

Attempts at answering these kinds of questions in the context of nonlinear dynamics have initially been made using some of the characterizations described earlier, such as fractal dimensions. The basic idea is to compute these quantities from different segments of the time-series, and if the resulting values are significantly different one can conclude, with some confidence, that the underlying dynamics has changed. Unfortunately, this approach has met with only limited success, due to the difficulty of computing the necessary values from limited amounts of data. It should be possible to derive more sensitive techniques based on prediction and signal separation, but, hitherto, there has been little systematic development along these lines. Very similar methods can also be used to determine whether two time-series are obtained from two different observables of the same dynamical systems, or whether they are generated by two completely different systems. This again may have some potentially interesting applications, which need further exploration.

Finally, another area that has received considerable attention in the last decade is that of 'chaotic control'. Introduced by Ott *et al.* (1990), it relies on the extreme sensitivity of chaotic systems to initial conditions to control such systems using only very small control signals. Intuitively, this is a very appealing approach to the control of complex systems. A similar idea was already used in the early 1980s to steer the ICE spacecraft to a rendezvous with the comet Giacobini–Zinober using the least possible amount of fuel (Farquhar *et al.* 1985). In order to implement this kind of technique, one needs a state space model of the system to be controlled. While this is not a problem in the case of spacecraft, in most other practical applications such a model is not directly available. Thus, once again, one has to rely on Takens' theorem to provide a reconstruction. In the last decade, a variety of practical algorithms adopting this approach has been successfully demonstrated in applications ranging from lasers to cardiac rhythms (see, for example, Ott *et al.* 1994).

9. Noise and external influences

The last few sections have demonstrated the fundamental significance of Takens' theorem to a wide range of applications. However, as already indicated, this theorem really only applies to a highly idealized class of mathematical models. More specifically, it assumes that the system and its observations are unaffected by noise or by outside events. Few real systems satisfy such conditions. Despite this, techniques based on the theorem have successfully been applied to many real problems, in other words, in circumstances where the theorem strictly speaking is not valid. From a practical point of view, this might seem to be unimportant. After all, who cares whether or not the hypotheses of an abstract mathematical theorem are satisfied, as long as the methods motivated by it work in practice? However, in the absence of a theoretical framework relevant to realistic systems, we have little insight into how and why the method of delays actually works and how much information it preserves. Furthermore, we do not know what its fundamental limits are and, hence, when it is likely to fail. Finally, in those applications where it does fail, we have little to guide us in how it might be successfully extended.

Fortunately, it turns out to be possible to generalize Takens' result to a much wider class of systems (Stark *et al.* 1997; Stark 1999), including both noisy systems and systems subject to external effects. Such effects can be deterministic, or essentially unpredictable. The former include deliberate periodic forcing as found in many laboratory experiments, or the inevitable seasonal forcing of measles epidemics or many ecosystems. The latter applies particularly to man-made systems: thus, in signal processing, the main role of filters is to process arbitrary input sequences; while systems such as telecommunications, traffic and energy distribution networks or stock markets and other financial institutions are subject to a variety of more-or-less arbitrary irregular shocks. This class of systems was named *input–output systems* by Casdagli (1992), who conjectured that a generalized form of Takens' theorem applies in the case where the input is a single variable time-series.

All of the above cases can be unified in a single mathematical framework. In particular, a very wide class of noisy systems can be thought of as deterministic systems driven by some random process. A particularly convenient way of presenting this framework is to make the function f that specifies the dynamical evolution depend on a vector of parameters ω, and to make a new choice ω_n of these parameters at each time-step n. The state of the system, x_n, thus evolves according to

$$x_{n+1} = f(x_n, \omega_n).$$

This is in contrast to the standard framework, where we had $x_{n+1} = f(x_n)$. Often, it is helpful to write $f_{\omega_n}(x_n)$ instead of $f(x_n, \omega_n)$. This suggests the interpretation that instead of applying the same function f every time, we choose a different function f_{ω_n} at each time-step. Both points of view are equally valid, and equally useful in different contexts. In terms of applications, ω_n might, for instance, denote the time of year in a model of a measles epidemic, or represent several million TV viewers switching on a kettle immediately after a favourite TV programme in a model of an electrical distribution network, or perhaps be a binary sequence encoding some message being processed by an electronic filter. The observed time-series is derived in exactly the same way from x_n as in the deterministic case, that is $\varphi_n = \varphi(x_n)$ (more generally, we might also incorporate noise or forcing on the observations, but this makes little fundamental difference).

Remarkably, it is possible to extend Takens' theorem to time-series derived in this way, and, hence, to reconstruct the original dynamics from the observed data. The most difficult case is, strangely enough, that of deterministic forcing, that is where ω_n is itself generated by some deterministic forcing system (see periodic forcing above). In such a case, one reconstructs both the forced and the forcing dynamics simultaneously (Stark 1999). This gives a time-series model of the form (7.1). On the other hand, when ω_n is either a noise sequence, or an arbitrary input sequence, we make no effort to reconstruct ω_n, since such an attempt would automatically be doomed to failure. The reconstructed dynamics takes the form

$$z_{n+1} = F(z_n, \omega_{n-d}, \omega_{n-d+1}, \ldots, \omega_{n-1}),$$

so that whereas f depends only on the current input ω_n, the reconstructed map F involves a segment of its past history (Stark *et al*. 1997). This gives a time-series model of the form

$$\varphi_n = G(\varphi_{n-d}, \varphi_{n-d+1}, \ldots, \varphi_{n-1}, \omega_{n-d}, \omega_{n-d+1}, \ldots, \omega_{n-1}). \tag{9.1}$$

Where the input sequence is known, this provides an explicit method for incorporating it in the modelling procedure. The practical applications of this are only just beginning to be explored (see, for example, Richter & Schreiber 1998). One that is worth highlighting is that of irregular sampling. This is impossible for the standard framework to handle, while here we simply let ω_n be the time between successive observations, and include these in the estimation of G, as in equation (9.1).

Where the input sequence is unknown, e.g. when it represents noise, it has to be estimated from the data. Although statistical methods exist for doing this, they are far from simple, and a great deal of further work in this direction is required. Nevertheless, at least the theorem presented here provides a starting point, and a mathematical justification for attempting such an enterprise.

10. Multivariable observations

Up until now, we have only discussed scalar time-series, produced by a single variable observable. In many applications, however, we are increasingly able to measure several observables simultaneously. In such a case, $\varphi(x)$ becomes a vector and $\varphi(x_0), \varphi(x_1), \ldots, \varphi(x_n), \ldots$, is then a *multivariable time-series*. In the context of our earlier examples, the components of $\varphi(x_n)$ might, for instance, specify the positions of a number of different planets, or the population of a number of different species, at each specified time. A more extreme case would be a digital video recording of some complex process, where the components of $\varphi(x_n)$ would be the intensities of the primary colours in each of the pixels making up the nth frame.

The extension of Takens' theorem to multiple observables poses few problems, though as far as we are aware, an explicit statement and proof of such a theorem has never been published. Nevertheless, it is clear to anyone familiar with the standard theorem that the modifications required are minimal. The same holds for the generalizations presented in § 9. In fact, increasing the number of independent observations actually makes life simpler, to the extent that if we have more than $2m + 1$ independent observables, there is no need to resort to any delays and we can guarantee that reconstruction is possible by the Whitney embedding theorem, a classical result in differential topology. Furthermore, if we have a sufficient number of independent measurement functions, we can even reconstruct the noise ω_n (Muldoon *et al*. 1998).

11. Spatio-temporal systems

More serious issues arise, however, when we consider more complex dynamical systems, where, intuitively, it seems that multiple observations might play a more significant role. A particularly important class is that of spatio-temporal systems. By this, we mean systems in which spatially distinct parts can evolve in time in distinct ways. We can thus think of such systems as consisting of many local dynamical systems, one at each spatial location, coupled together into one large system. Classically, such systems would be modelled by partial differential equations, such as the Navier–Stokes equation for fluid flow or reaction–diffusion equations for pattern formation in a reacting medium. In such models, space is represented as a continuum. Increasingly, however, there is interest in using discrete spatial locations. If these are arranged in a regular structure, one obtains coupled-map lattices and related systems, while in other cases, it may be more appropriate to use arbitrary arrangements leading to various kinds of networks (e.g. in transport or telecommunication systems).

The spatial aspect of these systems dramatically increases their structural complexity compared with the localized dynamical systems that we have been considering up to now. This leads to a much richer repertoire of possible behaviour, though not necessarily always to more complex behaviour. It also provides the flexibility to model a much larger variety of real systems, so that spatio-temporal systems are widely encountered in many applications, such as fluid flow, biological pattern formation, neuroscience, ecology, road and telecommunication traffic, and many others.

In some cases (e.g. weather prediction), there is a reasonable understanding of the underlying deterministic mechanisms governing the evolution of the system, and we can use this to construct an *a priori* state space model to which observations can be fitted. In many other situations, particularly in the biological and social sciences, this is not possible and we need to reconstruct the unknown dynamics from observed data in a similar fashion to that described above for localized systems.

Although Takens' theorem implies that in principle it might be possible to do this using a sufficient number of delays of a single observable, common sense suggests that this is unlikely to be feasible in practice. In particular, the amount and quality of data obtained by observing a single spatial location will almost certainly be insufficient to reconstruct the behaviour of the whole spatially extended system. Any serious approach to the reconstruction of spatio-temporal systems will, therefore, inevitably use multiple measurements, distributed in an appropriate way around the system. Delay reconstruction techniques are increasingly being applied to such data (see, for example, Muldoon *et al.* 1994; Little *et al.* 1996; Rand & Wilson 1997; Ørstavik & Stark 1998; Cao *et al.* 1998). However, since there is almost no theoretical framework in this case, there is little understanding of the basic principles of operation of these methods, of the properties of the underlying system that they preserve, or of the theoretical limits to their performance. Developing the necessary theory to address these issues is a major undertaking that raises a wide variety of both mathematical and practical issues.

Perhaps the most important of these is the question of what exactly we aim to reconstruct. The standard embedding approach requires the system that we wish to reconstruct to be autonomous, i.e. free of any outside influences. Since, in a non-trivial spatio-temporal system, the local subsystems interact with each other, the conventional framework obliges us to reconstruct the whole spatially extended system. This has a serious drawback, however. The state space of the whole system is typically very large and the dynamics depends on a huge number of variables. Attempting to reconstruct this results in extremely complex models and a large embedding dimension d. Such models typically do not perform well in practice; in particular, they suffer from the so-called 'curse of high dimensionality'. This refers to the fact that as d increases, we need more and more data to adequately sample the reconstructed state space.

This can be best illustrated in the context of forecasting by analogy, as in figure 6. If d is small, and we are given a particular current pattern z_n, then it is likely that we will find many similar patterns in our historical dataset. We should thus be able to make good predictions. On the other hand, suppose d is large, say of the order of 100, which is not unreasonable for a typical spatio-temporal system. Then, given a particular pattern z_n (which now consists of 100 successive values of the time-series), it is highly unlikely that we shall find a close match in our dataset. We thus have nothing on which to base our forecasts. Essentially the same problem arises whatever prediction method we use, no matter how clever. Thus, for instance, Ørstavik & Stark (1998) find that the best predictions are obtained for a very small embedding dimension ($d = 4$). Increasing either the number of measurements functions, or the number of delays, gives no advantage. In other words, when we are restricted to using observed data, the best model for the data may be one that is much simpler than the system that originally generated the data.

To resolve this apparent paradox, note that in most spatio-temporal systems a given subsystem will be influenced far more strongly by some subsystems than by others (typically, nearby subsystems will have a greater effect; see, for example, Ruelle (1982) and Carretero-González *et al*. (1999*a*)). We may thus be able to make reasonable predictions of the behaviour in a given region by only considering the dynamics in that region, and ignoring the rest of the system. Indeed, it turns out that in some examples if we only observe one locality we cannot distinguish a large extended system from a small local system driven by a simple random process (Carretero-González *et al*. 1999*b*). In such a case, we can replace the effect of all the remote components of a spatio-temporal system by noise. This suggests that some of the theoretical models being developed to study spatio-temporal behaviour may be too complex to be used in the context of observed data. In particular, we may be unable to observe the difference between a complex deterministic model and a much simpler noisy model. Hence, although, in principle, the deterministic model is capable of making perfect predictions (i.e. in the case of an infinite amount of data), in practice, the simple model may be preferable.

At least informally, we can use the stochastic version of Takens' theorem described above to provide a reconstruction framework for the noisy local system. The variable ω_n now represents those parts of the system we choose not to reconstruct. A crucial question in such an approach is how to determine the size of the local system, in other words, what behaviour to assign to the deterministic dynamics x_n and what to assign to the random process ω_n? This exemplifies the trade-off between simplicity and complexity that lies at the heart of the issues examined in this section. If we make the local system too small, it fails to capture the phenomena we are interested in; if we make it too large, we run into the problems described above.

Finally, we consider how to approach the problem of predicting the evolution of the whole system, rather than just a localized part. If we attempt to do this using a single large model, we encounter the 'curse of high dimensionality', just as above. A possible alternative, hitherto untried, is to predict each localized subsystem independently, using the ideas outlined above. In many cases, we have reason to suppose that the subsystems at each locality should be identical, in which case we can fit a single model using observations from all the different locations. This has the advantage of substantially increasing the amount of data available to fit the model.

Much remains to be done in order to make these ideas rigorous, to explore their practical ramifications, and to develop and evaluate the resulting algorithms. We believe, however, that even the most tentative results in these directions would be extremely valuable, greatly enhancing our ability to devise simple models for complex data.

12. Conclusion: seeing simplicity

Nonlinear dynamical systems provide a broad framework for the analysis of many complex phenomena, ranging from simple electrical circuits to whole ecosystems. Often, the structure of such systems is far simpler than the observed data they can generate. In practice, it is rare to have direct access to such structure, and our knowledge of a system is essentially restricted to the observed data. This may make some phenomena appear more complicated than they really are. Takens' theorem allows us to reconstruct the structure of a system from repeated observations, and, hence, offers the possibility of finding simple explanations for apparently complicated behaviour. We have outlined this theorem, its more important consequences, and described some of its range of applications.

Recent extensions of the theorem allow us to incorporate noise and other external influences in our model. At least informally, these allow us to apportion observed behaviour between a deterministic component and a random component. We have argued that if the deterministic part is too complex, then it cannot be distinguished using a reasonable amount of measured data. This suggests that no matter how complicated a system is, the process of observation may restrict us to seeing only a limited part of its complexity, the remainder inescapably appearing as random noise.

Many people have contributed to the development of the ideas discussed here. I would particularly like to thank David Broomhead, Jerry Huke, Sakse Ørstavik and Ricardo Carretero-González. I would also like to express my gratitude to the Royal Society and the Leverhulme Trust for the award of a Senior Research Fellowship, during which much of the work on spatio-temporal systems described in this paper was carried out, and to the EPSRC for current financial support under grant GR/L42513.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY A
PHILOSOPHICAL TRANSACTIONS OF

References

Abarbanel, H. D. I. 1995 *Analysis of observed chaotic data*. Springer.

Aeyels, D. 1981 Generic observability of differentiable systems. *SIAM J. Control Optimization* **19**, 595–603.

Cao, L., Mees, A. & Judd, K. 1998 Dynamics from multivariate time series. *Physica* D **121**, 75–88.

Carretero-González, R., Ørstavik, S., Huke, J., Broomhead, D. S. & Stark, J. 1999*a* Scaling and interleaving of sub-system Lyapunov exponents for spatio-temporal systems. *Chaos* **9**, 466–482.

Carretero-González, R., Ørstavik, S., Huke, J., Broomhead, D. S. & Stark, J. 1999*b* Thermodynamic limit from small lattices of coupled maps. *Phys. Rev. Lett.* **83**, 3633–3636.

Casdagli, M. 1992 A dynamical systems approach to modelling input–output systems. In *Nonlinear modelling and forecasting* (ed. M. Casdagli & S. Eubank). Reading, MA: Addison-Wesley.

Davies, M. E. & Stark, J. 1994 A new technique for estimating the dynamics in the noise-reduction problem. In *Nonlinearity and chaos in engineering dynamics* (ed. J. M. T. Thompson & S. Bishop), pp. 329–344. Wiley.

Farquhar, R., Muhonen, D. & Church, L. C. 1985 Trajectories and orbital maneuvers for the ISEE-3/ICE comet mission. *J. Astronaut. Sci.* **33**, 235–254.

Kantz, H. & Schreiber, T. 1998 *Nonlinear time series analysis*. Cambridge University Press.

Kostelich, E. J. & Yorke, J. A. 1988 Noise reduction in dynamical systems. *Phys. Rev.* A **38**, 1649–1652.

Little, S., Ellner, S., Pascual, M., Neubert, M., Kaplan, D., Sauer, T., Caswell, H. & Solow, A. 1996 Detecting nonlinear dynamics in spatio-temporal systems. Examples from ecological models. *Physica* D **96**, 321–333.

Lorenz, E. N. 1963 Deterministic non-periodic flows. *J. Atmos. Sci.* **20**, 130–141.

Muldoon, M., Broomhead, D. S. & Huke, J. P. 1994 Delay reconstruction for multiprobe signals. *IEE Digest* **143**, 3/1–3/5.

Muldoon, M. R., Broomhead, D. S., Huke, J. P. & Hegger, R. 1998 Delay embedding in the presence of dynamical noise. *Dyn. Stab. Sys.* **13**, 175–186.

Ørstavik, S. & Stark, J. 1998 Reconstruction and cross-prediction in coupled map lattices using spatiotemporal embeddings. *Phys. Lett.* A **247**, 145–160.

Ott, E., Crebogi, C. & Yorke, J. A. 1990 *Controlling chaos*. *Phys. Rev. Lett.* **64**, 1196–1199.

Ott, E., Sauer, T. & Yorke, J. A. 1994 *Coping with chaos*. Wiley.

Packard, N., Crutchfield, J., Farmer, D. & Shaw, R. 1980 Geometry from a time series. *Phys. Rev. Lett.* **45**, 712–715.

Rand, D. A. & Wilson, H. B. 1997 Using spatio-temporal chaos and intermediate-scale determinism to quantify spatially extended ecosystems. *Proc. R. Soc. Lond.* B **259**, 111–117.

Richter, M. & Schreiber, T. 1998 Phase space embedding of electrocardiograms. *Phys. Rev.* E **58**, 6392–6398.

Ruelle, D. 1982 Large volume limit distribution of characteristic exponents in turbulence. *Comm. Math. Phys.* **87**, 287–302.

Sauer, T., Yorke, J. A. & Casdagli, M. 1991 Embedology. *J. Stat. Phys.* **65**, 579–616.

Schreiber, T. & Kaplan, D. T. 1996 Signal separation by nonlinear projections: the fetal electrocardiogram. *Phys. Rev.* E **53**, R4326–R4329.

Stark, J. 1999 Delay embeddings of forced systems. I. Deterministic forcing. *J. Nonlinear Sci.* **9**, 255–332.

Stark, J., Broomhead, D. S., Davies, M. E. & Huke, J. 1997 Takens embedding theorems for forced and stochastic systems. *Nonlinear Analysis* **30**, 5303–5314.

Takens, F. 1980 Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick* (ed. D. A. Rand & L.-S. Young). Lecture Notes in Mathematics, no. 898. Springer.

AUTHOR PROFILE

J. Stark

Jaroslav Stark graduated in mathematics from the University of Cambridge in 1982. He obtained a PhD from the Mathematics Institute at the University of Warwick in 1986, staying on for another year as a Research Assistant at the newly formed Nonlinear Systems Laboratory. This was followed by a SERC Fellowship at Imperial College and then four years at the GEC–Marconi Hirst Research Centre. He joined the Centre for Nonlinear Dynamics and its Applications at University College London in 1992, being promoted to Professor in 1999. He held a Royal Society Leverhulme Trust Senior Research Fellowship in 1997–98 and has been joint Editor-in-Chief of *Dynamics and Stability of Systems* since 1995. Aged 39, his research interests include nonlinear dynamics and its applications, particularly to the life sciences.