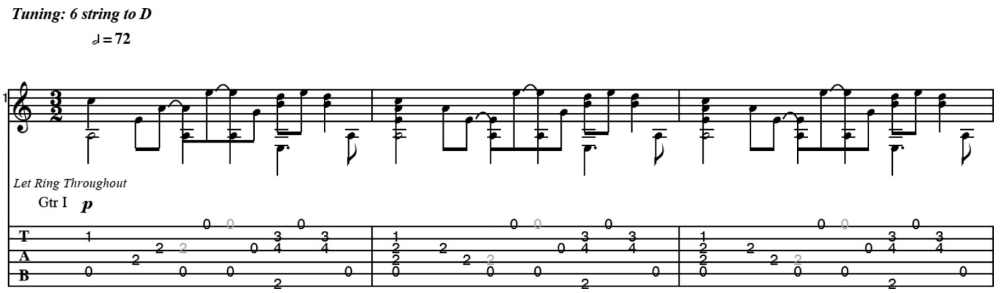
5

Analysing Data from Stochastic Dynamical Systems



Let the data speak and conquer the noise.
DADGBE, Andecy, Andrew York

In the previous chapters, we were given a stochastic model for a random variable X_t and computed the statistics of the solutions from the properties of the equations. However, often only a time series (e.g., from observations) is given, and we a priori may not know the underlying equations. Hence we have to estimate the statistical quantities we are interested in from the time series itself.

We start with a brief overview of classical methods of univariate, bivariate and multivariate time series analysis. The literature on this subject is very broad, and for climate research, one of the main references is Von Storch and Zwiers (1999). It is useful here, however, to repeat some of the main results, as they are needed to understand modern nonlinear techniques, of which some are presented later in this chapter.

5.1 Classical univariate methods

In general, we have a time series at discrete time intervals Δt, the sampling interval. If we denote the variable by X, then we have a discrete time series X_t at points $t = k\Delta t$.

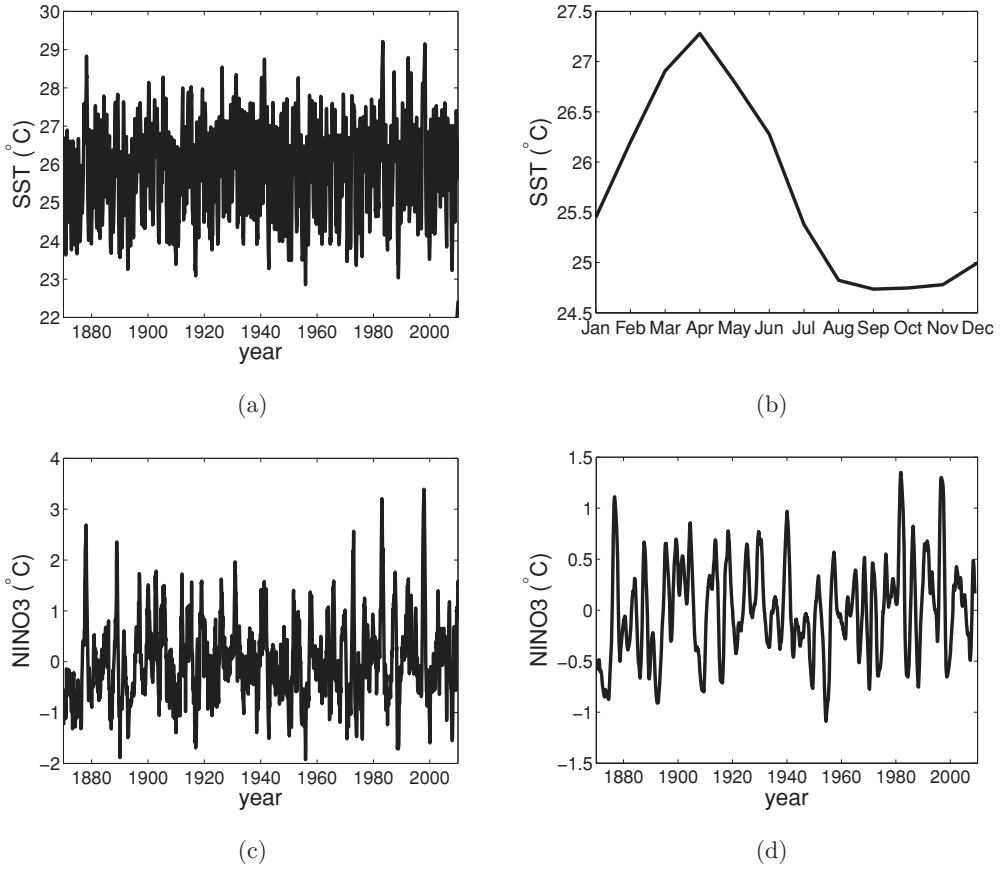


Figure 5.1 (a) Time series over the period 1870–2010 of the sea surface temperature (SST) in the Eastern Equatorial Pacific averaged over the domain [150°W–90°W] × [5°S–5°N] as provided in the HadISST1 data set (Rayner et al., 2003). (b) The mean seasonal cycle obtained from the time series in (a) by averaging over all January months, February, . . . , December. (c) Difference of the time series in (a) and (b) giving the NINO3 time series. (d) Low-pass-filtered time series of (c) with a cut-off value of 24 months.

We illustrate the methodology of analysing such a time series with help of a specific time series (Fig. 5.1a): the sea surface temperature in the Eastern Equatorial Pacific averaged over the domain [150°W–90°W] × [5°S–5°N].

The process of analysis of the time series consists of the following four steps (Chatfield, 2004):

(i) Plot and inspect the time series. This may seem trivial, but it is important to get an impression of trends, periodicities and rapid changes.
(ii) Remove any undesired known periodic cycles (diurnal, seasonal) and long-term trends. Filter the time series, if needed.

(iii) Analyse the result after (ii) in the time domain (correlation analysis) and frequency domain (spectral analysis).
(iv) Obtain the 'signal' from the 'noise'. Formulate a null hypothesis and try to reject each null hypothesis based on a test statistic and a stochastic process representing the noise.

5.1.1 Trend removal and filtering

In many cases, the trend may be just the quantity one is after. Think of the warming trend of surface temperatures due to the increase of greenhouse gases, such as in Fig. 1.3. Or one may just be interested in the seasonal cycle of a particular quantity, say, of the surface pressure in station De Bilt in the Netherlands.

Often, however, climate scientists are interested in the processes causing variability in a particular frequency band, for example, the interannual variability associate with the El Niño/Southern Oscillation (ENSO) phenomenon. A question here may be what the dominant frequency of the ENSO variability is in a particular time series such as that of the SST in Fig. 5.1a. In this case, one wants to remove the long-term trend (maybe caused by global warming) and the seasonal cycle.

There are many methods of linear trend removal. The general model is that the time series can be represented by

$$X_t = X_t^d + \alpha + \beta t + \epsilon_t, \tag{5.1}$$

with constants α and β, a noise term ϵ_t and X_t^d as the detrended time series. There are now two cases:

(i) The observations have a seasonal signal. In this case, one computes annual averages of the data as a new time series Y_t, fits a line to these annual averages to determine α and β in (5.1) and subtracts this line from the original time series.
(ii) When the data have no seasonal signal, one computes the difference time series

$$Y_t = X_{t+1} - 2X_t + X_{t-1}$$

to remove the linear trend and then puts $X_t^d = Y_t$.

To remove a seasonal signal S_t, one first needs to know whether the seasonal signal is additive or multiplicative. In the (most common) additive case, the general model is

$$X_t = X_t^d + S_t + \epsilon_t. \tag{5.2}$$

One can test whether this model is appropriate by checking whether

$$\sum_{t=0}^{n\Delta t} S_t \approx 0, \tag{5.3}$$

where n is the number of observations within a year. In the multiplicative case, the model is $X_t = X_t^d S_t + \epsilon_t$, and the sum (5.3) is approximately unity. For the time series in Fig. 5.1a, the trend is very small, and the seasonal cycle is shown in Fig. 5.1b. The SST anomaly time series with respect to this seasonal cycle is plotted in Fig. 5.1c; this time series is also known as the NINO3 time series. When one has monthly data, one can also remove the seasonal cycle (in the additive case) by computing

$$X_t^d = \frac{1}{12}\left(\frac{1}{2}X_{t-6} + X_{t-5} + \cdots + X_{t+5} + \frac{1}{2}X_{t+6}\right), \tag{5.4}$$

where the coefficients are chosen in such a way as to add up to 12.

The procedure (5.4) is an example of a so-called running mean or moving average filter. A much-used expression is

$$Sm(X_t) = \frac{1}{2q+1}\sum_{r=-q}^{q} X_{t+r}, \tag{5.5}$$

for $t = q + 1$ up to $t = N - q$. In a low-pass filter, only the low-frequency components are retained. We can, for example, study the low-frequency variability of the NINO3 monthly mean time series (Fig. 5.1d) by computing a running mean filter with a time scale of two years ($q = 24$). In a high-pass filter, we perform the same operation but then consider the time series $X_t - Sm(X_t)$.

5.1.2 Elementary statistical properties

The detrended (and maybe filtered) time series is the starting point for further statistical analysis. A first issue to check is whether the time series is stationary, that is, for which the statistical quantities such as mean and variance do not explicitly depend on time. One method to test this is to divide the time series in several blocks and perform the analysis of mean and variance on each block separately. In the following, we assume that we have a stationary time series.

Given a time series $X_t, t = \Delta t, \ldots, i\Delta t, \ldots, N\Delta t$, or for convenience $X_i, i = 1, \ldots, N$, the so-called estimators (denoted by a quantity with a hat) for the mean and variance are given by

$$\hat{\mu}_X = \frac{1}{N}\sum_{i=1}^{N} X_i, \tag{5.6a}$$

$$\hat{\sigma}_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2. \tag{5.6b}$$

We can immediately compute the mean and variance of these estimators when it is assumed that every X_i is from a Gaussian distribution $N(\mu, \sigma^2)$ and that they are independent. In this case, we find for the mean, using the expectation E and variance

Var operators,

$$E[\hat{\mu}_X] = \mu, \tag{5.7a}$$

$$Var[\hat{\mu}_X] = \frac{\sigma^2}{N}. \tag{5.7b}$$

To determine $E[\hat{\sigma}_X^2]$, we first consider

$$\begin{aligned}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2 &= \sum_{i=1}^{N}(X_i - \mu + \mu - \hat{\mu}_X)^2 \\ &= \sum_{i=1}^{N}(X_i - \mu)^2 + 2\sum_{i=1}^{N}(X_i - \mu)(\mu - \hat{\mu}_X) + N(\mu - \hat{\mu}_X)^2 \\ &= \sum_{i=1}^{N}(X_i - \mu)^2 - N(\hat{\mu}_X - \mu)^2,\end{aligned}$$

and hence

$$\begin{aligned}E\left[\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \hat{\mu}_X)^2\right] &= \frac{1}{N-1}\sum_{i=1}^{N}E[(X_i - \hat{\mu}_X)^2] \\ &= \frac{1}{N-1}\sum_{i=1}^{N}E[(X_i - \mu)^2] - \frac{N}{N-1}E[(\hat{\mu}_X - \mu)^2],\end{aligned}$$

and using (5.7b), we find

$$E[\hat{\sigma}_X^2] = \frac{N}{N-1}\sigma^2 - \frac{1}{N-1}\sigma^2 = \sigma^2.$$

From a similar calculation for the variance of the estimator $\hat{\sigma}_X^2$, it is found that

$$E[\hat{\sigma}_X^2] = \sigma^2, \tag{5.8a}$$

$$Var[\hat{\sigma}_X^2] = \frac{2}{N-1}\sigma^4. \tag{5.8b}$$

Estimators preferably should be unbiased for a Gaussian time series. If $\hat{\alpha}$ is an estimator of a quantity α, then the bias $B(\hat{\alpha})$ of $\hat{\alpha}$ is defined as

$$B(\hat{\alpha}) = E[\hat{\alpha}] - \alpha. \tag{5.9}$$

Suppose that every X_i is from a Gaussian distribution $N(\mu, \sigma^2)$ and that they are independent; we then compute the bias of the estimators (5.6) as

$$B(\hat{\mu}_X) = E[\hat{\mu}_X] - \mu = 0, \tag{5.10a}$$

$$B(\hat{\sigma}_X^2) = E[\hat{\sigma}_X^2] - \sigma^2 = 0, \tag{5.10b}$$

and hence both are unbiased.

Once the estimator of the mean $\hat{\mu}_X$ is determined, one usually proceeds to subtract this mean from the signal (if only interested in the variability). The next phase is to analyse the degree of correlation in the time resulting series. When we anticipate a periodic signal, the time series will repeat itself after some lag k (which corresponds to a period $k\Delta t$). To determine this correlation, one determines estimators of the covariance coefficients $\hat{c}_k$ with

$$\hat{c}_k = \frac{1}{N}\sum_{i=1}^{N-k}(X_i - \hat{\mu}_X)(X_{i+k} - \hat{\mu}_X), \tag{5.11}$$

and correlation coefficients $\hat{r}_k$ defined by

$$\hat{r}_k = \frac{\hat{c}_k}{\hat{c}_0}. \tag{5.12}$$

It is instructive to inspect the plot of $\hat{r}_k$ versus k. For a random Gaussian process, we expect $\hat{r}_k = 0$ for every $k \neq 0$, whereas $\hat{r}_k$ will oscillate in k when the time series has an oscillatory component. A plot of the function $\hat{r}_k$ versus k for the NINO3 time series in Fig. 5.1c (restricted to the years 1950–2000) is plotted in Fig. 5.2a. Here one can see that there is a dominant periodic signal (with a period of about 50 months) in this time series.

5.1.3 Spectral analysis

Spectral analysis is used to determine dominant frequencies in time series. A most obvious method would be to compute the discrete Fourier transform of the covariance coefficients $\hat{c}_k$. However, it turns out that this is no unbiased estimator of the spectrum and that windowing methods have to be used. The Fourier transform actually used is

$$\hat{f}(\omega) = \frac{1}{\pi}\left[\lambda_0\hat{c}_0 + 2\sum_{k=1}^{M}\lambda_k\hat{c}_k\cos\omega k\right], \tag{5.13}$$

where ω is the frequency and the λ_k are windowing coefficients. If the sampling time is Δt, the largest frequency that can be resolved is the Nyquist frequency $\omega = \pi/\Delta t$ (or the period $\tau = 2\Delta t$), and hence we can take $\omega_j = (j\pi/M\Delta t)$, $j = 1, \ldots, M$ in (5.13).

Several windows are in use and give biased estimators of the amplitude of the spectral coefficients $|\hat{f}(\omega)|$. In the so-called Tukey window, the coefficients λ_k are taken as

$$\lambda_k = \frac{1}{2}\left(1 + \cos\frac{\pi k}{M}\right), \quad k = 0, \ldots, M. \tag{5.14}$$

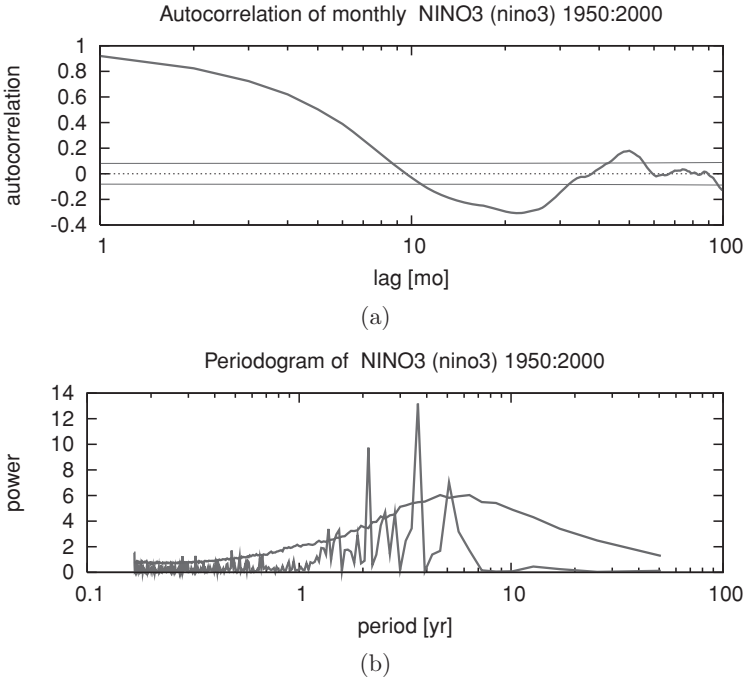


Figure 5.2 (a) Autocorrelation function of the NINO3 time series over the period 1950–2000. The horizontal lines give the 95% significance for a single point in the case of white noise, assuming all measurements are independent. (b) Fourier spectrum (periodogram, light curve) of the NINO3 time series in Fig. 5.1c over the period 1950–2000. The dark curve denotes the 95% highest spectrum of 2669 AR(1) processes with the same autocorrelation (0.939). The most significant peaks are at 2.12 and 3.64 years.

In the Parzen window, the coefficients are chosen as

$$\lambda_k = 1 - 6\left(\frac{k}{M}\right)^2 + 6\left(\frac{k}{M}\right)^3, 0 \le k \le M/2, \tag{5.15a}$$

$$\lambda_k = 2 - 6\left(1 - \frac{k}{M}\right)^3, \quad M/2 \le k \le M. \tag{5.15b}$$

A 'rule of thumb' is the choice $M \approx 2\sqrt{N}$. A spectrum (without window, this is a periodogram) for the NINO3 time series is shown in Fig. 5.2b. The question now arises: are the peaks in this spectrum statistically significant? In other words, can these peaks also be produced by just some noise process?

5.2 Hypothesis testing and significance

The general idea behind hypothesis testing is that we pose some null hypothesis that we try to reject using either a statistical test variable or a computed distribution

obtained through (e.g., Monte Carlo) simulation. We illustrate these procedures with two examples.

5.2.1 Test variables

Suppose we have two time series $X_i, i = 1, \ldots, N_x$ and $Y_j, j = 1, \ldots, N_y$, and we want to deduce the relation between the mean μ_X and μ_Y of these two time series. The null hypothesis H_0 is

$$H_0 : \mu_X = \mu_Y. \tag{5.16}$$

The optimal test variable for this null hypothesis is given by (Von Storch and Zwiers, 1999)

$$\hat{T} = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\hat{S}_p \sqrt{\frac{1}{N_x} + \frac{1}{N_y}}}, \tag{5.17}$$

where

$$\hat{S}_p^2 = \frac{\sum_{i=1}^{N_x}(X_i - \hat{\mu}_X)^2 + \sum_{j=1}^{N_y}(Y_j - \hat{\mu}_Y)^2}{N_x + N_y - 2}$$

and $\hat{\mu}_X$ and $\hat{\mu}_Y$ are determined from (5.6a).

Consider the following assumptions:

(i) Every realization of X is independent of every other realization. Same for Y.
(ii) The distribution that generates X_i is the same for every i. Same for Y.
(iii) The distributions of X and Y are both normal.

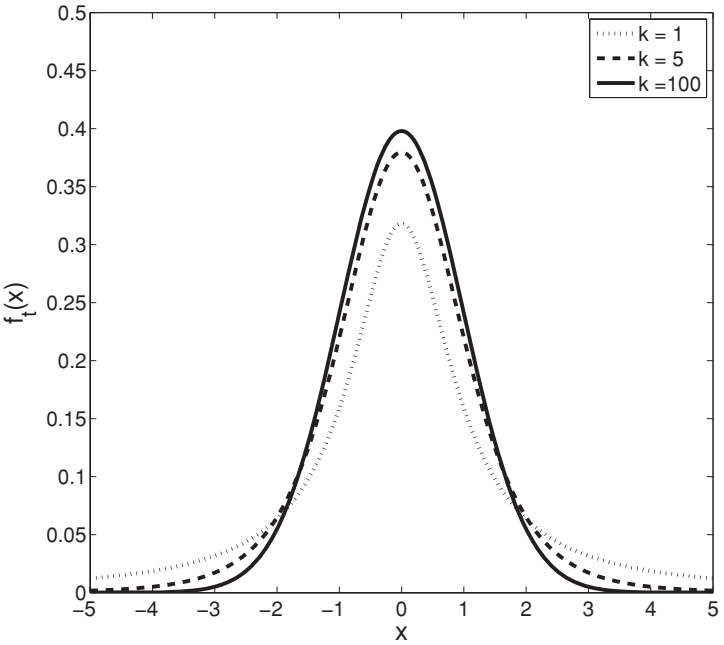
Under these assumptions, the test variable $\hat{T}$ in (5.17) has a so-called t-distribution with $k = N_x + N_y - 2$ degrees of freedom, given by

$$f_t(x) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)}, \tag{5.18}$$

where Γ, the Gamma function, is defined by

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt. \tag{5.19}$$

The probability density function $f_t(x)$ is plotted for different k in Fig. 5.3a. In Fig. 5.3b, df indicates the number of degrees of freedom, and the values in the table indicate



(a)

	$\tilde{p}$						
df	0.750	0.900	0.950	0.975	0.990	0.995	0.999
1	1.000	3.078	6.314	12.706	31.821	63.657	318.313
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552

(b)

Figure 5.3 (a) Probability density of the t-distribution, with k being the number of degrees of freedom. (b) Values of T_0 for different degrees of freedom ($df = k$) and significance levels ($\tilde{p}$).

the value of $\hat{T}$, say T_0, for which $P(\hat{T} > T_0) = 1 - \tilde{p}$, for a specific significance level $\tilde{p}$. The test procedure is now as follows:

(i) Choose a significance level $\tilde{p}$.
(ii) Compute the number of degrees of freedom $N_x + N_y - 2$ and the test variable $\hat{T}$ from (5.17).
(iii) Determine from the table the value of T_0 at the particular value of $\tilde{p}$.
(iv) If $\hat{T} \geq T_0$, then the null hypothesis can be rejected at the $100\tilde{p}\%$ significance level. If $\hat{T} < T_0$, then the null hypothesis cannot be rejected at that significance level.

5.2.2 Stochastic simulation

In many situations, it is difficult to determine test variables such as $\hat{T}$ in (5.17). In that case, we can sometimes choose a stochastic model, representing our knowledge of the noise, and use the sample paths of this model to test the null hypothesis H_0. As an example, consider the significance of a peak of the spectrum from a certain time series X. We want to be sure that this peak cannot be caused by a certain stochastic process, which we think is a good model for the noise. There are many discrete stochastic models around, for example, white noise, random walk (Wiener process), red noise, autoregressive (AR) processes, moving average (MA) processes and combinations of the latter two processes (ARMA). For easy reference, we list two of these processes and their properties.

(i) White noise. In this case,

$$X_t = Z_t, \tag{5.20}$$

where Z_t are independent identically distributed Gaussian $N(0, \sigma)$ stochastic variables. For the autocovariance, we find directly from (5.11) that

$$\hat{c}_k = 0, k = \pm 1, \pm 2, \ldots, \tag{5.21}$$

and the autocorrelation $\rho_k = 1$ for $k = 0$ and zero otherwise.

(ii) Red noise. This is a so-called autoregressive process of order one (also abbreviated as AR(1)), defined by

$$X_t = \alpha X_{t-1} + Z_t, \tag{5.22}$$

where $\alpha \in (0, 1)$ and Z_t has a $N(0, \sigma^2)$ distribution. To compute the autocovariance and autocorrelation functions, we realise that

$$X_t = \alpha(\alpha X_{t-2} + Z_{t-1}) + Z_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \cdots,$$

and hence

$$E[X_t] = 0\ ;\ Var[X_t] = \sigma^2(1 + \alpha^2 + \alpha^4 + \cdots) = \frac{\sigma^2}{1 - \alpha^2}, \tag{5.23}$$

and, finally,

$$\hat{c}_k = \sigma^2 \alpha^k \sum_{i=0}^{\infty} (\alpha^i)^2 = \frac{\alpha^k \sigma^2}{1-\alpha^2}, \tag{5.24a}$$

$$\rho_k = \alpha^k. \tag{5.24b}$$

Once the discrete stochastic model is chosen, say as AR(1), the procedure to compute significance levels is as follows:

(i) Run many realizations of the AR(1) process with the same time step as that in the data. Choose α such that the autocorrelation function ρ_k has the same decay as that of the data.
(ii) Determine spectra for each realization and compute the distribution function of the spectra.
(iii) Choose a significance level $\tilde{p}$ and plot the spectrum that has a probability $1 - \tilde{p}$. The peaks that extend above the latter spectrum are significant at the $100\tilde{p}\%$ level.

The spectrum at the 95% ($\tilde{p} = 0.95$) significance level of an AR(1) process is plotted for the NINO3 time series as the dark curve in Fig. 5.2b. Here 2,669 AR(1) sample paths were computed.

5.3 Bivariate and multivariate analysis

In this section, the results of the univariate analysis are extended to study statistical relations between two time series or between time series of fields, such as spatial fields of climate variables.

5.3.1 Cross-covariance and co-spectra

Suppose we have two time series, for example, of NINO3 (Fig. 5.1) and of the global mean temperature (Fig. 1.3), and ask the question whether there is any causal relation between the data in these series. In the time domain, we are hence interested in the cross-covariance of both time series. When the time series are denoted by $X_i, i = 1, \ldots, N_x$ and $Y_j, j = 1, \ldots, N_y$, the cross-covariance coefficients with lag k then are defined as (for convenience, we assume that $N_x = N_y = N$)

$$k = 0, \ldots, N-1 : \hat{c}_{xy}(k) = \frac{1}{N} \sum_{i=1}^{N-k} (X_i - \hat{\mu}_X)(Y_{i+k} - \hat{\mu}_Y), \tag{5.25a}$$

$$k = -(N-1), \ldots, -1 : \hat{c}_{xy}(k) = \frac{1}{N} \sum_{i=1-k}^{N} (X_i - \hat{\mu}_X)(Y_{i+k} - \hat{\mu}_Y), \tag{5.25b}$$

and the cross-correlation coefficients are defined by

$$\hat{r}_{xy}(k) = \frac{\hat{c}_{xy}(k)}{\sqrt{\hat{c}_{xx}(0)\hat{c}_{yy}(0)}}. \tag{5.26}$$

In addition, the co-spectrum $\hat{c}$ and the quadrature spectrum $\hat{q}$ are defined as

$$\hat{c}(\omega) = \frac{1}{\pi}\left[\sum_{k=-M}^{M} \lambda_k \hat{c}_{xy}(k)\cos\omega k\right], \tag{5.27a}$$

$$\hat{q}(\omega) = \frac{1}{\pi}\left[\sum_{k=-M}^{M} \lambda_k \hat{c}_{xy}(k)\sin\omega k\right], \tag{5.27b}$$

and the cross-amplitude, phase and coherency spectra follow from

$$\hat{\alpha}(\omega) = \sqrt{\hat{c}^2(\omega) + \hat{q}^2(\omega)}, \tag{5.28a}$$

$$\tan\hat{\phi}(\omega) = -\frac{\hat{q}(\omega)}{\hat{c}(\omega)}, \tag{5.28b}$$

$$\hat{C}(\omega) = \frac{\hat{\alpha}^2(\omega)}{\hat{f}_x(\omega)\hat{f}_y(\omega)}, \tag{5.28c}$$

where $\hat{f}_x$ and $\hat{f}_y$ are estimates of the individual power spectra of X and Y.

For fields of variables, such relations can be determined between quantities of two different locations. The analysis can be easily extended to deal with correlations in the whole field simultaneously, as we see in the next section.

5.3.2 Empirical orthogonal functions

Suppose we now have a vector time series $\mathbf{X}^k, k = 1, \ldots, N$ where each $\mathbf{X}^k$ has a dimension d. Such vector time series arise naturally from grid point values of a climate quantity, such as the monthly mean sea-surface temperature T. For example, if we sample T at a $I \times J$ grid for N years, then $d = I \times J$, and each vector $\mathbf{X}^k = (T^k_{1,1}, \ldots, T^k_{I,J})$ for $k = 1, \ldots, N$. Just as in the bivariate case, we would like to analyse the covariance between the time series at each grid point.

The bivariate approach to calculate cross-covariance coefficients is easily generalised. We first estimate elements of the covariance matrix $\hat{\Sigma}$, with

$$\hat{\Sigma} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{X}^k - \hat{\mu})(\mathbf{X}^k - \hat{\mu})^T, \tag{5.29}$$

where the vector mean $\hat{\mu}$ is determined from

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{X}^k, \tag{5.30}$$

and hence an element of $\hat{\Sigma}$, say $\hat{\Sigma}_{ij}$, is given by

$$\hat{\Sigma}_{ij} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{X}_i^k - \hat{\mu}_i)(\mathbf{X}_j^k - \hat{\mu}_j), \tag{5.31}$$

which is a similar expression as for the cross-covariance coefficients (5.25).

As the $d \times d$ real matrix $\hat{\Sigma}$ is symmetric and positive definite, it has real positive eigenvalues $\hat{\lambda}_i$, $i = 1, \ldots, d$. Let these eigenvalues be ordered by $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_d$ and the corresponding eigenvectors be indicated by $\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_d$. These eigenvectors are called the empirical orthogonal functions (EOFs). As the EOFs form an orthonormal basis in $\mathbb{R}^d$, every vector $\mathbf{X}^k$ can be represented as

$$\mathbf{X}^k = \sum_{i=1}^{d} \hat{\alpha}_i^k \hat{\mathbf{e}}_i. \tag{5.32}$$

For fixed i, the coefficient $\hat{\alpha}_i^k$ is called the principal component (PC) corresponding to the ith EOF. It can be shown (Von Storch and Zwiers, 1999) that

$$\hat{\alpha}_i^k = < \mathbf{X}^k, \hat{\mathbf{e}}_i >, \tag{5.33}$$

with $<, >$ the inner product in $\mathbb{R}^d$ minimizes the error

$$\hat{\epsilon}_M = \sum_{k=1}^{N} |\mathbf{X}^k - \sum_{i=1}^{M} \hat{\alpha}_i^k \hat{\mathbf{e}}_i|^2, \tag{5.34}$$

for every M. This means that the EOFs are optimal in representing the variance in the stochastic process $\mathbf{X}^1, \ldots, \mathbf{X}^N$. Note that each EOF represents a spatial pattern on the grid $I \times J$ if each $\mathbf{X}^k$ is derived from a grid-point representation of a continuous variable.

If for fixed i, $\hat{\alpha}_i$ indicates the stochastic variable associated with the time series $\hat{\alpha}_i^k$, $k = 1, \ldots, N$, then it can also be shown that for each i,

$$Var[\hat{\alpha}_i] = \hat{\lambda}_i. \tag{5.35}$$

The variance $\hat{\sigma}_i^2$ accounted for by the ith EOF when a total of M EOFs are used to represent the signal is usually written as

$$\hat{\sigma}_i^2 = \frac{\hat{\lambda}_i}{\sum_{j=1}^{M} \hat{\lambda}_j} \times 100\%. \tag{5.36}$$

In many practical problems, a large part of the variance can already be accounted for by only a few EOFs, which makes this method an attractive approach to data reduction.

Table 5.1. *First column: first five eigenvalues $\hat{\lambda}_i$ of the covariance matrix for the SST data over the period January 1950 to December 2000 from the HadISST1 data set. Second column: explained variance of the EOF. Third column: cumulative explained variance*

# EOF	Eigenvalue	Explained variance (%)	Cumulative (%)
1	1078.8	55.44	55.44
2	228.91	11.76	67.20
3	98.110	5.04	72.24
4	69.068	3.55	75.79
5	55.494	2.85	78.65

Measures of uncertainty in the EOFs $\hat{\mathbf{e}}_i$ and eigenvalues $\hat{\lambda}_i$ are given (North et al., 1982) by

$$\Delta\hat{\lambda}_i = \sqrt{\frac{2}{N^*}}\hat{\lambda}_i \; ; \; \Delta\hat{\mathbf{e}}_i = \frac{\Delta\hat{\lambda}_i}{\hat{\lambda}_j - \hat{\lambda}_i}\hat{\mathbf{e}}_i, \tag{5.37}$$

where $\hat{\lambda}_j$ is the closest eigenvalue to $\hat{\lambda}_i$ and N^* is the number of independent observations in the sample. This indicates that one has to be careful if the variance explained by two EOFs is nearly equal.

Example 5.1 EOFs of Pacific SST Consider monthly mean data of the sea-surface temperature (SST) in the equatorial Pacific ([120°E–70°W] × [20°S–20°N]) over the period January 1950 to December 2000 from the HadISST1 data set (Rayner et al., 2003). These are available on a $1^\circ \times 1^\circ$ grid, but we will compute the EOFs for data on a $2^\circ \times 2^\circ$ grid. In this case, $N = 12 \times 51 = 612$ and $d = 170/2 \times 40/2 = 85 \times 20 = 1{,}700$. The seasonal cycle is calculated by averaging over all January months, February months, and so forth, and then subtracting from the original data; the resulting vector time series consists only of SST anomalies.

The eigenvalues of the covariance matrix are shown in Table 5.1. The first two EOFs already explain about 67% of the variance in the data and are very dominant compared with the rest of the EOFs. The uncertainty in the eigenvalues can be computed from (5.37) with $N^* \approx N$. The pattern of the first EOF (Fig. 5.4a) is the well-known ENSO pattern with large amplitudes of the SST variability in the eastern Pacific. The SST variability associated with the second EOF is mainly confined to the coast of South America. The time series of the first two principal components (Fig. 5.5) show the characteristic interannual variability of SST in the Eastern Equatorial Pacific; PC1 is well correlated with NINO3. ■

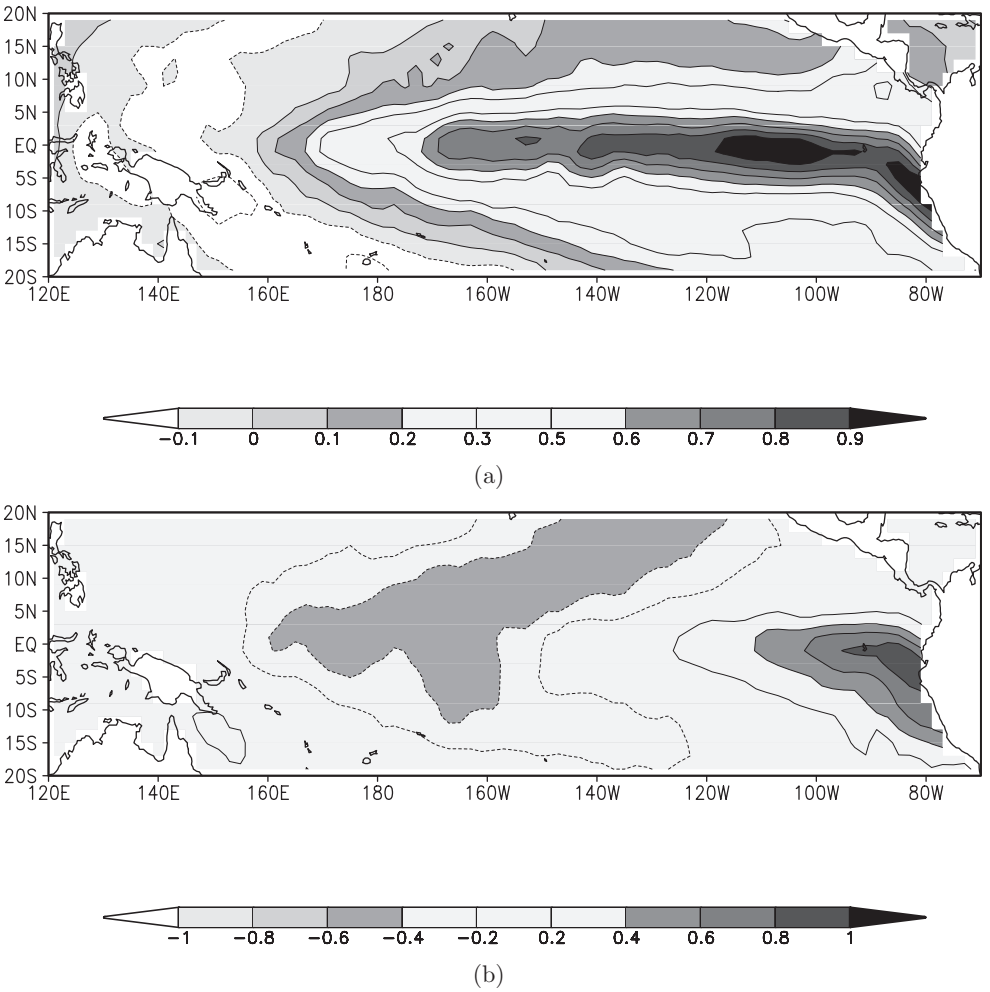
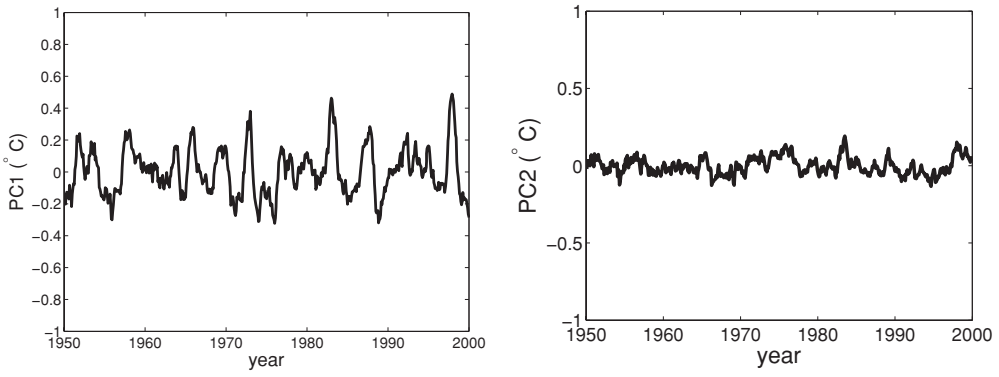


Figure 5.4 The first (a) and the second (b) EOF of equatorial SST from the HadISST1 data set over the period 1950–2000.

5.3.3 Singular spectrum analysis

In recent years, a combination of time series analysis and dynamical systems ideas has led to spectral techniques by which important information (such as dominant frequencies and patterns of variability) can be extracted from short and noisy time series (Ghil et al., 2002). Singular spectrum analysis (SSA) allows one to unravel the information embedded in a delay-coordinate state space (as described next) by decomposing the sequence of augmented vectors thus obtained into elementary patterns of behaviour. It does so by providing data-adaptive filters that help separate the time series into components that are statistically independent, at zero lag, in the augmented vector space of interest. These components can be classified essentially into trends,



(a) (b)

Figure 5.5 Time series of the leading two PCs corresponding to the leading two EOFs, (a) PC1 and (b) PC2.

oscillatory patterns and noise. It is an important feature of SSA that the trends need not be linear and that the oscillations can be amplitude and phase modulated.

The starting point of SSA is to embed a time series $X^k, k = 1, \ldots, N$ in a vector space of dimension M, that is, to represent it as a trajectory in the phase space of the hypothetical system that generated X^k. In concrete terms, this is equivalent to representing the behavior of the system by a succession of overlapping views of the series through a sliding M point window.

The embedding procedure applied so constructs a sequence $\tilde{\mathbf{X}}^k$ of M-dimensional vectors from the original time series, by using lagged copies of the scalar data X^k, as follows:

$$\tilde{\mathbf{X}}^l = (X^l, X^{l+1}, \ldots, X^{l+M-1}), \tag{5.38}$$

where the vectors $\tilde{\mathbf{X}}^l$ are indexed by $l = 1, \ldots, N'$, where $N' = N - M + 1$. The $M \times M$ covariance matrix C_X of the vector time series (5.38) is a so-called Toeplitz matrix with elements

$$\hat{c}_{ij} = \frac{1}{N - |i - j|} \sum_{l=1}^{N-|i-j|} X^l X^{l+|i-j|}, \tag{5.39}$$

for $i = 1, \ldots, M$ and $j = 1, \ldots, M$.

In SSA, the eigenvalues λ_m and eigenvectors ρ_m, $m = 1, \ldots, M$ of the matrix C_X are determined by solving the problem

$$C_X \rho_m = \lambda_m \rho_m. \tag{5.40}$$

Because the ρ_m are also eigenvectors of a covariance matrix, they are indicated as t-EOFs and the corresponding PCs as t-PCs.

Example 5.2 SSA of NINO3 For the NINO3 (1950–2000) time series in Fig. 5.1c, as derived from the HadISST1 data set (Rayner et al., 2003), the eigenvalues λ_m are plotted in Fig. 5.6a; here $N = 612$ and $M = 80$. The first two eigenvalues are paired and indicate an oscillatory signal in the time series. Statistical tests are available (Ghil et al., 2012) to determine the significance of this pair (e.g., under a red-noise hypothesis). The error bars on the eigenvalues are computed here by $\sqrt{2/N}\lambda_m$.

The first two t-EOFs and t-PCs are shown in Fig. 5.6b–e, where the $N - M$ number t-PCs A_m are calculated from

$$A_m^l = \sum\nolimits_{j=1}^{M} X^{l+j-1} \rho_m^j \tag{5.41}$$

and clearly indicate that t-PC1 and t-PC2 are in quadrature (they are shifted by a quarter period, here about 10 months).

As the dominant signal seems to be represented by the first few t-EOFs, we can restrict the signal to a small set of t-EOFs and obtain the reconstructed time series (with a length N). Using six tEOFs, the reconstructed component of the NINO3 time series is plotted in Fig. 5.7a, together with the original time series. A spectrum of the reconstructed component (Fig. 5.7b) now shows two spectral peaks at frequencies 0.019/month (period 52 months) and 0.034/month (period 29 months). ■

Multichannel singular spectral analysis (M-SSA) (Plaut and Vautard, 1994) is a direct extension of the SSA technique, which aims to extract propagating patterns that are optimal in representing variance. M-SSA is mathematically equivalent to extended EOF analysis (EEOF) (Weare and Nasstrom, 1982), but in M-SSA, focus is on the temporal structure of the variability, whereas in EEOF, the spatial variability is emphasized.

Let a data set $\mathbf{X}$ consist of a multichannel time series $X_l^i,\ i = 1, \ldots, N;\ l = 1, \ldots, L$, where i represents time and l the channel number. Index l may represent a point number on a specific grid or a principal component (PC) if the data are prefiltered with principal component analysis (Preisendorfer, 1988). We assume that $\mathbf{X}$ has zero mean and is stationary. By making M lagged copies of $\mathbf{X}$, the state vector at time i is given by

$$(X_1^{i+1}, X_1^{i+2}, \ldots, X_1^{i+M}, \ldots, X_L^{i+1}, \ldots, X_L^{i+M})^T, \tag{5.42}$$

where M is the window length. The covariance matrix T for a chosen window length M has a general block-Toeplitz form in which each block $\mathsf{T}_{ll'}$ is the lag covariance matrix (with maximum lag M) between channel l and channel l'. The $L \times M$ real eigenvalues λ_k of the symmetric matrix T are sorted in decreasing order where an eigenvector (referred to as a ST-EOF) $\mathbf{E}^k$ is associated with the k-th eigenvalue λ_k. The $\mathbf{E}^k$ are M-long time sequences of vectors, describing space-time patterns of decreasing importance as their order k increases. A space-time principal component (referred to as a ST-PC) α^k can be computed by projecting $\mathbf{X}$ onto $\mathbf{E}^k$; λ_k is the

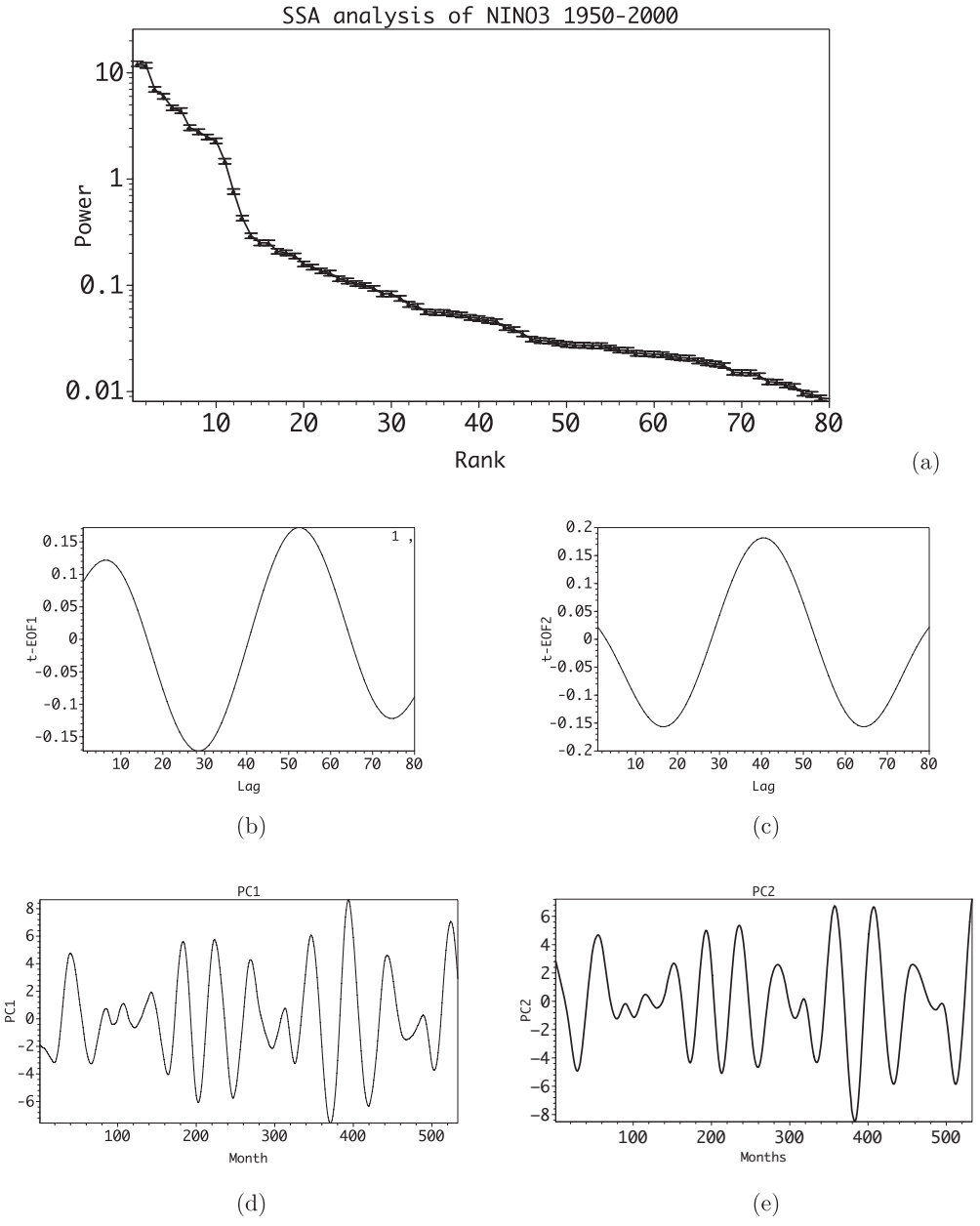
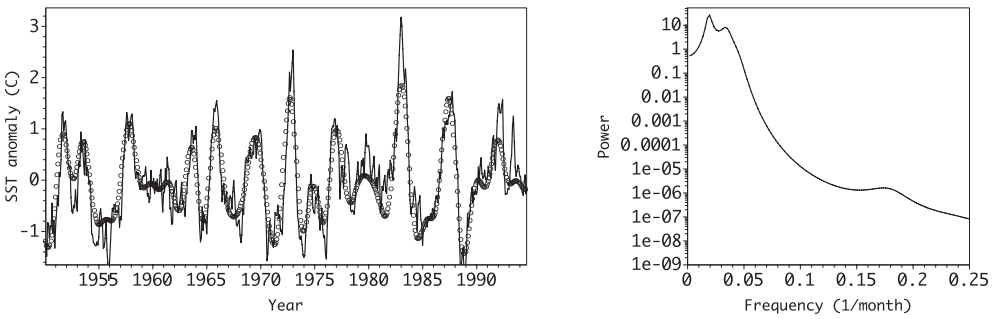


Figure 5.6 (a) The eigenvalues of the problem (5.40) for the NINO3 time series over the period 1950–2000 with a window $M = 80$. (b–c) The two leading t-EOFs and (d–e) the corresponding t-PCs.



(a) (b)

Figure 5.7 (a) Reconstructed component (using the first six t-EOFs) for the NINO3 time series over the period 1950–2000 with a window $M = 80$. (b) Maximum entropy spectrum of the time series in (a) with order 20; this spectrum was made using the software package kSpectra (http://www.spectraworks.com/).

variance in α^k. In this way, the M-SSA expansion of the original data series is given by

$$X_l^{i+j} = \sum_{k=1}^{L\times M} \alpha_i^k \mathbf{E}_{lj}^k, \quad j = 1, \ldots, M. \tag{5.43}$$

Principal component analysis and SSA (Vautard and Ghil, 1989; Vautard et al., 1992) are particular cases of M-SSA: PCA can be derived from M-SSA with $M = 1$ and from SSA with $L = 1$.

When two consecutive eigenvalues are nearly equal and the two corresponding $\mathbf{E}^k$ as well as the associated α^k are in quadrature, then the data possess an oscillation whose period is given by that of α^k and whose spatial pattern is that of $\mathbf{E}^k$ (Plaut and Vautard, 1994). The sum on the right-hand side of (5.43), restricted to one or several terms, describes the part of the signal behaving as the corresponding $\mathbf{E}^k$. The components constructed in this way are called reconstructed components (RCs) and the part of the signal involved with an oscillation can be isolated. The original signal is exactly the sum of all the RCs.

As Allen and Robertson (1996) have pointed out, the presence of an eigenvalue pair in M-SSA is not sufficient grounds to conclude that the data exhibit an oscillation. Moreover, low-frequency eigenvalue pairs, which are entirely due to red noise, will appear high in the eigenvalue rank order. A Monte Carlo red-noise significance test for M-SSA was therefore constructed (Allen and Robertson, 1996). This is an objective hypothesis test for the presence of oscillations at low signal-to-noise ratios in multivariate data. Rejection of the red-noise null hypothesis using the test should be considered a necessary condition for M-SSA to have detected an oscillation, although in certain situations, non-oscillatory processes might also lead to rejection.

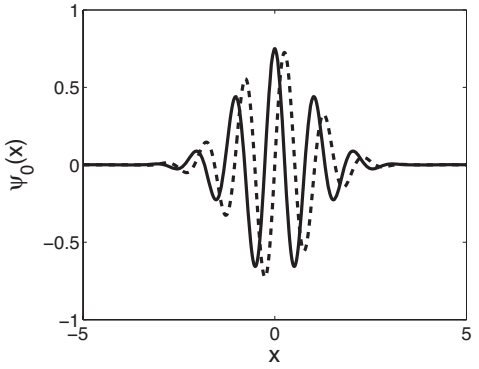


Figure 5.8 Plot of the real (drawn) and imaginary (dashed) part of the Morlet wavelet (5.44) with $\omega_0 = 6$.

5.4 Nonstationary methods: wavelets

Many time series are nonstationary, with, for example, the mean and variance changing in time. For the analysis of these time series, so-called wavelet techniques have been developed; with these techniques, the changes in the amplitude and period of an oscillation with time can be determined. The wavelet basis differs from the Fourier basis in that the functions have limited extent and can be stretched and translated. As an example, consider the much used Morlet wavelet, where the mother wavelet is given by

$$\psi_0(\eta) = \pi^{-\frac{1}{4}} e^{i\omega_0\eta} e^{-\frac{\eta^2}{2}}. \tag{5.44}$$

The function ψ_0 is a complex valued function of which the real and imaginary parts are plotted in Fig. 5.8 for $\omega_0 = 6$. The independent quantity η will be used to stretch and translate the waveform as plotted in Fig. 5.8.

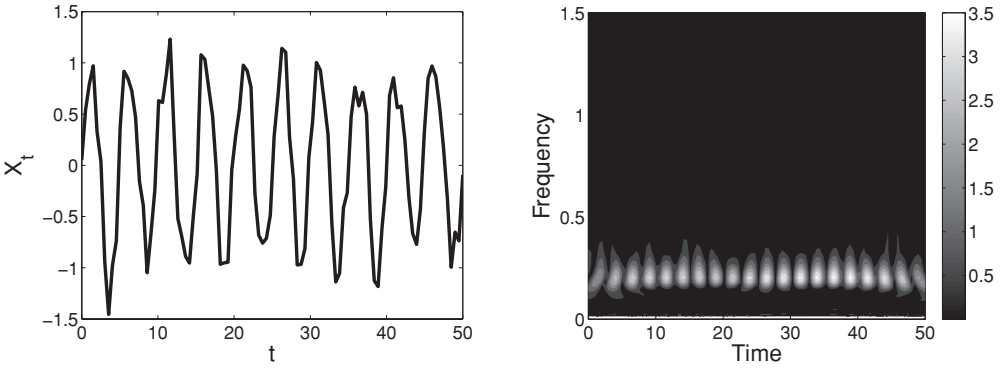
After choosing a certain mother wavelet, a range of scales must be selected to adequately sample the time series. For example, if the data are sampled with a time interval (Δt), then the smallest period that can be resolved (the Nyquist period) is $2\Delta t$, which is usually chosen as the scale s_0. The larger scales (longer periods) are then chosen as

$$s_j = 2^j s_0, \quad j = 1, \ldots, J, \tag{5.45}$$

where J is determined such that the largest period is less than half the length of the time series.

Once these scales are chosen, the wavelet transform $W_n(s)$ of the time series $X_k, k = 1, \ldots, N$ is calculated (for each $s = s_j$) from

$$W_n(s) = \sqrt{\frac{\Delta t}{s}} \sum_{l=1}^{N} X_l \, \psi_0^{\dagger}\left(\frac{l-n}{s}\Delta t\right), \tag{5.46}$$



(a) (b)

Figure 5.9 (a) Plot of the simple time series (5.47). (b) Wavelet spectrum of the time series in (a).

where the † superscript indicates the complex conjugate. Hence $n = 1, \ldots, N$ serves as a translation and s as the stretching (determining the scale) parameter. While 'moving over' the time series, the wavelet 'senses' the power $|W_n(s)|^2$ at the scales s.

Example 5.3 Examples of wavelet transforms As an example (Trauth, 2007), consider the pure sine wave with a period 5 with added Gaussian noise, defined by

$$X(t) = \sin\left(\frac{2\pi t}{5}\right) + \epsilon\zeta(t), \tag{5.47}$$

where ζ has the standard normal distribution $N(0, 1)$. The (stationary) time series (for $N = 100$, so with $\Delta t = 1/2$ and $s_0 = 1$) is plotted in Fig. 5.9a for $\epsilon = 0.2$. We use the Morlet wavelet over scales $s_j = 2^j$ and compute the wavelet spectrum (using MATLAB) as shown in Fig. 5.9b. Indeed, over the whole time period, the frequency 0.2 is the dominant one.

An example from climate data is that of the NINO3 time series as plotted in Fig. 5.1c. Here the number of data points $N = 612$ and $\Delta t = 1/12$ years. Hence $s_0 = 1/6$ years, and scales up to 26 years are considered. The wavelet spectrum is plotted in Fig. 5.10. The drawn contour in Fig. 5.10 indicates the region of frequencies where end effects (due to the finite length of the time series) are deteriorating the spectral analysis; this contour itself is the so-called cone of influence. ■

5.5 Nonlinear methods

Typical time series from observations of climate variables result from nonlinear deterministic processes (usually at large scales) combined with noise (usually from the small scales). Nonlinear deterministic processes may give rise to chaotic dynamics

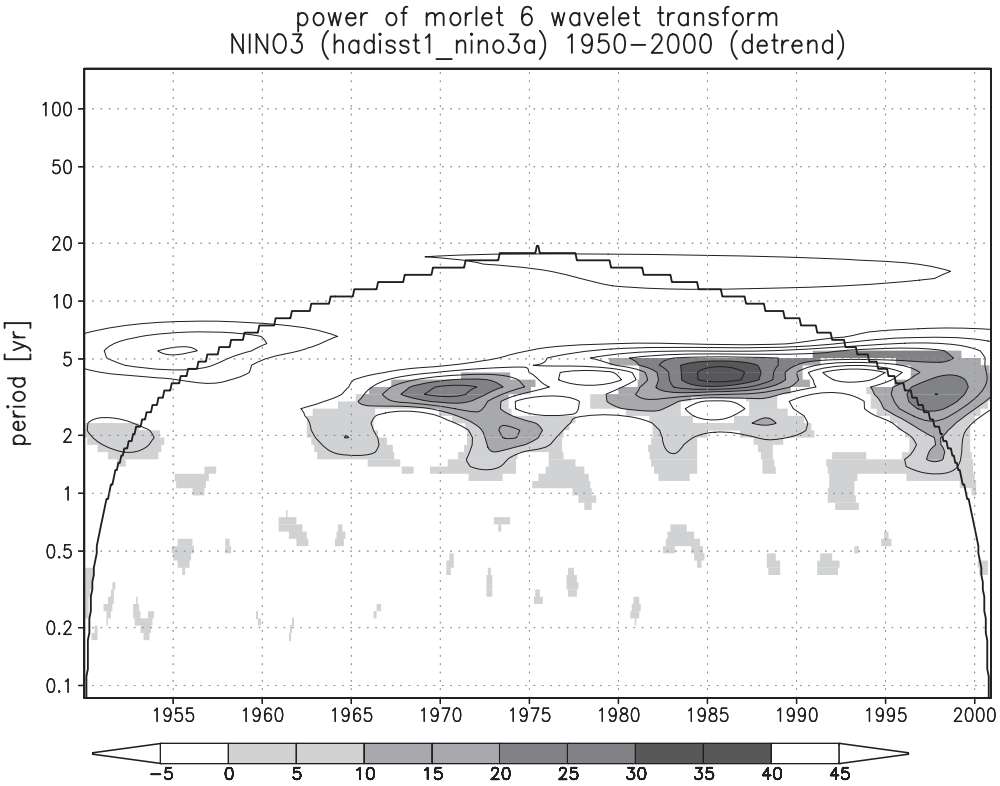


Figure 5.10 Wavelet spectrum of the NINO3 time series in Fig. 5.1c showing contour values of the power $|W_n(s)|^2$ (figure made with the Climate Explorer, http://climexp.knmi.nl).

with extremely complicated time series. In this section, an introduction will be given into techniques to analyze such time series, following Abarbanel (1996). Computations can be done, for example, by using the software package TISEAN, which can be downloaded at http://www.mpipks-dresden.mpg.de/~tisean/.

5.5.1 Attractor embedding

Let $\mathbf{X}_k, k = 1, \ldots, n$ indicate a vector of climate variables with sampling times indicated by $t_k = k\tau_s$. Assume that we observe only a single time series from this climate variable, that is,

$$s_k = s(t_k) = h(\mathbf{X}_k), \tag{5.48}$$

where h acts as an observation mapping. Although this seems rather limited information about the full state vector $\mathbf{X}_k$, in the early 1980s, it was realised that any subsequent element in the series s_k is related to all the other variables in the system (i.e., also those not observed) through nonlinear couplings.

All time derivatives of the underlying observable $s(t)$, for example, $\frac{ds}{dt}$, $\frac{d^2s}{dt^2}$, are also represented by the series $s(t)$ as

$$\frac{ds}{dt}(t) \approx \frac{s(t+\tau_s) - s(t)}{\tau_s}, \tag{5.49a}$$

$$\frac{d^2s}{dt^2}(t) \approx \frac{s(t+\tau_s) - 2s(t) + s(t-\tau_s)}{\tau_s^2}. \tag{5.49b}$$

A famous embedding theorem (Abarbanel, 1996) states that for a certain delay time $T = l\tau_s$ and a dimension d, the trajectories of the d-dimensional vector,

$$\mathbf{y}_k = (s(t_k), s(t_k + l\tau_s), s(t_k + 2l\tau_s), \ldots, s(t_k + l\tau_s(d-1)))^T, \tag{5.50}$$

have the 'same' behavior as that of the original state vector $\mathbf{X}_k$. The theorem does not provide a recipe for choosing the so-called embedding dimension d and the time delay T; the data have to be analysed to provide good choices.

An attractor of an autonomous dynamical system will be a geometrical set $\mathcal{A}$ (possibly with a complicated fractal structure); let a proper fractal dimension measure (such as the box-counting dimension D_0, defined in Example 2.3) indicate that the set $\mathcal{A}$ has dimension d_A. Orbits on this attractor do not have self-intersections, and hence in our reconstructed state space of dimension d, the trajectories should also not self-intersect; obviously $d > d_A$. Consider the subspace of dimension d_A within d. When two subspaces of dimension d_1 and d_2 intersect, the dimension of the intersection has a dimension $d_1 + d_2 - d$ (e.g., the intersection of two planes $[d_1 = d_2 = 2]$ in three-dimensional space $[d = 3]$ is in general a line $d_1 + d_2 - d = 1$). Hence a sufficient condition that the subset of dimension d_A does not self-intersect is that

$$d_A + d_A - d < 0 \rightarrow d > 2d_A. \tag{5.51}$$

However, this is an upper boundary for d, as we may embed the attractor without self intersections into a smaller space than that determined by $d > 2d_A$. In addition, this estimate is only useful if we have an estimate of the attractor dimension d_A.

To determine a minimal value for d, the method of false nearest neighbours (FNN) is useful. Here the distances between nearby points on the reconstructed trajectories are compared in dimension d and $d+1$. Let these points be indicated by $\mathbf{y}_k$ and $\mathbf{y}_k^{NN}$, where the subscript NN indicates nearest neighbour. We then have for their distance R^2 in dimensions d and $d+1$,

$$R_d^2(k) = \sum_{m=1}^{d} (s(t_k + (m-1)T) - s^{NN}(t_k + (m-1)T))^2, \tag{5.52a}$$

$$R_{d+1}^2(k) = R_d^2(k) + (s(t_k + dT) - s^{NN}(t_k + dT))^2, \tag{5.52b}$$

where T is again the time delay. If the relative distance error F, given by

$$F = \sqrt{\frac{R_{d+1}^2(k) - R_d^2(k)}{R_d^2(k)}} = \frac{|(s(t_k + dT) - s^{NN}(t_k + dT))|}{R_d(k)}, \tag{5.53}$$

is small, then the points are 'true' neighbours in going from dimension d to $d + 1$. However, when F is large apparently close points in dimension d are not close in dimension $d + 1$.

Example 5.4 The method of false nearest neighbours Consider the time series given by (with an amplitude $A > 0$)

$$s(t_k) = A \sin t_k, \tag{5.54}$$

for $t_k = k\tau_s \in (0, 2\pi]$ observed by monitoring the displacement component $s(t) = x(t)$ of the harmonic oscillator, which is a solution to the differential equation

$$\frac{d^2x}{dt^2} + x = 0. \tag{5.55}$$

Clearly the phase space of this deterministic model is two-dimensional, and trajectories are circles.

With embedding dimensions $d = 1$ and $d = 2$, we have the reconstructions

$$y_k = A \sin t_k, \tag{5.56a}$$

$$\mathbf{y}_k = (A \sin t_k, A \sin(t_k + T))^T, \tag{5.56b}$$

for a certain time delay T. For the case $d = 1$, the 'orbits' move in an interval, whereas in the $d = 2$ case, the orbits move on closed curves (Fig. 5.11). An example of a pair of such FNN for $d = 1$ is shown in Fig. 5.11; in this case R_1 is small, whereas R_2 is large and hence F is large. When we consider $d = 3$ for this simple example, we notice that all points are real neighbours, and hence $R_2 \approx R_3$ and F is small. The necessary embedding dimension d is hence obtained from the data at the dimension where F rapidly decreases. ■

At this point, the delay time T is still unknown. Based on classical analysis, one could base a choice of T on the autocorrelation function (e.g., take T to be the lag where the autocorrelation has the first zero), but in a chaotic system, this is usually not very revealing. Intuitively T should not be too small, because otherwise $s(t_k)$ and $s(t_k + T)$ are too much related, and the attractor is not sampled well. However, T should not be too large, because otherwise $s(t_k)$ and $s(t_k + T)$ are totally uncorrelated. To determine T, a nonlinear extension of the autocorrelation function, called the mutual information $I(T)$, is borrowed from information theory.

Consider realizations $A_i, i = 1, \ldots, N_A$ and $B_j, j = 1, \ldots, N_B$ of event A and event B, and we ask the question: what amount of bits is learned from the measurement

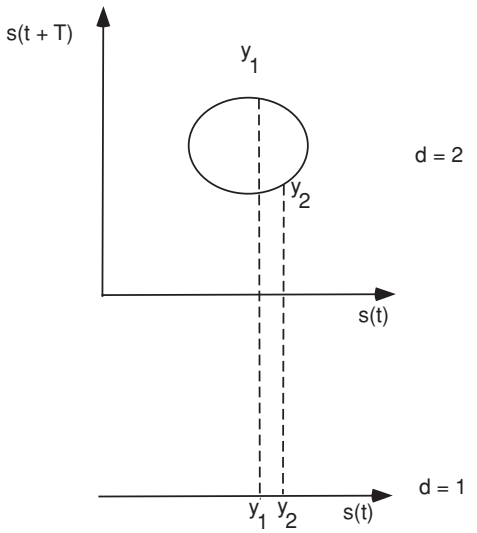


Figure 5.11 Sketch of the nearest neighbours in the reconstruction of the phase space of the harmonic oscillator.

of A_i about the measurement of B_j? From information theory, a measure of this amount of information is defined as

$$I_{AB} = \sum_{A_i, B_j} \log_2 \frac{P_{AB}(A_i, B_j)}{P_A(A_i) P_B(B_j)}, \tag{5.57}$$

where P_{AB} is the joint probability for the measurements of A and B and P_A (P_B) is the individual probability of A (B). If the A_i and B_j are uncorrelated, then $P_{AB}(A_i, B_j) = P_A(A_i) P_B(B_j)$ and $I_{AB} = 0$. The average mutual information is given by

$$\bar{I}_{AB} = \sum_{A_i, B_j} P_{AB}(A_i, B_j)\, I_{AB}. \tag{5.58}$$

If we now take for event A the measurements $s(t_k)$ and for B the events $s(t_k + T)$, then the average mutual information is defined as

$$I(T) = \sum_{s(t_k), s(t_k+T)} P(s(t_k), s(t_k + T)) \log_2 \frac{P(s(t_k), s(t_k + T))}{P(s(t_k)) P(s(t_k + T))}. \tag{5.59}$$

To determine $P(s(t_k))$, we make a histogram of the values of $s(t_k)$ versus k; same for $P(s(t_k + T))$. For the joint probability, we compute a two-dimensional histogram in the same way. The first minimum of the function $I(T)$ versus T will give a proper estimate of the time lag T.

5.5.2 Estimation of attractor properties

Once we have found an appropriate T and an embedding dimension d and the corresponding reconstruction of the dynamical system, that is,

$$\mathbf{y}_k = (s(t_k), s(t_k + T), \ldots, s(t_k + T(d-1)))^T, \tag{5.60}$$

for $k = 1, \ldots, N$, we are able to determine further properties of the attractor. A generalised view of the geometric structure of the attractor can be obtained by determining the number of data points $n(\mathbf{x}, r)$ within a radius r of a certain chosen point $\mathbf{x}$ in the d-dimensional space. With $\mathcal{H}$ being the Heaviside function, we have

$$n(\mathbf{x}, r) = \frac{1}{N} \sum_{k=1}^{N} \mathcal{H}(r - |\mathbf{y}_k - \mathbf{x}|), \tag{5.61}$$

where the norm is taken in $\mathbb{R}^d$. When we take $\mathbf{x}$ equal to all the $\mathbf{y}_k$, an average density is given by the correlation function

$$C(q, r) = \frac{1}{N} \sum_{k=1}^{N} \left[\frac{1}{K} \sum_{l=1}^{K} \mathcal{H}(r - |\mathbf{y}_k - \mathbf{y}_l|) \right]^{q-1}. \tag{5.62}$$

If the number of points $n(\mathbf{x}, r) \approx r^{D_q}$, then we expect an average density

$$C(q, r) \approx r^{(q-1)D_q} = e^{(q-1)D_q \ln r} \rightarrow \ln C(q, r) \approx (q-1) D_q \ln r. \tag{5.63}$$

Hence by plotting $\ln C(q, r)$ versus $\ln r$, we may find a slope $(q-1)D_q$. As we have chosen q, we find an estimate of the so-called correlation dimension of the attractor by

$$D_q = \lim_{r \to 0} \frac{\ln C(q, r)}{(q-1) \ln r}. \tag{5.64}$$

The box-counting dimension D_0 discussed in (2.43) is a special case. It is defined as

$$D_0 = \lim_{r \to 0} \frac{\ln C(0, r)}{-\ln r} \; ; \; C(0, r) \approx N(r). \tag{5.65}$$

The dimension D_2 is usually referred to as the Grassberger-Procaccia dimension.

As we saw in Section 2.4.3, an important property of an attractor is whether there is sensitivity to initial conditions; if so, there should be a positive Lyapunov exponent. The issue here is that now we have to determine these exponents from the data using the reconstructed attractor. This is not an easy issue in practice, but the procedure is as follows. The Lyapunov exponents require an estimate of the Jacobian matrix. In the reconstructed space, the different points on the trajectory are related through

$$\mathbf{y}_{k+1} = \mathbf{F}(\mathbf{y}_k), \tag{5.66}$$

for some mapping $\mathbf{F}$. When a perturbation Δ_k is assumed on a certain $\bar{\mathbf{y}}_k$, that is, $\mathbf{y}_k = \bar{\mathbf{y}}_k + \Delta_k$, then linearization leads to

$$\Delta_{k+1} = J(\bar{\mathbf{y}}_k)\Delta_k, \tag{5.67}$$

where J is the local Jacobian matrix. We hence find through iteration that

$$\Delta_{k+n} = J^n(\bar{\mathbf{y}}_k)\Delta_k. \tag{5.68}$$

At this point a famous theorem from ergodic theory due to Oseledec (Abarbanel, 1996) can be used, which states that:

(i) The length of the vector $|\Delta_{k+n}|$ is given by

$$|\Delta_{k+n}|^2 = \Delta_k^T (J^n(\bar{\mathbf{y}}_k))^T J^n(\bar{\mathbf{y}}_k)\Delta_k. \tag{5.69}$$

(ii) The Lyapunov exponents λ_i are given by

$$\lambda_i = \ln \sigma_i, \tag{5.70}$$

where σ_i are the eigenvalues of the Oseledec matrix

$$\mathcal{S} = \lim_{n\to\infty} \left[(J^n)^T J^n\right]^{\frac{1}{2n}}. \tag{5.71}$$

(iii) When we diagonalise the Oseledec matrix $\mathcal{S}$ into the form

$$\mathcal{S} = P^T \Sigma P,$$

where Σ is the diagonal matrix with the eigenvalues σ_i, then from (i) we find for large n

$$|\Delta_{k+n}|^2 \approx \Delta_k^T P^T (\Sigma)^{2n} P \Delta_k$$

and hence with $\tilde{\Delta}_k = P\Delta_k$, eventually

$$|\tilde{\Delta}_{k+n}|^2 \approx \sigma_1^{2n} |\tilde{\Delta}_k|^2 = e^{2n\lambda_1} |\tilde{\Delta}_k|^2, \tag{5.72}$$

where λ_1 is the largest Lyapunov exponent. This expression shows that the norm of the perturbations exponentially diverges if $\lambda_1 > 0$.

To compute the Oseledec matrix, we must determine an approximation of the Jacobian matrix from the data. Thereto we make use of the fact that every point on the attractor will be visited in time (because of the recurrence property). Choosing $\mathbf{y}_k$ and its N_B nearest neighbours $\mathbf{y}_k^r$, $r = 1, \ldots, N_B$, we next follow all the orbits of these nearest neighbours and choose a polynomial basis $\phi_m(\mathbf{x})$ to represent every vector into the form

$$\mathbf{y}_k = \sum_{m=1}^{M} c_{m,k} \phi_m(\mathbf{y}_k), \tag{5.73}$$

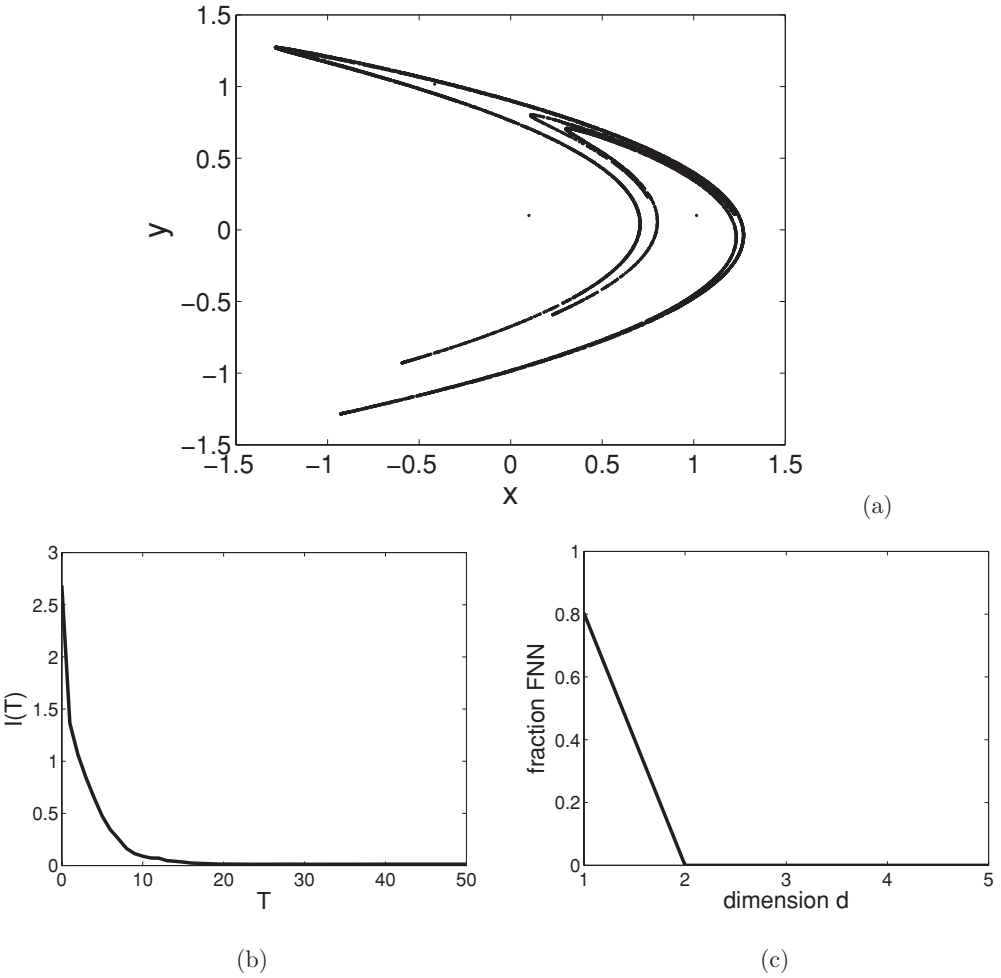


Figure 5.12 (a) Plot of the Hénon attractor obtained by iteration of the map (5.74). (b) Average mutual information for the Hénon map. (c) False nearest neighbours versus the embedding dimension d for the Hénon map (with a delay $T = 1$).

where the coefficients $c_{m,k}$ are determined through least squares. In this way we can compute the derivatives and hence an approximation of the Jacobian matrix.

Example 5.5 The Hénon map The equations for the Hénon map are given by

$$x_{k+1} = 1 - ax_k^2 + y_k, \tag{5.74a}$$

$$y_{k+1} = bx_k, \tag{5.74b}$$

with two parameters $a = 1.4$ and $b = 0.3$. By applying the iteration from any starting value, one eventually obtains the picture Fig. 5.12a, which is the famous Hénon attractor.

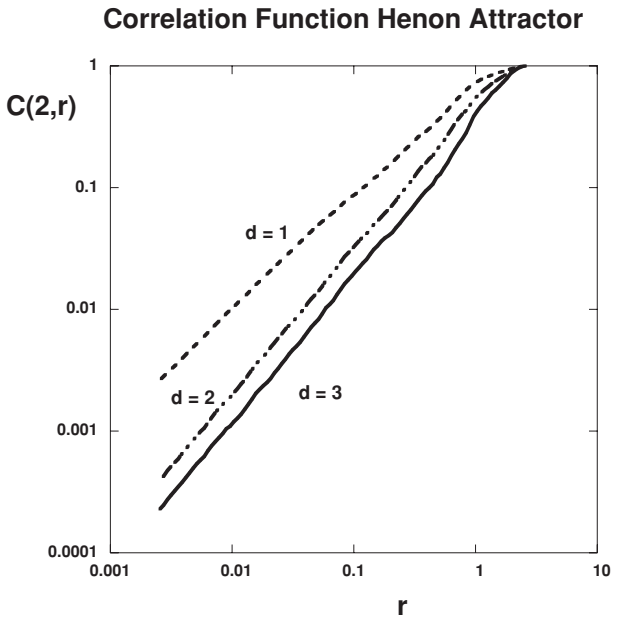


Figure 5.13 Correlation dimension of the Hénon map for different embedding dimensions d.

Here the natural delay time is the iteration step and hence $T = 1$. When we plot the average mutual information (Fig. 5.12b), we see that $I(T)$ is monotonically decreasing, and hence there is indeed no better choice than $T = 1$. With this delay, the FNN method is applied, and the percentage of FNN versus the embedding dimension (Fig. 5.12c) indicates that the Hénon attractor can be embedded into two-dimensional space, and hence $d = 2$ is a good choice.

In a logarithmic plot of $C(2, r)$ (calculated with TISEAN) for different embedding dimensions (Fig. 5.13), we obtain from the slope of $\log C(2, r)$ versus $\log r$ (and from a power-law fit for both $d = 2$ and $d = 3$) an estimate $D_2 = 1.23$. This is close to the standard value of 1.21 ± 0.01; the Hénon attractor is hence a fractal set. For the Hénon attractor, the TISEAN program provides values $\lambda_1 = 0.42$ and $\lambda_2 = -1.58$ for the Lyapunov exponents and hence the Hénon dynamical system is a chaotic system. ■

Of course, these approaches have their pitfalls. One of them is that one generally needs an enormous amount of data to accurately determine the desired properties of the attractors. Climate time series from observations are often too short, and the false nearest neighbour method may indicate a finite dimension of the attractor just because of lack of data.

5.5.3 Detection of critical transitions

In Subsection 3.5.5, we have seen that the approach to critical conditions in fast-slow systems is associated with an increase of variance in the time series. In this section, methodology is presented to detect the occurrence of such critical conditions from time series.

Suppose we have a time series of a certain quantity $x_n = x(t_n)$, where n indicates the time index. Near critical conditions, we know that one eigenvalue of the Jacobian matrix of the dynamical system that has generated x_n moves through the imaginary axis, and hence one normal mode switches from being damped to being amplified. To mimic this change in damping, suppose x is generated by the stochastic Itô SDE of the form (3.47) (Section 3.4.1),

$$X_t = X_0 + \int_0^t \lambda X_s ds + \sigma \int_0^t dW_s, \tag{5.75}$$

describing an Ornstein-Uhlenbeck process with damping coefficient λ. The solution to this equation is (3.49) with $\gamma = -\lambda$,

$$X_t = e^{\lambda t}\left(X_0 + \sigma \int_0^t e^{-\lambda s} dW_s\right). \tag{5.76}$$

The discrete version of (5.76) is (at t_{n+1} and t_n),

$$x_{n+1} = e^{\lambda t_{n+1}} X_0 + \sigma e^{\lambda t_{n+1}} \int_0^{t_{n+1}} e^{-\lambda s} dW_s, \tag{5.77a}$$

$$x_n = e^{\lambda t_n} X_0 + \sigma e^{\lambda t_n} \int_0^{t_n} e^{-\lambda s} dW_s \tag{5.77b}$$

and subtracting both equations gives

$$x_{n+1} - e^{\lambda \Delta t} x_n = \sigma \epsilon_n, \tag{5.78}$$

where ϵ_n represents the stochastic part and $\Delta t = t_{n+1} - t_n$. This is an AR(1) process (Section 5.2.2) of which the variance and autocorrelation follow from (5.24) as

$$Var[X] = \frac{\sigma^2}{1 - e^{2\lambda \Delta t}}, \tag{5.79a}$$

$$\rho_n = e^{n\lambda \Delta t}. \tag{5.79b}$$

This shows that the variance of the time series increases as $\lambda \to 0$ and also that long-range correlations do appear.

There are several techniques to detect long-range correlations and variance increase in time series, of which we discuss two: degenerate fingerprinting (Held and Kleinen, 2004) and detrented fluctuation analysis (Kantelhardt et al., 2001; Livina and Lenton, 2007). We see applications of these techniques in Section 10.3 when we discuss the Dansgaard-Oeschger oscillations.

Degenerate fingerprinting

The basis for this method is the equation (5.78), written in the form

$$x_{n+1} = cx_n + \sigma\epsilon_n, \tag{5.80}$$

with $c = e^{\lambda\Delta t}$. After the linear trend of the time series is removed, the quantity c is obtained by fitting. The error bars on c are given by (Held and Kleinen, 2004)

$$\Delta c^2 \approx (1 - c^2)\frac{\Delta t}{T} \rightarrow \Delta\lambda \cong \sqrt{\frac{2\lambda}{T}}, \tag{5.81}$$

where T is the total length of the time series. From this equation, one can also deduce the minimum length of the time series to obtain a specific accuracy in λ.

Detrended fluctuation analysis

This method is designed to determine long-range correlations in time series. In general, we have a time series $X(i)$, $i = 1, \ldots, N$ and determine the autocorrelation as

$$C(s) = \frac{1}{N - s}\sum_{i=1}^{N-s} X(i)X(i + s), \tag{5.82}$$

For processes with short-range correlations, we expect $C(s) \approx e^{-s/\tau}$, where τ is the decay time scale. However, for long-range correlations, an autocorrelation of the form $C(s) = s^{-\gamma}$, with $0 < \gamma < 1$ can be expected.

The detrended fluctuation analysis essentially consists of three steps:

1. Let $Y(i) = \sum_{k=1}^{i} X(k)$. The profile of $Y(i)$ is cut into precisely $N_s = N/s$ nonoverlapping segments of equal length s.
2. Determine the local trend P_ν of each segment ν by a least-squares polynomial fit of order n to the data. The detrended time series for a segment of duration s, denoted by $Y_s(i)$, is then given by

$$Y_s(i) = Y(i) - P_\nu(i). \tag{5.83}$$

   An example of the construction of the $Y_s(i)$ for $s = 100$ and $s = 200$ is shown in Fig. 5.14.
3. Calculate for each of the segments the quantity

$$F_s^2(\nu) = \frac{1}{s}\sum_{i=1}^{s}(Y_s((\nu - 1)s + i))^2, \tag{5.84}$$

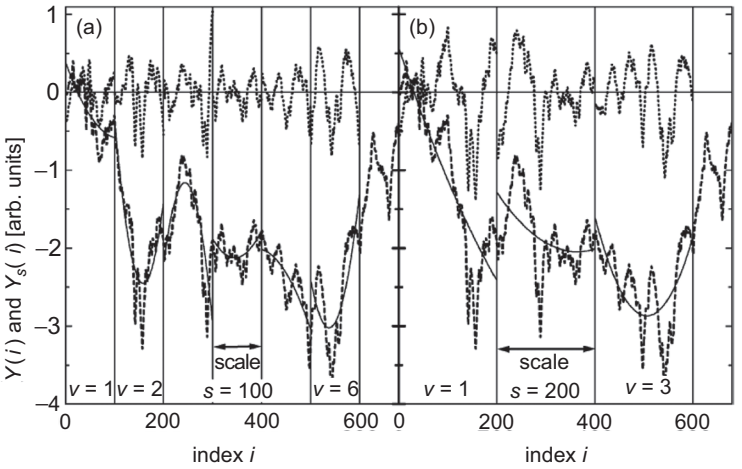


Figure 5.14 Illustration of the detrending procedure in the detrended fluctuation analysis for (a) $s = 100$ and (b) $s = 200$. The time series $Y(i)$ is the dashed curve, the drawn curves are the least-squares quadratic fits $P_2(i)$ and the dotted curve is the time series $Y_s(i)$ (figure from Kantelhardt et al., 2001).

and then from this the fluctuation function

$$F(s) = \left[\frac{1}{N_s} \sum_{v=1}^{N_s} F_s^2(v) \right]^{1/2}, \tag{5.85}$$

for several different polynomial orders n.

From Fig. 5.14, it is seen that the variance in Y_s, as measured by $F(s)$ increases with s. Suppose the original time series has zero mean and no trend; then $Y_s(i) = Y(i)$ and hence

$$Y^2(i) = \left(\sum_{k=1}^{i} X(k) \right)^2 = \sum_{k=1}^{i} (X(k))^2 + \sum_{k \neq j}^{j,k \leq i} X(j)X(k). \tag{5.86}$$

The second term can be written as

$$\sum_{k \neq j}^{j,k \leq i} X(j)X(k) = 2 \sum_{k=1}^{i-1} (i-k)C(k),$$

with $C(k)$ defined in (5.82). For long-range correlations, we have $C(k) \sim k^{-\gamma}$, for $0 < \gamma < 1$, and hence

$$\sum_{k=1}^{i-1} kC(k) \sim \sum_{k=1}^{i} k^{1-\gamma} \sim \int_1^i k^{1-\gamma} dk \sim i^{2-\gamma}, \tag{5.87}$$

which will dominate the variance in Y for large i.

Substituting this result into (5.85), the dependence of $F(s)$ on s is given by

$$F(s) \sim (Y_s^2)^{1/2} \sim s^{1-\gamma/2} = s^{\alpha}, \tag{5.88}$$

where $\alpha = 1 - \gamma/2$. For short-range correlations (or uncorrelated) data, $C(s)$ decays exponentially, and the first term on the right-hand side of (5.86) dominates. As $Y^2(i) \sim i$, this gives $F(s) \sim s^{1/2}$, and hence $\alpha = 1/2$. Indications for long-range correlations in the time series are hence found when $\alpha > 1/2$.