CHAPTER 15

MULTISCALE ANALYSIS BY THE SCALE-DEPENDENT LYAPUNOV EXPONENT (SDLE)

In Chapter 1, we emphasized the importance of developing scale-dependent measures to simultaneously characterize behaviors of complex multiscaled signals on a wide range of scales. This has helped us uncover nonlinear structures in sea clutter, as demonstrated in Chapter 13. We now develop an effective algorithm to compute an excellent measure, the scale-dependent Lyapunov exponent (SDLE), and show that the SDLE can readily classify known types of complex motions, neatly solve the classic problem of distinguishing chaos from noise, effectively deal with non-stationarity, and aptly characterize complex real-world signals, including financial time series, EEG, heart rate variability (HRV) data, and sea clutter data.

15.1 BASIC THEORY

The SDLE is a variant of the finite-size Lyapunov exponent (FSLE) (see Sec. C.1 of Appendix C). The latter is closely related to another scale-dependent measure, the ϵ-entropy, which has been discussed in Sec. 13.2.3. As described in Sec. C.1, the algorithm for calculating the FSLE is very similar to Wolf et al.'s algorithm. It computes the average r-fold time by monitoring the divergence between a reference trajectory and a perturbed trajectory. To do so, it needs to define the nearest neighbors, as well as to perform a renormalization whenever the distance between the

reference and the perturbed trajectory becomes too large. Such a procedure requires very long time series and therefore is not practical. To facilitate the derivation of a fast algorithm that works on short data, as well as to ease discussion of continuous but nondifferentiable stochastic processes, we define the SDLE as follows.

Consider an ensemble of trajectories. Denote the initial separation between two nearby trajectories by ϵ_0 and their *average separation* at times t and $t + \Delta t$ by ϵ_t and $\epsilon_{t+\Delta t}$, respectively. Being defined in an average sense, ϵ_t and $\epsilon_{t+\Delta t}$ can be readily computed from any processes, even if they are nondifferentiable. Next, we examine the relation between ϵ_t and $\epsilon_{t+\Delta t}$ where Δt is small. When $\Delta t \to 0$, we have

$$\epsilon_{t+\Delta t} = \epsilon_t e^{\lambda(\epsilon_t)\Delta t} \tag{15.1}$$

or

$$\lambda(\epsilon_t) = \frac{\ln \epsilon_{t+\Delta t} - \ln \epsilon_t}{\Delta t}, \tag{15.2}$$

where $\lambda(\epsilon_t)$ is the SDLE. Equivalently, we have a differential equation for ϵ_t,

$$\frac{d\epsilon_t}{dt} = \lambda(\epsilon_t)\epsilon_t. \tag{15.3}$$

Given time series data, the smallest Δt possible is the sampling time τ.

The definition of the SDLE suggests a simple ensemble average-based scheme to compute it. A straightforward way would be to find all the pairs of vectors in the phase space whose distance is approximately ϵ and then calculate their average distance after a time Δt. The first half of this description amounts to introducing a shell (indexed as k),

$$\epsilon_k \leq \|V_i - V_j\| \leq \epsilon_k + \Delta \epsilon_k,$$

where V_i, V_j are reconstructed vectors and ϵ_k (the radius of the shell) and $\Delta \epsilon_k$ (the width of the shell) are arbitrarily chosen small distances. Such a shell may be considered as a differential element that would facilitate computation of the conditional probability. To expedite the computation, it is advantageous to introduce a sequence of shells, $k = 1, 2, 3, \cdots$. We thus arrive at the same computational procedure for computing the time-dependent exponent (TDE) curves discussed in Chapter 13. Specifically, with all these shells, we can then monitor the evolution of all of the pairs of vectors (V_i, V_j) within a shell and take the average. When each shell is very thin, by assuming that the order of averaging and taking the logarithm in Eq. (15.2) can be interchanged, we have

$$\lambda(\epsilon_t) = \frac{\left\langle \ln \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| - \ln \|V_{i+t} - V_{j+t}\| \right\rangle}{\Delta t}, \tag{15.4}$$

where t and Δt are integers in a unit of the sampling time, and the angle brackets denote the average within a shell. Note that contributions to the SDLE at a specific scale from different shells can be combined, with the weight for each shell being determined by the number of pairs of vectors (V_i, V_j) in that shell. In the following,

to see more clearly how each shell characterizes the dynamics of the data on different scales, we shall plot the $\lambda(\epsilon)$ curves for different shells separately.

In the above formulation, it is assumed that the initial separation, $\| V_i - V_j \|$, aligns with the most unstable direction instantly. For high-dimensional systems this is not true, especially when the growth rate is nonuniform and/or the eigenvectors of the Jacobian are nonnormal. Fortunately, the problem is not as serious as one might think, since our shells are not infinitesimal. When computing the TDE, we have found that when difficulties arise, it is often sufficient to introduce an additional condition,

$$|j - i| \geq (m - 1)L, \tag{15.5}$$

when finding pairs of vectors within each shell. This amounts to removing the time scale k_a shown in Fig. 13.9(a). Such a scheme also works well when computing the SDLE. This means that after taking a time comparable to the embedding window $(m - 1)L$, it would be safe to assume that the initial separation has evolved to the most unstable direction of the motion. In turn, the time index t in Eq. (15.4) also has to satisfy a similar inequality, $t \geq (m - 1)L$.

What is the relation between the SDLE and other complexity measures such as the largest positive Lyapunov exponent λ_1? To find the answer, we recall that Wolf et al.'s algorithm for computing λ_1 involves monitoring the exponential divergence between a reference and a perturbed trajectory. It involves a scale parameter ϵ^* such that whenever the divergence exceeds this chosen scale, a renormalization procedure is performed. Therefore, λ_1 estimated by Wolf et al.'s algorithm can be computed from the SDLE by the following relation

$$\lambda_1 = \int_0^{\epsilon^*} \lambda(\epsilon)p(\epsilon)d\epsilon, \tag{15.6}$$

where $p(\epsilon)$ is the probability density function for the scale ϵ. Since our algorithm involves shells described by condition (15.4), $p(\epsilon)$ is proportional to the derivative of the Grassberger-Procaccia's correlation integral,

$$p(\epsilon) = Z\frac{dC(\epsilon)}{d\epsilon}, \tag{15.7}$$

where Z is a normalization constant satisfying

$$\int_0^{\epsilon^*} p(\epsilon)d\epsilon = 1 \tag{15.8}$$

and $C(\epsilon)$ is the correlation integral given by Eq. (13.14).

Before proceeding, we wish to emphasize the major difference between our algorithm and the standard method for calculating the FSLE. As we have pointed out, to compute the FSLE, two trajectories, one as reference, another as perturbed, have to be defined. This requires huge amounts of data. Our algorithm avoids this by employing two critical operations to fully utilize information about the time

evolution of the data: (1) The reference and perturbed trajectories are replaced by time evolution of all pairs of vectors satisfying Inequality (15.5) and falling within a shell, and (2) introduction of a sequence of shells ensures that the number of pairs of vectors within the shells is large and the ensemble average within each shell is well defined. Let the number of points needed to compute the FSLE by standard methods be N. These two operations imply that the method described here requires only about $\sqrt{N}$ points to compute the SDLE. In the following, we shall illustrate the effectiveness of our algorithm by examining various types of complex motions.

15.2 CLASSIFICATION OF COMPLEX MOTIONS

To understand the SDLE as well as appreciate its power, we apply it to classify seven major types of complex motions.

15.2.1 Chaos, noisy chaos, and noise-induced chaos

Obviously, for truly low-dimensional chaos, $\lambda(\epsilon)$ equals the largest positive Lyapunov exponent and, hence, must be independent of ϵ over a wide range of scales. For noisy chaos, we expect $\lambda(\epsilon)$ to depend on small ϵ. To illustrate both features, we consider the chaotic Lorenz system with stochastic forcing described by Eq. (13.16). Figure 15.1(a) shows five curves for the cases of $D = 0, 1, 2, 3, 4$. The computations are done with 10,000 points and $m = 4, L = 2$. We observe a few interesting features:

- For the clean, chaotic signal, $\lambda(\epsilon)$ fluctuates slightly around a constant (which numerically equals the largest positive Lyapunov exponent) when ϵ is smaller than a threshold value that is determined by the size of the chaotic attractor. The reason for the small fluctuations in $\lambda(\epsilon)$ is that the divergence rate varies from one region of the attractor to another.

- When there is stochastic forcing, $\lambda(\epsilon)$ is no longer a constant when ϵ is small, but increases as $-\gamma \ln \epsilon$ when the scale ϵ is decreased. The coefficient γ does not seem to depend on the strength of the noise. This feature suggests that entropy generation is infinite when the scale ϵ approaches zero. Note that the relation of $\lambda(\epsilon) \sim -\gamma \ln \epsilon$ has also been observed for the FSLE and the ϵ-entropy. In fact, such a relation can be readily proven for the ϵ-entropy.

- When the noise is increased, the part of the curve with $\lambda(\epsilon) \sim -\gamma \ln \epsilon$ shifts to the right. In fact, little chaotic signature can be identified when D is increased beyond 3. When noise is not too strong, this feature can be readily used to quantify the strength of noise.

Next, we consider noise-induced chaos. To illustrate the idea, we study the noisy logistic map

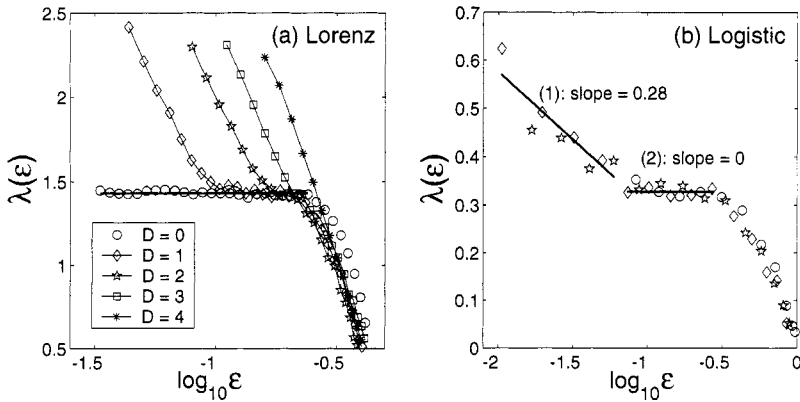$$x_{n+1} = \mu x_n(1 - x_n) + P_n, \quad 0 < x_n < 1, \tag{15.9}$$

Figure 15.1. $\lambda(\epsilon)$ curves for (a) the clean and noisy Lorenz systems and (b) the noise-induced chaos in the logistic map. Curves from different shells are designated by different symbols.

where μ is the bifurcation parameter and P_n is a Gaussian random variable with zero mean and standard deviation σ. It has been found that at $\mu = 3.74$ and $\sigma = 0.002$, noise-induced chaos occurs, and it may be difficult to distinguish noise-induced chaos from clean chaos. In Fig. 15.1(b), we have plotted the $\lambda(\epsilon_t)$ for this particular noise-induced chaos. The computation was done with $m = 4, L = 1$ and 10,000 points. We observe that Fig. 15.1(b) is very similar to the curves of noisy chaos plotted in Fig. 15.1(a). Hence, noise-induced chaos is similar to noisy chaos but different from clean chaos.

At this point, two comments are noteworthy: (1) On very small scales, the effect of measurement noise is similar to that of dynamic noise. (2) The $\lambda(\epsilon)$ curves shown in Fig. 15.1 are based on a fairly small shell. The curves computed based on larger shells collapse on the right part of the curves shown in Fig. 15.1. For this reason, for chaotic systems, one or a few small shells are sufficient. If one wishes to know the behavior of λ on ever smaller scales, one has to use longer and longer time series.

Finally, we consider the Mackey-Glass delay differential system,

$$dx/dt = \frac{ax(t + \Gamma)}{1 + x(t + \Gamma)^c} - bx(t).$$

When $a = 0.2, b = 0.1, c = 10, \Gamma = 30$, it has two positive Lyapunov exponents, with the largest Lyapunov exponent close to 0.007. Having two positive Lyapunov exponents while the value of the largest Lyapunov exponent of the system is not much greater than 0, one might be concerned that it may be difficult to compute the SDLE of the system. This is not the case. In fact, this system can be analyzed as straightforwardly as other dynamical systems, including the Henon map and the Rossler system. An example of the $\lambda(\epsilon)$ curve is shown in Fig. 15.2, where we have used $m = 5, L = 1$, and 5000 points sampled with a time interval of 6. We observe a well-defined plateau, with its value close to 0.007. This example illustrates that
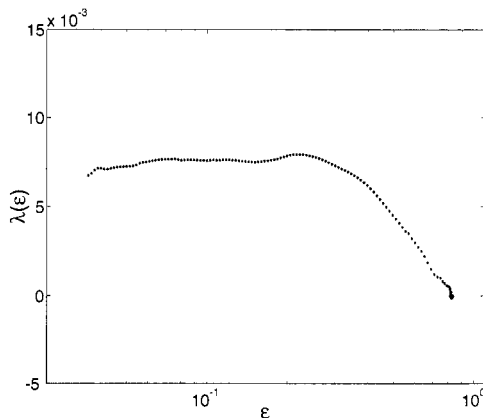
Figure 15.2.   The $\lambda(\epsilon)$ curve for the Mackey-Glass system. The computation was done with $m = 5, L = 1$, and 5000 points were sampled with a time interval of 6.

when computing the SDLE, one does not need to be concerned about a nonuniform growth rate in high-dimensional systems.

15.2.2 $1/f^\beta$ processes

As we have discussed in depth in Chapters 6 and 8, a $1/f^\beta$ process is one of the most important classes of random fractals. Two important prototypical models for such processes are the fractional Brownian motion (fBm) process and the ON/OFF intermittency with power-law distributed ON and OFF periods. Depending on whether $H = (\beta - 1)/2$ is smaller than, equal to, or larger than $1/2$, the process is said to have antipersistent correlation, short-range correlation, or persistent long-range correlation.

It is well known that the variance of such stochastic processes increases with t as t^{2H}. Translating into the average distance between nearby trajectories, we can write

$$\epsilon_t = \epsilon_0 t^H. \tag{15.10}$$

Using Eq. (15.2), we readily find $\lambda(\epsilon_t) \sim H/t$. Expressing t by ϵ_t, we finally find

$$\lambda(\epsilon_t) \sim H\epsilon_t^{-1/H}. \tag{15.11}$$

Note that the same functional relation can be derived for ϵ-entropy. However, using ϵ-entropy, it is not easy to estimate H through analysis of short time series. In contrast, Eq. (15.11) can be conveniently used to estimate H. To illustrate this, we first study fBm processes $B_H(t)$. Figure 15.3(a) shows three curves for $H = 0.33, 0.5$ and 0.7, where the calculation is done with 2^{14} points and $m = 2, L = 1$. We observe that the estimated $1/H$ clearly match those used in simulating these processes.
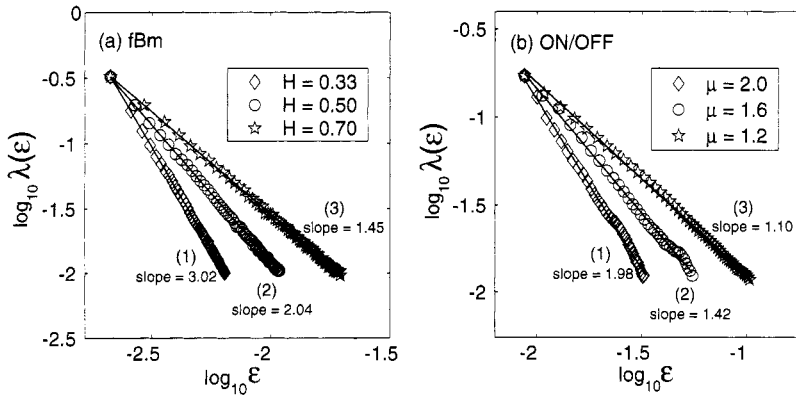
Figure 15.3. $\lambda(\epsilon)$ curves for (a) fBm processes with $H = 0.33, 0.50$, and 0.70, and (b) ON/OFF models with $\mu = 1.2, 1.6$, and 2.0.

Next, we study ON/OFF models with power-law distributed ON and OFF periods,

$$P[X \geq x] = \left(\frac{b}{x}\right)^{\mu}, \quad x \geq b > 0, \ 0 < \mu \leq 2.$$

Figure 15.3(b) shows three curves for $\mu = 1.2, 1.6, 2.0$, which correspond to $H = 0.9, 0.7$, and 0.5, respectively. Recalling that $H = (3 - \mu)/2$, again we find that the estimated H accurately reflect the values of μ used in the simulations.

What is the meaning of small ϵ in the power-law distributed ON/OFF model? As we have pointed out, when $\mu < 2$, ON/OFF processes have infinite variance. To effectively characterize ON/OFF processes, the absolute value of ϵ also has to be large. Therefore, ϵ being small is only relative to the largest scale resolved by the dataset. This note is also pertinent to the Levy flights to be studied next. In fact, the example of a Levy flight to be discussed has not only infinite variance but also infinite mean. Interestingly, the defining parameter for such a Levy flight can still be readily estimated by our method.

Finally, we discuss in which scale range Eq. (15.11) may be valid. Let Eq. (15.10) be valid between two time scales, t_1 and t_2, where $t_1 < t_2$. Using Eq. (15.10), we have $\epsilon_2/\epsilon_1 = (t_2/t_1)^H$. This is the scale range for Eq. (15.11) to be valid. Let t_2/t_1 be fixed, it becomes smaller when H is decreased.

15.2.3 Levy flights

We now consider Levy processes, another important type of random fractal models that have found numerous applications.

There are two types of Levy processes. One is Levy flights, which are random processes consisting of many independent steps, each characterized by a stable law and consuming a unit time regardless of its length. The other is Levy walks, where each step takes time proportional to its length. A Levy walk can be viewed
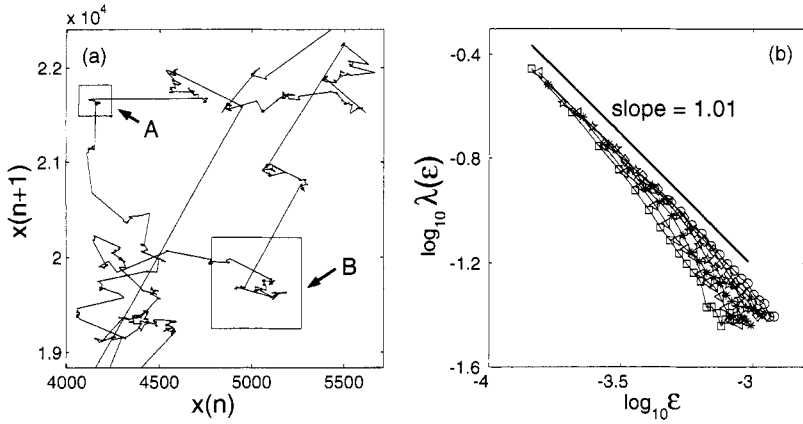
Figure 15.4. (a) Two-dimensional Levy flights with $\alpha = 1.0$, (b) $\lambda(\epsilon)$ curves for the Levy process (curves from different shells are designated by different symbols).

as sampled from a Levy flight with a uniform speed. The increment process of a Levy walk, obtained by differencing the Levy walk, is very similar to an ON/OFF train with power-law distributed ON and OFF periods. Therefore, in the following, we shall not be concerned about it. We shall focus on Levy flights. As we have seen in Chapter 8, Levy flights, having independent steps, are memoryless processes characterized by $H = 1/2$, irrespective of the value of the exponent α characterizing the stable laws. In other words, methods such as DFA cannot be used to estimate the α parameter.

In order to derive an analytic expression for $\lambda(\epsilon)$ to estimate α, it is sufficient to note Eq. (14.12) derived in Chapter 14:

$$\Delta L(\Delta t) \propto \Delta t^{1/\alpha}.$$

Comparing this to $1/f^{2H+1}$ processes, we find that $1/\alpha$ plays the role of H. Therefore, the scaling for the SDLE is

$$\lambda(\epsilon_t) \sim \frac{1}{\alpha}\epsilon_t^{-\alpha}. \tag{15.12}$$

We have simulated a number of Levy flights with different α. One realization for the flights with $\alpha = 1$ is shown in Fig. 15.4(a). The computed $\lambda(\epsilon_t)$ curves (based on 2^{15} points and $m = 2, L = 1$) are shown in Fig. 15.4(b). In fact, the $\lambda(\epsilon_t)$ curves from a number of different shells are plotted together. We observe that the slope of the *envelope* is -1, consistent with the value of α chosen in simulating the flights.

To understand why the envelope of $\lambda(\epsilon_t)$ gives a better estimate of α, we resort to the two small boxes, denoted A and B, in Fig. 15.4(a). When those boxes are enlarged, they show patterns similar to those in Fig. 15.4(a). Obviously, the scales involved in A and B are different. Those different scales are captured by shells of different sizes. Since the mean and variance of the flights are both infinite, while

the number of points in each small box is not large, $\lambda(\epsilon_t)$ from each shell cannot have a long scaling region. However, when $\lambda(\epsilon_t)$ from different shells are plotted together, they form an envelope with a long scaling region. Using this argument, it is clear that when computing $\lambda(\epsilon_t)$ for a Levy flight, a number of shells are more advantageous than a single shell, especially when the time series is not too long.

15.2.4   SDLE for processes defined by PSIC

In Chapter 14, we saw that PSIC provides a common foundation for chaos and random fractals. Now that we have characterized chaos, $1/f^\beta$ noise, and Levy processes by the SDLE, we derive a simple equation relating the λ_q and q of PSIC to the SDLE.

First, we recall that PSIC is defined by

$$\xi_t = \lim_{||\Delta V(0)|| \to 0} \frac{||\Delta V(t)||}{||\Delta V(0)||} = \left[1 + (1 - q)\lambda_q^{(1)} t \right]^{1/(1-q)}$$

Since $\epsilon_t = ||\Delta V(0)||\xi_t$, it is now more convenient to express the SDLE as a function of ξ_t. Using Eq. (15.2), we find that

$$\lambda(\xi_t) = \frac{\ln \xi_{t+\Delta t} - \ln \xi_t}{\Delta t}. \tag{15.13}$$

When $\Delta t \to 0$, $1 + (1 - q)\lambda_q^{(1)} t \gg (1 - q)\lambda_q^{(1)} \Delta t$. Simplifying Eq. (15.13), we obtain

$$\lambda(\xi_t) = \lambda_q^{(1)} \xi_t^{q-1}. \tag{15.14}$$

We now consider three cases:

- For chaotic motions, $q = 1$; therefore, $\lambda(\xi_t) = \lambda_q = $ constant.
- For $1/f^\beta$ noise, using Eqs. (14.10) and (14.11), we have $q = 1 - \frac{1}{H}$ and $\lambda_q^{(1)} = H$. Therefore,

$$\lambda(\xi_t) = H\xi_t^{-\frac{1}{H}}.$$

  This is equivalent to Eq. (15.11).

- For Levy flights, using Eqs. (14.13) and (14.14), we have $q = 1 - \alpha$ and $\lambda_q^{(1)} = 1/\alpha$. Therefore,

$$\lambda(\xi_t) = \frac{1}{\alpha}\xi_t^{-\alpha}.$$

  This is equivalent to Eq. (15.12).

15.2.5   Stochastic oscillations

Stochastic oscillation is an important type of complex motion that has been observed in many different disciplines of science and engineering. A stochastic oscillator,

having structures due to oscillatory motions but not having closed orbits in the phase space, may be interpreted as chaos. Which functional form of $\lambda(\epsilon_t)$ characterizes such motions? To gain insights into this problem, we study a stochastically driven Van der Pol's oscillator:

$$
\begin{aligned}
dx/dt &= y + D_1\eta_1(t), \\
dy/dt &= -(x^2 - 1)y - x + D_2\eta_2(t),
\end{aligned}
\tag{15.15}
$$

where $< \eta_i(t) >= 0$, $< \eta_i(t)\eta_j(t') >= \delta_{ij}\delta(t - t')$, $i, j = 1, 2$, and the parameters D_i, $i = 1, 2$ characterize the strength of noise. Figure 15.5(a) shows short segment of the $y(t)$ time series for $D_1 = D_2 = 0.02$. Since the noise is small, the reconstructed phase diagram is only slightly diffused, as shown in Fig. 15.5(b). Interestingly, we have found two generic functional forms of $\lambda(\epsilon_t)$ for the $y(t)$ (and $x(t)$) time series of the stochastic Van der Pol's oscillator as well as other stochastic oscillators, independent of the noise level (so long as the noise is not extremely small). One functional form for $\lambda(\epsilon_t)$ is $\lambda(\epsilon) \sim -\ln \epsilon$, observed when the embedding dimension m and delay time L are both small. This is shown in Fig. 15.5(c). Another type of behavior is $\lambda(\epsilon) \sim \epsilon^{-1/H}$, where $H \approx 1/2$, observed when $(m - 1)L$ is comparable to half of the period of the oscillation. An example of the latter is shown in Fig. 15.5(d).

The relation $\lambda(\epsilon) \sim -\ln \epsilon$ for a small embedding window suggests that locally, the dynamics of a stochastic oscillator are just like other noisy dynamics, such as those shown in Fig. 15.1. The relation $\lambda(\epsilon) \sim \epsilon^{-2}$ for a large embedding window suggests that the motion is like a Brownian motion (Bm). This reflects the variation in amplitude of the oscillation on longer time scales. Such variations have also been observed in experimental data, such as pathologic tremor data and velocity fluctuations in the near wake of a circular cylinder. In fact, pathologic tremor data are often characterized by more general relations $\lambda(\epsilon) \sim \epsilon^{-1/H}$, with $H \neq 1/2$.

Up to now, we have focused on the positive portion of $\lambda(\epsilon_t)$. It turns out that when t is large, $\lambda(\epsilon)$ becomes oscillatory, with a mean about 0. Such oscillatory behavior can be in fact identified from Fig. 15.5(c). Denote the corresponding scales by ϵ_∞ and call them the limiting or characteristic scales. They are the stationary portion of ϵ_t, and hence, they may still be a function of time. When we take the Fourier transform of the limiting scales, we find that its power spectral density (PSD), shown in Fig. 15.5(f), is very similar to that of the original $y(t)$ time series shown in Fig. 15.5(e). This simple discussion suggests that the limiting scales capture the structured component of the data. This feature will be further illustrated in Secs. 15.4 and 15.7.2 when we discuss identification of hidden frequencies and analysis of HRV data. Therefore,the positive portion of $\lambda(\epsilon_t)$ and the concept of limiting scale provide a comprehensive characterization of the signals.

15.2.6   Complex motions with multiple scaling behaviors

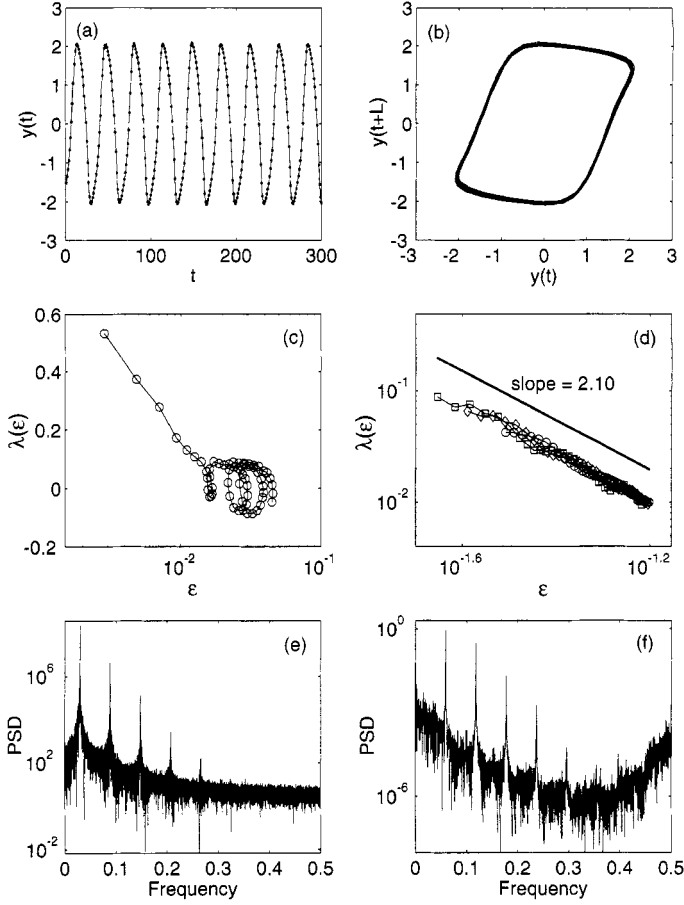Some dynamical systems may exhibit multiple scaling behaviors, such as chaotic

Figure 15.5.   (a) $y(t)$ time series of the Van der Pol's oscillator with $D_1 = D_2 = 0.02$. (b) the reconstructed 2-D phase diagram. (c,d) $\lambda(\epsilon)$ curves for the $y(t)$ time series, where embedding parameters are $m = 4$, $L = 1$ for (c) and $m = 8$, $L = 2$ for (d). In (d), three curves from three different shells are designated by different symbols. (e,f) are, respectively, the PSD for $y(t)$ data and the stationary portion of ϵ_t (which corresponds to $\lambda_\epsilon \approx 0$). The data were sampled with a time interval of 0.2. The period of the oscillation is about 30 sample points. The embedding parameters of $m = 8$, $L = 2$ yield an embedding window $(m - 1)L$ of about half of the period.

behavior on small scales but diffusive behavior on large scales. To see how the SDLE can characterize such systems, we study the following map:

$$x_{n+1} = [x_n] + F(x_n - [x_n]) + \sigma\eta_t, \qquad (15.16)$$

where $[x_n]$ denotes the integer part of x_n, η_t is a noise uniformly distributed in the interval $[-1, 1]$, σ is a parameter quantifying the strength of noise, and $F(y)$ is
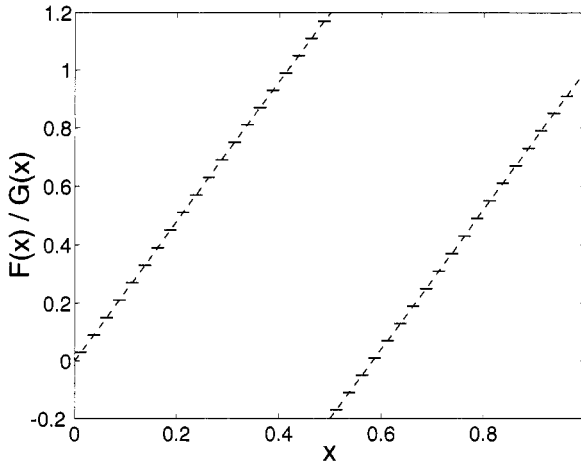
Figure 15.6. The function $F(x)$ (Eq. (15.17)) for $\Delta = 0.4$ is shown as the dashed lines. The function $G(x)$ (Eq. (15.18)) is an approximation of $F(x)$, obtained using 40 intervals of slope 0. In the case of noise-induced chaos discussed in [64], $G(x)$ is obtained from $F(x)$ using 10^4 intervals of slope 0.9.

given by

$$F(y) = \begin{cases} (2 + \Delta)y & \text{if } y \in [0, 1/2) \\ (2 + \Delta)y - (1 + \Delta) & \text{if } y \in (1/2, 1] \end{cases}. \tag{15.17}$$

The map $F(y)$ is shown in Fig. 15.6 as the dashed lines. It shows chaotic dynamics with a positive Lyapunov exponent $\ln(2 + \Delta)$. On the other hand, the term $[x_n]$ introduces a random walk on integer grids.

It turns out that this system is very easy to analyze. When $\Delta = 0.4$, with only 5000 points and $m = 2, L = 1$, we can resolve both the chaotic behavior on very small scales, and the normal diffusive behavior (with slope -2) on large scales. See Fig. 15.7(a).

We now ask a question: Given a small dataset, which type of behavior, chaotic or diffusive, is resolved first? To answer it, we have tried to compute the SDLE with only 500 points. The result is shown in Fig. 15.7(b). It is interesting to observe that chaotic behavior can be well resolved by only a few hundred points. However, diffusive behavior needs much more data to resolve. Intuitively, this makes sense, since diffusive behavior amounts to a Bm on the integer grids and is of much higher dimension than the small-scale chaotic behavior. Therefore, more data are needed to resolve it. As will be shown in the next section, this important result will allow us to determine whether a deterministic Bm is of low or high dimension.

We have also studied the noisy map. The resulting SDLE for $\sigma = 0.001$ is shown in Fig. 15.7(a) as squares. We have again used 5000 points. While the behavior of the SDLE suggests noisy dynamics, with 5000 points we are not able to resolve
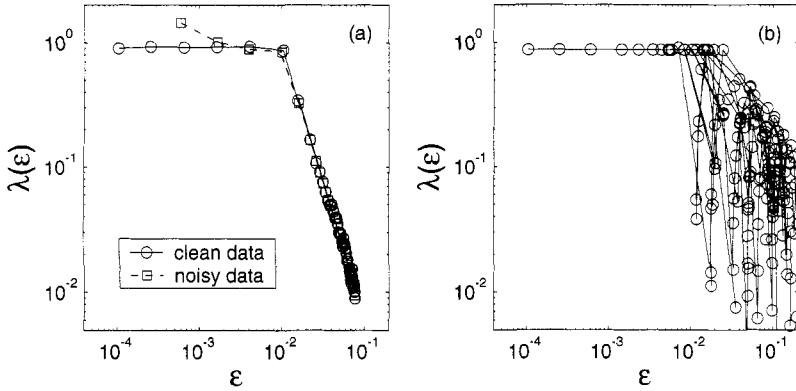
Figure 15.7. $\lambda(\epsilon)$ for the model described by Eq. (15.16): (a) 5000 points were used; for the noisy case, $\sigma = 0.001$; (b) 500 points were used.

well the relation $\lambda(\epsilon) \sim -\ln \epsilon$. This indicates that for the noisy map, on very small scales, the dimension is very high.

Map (15.16) can be modified to produce an interesting system with noise-induced chaos. This can be done by replacing the function $F(y)$ in map (15.16) by $G(y)$ to obtain the following map:

$$x_{t+1} = [x_t] + G(x_t - [x_t]) + \sigma \eta_t, \qquad (15.18)$$

where η_t is a noise uniformly distributed in the interval $[-1, 1]$, σ is a parameter quantifying the strength of noise, and $G(y)$ is a piecewise linear function that approximates $F(y)$ in Eq. (15.17). An example of $G(y)$ is shown in Fig. 15.6. In our numerical simulations, we have followed Cencini et al. and used 10^4 intervals of slope 0.9 to obtain $G(y)$. With such a choice of $G(y)$, in the absence of noise the time evolution described by map (15.18) is nonchaotic, since the largest Lyapunov exponent $\ln(0.9)$ is negative. With an appropriate noise level (e.g., $\sigma = 10^{-4}$ or 10^{-3}), the SDLE for the system becomes indistinguishable from the noisy SDLE shown in Fig. 15.7 for map (15.16). Having a diffusive regime on large scales, this is a more complicated noise-induced chaos than the one we have found from the logistic map.

15.3   DISTINGUISHING CHAOS FROM NOISE

15.3.1   General considerations

Distinction between chaos and noise is a classic problem arising from the life sciences, finance, ecology, physics, fluid mechanics, and geophysics. Although tremendous efforts have been made to solve this problem, it remains wide open. Two major difficulties in solving this problem are that (1) chaos can be induced by

noise and (2) standard Bm's may have a deterministic origin. To overcome these two difficulties, we ask two questions: (1) What are the fundamental differences among clean low-dimensional chaos, noisy chaos, and noise-induced chaos? (2) When a Bm has a deterministic origin, can we determine whether it is from a low-dimensional deterministic system or a high-dimensional deterministic system? The above questions motivate us to propose the following algorithmic scheme to solve the problem of distinguishing chaos from noise. Denote two scales resolvable by the resolution of the data by ϵ_1 and ϵ_2, where $\epsilon_1 < \epsilon_2$. On this scale range, if the behavior of the data is the same as that of chaotic data, then we say that the data are chaotic. If the behavior is like that of a Bm, then we say that the data are a Bm. Of course, there may exist scale ranges disjoint with (ϵ_1, ϵ_2), where the data behave neither like a chaotic motion nor like a Bm, but have features that can define other types of motion. The data on those scales will be classified accordingly. The feasibility of such a scheme depends critically on whether, by analyzing short noisy time series, one can (1) classify different types of motions and (2) characterize the time series by automatically identifying different scale ranges where different types of motion are manifested. Note that when dealing with complex multiscaled data, the choice of a scale-dependent classification scheme is most natural. However, when a Bm is generated by a very high-dimensional deterministic system, we shall simply treat it as a type of stochastic process, so long as it has the defining properties for a Bm.

At this point, we must comment on the concept of a resolvable scale. Denote a time series under investigation by $x(1), x(2), \cdots, x(n)$. Using time delay embedding, one obtains vectors of the form $V_i = [x(i), x(i + L), \cdots, x(i + (m - 1)L)]$, where the embedding dimension m and the delay time L are chosen according to certain optimization criteria. In the reconstructed phase space, the dataset determines two scales: the maximum and the minimum of the distances between two vectors, $\|V_i - V_j\|$, where $i \neq j$. Denote them by $\epsilon_{\max}$ and $\epsilon_{\min}$, respectively. This is the resolvable scale range. Analysis of data has to be confined within these two scales. Of course, with more data available, the resolvable scale range can be enlarged. Note that one can treat $\epsilon_{\max}$ as one unit. Alternatively, one may normalize the time series $x(1), x(2), \cdots, x(n)$ into the unit interval $[0, 1]$ before further analysis.

15.3.2   A practical solution

Now that we understand that the key to distinguish chaos from noise is to identify different scale ranges where different types of motion are manifested, we are ready to determine how chaos can be distinguished from noise. In order for our discussion to be useful for practical applications, we assume that our dataset is not only finite, but of small or medium size. Note that if the dataset can be extremely large, then on scales smaller than those where the behavior of $\lambda(\epsilon) \sim -\gamma \ln \epsilon$ is observed, one can observe another plateau, indicating the existence of a high-entropic chaotic state.

Nonlinear maps used as pseudorandom number generators belong in this category. But we shall not be concerned about this here.

We first ask a question. When all three distinctive behaviors of the SDLE,

(1) $\lambda(\epsilon) \sim -\ln \epsilon$,

(2) $\lambda(\epsilon) \sim$ constant,

(3) $\lambda(\epsilon) \sim \epsilon^{-1/H}$,

coexist, is it typically true that the scales where they occur are in ascending order? The answer is yes, as can be readily understood by the following argument. Behavior (1) amounts to stochastic forcing. Its effect is to kick the dynamics to larger scales. For the chaotic motion to be resolvable, its effect has to be limited to scales that are smaller than the scales showing chaotic motion. To understand why diffusive behavior (3) has to occur on scales larger than those showing chaotic motions, it suffices to note that diffusive motion is a nonstationary process. It needs a huge or even an unbounded region of phase space to play out. It is unimaginable to have chaotic motion beyond those scales still resolvable by finite data. Therefore, the scales for the three behaviors to occur have to be in increasing order.

The above discussion suggests that if a Bm is generated by a deterministic system, then the dimension of the system can be readily determined. This is because, for deterministic low-dimensional systems such as map (15.16) to generate large-scale diffusive motions, its local dynamics have to be unstable. Low-dimensional local unstable dynamics amount to low-dimensional chaotic motion. Therefore, we must have a constant $\lambda(\epsilon)$ on small scales. In other words, coexistence of $\lambda(\epsilon) \sim$ constant and $\lambda(\epsilon) \sim \epsilon^{-2}$ would indicate that the Bm is from a low-dimensional deterministic system. We emphasize that $\lambda(\epsilon) \sim$ constant should occur on the scales smaller than those for diffusion to occur and that the behavior of $\lambda(\epsilon) \sim -\ln \epsilon$ should not occur, since the system is low-dimensionally deterministic.

We are now ready to determine to what extent chaos can be distinguished from noise. First, in order to say that a time series under study has a chaotic scaling regime, $\lambda(\epsilon)$ has to show a plateau (i.e., almost constant) for a scale range (ϵ_1, ϵ_2). Since many arbitrary local functions could be treated as a constant, if $\lambda(\epsilon)$ is constant only for ϵ_2 very close to ϵ_1, then it should not be interpreted as a plateau. In other words, ϵ_2 has to be considerably larger than ϵ_1. Unfortunately, it is not easy to state exactly how much larger ϵ_2 should be than ϵ_1, since this depends on the dimension of the data. We believe that $\log_{10}(\epsilon_2/\epsilon_1)$ has to be at least around 1/2.

Once we know about the existence of a chaotic scaling regime in the data, it is natural to ask if the dynamics are deterministically chaotic or not. This can be readily inferred by checking the behavior of $\lambda(\epsilon)$ on scales smaller than ϵ_1. If, on those scales, $\lambda(\epsilon)$ is above the plateau, then we have good reason to believe that the data are noisy. If, on those scales, $\lambda(\epsilon) \sim -\ln \epsilon$, then we can be sure that the data are noisy chaos or noise-induced chaos. However, if all we know is the time series, then we will not be able to distinguish further between noisy chaos and noise-induced chaos.

In most situations, when the data are found to be noisy, it is important to find out which type of noise process the data represent. The power-law behavior of the SDLE is especially interesting for this purpose. For example, we can combine analysis using the SDLE with spectral analysis and distributional analysis to determine whether the data are a type of $1/f^\beta$ process, or Levy processes, or stochastic oscillations. More precisely, power-law behavior of the SDLE together with sharp spectral peaks revealed by Fourier analysis would indicate a stochastic oscillation, whereas power-law behavior of the SDLE together with a Gaussian-like distribution would indicate fBm-like data. If, instead, heavy-tailed distributions are observed, then the data are an ON/OFF intermittency type or are similar to Levy processes. While many excellent methods for estimating the Hurst parameter have been proposed (see the recent comparative study by Gao et al. [157]), the SDLE offers another effective approach. In fact, as we pointed out in Sec. 15.2, the SDLE provides better characterization of Levy flights and stochastic oscillations, since most other methods only give $H = 0.5$ for Levy flights, irrespective of the values of the defining parameter α, and fail to characterize the power-law behavior of the SDLE for stochastic oscillations.

15.4 CHARACTERIZING HIDDEN FREQUENCIES

Defining and characterizing large-scale, orderly motions is a significant problem in many scientific disciplines. One of the most important types of large-scale, orderly motions is oscillatory motions. An interesting type of oscillatory motions is associated with the so-called hidden frequency phenomenon. That is, when the dynamics of a complicated system are monitored through the temporal evolution of a variable x, Fourier analysis of $x(t)$ may not suggest any oscillatory motions. However, if the dynamics of the system are monitored through the evolution of another variable, say, z, then the Fourier transform of $z(t)$ may contain a well-defined spectral peak indicating oscillatory motions. An example is the chaotic Lorenz system described in Eq. (13.16). Figures 15.8(a–c) show the PSD of the x, y, z components of the system. We observe that the PSD of $x(t)$ and $y(t)$ are simply broad. However, the PSD of $z(t)$ shows a very sharp spectral peak. Recall that geometrically the Lorenz attractor consists of two scrolls (see Fig. 15.9). The sharp spectral peak in the PSD of $z(t)$ of the Lorenz system is due to the circular motions along either of the scrolls.

The above example illustrates that if the dynamics of a system contain a hidden frequency that cannot be revealed by the Fourier transform of a measured variable (say, $x(t)$), then in order to reveal the hidden frequency, one has to embed $x(t)$ in a suitable phase space. This idea has led to the development of two interesting methods for identifying hidden frequencies. One method, proposed by Ortega [326, 327], involves computing the temporal evolution of density measures in the reconstructed phase space. Another, proposed by Chern et al. [74], involves
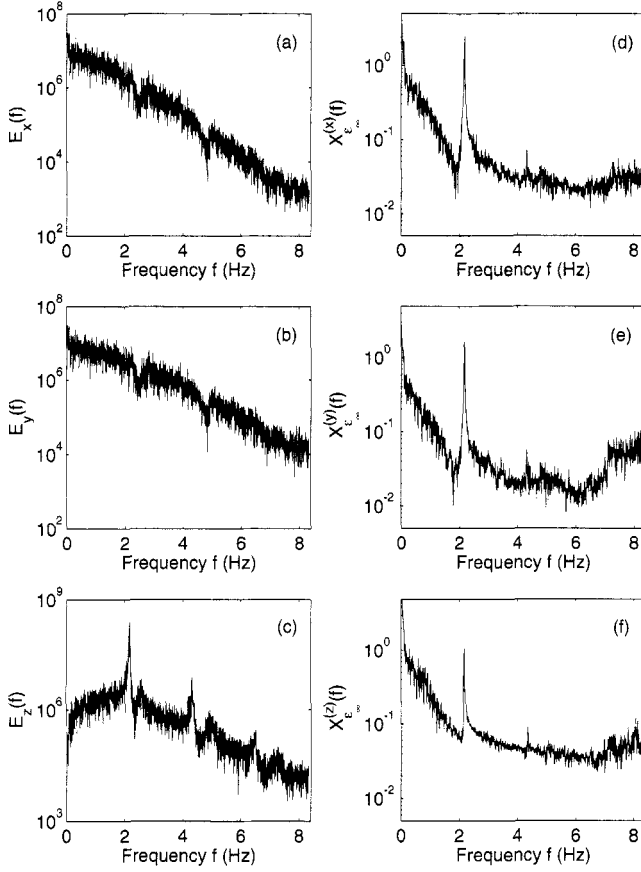
Figure 15.8.   Power-spectral density (PSD) for (a) $x(t)$, (b) $y(t)$, (c) $z(t)$, and the magnitude of the Fourier transform for (d) $\epsilon_\infty^{(x)}(t)$, (e) $\epsilon_\infty^{(y)}(t)$, and (f) $\epsilon_\infty^{(z)}(t)$. The SDLE was computed with a shell of size $(2^{-9/2}, 2^{-5})$, 5000 points sampled at $\tau = 0.06$, and embedding parameters $m = 4, L = 2$.

taking singular value decomposition (see Appendix B) of local neighbors. Ortega's method has been applied to an experimental time series recorded from a far-infrared laser in a chaotic state (see Sec. A.4 in Appendix A). The laser dataset, shown in Fig. 15.10(a), contains 10,000 points, sampled with a time interval of 80 ns. The PSD of the data is shown in Fig. 15.10(b), where one observes a sharp peak around 1.7 MHz. Figure 15.10(c) shows the PSD of the density time series, where one notes an additional spectral peak around 37 kHz. This peak is due to the envelope of chaotic pulsations, which is discernible from Fig. 15.10(a).

To appreciate the strength of the additional spectral peak in Fig. 15.10(c), we have to note that the units of the PSD in Figs. 15.10(b,c) are arbitrary; the largest peak is treated as 1 unit, as was done by Ortega. In fact, the actual energy of the original laser intensity time series is much greater than that of the density time
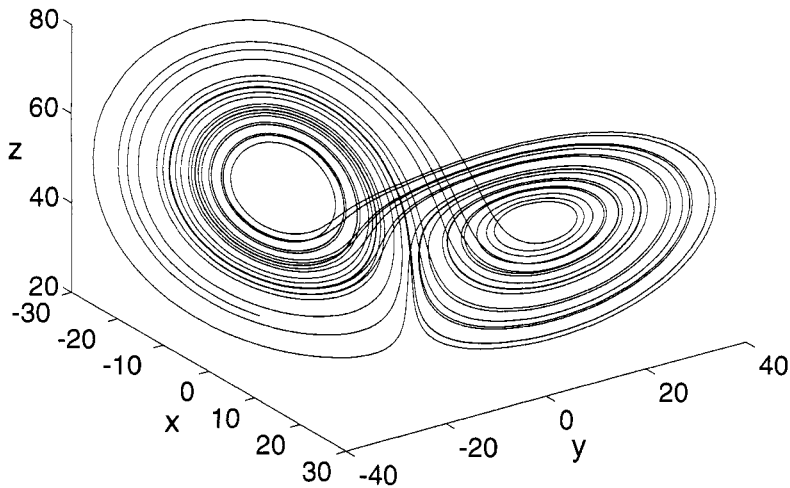
Figure 15.9. Lorenz attractor.

series. Therefore, the additional spectral peak of 37 kHz in Fig. 15.10(c) is really quite weak. In other words, although Ortega's method is able to reveal hidden frequencies, it is not very effective. In fact, Chern et al. have pointed out that the effectiveness of their method, as well as that of Ortega's method, in experimental data analysis remains uncertain.

One of the major difficulties with the above two methods is conceptual; both methods are based on very small-scale neighbors, while the phenomenon itself is large-scale. Recognizing this, one can readily understand that the concept of limiting scale, developed in Sec. 15.2.5, provides an excellent solution to the problem. In other words, one can simply take the Fourier transform of the temporal evolution of the limiting scale and expect additional well-defined spectral peaks if the dynamics contain hidden frequencies. To illustrate the idea, we have shown in Figs. 15.8(d–f) the magnitude of the Fourier transform of the limiting scales of $x(t), y(t)$, and $z(t)$ of the Lorenz system. We observe very well-defined spectral peaks in all three cases. In fact, now the z component no longer plays a more special role than the x and y components. The method is also amazingly effective in identifying the hidden frequency from the laser data, as shown in Fig. 15.10(d). We now not only observe an additional spectral peak around 37 KHz in (d), but also that this peak is even more dominant than the peak around 1.7 MHz. The reason for the exchange of this dominance is the fairly large sampling time; although 80 ns is small, *it is only* able to sample about eight points in each oscillation. When the embedding window,
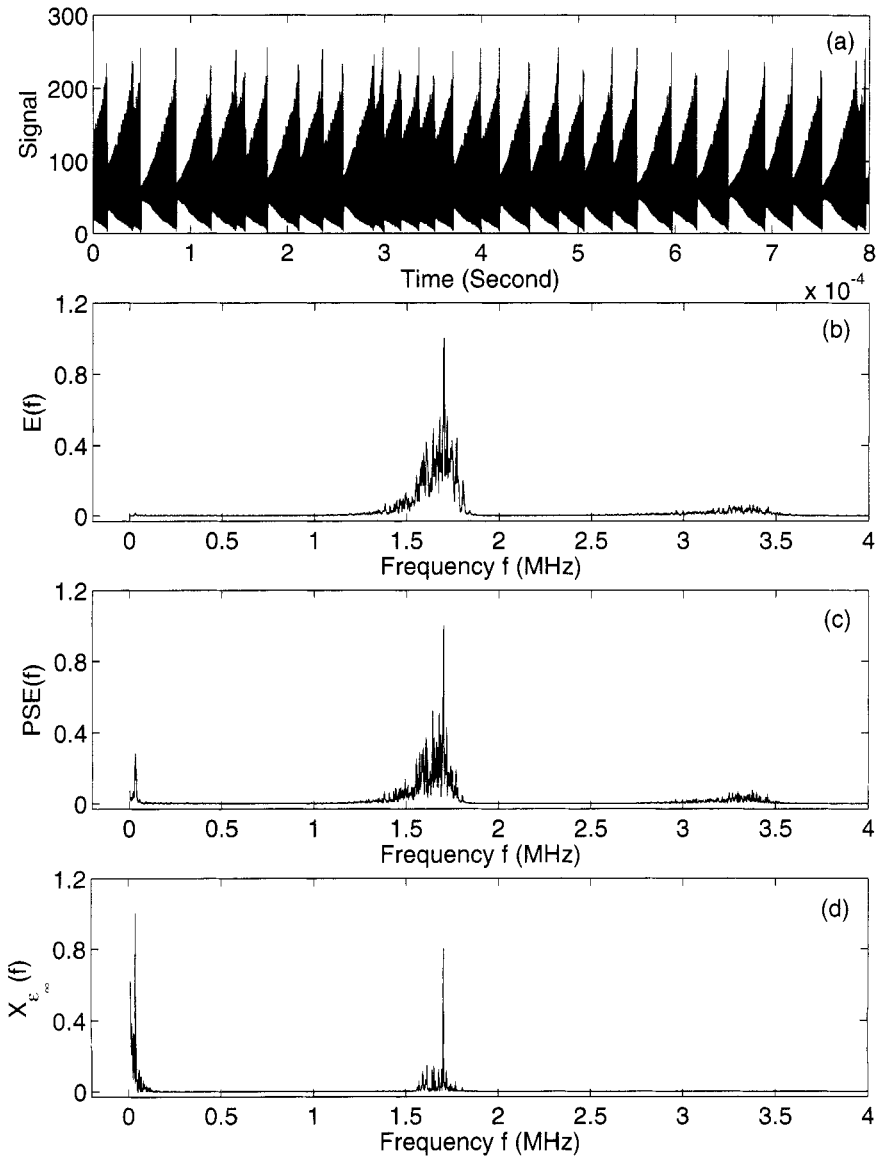
Figure 15.10. (a) Laser intensity data, (b) the PSD of the original laser intensity time series, (c) the PSD of the density measure of the phase space reconstructed from the laser data, and (d) the magnitude of the Fourier transform of the limiting scale of the laser data. In (c) and (d), the embedding parameters are $m = 4, L = 1$.

$(m - 1)L$, becomes greater than 4, the Nyquist sampling theorem is violated in the reconstructed phase space; then this peak may even diminish.

It is important to note that the limiting scale is not greatly affected by either measurement noise or dynamic noise. Therefore, one can expect that the hidden frequencies revealed by limiting scales will not be greatly affected by noise either.

15.5   COPING WITH NONSTATIONARITY

We have observed in Chapter 1 and other chapters that many real-world data such as HRV data, sea clutter data, and economic time series are highly nonstationary. One important source of nonstationarity is sudden jumps or outliers. Some of the jumps are intrinsic to the system, while others may be caused by errors during measurement. Such jumps are salient features of HRV data as well as economic time series. Another source of nonstationarity is the existence of oscillatory components from time to time. For example, respiration can contribute an oscillatory component to HRV data, and business cycles or other oscillatory events in financial markets can contribute oscillatory components to financial data.

In the literature, HRV data have been characterized by $1/f^{2H+1}$ processes with $0 < H < 1/2$. As we shall see in Sec. 15.7.3, economic time series also resemble $1/f^{\beta}$ processes. To better understand how the SDLE can handle nonstationarity, below we first study complicated processes obtained by applying the following two steps to a $1/f^{\beta}$ process:

1. Shift a $1/f^{\beta}$ process downward or upward at randomly chosen points in time by an arbitrary amount. For convenience, we call this procedure type-1 nonstationarity and the processes obtained broken-$1/f^{\beta}$ processes.

2. At randomly chosen time intervals, concatenate randomly broken-$1/f^{\beta}$ processes and oscillatory components or superimpose oscillatory components on broken-$1/f^{\beta}$ processes. This procedure causes a different type of nonstationarity, which for convenience we shall call type-2 nonstationarity.

We call the resulting random processes perturbed $1/f^{\beta}$ processes. A number of examples of the $\lambda(\epsilon)$ curves for such processes are shown in Fig. 15.11. We observe that Eq. (15.11) still holds very well when $\lambda(\epsilon) > 0.02$.

To understand why the SDLE can deal with type-1 nonstationarity, it suffices to note that type-1 nonstationarity causes shifts of the trajectory in phase space; the greater the nonstationarity, the larger the shifts. The SDLE, however, cannot be affected much by shifts, especially large ones, since it is based on the coevolution of pairs of vectors within chosen small shells. In fact, the effect of shifts is to exclude a few pairs of vectors that were originally counted in the ensemble average. Therefore, so long as the shifts are not frequent, the effect of shifts can be neglected, since ensemble averaging within a shell involves a large number of pairs of vectors.
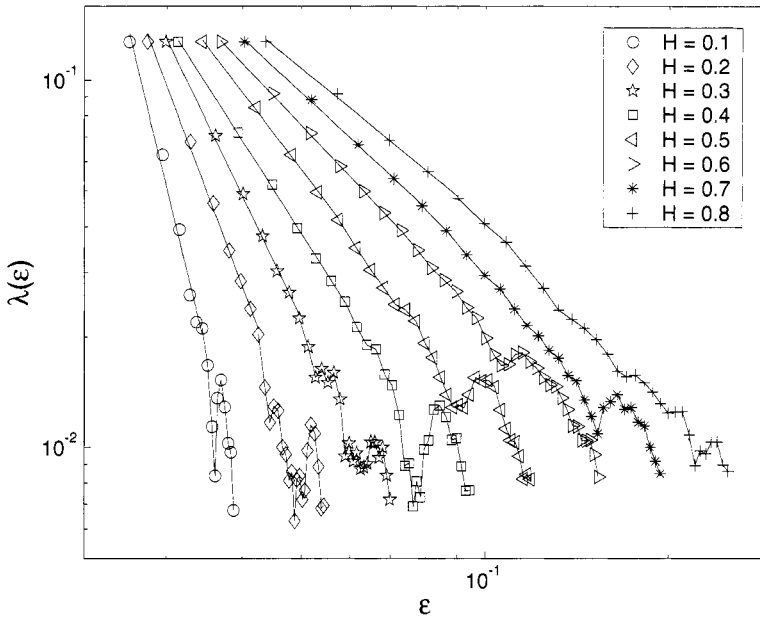
Figure 15.11. $\lambda(\epsilon)$ vs. ϵ curves for perturbed $1/f$ processes. Eight different H are considered. To put all the curves on one plot, the curves for different H (except the smallest one considered here) are arbitrarily shifted rightward.

Let us now turn to type-2 nonstationarity, which involves oscillatory compo-
nents. Being regular, these components can only affect $\lambda(\epsilon)$ where it is close to 0.
Therefore, type-2 nonstationarity cannot affect the positive portion of $\lambda(\epsilon)$ either.

Finally, we examine how the two perturbations may affect deterministically
chaotic or noisy chaotic signals. We find that the characteristic features of the
SDLE for chaotic signals are not affected much either, as shown in Fig. 15.12.

15.6   RELATION BETWEEN SDLE AND OTHER COMPLEXITY MEASURES

In Chapter 1, we mentioned that in order to understand brain dynamics and diagnose
brain pathologies, a number of complexity measures from information theory, chaos
theory, and random fractal theory have been used to analyze EEG data. Since these
theories have different foundations, it has been difficult to compare studies based
on different complexity measures. We shall now show that this problem can be
solved by relating those complexity measures to the value of the SDLE at specific
scales. For coherence of the presentation, we shall directly explain the results,
postponing descriptions of the complexity measures to Appendix C, if they have
not been already discussed.

To compare the SDLE with other complexity measures, seven patients' *multiple-
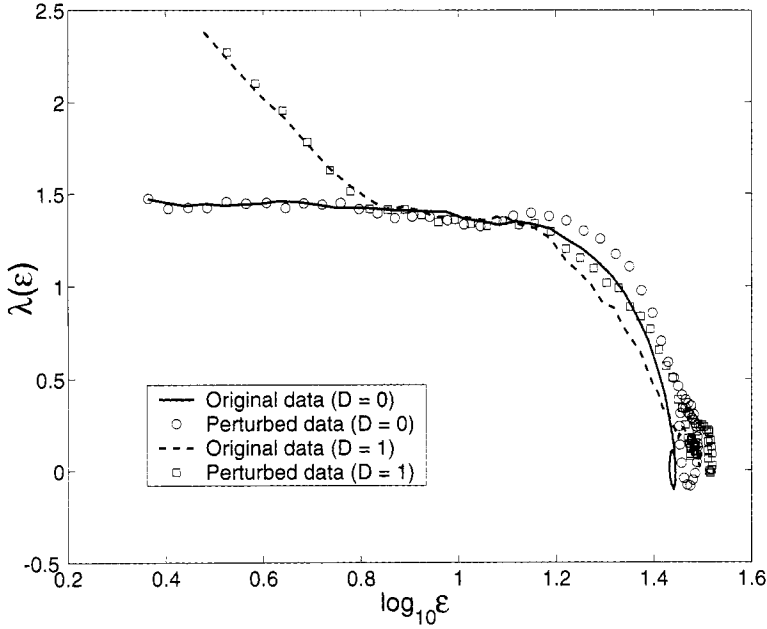channel* intracranially measured EEG data, each with a duration of a few hours, with

Figure 15.12. $\lambda(\epsilon)$ vs. ϵ curves for the perturbed Lorenz data. It may be useful to compare this figure with Fig. 15.1(a).

a sampling frequency of 200 Hz, have been analyzed. As is typically done, the long EEG signals are partitioned into short windows of length W points, and the measures of interest for each window are calculated. While the following results are based on $W = 2048$ points, it has been found that the variations in these measures with time are largely independent of the window size W. The relations among the measures studied here are the same for all patients' EEG data, so we illustrate the results based on only one patient's EEG signals.

We first consider the Lempel-Ziv (LZ) complexity, denoted as C_{LZ}. As is explained in Sec. C.2 of Appendix C, the LZ complexity is a popular measure used to characterize the randomness of a signal. It is closely related to the Kolmogorov complexity and the Shannon entropy. Figure 15.13(a) shows the variation in C_{LZ} with time. The vertical red dashed lines indicate seizure occurrence times. We note that at the onset of seizure, C_{LZ} increases sharply; then it decreases sharply, followed by a gradual increase. While the sudden sharp increase in C_{LZ} is not observed in all seven patients' data, the sharp decrease and then the gradual increase in C_{LZ} are typical features. This indicates that the dynamics of the brain first become more regular right after the seizure; then their irregularity increases as they approach the normal state.

Next, we examine the permutation entropy (PE), denoted as E_p. As is explained in Sec. C.3 of Appendix C, it is a measure from chaos theory. It measures the

departure of the time series under study from a completely random one: the smaller the value of E_p, the more regular the time series is. Figure 15.13(b) shows the $E_p(t)$ for the same patient's EEG data. We observe that the variation of E_p with t is similar to that of C_{LZ} with t.

Third, we consider the correlation entropy, K_2. As is discussed in Chapter 13, K_2 is a tight lower bound of the Kolmogorov entropy. It is zero, positive, and infinite for regular, chaotic, and random data, respectively. Its variation with time is shown in Fig. 15.13(c). Its pattern is similar to that of Figs. 15.13(a,b). This is as expected, since all three measures are entropy measures.

As the fourth measure, we examine the correlation dimension, D_2. First, we recall from Chapter 13 that D_2 is a tight lower bound of the box-counting dimension and measures the minimal number of variables needed to fully describe the dynamics of the data. It is independent of the correlation entropy. However, from Fig. 15.13(d), we find that its variation with time is similar to $K_2(t)$ of Fig. 15.13(c) (as well as Figs. 15.13(a,b)). This is a puzzling observation.

It turns out that the relation between the largest Lyapunov exponent λ_1 and the correlation entropy is reciprocal, as can be seen by comparing Figs. 15.13(e) and 15.13(c). This is an even bigger puzzle for the following reason. The summation of the positive Lyapunov exponents is typically equal to the Kolmogorov entropy. As a major component contributing to the Kolmogorov entropy, $\lambda_1(t)$ is expected to be similar to $K_2(t)$. However, we have observed a reciprocal relation between $\lambda_1(t)$ and $K_2(t)$.

Finally, we consider the calculation of the Hurst parameter H using DFA, one of the most reliable methods for estimating H. The variation of H with time is shown in Fig. 15.13(f). We observe that $H(t)$ and $\lambda_1(t)$ are similar. This is another surprising relation, since the Hurst parameter is a measure from the random fractal theory, while $\lambda_1(t)$ is a measure from chaos theory.

We now calculate the SDLE for each segment of the EEG data. Two representative examples for seizure and nonseizure segments are shown in Fig. 15.14. We observe that on a specific scale ϵ^*, the two curves cross. Loosely, we may term any $\epsilon < \epsilon^*$ small scale, while any $\epsilon > \epsilon^*$ is large scale. Therefore, on small scales, $\lambda(\epsilon)$ is smaller for seizure than for nonseizure EEGs, while on large scales the opposite is true. The variations of $\lambda_{\text{small}-\epsilon}$ and $\lambda_{\text{large}-\epsilon}$ with time in this patient's data, where small $-\epsilon$ and large $-\epsilon$ stand for (more or less arbitrarily) chosen fixed small and large scales, are shown in Figs. 15.15(a) and 15.15(b), respectively. We observe that the pattern of variation of $\lambda_{\text{small}-\epsilon}(t)$ is the reciprocal of that of $\lambda_{\text{large}-\epsilon}(t)$. This result can be expected from Fig. 15.14. More importantly, the patterns of $\lambda_{\text{small}-\epsilon}(t)$ and $\lambda_{\text{large}-\epsilon}(t)$ encompass the two patterns for the six measures presented in Fig. 15.13.

We are now ready to resolve all of the curious relations observed among the six measures presented in Fig. 15.13.
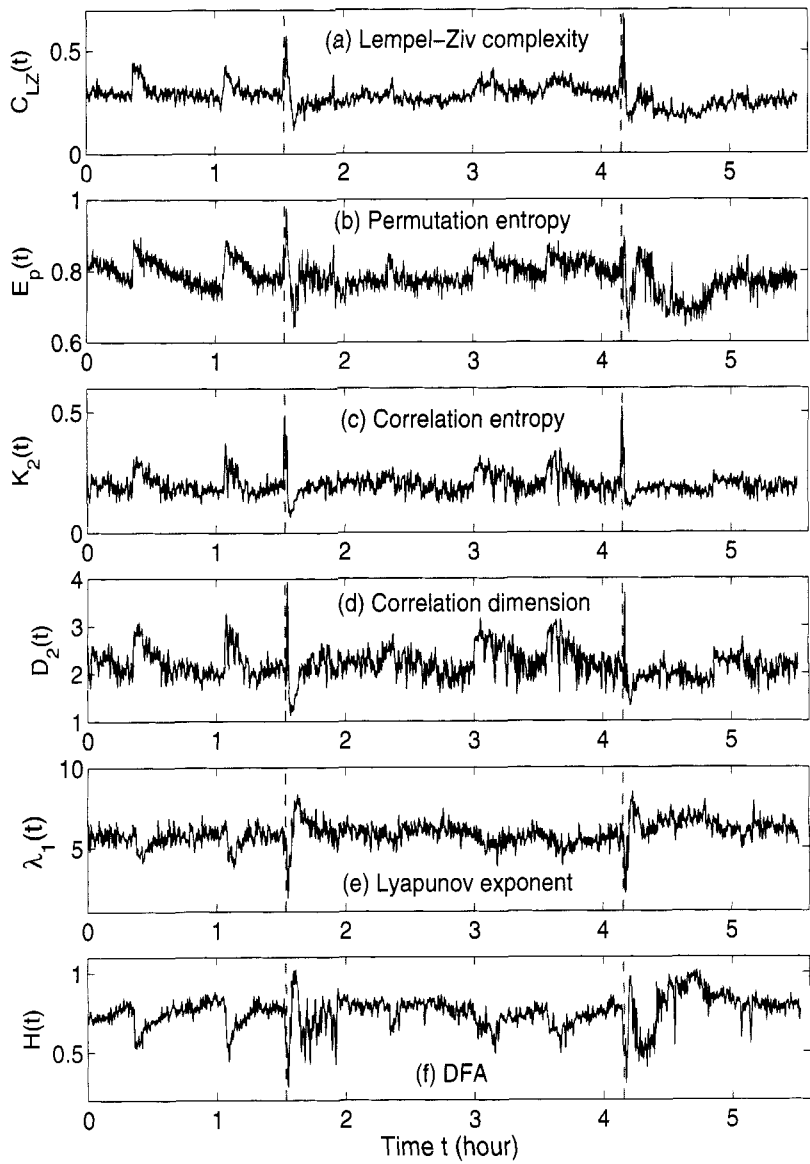
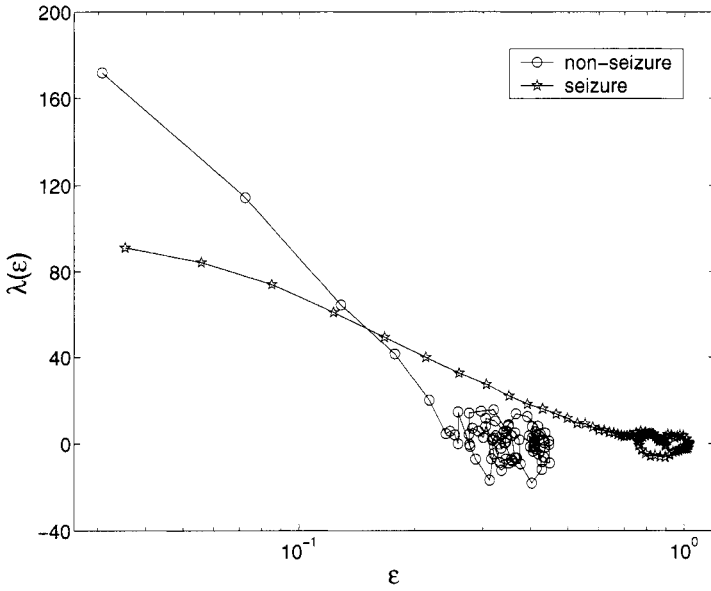**Figure 15.13.** Complexity measures for EEG.

Figure 15.14.   Representative $\lambda(\epsilon)$ (per second) vs. ϵ for seizure and nonseizure EEG segments.

1. Generally, entropy measures the randomness of a dataset. This pertains to small scale. Therefore, $C_{LZ}(t)$, $E_P(t)$, and $K_2(t)$ should be similar to $\lambda_{small-\epsilon}(t)$. This is indeed the case. Therefore, we can conclude that the variation of entropy is represented by $\lambda_{small-\epsilon}(t)$, regardless of how entropy is defined.

2. To understand why $\lambda_1(t)$ calculated by the algorithm of Wolf et al. [478] corresponds to $\lambda_{large-\epsilon}(t)$, it is sufficient to note that when the algorithm of Wolf et al. is applied to a time series with only a few thousand points, in order to obtain a well-defined Lyapunov exponent, a fairly large scale parameter has to be chosen. Typically, the probability in Eq. (15.6) is larger on large scales than on small scales, therefore, the Lyapunov exponent and $\lambda_{large-\epsilon}$ are similar. In fact, the scale we have chosen to calculate $\lambda_1(t)$ is even larger than that for calculating $\lambda_{large-\epsilon}(t)$. This is the reason that the value of $\lambda_1(t)$ shown in Fig. 15.13(e) is smaller than that of $\lambda_{large-\epsilon}(t)$ shown in Fig. 15.15(b).

3. It is easy to see that if one fits the $\lambda(\epsilon)$ curves shown in Fig. 15.14 by a straight line, then the variation of the slope with time should be similar to $\lambda_{small-\epsilon}(t)$ but the reciprocal of $\lambda_{large-\epsilon}(t)$. Such a pattern will be preserved even if one takes the logarithm of $\lambda(\epsilon)$ first and then does the fitting. Note that the slope plays the role of the Hurst parameter. Therefore, even if the EEG is not ideally of the $1/f^{2H+1}$ type, qualitatively the relation $\lambda(\epsilon) \sim \epsilon^{-1/H}$ holds. This, in turn, implies that $D_2 \sim 1/H$.
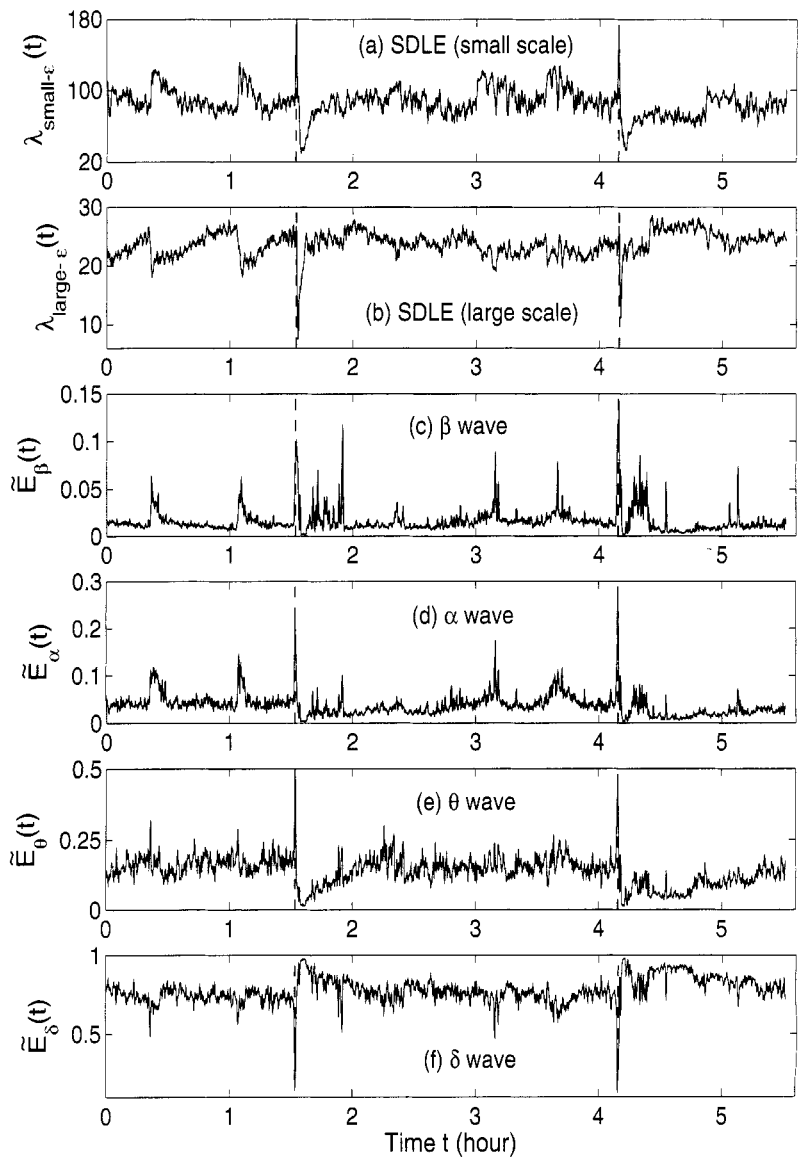
Figure 15.15.   The SDLE and the four waves for the EEG signals.

With the above arguments, it is clear that the seemingly puzzling relations among the measures considered here can be readily understood by the $\lambda(\epsilon)$ curves. Most importantly, we have established that commonly used complexity measures can be related to the values of the SDLE at specific scales.

Before ending this section, we make two comments. (1) To comprehensively characterize the complexity of complicated data such as EEG data, a wide range of scales has to be considered, since the complexity may be different on different scales. For this purpose, the entire $\lambda(\epsilon)$ curve, where ϵ is such that $\lambda(\epsilon)$ is positive, provides a good solution. This point is particularly important when one wishes to compare the complexity between two signals; the complexity of one signal may be higher on some scales but lower on other scales. The situation shown in Fig. 15.14 may be considered one of the simplest. (2) For detecting important events such as epileptic seizures, $\lambda_{\text{small}-\epsilon}$ and $\lambda_{\text{large}-\epsilon}$ appear to provide better-defined features than other commonly used complexity measures. This may be due to the fact that $\lambda_{\text{small}-\epsilon}$ and $\lambda_{\text{large}-\epsilon}$ are evaluated at fixed scales, while other measures are not. In other words, scale mixing may blur the features of events being detected, such as seizures.

15.7 BROAD APPLICATIONS

In Sec. 15.2.5, we have introduced the notion of limiting scales, ϵ_∞, and pointed out that they contain a lot of useful information about the structured components (i.e., regularly oscillating components) of the data. More generally, we can say that they are closely related to the total variation or the energy of the signal. For example, for a chaotic system, ϵ_∞ defines the size of the chaotic attractor. If one starts from $\epsilon_0 \ll \epsilon_\infty$, then, regardless of whether the data are deterministically chaotic or simply random, ϵ_t will initially increase with time and gradually settle at ϵ_∞. Consequently, $\lambda(\epsilon_t)$ will be positive before ϵ_t reaches ϵ_∞. On the other hand, if one starts from $\epsilon_0 \gg \epsilon_\infty$, then ϵ_t will simply decrease, yielding negative $\lambda(\epsilon_t)$, again regardless of whether the data are chaotic or random. When $\epsilon_0 \sim \epsilon_\infty$, $\lambda(\epsilon_t)$ will remain around 0. With these comments, we are now ready to tackle various types of complex data.

15.7.1 EEG analysis

In Sec. 15.6, we analyzed EEG by computing a number of different complexity measures. We complete our study of EEG by trying to understand the four waves — β, α, θ and δ of the EEG.

In Sec. 4.1, we pointed out that these waves are not characterized by sharp spectral peaks in the frequency domain. Therefore, it would make more sense to find the total energy of each type of wave. Denote them by $E_\beta, E_\alpha, E_\theta$, and E_δ, respectively. They can be obtained by integrating the PSD over their defining frequency bands. Furthermore, we normalize them by the total energy in the frequency band $(0, 40)$

Hz. Denote the normalized energy, or fraction of energy, by $\tilde{E}_\beta, \tilde{E}_\alpha, \tilde{E}_\theta$, and $\tilde{E}_\delta$. Note that normalization does not change the functional form for temporal variations of the energy of each type of wave. The variations of these normalized energies with time in the EEG data of the patient discussed in Sec. 15.6 are shown in Figs. 15.15(c–f). Comparing with Figs. 15.15(a,b), we observe that, overall, the variations of $\tilde{E}_\beta(t), \tilde{E}_\alpha(t)$, and $\tilde{E}_\theta(t)$ are similar to that of $\lambda_{\text{small}-\epsilon}(t)$, while the variation of $\tilde{E}_\delta(t)$ is similar to that of $\lambda_{\text{large}-\epsilon}(t)$. This suggests that small $-\epsilon$ corresponds to high frequency, or small temporal scales, while large $-\epsilon$ corresponds to low frequency, or large temporal scales. It should be emphasized, however, that the features uncovered by spectral analysis are not as good as those uncovered by the complexity measures discussed here. Furthermore, when the high-frequency bands, such as the β wave band, are used for seizure detection, the variation of the corresponding energy with time shows activities that are not related to seizures but are related to noise such as motion artifacts. Therefore, for the purpose of epileptic seizure detection/prediction, spectral analysis (or other methods, such as wavelet-based methods, for extracting equivalent features) may not be as effective as the complexity measures discussed here, especially the SDLE.

Let us now quantitatively understand the relation between $\tilde{E}_\beta(t), \tilde{E}_\alpha(t), \tilde{E}_\theta(t)$, and $\lambda_{\text{small}-\epsilon}(t)$. In Sec. 15.6, we pointed out that EEG signals may be approximated as a type of $1/f^{2H+1}$ process. Integrating $1/f^{2H+1}$ over a frequency band of $[f_1, f_2]$, we obtain the energy

$$E(f_1 \rightarrow f_2) = \int_{f_1}^{f_2} \frac{1}{f^{2H+1}} df = \frac{1}{2H}\left[f_1^{-(2H+1)} - f_2^{-(2H+1)}\right].$$

This suggests that the variation of $E(f_1 \rightarrow f_2)$ is similar to $1/H(t)$. Since the variation of H with t is reciprocal to that of $\lambda_{\text{small}-\epsilon}$, as can be seen in Figs. 15.13(f) and 15.15(a), we thus see that β, α, and θ waves can be approximated by $1/f^{2H+1}$-type processes. The δ wave, however, is not of the $1/f^{2H+1}$ type.

15.7.2   HRV analysis

We now examine HRV data and show how the SDLE can readily characterize the hidden differences in HRV under healthy and disease conditions.

We examine two types of HRV data, one for healthy subjects, another for subjects with congestive heart failure (CHF), a life-threatening disease. The data were downloaded from PhysioNet. There are 18 healthy subjects and 15 subjects with CHF. Parts of these datasets were analyzed by other methods. In particular, 12 of the 15 CHF datasets were analyzed by wavelet-based multifractal analysis and point process adaptive filtering for the purpose of distinguishing healthy subjects from CHF patients. For ease of comparison, we take the first 3×10^4 points of both groups of HRV data for analysis. Figures 15.16(a,b) show two typical $\lambda(\epsilon)$ vs. ϵ curves, one for a healthy subject and another for a patient with CHF. We observe that for the healthy subject, $\lambda(\epsilon)$ linearly decreases with $\ln \epsilon$ before λ approximately
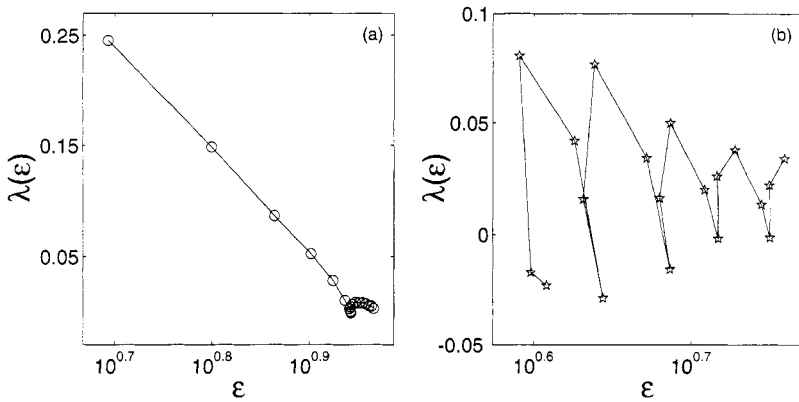
Figure 15.16. $\lambda(\epsilon)$ (per beat) vs. ϵ (on a semilog scale) for HRV data of (a) a healthy subject and (b) a subject with CHF.

reaches 0 or before ϵ settles around the characteristic scale, ϵ_∞. Recall that this is a characteristic of noisy dynamics (Fig. 15.1). For the CHF case plotted in Fig. 15.16(b), we observe that $\lambda(\epsilon)$ is oscillatory, with its value always close to 0; hence, the only scale resolvable is around ϵ_∞. Since the length of the time series used in our analysis for the healthy and CHF subjects is the same, the inability to resolve the $\lambda(\epsilon)$ behavior on scales much smaller than ϵ_∞ for patients with CHF strongly suggests that the dimension of the dynamics of the cardiovascular system for CHF patients is considerably higher than that of healthy subjects.

We now discuss how to distinguish between healthy subjects and patients with CHF from HRV analysis. We have devised two simple measures, or features. One is to characterize how well the linear relation between $\lambda(\epsilon)$ and $\ln \epsilon$ can be defined. We have quantified this by calculating the error between a fitted straight line and the actual $\lambda(\epsilon)$ vs. $\ln \epsilon$ plots of Figs. 15.16(a,b). The second feature is to characterize how well the characteristic scale ϵ_∞ is defined. This is quantified by the ratio between two scale ranges, one from the 2nd to the 6th point of the $\lambda(\epsilon)$ curves and the other from the 7th to the 11th point of the $\lambda(\epsilon)$ curves. Now each subject's data can be represented as a point in the feature plane, as shown in Fig. 15.17. We observe that for healthy subjects, feature 1 is generally very small but feature 2 is large, indicating that the dynamics of the cardiovascular system are like those of a nonlinear system with stochasticity, with resolvable small-scale behaviors and a well-defined characteristic scale ϵ_∞. The opposite is true for the patients with CHF: feature 1 is large but feature 2 is small, indicating not only that small-scale behaviors of the $\lambda(\epsilon)$ curves cannot be resolved but also that the characteristic scale ϵ_∞ is not well defined. Very interestingly, these two simple features completely separate the normal subjects from patients with CHF. In fact, each feature alone can almost perfectly separate the two groups of subjects studied here.
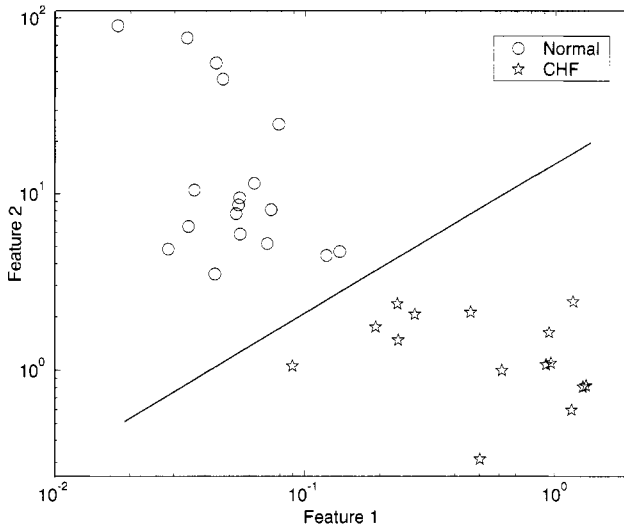
Figure 15.17.   Feature plane separating normal subjects from subjects with CHF.

In recent years, many efforts have been made to search for cardiac chaos. Due to the inability to unambiguously distinguish deterministic chaos from noise by calculating the largest positive Lyapunov exponent and the correlation dimension, it is still unclear whether the control mechanism of the cardiovascular system is truly chaotic or not. Our analysis here strongly suggests that if cardiac chaos does exist, it is more likely to be identified in healthy subjects than in pathologic groups. This is because the dimension of the dynamics of the cardiovascular system appear to be lower for healthy than for pathologic subjects. Intuitively, such an implication makes sense, because a healthy cardiovascular system is a tightly coupled system with coherent functions, while components in a malfunctioning cardiovascular system are somewhat loosely coupled and function incoherently.

15.7.3   Economic time series analysis

In Chapter 1, we pointed out that although economic time series are very complicated, recent studies have reported negative largest Lyapunov exponents for various types of economic time series. We now show that the reported negative largest Lyapunov exponent may correspond to the value of the SDLE at fairly large scales and therefore does not necessarily imply regular economic dynamics.

15.7.3.1   Analysis of a chaotic asset pricing model   We first analyze a chaotic asset pricing model with heterogeneous beliefs driven by dynamic noise proposed by Brock and Hommes [58]. The model is a nonlinear map of the form

$$x_t = F(x_{t-1}, x_{t-2}, x_{t-3}) + \sigma\eta_t. \tag{15.19}$$

Specifically, the model is described by the following equations:

$$x_t = \frac{1}{R} \sum_{h=1}^{4} n_{h,\,t}(g_h x_{t-1} + b_h) + \sigma \eta_t \qquad (15.20)$$

$$n_{h,\,t} = \frac{e^{\beta U_{h,\,t-1}}}{\sum_{j=1}^{4} e^{\beta U_{j,\,t-1}}} \qquad (15.21)$$

$$U_{h,\,t-1} = (x_{t-1} - R x_{t-2})(g_h x_{t-3} + b_h - R x_{t-2}). \qquad (15.22)$$

Here x_t denotes the deviation of the price of a risky asset from its benchmark fundamental value (the discounted sum of expected future dividends), $R > 1$ is the constant risk-free rate, $n_{h,\,t}$ represents the discrete choice of agents using belief type h, $U_{h,\,t-1}$ is the profit generated by strategy h in the previous period, g_h and b_h characterize the linear belief with one time lag of strategy h, and the noise term $\sigma \eta_t$ is standard normally distributed noise with variance σ^2. For suitable choice of the parameter values, the model exhibits chaotic dynamics. In particular, Brock and Hommes have studied the model at parameter values $\beta = 90$, $R = 1.1$, $g_1 = b_1 = 0$, $g_2 = 1.1$, $b_2 = 0.2$, $g_3 = 0.9$, $b_3 = -0.2$, $g_4 = 1.21$, $b_4 = 0$. They found that the motion is chaotic with the largest Lyapunov exponent 0.1. More recently, Hommes and Manzan [222] studied the model at parameter values $\beta = 90$, $R = 1.01$, $g_1 = b_1 = 0$, $g_2 = 0.9$, $b_2 = 0.2$, $g_3 = 0.9$, $b_3 = -0.2$, $g_4 = 1.01$, $b_4 = 0$. When $\sigma = 0$, they found that the motion is chaotic with the largest Lyapunov exponent around 0.135. When σ was increased beyond 0.1, they found that the largest Lyapunov exponent becomes negative. Here we shall focus on these two parameter sets. For convenience, we call them the BH and HM parameter sets, respectively. We have calculated the $\lambda(\epsilon)$ curves for the model with the BH and HM parameter sets for various levels of noise. The results are shown in Figs. 15.18(a,b). We observe that they are very similar to those in Figs. 15.1. Specifically, we note the following:

1. The value of about 0.1 for the largest positive Lyapunov exponent shown in Fig. 15.18(b) is consistent with the Lyapunov exponent obtained by Brock and Hommes, but the value of about 0.3 for the largest positive Lyapunov exponent shown in Fig. 15.18(a) is larger than the value of the Lyapunov exponent obtained by Hommes and Manzan using the neural network–based Lyapunov exponent estimator (as well as the algorithm of Wolf et al.). We shall explain why this discrepancy exists shortly.

2. The scale range where $\lambda(\epsilon)$ is almost constant is much wider for the HM parameter set than for the BH parameter set. In fact, only 10,000 points were used to calculate Fig. 15.18(a), while 30,000 points were used for Fig. 15.18(b). The reason is that the correlation dimension for the chaotic attractor of the HM parameter set is only about 0.4, while that of the BH parameter set is about 2.0. This means that in order to resolve the behavior of $\lambda(\epsilon)$ on ever smaller scales, longer and longer time series have to be used.
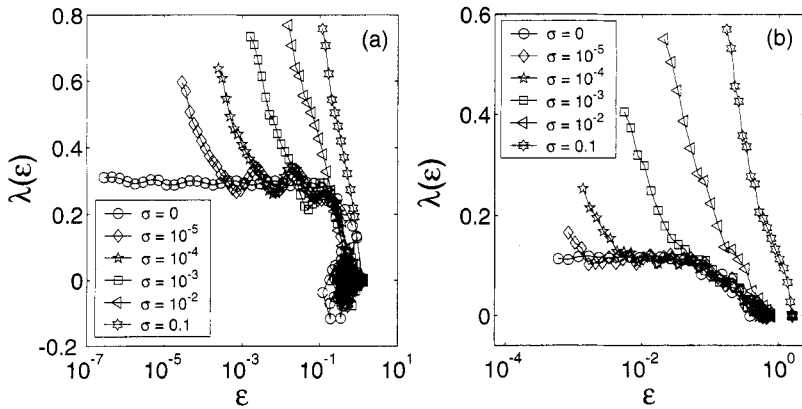
Figure 15.18. $\lambda(\epsilon)$ curves for the clean and the noisy asset pricing model with (a) the HM parameter set and (b) the BH parameter set.

More precisely, for a given dataset, if the smallest resolvable scale is ϵ_0, in order to resolve a smaller scale ϵ_0/r, where $r > 1$, a larger dataset has to be used. The larger the dimension of the attractor, the longer the time series has to be.

We can now understand why the Lyapunov exponent calculated by Hommes and Manzan using the neural network–based Lyapunov exponent estimator and the algorithm of Wolf et. al. is smaller than our estimate in the noise-free case. The answer is given by Eq. (15.6), noting the fact that the probability in Eq. (15.6) is typically larger on large scales than on small scales. As we pointed out earlier, when the algorithm of Wolf et al. is applied to a time series of finite length, in order to obtain a well-defined Lyapunov exponent, a fairly large-scale parameter has to be chosen. The discrepancy between our estimate and that of Hommes and Manzan suggests that the scale parameter used by them, using the algorithm of Wolf et al., is larger than the scales where $\lambda(\epsilon)$ is almost constant.

We now discuss the neural network–based Lyapunov exponent estimator. Here the Lyapunov exponent is estimated by the Jacobian matrix method. The main function of the neural network is to estimate a trajectory in the phase space so that the Jacobian matrices can be computed. When estimating the trajectory in the phase space, global optimization is employed such that the error between the given and estimated time series is minimized. Global optimization involves a scale parameter. Hommes and Manzan's analysis suggests that this scale parameter is comparable to the one used by the algorithm of Wolf et al.

Finally, we discuss why the estimated Lyapunov exponent can be negative in the strongly noise-driven case. The reason is as follows. In the strongly noise-driven case, the scales that can be resolved by a finite dataset are always close to the characteristic scale ϵ_∞. In order for an algorithm to numerically return a non-null

result, the chosen scale parameter has to be larger than ϵ_{∞}. This then produces a negative Lyapunov exponent.

Our discussion up to now has made it clear that an estimated negative Lyapunov exponent may not necessarily imply absence of chaos and, therefore, regular dynamics. On the contrary, the dynamics are simply stochastic. Entropy generation is high even on fairly large scales and is much higher than the case of noise-free, low-dimensional chaos.

15.7.3.2   Analysis of U.S. foreign exchange rate data   Now that we understand the generic behavior of the $\lambda(\epsilon)$ curves for clean and noisy chaotic systems, we analyze some of the U.S. foreign exchange rate data. Altogether, we have analyzed about 20 datasets with a sampling time of 1 day. Three examples are shown in Figs. 15.19(b,d,f), for U.S.-Canada, U.S.-Mexico, and U.S.-Korea exchange rate data. We conclude that foreign exchange rate data, when sampled at a 1 day resolution, are like random walk–type processes and are well characterized by Eq. (15.11), with the Hurst parameter being able to take all types of values (close to, smaller than, and larger than 1/2 in Figs. 15.19(b,d,f), respectively). We note that among the 20 datasets, $H \approx 1/2$ is most prevalent. This indicates that a market is often efficient. However, deviations from the efficient market assumption do exist, as clearly shown in Figs. 15.19(d,f). Such deviations might be due to political factors. If so, then such deviations might be used to study economic/political ties between two nations.

15.7.4   Sea clutter modeling

In Sec. 8.9.1, we discussed target detection within sea clutter using the Hurst parameter. While the detection accuracy is high, we have emphasized that the scaling behavior is not well defined for sea clutter data. It turns out that sea clutter data are more complicated than other data examined in this book in the sense that the $\lambda(\epsilon)$ curves corresponding to different shells do not collapse together. Interestingly, each $\lambda(\epsilon)$ curve behaves very well. Figure 15.20(a) shows one example for a fairly large shell. When the slope of the $\lambda(\epsilon)$ curve is used to detect targets within sea clutter, we obtain figures like that shown in Fig. 15.20(b). Interestingly, if we form histograms of the slopes for data with the primary target and without the target, the histograms completely separate, as shown in the right subplot of Fig. 8.14. More importantly, the time scale for $\lambda(\epsilon) = -\gamma \ln \epsilon$ to be valid is up to 0.01 s, smaller than the time scale of $0.04 - 1$ s identified when we performed fractal analysis in Sec. 8.9.1. Furthermore, the pattern of the $\lambda(\epsilon)$ curves where they are about 0 is quite different for sea clutter data with and without a target. All this information can be combined to improve the accuracy of target detection.
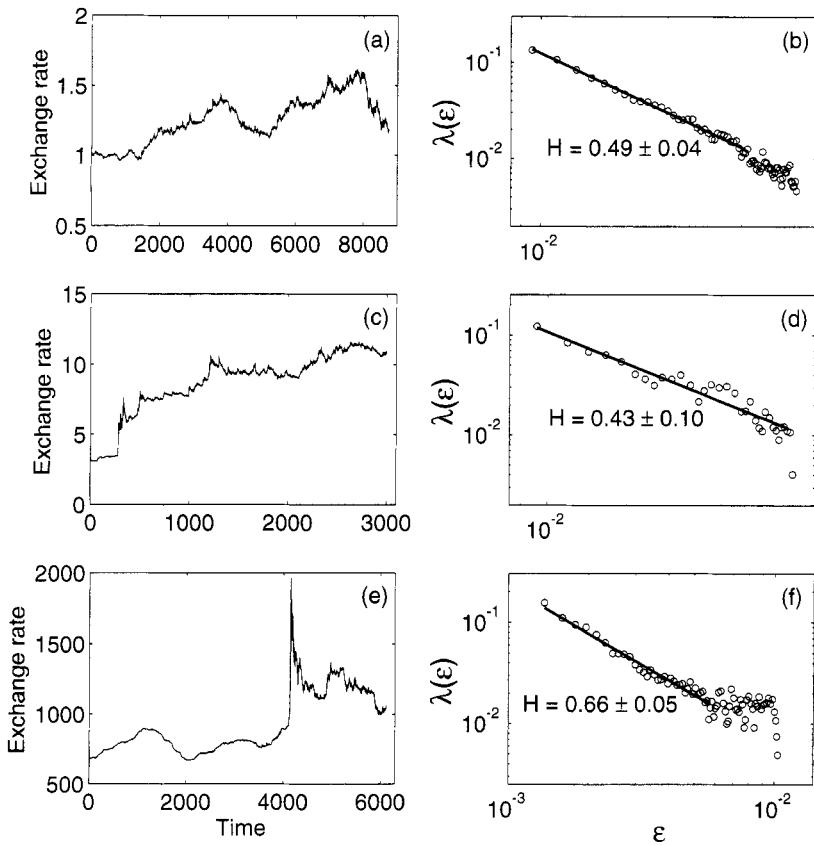
Figure 15.19. Foreign exchange rate data between the United States and (a) Canada, (c) Mexico, and (e) Korea; (b,d,f) are the corresponding $\lambda(\epsilon)$ curves.

15.8 BIBLIOGRAPHIC NOTES

Most of the materials in this chapter are new. They have been developed based on the concepts of the SDLE [156] and time-dependent exponent curves [178–180]. The former is closely related to the finite-size Lyapunov exponent [16, 53, 64, 262] and (ϵ, τ)-entropy [183]. See also approximate entropy and sample entropy in Sec. 13.2.3. For distinguishing chaos from noise, we refer to [28, 104, 105, 179, 180, 182, 197, 229, 251, 256, 339, 352, 358, 380, 416, 451, 469]. In particular, [99, 135, 241, 354, 371, 410, 433] are about whether EEG signals are chaotic or not, while [11, 14, 220, 242, 275, 300, 356] discuss the appropriateness of using nonlinear measures for epileptic seizure forewarning. For complexity of biomedical signals, we refer to [2, 10, 208, 272, 317, 325, 332, 360, 365, 367, 412, 419, 442–445, 468, 484, 485, 489–494]. For HRV analysis, we refer to [6, 15, 17, 29, 49, 52, 123, 127, 137, 194, 195, 221, 223, 232, 235, 236, 252, 261, 264, 292, 303, 342,
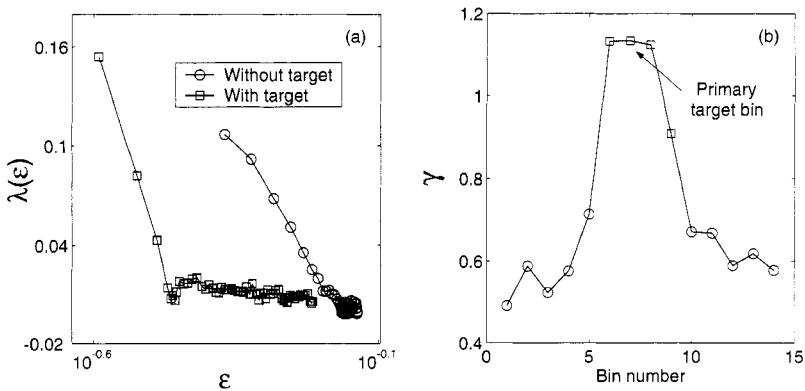
Figure 15.20. (a) $\lambda(\epsilon)$ curves for sea clutter. The embedding parameters are $m = 4, L = 1$. (b) Target detection using γ parameter.

343, 346, 386, 427, 453, 486]. For chaos in cardiovascular systems, we refer to [17, 138, 181, 194, 195, 244, 252, 303]. For chaos in economic time series, we refer to [21, 34, 57–59, 97, 112, 200, 201, 222, 273, 389, 395, 401, 402]. For memory and multifractal features of economic time series, we refer to [329, 498]. For analysis of nonstationary signals, we refer to [115, 146, 147, 149, 152, 177, 446, 470]. For characterizing hidden frequencies in time series, we refer to [74, 326, 327]. For nonstationarity, we refer to [149, 230, 254, 299, 392].