

# Time-Series Analysis

---

## 6.1 STARTING WITH DATA

Up until now we have examined mathematical descriptions of dynamical systems and seen how different types of behavior can be generated, such as fixed points, limit cycles, and chaos. The goal of applied dynamics is to relate these mathematical systems to physical or biological systems of interest. The approach we have taken so far is model building—we use our understanding of the physical system to write dynamical equations. For example, we used our understanding of the interaction of predators and prey to motivate the Lotka-Volterra equations. These equations then suggested the types of dynamics we were likely to observe in the field, such as population oscillations around a fixed point, or extinction.

In this chapter, we shall take the opposite approach. Starting with a sequence of measurements—a **time series**—we want to see what the data themselves can tell us about the dynamics. In particular, we will introduce some tools from time-series analysis (often termed **signal processing**) that can sometimes be used to suggest what types of equations are appropriate, or to compare the predictions made by mathematical models to measurements made in the field.

The ultimate goal for time-series analysis might be to construct a computer program that, without any knowledge of the physical system from which the data come, can take the measured data as input and provide as output a mathematical model describing the data. This can be done with current technology (see Section 6.7), but the method has a severe shortcoming: The resulting mathematical model generally does not have identifiable components that can be given physical meaning. Thus, it is not possible to use such data-generated mathematical models to determine the effect of changing some aspect of the physical system, which is often the motivation for studying dynamics in the first place.

In practice, the approach that is taken is a combination of model building and time-series analysis. Model building based on our knowledge of the physical system is used to suggest what features to look for in the data; time-series analysis is used to detect and quantify these features or to refute their existence, thus motivating changes in the model.

In this chapter, we shall mimic this process; a series of models will be proposed, data will be generated from these models, and time-series analysis techniques will be introduced to show how the models and data can be related to one another. The choice of models here is intended to illustrate various aspects of time-series analysis and does not include the physical and biological information that would motivate realistic models of specific phenomena.

## 6.2 DYNAMICS, MEASUREMENTS, AND NOISE

In the previous sections of this book, we have dealt extensively with dynamics. By now, we are familiar with equations of the form

$$x_{t+1} = f(x_t)$$

and

$$\begin{aligned}\frac{dx}{dt} &= g(x, y) \\ \frac{dy}{dt} &= h(x, y).\end{aligned}$$

The functions  $f()$ ,  $g()$ , and  $h()$  govern the dynamics of the systems, and given the functions, we know how to look for dynamical behavior such as fixed points, cycles, and chaos.

When dealing with data, we need to introduce two new concepts: **measurement** and **noise**.

In conducting an experiment or making measurements in the field, we can measure only a limited set of quantities and are able to make those measurements

with limited precision. For example, an ecologist studying predator–prey dynamics might be able to count the population of the predator only, even though it is clear from models such as the Lotka-Volterra equations that both the predator and prey play a role in the system dynamics.

When constructing a mathematical model of observed dynamics, it is essential to include an equation that describes how the actual measurements are related to the dynamical variables. For instance, in the Fitzhugh-Nagumo model of nerve cell dynamics (Eq. 5.29), the transmembrane voltage  $v$  is usually measured in experiments, while the recovery variable  $w$  cannot be measured directly. In this chapter, an additional equation will be added to dynamical models, describing how the measurement at time  $t$ , denoted as  $D_t$  or  $D(t)$ , is related to the variables in the dynamical system.

The measurements approximate the true dynamical variables; the difference between the two is called the **measurement error**. The measurement error arises from several factors: systematic bias, measurement noise, and dynamical noise.

**Systematic bias** results from a flaw in the measurement process. For instance, suppose one tried to measure the use of a university's library by counting the number of students in the library just before exams at the end of the semester. Such a count would probably seriously overestimate library usage over the course of a year. Such systematic bias will not be discussed further here.

**Measurement noise** refers to fluctuations in measurements that arise from chance. Even if there were a well-defined average level of library use, the number of students at any particular moment would likely differ from this average.

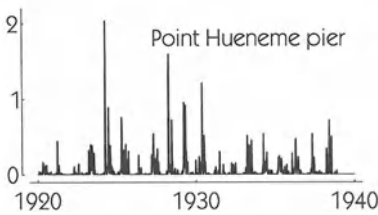
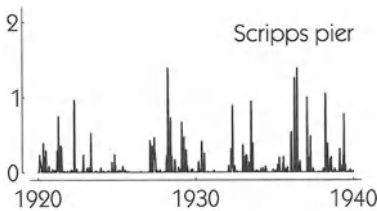
**Dynamical noise** is another important source of noise in data. Real-world systems do not exist in isolation. They are affected by outside influences. For example, the population of prey depends not just on the population of predators, but also on environmental variables such as the temperature and precipitation, which themselves fluctuate. One would like to include such outside influences in dynamical models. This is often done by regarding the outside influences as random noise that affects the dynamical variables.

## DYNAMICS IN ACTION

### 16 FLUCTUATIONS IN MARINE POPULATIONS

In order to study the dynamics of phytoplankton, marine biologist W. E. Allen made daily measurements from 1920 to 1939 of the total number of diatoms per liter

of water off of two piers in California, at the Scripps Institute of Oceanography and Point Hueneme. These impressive data sets are displayed in the figure. Like all measurements, there are flaws and shortcomings in this data. There is certainly a component of measurement noise here: Not every liter of water contains the same number of diatoms. Systematic biases in the measurements may also exist. Important dynamical variables, such as the number of organisms that eat phytoplankton, were not measured.



Weekly averages of daily counts of total number of diatoms (phytoplankton) (millions of cells per liter) at the Scripps and Point Hueneme piers, California, 1920–1939, collected by W. E. Allen. Data from Tont (1986).

There are many outside influences that affect the dynamics: the amount of sunlight, the water temperature, and the amount of nutrients in the water. These were not measured. Even if these variables had been measured at the piers, the fact that ocean currents carry phytoplankton from place to place makes it unclear how to interpret measurements made in a single place.

## GAUSSIAN WHITE NOISE

A source of random numbers with which everyone is familiar is a deck of cards. Imagine that you have a very large deck of cards and that each card has a number from  $-1$  to  $1$  written on it. The deck has been thoroughly shuffled so that the cards are in random order. Each card that you draw from the deck tells you virtually nothing about either the previous cards that were drawn or the subsequent cards yet to be drawn. In this situation, the drawn cards are said to be **independent** of one another. The resulting numbers are said to be “drawn from

a distribution” of numbers. Here the distribution is the set of all the numbers in the deck.

It is easy to imagine a situation in which numbers drawn from a distribution would not be independent of one another. Suppose the cards in the deck were sorted in ascending order. Then each card would give you a good idea of what the value on the following card would be. When the random numbers drawn from the deck are independent of one another—when the deck is shuffled randomly—the numbers form a source of **white noise**.

White noise is often a good model of measurement and dynamical noise. But what is the distribution from which the white noise is drawn? It might seem that the distribution will depend on details of the system being studied, but for reasons described ahead, it happens that a very commonly encountered distribution in practice is the **Gaussian distribution**.

The random variability in a measurement or a random outside influence is often the sum of many different types of random variability. For example, in measuring the population of flies in a field, there are many potentially random events: the number of flies that happen to be near the capturing net, the temperature and wind velocity at the time the measurement was made (which influences the number of flies who are up and about), and so on. Careful experimental design can minimize the influence of such factors, but whichever ones remain often tend to add up.

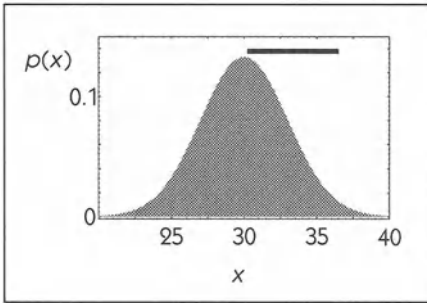
In terms of the deck-of-cards analogy, this means that each measurement error or outside perturbation is not a single card drawn from a deck, but instead results from drawing several cards at once and adding up the numbers on the cards. *Dynamics in Action 7* describes a **random walk**, a process in which independently drawn random numbers are added up to give a final result. As seen in Appendix A, the probability distribution for a random walk is the bell-shaped **Gaussian distribution** shown in Figure 6.1.

$$p(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x - M)^2}{2\sigma^2} dx. \quad (6.1)$$

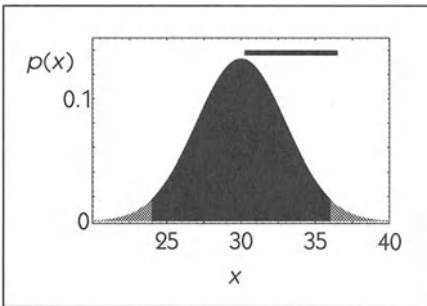
$M$  and  $\sigma$  are constants:  $M$  is the **mean** value, and  $\sigma$  is called the **standard deviation**.

Equation 6.1 is to be interpreted in the following way: The probability that a value drawn from a Gaussian distribution will fall into the range  $x$  to  $x + dx$  is  $p(x)dx$  when  $dx$  is small.  $p(x)$  is called the probability density. If we want to know the probability of noise falling in a larger range, it is necessary to calculate the integral—the probability that the noise is in the range  $a \leq x \leq b$  is

$$\int_a^b p(x)dx. \quad (6.2)$$



**Figure 6.1** A Gaussian probability density, Eq. 6.1, with mean  $M = 30$  and standard deviation  $\sigma = 3$ . The black bar has length  $2\sigma$ .



**Figure 6.2** The probability of a single measurement falling into a specified range is the integral of the probability density over that range. The probability of the measurement falling in the range  $[M - 2\sigma, M + 2\sigma]$  is the area shown in black, which is approximately 0.95.

**Table 6.1** The probability that a single measurement, drawn from a Gaussian distribution with mean  $M$  and standard deviation  $\sigma$ , falls into the indicated interval.

Interval			Prob.
$M - 0.5\sigma$	to	$M + 0.5\sigma$	0.383
$M - \sigma$	to	$M + \sigma$	0.683
$M - 1.5\sigma$	to	$M + 1.5\sigma$	0.866
$M - 2\sigma$	to	$M + 2\sigma$	0.954
$M - 2.5\sigma$	to	$M + 2.5\sigma$	0.988
$M - 3\sigma$	to	$M + 3\sigma$	0.997

The integral in Eq. 6.2 is so important in practice that tables of its values are widely published. Using one of these tables, such as Table 6.1, we can see that the probability of the noise falling into the range  $M - 2\sigma$  to  $M + 2\sigma$  is roughly 0.954, or about 95 percent. See Figure 6.2.

---

➔ **MODEL ONE**

The behavior of the finite-difference equation

$$x_{t+1} = A + \rho x_t \quad (6.3)$$

is easily studied with the methods presented in Chapter 1. There is a steady state at

$$x_t = A/(1 - \rho) = M$$

that is stable if  $|\rho| < 1$ , which is the case we shall assume here. (We use the variable  $M$  as shorthand for  $\frac{A}{(1-\rho)}$ .) The solution to the finite-difference equation is exponential decay to the steady state: After the transient passes, we have steady-state behavior  $x_t = M$ .

For simplicity, we will assume that a direct measurement of the dynamical variable  $x_t$  is made, but since there is measurement noise the measurement at time  $t$  is

$$D_t = x_t + W_t, \quad (6.4)$$

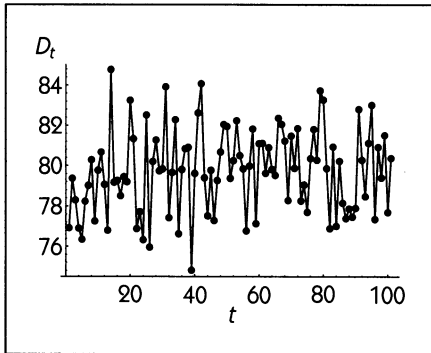
where  $W_t$  is a random number drawn independently at each  $t$  from a Gaussian probability distribution with a mean of zero and standard deviation  $\sigma$ .

Figure 6.3 shows data  $D_t$  generated from this model, with  $A = 4$ ,  $\rho = 0.95$ , and consequently  $M = \frac{A}{(1-\rho)} = 80$ .  $W_t$  is Gaussian white measurement noise with a standard deviation of  $\sigma = 2$ .

This model might serve as a description of a system where there is some quantity (e.g., population level or amount of a circulating hormone) that is maintained at a steady level. The model assumes that no outside perturbation affects  $x_t$ —the dynamics of the model are completely trivial once the transient has died out: steady state.

Using the model as a motivation in interpreting measured data, we might ask the following questions:

- What is the value of the steady state in the data?
- What is the level of measurement noise in the data?



**Figure 6.3**

Data  $D_t$  from the model  $x_{t+1} = A + \rho x_t$  with measurement  $D_t = x_t + W_t$ .  $A = 4$ ,  $\rho = 0.95$ , and the standard deviation of  $W_t$  is  $\sigma = 2$ .

We also might want to decide if the model is good for describing the measured data:

- Is there evidence that there really is a steady state?
- Is there evidence that there is only measurement noise and no outside perturbations to the state  $x_t$ ?

□

### 6.3 THE MEAN AND STANDARD DEVIATION

We make a series of measurements, as in Figure 6.3 and we have a model in mind such as Model One, which suggests that the system is at a stable steady state. How do we estimate the value  $M$  of this steady state from the measurements? Intuition tells us that we should **average** all the  $N$  measurements  $D_1, D_2, \dots, D_N$  rather than take just a single measurement, say  $D_7$ , as our estimate of  $M$ . Because we cannot measure  $M$  directly, but rather estimate it from  $D_t$ , we will denote the quantity we estimate as  $M_{\text{est}}$ . Although  $M$  depends only on the dynamical equation 6.3 and—according to the model—is constant,  $M_{\text{est}}$  may vary depending on how many data points  $D_t$  we collect and on when they are collected.

To see where the idea of averaging comes from, consider trying to find the value  $M_{\text{est}}$  that is closest to all of the measurements  $D_1, \dots, D_N$ . We take the separation between  $M_{\text{est}}$  and  $D_t$  to be  $(D_t - M_{\text{est}})^2$ . To make  $M_{\text{est}}$  as close as possible to all the measurements, we minimize the total separation  $E$ ,

$$E = \sum_{t=1}^N (D_t - M_{\text{est}})^2. \quad (6.5)$$



To perform the minimization, take  $\frac{dE}{dM_{\text{est}}}$  and set it equal to zero (remember, we're trying to find the value of  $M_{\text{est}}$  that gives the smallest value of  $E$ ):

$$\frac{dE}{dM_{\text{est}}} = 0 = 2 \sum_{t=1}^N (D_t - M_{\text{est}}). \quad (6.6)$$

Rearranging the right-hand side of Eq. 6.6, we find

$$M_{\text{est}} = \frac{1}{N} \sum_{t=1}^N D_t. \quad (6.7)$$

This is the familiar formula for averaging.  $M_{\text{est}}$  is termed the **sample mean** of the set of measurements  $D_t$ .

### STANDARD DEVIATION

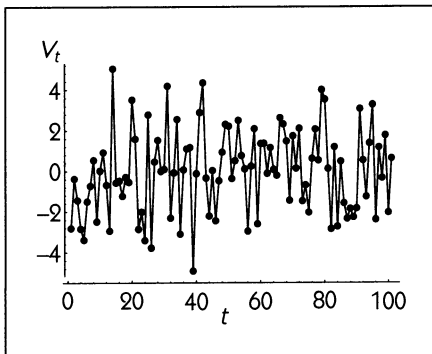
By calculating the mean of the measured data, we now have an estimate,  $M_{\text{est}}$ , of the value of the steady state  $M$ . We are now interested in the fluctuations  $V_t$  of the measurements around the mean,

$$V_t = D_t - M_{\text{est}}.$$

Model One interprets these fluctuations as noise. One of the goals of time-series analysis of the Model One data is to assess the validity of this interpretation.

As a first step, we want to characterize the size of the fluctuations. One way (which will turn out not to be very useful) is to consider the mean value of the fluctuations:

$$\frac{1}{N} \sum_{t=1}^N V_t = \frac{1}{N} \sum_{t=1}^N (D_t - M_{\text{est}}) = \left( \frac{1}{N} \sum_{t=1}^N D_t \right) - M_{\text{est}} = M_{\text{est}} - M_{\text{est}} = 0$$



**Figure 6.4**

The sample mean  $M_{\text{est}}$  of the data shown in Figure 6.3 is 79.74. Subtracting this value from each data point  $D_t$  gives the fluctuations about the mean,  $V_t = D_t - M_{\text{est}}$ , as plotted here. The standard deviation of these fluctuations is 2.06.

The mean value of the fluctuations is always zero! This isn't so remarkable when we remember that the fluctuations are defined to be the difference between each measurement  $D_t$  and the mean  $M_{\text{est}} = \sum_{t=1}^N \frac{D_t}{N}$ . (The fact that the mean of the fluctuations around  $M_{\text{est}}$  is *always* zero, even though the fluctuations are hypothesized to be random, points out that  $M_{\text{est}}$  is only an *estimate* of the fixed point  $M$ —the fluctuations around  $M$  are unlikely to average out to be exactly zero.)

More useful is the mean value of the square of the fluctuations:

$$\sigma^2 = \frac{1}{N} \sum_{t=1}^N V_t^2 = \frac{1}{N} \sum_{t=1}^N (M_{\text{est}} - D_t)^2. \quad (6.8)$$

$\sigma^2$  is called the **variance**. The square root of the variance,  $\sigma$ , is the **standard deviation**. Note that  $N\sigma^2$  is the same quantity that we minimized in Eq. 6.6 in order to find the mean, so the mean might be defined as “the value that minimizes the variance.”

### STANDARD ERROR OF THE MEAN

Although  $M_{\text{est}}$  is easy to calculate, it is only an estimate of the true mean  $M$ . Why only an estimate? Consider the limiting case where only a single measurement  $D_1$  is made. In this case,  $M_{\text{est}} = D_1$ , and clearly any noise in  $D_1$  is duplicated in  $M_{\text{est}}$ . With two measurements,  $D_1$  and  $D_2$ , there is some chance that the noise will cancel out, but it probably will not cancel out exactly.

Intuition tells us that the more measurements we use in averaging, the better our estimate  $M_{\text{est}}$  will be. We can quantify this intuition. A good way to interpret  $M_{\text{est}}$  is that it is the sum of the true value  $M$  plus some uncertainty,

$$M_{\text{est}} = M + \text{uncertainty}. \quad (6.9)$$

The uncertainty in  $M_{\text{est}}$  comes from averaging the noisy components of the individual measurements. Very often the amplitude of the uncertainty is well described by a Gaussian probability distribution. The standard deviation of this uncertainty is

$$\frac{\sigma}{\sqrt{N}}, \quad (6.10)$$

which is called the **standard error of the mean**. Note that the  $\frac{1}{\sqrt{N}}$  dependence of the standard error of the mean implies that taking more measurements reduces

the uncertainty in the estimate of  $M$ , but that in order to reduce the uncertainty by a factor of 2, one needs to collect four times as much data.

An important assumption that goes into the derivation of the formula for the standard error of the mean given in 6.10 is that the measurements are independent of one another. In the next several sections, we will see various ways to test for such independence. When measurements are not independent, the uncertainty in the estimate of the mean may vary with  $N$  in different ways. An extreme case is that of  $\frac{1}{f}$  noise, described in *Dynamics in Action* 6. For  $\frac{1}{f}$  noise, the variance increases as  $N$  increases, and so the uncertainty in the estimate of the mean increases as more data are collected!

### □ EXAMPLE 6.1

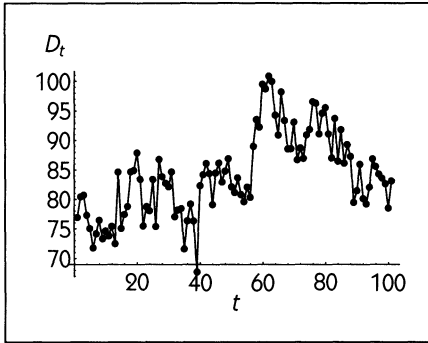
Are the data plotted in Figure 6.3 consistent with Model One? More specifically, does the mean of the data correspond to the theoretical value of the steady state for the parameters used in Model One?

**Solution:** The mean of the 100 data points plotted in Figure 6.3 is found to be  $M_{\text{est}} = 79.74$ , and the standard deviation is 2.06. The theoretical value of the steady state for the parameters used is  $M = \frac{A}{(1-\rho)} = 80$ . So now the question is whether 79.74 is close enough to 80 for us to conclude that the data and the model are consistent.

Since there are 100 data points, the standard error of the mean is  $\frac{2.06}{\sqrt{100}} = 0.206$ . This standard error describes the uncertainty in the estimate of the mean—how much estimated mean might deviate from the true mean just because of chance fluctuations in the data. As a rule of thumb, the difference between a number and  $M_{\text{est}}$  is only statistically significant if the difference is greater than twice the standard error of the mean. (This is only a guideline. A more accurate and precise statement of the meaning of statistical significance is given in statistics textbooks such as Snedecor and Cochran (1989).) In this case, the difference between  $M$ , the theoretical value of the steady state, and  $M_{\text{est}}$  is  $|79.74 - 80| = 0.26$ , which is less than twice the standard error of the mean. Therefore, we conclude that the difference between  $M_{\text{est}}$  and  $M$  is statistically insignificant: The data are consistent with the model. □

### ✈ MODEL TWO

A possible deficiency with Model One is that it does not include any outside influences on the state variable  $x_t$ . For this reason, all the observed variability is modeled as measurement noise.



**Figure 6.5**

Data  $D_t$  from the model  $x_{t+1} = A + \rho x_t + v_t$  with measurement  $D_t = x_t + W_t$ .  $A = 4$ ,  $\rho = 0.95$ . The standard deviation of  $W_t$  is  $\sigma = 2$  and that of  $v_t$  is 3.

A simple way to modify Model One to include outside influences is to write the finite-difference equation

$$x_{t+1} = A + \rho x_t + v_t. \quad (6.11)$$

This incorporates a random influence  $v_t$  on the state variable. As before, the measurement function  $D_t$  will be taken, for simplicity, to be the state variable  $x_t$  itself, plus random measurement noise  $W_t$ ,

$$D_t = x_t + W_t. \quad (6.12)$$

We now have two different sources of noise in the model. We will assume that these two sources are completely independent and that each has its own mean and standard deviation.

Some simulated data from this model, with  $A = 4$ ,  $\rho = 0.95$  (i.e.,  $M = 80$ , as in Model One) are shown in Figure 6.5. Here we again take the standard deviation of  $W_t$  to be 2, and we will assume that the standard deviation of  $v_t$  is 3. The mean of each of the random influences is assumed to be zero.

Some differences between the data from Model One and Model Two can be seen: Model Two produces a much greater range of variability than Model One and shows slow trends, whereas Model One does not.

In interpreting the measured data according to Model Two, we might ask:

- What are the dynamics of movement toward the stable fixed point after an outside perturbation? In particular, can we estimate the time constant of exponential decay,  $\rho$ , from the data?
- How much of the variability in the data is due to measurement error, and how much is due to outside perturbation?



## 6.4 LINEAR CORRELATIONS

One of the assumptions in Model One is that the observed fluctuations around the steady state are a result of white measurement noise; that is, the noise in each measurement is independent of the noise in every other measurement. How can we test the validity of this assumption in the data? If the fluctuations were not independent, how could we quantify their dependence?

So far, we have characterized fluctuations by their mean and standard deviation. These two statistics have an important property: They do not depend on the order in which the data occur. That is, if each measurement was written on its own card, and the stack of cards was shuffled, sorted, or rearranged in any way whatsoever, the mean and standard deviation would remain exactly the same.

As we discussed in Section 6.2, a randomly shuffled deck of cards generates values that are independent of one another. Since the mean and standard deviation are not influenced by the order of cards in the deck, they are of no use in deciding whether fluctuations are independent of each other.

In order to quantify the degree of dependence or independence, consider two limiting cases. Recall that the fluctuations around the mean are denoted  $V_t$ . If the fluctuations are white noise—this is the case of complete independence—then we can model them as

$$V_t = \mathcal{W}_t, \quad (6.13)$$

where  $\mathcal{W}_t$  is white noise. (We write  $\mathcal{W}_t$  instead of  $W_t$  in order to distinguish this model of  $V_t$  from the white noise used in Models One and Two that affected the variable  $D_t$ . You can think of  $W_t$  and  $\mathcal{W}_t$  as different decks of cards.)

At the other extreme,  $V_{t+1}$  might be completely dependent on  $V_t$ , that is,

$$V_{t+1} = f(V_t).$$

Before moving on to nonlinear forms of the function  $f(V_t)$ , we will start here with the simplifying assumption that  $f(V_t)$  is *linear*:

$$V_{t+1} = \rho V_t. \quad (6.14)$$

Combining these two extreme cases of Eqs. 6.13 and 6.14 into one model of the fluctuations, we can write

$$V_{t+1} = \rho V_t + \mathcal{W}_t. \quad (6.15)$$

Note that Eq. 6.15 hasn't been derived from any calculation; it is just a convenient way of writing a model having a parameter  $\rho$  that indicates the degree of dependence (with the assumption of linearity) between  $V_{t+1}$  and  $V_t$ . When  $\rho = 0$ ,  $V_{t+1}$  is independent of  $V_t$ . If  $\rho$  is close to 1, then  $V_{t+1}$  is almost the same as  $V_t$ ; if  $\rho$  is close to  $-1$ , then  $V_{t+1}$  is again almost the same as  $V_t$ , but with a value of the opposite sign. Remember that if  $|\rho| > 1$ , the steady state at  $V = 0$  in Eq. 6.14 is unstable. If the data are not blowing up to  $\infty$ , then the model of Eq. 6.14 must have  $|\rho| < 1$ .

How can we estimate  $\rho$  from measured data? We can take the following approach: Look for a value  $\rho_{\text{est}}$  that makes the square of the difference between  $V_{t+1}$  and  $\rho V_t$  as small as possible—a value that fits the equation  $V_{t+1} = \rho_{\text{est}} V_t$  as closely as possible. We will do this using a least-squares criterion:

$$E = \sum_{t=1}^{N-1} (V_{t+1} - \rho_{\text{est}} V_t)^2. \tag{6.16}$$

Finding the minimum by taking the derivative of  $E$  with respect to  $\rho_{\text{est}}$  and setting this equal to zero, we get

$$\frac{dE}{d\rho_{\text{est}}} = 0 = \sum_{t=1}^{N-1} (V_{t+1} - \rho_{\text{est}} V_t) V_t, \tag{6.17}$$

which implies

$$\rho_{\text{est}} = \frac{\sum_{t=1}^{N-1} V_{t+1} V_t}{\sum_{t=1}^{N-1} V_t V_t}. \tag{6.18}$$

$\rho_{\text{est}}$  is called the **correlation coefficient**.

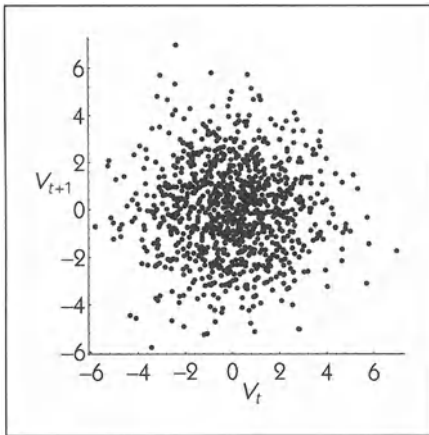
**EXAMPLE 6.2**

In the data from Models One and Two, are the fluctuations around the fixed point consistent with the assumption that they are due to white measurement noise?

**Solution:** We have already found the mean of the data  $D_t$  from Model One to be  $M_{\text{est}} = 79.74$ . The fluctuations around the mean are therefore

$$V_t = D_t - 79.74.$$

Using measured data in the formula for the correlation coefficient in Eq. 6.18, we find that  $\rho_{\text{est}} = -0.0026$ , which is close to zero and therefore consistent with



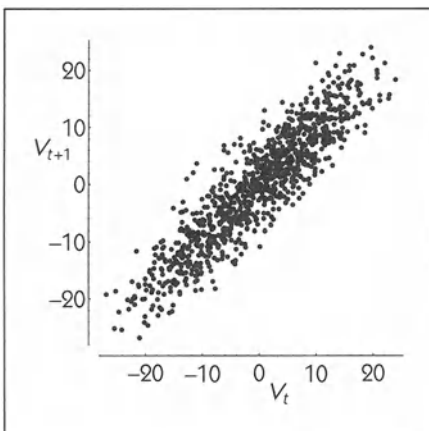
**Figure 6.6**

A scatter plot of the fluctuations around the mean:  $V_{t+1}$  versus  $V_t$  from Model One. (One thousand data points are shown.)

a claim that  $V_{t+1}$  is independent of  $V_t$ . (More advanced texts on statistics give a precise meaning of “close to zero” in terms of the uncertainty in the estimate  $\rho_{\text{est}}$ . See, for example, Box and Jenkins (1976).) Figure 6.6 shows  $V_{t+1}$  plotted against  $V_t$  for the Model One data. The round cloud of points is typical of a lack of correlation between the two variables.

The mean of the measurements from Model Two is  $M_{\text{est}} = 84.10$ . Calculating the correlation coefficient by applying Eq. 6.18, we find  $\rho_{\text{est}} = 0.786$ . This indicates a substantial degree of correlation between  $V_{t+1}$  and  $V_t$ , as shown by the cigar-shaped cloud of Figure 6.7. This leads us to conclude that the fluctuations from the mean in Model Two are not entirely the result of white noise measurement error.

The measurement noise  $W_t$  and the dynamical noise  $v_t$  in Model Two are both Gaussian white noise, but they play different roles. The measurement noise



**Figure 6.7**

A scatter plot of the fluctuations around the mean:  $V_{t+1}$  versus  $V_t$  from Model Two. (One thousand data points are shown.)

is completely forgotten from one time step to the next— $D_{t+1}$  contains no information about the measurement noise at time  $t$ . The dynamical noise  $v_t$ , however, changes the value of the state variable  $x_t$ . Imagine that after time  $t$  the dynamical noise  $v_t$  were turned off. The noise at time  $t$  would be remembered while the state variable moved back exponentially to its fixed point. This “memory” of past dynamical noise, which can be characterized by the **impulse response function** studied in Section 4.7 creates the correlation between  $V_t$  and  $V_{t+1}$ . In contrast, there is no mechanism to preserve memory of the measurement noise from one time to another.

The calculated value of the correlation coefficient,  $\rho_{\text{est}} = 0.786$ , tells us that consecutive measurements are not independent of one another. We might want to go further, and use  $\rho_{\text{est}}$  as an estimate of the value of  $\rho$  in the Model Two dynamics (Eq. 6.11), which we know to be  $\rho = 0.95$ . The difference between  $\rho_{\text{est}}$  and  $\rho$  arises mostly from the influence of the measurement noise  $W_t$  in Eq. 6.12. Since the measurement noise is incorporated in  $\rho_{\text{est}}$ ,  $\rho_{\text{est}}$  cannot be used by itself to estimate  $\rho$ . In Example 6.3 we will see one way to estimate  $\rho$  from the measurements.  $\square$

### ✦ MODEL THREE

Models One and Two display fixed points and exponential decay to a fixed point, respectively. Another type of behavior frequently encountered is oscillations. For example, consider the two coupled differential equations

$$\frac{dx}{dt} = y + v(t), \quad \frac{dy}{dt} = -ay - bx, \quad (6.19)$$

where  $v(t)$  is random noise. If we neglect the dynamical noise  $v(t)$ , we can use the tools from Chapter 5 and write down the characteristic equation for this differential equation, and then find the eigenvalues. They are

$$\lambda = \frac{-a}{2} \pm \frac{\sqrt{a^2 - 4b}}{2}, \quad (6.20)$$

so for  $a > 0$  and  $b > \frac{a^2}{4}$  the equation produces oscillations of exponentially decaying amplitude. The frequency of the oscillations are  $\omega = \frac{\sqrt{a^2 - 4b}}{2}$  and the time constant of the exponential decay is  $\frac{2}{a}$ .

In this case, we have two dynamical variables,  $x(t)$  and  $y(t)$ . We shall assume that we measure only one of them,  $x(t)$ , along with Gaussian white measurement noise  $W_i$ . The measurements  $D_i$  are made at discrete times, every  $T$  time units,

$$D_i = x(iT) + W_i. \quad (6.21)$$



Figure 6.8 shows a measured time series from Model Three with  $a = 0.5$ ,  $b = 3$ , and  $T = 0.1$ . We set the standard deviation of the dynamical noise  $\nu(t)$  to be 3 and that of the measurement noise  $W_i$  to be 1.

Although there are ups and downs in the data, it would be hard to claim from Figure 6.8 that the dynamics have much to do with an exponential decay in the amplitude of the oscillations. Nonetheless, motivated by the model we might ask

- What are the dynamics of movement toward the stable fixed point after an outside perturbation? In particular, what is the intrinsic frequency of the oscillation and the time constant of exponential decay?
- How much of the variability in the data is due to measurement error, and how much is due to outside perturbation?

To evaluate whether the model is appropriate for describing the data, ask:

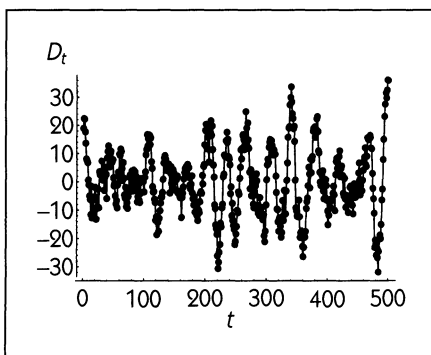
- What is the evidence that there are oscillations in the dynamics, as opposed to random perturbations and exponential decay as in Model Two?

□

## THE AUTOCORRELATION FUNCTION

The dynamics of Model Two involve exponential approach to the fixed point. The dynamics of Model Three involve sine-wave oscillations with an amplitude that decays exponentially. The data generated from the models do not show these dynamics very clearly but do indeed contain within them information about the exponential decay and sine-wave oscillations. We can use coefficients of correlation to reveal the dynamics obscured by noise.

Recall that  $V_t$  denotes fluctuations of the measured values around the mean,  $V_t = D_t - M_{\text{est}}$ . The correlation coefficient fits the relationship between  $V_{t+1}$



**Figure 6.8**

A time series from Model Three, with  $a = 0.5$ ,  $b = 3$ , and  $T = 0.1$ . The differential equations were integrated numerically using the Euler method with a time step  $\Delta = 0.1$ . The standard deviation of the dynamical noise  $\nu(t)$  is 3; that of the measurement noise  $W(t)$  is 1.

and  $V_t$  to the equation  $V_{t+1} = \rho V_t + \mathcal{W}_t$ . We can easily generalize the correlation coefficient to describe the relationship between  $V_{t+k}$  and  $V_t$ , giving us the **autocorrelation function**  $R(k)$ ,

$$R(k) = \frac{\sum_{t=1}^{N-k} V_{t+k} V_t}{\sum_{t=1}^{N-k} V_t V_t}. \quad (6.22)$$

$k$  is called the **lag** between the variables  $V_t$  and  $V_{t+k}$ . Note that the variable  $t$  is a “dummy variable”—it is used purely as an accounting device in the summations.  $R(k)$  is quite simple to calculate from data: One repeats basically the same calculation for several different values of  $k$ .

---

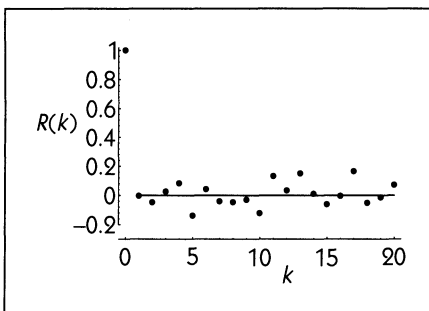
□ **EXAMPLE 6.3**

Use the autocorrelation function to show that the measured data from Models One, Two, and Three show distinct dynamics for the three models.

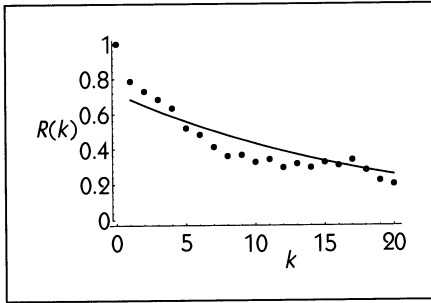
**Solution:** Figure 6.9 shows the autocorrelation function  $R(k)$  for the data from Model One. The autocorrelation function for this data, and for all data, takes the value 1 at  $k = 0$ , that is,  $R(0) = 1$ . The reason for this can be seen by inspecting Eq. 6.22; when  $k = 0$ , the numerator is the same as the denominator.

For the Model One data,  $R(k)$  is approximately zero for  $k > 0$ . (The deviations from zero are due to the finite length of the data used in calculating  $R(k)$ . See Exercise 6.5.) This is consistent with the model of the fluctuations as resulting from white measurement noise. In fact, this shape for  $R(k)$  is often taken as the definition of white noise, especially in older textbooks written before the current appreciation of nonlinear dynamics and chaos.

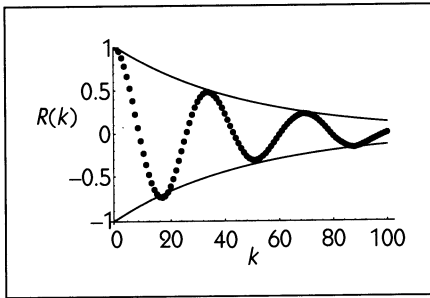
For the Model Two data,  $R(k)$  has a different shape, as Figure 6.10 shows. From its value of 1 at  $k = 0$ ,  $R(k)$  falls off sharply to approximately 0.8 at  $k = 1$ . For  $k \geq 1$ ,  $R(k)$  falls off exponentially. From the parameters used in Model Two, we know that the exponential dynamics in the absence of noise have the form



**Figure 6.9**  
The autocorrelation function  $R(k)$  versus  $k$  for the data from Model One.



**Figure 6.10**  
The autocorrelation function  $R(k)$  versus  $k$  for the data from Model Two.



**Figure 6.11**  
The autocorrelation function  $R(k)$  versus  $k$  for the data from Model Three.

$x_t = 0.95^t x_0$ . The thin line in Figure 6.10 plots this theoretical exponential decay and as can be seen, it fits the autocorrelation function very closely.

The sharp fall-off in  $R(k)$  from  $k = 0$  to  $k = 1$  reflects the white measurement noise in Model Two. This fall-off can be used to estimate the variance of the measurement noise. Without going into detail, we note that the fall-off has an amplitude of roughly 0.2. This means that 20 percent of the total variance of the Model Two data can be ascribed to white measurement noise. Since the total variance can be calculated from Eq. 6.8 to be 50.1, the variance of the measurement noise is estimated to be roughly  $50.1 \times 0.2 \approx 10$ . This gives an estimated standard deviation of 3.2, consistent with the theoretical value of 3 used in generating the data.

The autocorrelation function for the Model Three data is shown in Figure 6.11. It consists of a sine wave of exponentially decaying amplitude. From the figure, the period of the sine wave is easily found to be roughly 36 time units. The thin lines show an exponentially decaying envelope of the form  $0.98^t$ . This compares well with the theoretical form for the noiseless dynamics as  $0.975^t \sin\left(\frac{2\pi t}{36.6}\right) + \cos\left(\frac{2\pi t}{36.6}\right)$ , where  $A$  and  $B$  are set by the initial conditions.  $\square$

## 6.5 POWER SPECTRUM ANALYSIS

Consider a slight modification of Model Three: The dynamics and measurement process are the same, except that in addition to measuring  $D_t$ , let us also measure the dynamical noise  $\nu_t$ . Perhaps in this situation it is impolite to call  $\nu_t$  “noise” since we know what it is. Thus we will now call  $\nu_t$  the **input** to the system, and we’ll call  $D_t$  the **output**. We are interested in the input/output relationship.

### THE FOURIER TRANSFORM

In Section 4.7, we saw how a signal could be broken down, or **decomposed**, into the sum of simpler signals. For instance, the signal shown Figure 4.17 can be decomposed into the four simpler signals shown in Figure 4.18. This type of decomposition can be performed in any number of ways.

One incredibly powerful decomposition is into sine waves of different frequencies. Recall from Section 4.7 the following facts for linear systems:

1. The output that results from a sine-wave input of frequency  $\omega$  is a sine wave of the same frequency  $\omega$  but perhaps of different amplitude and phase. The amplitude of the output sine wave  $A_{\text{output}}(\omega)$  is proportional to the amplitude of the input sine wave  $A_{\text{input}}(\omega)$ :

$$A_{\text{output}}(\omega) = G(\omega)A_{\text{input}}(\omega).$$

For any input phase  $\phi_{\text{input}}(\omega)$ , the output phase  $\phi_{\text{output}}(\omega)$  is shifted by a fixed amount at each frequency,

$$\Phi(\omega) = \phi_{\text{output}}(\omega) - \phi_{\text{input}}(\omega).$$

$G(\omega)$  is called the **gain** of the system, and it may be different at different frequencies.  $\Phi(\omega)$  is called the **phase shift** and may also differ at different frequencies.

2. **Linear superposition of inputs** says that if the input can be written as a sum of sine waves of different frequencies, then the output is the sum of sine waves of those same frequencies. The amplitude and phase of the sine wave at each frequency in the output are exactly the same as if the input had been purely the single corresponding sine wave in the input.

The method for decomposing a signal into sine waves of different frequencies is called the **Fourier transform**. The details of how this is done are covered in many texts (see Press et al. (1992)). Here we simply point out that any signal

can be decomposed into sine waves and that the result is an amplitude and phase at each frequency.

An especially important case is when the input is white noise. For white noise,  $A_{\text{input}}(\omega)$  is a constant for all  $\omega$ . The constant is proportional to the standard deviation of the white noise. The phase  $\phi_{\text{input}}(\omega)$  varies from one frequency to another, and is generally regarded as random. Since  $A_{\text{input}}(\omega)$  is constant, white noise can be considered as a sum of signals of all different frequencies. This is where the name “white” comes from, by analogy to the fact that white light is a mixture of equal parts of many different frequencies of light.

### THE TRANSFER FUNCTION

Having measured the input and output signals from Model Three, we use the Fourier transform to decompose each of the two signals into a sum of sine waves of different frequencies. At each frequency  $\omega$ , we have amplitudes  $A_{\text{input}}(\omega)$  and  $A_{\text{output}}(\omega)$  and phases  $\phi_{\text{input}}(\omega)$  and  $\phi_{\text{output}}(\omega)$ . We can easily calculate

$$G(\omega) = \frac{A_{\text{output}}(\omega)}{A_{\text{input}}(\omega)} \quad \text{and} \quad \Phi(\omega) = \phi_{\text{output}}(\omega) - \phi_{\text{input}}(\omega).$$

Note that  $G(\omega)$  and  $\Phi(\omega)$  are functions of frequency  $\omega$ . This pair of functions is called the **transfer function** of the system. If we know the transfer function for a linear system, then we can calculate the output for any given input, or vice versa (as long as  $G(\omega) \neq 0$ ).

You may recall from Section 4.7 that an input/output system is described by its **impulse response**. The transfer function and impulse response are different ways of looking at exactly the same thing. In fact, the transfer function is the Fourier transform of the impulse response.

### THE POWER SPECTRUM

Suppose that we do not actually measure the input but that we know or assume that it is white noise. This tells us that  $A_{\text{input}}(\omega)$  is constant. Knowing this, we can calculate the gain  $G(\omega)$  to within a constant of proportionality, even without having measured the input:

$$G(\omega) = \text{const } A_{\text{output}}(\omega).$$

However, since we don't know anything about  $\phi_{\text{input}}(\omega)$ , we cannot calculate  $\Phi(\omega)$ . The square of  $G(\omega)$  is called the **power spectrum**.

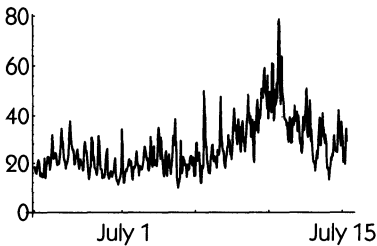
The power spectrum contains exactly the same information as the auto-correlation function—the power spectrum is in fact the Fourier transform of

the autocorrelation function. Although the information is the same, the different format of the information sometimes makes it advantageous to use the power spectrum rather than the autocorrelation function for analyzing data.

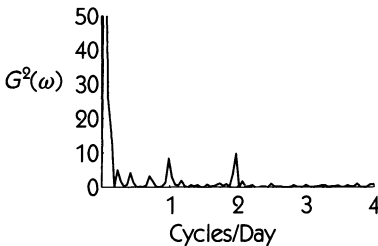
## DYNAMICS IN ACTION

### 17 DAILY OSCILLATIONS IN ZOOPLANKTON

The top figure here shows hourly measurements of zooplankton density. The power spectrum, shown in the bottom figure, displays the square of the amplitude of the oscillations at each frequency. Here, instead of measuring frequency in units of cycles/second (Hertz), we use cycles/day to reflect the time scale over which zooplankton density changes significantly.



Hourly measurements of zooplankton density (in  $\text{g/m}^3$ ) measured in the Middle Atlantic Bight starting on 25 June 1988. (See Ascioiti et al., 1993)



The power spectrum  $G^2(\omega)$  of the data from the above figure.

The power spectrum  $G^2(\omega)$  from this data shows a peak at 1 cycle per day. This peak corresponds to the daily changes in zooplankton density that come from the day/night cycle. (There is also a peak in  $G^2(\omega)$  at 2 cycles per day. This suggests that the daily cycle is not a simple sine wave, but that each cycle has some other shape.)

In addition to the daily changes at 1 and 2 cycles per day, the zooplankton data contain other variations. For instance, there is a week-long buildup that reaches to a maximum on July 10. Such slow variability in the zooplankton density appears as the large values of  $G^2(\omega)$  for low frequencies ( $\ll 1$  cycle/day). The power spectrum is often used, as in this case, to display periodic variability at a given frequency that might be hidden by other forms of variability.

#### ➤ MODEL FOUR

Models One, Two, and Three have linear dynamics. The parameters used in the models have been set so that, in the absence of dynamical noise, the stable fixed point is approached asymptotically. Nonlinear models can have nonfixed asymptotic behavior. As we saw in Chapter 1, the quadratic map

$$x_{t+1} = \mu x_t(1 - x_t) \quad (6.23)$$

can show a variety of behaviors from stable fixed points, to stable periodic cycles, to chaos. In particular, for  $\mu = 4.0$  the dynamics are chaotic, while for  $\mu = 3.52$  there is a stable cycle of period 4. Equation 6.23 involves no dynamical noise.

In order to emphasize the difference between the chaotic dynamics of Model Four and the noisy linear dynamics of Models One, Two, and Three, we shall assume that there is no measurement noise:

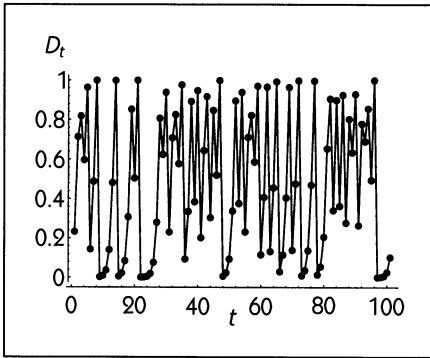
$$D_t = x_t. \quad (6.24)$$

Figure 6.12 shows a time series taken from Model Four.

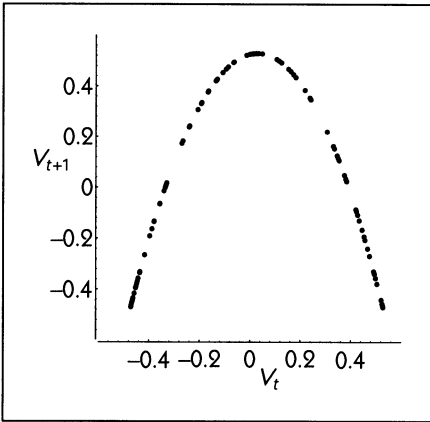
Since there is neither dynamical nor measurement noise, the model is completely **deterministic**. This means that, in principle, if we know the initial condition we can calculate all future values. Of course, if the model is chaotic, there may be practical limitations on our ability to do this.

With this model as a hypothesis, we might ask the following questions about our data:

1. What evidence is there that a deterministic process generates the data?
2. What evidence is there that the data involve a nonlinear process?
3. If the data are indeed chaotic, how large is the sensitive dependence on initial conditions?

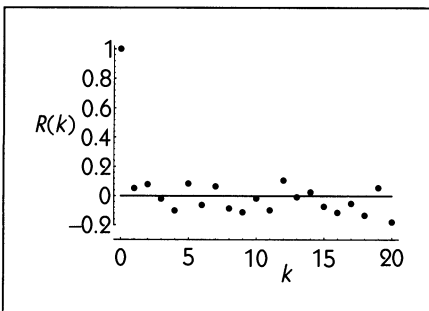


**Figure 6.12**  
A simulated time series from Model Four, with  $\mu = 4.0$ . There is no measurement noise.



**Figure 6.13**  
A scatter plot of fluctuations around the mean:  $V_{t+1}$  versus  $V_t$  from Model Four.

We can start the data analysis with the tools already at our disposal. The mean of the data in Figure 6.12 is  $M_{est} = 0.471$ . The fluctuations about the mean  $V_t = D_t - M_{est}$  can be used to calculate the correlation coefficient between  $V_{t+1}$  and  $V_t$ . This is  $\rho_{est} = 0.054$ , close to zero even though a scatter plot of  $V_{t+1}$  versus  $V_t$  does not look like a ball. (Compare Figure 6.13 with Figure 6.6. Both scatter



**Figure 6.14**  
The autocorrelation function  $R(k)$  calculated from the data in Model Four.



plots produce  $\rho_{\text{est}}$  near zero.) In fact, the autocorrelation function for the data is very similar to that found for the data from Model One (compare Figure 6.9 to Figure 6.14). This suggests that the data from Model Four are white noise, apparently contradicting the fact that the data are from a deterministic model.

The resolution to this paradox can be seen if we remember that the correlation coefficient and the autocorrelation function measure linear correlations in the data. The scatter plot of  $V_{t+1}$  versus  $V_t$  shows a very strong relationship, but the relationship is nonlinear and hence not accurately represented by the correlation coefficient and autocorrelation function. □

## 6.6 NONLINEAR DYNAMICS AND DATA ANALYSIS

In the previous section we saw that statistics such as the correlation coefficient and the autocorrelation function are not able to distinguish between the data from the linear Model One and those from the nonlinear Model Four. In this section we will describe data-analysis methods that are appropriate for nonlinear systems. Nearly all of the techniques have been developed since 1980, and new developments are made on an almost daily basis.

Most techniques for nonlinear data analysis involve two steps. In the first step, the data are used to **reconstruct** the dynamics of the system. This is the subject of the present section. The second step involves **characterization** of the reconstructed dynamics and will be the subject of Sections 6.7 and 6.8.

### RECONSTRUCTING FINITE-DIFFERENCE EQUATIONS: RETURN MAPS

Model Four is a finite-difference equation (the quadratic map that we studied in Chapter 1). Compare Figure 6.13 to Figure 1.16. The scatter plot derived from data reproduces the parabolic form of the graph drawn from the finite-difference equation. This shouldn't be surprising. A finite-difference equation like  $x_{t+1} = f(x_t)$  describes the relationship between  $x_{t+1}$  and  $x_t$ . A scatter plot of the measured data,  $D_{t+1}$  versus  $D_t$ , describes exactly the same relationship. Since in Model Four we defined  $D_t = x_t$ , each of the dots in the scatter plot falls on the function  $f(\cdot)$ , and the dots do a good job of indicating the parabolic geometry of  $f(\cdot)$ . (In Figure 6.13 we plot  $V_{t+1}$  versus  $V_t$ . This is this is the same thing as plotting  $D_{t+1}$  versus  $D_t$  but translating both axes by the mean  $M_{\text{est}}$ .)

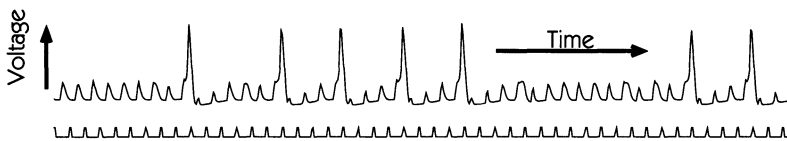
The idea of using a scatter plot to display the relationship between successive measurements is fundamental to the analysis of data from nonlinear systems. We will call the scatter plot a **return plot**, but other names found in the technical

literature for this type of scatter plot are *first-return plot*, *Poincaré return map*, and *return map*.

## DYNAMICS IN ACTION

### 18 RECONSTRUCTING NERVE CELL DYNAMICS

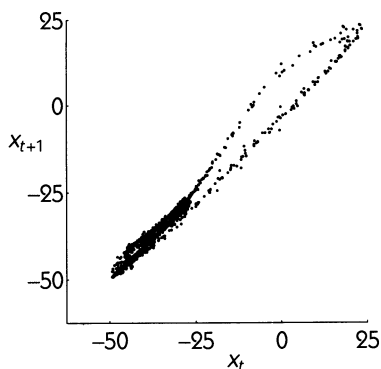
The top figure shows a recording of the voltage across the membrane of a giant axon (the axon is a part of a nerve cell) from a squid. These data were collected by Alvin Shrier and John Clay, at the Woods Hole Oceanographic Institution.



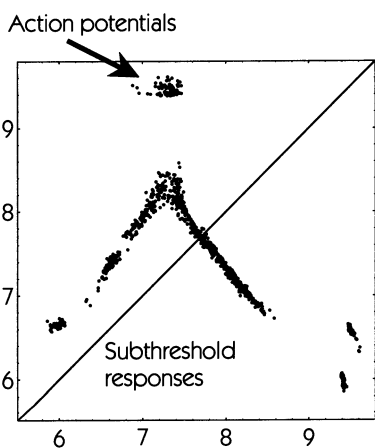
**Transmembrane voltage from a periodically stimulated squid giant axon. The times of the stimuli are also indicated. The bottom trace shows the stimulation current. Stimuli were applied every 10 msec. These data were provided by Drs. A. Shrier and J. Clay.**

An electrode has been inserted into the cell, and a periodic stimulus has been applied. In response to each stimulation, the axon has either a small response (a “subthreshold response”) or a big one (an “action potential”). The transmembrane voltage has been sampled by a computer 10,000 times per second. Since the voltage does not change much over 0.0001 seconds, a return plot of the voltage  $x_{t+1}$  versus  $x_t$  stays very close to the line of identity, and there is no evidence in this plot for a single-valued nonlinear function (see the next figure).

One technique for generating a return plot appropriate for the squid axon data is to reduce the time series into a set of discrete measurements made at a time interval having a relationship to the systems’s dynamics. In the squid axon case, a sensible time interval is the time between stimuli, rather than the time between successive voltage samples taken by the computer. Several types of measurements might be taken once per stimulus. For example, we might take a single measurement from the recording some fixed time (say, 20 msec) after each stimulus. Or we might choose to measure the recording only at the peak of the response to each stimulus.



A return plot  $x_{t+1}$  versus  $x_t$  for the voltage across the membrane of the squid axon.



$A_{i+1}$  versus  $A_i$  for the squid axon data, where  $A_i$  is the logarithm of the area under response  $i$ .

The bottom figure shows a return plot made by calculating the logarithm of the area under each stimulus response. The plot shows that the repetitive subthreshold responses seen in the time series result from an unstable fixed point; action potentials are generated only when the dynamics move away from this fixed point. The action potentials that appear to occur at random intervals in the top figure are really generated by a nonlinear dynamical system that can be largely characterized by the return plot.

For other systems, it may not be obvious how frequently to make measurements for the purpose of drawing a return plot. For example, the sunspot data shown in the preface were collected once per month. Nothing about a one-month interval relates to the dynamics of the sun—if we want to extract information about the dynamics, another measurement interval might be more appropriate.

In many cases, data have been collected from a continuous-time dynamical system properly described by differential equations rather than by finite-difference equations. In such cases, it may be appropriate to use the *phase-plane* or *embedding* reconstruction techniques described in the following sections. Sometimes, however, a return map does describe effectively the dynamics behind continuous-time data. In drawing such a return map, take care to select a time interval that reflects some important aspect of the dynamics.

### RECONSTRUCTING THE PHASE PLANE

Consider data generated from the second-order differential equation describing a harmonic oscillator:

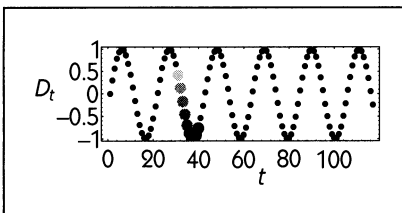
$$\frac{d^2x}{dt^2} = -bx. \quad (6.25)$$

As shown in Section 5.4, this equation can be rewritten in terms of two first-order differential equations,

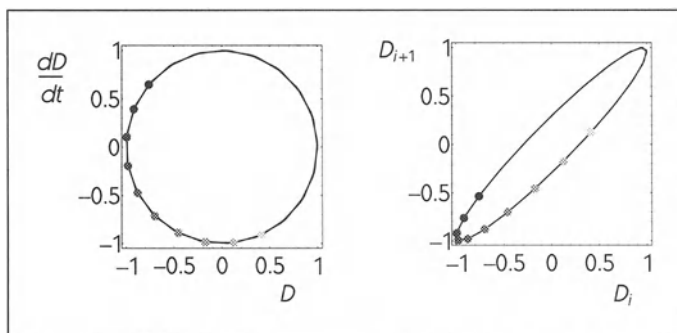
$$\begin{aligned} \frac{dx}{dt} &= y \\ \frac{dy}{dt} &= -bx. \end{aligned} \quad (6.26)$$

The variables  $x$  and  $y$  form the **phase plane**, and Eq. 6.26 describes the flow of the dynamics on this plane.

Suppose that we measure a time series  $D(t) = x(t)$  from Eq. 6.25 (see Figure 6.15). How can we reconstruct the phase plane and the flow on it from the measured data? At any instant, the position of the system on the phase plane is given by the coordinates  $(x, y)$ . The time series itself gives us  $D$  at every instant.



**Figure 6.15**  
The quantity  $D_t$  measured from Eq. 6.25.



**Figure 6.16** Two versions of the reconstructed phase plane for the data generated from Eq. 6.25. The gray dots indicate the position in the phase plane at the corresponding times in Figure 6.15.

We can measure  $y(t)$  from  $D(t)$  by noticing in Eq. 6.26 that  $y = \frac{dD}{dt}$ . If we plot out  $\frac{dD}{dt}$  versus  $D$ , we get the **trajectory** of the system in the phase plane—this describes the flow based on the measured data (see Figure 6.16).

Given a time series  $x(t)$ , how do we calculate  $\frac{dx}{dt}$ ? There are simple electronic circuits that act as differentiators, and in the past such a circuit might have been used to sketch out the trajectory on an oscilloscope screen. Today data typically are collected by computer, and so the measurement  $D(t)$  actually consists of a sequence of measurements made at discrete times  $D_0, D_1, D_2, \dots$ . Using the textbook definition of the derivative of  $x$  at time  $t$ ,

$$\frac{dx(t)}{dt} = \lim_{h \rightarrow 0} \frac{x(t+h) - x(t)}{h},$$

we are motivated to approximate the derivative at time  $t$  as

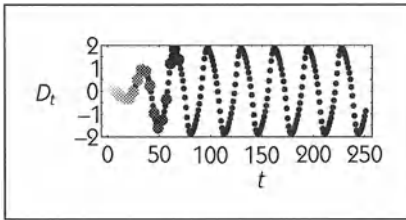
$$\frac{dD_t}{dt} = \frac{D_{t+h} - D_t}{h}.$$

For the discrete-time measurements,  $h$  can only take on the values 0, 1, 2, 3,  $\dots$ —it cannot have a fractional value. The smallest useful value is  $h = 1$ , but sometimes, as we will see below, it is appropriate to select larger  $h$ .

Reconstructing the phase plane is thus a matter of plotting  $\frac{D_{t+h} - D_t}{h}$  versus  $D_t$ . Notice that only two quantities are involved:  $D_{t+h}$  and  $D_t$ . They contain all the information in the plot, and it is effective simply to plot  $D_{t+h}$  versus  $D_t$ .

Equation 6.26 is a special case because  $\frac{dx}{dt}$  gives us  $y$ . In general, dynamics on the phase plane are given by the pair of coupled differential equations (see Eq. 5.18)

$$\frac{dx}{dt} = f(x, y), \quad \frac{dy}{dt} = g(x, y). \quad (6.27)$$



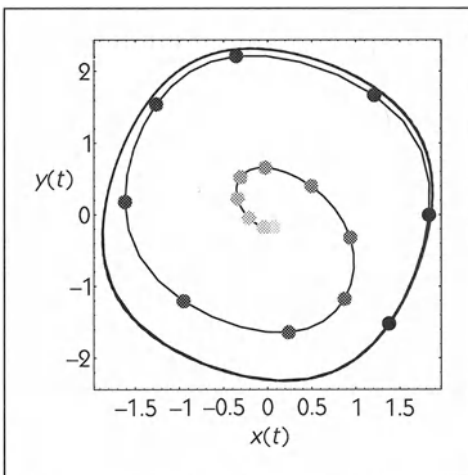
**Figure 6.17**  
The quantity  $D_t$  measured from the van der Pol system, Eq. 5.25

If we measure only  $x(t)$ , how can we calculate the value of  $y$ ? The answer is that often we cannot, but we do not need to in order to display relevant information about the dynamics in the phase plane. Notice that if we measure  $x(t)$  and calculate  $\frac{dx}{dt}$ , we have both a direct measurement of  $x$  and a calculated value of  $f(x, y)$ . Some information about  $y$  is contained in the value of  $f(x, y)$ , and often this information is enough to allow us to get a good idea of the dynamics. Figures 6.17 through 6.19 give an example that shows how the reconstructed  $(D_t, D_{t+1})$  phase plane compares to the original  $(x, y)$  phase plane.

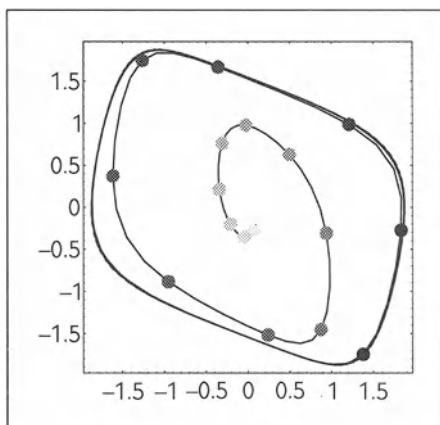
To summarize, by making a series of measurements  $D_t$  and plotting  $D_{t+h}$  versus  $D_t$ , we can often reconstruct the phase-plane dynamics of a system, even though we never make direct measurements of the dynamical variable  $y$ .

**EMBEDDING A TIME SERIES**

As we saw in Chapter 4, a continuous-time system of ordinary differential equations that generates chaos must involve at least three equations. This means that the two-dimensional dynamics in a phase plane cannot represent chaotic be-



**Figure 6.18**  
Dynamics in the original  $x, y$  phase plane for the van der Pol equation.



**Figure 6.19**  
Dynamics of the van der Pol equation in the reconstructed  $D_{t+10}$  versus  $D_t$  phase plane.

havior. In order to reconstruct the geometry of a continuous-time chaotic system from a time series, we can extend the technique developed for reconstructing the phase plane. The phase-plane reconstruction involved plotting successive points in a two-dimensional space. To reconstruct the dynamics in a three-dimensional space, we plot the points as a three-dimensional coordinate:

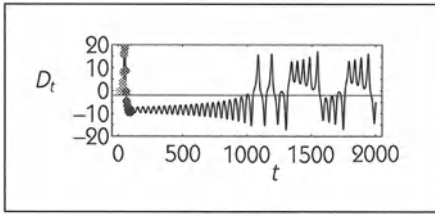
$$(D_t, D_{t-h}, D_{t-2h}).$$

More generally, we can **embed** the time series in a  $p$ -dimensional space by taking  $p$ -coordinates,

$$\mathbf{D}_t = (D_t, D_{t-h}, D_{t-2h}, \dots, D_{t-(p-1)h}). \quad (6.28)$$

We use the boldface  $\mathbf{D}_t$  to denote the embedded measurements, to differentiate from  $D_t$ , which denotes a single measurement at time  $t$ .  $\mathbf{D}_t$  incorporates measurements made at different times, ranging from  $t$  to  $t - (p - 1)h$ , but the index  $t$  is used for notational convenience.

This technique of representing a measured time series as a sequence of points in a  $p$ -dimensional space is called **time-lag embedding**. There is an important theorem (Taken's embedding theorem) that says the reconstructed dynamics are geometrically similar to the original for both continuous-time and discrete-time systems. The sequence of points created by embedding a time series is called the trajectory of the time series.  $p$  is called the **embedding dimension**, and  $h$  is the **embedding lag**.

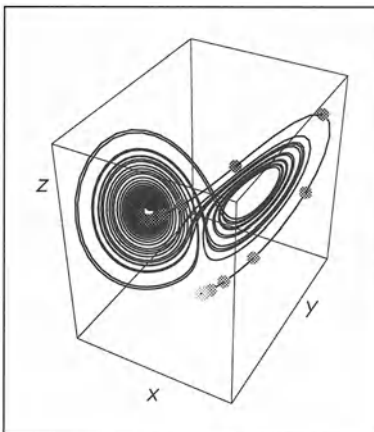


**Figure 6.20**  
A measured signal  $D_t$  from the Lorenz system (Eq. 6.29).

As an example, consider the Lorenz system of three ordinary nonlinear differential equations that produce chaos:

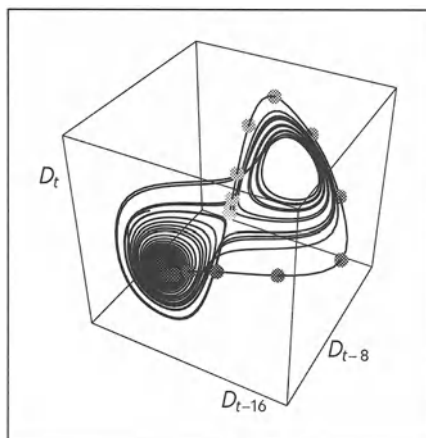
$$\begin{aligned}\frac{dx}{dt} &= 10(y - x), \\ \frac{dy}{dt} &= 28x - y - xz, \\ \frac{dz}{dt} &= 28xy - \frac{8z}{3}.\end{aligned}\tag{6.29}$$

If one could measure  $x(t)$ ,  $y(t)$ , and  $z(t)$  simultaneously, in a physical system, then by plotting out the three-dimensional coordinate  $(x(t), y(t), z(t))$ , we can reconstruct the dynamics in the three-dimensional **phase space**. But if we measure only one of the variables, so that  $D(t) = x(t)$ , we can create a reconstruction that is faithful to the geometry of the original, as shown in Figures 6.20 through 6.22.



**Figure 6.21**  
The trajectory of Eq. 6.29 in the original  $x, y, z$  phase space.





**Figure 6.22**  
The reconstructed trajectory  
of Eq. 6.29 using  $D_t$ ,  $D_{t-8}$ ,  $D_{t-16}$ .

#### ✈ MODEL FIVE

The chaotic dynamics of Model Four were generated by a single equation involving a single state variable. Chaotic dynamics can also be generated by systems with more state variables, for instance the Lorenz equations (Eq. 6.29) have three state variables producing a chaotic attractor with a fractal dimension of approximately 2.06.

Although the dimension of a chaotic attractor may be less than the number of state variables, it can never exceed the number of state variables. In order to illustrate some of the properties of high-dimensional chaotic systems, we introduce a new model that produces chaotic dynamics, the Ikeda map:

$$\begin{aligned}x_{t+1} &= 1 + \mu(x_t \cos m_t - y_t \sin m_t), \\y_{t+1} &= \mu(x_t \sin m_t + y_t \cos m_t),\end{aligned}\tag{6.30}$$

where  $m_t = 0.4 - \frac{6.0}{(1+x_t^2+y_t^2)}$  and  $\mu = 0.7$ . The Ikeda map has two dynamical variables,  $x_t$  and  $y_t$ . ( $m_t$  is just a convenience variable and can easily be eliminated from the equations by substitution.)

Since Eq. 6.30 has just two dynamical variables, any attractor it has can be at most two-dimensional. This is not very high, so let us consider another chaotic dynamical system, the Henon map:

$$\begin{aligned}z_{t+1} &= 1.4 + 0.3v_t - z_t^2, \\v_{t+1} &= z_t.\end{aligned}\tag{6.31}$$

This equation also has two variables, so its attractor can also be at most two-dimensional.

Taken together, however, the Ikeda map and the Henon map have four dynamical variables,  $v_t, w_t, x_t, y_t$ , so the attractor for the combined system can be at most four-dimensional. Suppose we measure

$$D_t = x_t + \beta z_t + W_t, \quad (6.32)$$

where  $W_t$  is random Gaussian white measurement noise. Our measured data will reflect the dynamics of both the Henon and Ikeda maps, and also the random noise  $W_t$ . Similarly, a natural or experimental time series may reflect the dynamics of several subsystems. Here, we will somewhat arbitrarily pick  $\beta = 0.3$  and set the level of measurement noise to a standard deviation of 0.05 (see Figure 6.23).

This trick of adding signals from unrelated chaotic systems allows us to make a chaotic system of a higher dimension than any of the individual systems. We could add any number of such systems. Surprisingly, we could even add two or more copies of the same chaotic system, as long as the initial conditions were different in each copy.

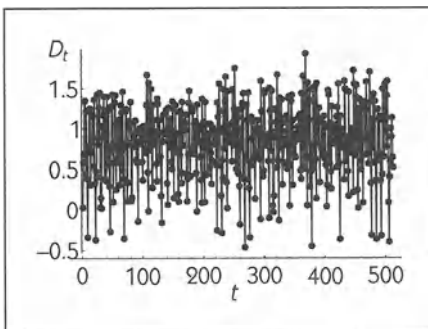
Sometimes the dynamics of subsystems are coupled together so that one subsystem affects another. All of the sets of equations examined previously in this book have been this way. For Model Five, we will linearly couple the  $x$  variable of the Ikeda map to the  $z$  variable in the Henon map,

$$x_{t+1} = 1 + \mu(x_t \cos m_t - y_t \sin m_t) + 0.2z_t, \quad (6.33)$$

and leave the dynamical equations for the other variables as they are in Eq. 6.31.

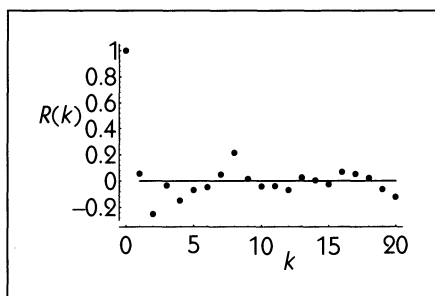
Using a model of this sort in interpreting a measured time series, we might ask the following questions:

- How many variables are involved in the dynamics?
- Is there an attractor, and what is its dimension?



**Figure 6.23**

A simulated time series from Model Five,  $D_t = x_t + \beta z_t + W_t$ . The measurement noise  $W_t$  has standard deviation 0.05.

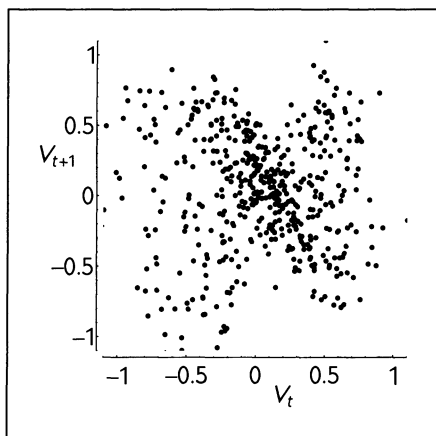


**Figure 6.24**  
The autocorrelation function of the Model-Five data shown in Figure 6.23.

- Can we distinguish between the measurement noise and the deterministic dynamics?
- If the system is very high dimensional, is it even possible to detect the deterministic dynamics?

Again, we start our analysis of the data with the tools we have already introduced. The mean of the Model Five data is  $M_{\text{est}} = 0.828$ , and the standard deviation is  $\sigma = 0.455$ . The fluctuations about the mean are  $V_t = D_t - M_{\text{est}}$ . The autocorrelation function  $R(k)$ , shown in Figure 6.24, is consistent with white noise.

Although there is no dynamical noise in the equations, a scatter plot of  $V_{t+1}$  versus  $V_t$  for the Model Five data (Figure 6.25) does not show the simple geometry that was evident in the Model Four data (Figure 6.13). For Model Five,  $V_{t+1}$  is clearly not a function of  $V_t$ , even though we know that deterministic dynamics are at work. As we shall see, by using a two- or higher-dimensional embedding,



**Figure 6.25**  
A return plot of fluctuations about the mean:  $V_{t+1}$  versus  $V_t$  for the Model Five data.

the deterministic relationship between  $V_{t+1}$  and previous values ( $V_t, V_{t-1}, \dots$ ) becomes clearer. □

---

## 6.7 CHARACTERIZING CHAOS

In Chapter 1, chaos was defined to be bounded, deterministic dynamics that are aperiodic and display sensitive dependence on initial conditions. In this section we will study time-series analysis techniques that allow us to investigate each of these characteristics in data.

### BOUNDEDNESS

According to Chapter 1, dynamics are bounded if they stay in a finite range and do not approach  $\infty$  or  $-\infty$  as time increases. In practice, things are more subtle than this. An example is given by the simple linear system  $x_{t+1} = Rx_t$ . The solution is  $x_t = R^t x_0$ , that is,  $x_t$  grows or decays exponentially. Suppose that we measure  $D_t = \frac{1}{x_t}$ . For  $|R| > 1$ , the dynamics of  $x$  are unbounded, but  $D_t$  will go to zero as  $t \rightarrow \infty$ . For  $|R| < 1$ , the dynamics are bounded but  $D_t \rightarrow \pm\infty$ . This example shows that when dealing with measured data it is not sufficient to say that dynamics are bounded if a measured time series stays in a finite range, or unbounded if the time series blows up. In fact, if for no other reason than not having the opportunity to wait until  $t \rightarrow \infty$ , we can never definitively know from measurements whether the “true,” unmeasured state variables stay bounded.

The definition of bounded as “staying in a finite range” is not very useful when dealing with data; any measured data will be in a finite range, since the mass and energy of the universe are finite. Infinity is a mathematical concept, not a physical one.

A different, but related concept for assessing boundedness in data is **stationarity**. We say that a time series is stationary when it shows similar behavior throughout its duration. One useful definition of “similar behavior” is that the mean and standard deviation remain the same throughout the time series. An operational definition might be that the mean and standard deviation in one third of the signal are not significantly different from those in the other two thirds—or one might prefer to use quarters or tenths, and so on.

If a time series is nonstationary, then it is questionable whether the techniques described in the following sections can be applied meaningfully. In this case we can attempt to generate stationarity by altering the time series. A simple and often effective technique is to **first-difference** the time series. That is, if the

measurements are  $\dots, x_i, x_{i+1}, x_{i+2}, \dots$ , then define  $y_i = x_i - x_{i-1}$  and use  $y_i$  for further analysis.

Another technique for attempting to create stationarity is motivated by exponential growth. If  $x_{t+1} = Rx_t$  for  $|R| > 1$ , then  $x_t$  will be nonstationary as will the first difference  $y_t = x_t - x_{t-1}$ . However,  $\frac{x_t}{x_{t-1}}$  will be stationary.

### APERIODICITY

Chaotic behavior is aperiodic. It might seem that the question of aperiodicity is one with a yes-or-no answer: Either a time series is periodic or it is not. However, in the presence of measurement noise, a measured time series from a truly periodic system can appear aperiodic. Because aperiodic systems can differ in their aperiodicity, it can be meaningful to quantify “how aperiodic” a time series is. Recall that aperiodicity means that the state variables never return to their exact previous values. However, in an aperiodic system, variables may return quite close to previous values. We can characterize aperiodicity by asking “How close?” and “How often?”

Since we often do not directly measure all of the state variables of a system, we need to use the embedding technique to represent all of our measured data’s state variables. Recall that  $D_t$  is the measurement made at time  $t$ . By *embedding* the time series, we create a sequence

$$\mathbf{D}_t = (D_t, D_{t-h}, \dots, D_{t-(p-1)h}),$$

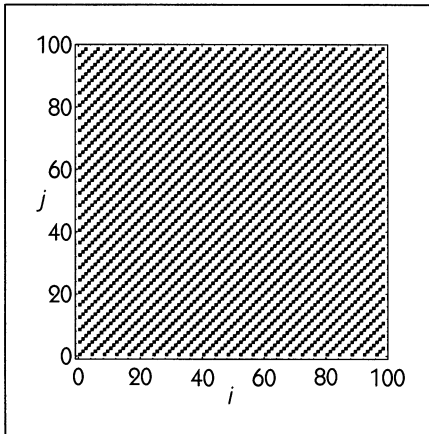
where  $p$  is the embedding dimension and  $h$  is the embedding lag. Each  $\mathbf{D}_t$  is a point in the  $p$ -dimensional embedding space, and the embedded time series can be regarded as a sequence of points, one point at each time  $t$ . Each point represents the state of the system at that time.

We can calculate the distance between the two points at times  $i$  and  $j$ :

$$\delta_{i,j} = |\mathbf{D}_i - \mathbf{D}_j|.$$

If the time series were periodic with period  $T$ , then  $\delta_{i,j} = 0$  when  $|i - j| = nT$ , for  $n = 0, 1, 2, 3, \dots$ . In contrast, for an aperiodic time series,  $\delta_{i,j}$  will not show this pattern. Suppose we pick some distance  $r$ , and ask when  $|\mathbf{D}_i - \mathbf{D}_j| < r$ . One way to do this is to make a plot where  $i$  is on the horizontal axis,  $j$  is on the vertical axis, and a dot is placed at coordinate  $(i, j)$  if  $|\mathbf{D}_i - \mathbf{D}_j| < r$ . Such plots are called **recurrence plots** because they depict how the reconstructed trajectory recurs or repeats itself (see Figures 6.26 and 6.27).

For a periodic signal of period  $T$ , the plot looks like Figure 6.26 for very small  $r$ . This is a series of stripes at 45 degrees, with the stripes separated by a

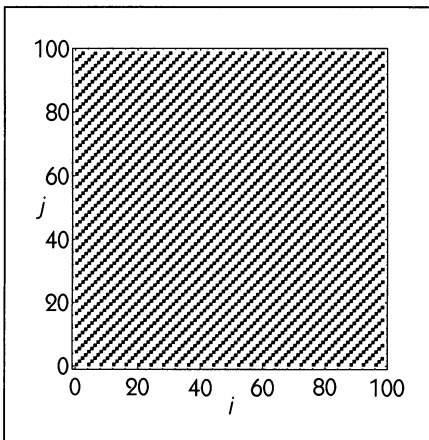


**Figure 6.26**  
 A recurrence plot for the quadratic map  $x_{t+1} = 3.52x_t(1 - x_t)$ . A black dot appears whenever  $|D_i - D_j| < r$ . The trajectory has a period of 4, so the recurrence plot consists of diagonal stripes separated by 4. The embedding dimension  $p = 2$ , and  $r = 0.001$ .

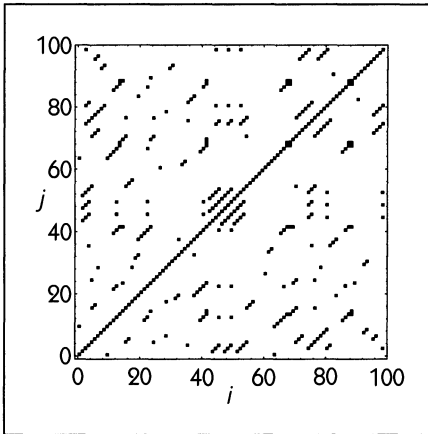
distance of  $T$  in the vertical and horizontal directions. (In all recurrence plots, there is a stripe along the diagonal corresponding to  $i = j$ .)

For a chaotic time series, the recurrence plot has a more complicated structure, sometimes with hints of almost periodic trajectories—one can see brief episodes where there are parallel stripes at 45 degrees (see Figures 6.28 and 6.29). For randomly generated numbers, such a structure is not evident (see Figures 6.30 and 6.31).

One thing to keep in mind is that the number of dots in a recurrence plot tells how many times the trajectory came within distance  $r$  of a previous value. The **correlation integral**  $C(r)$  is defined to be the fraction of pairs of times  $i$  and



**Figure 6.27**  
 The same as Figure 6.26, but  $r$  is ten times bigger:  $r = 0.01$ . The plot is identical to Figure 6.26.

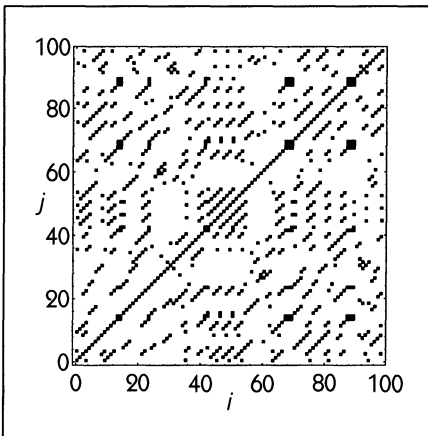


**Figure 6.28**  
 Recurrence plot for the chaotic  
 time series generated by  
 $x_{t+1} = 4x_t(1 - x_t)$ ;  $p = 2$ ,  
 $r = 0.001$ .

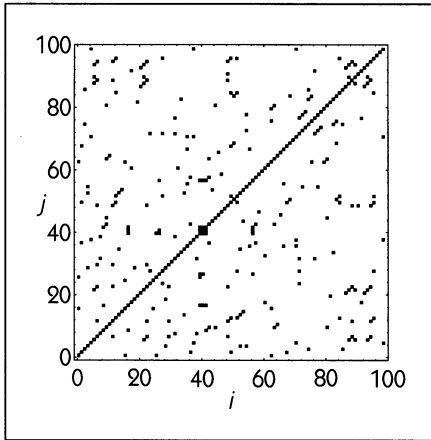
$j$  where  $D_i$  and  $D_j$  are closer than  $r$  for  $i \neq j$ :

$$C(r) = \frac{\text{number of times } |D_i - D_j| < r}{N(N - 1)}. \quad (6.34)$$

You can think of  $C(r)$  as the density of ink in a recurrence plot. The numerator is the actual number of dots in the plot, and the  $N(N - 1)$  in the denominator is the maximum possible number of dots. (Remember, we exclude the cases where  $i = j$ . Otherwise, the denominator would be  $N^2$ .)



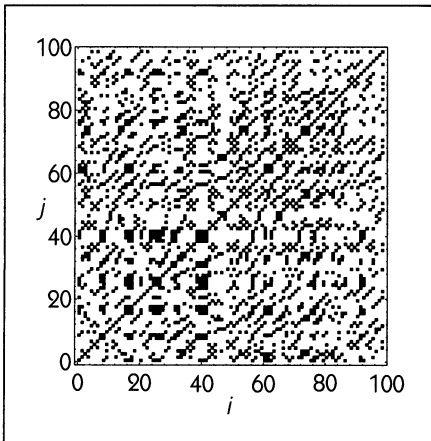
**Figure 6.29**  
 Same as Figure 6.28, but  $r$  is ten  
 times bigger.



**Figure 6.30**  
Recurrence plot for random  
white noise;  $p = 2, r = 0.1$ .

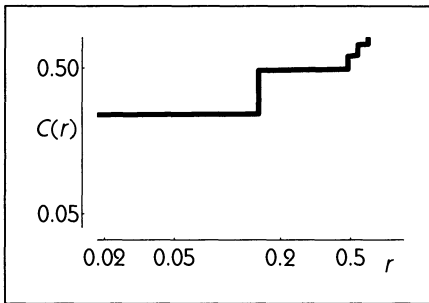
### THE CORRELATION DIMENSION

The correlation integral is one of the most fundamental quantities in chaotic time-series analysis. What is important is not the value of  $C(r)$  at any particular single value of  $r$ , but how  $C(r)$  changes with  $r$ . As  $r$  is increased, more dots appear in the recurrence plots and so  $C(r)$  increases. Figures 6.32, 6.33, and 6.34 show the correlation integral for the periodic data, the chaotic data, and the random white noise. For a perfectly periodic system, increasing  $r$  a little does not change the number of dots very much—compare Figures 6.26 and 6.27. For the chaotic data of Model Four, increasing  $r$  by the same amount causes more dots to appear (Figures 6.28 and 6.29), but the most dramatic increase occurs in the random white noise (Figures 6.30 and 6.31).  $C(r)$  is flat for the periodic system

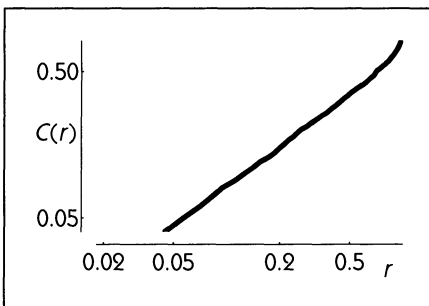


**Figure 6.31**  
Same as Figure 6.30, but  $r$  is ten  
times bigger.

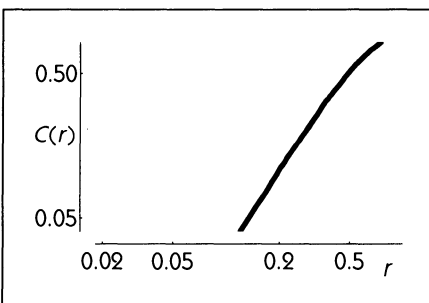


**Figure 6.32**

The correlation integral  $C(r)$  of a periodic time series with period 4 generated from  $x_{t+1} = 3.52x_t(1 - x_t)$ .  $N = 100$  data points were used, in an embedding dimension of  $p = 2$ . Note that  $C(r)$  is plotted on a log-log scale.

**Figure 6.33**

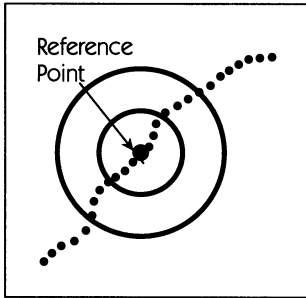
The correlation integral  $C(r)$  from a chaotic time series generated from Model Four ( $x_{t+1} = 4x_t(1 - x_t)$ );  $N = 100$ ,  $p = 2$ .

**Figure 6.34**

The correlation integral  $C(r)$  from a time series produced by a computer random-number generator;  $N = 100$ ,  $p = 2$ .

(Figure 6.32), has a gentle slope for the chaotic system (Figure 6.33), and has a steeper slope for the random system (Figure 6.34).

There is a close relationship between the correlation integral  $C(r)$  and the concept of fractal dimension introduced in Section 3.3. Imagine for a moment that you have a set of points scattered more or less uniformly on a one-dimensional curve, as in Figure 6.35. Pick one of the points as a reference, and count how many of the other points are within distance  $r$  of the reference. As  $r$  is increased, the number of points within distance  $r$  will increase directly as the length  $r$ . Now imagine that the points are scattered more or less uniformly on a two-dimensional surface (Figure 6.36). Choosing one of the points as a reference, we can see that



**Figure 6.35**  
When points are scattered along a one-dimensional curve, the number of points closer than distance  $r$  to a reference point increases linearly with  $r$ .

the number of points within distance  $r$  of the reference will be related to the area of a circle of distance  $r$ , that is,  $\pi r^2$ . Similarly, if the points were scattered throughout a three-dimensional volume, the number of points within distance  $r$  of a reference point would be related to the volume of a sphere of radius  $r$ , that is,  $\frac{4}{3} \pi r^3$ . In general, for points scattered throughout a  $\nu$ -dimensional object, the number of points closer than distance  $r$  to a reference point is proportional to  $r^\nu$ .

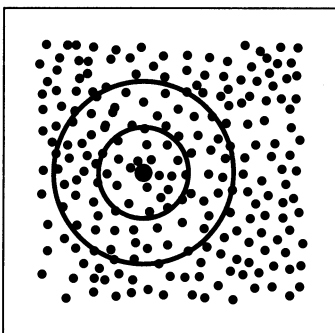
In calculating the correlation integral of a set of points, one uses each of the points as a reference and counts how many of the other points are within distance  $r$ . This suggests that the correlation integral of a scattering of points throughout a  $\nu$ -dimensional volume will be proportional to  $r^\nu$ , that is,

$$C(r) = Ar^\nu, \tag{6.35}$$

where  $A$  is a constant of proportionality. Taking the logarithm of both sides of Eq. 6.35 gives

$$\log C(r) = \nu \log r + \log A. \tag{6.36}$$

In order to find  $\nu$ , we simply need to plot  $\log C(r)$  versus  $\log r$  and find the slope of the resulting line. This procedure can also be applied to estimate the



**Figure 6.36**  
When points are scattered on a two-dimensional surface, the number of points closer than distance  $r$  to a reference point increases as the area of a circle of radius  $r$ .

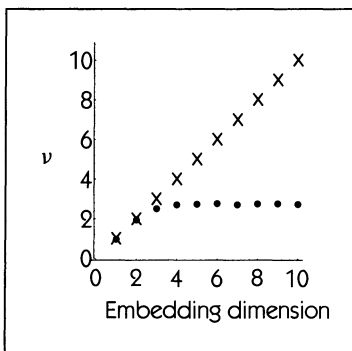
fractal dimension of an object, in place of the box-counting technique described in *Dynamics in Action* 5.

One way that the correlation dimension has been used in time-series analysis is to look for attractors in time series. The initial idea, proposed by Grassberger and Procaccia (1983), was based on the observation that the attractors of chaotic systems are often self-similar and can be described by a fractal dimension. If a time series comes from a dynamical system that is on an attractor, then the trajectory made from the time series by embedding will have the same topological properties as the original attractor—as long as the embedding dimension is large enough. In particular, the reconstructed trajectory will have the same dimension as the original one. Takens (1981) proved that if the original attractor has dimension  $\nu$ , then an embedding dimension of  $p = 2\nu + 1$  will be adequate for reconstructing the attractor. In practice,  $p \geq \nu$  will often be adequate, but the only guarantee comes when  $p \geq 2\nu + 1$ .

Since the objective of the Grassberger-Procaccia analysis is to find the dimension  $\nu$ , one does not know at the outset what embedding dimension  $p$  to use. The solution to this problem is to calculate  $\nu$  from the correlation integral at many different values of  $p$ , as shown in Figure 6.37.

For a time series from a system that is on a  $\nu$ -dimensional attractor, the correlation dimension of the time series offers a means to estimate the dimension of the attractor. However, for systems that are not on an attractor, the interpretation of  $\nu$  can be much more difficult.

One relatively simple case for interpreting  $\nu$  is random white noise. Consider a sequence of random white noise measurements, such as those from Model One. From Figure 6.6, you can see that when the data from Model One are embedded with  $p = 2$ , they create a solid-looking blob. Since this blob covers the whole plot with ink (or at least it would if there were many more data points), it is two-dimensional (i.e.,  $\nu = 2$  when  $p = 2$ ). Similarly, an embedding with  $p = 3$  would create a three-dimensional blob that would completely fill a three-dimensional volume, and so  $\nu = 3$  when  $p = 3$ . In the ideal case for random



**Figure 6.37**

An idealized case of correlation dimension  $\nu$  versus embedding dimension. The dots display  $\nu$  for a time series with attractor dimension 2.7, while the x symbol gives  $\nu$  for random white noise.

white noise,  $\nu = p$ . Here, “ideal” means “we have an infinite amount of data.” Of course, it is impossible to have an infinite amount of data, but even for small data sets the relationship  $\nu \approx p$  may hold. One theoretical rule of thumb is that  $10^p$  data points are needed to show that  $\nu \approx p$  is true up to any given  $p$ ; thus 1000 (or  $10^3$ ) points are needed to show that  $\nu = p$  for  $p = 3$ . In practice, this rule is quite conservative, and  $\nu$  can be shown to increase with  $p$  even for much shorter time series.

When introduced in the early 1980s, the correlation dimension was greeted with incredible enthusiasm and optimism. Now it has fallen into some disrepute. As shown in Figure 6.37, for a time series from a system on an attractor,  $\nu$  levels out with increasing  $p$  once  $p$  is large enough, while for random white noise,  $\nu$  increases with  $p$ . There has been a strong temptation for people to invert this logic and to believe that if  $\nu$  levels out with increasing  $p$ , then the time series reflects an attractor. Sometimes this is the case, but sometimes it is not. There are cases where  $\nu$  levels out but there is no attractor. A particularly important case is that of  $\frac{1}{f}$  noise (see *Dynamics in Action* 6). To guard against the incorrect interpretation of the correlation integral, it is important to use surrogate data, as described in Section 6.8.

---

□ **EXAMPLE 6.4**

Estimate the dimension of the chaotic attractor underlying the Model Five data.

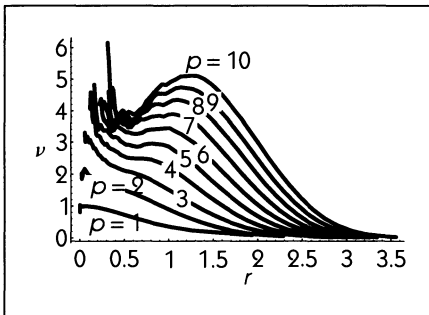
**Solution:** The first step is to embed the time series. This requires the choice of an embedding dimension  $p$  and an embedding lag  $h$ . One way to choose an embedding lag is to take the smallest value of  $h$  at which the autocorrelation function  $R(h) \approx 0$ . In this case, Figure 6.24 shows that  $R(1) \approx 0$ , so we will use  $h = 1$ .

Rather than picking a single embedding dimension, we will repeat the calculations for  $p = 1, 2, \dots, 10$ . At each of these embedding dimensions, we repeat the same steps:

1. Calculate the correlation integral  $C(r)$  using Eq. 6.34.
2. Using Eq. 6.36, use  $C(r)$  to calculate the dimension  $\nu$ . One way to do this is to plot  $\log C$  versus  $\log r$ .  $\nu$  is the slope of this graph:

$$\nu = \frac{d \log C}{d \log r}.$$

However, this slope generally depends on the value of  $r$  selected. To avoid this problem for the moment, we will plot out the slope  $\nu$  as a function of  $r$ .



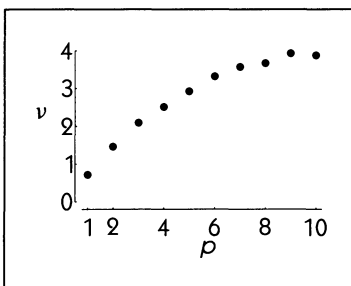
**Figure 6.38**  
Slope  $\nu$  versus  $r$  for embedding dimensions  $p = 1$  through  $p = 10$ .

Figure 6.38 shows how the slope  $\nu$  changes with  $r$  for different embedding dimensions. For the largest values of  $r$  found in the embedded time series, the slope approaches zero, regardless of the embedding dimension used. Estimating a dimension by using large  $r$  is somewhat like looking at the object from a great distance; no matter what the object, it will look like a single point—an object of dimension 0—just as a distant star looks like a single point of light.

For the smallest values of  $r$ , the slope  $\nu$  depends on and increases with the embedding dimension. This is characteristic of noise and reflects the measurement noise in the Model Five data.

For  $r \approx 0.5$  the slope is roughly the same for many different embedding dimensions. Figure 6.39 shows  $\nu$  (at  $r = 0.5$ ) for  $p = 1, 2, \dots, 10$ . The pattern is similar to that seen in Figure 6.37, and we conclude that the attractor of the Model Five data has a dimension of approximately 3.9. A range of values of  $r$  at which  $\nu$  is fairly constant is often called a **scaling region**.

The need to pick a specific range for  $r$  is one of the difficulties, and a great weakness, of estimating dimensions. In this case, we chose a range near  $r = 0.5$  because that value of  $r$  gives us the results closest to the ideal form shown in Figure 6.37. Here we have the advantage of knowing that there is an attractor, and therefore we have good reason to believe we are justified in our choice of  $r$ . Without this information, interpreting the meaning of the dimension calculation



**Figure 6.39**  
The correlation dimension versus embedding dimension  $p$  for  $r = 0.5$ .

could be quite challenging and problematic, and the results should be treated with skepticism. □

### DETERMINISM

We say that a system is deterministic when future events are causally set by past events. A finite-difference equation like  $x_{t+1} = f(x_t)$  is deterministic as long as  $f(x_t)$  has only one value for each possible value of  $x_t$ ; given the past value  $x_t$ , the function  $f(\cdot)$  determines the future value  $x_{t+1}$ . For Model Four, which is a finite-difference equation that produces chaos, if we knew  $x_0$  then by iteration we could calculate all future values of  $x_t$  using Eq. 6.23.

But, of course, we do not know  $x_0$  exactly, since our measurements are made with noise:  $D_t = x_t + W_t$ . If we take  $D_0$  as our estimate of the initial condition  $x_0$ , and iterate from this using Eq. 6.23, then sensitive dependence on initial conditions will cause our predictions to become faulty. For instance, suppose the true initial condition is  $x_0 = 0.37$  but that our measurement of

**Table 6.2**  $x_t$  and  $y_t$  from two identical finite-difference equations,  $x_{t+1} = 4x_t(1 - x_t)$  and  $y_{t+1} = 4y_t(1 - y_t)$ . Although  $x_0 \approx y_0$ , by time  $t = 5$  the values of  $x$  and  $y$  have moved far apart.

$x_t$	$y_t$	$t$
0.370	0.380	0
0.932	0.942	1
0.252	0.217	2
0.754	0.680	3
0.741	0.870	4
0.767	0.451	5
0.715	0.990	6
0.814	0.038	7
0.605	0.147	8
0.956	0.501	9
0.167	0.999	10

the initial condition is  $D_0 = 0.38$ . As shown in Table 6.2, our predictions based on Eq. 6.23 will diverge from the true values. At first the predictions will be quite good, but after five or so time steps, they become completely wrong. Even when the predictions are good, they are not perfect because of the measurement noise.

We can decide from data whether an underlying deterministic system is present: Use the data to construct a model of the dynamics, and then see whether the predictions made from this model are accurate. If the predictions are perfect, then the system is *completely* deterministic. If the predictions are good, but not perfect, then the system has a deterministic *component*. If the predictions are terrible, then the system is *not* deterministic at all.

We can construct dynamical models from data in a number of different ways. One of the simplest methods works as follows. Suppose that we make our measurements up to time  $T$  and that we want to make a prediction of the value at time  $T + 1$ .

1. Embed the time series to produce  $\mathbf{D}_T$ .
2. Take the embedded point at time  $T$ ,

$$\mathbf{D}_T = (D_T, D_{T-h}, \dots, D_{T-(p-1)h}),$$

and look through the rest of the embedded time series to find the point that is closest to  $\mathbf{D}_T$ . Let's say that this closest point has time index  $a$ . This means that  $\mathbf{D}_a$  is closer to  $\mathbf{D}_T$  than any other  $\mathbf{D}_i$ .

3. The definition of determinism is that future events are set causally by past events.  $\mathbf{D}_T$  describes the past events to  $D_{T+1}$ . Similarly  $\mathbf{D}_a$  describes the past events to the measurement  $D_{a+1}$ . If  $\mathbf{D}_T$  is close to  $\mathbf{D}_a$ , and if the system is deterministic, then we expect that  $D_{a+1}$  will be close to  $D_{T+1}$ . So we take as our prediction of  $D_{T+1}$  the measured value  $D_{a+1}$ . We will call this prediction  $\mathcal{P}_{T+1}$ .

This is a funny kind of model. Our previous models have consisted of sets of explicit equations. This model, which is used for prediction, consists of a data set (the measured time series) and a set of instructions (e.g., "find the nearest point  $\mathbf{D}_a$ "). The set of instructions is called an **algorithm**, and the model exists implicitly in the set of data and the algorithm. Such data-implicit models were uncommon before the advent of computers, but now they are commonplace and of increasing importance.

There are many variations on this simple model of dynamics. One elaboration is to take not just the time  $a$  where  $\mathbf{D}_a$  is closest to  $\mathbf{D}_T$ , but to take  $K$  different times  $a_1, a_2, \dots, a_K$  where  $\mathbf{D}_{a_1}, \mathbf{D}_{a_2}, \dots, \mathbf{D}_{a_K}$  are all close to  $\mathbf{D}_T$ . Then

the prediction of  $D_{T+1}$  is taken as the average of  $D_{a_1+1}, D_{a_2+1}, \dots, D_{a_K+1}$ :

$$\mathcal{P}_{T+1} = \frac{1}{K} \sum_{i=1}^K D_{a_i+1}. \quad (6.37)$$

Given a method for making a prediction  $\mathcal{P}_{T+1}$ , we need to make an actual measurement of  $D_{T+1}$  in order to decide if the prediction is good or bad. The difference between  $\mathcal{P}_{T+1}$  and  $D_{T+1}$  is the **prediction error**, which tells us about the quality of the prediction. Of course, a single prediction might be good or bad just by chance. To give a more meaningful indication of the determinism in the data, we can take the average of many prediction errors. Suppose we make  $2T$  measurements of a time series. We take the first half of the time series to construct a data-implicit model of the dynamics. Then we use the model to predict the values of the second half of the time series.

There are two ways to do this. One is to use the model to predict the value at time  $T + 1$ . Then, we construct a new embedded point using this predicted value  $\mathcal{P}_{T+1}$ :

$$\mathbf{D}_{T+1} = (\mathcal{P}_{T+1}, D_{T+1-h}, \dots, D_{T+1-(p-1)h}).$$

We then find the nearest points to  $\mathbf{D}_{T+1}$  to make a prediction of the value at time  $T + 2$ , which we call  $\mathcal{P}_{T+2}$ . This process can be iterated—we use past predictions to make future predictions. This method can in fact be used to extrapolate a time series beyond its measured values.

Although such extrapolation is useful to make predictions far in the future, for the purposes of assessing determinism in data it is better to use the measured data directly. In this second way of making predictions, in order to predict the value at time  $T + 2$ , we make the embedded point

$$\mathbf{D}_{T+1} = (D_{T+1}, D_{T+1-h}, \dots, D_{T+1+(p-1)h}).$$

Note that here the measurement at time  $T + 1$  is used, and not the prediction  $\mathcal{P}_{T+1}$ ; we are not using the past predictions to make future predictions.

Once we have made predictions for the second half of the time series, we can calculate a mean prediction error,  $\mathcal{E}$ :

$$\mathcal{E} = \frac{1}{T} \sum_{k=1}^T (D_{T+k} - \mathcal{P}_{T+k})^2. \quad (6.38)$$

Very large  $\mathcal{E}$  means the predictions are bad and the system is not deterministic. Conversely, small  $\mathcal{E}$  suggests that the system is deterministic.



How do we decide if  $\mathcal{E}$  is large or small? What do we compare it to as a standard? Suppose that we are very lazy and that instead of calculating a new prediction for each time  $T + k$ , we made the same prediction for all times. Presumably this would be a bad method of prediction, since it completely ignores the dynamics of the data. What is the best value to use for this bad method of prediction? We want to choose a value  $\mathcal{P}_{\text{lazy}}$  that minimizes the mean prediction error,

$$\mathcal{E}_{\text{lazy}} = \frac{1}{T} \sum_{k=1}^T (D_{T+k} - \mathcal{P}_{\text{lazy}})^2. \quad (6.39)$$

Compare Eq. 6.39 to Eq. 6.5. They are very similar; the minimization we are doing here is almost identical to the minimization we performed to find the sample mean  $M_{\text{est}}$ . In fact, a good value for  $\mathcal{P}_{\text{lazy}}$  is the sample mean of the time series. Given that we set  $\mathcal{P}_{\text{lazy}} = M_{\text{est}}$ , the mean prediction error  $\mathcal{E}_{\text{lazy}}$  in Eq. 6.39 is very similar to the variance of the time series,  $\sigma^2$ , as can be seen by comparison to Eq. 6.8.

A convenient way to decide if  $\mathcal{E}$  is large or small is to compare it to  $\mathcal{E}_{\text{lazy}}$ , or rather to the variance of the time series  $\sigma^2$ . We can do this by taking the ratio

$$\frac{\mathcal{E}}{\sigma^2}.$$

If this ratio is close to one, then the mean prediction error is large. If the ratio is close to zero, then the mean prediction error is small.

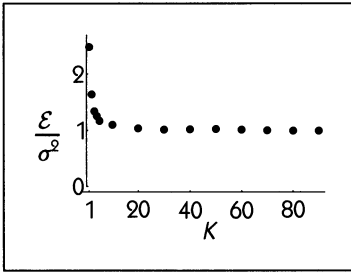
---

□ **EXAMPLE 6.5**

We can examine the data sets from Models One through Four to look for determinism. Rather than using just the closest neighboring point in the embedded time series to make the prediction, we will use  $K$  nearby points, as in Eq. 6.37.

**Solution:** For the example here, we will use an embedding dimension  $p = 1$  so that  $\mathbf{D}_t = D_t$ . As discussed in Section 6.6, it often makes sense to pick  $p > 1$ .

For the Model-One data, Figure 6.40 shows the ratio of the prediction error to the variance,  $\frac{\mathcal{E}}{\sigma^2}$ , as a function of the number of neighbors  $K$  used to make the prediction. When  $K$  is small, the ratio is greater than 1. This says that the model makes worse predictions than the lazy method of simply predicting the mean. As  $K$  becomes large, the ratio goes to unity.

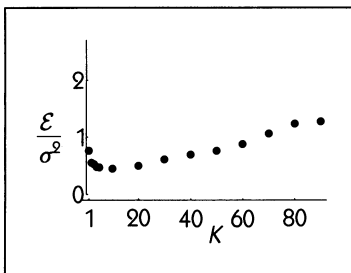


**Figure 6.40**  
 Prediction error  $\frac{\mathcal{E}}{\sigma^2}$  versus number of neighbors  $K$  used in the prediction for the data from Model One.

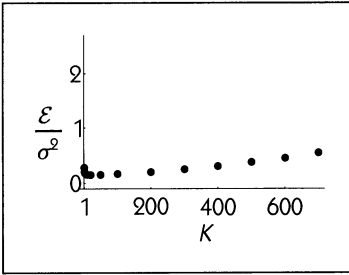
Recall that there are no active dynamics in Model One. The system is at a fixed point and all of the variability in the measurements is due to the random measurement noise  $W_t$ . There are no deterministic dynamics to predict, so it is not surprising that prediction is ineffective. Since there are 100 points in the time series, when  $K$  is near 100 virtually all of the points in the time series are being averaged together to produce the prediction, and Eq. 6.37 yields a prediction that is basically the mean,  $M_{\text{est}}$ . This is therefore the same as the lazy prediction method of using  $M_{\text{est}}$ , and so  $\mathcal{E} = \sigma^2$ . In fact, since there are no dynamics to the data, we cannot do better than using  $M_{\text{est}}$  to predict the time series, since  $M_{\text{est}}$  is the quantity that minimizes the prediction error, as in Eq. 6.5. When  $K$  is small, the number the prediction algorithm generates will vary around the mean since only a few of the points are used in the averaging. As described in Section 6.3, the fewer points used, the more the prediction will vary around the mean. Since the mean gives the best possible prediction for this data, any deviation from the mean will give worse predictions.

For the Model-One data, the prediction error is large (i.e.,  $\frac{\mathcal{E}}{\sigma^2} \geq 1$ ) and we are justified in concluding that the data are random, consistent with the known mechanism of the model.

Model Two, in contrast, shows definite predictability. The ratio  $\frac{\mathcal{E}}{\sigma^2}$  is shown in Figure 6.41. Here the ratio is less than unity for small  $K$ , and approaches unity as  $K$  approaches 100, the number of points in the time series. At very small  $K$ , the predictions are worse than at intermediate  $K$ —averaging the five to ten nearest



**Figure 6.41**  
 $\frac{\mathcal{E}}{\sigma^2}$  versus  $K$  for the data from Model Two.

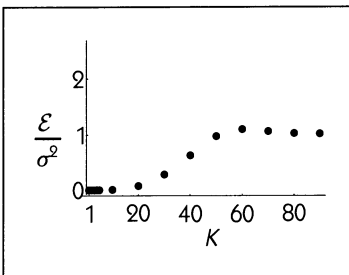


**Figure 6.42**  
 $\frac{\varepsilon}{\sigma^2}$  versus  $K$  for the data from Model Three.

points produces a better prediction than simply using the single nearest point. This prediction is substantially better than using  $M_{\text{est}}$ . Therefore, we conclude that the data from Model Two contain some determinism but are not completely deterministic. This is consistent with Eq. 6.11, which has a deterministic component to its dynamics ( $\rho x_t$ ) in addition to the random dynamical noise ( $v_t$ ) and measurement noise ( $W_t$ ). The 1000 data points from Model Three produce quite similar results, as shown in Figure 6.42.

Model Four is completely deterministic. The data analysis using prediction error confirms this; the prediction error is virtually zero for small  $K$  and, as expected, the ratio  $\frac{\varepsilon}{\sigma^2}$  approaches unity as  $K$  approaches the number of points in the time series (see Figure 6.43).

The prediction results for Models One, Two, and Three could have been anticipated from the the autocorrelation function. The autocorrelation function  $R(k)$  for Model One shows no correlation between measured values  $D_T$  and  $D_{T+1}$  (Figure 6.9). The autocorrelation functions for the data from Models Two and Three show quite strong correlations between  $D_T$  and  $D_{T+1}$  (Figures 6.10 and 6.11). This is not the whole story, however. Note that although the autocorrelation function for Model Four (Figure 6.14) is much the same as that for Model One, the prediction results are completely different. The prediction method is sensitive to the nonlinearity in Model Four, whereas the autocorrelation function is not.



**Figure 6.43**  
 $\frac{\varepsilon}{\sigma^2}$  versus  $K$  for the data from Model Four.

[Technical note: In this example we have made a slight modification to the calculation of  $\mathcal{E}$  described in the body of the text. There we divided the time series into two halves and used the first half to construct the data-implicit model, while using the second half to evaluate the predictive abilities of the model. Here we do not divide the time series. We use all of the data in constructing the model, and we evaluate the predictions based on the same data. However, when making a prediction  $\mathcal{P}_{T+1}$ , we exclude the point  $\mathbf{D}_T$  from the data used to generate the model. If we did not do this, then the closest point to  $\mathbf{D}_T$  would obviously be itself, which would give  $\mathcal{P}_{T+1} = D_{T+1}$  (for  $K = 1$ ). This would be a perfect prediction, but completely worthless since we would be predicting what we already knew. By excluding  $\mathbf{D}_T$ , we avoid this problem and can use all of the data at once.]  $\square$

## DYNAMICS IN ACTION

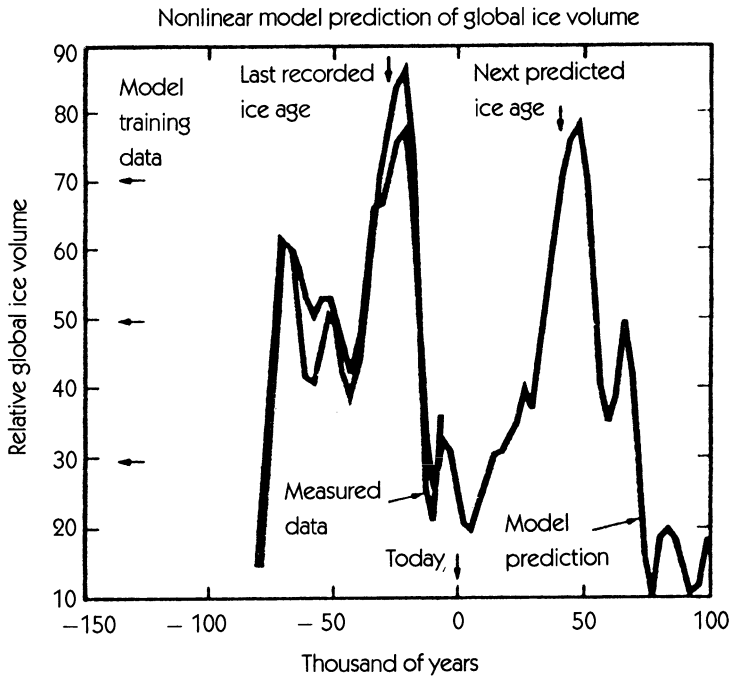
### 19 PREDICTING THE NEXT ICE AGE

Over the past millions of years, glaciers have repeatedly built up in the northern hemisphere, covering land that is now in temperate climates. The last ice age ended roughly 10,000 years ago.

When will the next ice age come? This is a question for climate prediction, as opposed to the short-term weather prediction with which we are all familiar from the nightly news. Unlike the day-to-day weather, the fundamental principles underlying climate are largely unknown. This means that data-implicit models may have an important role to play. The following application of data-implicit modeling to ice-age prediction is drawn from Hunter (1992).

An indirect record of global ice volume is contained in the ratio of two oxygen isotopes,  $\text{O}^{16}$  and  $\text{O}^{18}$ , found in the shells of Foraminifera that are found at the bottom of the ocean. Cores taken at the ocean bottom indicate this  $\frac{\text{O}^{16}}{\text{O}^{18}}$  ratio over the past 800,000 years, a period that includes roughly eight ice ages.

The mechanism that causes ice ages is not known. The Milankovitch theory is based on the idea that the amount of summer sunlight in the Northern hemisphere is a dominant factor. If summer sunlight is too low, snow that fell during the winter cannot all melt in the summer, and so ice gradually accumulates.



Measured global ice volume and predictions from a data-implicit model. Redrawn from Hunter (1992).

The amount of summer sunlight depends on the luminosity of the Sun, but also on parameters of the Earth's orbit. The closer the Earth is to the Sun, the more light falls on the Earth. The Sun-Earth distance varies over the course of the year and is governed by the **eccentricity** of the Earth's orbit. The axis of the Earth's rotation is slightly inclined with respect to its orbital plane; this **obliquity** is what causes the yearly seasonal cycle. The angle of inclination changes over time, just as a spinning top wobbles. This is called *precession*. Scientists have a good understanding of these orbital parameters and are able to calculate their past and future values.

What is not known, however, is how the accumulation of snow depends on these parameters. What seems to be important is the eccentricity, which modulates the amplitude of precession. The eccentricity is therefore related to the variability in the angle of inclination. When there is much variability, past accumulations of snow have

an occasional chance to melt; when there is little variability, snow accumulations may not have the opportunity to melt completely.

Other factors also make modeling difficult. When there is snow on the ground over large fractions of the Earth's surface, more sunlight is reflected back into space, increasing the propensity for snow to accumulate. However, changes in ocean level and cloud cover may have countervailing effects.

Hunter used past records of global ice volume, as inferred from measured  $\frac{O^{16}}{O^{18}}$  ratios in ocean cores, and constructed a data-implicit model of global ice volume (see the figure on the previous page). Since future orbital parameters are well known, he used them as well.

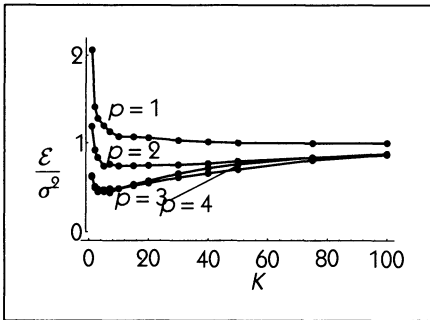
The indirectly measured global ice volume and the calculated orbital parameters for 800,000 years ago until 60,000 years ago were used to construct a data-implicit model. This model was iterated until the present, in order to confirm that the model could have predicted the last ice age, which occurred 10,000 years ago. Starting at the present, and using calculated future values of orbital parameters, the model is iterated to predict future ice volume. According to the model's predictions, the next ice age will start in 30,000 years and will peak in 50,000 years.

### □ EXAMPLE 6.6

From the data from Model Five, use nonlinear predictability to estimate how many variables are involved in the dynamics.

**Solution:** Nonlinear prediction allows us to look for a functional relationship between  $V_{t+1}$  and previous values of  $V$ . If the predictability is good—if  $\mathcal{E}$  is small—then  $V_{t+1}$  is determined by previous values of  $V$ . Figure 6.44 shows  $\mathcal{E}$  for embedding dimensions  $p = 1, 2, 3,$  and  $4$ . The best predictions are made when roughly ten nearest neighbors are used, and these predictions appear to be improving as  $p$  is increased. Figure 6.45 shows  $\mathcal{E}$  for  $k = 10$  nearest neighbors for embedding dimensions  $p = 1$  through  $p = 10$ : The predictions are best for  $p = 3$  or  $p = 4$ , suggesting that three or four previous values of  $V$  do the best job of determining  $V_{t+1}$ . This is consistent with the fact that four coupled dynamical equations were used to generate the data.

Perhaps it is surprising that using more than four previous values of  $V$  does not lead to a better prediction. After all, using more information can't hurt, can it? Due to the sensitive dependence on initial conditions in chaotic systems, values



**Figure 6.44**

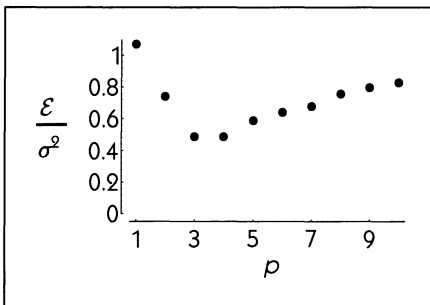
Nonlinear predictability  $\frac{\varepsilon}{\sigma^2}$  versus the number of nearest neighbors  $K$  used to make the prediction. Results are shown for embedding dimensions  $p = 1$  through  $p = 4$ .

of  $V$  from far in the past do not contain much information about the value of  $V_{t+1}$ . If we use these irrelevant values in finding nearest neighbors, predictions will be poor. This irrelevance of the distant past limits our ability to investigate high-dimensional chaotic systems using time-lag embedding. □

### SENSITIVE DEPENDENCE ON INITIAL CONDITIONS: LYAPUNOV EXPONENTS

Suppose that we have two copies of Model Four—one using the variable  $x$ , the other using  $y$ —that are identical except that their initial condition can be made to differ. We start with  $x_0$  and  $y_0$  very close together. As we iterate the system from the two initial times,  $x_i$  and  $y_i$  start to move apart, slowly at first and then more rapidly. Eventually,  $x_i$  and  $y_i$  show no correlation with one another, yet the dynamics of both arise from the same equation (see Table 6.2 on page 326).

This “stretching apart” of the distance between initially nearby points is called **sensitive dependence on initial conditions**. One way to characterize a chaotic dynamical system is to measure the strength of this sensitive dependence.



**Figure 6.45**

Nonlinear predictability  $\frac{\varepsilon}{\sigma^2}$  versus embedding dimension.  $K = 10$  nearest neighbors were used to make the prediction.

In order to develop some ideas about how to measure sensitive dependence on initial conditions, let's assume that we have perfectly deterministic dynamics

$$x_{t+1} = f(x_t). \quad (6.40)$$

If we have two initial conditions  $x_0$  and  $y_0$  whose initial separation  $|x_0 - y_0|$  is very small, the separation after one time step is

$$|x_1 - y_1| = |f(x_0) - f(y_0)| = \frac{|f(x_0) - f(y_0)|}{|x_0 - y_0|} |x_0 - y_0| \approx \left| \frac{df}{dx} \right|_{x_0} |x_0 - y_0|,$$

where we make use of the definition of the derivative

$$\left. \frac{df}{dx} \right|_{x_0} = \lim_{y_0 \rightarrow x_0} \frac{f(x_0) - f(y_0)}{x_0 - y_0}.$$

The strength of the sensitive dependence on initial conditions is therefore  $\left| \frac{df}{dx} \right|_{x_0}$ . Clearly, this depends on the initial condition  $x_0$ . What we want, though, is a number that describes sensitive dependence for the map as a whole, and not just at one initial condition; we want to “average” all initial conditions. We can motivate the proper form of averaging by noting that

$$|x_2 - y_2| \approx \left| \frac{df}{dx} \right|_{x_1} |x_1 - y_1| \approx \left| \frac{df}{dx} \right|_{x_1} \left| \frac{df}{dx} \right|_{x_0} |x_0 - y_0|$$

and, by iteration,

$$|x_n - y_n| \approx \left| \prod_{t=0}^{n-1} \frac{df}{dx} \right|_{x_t} |x_0 - y_0|$$

(where  $\prod$  means multiplication in the same way that  $\sum$  means summation). Recalling from Chapter 1 that the solution to the linear finite-difference equation  $x_{t+1} = ax_t$  is  $x_n = a^n x_0$ , we see that the average separation per iteration (which is  $a$  for the linear system) is

$$\left( \left| \prod_{t=0}^{n-1} \frac{df}{dx} \right|_{x_t} \right)^{\frac{1}{n}}.$$

This is the **geometric mean** of the quantities  $\left| \frac{df}{dx} \right|_{x_t}$ . The term **Lyapunov exponent** is used for the logarithm of this average separation per iteration.



A procedure, then, for quantifying the sensitive dependence on initial conditions from a one-dimensional finite-difference map is as follows:

1. Iterate the map to generate a sequence of values  $x_0, x_1, x_2, \dots, x_{n-1}$ .
2. Calculate the slope of the map at each of the points  $x_0, \dots, x_{n-1}$ .
3. Calculate the absolute value of the geometric mean of the values in step (2). (If you are doing this on a computer, beware of round-off errors.) This value represents the sensitive dependence on initial conditions of the map as a whole.

□ **EXAMPLE 6.7**

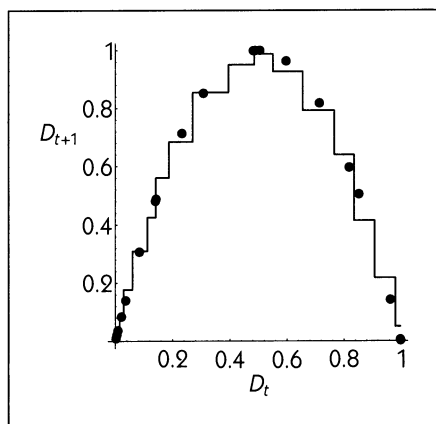
Estimate a Lyapunov exponent for the Model Four data.

**Solution:** The first step is to use the data to construct a prediction model as described earlier in this section. We can then use this prediction model as the function  $f(x)$  in Eq. 6.40.

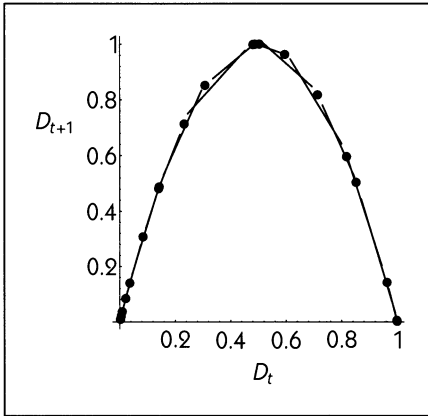
A prediction model of the form of Eq. 6.37 is not adequate for the purpose of finding Lyapunov exponents from data. Although the data-implicit model of Eq. 6.37 provides us with a value of  $f(\cdot)$  at each data point, it does not tell us what  $\frac{df}{dx}$  is, and this information is needed to find the Lyapunov exponent.

Instead, we need to fit a prediction model that does specify the slope  $\frac{df}{dx}$  at each data point. Many different models could do this job. Possibly the simplest is to fit short line segments to the data—a **locally linear model**. The slope of these line segments then gives  $\frac{df}{dx}$ .

Figures 6.46 and 6.47 show a locally constant model (Eq. 6.37) and a locally linear model fit to some of the Model Four data in a return plot (a one-dimensional embedding). For these plots,  $K = 3$  points were used. In a two-dimensional



**Figure 6.46**  
A model of the form Eq. 6.37 fits the data to short, level line segments—a locally constant model.



**Figure 6.47**

A locally linear model fits the data to short line segments. The slope of the line segment at each point gives  $\frac{df}{dx}$  at that point.

embedding, small portions of planes would be used instead of line segments; in a higher-dimensional embedding, hyperplanes would be used. □

### ➤ MODEL SIX

Models One, Two, and Three are linear models. A general form of a multi-dimensional, linear, finite-difference equation is the **autoregressive** model,

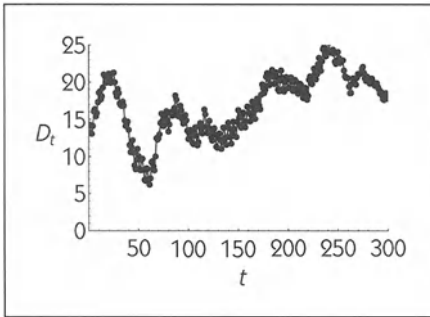
$$x_{t+1} = a_0x_t + a_1x_{t-1} + a_2x_{t-2} + \cdots + a_{p-1}x_{t-(p-1)} + v_t. \quad (6.41)$$

In this equation,  $x_{t+1}$  depends on the  $p$  previous values:  $x_t, \dots, x_{t-(p-1)}$ . The parameters  $a_0, \dots, a_{p-1}$  are fixed in time and play the same role as  $\rho$  in Eq. 6.3 or Eq. 6.11. The dynamical noise at time  $t$  is  $v_t$  and is almost always assumed to be Gaussian white noise. Here we will assume that there is no measurement noise, that is,  $D_t = x_t$ .

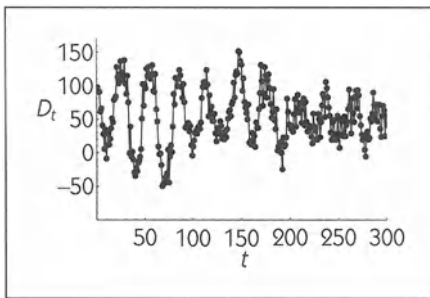
Equation 6.41 is capable of producing many different types of output, depending on the values of the parameters  $a_0, \dots, a_{p-1}$ . Three different examples are shown in Figures 6.48, 6.49, and 6.50.

The analysis of Eq. 6.41 is central to a number of fields in science and technology, and, correspondingly, there are a number of different names that can be found in the technical literature: Statisticians tend to use the term “autoregressive (AR) model,” control engineers use “all-pole model,” signal processing engineers use “infinite impulse-response filter,” while physicists prefer “maximum entropy model.”

Whatever the name, the dynamics displayed by the model are those we have already seen in linear models: exponential growth and decay, and oscillations whose amplitudes either grow or decay exponentially. However, in contrast to



**Figure 6.48**  
An output of Model Six for  
 $a_0 = 1.39$ ,  $a_1 = -0.703$ ,  
 $a_2 = 0.038$ ,  $a_3 = 0.735$ ,  
 $a_4 = -0.46$ .

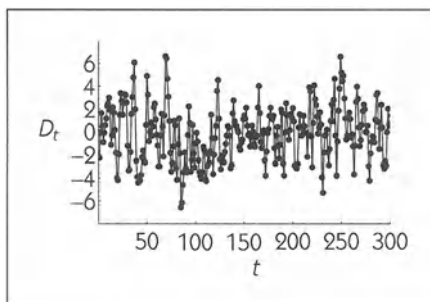


**Figure 6.49**  
An output of Model Six for  
 $a_0 = 0.677$ ,  $a_1 = 0.175$ ,  
 $a_2 = 0.297$ ,  $a_3 = 0.006$ ,  
 $a_4 = -0.114$ ,  $a_5 = -0.083$ ,  
 $a_6 = -0.025$ .

the linear models we have already studied, Eq. 6.41 can produce several different frequencies of oscillation at the same time, with the amplitudes of the different frequencies growing or decaying exponentially at different rates. The result is that Eq. 6.41 is quite general, suitable for modeling many diverse types of data.

For any given time series, the question of how to find the best  $a_0, a_1, \dots, a_{p-1}$  can be addressed in the spirit of the prediction models we have already studied. We use Eq. 6.41 to make a **linear prediction** at time  $T + 1$  using the measurements  $D_t$  made prior to that time:

$$\mathcal{P}_{T+1} = a_0 D_T + a_1 D_{T-1} + \dots + a_{p-1} D_{T-(p-1)}.$$



**Figure 6.50**  
An output of Model Six for  
 $a_0 = 1.05$ ,  $a_1 = -0.5$ .

We want to select the parameters  $a_0, a_1, \dots, a_{p-1}$  to minimize the prediction error using a least-squares criterion:

$$\begin{aligned}\mathcal{E} &= \sum_{i=p}^{N-1} (\mathcal{P}_i - D_i)^2 \\ &= \sum_{i=p}^{N-1} (v_i)^2.\end{aligned}\tag{6.42}$$

Finding the parameters  $a_0, a_1, \dots, a_{p-1}$  that minimize  $\mathcal{E}$  is somewhat technical, but the upshot is that there is a formula that specifies the parameters in terms of the autocorrelation function  $R(k)$  of the data. This means that the autocorrelation function uniquely specifies a linear model of the data in the form of Eq. 6.41. This model is sometimes called an **optimal linear model** because it uses the parameters  $a_0, a_1, \dots, a_{p-1}$  that minimize  $\mathcal{E}$ , but it should be understood that the model is “optimal” only relative to other linear models and that nonlinear models might produce a smaller prediction error.

The question of how to select  $p$ , called the **model order**, is more subtle. Ultimately, this is a philosophical question, and the technical issues surrounding it are well beyond the scope of this text.

Fortunately, we will see that the model order  $p$  is not important for our purposes in this chapter. The only facts about the autoregressive model that we need to keep in mind are

1. It is a model with linear dynamics.
2. Optimal model parameters  $a_0, a_1, \dots, a_{p-1}$  can be selected that minimize the prediction error  $\mathcal{E}$  for any given time series.
3. The optimal parameters can be calculated from the autocorrelation function  $R(k)$ .

In particular, this last point means that if two time series have the same autocorrelation function, then they have the same optimal linear model.

The main question we will attempt to answer using Model Six is, are the data well described by a linear model, or is there evidence in the data for nonlinear dynamics? □

## 6.8 DETECTING CHAOS AND NONLINEARITY

We have a time series and we want to know whether the system that produced it is chaotic. How can we tell?

The answer is easy: We cannot. Any finite amount of data might come from a chaotic system, or might come from a random system. The situation is similar to the famous scene of monkeys at typewriters: If we put enough monkeys at enough typewriters for long enough, all the works of Shakespeare will eventually be produced. So, given a Shakespearean tragedy, how can we know for sure whether it was produced by the Immortal Bard, or by a monkey pecking randomly at a typewriter?

O heavy lightness, serious vanity,  
 Misshapen chaos of well-seeming forms,  
 Feather of lead, bright smoke, cold fire, sick health,  
 Still-waking sleep, that is not what it is!

*William Shakespeare (1564–1616), Romeo and Juliet, Act 1, Scene 1*

,Fs a teetsdl,ss t cfrihohpincusfs l e  
 od egt acgl,tm gkiwivt waolakate ihoe-s to mheii eeia mslsy  
 sihnofsyStena ovs  
 ilwi,rrb elehilOg h ! i,Mhskohr ,pnf rlana hset-eea n

*A typing monkey (simulated)*

Of course, only a fool would claim that the works of Shakespeare were generated by a typing monkey. We can look at even a small fragment of Shakespeare's works, and see structure such as words and syntax, and divine more abstract structure such as meaning. The chances of seeing such structure in randomly generated letters are so small that we discount the very possibility as absurd.

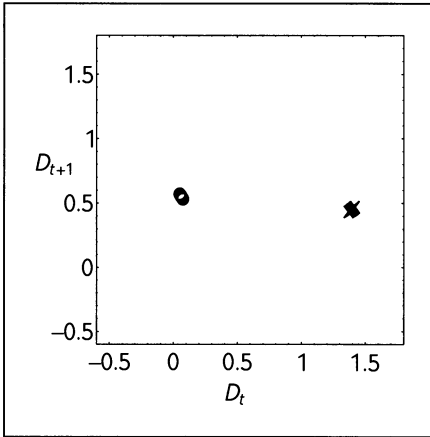
We can look at time series in a similar way. Suppose that we look for chaotic structure in a time series. If we see it, then we can argue that the time series is unlikely to have been generated by random noise.

But what is "chaotic structure"? We have already seen that the definition of chaos includes three elements:

- determinism,
- aperiodicity, and
- sensitive dependence on initial conditions.

In Section 6.7 we introduced several ways to quantify these elements in a time series.

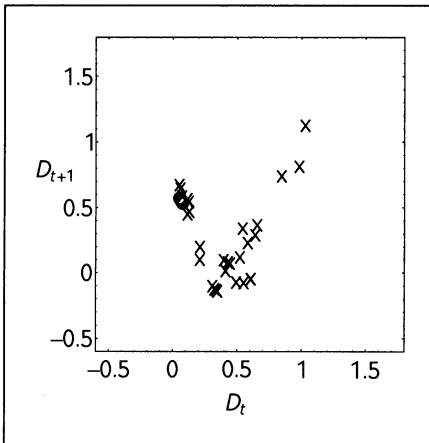
The fourth element in the definition of chaos, "boundedness," is not of much use to us here. It is easy to generate random numbers that are bounded. For instance, if we write each number from a time series on a deck of cards, and then shuffle the cards, the deck will serve as a random-number generator that is bounded. So, whatever structure there is in "boundedness" cannot distinguish between chaos and randomness.



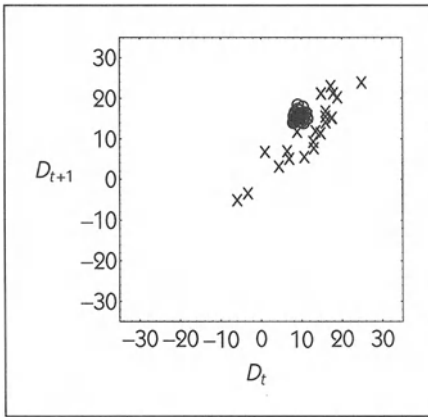
**Figure 6.51**  
 The Model-Five dynamics started at several nearby initial condition. Each o marks an initial condition plotted as  $D_{t+1}$  versus  $D_t$ ; there are 30 of them. The x's mark the state after iterating the Model-Five dynamics for 5 time steps, from each of the 30 initial conditions.

Actually, we have to be careful even in using the other three elements in the definition of chaos. Consider the use of determinism in detecting chaos. Models Two and Three show deterministic structure (see Figures 6.41 and 6.42) even though they have only linear dynamics and are therefore incapable of producing chaos.

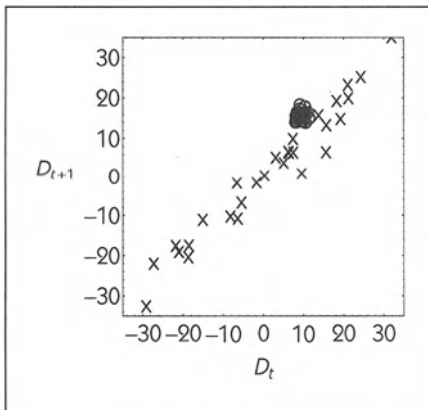
A similar problem arises when looking at sensitive dependence on initial conditions. Figures 6.51 and 6.52 show what happens to two small clouds of initial conditions in the dynamics of Model Five. The sensitive dependence on initial conditions in the chaotic dynamics causes the clouds to broaden over time. Very similar behavior can be observed in Figures 6.53 and 6.54, which show the linear dynamics of Model Three. In this case, the cloud broadening is due to dynamical noise, not chaos.



**Figure 6.52**  
 After 200 time steps of the chaotic Model Five dynamics, the cloud of initial conditions spread out, showing the sensitive dependence on initial conditions.



**Figure 6.53** The Model-Three dynamics, started at several nearby initial conditions. Each of the o's marks one initial condition, plotted as  $D_{t+1}$  versus  $D_t$ . The x's mark the position that the Model-Three dynamics take each initial condition after 5 time steps. The cloud of initial conditions has spread out, reflecting the influence of the noise in the Model-Three dynamics, but visually, it is hard to distinguish this from sensitive dependence on initial conditions.



**Figure 6.54** After 200 time steps of the linear Model Three dynamics

## TOWARD AN APPROPRIATE NULL HYPOTHESIS

Scientists often work by putting forward a hypothesis and then trying to find an example to refute it, to show that the hypothesis is incorrect. As applied to chaos and time series, following this procedure means that one does not try to prove that a time series is chaotic, but rather to refute or reject some other hypothesis. The hypothesis that one is trying to reject is called the **null hypothesis**.

What is an appropriate null hypothesis, when thinking about the possibility of chaos in a time series? Up until now, we have pointed to the dichotomy between

chaos and randomness, suggesting that we should use randomness as the null hypothesis. But what do we mean by randomness?

One possibility is white noise: Each measurement  $D_t$  is independent of every other measurement. This is a convenient hypothesis because it readily suggests a test; if we can find any dependence between measurements, then we can reject the null hypothesis. Unfortunately, we have already seen cases where we can reject the white noise null hypothesis even when there is no chaos. For example, in the case of Models Two and Three, the autocorrelation function shows that there is dependence between successive measurements, even though these models involve only linear dynamics and therefore cannot produce chaos.

The white noise null hypothesis is somewhat like the hypothesis of typing monkeys. It is not too often that we have to decide whether a sentence was written by Shakespeare or by a typing monkey. More likely, the problem is one of deciding whether a work was written by Shakespeare or, say, by Alexander Pope—typing monkeys can be ruled out from the very beginning, and doing so tells us nothing useful about the true author. White noise is simply too restrictive to be a good null hypothesis when testing for chaos.

Lo! thy dread Empire CHAOS! is restor'd;  
 Light dies before thy uncreating word;  
 Thy hand, great Anarch! lets the curtain fall;  
 And Universal Darkness buries All.

*Alexander Pope (1688–1744), The Dunciad*

A better null hypothesis is provided by Model Six, which gives an optimal linear model of any data set. The hypothesis here is that the dynamics are linear, with Gaussian white noise random inputs. We will therefore call this the **linear-dynamics null hypothesis**. This null hypothesis is inconsistent with the possibility of chaos, since linear dynamics cannot produce chaos. This means that if a time series is chaotic, we should in principle be able to reject the null hypothesis.

### TESTING THE NULL HYPOTHESIS WITH SURROGATE DATA

We have a measured time series and a null hypothesis. How do we test whether the time series is inconsistent with the null hypothesis? We use what is called a **discriminating statistic**, some quantity that can be computed both from the measured time series and also from a time series that is consistent with the null hypothesis. Three discriminating statistics that are relevant to chaos are the nonlinear predictability  $\mathcal{E}$ , the Lyapunov exponent, and the correlation dimension, but other discriminating statistics can potentially be used.

We test whether the time series is consistent with the null hypothesis in the following way: First, calculate the value of the discriminating statistic on the



measured time series. We will call this value  $\mathcal{D}$ . Then find the range of values for the discriminating statistic for time series that are consistent with the null hypothesis. If  $\mathcal{D}$  falls within this range, then the discriminating statistic cannot distinguish between the null hypothesis and the measured time series. On the other hand, if  $\mathcal{D}$  falls outside the range, then the time series is inconsistent with the null hypothesis.

One way of finding the range of values of the discriminating statistic for a time series consistent with the null hypothesis is this: Generate many different time series that are consistent with the null hypothesis, and then calculate the value of the discriminating statistic for each of these time series. We will call this value  $\mathcal{S}_i$  for each of the null hypothesis time series. Data generated to be consistent with the null hypothesis are called **surrogate data**. This process of using surrogate data to find the range of values for data consistent with the null hypothesis is called **bootstrapping**.

There is a particularly simple method for generating surrogate data consistent with the null hypothesis of linear dynamics with Gaussian white noise inputs. In Section 6.5, we saw that a linear dynamical system can be characterized by a transfer function, which consists of two parts: the transfer gain  $G(\omega)$  and the transfer phase  $\Phi(\omega)$ . The transfer function describes the relationship between any input and the output of the linear dynamical system. In order to calculate the transfer function, we need to measure both the input and the output. However, if the input is Gaussian white noise, then even if we do not measure the input we can calculate the transfer gain  $G(\omega)$ . The transfer phase  $\Phi(\omega)$  will be random numbers between 0 and  $2\pi$  at each  $\omega$ . Or, to be more precise, since we don't measure the input, the phases  $\Phi(\omega)$  look random to us, even though they are determined by the input.

If we took the same linear dynamical system and gave as input a new sequence of Gaussian white noise random inputs, then the transfer gain would be the same as before, but the transfer phase would be a new set of random numbers. In order to simulate this, we can take the following steps:

1. Compute the Fourier transform of the original time series. This will consist of an amplitude  $A_{\text{output}}(\omega)$  and a phase  $\phi_{\text{output}}(\omega)$  at each frequency  $\omega$ .
2. Replace the phases  $\phi_{\text{output}}(\omega)$  with random numbers ranging between 0 and  $2\pi$ . Note that this has no effect on the amplitude  $A_{\text{output}}(\omega)$ . (Technical note: In the original time series,  $\phi_{\text{output}}(\omega) = -\phi_{\text{output}}(-\omega)$ , and this symmetry should be maintained when assigning random phases.)
3. Compute the inverse Fourier transform of  $A_{\text{output}}(\omega)$  and the randomized  $\phi_{\text{output}}(\omega)$ . This produces a new time series, the surrogate data.

The surrogate data has the same  $A_{\text{output}}(\omega)$  as the original time series. Since the power spectrum is proportional to  $A_{\text{output}}^2(\omega)$ , the surrogate data time series has exactly the same power spectrum as the original. Since the autocorrelation function is the Fourier transform of the power spectrum, the surrogate data also have exactly the same autocorrelation function as the original time series. This means that it is impossible to discriminate between the surrogate and the original based on the autocorrelation function or anything that is derived from the autocorrelation function.

Recall from Model Six that the parameters in the optimal linear model of a time series are derived from the autocorrelation function. Since the surrogate data and the original data have exactly the same autocorrelation function, both the surrogate data and the original have identical optimal linear models. In generating surrogate data in this manner, we do not need to specify the model order  $p$  of the linear model: The surrogate data and the original time series have the same optimal linear model for any model order.

Because we want to find the range of values for the discriminating statistic for data consistent with the null hypothesis, we will want to make many different surrogate data time series. Each one is called a **realization** of the null hypothesis. If we want to make many different realizations of the null hypothesis, then we follow the same process, using different random numbers for the phases  $\phi_{\text{output}}(\omega)$  in step 2. Typically, 10 to 100 different realizations are used.

Now the procedure is easy: Calculate the value of the discriminating statistic for the original time series and for each of the surrogate data time series. If the value for the original time series is outside the range of values found for the surrogates, then the original time series is inconsistent with the null hypothesis.

We have considerable latitude in choosing a discriminating statistic. If one is interested in chaos, then an appropriate discriminating statistic is the nonlinear predictability  $\mathcal{E}$ , or the Lyapunov exponent, or the correlation dimension. However, in principle, any discriminating statistic could be used, even if it has nothing whatsoever to do with chaos. Whatever discriminating statistic is being used, the result indicates whether the original time series is consistent with the null hypothesis of linear dynamics with Gaussian white noise inputs.

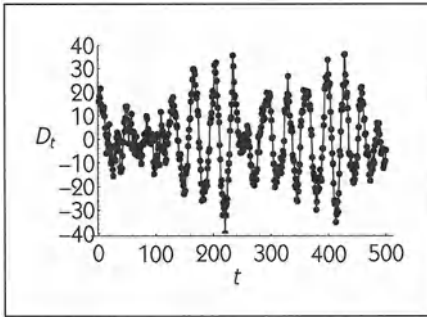
---

□ **EXAMPLE 6.8**

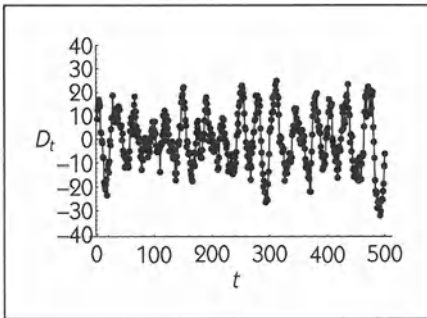
Use surrogate data to indicate whether the data from Model Three and the data from Model Five reflect linear or nonlinear dynamics.

**Solution:** We follow this sequence of steps:

- Generate many different realizations of surrogate data for each of the data sets. In this case, we will use ten realizations for each data set. Some examples are shown in Figures 6.9 through 6.14.



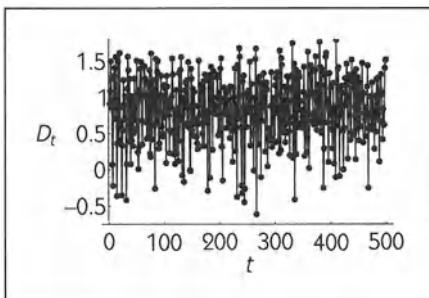
**Figure 6.55**  
A realization of surrogate data  
for Model Three.



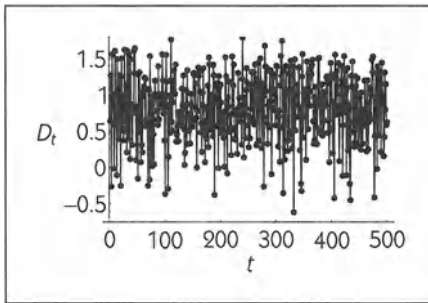
**Figure 6.56**  
A second realization of surrogate  
data for Model Three.

- Calculate a discriminating statistic on the original data and on the corresponding surrogates. For this example, we will use nonlinear predictability  $\frac{\xi}{\sigma^2}$  as a discriminating statistic. We will use an embedding dimension of  $p = 4$  for both the Model Three and Model Four data.
- See whether the value of the discriminating statistic for the original data lies outside the range for the many realizations of the surrogate data.

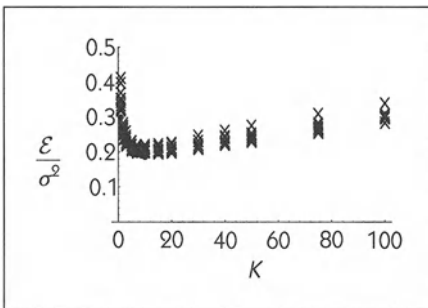
Figure 6.59 shows the nonlinear predictability  $\frac{\xi}{\sigma^2}$  for the Model Three data and for ten surrogates generated from this data. The Model Three data's predictability lies within the range of the surrogates. This means that nonlinear



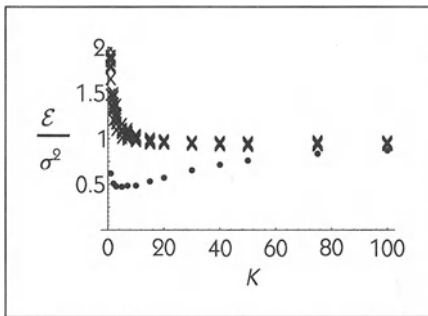
**Figure 6.57**  
A realization of surrogate data  
for Model Five.



**Figure 6.58**  
A second Model Five surrogate.



**Figure 6.59**  
Nonlinear predictability  $\frac{\epsilon}{\sigma^2}$  versus  $K$  for the Model Three data (dots) and ten surrogate data sets (x). An embedding dimension of  $p = 4$  was used.



**Figure 6.60**  
Nonlinear predictability  $\frac{\epsilon}{\sigma^2}$  versus  $K$  for the Model Five data (dots) and ten surrogate data sets (x).

predictability does not refute the null hypothesis that the data arise from a linear model. Note that  $\frac{\epsilon}{\sigma^2} \ll 1$  for the Model Three data. This means that there is some determinism in Model Three. However, the surrogate data analysis tells us that this determinism is consistent with linear dynamics.

The results for the Model Five data are different. (See Figure 6.60). For small values of  $K$ , the Model Five data's predictability lies well outside the range of values found for the surrogate data. This allows us to reject the linear dynamics null hypothesis.



## 6.9 ALGORITHMS AND ANSWERS

Suppose you have a time series from some field measurement, for example the Standard & Poor's stock price index measured each day. Suppose also that you have two computer programs, one to calculate a Lyapunov exponent and one to calculate a correlation dimension. You run the programs on your time-series data, setting parameters such as the number of nearest neighbors to use or the scaling region. The computer prints a message saying that the Lyapunov exponent is 0.2 and the correlation dimension is 3.1.

Does this mean that stock prices have a self-similar attractor with a dimension of 3.1 and that there is sensitive dependence on initial conditions? Not at all! Most computer programs are written to provide an output for any input. In designing the algorithm, the programmer makes certain assumptions. For example, the algorithm for quantifying sensitive dependence on initial conditions, described in Section 6.7, assumes that data are well described by a deterministic finite-difference equation. The algorithm for calculating the correlation dimension assumes that the trajectory lies on an attractor. If the data do not satisfy these assumptions, then the output of the algorithms should not be interpreted as answers to questions such as "Is there sensitive dependence on initial conditions?" or "Is there an attractor with a fractal dimension?"

The advantage of using surrogate data and testing the null hypothesis is that the assumptions behind the algorithms become unimportant. This is because we are no longer answering questions such as "Is there an attractor with a fractal dimension?" Instead, we are asking whether or not the time series is consistent with the null hypothesis.

It is tempting to believe that if we use a discriminating statistic that is motivated by chaos—for instance, the nonlinear predictability  $\mathcal{E}$ —then finding that the original time series is inconsistent with the linear dynamics null hypothesis means that the time series is chaotic. This is incorrect. All we can conclude, no matter which discriminating statistic is used, is that the time series is inconsistent with the null hypothesis. Some nonchaos phenomena that can lead to rejection of the null hypothesis are

- nonstationarity of the data;
- non-Gaussian white noise random inputs;
- nonlinearities in the measurement process;
- nonlinearities in the dynamics that do not involve chaos, such as the nonlinearity seen in the Lotka-Volterra equations, Section 5.5.

If the linear dynamics null hypothesis is rejected, then we have still not proved that the dynamics are chaotic. As of this writing, there is no general and

standard test for any of the above nonchaos phenomena in time series that involve measurement noise or dynamical noise.

Conversely, if we cannot reject the null hypothesis, then all we can say is that “the measured data are consistent with the null hypothesis.” This does not mean that the null hypothesis is correct—it means that we don’t have any evidence that the hypothesis is incorrect. There might be too much measurement noise to reject the linear dynamics null hypothesis, or the dynamics might be chaotic but of such high dimension or high Lyapunov exponent that we do not have enough data to see the chaotic structure.

---

## SOURCES AND NOTES

There are many branches in the literature surrounding time series analysis, but the root of time series analysis emerges from statistics. A knowledge of basic statistics is indispensable to the study of time series, and there are many introductory statistics textbooks such as Snedecor and Cochran (1989). The “standard toolbox” of techniques for analysis of time series from linear systems is based in spectral analysis and the auto- and cross-correlation functions. The principles are laid out in Box and Jenkins (1976) and Jenkins and Watts (1968). Somewhat less comprehensive introductions that provide an introduction to probability theory are Bendat and Piersol (1971) and Peebles (1987).

Time series analysis is particularly important in the closely-related engineering fields of signal processing and control. There are a large number of textbooks in this area, including Oppenheim and Schaffer (1989) and Rabiner and Gold (1975) for signal processing, and Kailath (1980) for linear systems control theory. An important subject is estimation—how one deduces the values of unmeasured variables from measured ones—and also has a large literature. An overview is provided by Gelb et al. (1974). There is also a large engineering literature dealing with nonlinear control systems; Isidori (1989) provides an introduction.

The subject of time series analysis of chaotic systems is quite new, and there are no standard texts on the subject. Instead, one must resort to the technical literature, which can be quite intimidating. A good place to start is with review articles; two excellent ones are Grassberger et al. (1991) and Abarbanel et al. (1993). An intermediate-level introduction to chaos and randomness is given in Eubank and Farmer (1990).

The review articles mentioned above contain many references to the research literature. Here, we mention some articles that are particularly germane to the presentation of this chapter. The basis for almost all nonlinear dynamics time series analysis methods is time-lag-embedding of data. The first application

of this technique to chaotic time series was by Glass and Mackey (1979) in the context of delay-differential equations, and the technique was introduced more generally in Packard et al. (1980) based on a suggestion by Ruelle. An important theorem was proved by Takens (1981) and extended and elaborated upon by Sauer et al. (1991). The influence of noise is considered in Casdagli et al. (1991).


Recurrence plots were introduced by Eckmann et al. (1987). Somewhat earlier, Grassberger and Procaccia (1983) had shown that the correlation dimension was a practical means of characterizing chaotic attractors. This technique has been widely used in applications, despite difficulties in interpreting results in data that may not be chaotic. Nonlinear prediction techniques were introduced by Farmer and Sidorowich (1987) in part to overcome this difficulty in interpretation. A paper written for non-specialists is Sugihara and May (1990). The presentation given in Chapter 6 is strongly influenced by Casdagli (1989), and the ice-age example is drawn from Hunter (1992). Methods for detecting determinism without constructing prediction models are described in Kaplan and Glass (1992, 1993) and Kennel et al. (1992). Of course, prediction techniques may be of ultimate use in forecasting the future. Many of the scientists involved in developing these methods have left academia for Wall Street. A popular review of the possible connections between chaos and finance is given in *The Economist* (Oct. 9, 1993).

The use of surrogate data is essential for deciding whether an irregular time series arises from nonlinear deterministic chaos or linear stochastic dynamics. The method was introduced by Theiler et al. (1992). An early application of phase-randomization to biological data is found in Kaplan and Cohen (1990).

The nonlinear techniques described in Chapter 6 are still part of the ongoing research enterprise. We do not know which methods will grow in use and which will wither as useless historical diversions. We also do not know what new techniques will emerge as important to using nonlinear dynamics to understand time series, but given that the field had its inception as recently as the late 1970s, it is likely that changes will be dramatic.

---

 EXERCISES

 **6.1** Time Series A (see Table 6.3) was produced by a computer random number generator. The mean of the entire time series is  $M_{\text{est}} = 0.178$ , and the standard deviation is  $\sigma^2 = 1.045$ . Is  $M_{\text{est}}$  significantly different than zero?

Calculate the mean of the first ten points of Time Series A, and the last ten points. Are the two means statistically different from zero? (If we want to know