4

Spectral Analysis and Filtering

4.1 Introduction

The notion that a time series exhibits repetitive or regular behavior over time is of fundamental importance because it distinguishes time series analysis from classical statistics, which assumes complete independence over time. We have seen how dependence over time can be introduced through models that describe in detail the way certain empirical data behaves, even to the extent of producing forecasts based on the models. It is natural that models based on predicting the present as a regression on the past, such as are provided by the celebrated ARIMA or state-space forms, will be attractive to statisticians, who are trained to view nature in terms of linear models. In fact, the difference equations used to represent these kinds of models are simply the discrete versions of linear differential equations that may, in some instances, provide the ideal physical model for a certain phenomenon. An alternate version of the way nature behaves exists, however, and is based on a decomposition of an empirical series into its regular components.

In this chapter, we argue that the concept of regularity of a series can best be expressed in terms of periodic variations of the underlying phenomenon that produced the series, expressed as Fourier frequencies being driven by sines and cosines. Such a possibility was discussed in Chapters 1 and 2. From a regression point of view, we may imagine a system responding to various driving frequencies by producing linear combinations of sine and cosine functions. Expressed in these terms, the time domain approach may be thought of as regression of the present on the past, whereas the frequency domain approach may be considered as regression of the present on periodic sines and cosines.

Frequency domain approaches are the focus of this chapter. To illustrate the two methods for generating series with a single primary periodic component, consider Figure 1.9, which was generated from a simple second-order autoregressive model, and the middle and bottom panels of Figure 1.11, which were generated by adding a cosine wave with a period of 50 points to white

noise. Both series exhibit strong periodic fluctuations, illustrating that both models can generate time series with regular behavior. As discussed in Example 2.8, a fundamental objective of spectral analysis is to identify the dominant frequencies in a series and to find an explanation of the system from which the measurements were derived.

Of course, the primary justification for any alternate model must lie in its potential for explaining the behavior of some empirical phenomenon. In this sense, an explanation involving only a few kinds of primary oscillations becomes simpler and more physically meaningful than a collection of parameters estimated for some selected difference equation. It is the tendency of observed data to show periodic kinds of fluctuations that justifies the use of frequency domain methods. Many of the examples in §1.2 are time series representing real phenomena that are driven by periodic components. The speech recording of the syllable *aa...hh* in Figure 1.3 contains a complicated mixture of frequencies related to the opening and closing of the glottis. Figure 1.5 shows the monthly SOI, which we later explain as a combination of two kinds of periodicities, a seasonal periodic component of 12 months and an El Niño component of about three to five years. Of fundamental interest is the return period of the El Niño phenomenon, which can have profound effects on local climate. Also of interest is whether the different periodic components of the new fish population depend on corresponding seasonal and El Niño-type oscillations. We introduce the coherence as a tool for relating the common periodic behavior of two series. Seasonal periodic components are often pervasive in economic time series; this phenomenon can be seen in the quarterly earnings series shown in Figure 1.1. In Figure 1.6, we see the extent to which various parts of the brain will respond to a periodic stimulus generated by having the subject do alternate left and right finger tapping. Figure 1.7 shows series from an earthquake and a nuclear explosion. The relative amounts of energy at various frequencies for the two phases can produce statistics, useful for discriminating between earthquakes and explosions.

In this chapter, we summarize an approach to handling correlation generated in stationary time series that begins by transforming the series to the frequency domain. This simple linear transformation essentially matches sines and cosines of various frequencies against the underlying data and serves two purposes as discussed in Example 2.8 and Example 2.9. The periodogram that was introduced in Example 2.9 has its population counterpart called the power spectrum, and its estimation is a main goal of spectral analysis. Another purpose of exploring this topic is statistical convenience resulting from the periodic components being nearly uncorrelated. This property facilitates writing likelihoods based on classical statistical methods.

An important part of analyzing data in the frequency domain, as well as the time domain, is the investigation and exploitation of the properties of the time-invariant linear filter. This special linear transformation is used similarly to linear regression in conventional statistics, and we use many of the same terms in the time series context. We have previously mentioned the coherence

as a measure of the relation between two series at a given frequency, and we show later that this coherence also measures the performance of the best linear filter relating the two series. Linear filtering can also be an important step in isolating a signal embedded in noise. For example, the lower panels of Figure 1.11 contain a signal contaminated with an additive noise, whereas the upper panel contains the pure signal. It might also be appropriate to ask whether a linear filter transformation exists that could be applied to the lower panel to produce a series closer to the signal in the upper panel. The use of filtering for reducing noise will also be a part of the presentation in this chapter. We emphasize, throughout, the analogy between filtering techniques and conventional linear regression.

Many frequency scales will often coexist, depending on the nature of the problem. For example, in the Johnson & Johnson data set in Figure 1.1, the predominant frequency of oscillation is one cycle per year (4 quarters), or .25 cycles per observation. The predominant frequency in the SOI and fish populations series in Figure 1.5 is also one cycle per year, but this corresponds to 1 cycle every 12 months, or .083 cycles per observation. For simplicity, we measure frequency, ω, at cycles per time point and discuss the implications of certain frequencies in terms of the problem context. Of descriptive interest is the *period* of a time series, defined as the number of points in a cycle, i.e., $1/\omega$. Hence, the predominant period of the Johnson & Johnson series is $1/.25$ or 4 quarters per cycle, whereas the predominant period of the SOI series is 12 months per cycle.

4.2 Cyclical Behavior and Periodicity

As previously mentioned, we have already encountered the notion of periodicity in numerous examples in Chapters 1, 2 and 3. The general notion of periodicity can be made more precise by introducing some terminology. In order to define the rate at which a series oscillates, we first define a cycle as one complete period of a sine or cosine function defined over a unit time interval. As in (1.5), we consider the periodic process

$$x_t = A\cos(2\pi\omega t + \phi) \tag{4.1}$$

for $t = 0, \pm 1, \pm 2, \ldots$, where ω is a frequency index, defined in cycles per unit time with A determining the height or *amplitude* of the function and ϕ, called the *phase*, determining the start point of the cosine function. We can introduce random variation in this time series by allowing the amplitude and phase to vary randomly.

As discussed in Example 2.8, for purposes of data analysis, it is easier to use a trigonometric identity[1] and write (4.1) as

[1] $\cos(\alpha \pm \beta) = \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta)$.

$$x_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t), \tag{4.2}$$

where $U_1 = A\cos\phi$ and $U_2 = -A\sin\phi$ are often taken to be normally distributed random variables. In this case, the amplitude is $A = \sqrt{U_1^2 + U_2^2}$ and the phase is $\phi = \tan^{-1}(-U_2/U_1)$. From these facts we can show that if, and only if, in (4.1), A and ϕ are independent random variables, where A^2 is chi-squared with 2 degrees of freedom, and ϕ is uniformly distributed on $(-\pi, \pi)$, then U_1 and U_2 are independent, standard normal random variables (see Problem 4.2).

The above random process is also a function of its frequency, defined by the parameter ω. The frequency is measured in cycles per unit time, or in cycles per point in the above illustration. For $\omega = 1$, the series makes one cycle per time unit; for $\omega = .50$, the series makes a cycle every two time units; for $\omega = .25$, every four units, and so on. In general, for data that occur at discrete time points will need at least two points to determine a cycle, so the highest frequency of interest is .5 cycles per point. This frequency is called the *folding frequency* and defines the highest frequency that can be seen in discrete sampling. Higher frequencies sampled this way will appear at lower frequencies, called *aliases*; an example is the way a camera samples a rotating wheel on a moving automobile in a movie, in which the wheel appears to be rotating at a different rate. For example, movies are recorded at 24 frames per second. If the camera is filming a wheel that is rotating at the rate of 24 cycles per second (or 24 Hertz), the wheel will appear to stand still (that's about 110 miles per hour in case you were wondering).

Consider a generalization of (4.2) that allows mixtures of periodic series with multiple frequencies and amplitudes,

$$x_t = \sum_{k=1}^{q} \left[U_{k1} \cos(2\pi\omega_k t) + U_{k2} \sin(2\pi\omega_k t) \right], \tag{4.3}$$

where U_{k1}, U_{k2}, for $k = 1, 2, \ldots, q$, are independent zero-mean random variables with variances σ_k^2, and the ω_k are distinct frequencies. Notice that (4.3) exhibits the process as a sum of independent components, with variance σ_k^2 for frequency ω_k. Using the independence of the Us and the trig identity in footnote 1, it is easy to show[2] (Problem 4.3) that the autocovariance function of the process is

$$\gamma(h) = \sum_{k=1}^{q} \sigma_k^2 \cos(2\pi\omega_k h), \tag{4.4}$$

and we note the autocovariance function is the sum of periodic components with weights proportional to the variances σ_k^2. Hence, x_t is a mean-zero stationary processes with variance

[2] For example, for x_t in (4.2) we have $\mathrm{cov}(x_{t+h}, x_t) = \sigma^2\{\cos(2\pi\omega[t+h])\cos(2\pi\omega t) + \sin(2\pi\omega[t+h])\sin(2\pi\omega t)\} = \sigma^2\cos(2\pi\omega h)$, noting that $\mathrm{cov}(U_1, U_2) = 0$.

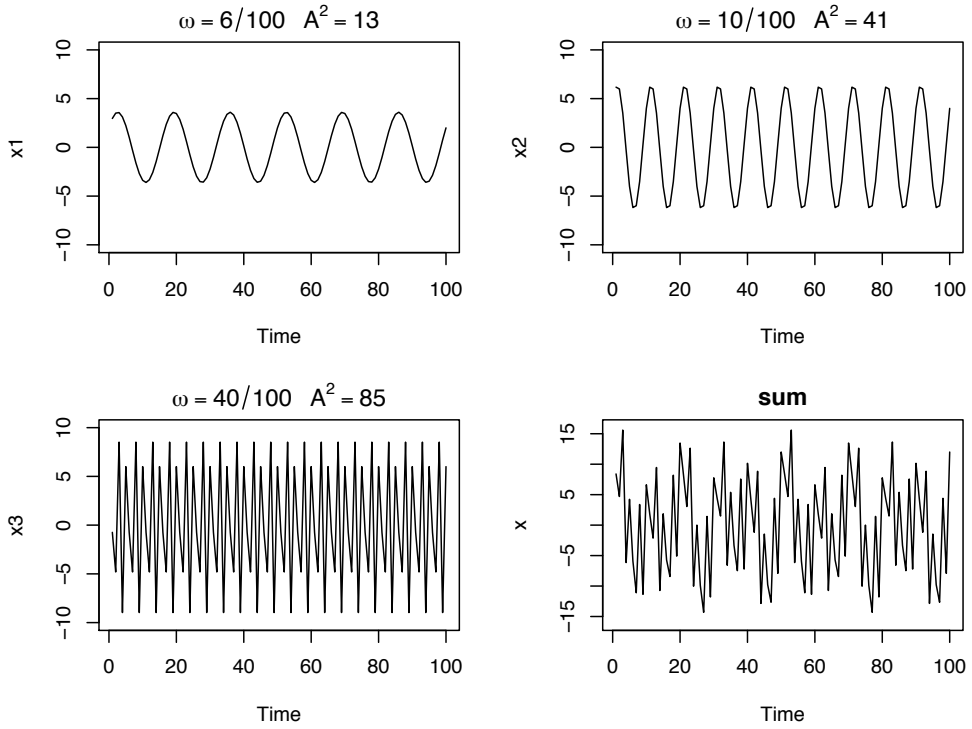

Fig. 4.1. Periodic components and their sum as described in Example 4.1.

$$\gamma(0) = E(x_t^2) = \sum_{k=1}^{q} \sigma_k^2, \tag{4.5}$$

which exhibits the overall variance as a sum of variances of each of the component parts.

Example 4.1 A Periodic Series

Figure 4.1 shows an example of the mixture (4.3) with $q = 3$ constructed in the following way. First, for $t = 1, \ldots, 100$, we generated three series

$$x_{t1} = 2\cos(2\pi t\, 6/100) + 3\sin(2\pi t\, 6/100)$$
$$x_{t2} = 4\cos(2\pi t\, 10/100) + 5\sin(2\pi t\, 10/100)$$
$$x_{t3} = 6\cos(2\pi t\, 40/100) + 7\sin(2\pi t\, 40/100)$$

These three series are displayed in Figure 4.1 along with the corresponding frequencies and squared amplitudes. For example, the squared amplitude of x_{t1} is $A^2 = 2^2 + 3^2 = 13$. Hence, the maximum and minimum values that x_{t1} will attain are $\pm\sqrt{13} = \pm 3.61$.

Finally, we constructed

$$x_t = x_{t1} + x_{t2} + x_{t3}$$

and this series is also displayed in Figure 4.1. We note that x_t appears to behave as some of the periodic series we saw in Chapters 1 and 2. The

systematic sorting out of the essential frequency components in a time series, including their relative contributions, constitutes one of the main objectives of spectral analysis.

The R code to reproduce Figure 4.1 is

```
x1 = 2*cos(2*pi*1:100*6/100) + 3*sin(2*pi*1:100*6/100)
x2 = 4*cos(2*pi*1:100*10/100) + 5*sin(2*pi*1:100*10/100)
x3 = 6*cos(2*pi*1:100*40/100) + 7*sin(2*pi*1:100*40/100)
x = x1 + x2 + x3
par(mfrow=c(2,2))
plot.ts(x1, ylim=c(-10,10), main=expression(omega==6/100~~~A^2==13))
plot.ts(x2, ylim=c(-10,10), main=expression(omega==10/100~~~A^2==41))
plot.ts(x3, ylim=c(-10,10), main=expression(omega==40/100~~~A^2==85))
plot.ts(x,  ylim=c(-16,16), main="sum")
```

Example 4.2 The Scaled Periodogram for Example 4.1

In §2.3, Example 2.9, we introduced the periodogram as a way to discover the periodic components of a time series. Recall that the scaled periodogram is given by

$$P(j/n) = \left(\frac{2}{n}\sum_{t=1}^{n} x_t \cos(2\pi t j/n)\right)^2 + \left(\frac{2}{n}\sum_{t=1}^{n} x_t \sin(2\pi t j/n)\right)^2, \qquad (4.6)$$

and it may be regarded as a measure of the squared correlation of the data with sinusoids oscillating at a frequency of $\omega_j = j/n$, or j cycles in n time points. Recall that we are basically computing the regression of the data on the sinusoids varying at the fundamental frequencies, j/n. As discussed in Example 2.9, the periodogram may be computed quickly using the fast Fourier transform (FFT), and there is no need to run repeated regressions.

The scaled periodogram of the data, x_t, simulated in Example 4.1 is shown in Figure 4.2, and it clearly identifies the three components x_{t1}, x_{t2}, and x_{t3} of x_t. Note that

$$P(j/n) = P(1 - j/n), \quad j = 0, 1, \ldots, n-1,$$

so there is a mirroring effect at the folding frequency of 1/2; consequently, the periodogram is typically not plotted for frequencies higher than the folding frequency. In addition, note that the heights of the scaled periodogram shown in the figure are

$$P(6/100) = 13, \quad P(10/100) = 41, \quad P(40/100) = 85,$$

$P(j/n) = P(1-j/n)$ and $P(j/n) = 0$ otherwise. These are exactly the values of the squared amplitudes of the components generated in Example 4.1. This outcome suggests that the periodogram may provide some insight into the variance components, (4.5), of a real set of data.

Assuming the simulated data, `x`, were retained from the previous example, the R code to reproduce Figure 4.2 is

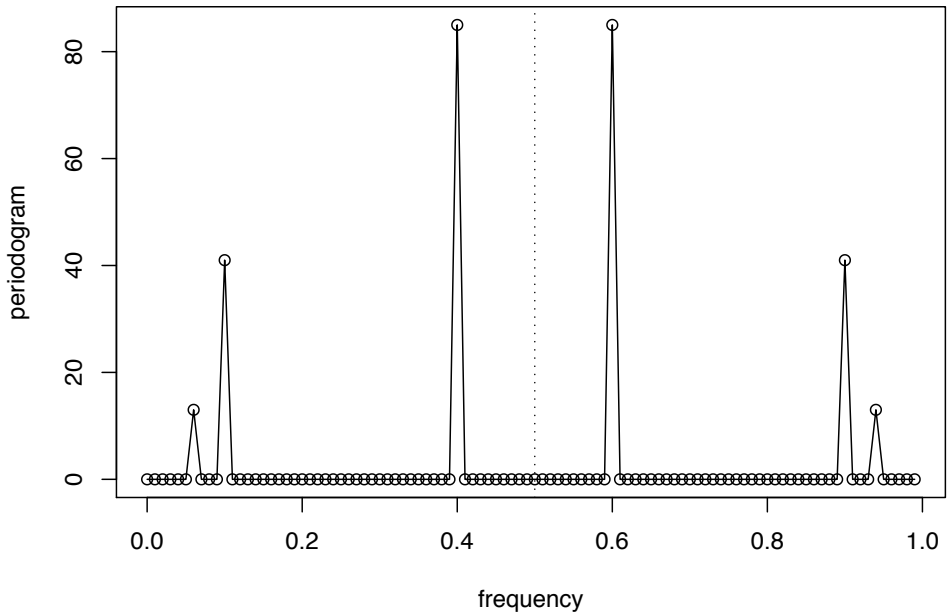

Fig. 4.2. Periodogram of the data generated in Example 4.1.

```
P = abs(2*fft(x)/100)^2;  Fr = 0:99/100
plot(Fr, P, type="o", xlab="frequency", ylab="periodogram")
```

If we consider the data x_t in Example 4.1 as a color (waveform) made up of primary colors x_{t1}, x_{t2}, x_{t3} at various strengths (amplitudes), then we might consider the periodogram as a prism that decomposes the color x_t into its primary colors (spectrum). Hence the term *spectral analysis*.

Another fact that may be of use in understanding the periodogram is that for any time series sample $x_1, \ldots, x_n$, where n is odd, we may write, *exactly*

$$x_t = a_0 + \sum_{j=1}^{(n-1)/2} \left[a_j \cos(2\pi t\, j/n) + b_j \sin(2\pi t\, j/n)\right], \tag{4.7}$$

for $t = 1, \ldots, n$ and suitably chosen coefficients. If n is even, the representation (4.7) can be modified by summing to $(n/2 - 1)$ and adding an additional component given by $a_{n/2}\cos(2\pi t\, 1/2) = a_{n/2}(-1)^t$. The crucial point here is that (4.7) is exact for any sample. Hence (4.3) may be thought of as an approximation to (4.7), the idea being that many of the coefficients in (4.7) may be close to zero. Recall from Example 2.9 that

$$P(j/n) = a_j^2 + b_j^2, \tag{4.8}$$

so the scaled periodogram indicates which components in (4.7) are large in magnitude and which components are small. We also saw (4.8) in Example 4.2.

The periodogram, which was introduced in Schuster (1898) and used in Schuster (1906) for studying the periodicities in the sunspot series (shown in

Figure 4.31 in the Problems section) is a sample based statistic. In Example 4.2, we discussed the fact that the periodogram may be giving us an idea of the variance components associated with each frequency, as presented in (4.5), of a time series. These variance components, however, are population parameters. The concepts of population parameters and sample statistics, as they relate to spectral analysis of time series can be generalized to cover stationary time series and that is the topic of the next section.

4.3 The Spectral Density

The idea that a time series is composed of periodic components, appearing in proportion to their underlying variances, is fundamental in the spectral representation. The result is quite technical because it involves stochastic integration; that is, integration with respect to a stochastic process. The essence of the result is that (4.3) is approximately true for any stationary time series. In other words, we have the following.

Property 4.1 Spectral Representation of a Stationary Process

In nontechnical terms, any stationary time series may be thought of, approximately, as the random superposition of sines and cosines oscillating at various frequencies.

Given that (4.3) is approximately true for all stationary time series, the next question is whether a meaningful representation for its autocovariance function, like the one displayed in (4.4), also exists. The answer is yes. The following example will help explain the result.

Example 4.3 A Periodic Stationary Process

Consider a periodic stationary random process given by (4.2), with a fixed frequency ω_0, say,

$$x_t = U_1 \cos(2\pi\omega_0 t) + U_2 \sin(2\pi\omega_0 t),$$

where U_1 and U_2 are independent zero-mean random variables with equal variance σ^2. The number of time periods needed for the above series to complete one cycle is exactly $1/\omega_0$, and the process makes exactly ω_0 cycles per point for $t = 0, \pm1, \pm2, \ldots$. It is easily shown that[3]

$$\begin{aligned}\gamma(h) &= \sigma^2 \cos(2\pi\omega_0 h) = \frac{\sigma^2}{2} e^{-2\pi i\omega_0 h} + \frac{\sigma^2}{2} e^{2\pi i\omega_0 h} \\ &= \int_{-1/2}^{1/2} e^{2\pi i\omega h} dF(\omega)\end{aligned}$$

[3] Some identities may be helpful here: $e^{i\alpha} = \cos(\alpha) + i\sin(\alpha)$ and consequently, $\cos(\alpha) = (e^{i\alpha} + e^{-i\alpha})/2$ and $\sin(\alpha) = (e^{i\alpha} - e^{-i\alpha})/2i$.

using a Riemann–Stieltjes integration, where $F(\omega)$ is the function defined by

$$F(\omega) = \begin{cases} 0 & \omega < -\omega_0, \\ \sigma^2/2 & -\omega_0 \le \omega < \omega_0, \\ \sigma^2 & \omega \ge \omega_0. \end{cases}$$

The function $F(\omega)$ behaves like a cumulative distribution function for a discrete random variable, except that $F(\infty) = \sigma^2 = \mathrm{var}(x_t)$ instead of one. In fact, $F(\omega)$ is a cumulative distribution function, not of probabilities, but rather of variances associated with the frequency ω_0 in an analysis of variance, with $F(\infty)$ being the total variance of the process x_t. Hence, we term $F(\omega)$ the *spectral distribution function*.

A representation such as the one given in Example 4.3 always exists for a stationary process. In particular, if x_t is stationary with autocovariance $\gamma(h) = E[(x_{t+h} - \mu)(x_t - \mu)]$, then there exists a unique monotonically increasing function $F(\omega)$, called the spectral distribution function, that is bounded, with $F(-\infty) = F(-1/2) = 0$, and $F(\infty) = F(1/2) = \gamma(0)$ such that

$$\gamma(h) = \int_{-1/2}^{1/2} \mathrm{e}^{2\pi i\omega h}\, dF(\omega). \tag{4.9}$$

A more important situation we use repeatedly is the case when the autocovariance function is absolutely summable, in which case the spectral distribution function is absolutely continuous with $dF(\omega) = f(\omega)\, d\omega$, and the representation (4.9) becomes the motivation for the property given below.

Property 4.2 The Spectral Density

If the autocovariance function, $\gamma(h)$, of a stationary process satisfies

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \tag{4.10}$$

then it has the representation

$$\gamma(h) = \int_{-1/2}^{1/2} \mathrm{e}^{2\pi i\omega h}\, f(\omega)\, d\omega \quad h = 0, \pm 1, \pm 2, \ldots \tag{4.11}$$

as the inverse transform of the spectral density, which has the representation

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h)\mathrm{e}^{-2\pi i\omega h} \quad -1/2 \le \omega \le 1/2. \tag{4.12}$$

This spectral density is the analogue of the probability density function; the fact that $\gamma(h)$ is non-negative definite ensures

$$f(\omega) \geq 0$$

for all ω. It follows immediately from (4.12) that

$$f(\omega) = f(-\omega)$$

verifying the spectral density is an even function. Because of the evenness, we will typically only plot $f(\omega)$ for $\omega \geq 0$. In addition, putting $h = 0$ in (4.11) yields

$$\gamma(0) = \mathrm{var}(x_t) = \int_{-1/2}^{1/2} f(\omega)\, d\omega,$$

which expresses the total variance as the integrated spectral density over all of the frequencies. We show later on, that a linear filter can isolate the variance in certain frequency intervals or bands.

Analogous to probability theory, $\gamma(h)$ in (4.11) is the characteristic function[4] of the spectral density $f(\omega)$ in (4.12). These facts should make it clear that, when the conditions of Property 4.2 are satisfied, *the autocovariance function, $\gamma(h)$, and the spectral density function, $f(\omega)$, contain the same information.* That information, however, is expressed in different ways. The autocovariance function expresses information in terms of lags, whereas the spectral density expresses the same information in terms of cycles. Some problems are easier to work with when considering lagged information and we would tend to handle those problems in the time domain. Nevertheless, other problems are easier to work with when considering periodic information and we would tend to handle those problems in the spectral domain.

We note that the autocovariance function, $\gamma(h)$, in (4.11) and the spectral density, $f(\omega)$, in (4.12) are Fourier transform pairs. In particular, this means that if $f(\omega)$ and $g(\omega)$ are two spectral densities for which

$$\gamma_f(h) = \int_{-1/2}^{1/2} f(\omega)\mathrm{e}^{2\pi i\omega h}\, d\omega = \int_{-1/2}^{1/2} g(\omega)\mathrm{e}^{2\pi i\omega h}\, d\omega = \gamma_g(h) \tag{4.13}$$

for all $h = 0, \pm 1, \pm 2, \ldots$, then

$$f(\omega) = g(\omega). \tag{4.14}$$

We also mention, at this point, that we have been focusing on the frequency ω, expressed in cycles per point rather than the more common (in statistics) alternative $\lambda = 2\pi\omega$ that would give radians per point. Finally, the absolute summability condition, (4.10), is not satisfied by (4.4), the example that we have used to introduce the idea of a spectral representation. The condition, however, is satisfied for ARMA models.

It is illuminating to examine the spectral density for the series that we have looked at in earlier discussions.

[4] If $M_X(\lambda) = E(e^{\lambda X})$ for $\lambda \in \mathbb{R}$ is the moment generating function of random variable X, then $\varphi_X(\lambda) = M_X(i\lambda)$ is the characteristic function.

Example 4.4 White Noise Series

As a simple example, consider the theoretical power spectrum of a sequence of uncorrelated random variables, w_t, with variance σ_w^2. A simulated set of data is displayed in the top of Figure 1.8. Because the autocovariance function was computed in Example 1.16 as $\gamma_w(h) = \sigma_w^2$ for $h = 0$, and zero, otherwise, it follows from (4.12), that

$$f_w(\omega) = \sigma_w^2$$

for $-1/2 \le \omega \le 1/2$. Hence the process contains equal power at all frequencies. This property is seen in the realization, which seems to contain all different frequencies in a roughly equal mix. In fact, the name white noise comes from the analogy to white light, which contains all frequencies in the color spectrum at the same level of intensity. Figure 4.3 shows a plot of the white noise spectrum for $\sigma_w^2 = 1$.

If x_t is ARMA, its spectral density can be obtained explicitly using the fact that it is a linear process, i.e., $x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$, where $\sum_{j=0}^{\infty} |\psi_j| < \infty$. In the following property, we exhibit the form of the spectral density of an ARMA model. The proof of the property follows directly from the proof of a more general result, Property 4.7 given on page 221, by using the additional fact that $\psi(z) = \theta(z)/\phi(z)$; recall Property 3.1.

Property 4.3 The Spectral Density of ARMA

If x_t is ARMA(p, q), $\phi(B)x_t = \theta(B)w_t$, its spectral density is given by

$$f_x(\omega) = \sigma_w^2 \frac{|\theta(\mathrm{e}^{-2\pi i\omega})|^2}{|\phi(\mathrm{e}^{-2\pi i\omega})|^2} \tag{4.15}$$

where $\phi(z) = 1 - \sum_{k=1}^{p} \phi_k z^k$ and $\theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k$.

Example 4.5 Moving Average

As an example of a series that does not have an equal mix of frequencies, we consider a moving average model. Specifically, consider the MA(1) model given by

$$x_t = w_t + .5w_{t-1}.$$

A sample realization is shown in the top of Figure 3.2 and we note that the series has less of the higher or faster frequencies. The spectral density will verify this observation.

The autocovariance function is displayed in Example 3.4 on page 90, and for this particular example, we have

$$\gamma(0) = (1 + .5^2)\sigma_w^2 = 1.25\sigma_w^2; \quad \gamma(\pm 1) = .5\sigma_w^2; \quad \gamma(\pm h) = 0 \text{ for } h > 1.$$

Substituting this directly into the definition given in (4.12), we have

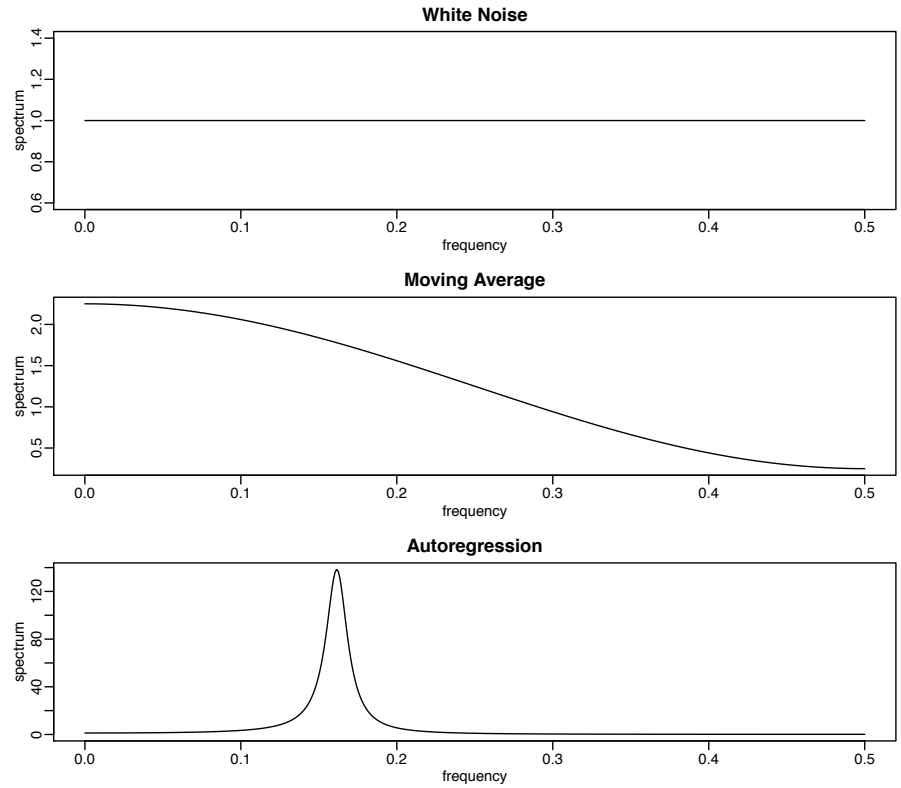

Fig. 4.3. Theoretical spectra of white noise (top), a first-order moving average (middle), and a second-order autoregressive process (bottom).

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h)\, \mathrm{e}^{-2\pi i\omega h} = \sigma_w^2 \left[1.25 + .5\left(\mathrm{e}^{-2\pi i\omega} + \mathrm{e}^{2\pi\omega}\right)\right] \\ = \sigma_w^2 \left[1.25 + \cos(2\pi\omega)\right]. \tag{4.16}$$

We can also compute the spectral density using Property 4.3, which states that for an MA, $f(\omega) = \sigma_w^2 |\theta(\mathrm{e}^{-2\pi i\omega})|^2$. Because $\theta(z) = 1 + .5z$, we have

$$|\theta(\mathrm{e}^{-2\pi i\omega})|^2 = |1 + .5\mathrm{e}^{-2\pi i\omega}|^2 = (1 + .5\mathrm{e}^{-2\pi i\omega})(1 + .5\mathrm{e}^{2\pi i\omega}) \\ = 1.25 + .5\left(\mathrm{e}^{-2\pi i\omega} + \mathrm{e}^{2\pi\omega}\right)$$

which leads to agreement with (4.16).

Plotting the spectrum for $\sigma_w^2 = 1$, as in the middle of Figure 4.3, shows the lower or slower frequencies have greater power than the higher or faster frequencies.

Example 4.6 A Second-Order Autoregressive Series

We now consider the spectrum of an AR(2) series of the form

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = w_t,$$

for the special case $\phi_1 = 1$ and $\phi_2 = -.9$. Figure 1.9 on page 14 shows a sample realization of such a process for $\sigma_w = 1$. We note the data exhibit a strong periodic component that makes a cycle about every six points.

To use Property 4.3, note that $\theta(z) = 1$, $\phi(z) = 1 - z + .9z^2$ and

$$\begin{aligned}|\phi(e^{-2\pi i\omega})|^2 &= (1 - \mathrm{e}^{-2\pi i\omega} + .9\mathrm{e}^{-4\pi i\omega})(1 - \mathrm{e}^{2\pi i\omega} + .9\mathrm{e}^{4\pi i\omega}) \\ &= 2.81 - 1.9(\mathrm{e}^{2\pi i\omega} + \mathrm{e}^{-2\pi i\omega}) + .9(\mathrm{e}^{4\pi i\omega} + \mathrm{e}^{-4\pi i\omega}) \\ &= 2.81 - 3.8\cos(2\pi\omega) + 1.8\cos(4\pi\omega).\end{aligned}$$

Using this result in (4.15), we have that the spectral density of x_t is

$$f_x(\omega) = \frac{\sigma_w^2}{2.81 - 3.8\cos(2\pi\omega) + 1.8\cos(4\pi\omega)}.$$

Setting $\sigma_w = 1$, the bottom of Figure 4.3 displays $f_x(\omega)$ and shows a strong power component at about $\omega = .16$ cycles per point or a period between six and seven cycles per point and very little power at other frequencies. In this case, modifying the white noise series by applying the second-order AR operator has concentrated the power or variance of the resulting series in a very narrow frequency band.

The spectral density can also be obtained from first principles, without having to use Property 4.3. Because $w_t = x_t - x_{t-1} + .9x_{t-2}$ in this example, we have

$$\begin{aligned}\gamma_w(h) &= \mathrm{cov}(w_{t+h}, w_t) \\ &= \mathrm{cov}(x_{t+h} - x_{t+h-1} + .9x_{t+h-2},\, x_t - x_{t-1} + .9x_{t-2}) \\ &= 2.81\gamma_x(h) - 1.9[\gamma_x(h+1) + \gamma_x(h-1)] + .9[\gamma_x(h+2) + \gamma_x(h-2)]\end{aligned}$$

Now, substituting the spectral representation (4.11) for $\gamma_x(h)$ in the above equation yields

$$\begin{aligned}\gamma_w(h) &= \int_{-1/2}^{1/2} \big[2.81 - 1.9(\mathrm{e}^{2\pi i\omega} + \mathrm{e}^{-2\pi i\omega}) + .9(\mathrm{e}^{4\pi i\omega} + \mathrm{e}^{-4\pi i\omega})\big]\mathrm{e}^{2\pi i\omega h} f_x(\omega)d\omega \\ &= \int_{-1/2}^{1/2} \big[2.81 - 3.8\cos(2\pi\omega) + 1.8\cos(4\pi\omega)\big]\mathrm{e}^{2\pi i\omega h} f_x(\omega)d\omega.\end{aligned}$$

If the spectrum of the white noise process, w_t, is $g_w(\omega)$, the uniqueness of the Fourier transform allows us to identify

$$g_w(\omega) = [2.81 - 3.8\cos(2\pi\omega) + 1.8\cos(4\pi\omega)]\, f_x(\omega).$$

But, as we have already seen, $g_w(\omega) = \sigma_w^2$, from which we deduce that

$$f_x(\omega) = \frac{\sigma_w^2}{2.81 - 3.8\cos(2\pi\omega) + 1.8\cos(4\pi\omega)}$$

is the spectrum of the autoregressive series.

To reproduce Figure 4.3, use the `spec.arma` script (see §R.1):

```
par(mfrow=c(3,1))
spec.arma(log="no", main="White Noise")
spec.arma(ma=.5, log="no", main="Moving Average")
spec.arma(ar=c(1,-.9), log="no", main="Autoregression")
```

The above examples motivate the use of the power spectrum for describing the theoretical variance fluctuations of a stationary time series. Indeed, the interpretation of the spectral density function as the variance of the time series over a given frequency band gives us the intuitive explanation for its physical meaning. The plot of the function $f(\omega)$ over the frequency argument ω can even be thought of as an analysis of variance, in which the columns or block effects are the frequencies, indexed by ω.

Example 4.7 Every Explosion has a Cause (cont)
In Example 3.3, we discussed the fact that explosive models have causal counterparts. In that example, we also indicated that it was easier to show this result in general in the spectral domain. In this example, we give the details for an AR(1) model, but the techniques used here will indicate how to generalize the result.

As in Example 3.3, we suppose that $x_t = 2x_{t-1} + w_t$, where $w_t \sim$ iid $\mathrm{N}(0, \sigma_w^2)$. Then, the spectral density of x_t is

$$f_x(\omega) = \sigma_w^2 \, |1 - 2e^{-2\pi i\omega}|^{-2}. \tag{4.17}$$

But,

$$|1 - 2e^{-2\pi i\omega}| = |1 - 2e^{2\pi i\omega}| = |(2e^{2\pi i\omega})\,(\tfrac{1}{2}e^{-2\pi i\omega} - 1)| = 2\,|1 - \tfrac{1}{2}e^{-2\pi i\omega}|.$$

Thus, (4.17) can be written as

$$f_x(\omega) = \tfrac{1}{4}\sigma_w^2 \, |1 - \tfrac{1}{2}e^{-2\pi i\omega}|^{-2},$$

which implies that $x_t = \frac{1}{2}x_{t-1} + v_t$, with $v_t \sim$ iid $\mathrm{N}(0, \frac{1}{4}\sigma_w^2)$ is an equivalent form of the model.

4.4 Periodogram and Discrete Fourier Transform

We are now ready to tie together the periodogram, which is the sample-based concept presented in §4.2, with the spectral density, which is the population-based concept of §4.3.

Definition 4.1 *Given data* $x_1, \ldots, x_n$*, we define the* **discrete Fourier transform (DFT)** *to be*

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i\omega_j t} \tag{4.18}$$

for $j = 0, 1, \ldots, n-1$*, where the frequencies* $\omega_j = j/n$ *are called the* **Fourier** *or* **fundamental frequencies**.

If n is a highly composite integer (i.e., it has many factors), the DFT can be computed by the fast Fourier transform (FFT) introduced in Cooley and Tukey (1965). Also, different packages scale the FFT differently, so it is a good idea to consult the documentation. R computes the DFT defined in (4.18) without the factor $n^{-1/2}$, but with an additional factor of $e^{2\pi i\omega_j}$ that can be ignored because we will be interested in the squared modulus of the DFT. Sometimes it is helpful to exploit the inversion result for DFTs which shows the linear transformation is one-to-one. For the inverse DFT we have,

$$x_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) \mathrm{e}^{2\pi i \omega_j t} \tag{4.19}$$

for $t = 1, \ldots, n$. The following example shows how to calculate the DFT and its inverse in R for the data set $\{1, 2, 3, 4\}$; note that R writes a complex number $z = a + ib$ as `a+bi`.

```
(dft = fft(1:4)/sqrt(4))
  [1]  5+0i  -1+1i  -1+0i  -1-1i
(idft = fft(dft, inverse=TRUE)/sqrt(4))
  [1] 1+0i  2+0i  3+0i  4+0i
(Re(idft))  # keep it real
  [1] 1  2  3  4
```

We now define the periodogram as the squared modulus[5] of the DFT.

Definition 4.2 *Given data $x_1, \ldots, x_n$, we define the* **periodogram** *to be*

$$I(\omega_j) = |d(\omega_j)|^2 \tag{4.20}$$

for $j = 0, 1, 2, \ldots, n-1$.

Note that $I(0) = n\bar{x}^2$, where $\bar{x}$ is the sample mean. In addition, because $\sum_{t=1}^{n} \exp(-2\pi i t \frac{j}{n}) = 0$ for $j \neq 0$,[6] we can write the DFT as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} (x_t - \bar{x}) e^{-2\pi i \omega_j t} \tag{4.21}$$

for $j \neq 0$. Thus, for $j \neq 0$,

$$\begin{aligned} I(\omega_j) = |d(\omega_j)|^2 &= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} (x_t - \bar{x})(x_s - \bar{x}) e^{-2\pi i \omega_j (t-s)} \\ &= n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) e^{-2\pi i \omega_j h} \\ &= \sum_{h=-(n-1)}^{n-1} \widehat{\gamma}(h) e^{-2\pi i \omega_j h} \end{aligned} \tag{4.22}$$

[5] Recall that if $z = a + ib$, then $\bar{z} = a - ib$, and $|z|^2 = z\bar{z} = a^2 + b^2$.

[6] $\sum_{t=1}^{n} z^t = z\frac{1-z^n}{1-z}$ for $z \neq 1$.

where we have put $h = t - s$, with $\widehat{\gamma}(h)$ as given in (1.34). [7]

Recall, $P(\omega_j) = (4/n)I(\omega_j)$ where $P(\omega_j)$ is the scaled periodogram defined in (4.6). Henceforth we will work with $I(\omega_j)$ instead of $P(\omega_j)$. In view of (4.22), the periodogram, $I(\omega_j)$, is the sample version of $f(\omega_j)$ given in (4.12). That is, we may think of the periodogram as the "sample spectral density" of x_t.

It is sometimes useful to work with the real and imaginary parts of the DFT individually. To this end, we define the following transforms.

Definition 4.3 *Given data $x_1, \ldots, x_n$, we define the* **cosine transform**

$$d_c(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t \cos(2\pi\omega_j t) \tag{4.23}$$

and the **sine transform**

$$d_s(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t \sin(2\pi\omega_j t) \tag{4.24}$$

where $\omega_j = j/n$ for $j = 0, 1, \ldots, n-1$.

We note that $d(\omega_j) = d_c(\omega_j) - i\, d_s(\omega_j)$ and hence

$$I(\omega_j) = d_c^2(\omega_j) + d_s^2(\omega_j). \tag{4.25}$$

We have also discussed the fact that spectral analysis can be thought of as an analysis of variance. The next example examines this notion.

Example 4.8 Spectral ANOVA

Let $x_1, \ldots, x_n$ be a sample of size n, where for ease, n is odd. Then, recalling Example 2.9 on page 67 and the discussion around (4.7) and (4.8),

$$x_t = a_0 + \sum_{j=1}^{m} \left[a_j \cos(2\pi\omega_j t) + b_j \sin(2\pi\omega_j t) \right], \tag{4.26}$$

where $m = (n-1)/2$, is exact for $t = 1, \ldots, n$. In particular, using multiple regression formulas, we have $a_0 = \bar{x}$,

$$a_j = \frac{2}{n} \sum_{t=1}^{n} x_t \cos(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_c(\omega_j)$$

$$b_j = \frac{2}{n} \sum_{t=1}^{n} x_t \sin(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_s(\omega_j).$$

Hence, we may write

[7] Note that (4.22) can be used to obtain $\widehat{\gamma}(h)$ by taking the inverse DFT of $I(\omega_j)$. This approach was used in Example 1.27 to obtain a two-dimensional ACF.

$$(x_t - \bar{x}) = \frac{2}{\sqrt{n}} \sum_{j=1}^{m} [d_c(\omega_j)\cos(2\pi\omega_j t) + d_s(\omega_j)\sin(2\pi\omega_j t)]$$

for $t = 1, \ldots, n$. Squaring both sides and summing we obtain

$$\sum_{t=1}^{n}(x_t - \bar{x})^2 = 2\sum_{j=1}^{m}\left[d_c^2(\omega_j) + d_s^2(\omega_j)\right] = 2\sum_{j=1}^{m} I(\omega_j)$$

using the results of Problem 2.10(d) on page 81. Thus, we have partitioned the sum of squares into harmonic components represented by frequency ω_j with the periodogram, $I(\omega_j)$, being the mean square regression. This leads to the ANOVA table for n odd:

Source	df	SS	MS
ω_1	2	$2I(\omega_1)$	$I(\omega_1)$
ω_2	2	$2I(\omega_2)$	$I(\omega_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
ω_m	2	$2I(\omega_m)$	$I(\omega_m)$
Total	$n-1$	$\sum_{t=1}^{n}(x_t - \bar{x})^2$	

This decomposition means that if the data contain some strong periodic components, the periodogram values corresponding to those frequencies (or near those frequencies) will be large. On the other hand, the corresponding values of the periodogram will be small for periodic components not present in the data.

The following is an R example to help explain this concept. We consider $n = 5$ observations given by $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 2, x_5 = 1$. Note that the data complete one cycle, but not in a sinusoidal way. Thus, we should expect the $\omega_1 = 1/5$ component to be relatively large but not exhaustive, and the $\omega_2 = 2/5$ component to be small.

```
x = c(1, 2, 3, 2, 1)
c1 = cos(2*pi*1:5*1/5);  s1 = sin(2*pi*1:5*1/5)
c2 = cos(2*pi*1:5*2/5);  s2 = sin(2*pi*1:5*2/5)
omega1 = cbind(c1, s1);  omega2 = cbind(c2, s2)
anova(lm(x~omega1+omega2))    # ANOVA Table
```

```
             Df   Sum Sq  Mean Sq
omega1        2  2.74164  1.37082
omega2        2   .05836   .02918
Residuals     0   .00000
```

```
abs(fft(x))^2/5       # the periodogram (as a check)
```

```
[1] 16.2  1.37082  .029179  .029179  1.37082
 #  I(0)   I(1/5)   I(2/5)   I(3/5)   I(4/5)
```

Note that $\bar{x} = 1.8$, and $I(0) = 16.2 = 5 \times 1.8^2 (= n\bar{x}^2)$. Also, note that

$$I(1/5) = 1.37082 = \texttt{Mean Sq}(\omega_1) \quad \text{and} \quad I(2/5) = .02918 = \texttt{Mean Sq}(\omega_2)$$

and $I(j/5) = I(1-j/5)$, for $j = 3, 4$. Finally, we note that the sum of squares associated with the residuals (SSE) is zero, indicating an exact fit.

We are now ready to present some large sample properties of the periodogram. First, let μ be the mean of a stationary process x_t with absolutely summable autocovariance function $\gamma(h)$ and spectral density $f(\omega)$. We can use the same argument as in (4.22), replacing $\bar{x}$ by μ in (4.21), to write

$$I(\omega_j) = n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \mu)(x_t - \mu) e^{-2\pi i \omega_j h} \tag{4.27}$$

where ω_j is a non-zero fundamental frequency. Taking expectation in (4.27) we obtain

$$E\left[I(\omega_j)\right] = \sum_{h=-(n-1)}^{n-1} \left(\frac{n-|h|}{n}\right) \gamma(h) e^{-2\pi i \omega_j h}. \tag{4.28}$$

For any given $\omega \neq 0$, choose a sequence of fundamental frequencies $\omega_{j:n} \to \omega$[8] from which it follows by (4.28) that, as $n \to \infty$[9]

$$E\left[I(\omega_{j:n})\right] \to f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i h \omega}. \tag{4.29}$$

In other words, under absolute summability of $\gamma(h)$, the spectral density is the long-term average of the periodogram.

To examine the asymptotic distribution of the periodogram, we note that if x_t is a normal time series, the sine and cosine transforms will also be jointly normal, because they are linear combinations of the jointly normal random variables $x_1, x_2, \ldots, x_n$. In that case, the assumption that the covariance function satisfies the condition

$$\theta = \sum_{h=-\infty}^{\infty} |h||\gamma(h)| \ < \infty \tag{4.30}$$

is enough to obtain simple large sample approximations for the variances and covariances. Using the same argument used to develop (4.28) we have

$$\operatorname{cov}[d_c(\omega_j), d_c(\omega_k)] = n^{-1} \sum_{s=1}^{n} \sum_{t=1}^{n} \gamma(s-t) \cos(2\pi\omega_j s) \cos(2\pi\omega_k t), \tag{4.31}$$

$$\operatorname{cov}[d_c(\omega_j), d_s(\omega_k)] = n^{-1} \sum_{s=1}^{n} \sum_{t=1}^{n} \gamma(s-t) \cos(2\pi\omega_j s) \sin(2\pi\omega_k t), \tag{4.32}$$

[8] By this we mean $\omega_{j:n} = j_n/n$, where $\{j_n\}$ is a sequence of integers chosen so that j_n/n is the closest Fourier frequency to ω; consequently, $|j_n/n - \omega| \leq \frac{1}{2n}$.

[9] From Definition 4.2 we have $I(0) = n\bar{x}^2$, so the analogous result of (4.29) for the case $\omega = 0$ is $E[I(0)] - n\mu^2 = n \operatorname{var}(\bar{x}) \to f(0)$ as $n \to \infty$.

and

$$\operatorname{cov}[d_s(\omega_j), d_s(\omega_k)] = n^{-1} \sum_{s=1}^{n} \sum_{t=1}^{n} \gamma(s-t) \sin(2\pi\omega_j s) \sin(2\pi\omega_k t), \tag{4.33}$$

where the variance terms are obtained by setting $\omega_j = \omega_k$ in (4.31) and (4.33). In Appendix C, §C.2, we show the terms in (4.31)-(4.33) have interesting properties under assumption (4.30), namely, for $\omega_j, \omega_k \neq 0$ or $1/2$,

$$\operatorname{cov}[d_c(\omega_j), d_c(\omega_k)] = \begin{cases} f(\omega_j)/2 + \varepsilon_n & \omega_j = \omega_k, \\ \varepsilon_n & \omega_j \neq \omega_k, \end{cases} \tag{4.34}$$

$$\operatorname{cov}[d_s(\omega_j), d_s(\omega_k)] = \begin{cases} f(\omega_j)/2 + \varepsilon_n & \omega_j = \omega_k, \\ \varepsilon_n & \omega_j \neq \omega_k, \end{cases} \tag{4.35}$$

and

$$\operatorname{cov}[d_c(\omega_j), d_s(\omega_k)] = \varepsilon_n, \tag{4.36}$$

where the error term ε_n in the approximations can be bounded,

$$|\varepsilon_n| \leq \theta/n, \tag{4.37}$$

and θ is given by (4.30). If $\omega_j = \omega_k = 0$ or $1/2$ in (4.34), the multiplier $1/2$ disappears; note that $d_s(0) = d_s(1/2) = 0$, so (4.35) does not apply.

Example 4.9 Covariance of Sine and Cosine Transforms

For the three-point moving average series of Example 1.9 and $n = 256$ observations, the theoretical covariance matrix of the vector $\boldsymbol{d} = (d_c(\omega_{26}), d_s(\omega_{26}), d_c(\omega_{27}), d_s(\omega_{27}))'$ is

$$\operatorname{cov}(\boldsymbol{d}) = \begin{pmatrix} .3752 & -.0009 & -.0022 & -.0010 \\ -.0009 & .3777 & -.0009 & .0003 \\ -.0022 & -.0009 & .3667 & -.0010 \\ -.0010 & .0003 & -.0010 & .3692 \end{pmatrix}.$$

The diagonal elements can be compared with half the theoretical spectral values of $\frac{1}{2}f(\omega_{26}) = .3774$ for the spectrum at frequency $\omega_{26} = 26/256$, and of $\frac{1}{2}f(\omega_{27}) = .3689$ for the spectrum at $\omega_{27} = 27/256$. Hence, the cosine and sine transforms produce nearly uncorrelated variables with variances approximately equal to one half of the theoretical spectrum. For this particular case, the uniform bound is determined from $\theta = 8/9$, yielding $|\varepsilon_{256}| \leq .0035$ for the bound on the approximation error.

If $x_t \sim \text{iid}(0, \sigma^2)$, then it follows from (4.30)-(4.36), Problem 2.10(d), and a central limit theorem[10] that

[10] If $Y_j \sim \text{iid}(0, \sigma^2)$ and $\{a_j\}$ are constants for which $\sum_{j=1}^{n} a_j^2 / \max_{1 \leq j \leq n} a_j^2 \to \infty$ as $n \to \infty$, then $\sum_{j=1}^{n} a_j Y_j \sim \text{AN}\left(0, \sigma^2 \sum_{j=1}^{n} a_j^2\right)$. AN is read *asymptotically normal*.

$$d_c(\omega_{j:n}) \sim \text{AN}(0, \sigma^2/2) \quad \text{and} \quad d_s(\omega_{j:n}) \sim \text{AN}(0, \sigma^2/2) \tag{4.38}$$

jointly and independently, and independent of $d_c(\omega_{k:n})$ and $d_s(\omega_{k:n})$ provided $\omega_{j:n} \to \omega_1$ and $\omega_{k:n} \to \omega_2$ where $0 < \omega_1 \neq \omega_2 < 1/2$. We note that in this case, $f_x(\omega) = \sigma^2$. In view of (4.38), it follows immediately that as $n \to \infty$,

$$\frac{2I(\omega_{j:n})}{\sigma^2} \xrightarrow{d} \chi^2_2 \quad \text{and} \quad \frac{2I(\omega_{k:n})}{\sigma^2} \xrightarrow{d} \chi^2_2 \tag{4.39}$$

with $I(\omega_{j:n})$ and $I(\omega_{k:n})$ being asymptotically independent, where χ^2_ν denotes a chi-squared random variable with ν degrees of freedom.

Using the central limit theory of §C.2, it is fairly easy to extend the results of the iid case to the case of a linear process.

Property 4.4 Distribution of the Periodogram Ordinates
If

$$x_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \tag{4.40}$$

where $w_t \sim iid(0, \sigma^2_w)$*, and* (4.30) *holds, then for any collection of* m *distinct frequencies* $\omega_j \in (0, 1/2)$ *with* $\omega_{j:n} \to \omega_j$

$$\frac{2I(\omega_{j:n})}{f(\omega_j)} \xrightarrow{d} \text{iid } \chi^2_2 \tag{4.41}$$

provided $f(\omega_j) > 0$*, for* $j = 1, \ldots, m$.

This result is stated more precisely in Theorem C.7 of §C.3. Other approaches to large sample normality of the periodogram ordinates are in terms of cumulants, as in Brillinger (1981), or in terms of mixing conditions, such as in Rosenblatt (1956a). Here, we adopt the approach used by Hannan (1970), Fuller (1996), and Brockwell and Davis (1991).

The distributional result (4.41) can be used to derive an approximate confidence interval for the spectrum in the usual way. Let $\chi^2_\nu(\alpha)$ denote the lower α probability tail for the chi-squared distribution with ν degrees of freedom; that is,

$$\Pr\{\chi^2_\nu \leq \chi^2_\nu(\alpha)\} = \alpha. \tag{4.42}$$

Then, an approximate $100(1-\alpha)\%$ confidence interval for the spectral density function would be of the form

$$\frac{2\ I(\omega_{j:n})}{\chi^2_2(1-\alpha/2)} \leq f(\omega) \leq \frac{2\ I(\omega_{j:n})}{\chi^2_2(\alpha/2)}. \tag{4.43}$$

Often, nonstationary trends are present that should be eliminated before computing the periodogram. Trends introduce extremely low frequency components in the periodogram that tend to obscure the appearance at higher frequencies. For this reason, it is usually conventional to center the data prior

to a spectral analysis using either mean-adjusted data of the form $x_t - \bar{x}$ to eliminate the zero or d-c component or to use detrended data of the form $x_t - \widehat{\beta}_1 - \widehat{\beta}_2 t$ to eliminate the term that will be considered a half cycle by the spectral analysis. Note that higher order polynomial regressions in t or nonparametric smoothing (linear filtering) could be used in cases where the trend is nonlinear.

As previously indicated, it is often convenient to calculate the DFTs, and hence the periodogram, using the fast Fourier transform algorithm. The FFT utilizes a number of redundancies in the calculation of the DFT when n is highly composite; that is, an integer with many factors of $2, 3$, or 5, the best case being when $n = 2^p$ is a factor of 2. Details may be found in Cooley and Tukey (1965). To accommodate this property, we can pad the centered (or detrended) data of length n to the next highly composite integer n' by adding zeros, i.e., setting $x^c_{n+1} = x^c_{n+2} = \cdots = x^c_{n'} = 0$, where x^c_t denotes the centered data. This means that the fundamental frequency ordinates will be $\omega_j = j/n'$ instead of j/n. We illustrate by considering the periodogram of the SOI and Recruitment series, as has been given in Figure 1.5 of Chapter 1. Recall that they are monthly series and $n = 453$ months. To find n' in R, use the command `nextn(453)` to see that $n' = 480$ will be used in the spectral analyses by default [use `help(spec.pgram)` to see how to override this default].

Example 4.10 Periodogram of SOI and Recruitment Series

Figure 4.4 shows the periodograms of each series, where the frequency axis is labeled in multiples of $\Delta = 1/12$. As previously indicated, the centered data have been padded to a series of length 480. We notice a narrow-band peak at the obvious yearly (12 month) cycle, $\omega = 1\Delta = 1/12$. In addition, there is considerable power in a wide band at the lower frequencies that is centered around the four-year (48 month) cycle $\omega = \frac{1}{4}\Delta = 1/48$ representing a possible El Niño effect. This wide band activity suggests that the possible El Niño cycle is irregular, but tends to be around four years on average. We will continue to address this problem as we move to more sophisticated analyses.

Noting $\chi^2_2(.025) = .05$ and $\chi^2_2(.975) = 7.38$, we can obtain approximate 95% confidence intervals for the frequencies of interest. For example, the periodogram of the SOI series is $I_S(1/12) = .97$ at the yearly cycle. An approximate 95% confidence interval for the spectrum $f_S(1/12)$ is then

$$[2(.97)/7.38,\ 2(.97)/.05] = [.26,\ 38.4],$$

which is too wide to be of much use. We do notice, however, that the lower value of .26 is higher than any other periodogram ordinate, so it is safe to say that this value is significant. On the other hand, an approximate 95% confidence interval for the spectrum at the four-year cycle, $f_S(1/48)$, is

$$[2(.05)/7.38,\ 2(.05)/.05] = [.01,\ 2.12],$$

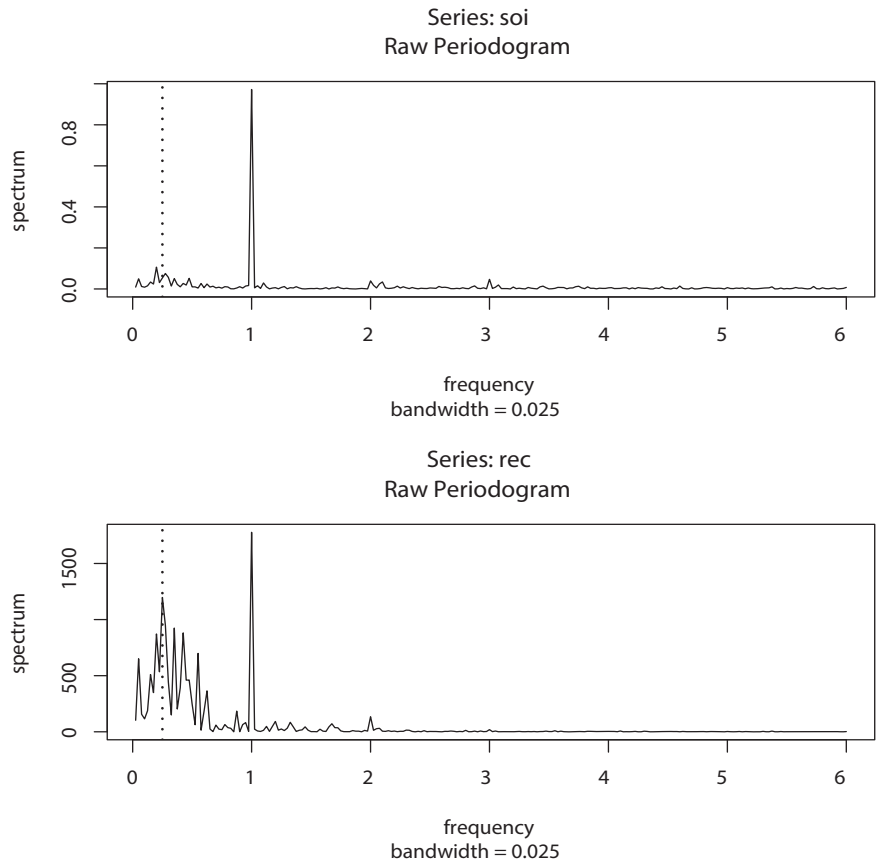

Fig. 4.4. Periodogram of SOI and Recruitment, $n = 453$ ($n' = 480$), where the frequency axis is labeled in multiples of $\Delta = 1/12$. Note the common peaks at $\omega = 1\Delta = 1/12$, or one cycle per year (12 months), and $\omega = \frac{1}{4}\Delta = 1/48$, or one cycle every four years (48 months).

which again is extremely wide, and with which we are unable to establish significance of the peak.

We now give the R commands that can be used to reproduce Figure 4.4. To calculate and graph the periodogram, we used the `spec.pgram` command in R. We note that the value of Δ is the reciprocal of the value of `frequency` used in `ts()` when making the data a time series object. If the data are not time series objects, `frequency` is set to 1. Also, we set `log="no"` because R will plot the periodogram on a $\log_{10}$ scale by default. Figure 4.4 displays a `bandwidth`. We will discuss bandwidth and tapering in the next section, so ignore these concepts for the time being.

```
require(astsa)  # needed for mvspec() - otherwise use spec.pgram()
par(mfrow=c(2,1))
soi.per = mvspec(soi, log="no")
abline(v=1/4, lty="dotted")
rec.per = mvspec(rec, log="no")
abline(v=1/4, lty="dotted")
```

The confidence intervals for the SOI series at the yearly cycle, $\omega = 1/12 = 40/480$, and the possible El Niño cycle of four years $\omega = 1/48 = 10/480$ can be computed in R as follows:

```
soi.per$spec[40]  # 0.97223; soi pgram at freq 1/12 = 40/480
soi.per$spec[10]  # 0.05372; soi pgram at freq 1/48 = 10/480
#  conf intervals  -  returned value:
U = qchisq(.025,2)    #  0.05063
L = qchisq(.975,2)    #  7.37775
2*soi.per$spec[10]/L  #  0.01456
2*soi.per$spec[10]/U  #  2.12220
2*soi.per$spec[40]/L  #  0.26355
2*soi.per$spec[40]/U  # 38.40108
```

The example above makes it clear that the periodogram as an estimator is susceptible to large uncertainties, and we need to find a way to reduce the variance. Not surprisingly, this result follows if we think about the periodogram, $I(\omega_j)$ as an estimator of the spectral density $f(\omega)$ and realize that it is the sum of squares of only two random variables for any sample size. The solution to this dilemma is suggested by the analogy with classical statistics where we look for independent random variables with the same variance and average the squares of these common variance observations. Independence and equality of variance do not hold in the time series case, but the covariance structure of the two adjacent estimators given in Example 4.9 suggests that for neighboring frequencies, these assumptions are approximately true.

4.5 Nonparametric Spectral Estimation

To continue the discussion that ended the previous section, we introduce a frequency band, $\mathcal{B}$, of $L << n$ contiguous fundamental frequencies, centered around frequency $\omega_j = j/n$, which is chosen close to a frequency of interest, ω. For frequencies of the form $\omega^* = \omega_j + k/n$, let

$$\mathcal{B} = \left\{\omega^* : \ \omega_j - \frac{m}{n} \le \omega^* \le \omega_j + \frac{m}{n}\right\}, \tag{4.44}$$

where

$$L = 2m + 1 \tag{4.45}$$

is an odd number, chosen such that the spectral values in the interval $\mathcal{B}$,

$$f(\omega_j + k/n), \quad k = -m, \ldots, 0, \ldots, m$$

are approximately equal to $f(\omega)$.

We now define an averaged (or smoothed) periodogram as the average of the periodogram values, say,

$$\bar{f}(\omega) = \frac{1}{L} \sum_{k=-m}^{m} I(\omega_j + k/n), \tag{4.46}$$

over the band $\mathcal{B}$. Under the assumption that the spectral density is fairly constant in the band $\mathcal{B}$, and in view of (4.41) we can show that under appropriate conditions,[11] for large n, the periodograms in (4.46) are approximately distributed as independent $f(\omega)\chi^2_2/2$ random variables, for $0 < \omega < 1/2$, as long as we keep L fairly small relative to n. This result is discussed formally in §C.2. Thus, under these conditions, $L\bar{f}(\omega)$ is the sum of L approximately independent $f(\omega)\chi^2_2/2$ random variables. It follows that, for large n,

$$\frac{2L\bar{f}(\omega)}{f(\omega)} \dot{\sim} \chi^2_{2L} \tag{4.47}$$

where $\dot{\sim}$ means *is approximately distributed as*.

In this scenario, where we smooth the periodogram by simple averaging, it seems reasonable to call the width of the frequency interval defined by (4.44),

$$B_w = \frac{L}{n}, \tag{4.48}$$

the bandwidth.[12] The concept of the bandwidth, however, becomes more complicated with the introduction of spectral estimators that smooth with unequal weights. Note (4.48) implies the degrees of freedom can be expressed as

$$2L = 2B_w n, \tag{4.49}$$

or twice the time-bandwidth product. The result (4.47) can be rearranged to obtain an approximate $100(1-\alpha)\%$ confidence interval of the form

$$\frac{2L\bar{f}(\omega)}{\chi^2_{2L}(1-\alpha/2)} \le f(\omega) \le \frac{2L\bar{f}(\omega)}{\chi^2_{2L}(\alpha/2)} \tag{4.50}$$

for the true spectrum, $f(\omega)$.

Many times, the visual impact of a spectral density plot will be improved by plotting the logarithm of the spectrum instead of the spectrum (the log

[11] The conditions, which are sufficient, are that x_t is a linear process, as described in Property 4.4, with $\sum_j \sqrt{|j|}\,|\psi_j| < \infty$, and w_t has a finite fourth moment.

[12] The bandwidth value used in R is based on Grenander (1951). The basic idea is that bandwidth can be related to the standard deviation of the weighting distribution. For the uniform distribution on the frequency range $-m/n$ to m/n, the standard deviation is $L/n\sqrt{12}$ (using a continuity correction). Consequently, in the case of (4.46), R will report a bandwidth of $L/n\sqrt{12}$, which amounts to dividing our definition by $\sqrt{12}$. Note that in the extreme case $L = n$, we would have $B_w = 1$ indicating that everything was used in the estimation; in this case, R would report a bandwidth of $1/\sqrt{12}$. There are many definitions of bandwidth and an excellent discussion may be found in Percival and Walden (1993, §6.7).

transformation is the variance stabilizing transformation in this situation). This phenomenon can occur when regions of the spectrum exist with peaks of interest much smaller than some of the main power components. For the log spectrum, we obtain an interval of the form

$$\big[\log \bar{f}(\omega) + \log 2L - \log \chi^2_{2L}(1-\alpha/2), \\ \log \bar{f}(\omega) + \log 2L - \log \chi^2_{2L}(\alpha/2)\big]. \quad (4.51)$$

We can also test hypotheses relating to the equality of spectra using the fact that the distributional result (4.47) implies that the ratio of spectra based on roughly independent samples will have an approximate $F_{2L,2L}$ distribution. The independent estimators can either be from different frequency bands or from different series.

If zeros are appended before computing the spectral estimators, we need to adjust the degrees of freedom and an approximation is to replace $2L$ by $2Ln/n'$. Hence, we define the adjusted degrees of freedom as

$$df = \frac{2Ln}{n'} \quad (4.52)$$

and use it instead of $2L$ in the confidence intervals (4.50) and (4.51). For example, (4.50) becomes

$$\frac{df\,\bar{f}(\omega)}{\chi^2_{df}(1-\alpha/2)} \le f(\omega) \le \frac{df\,\bar{f}(\omega)}{\chi^2_{df}(\alpha/2)}. \quad (4.53)$$

A number of assumptions are made in computing the approximate confidence intervals given above, which may not hold in practice. In such cases, it may be reasonable to employ resampling techniques such as one of the parametric bootstraps proposed by Hurvich and Zeger (1987) or a nonparametric local bootstrap proposed by Paparoditis and Politis (1999). To develop the bootstrap distributions, we assume that the contiguous DFTs in a frequency band of the form (4.44) all came from a time series with identical spectrum $f(\omega)$. This, in fact, is exactly the same assumption made in deriving the large-sample theory. We may then simply resample the L DFTs in the band, with replacement, calculating a spectral estimate from each bootstrap sample. The sampling distribution of the bootstrap estimators approximates the distribution of the nonparametric spectral estimator. For further details, including the theoretical properties of such estimators, see Paparoditis and Politis (1999).

Before proceeding further, we pause to consider computing the average periodograms for the SOI and Recruitment series, as shown in Figure 4.5.

Example 4.11 Averaged Periodogram for SOI and Recruitment

Generally, it is a good idea to try several bandwidths that seem to be compatible with the general overall shape of the spectrum, as suggested by the periodogram. The SOI and Recruitment series periodograms, previously computed in Figure 4.4, suggest the power in the lower El Niño frequency needs

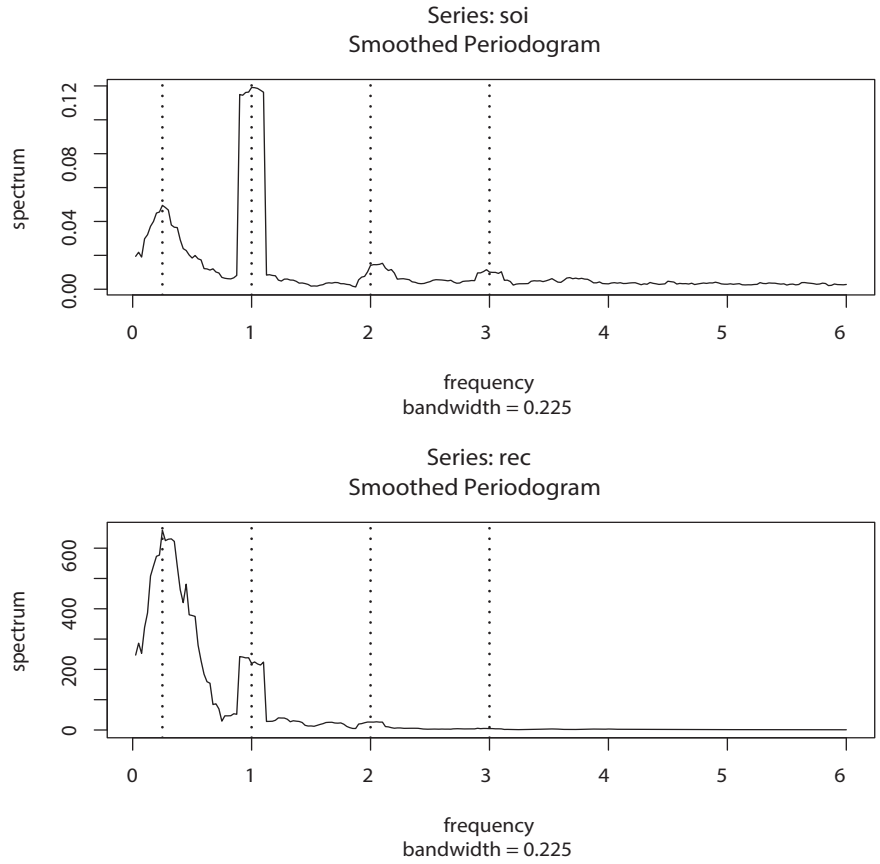

Fig. 4.5. The averaged periodogram of the SOI and Recruitment series $n = 453$, $n' = 480$, $L = 9$, $df = 17$, showing common peaks at the four year period, $\omega = \frac{1}{4}\Delta = 1/48$ cycles/month, the yearly period, $\omega = 1\Delta = 1/12$ cycles/month and some of its harmonics $\omega = k\Delta$ for $k = 2, 3$.

smoothing to identify the predominant overall period. Trying values of L leads to the choice $L = 9$ as a reasonable value, and the result is displayed in Figure 4.5.

The smoothed spectra shown in Figure 4.5 provide a sensible compromise between the noisy version, shown in Figure 4.4, and a more heavily smoothed spectrum, which might lose some of the peaks. An undesirable effect of averaging can be noticed at the yearly cycle, $\omega = 1\Delta$, where the narrow band peaks that appeared in the periodograms in Figure 4.4 have been flattened and spread out to nearby frequencies. We also notice, and have marked, the appearance of harmonics of the yearly cycle, that is, frequencies of the form $\omega = k\Delta$ for $k = 1, 2, \ldots$. Harmonics typically occur when a periodic component is present, but not in a sinusoidal fashion; see Example 4.12.

Figure 4.5 can be reproduced in R using the following commands. The basic call is to the function `mvspec`, which is available in `astsa`; alternately, use R's `spec.pgram`. To compute averaged periodograms, use the Daniell

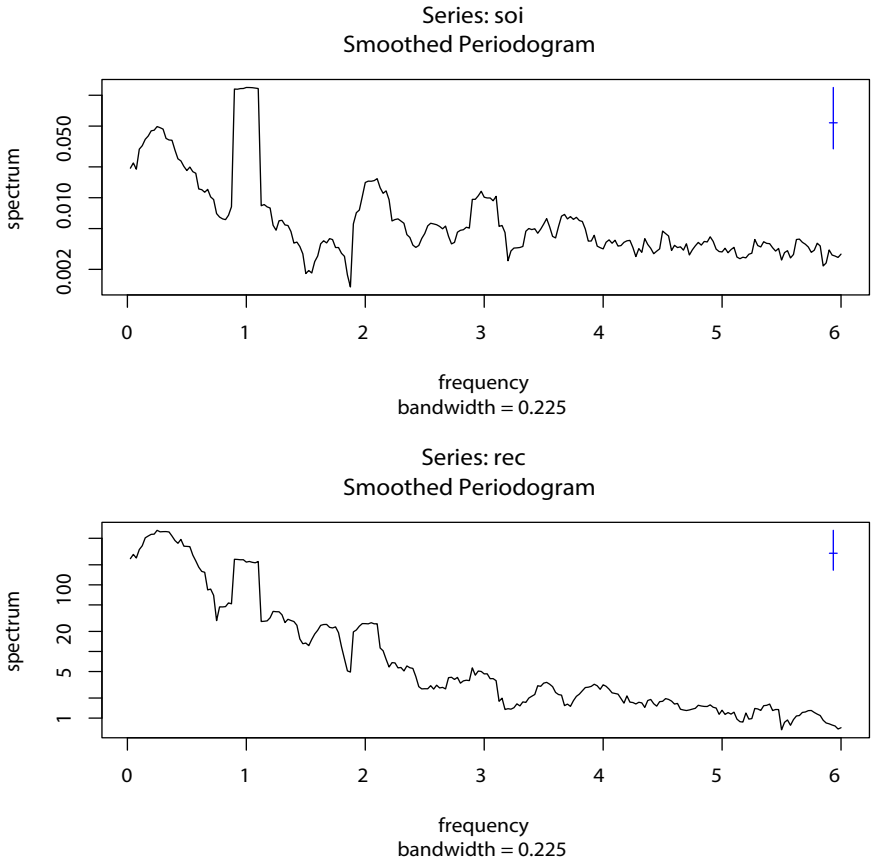

Fig. 4.6. Figure 4.5 with the average periodogram ordinates plotted on a $\log_{10}$ scale. The display in the upper right-hand corner represents a generic 95% confidence interval.

kernel, and specify m, where $L = 2m+1$ ($L = 9$ and $m = 4$ in this example). We will explain the kernel concept later in this section, specifically just prior to Example 4.13.

```
par(mfrow=c(2,1))
k = kernel("daniell", 4)
soi.ave = mvspec(soi, k, log="no")
abline(v=c(.25,1,2,3), lty=2)
# Repeat above lines using rec in place of soi on line 3
soi.ave$bandwidth       # = 0.225
```

The displayed bandwidth (.225) is adjusted for the fact that the frequency scale of the plot is in terms of cycles per year instead of cycles per month (the original unit of the data). Using (4.48), the bandwidth in terms of months is $9/480 = .01875$; the displayed value is simply converted to years, $.01875 \times 12 = .225$.

The adjusted degrees of freedom are $df = 2(9)(453)/480 \approx 17$. We can use this value for the 95% confidence intervals, with $\chi^2_{df}(.025) = 7.56$ and $\chi^2_{df}(.975) = 30.17$. Substituting into (4.53) gives the intervals in Table 4.1

Table 4.1. Confidence Intervals for the Spectra of the SOI and Recruitment Series

Series	ω	Period	Power	Lower	Upper
SOI	1/48	4 years	.05	.03	.11
	1/12	1 year	.12	.07	.27
Recruits	1/48	4 years	6.59	3.71	14.82
$\times 10^2$	1/12	1 year	2.19	1.24	4.93

for the two frequency bands identified as having the maximum power. To examine the two peak power possibilities, we may look at the 95% confidence intervals and see whether the lower limits are substantially larger than adjacent baseline spectral levels. For example, the El Niño frequency of 48 months has lower limits that exceed the values the spectrum would have if there were simply a smooth underlying spectral function without the peaks. The relative distribution of power over frequencies is different, with the SOI having less power at the lower frequency, relative to the seasonal periods, and the recruit series having relatively more power at the lower or El Niño frequency.

The entries in Table 4.1 for SOI can be obtained in R as follows:

```
df = soi.ave$df          #  df = 16.9875  (returned values)
U = qchisq(.025, df)     #  U = 7.555916
L = qchisq(.975, df)     #  L = 30.17425
soi.ave$spec[10]         #  0.0495202
soi.ave$spec[40]         #  0.1190800
# intervals
df*soi.ave$spec[10]/L    # 0.0278789
df*soi.ave$spec[10]/U    # 0.1113333
df*soi.ave$spec[40]/L    # 0.0670396
df*soi.ave$spec[40]/U    # 0.2677201
# repeat above commands with soi replaced by rec
```

Finally, Figure 4.6 shows the averaged periodograms in Figure 4.5 plotted on a $\log_{10}$ scale. This is the default plot in R, and these graphs can be obtained by removing the statement `log="no"`. Notice that the default plot also shows a generic confidence interval of the form (4.51) (with log replaced by $\log_{10}$) in the upper right-hand corner. To use it, imagine placing the tick mark on the averaged periodogram ordinate of interest; the resulting bar then constitutes an approximate 95% confidence interval for the spectrum at that frequency. We note that displaying the estimates on a log scale tends to emphasize the harmonic components.

Example 4.12 Harmonics

In the previous example, we saw that the spectra of the annual signals displayed minor peaks at the harmonics; that is, the signal spectra had a large peak at $\omega = 1\Delta = 1/12$ cycles/month (the one-year cycle) and minor peaks

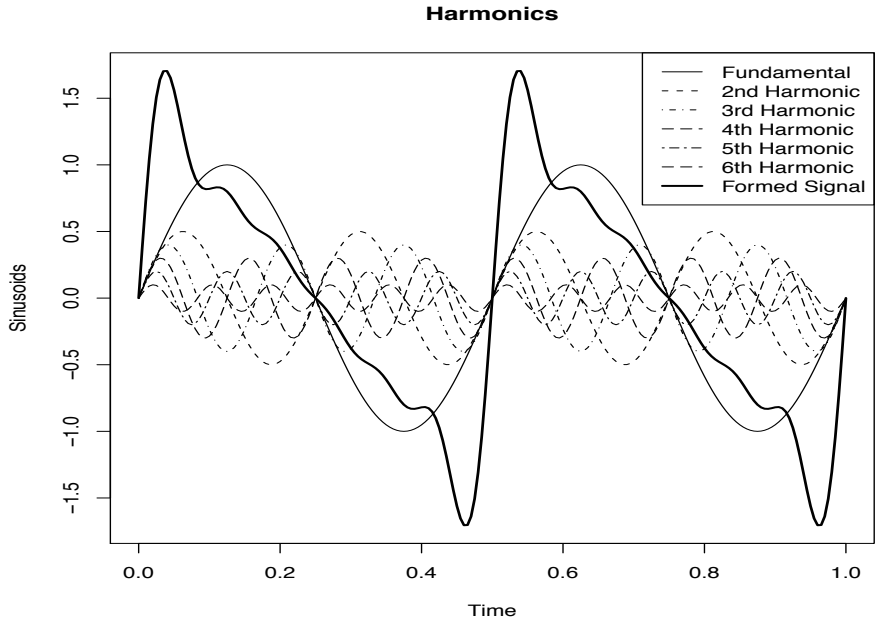

Fig. 4.7. A signal (thick solid line) formed by a fundamental sinusoid (thin solid line) oscillating at two cycles per unit time and its harmonics as specified in (4.54).

at its harmonics $\omega = k\Delta$ for $k = 2, 3, \ldots$ (two-, three-, and so on, cycles per year). This will often be the case because most signals are not perfect sinusoids (or perfectly cyclic). In this case, the harmonics are needed to capture the non-sinusoidal behavior of the signal. As an example, consider the signal formed in Figure 4.7 from a (fundamental) sinusoid oscillating at two cycles per unit time along with the second through sixth harmonics at decreasing amplitudes. In particular, the signal was formed as

$$\begin{aligned} x_t = {} & \sin(2\pi 2t) + .5\sin(2\pi 4t) + .4\sin(2\pi 6t) \\ & + .3\sin(2\pi 8t) + .2\sin(2\pi 10t) + .1\sin(2\pi 12t) \end{aligned} \tag{4.54}$$

for $0 \le t \le 1$. Notice that the signal is non-sinusoidal in appearance and rises quickly then falls slowly.

A figure similar to Figure 4.7 can be generated in R as follows.

```
t = seq(0, 1, by=1/200)
amps = c(1, .5, .4, .3, .2, .1)
x = matrix(0, 201, 6)
for (j in 1:6) x[,j] = amps[j]*sin(2*pi*t*2*j)
x = ts(cbind(x, rowSums(x)), start=0, deltat=1/200)
ts.plot(x, lty=c(1:6, 1), lwd=c(rep(1,6), 2), ylab="Sinusoids")
names = c("Fundamental","2nd Harmonic","3rd Harmonic","4th Harmonic",
    "5th Harmonic", "6th Harmonic", "Formed Signal")
legend("topright", names, lty=c(1:6, 1), lwd=c(rep(1,6), 2))
```

Example 4.11 points out the necessity for having some relatively systematic procedure for deciding whether peaks are significant. The question of deciding

whether a single peak is significant usually rests on establishing what we might think of as a baseline level for the spectrum, defined rather loosely as the shape that one would expect to see if no spectral peaks were present. This profile can usually be guessed by looking at the overall shape of the spectrum that includes the peaks; usually, a kind of baseline level will be apparent, with the peaks seeming to emerge from this baseline level. If the lower confidence limit for the spectral value is still greater than the baseline level at some predetermined level of significance, we may claim that frequency value as a statistically significant peak. To be consistent with our stated indifference to the upper limits, we might use a one-sided confidence interval.

An important aspect of interpreting the significance of confidence intervals and tests involving spectra is that typically, more than one frequency will be of interest, so that we will potentially be interested in simultaneous statements about a whole collection of frequencies. For example, it would be unfair to claim in Table 4.1 the two frequencies of interest as being statistically significant and all other potential candidates as nonsignificant at the overall level of $\alpha = .05$. In this case, we follow the usual statistical approach, noting that if K statements $S_1, S_2, \ldots, S_k$ are made at significance level α, i.e., $P\{S_k\} = 1 - \alpha$, then the overall probability all statements are true satisfies the Bonferroni inequality

$$P\{\text{all } S_k \text{ true}\} \geq 1 - K\alpha. \tag{4.55}$$

For this reason, it is desirable to set the significance level for testing each frequency at α/K if there are K potential frequencies of interest. If, a priori, potentially $K = 10$ frequencies are of interest, setting $\alpha = .01$ would give an overall significance level of bound of .10.

The use of the confidence intervals and the necessity for smoothing requires that we make a decision about the bandwidth B_w over which the spectrum will be essentially constant. Taking too broad a band will tend to smooth out valid peaks in the data when the constant variance assumption is not met over the band. Taking too narrow a band will lead to confidence intervals so wide that peaks are no longer statistically significant. Thus, we note that there is a conflict here between variance properties or bandwidth stability, which can be improved by increasing B_w and resolution, which can be improved by decreasing B_w. A common approach is to try a number of different bandwidths and to look qualitatively at the spectral estimators for each case.

To address the problem of resolution, it should be evident that the flattening of the peaks in Figure 4.5 and Figure 4.6 was due to the fact that simple averaging was used in computing $\bar{f}(\omega)$ defined in (4.46). There is no particular reason to use simple averaging, and we might improve the estimator by employing a weighted average, say

$$\widehat{f}(\omega) = \sum_{k=-m}^{m} h_k \, I(\omega_j + k/n), \tag{4.56}$$

using the same definitions as in (4.46) but where the weights $h_k > 0$ satisfy

$$\sum_{k=-m}^{m} h_k = 1.$$

In particular, it seems reasonable that the resolution of the estimator will improve if we use weights that decrease as distance from the center weight h_0 increases; we will return to this idea shortly. To obtain the averaged periodogram, $\bar{f}(\omega)$, in (4.56), set $h_k = L^{-1}$, for all k, where $L = 2m + 1$. The asymptotic theory established for $\bar{f}(\omega)$ still holds for $\widehat{f}(\omega)$ provided that the weights satisfy the additional condition that if $m \to \infty$ as $n \to \infty$ but $m/n \to 0$, then

$$\sum_{k=-m}^{m} h_k^2 \to 0.$$

Under these conditions, as $n \to \infty$,

(i) $E\left(\widehat{f}(\omega)\right) \to f(\omega)$

(ii) $\left(\sum_{k=-m}^{m} h_k^2\right)^{-1} \operatorname{cov}\left(\widehat{f}(\omega), \widehat{f}(\lambda)\right) \to f^2(\omega) \quad$ for $\omega = \lambda \neq 0, 1/2$.

In (ii), replace $f^2(\omega)$ by 0 if $\omega \neq \lambda$ and by $2f^2(\omega)$ if $\omega = \lambda = 0$ or $1/2$.

We have already seen these results in the case of $\bar{f}(\omega)$, where the weights are constant, $h_k = L^{-1}$, in which case $\sum_{k=-m}^{m} h_k^2 = L^{-1}$. The distributional properties of (4.56) are more difficult now because $\widehat{f}(\omega)$ is a weighted linear combination of asymptotically independent χ^2 random variables. An approximation that seems to work well is to replace L by $\left(\sum_{k=-m}^{m} h_k^2\right)^{-1}$. That is, define

$$L_h = \left(\sum_{k=-m}^{m} h_k^2\right)^{-1} \tag{4.57}$$

and use the approximation[13]

$$\frac{2L_h \widehat{f}(\omega)}{f(\omega)} \sim \dot{\chi}^2_{2L_h}. \tag{4.58}$$

In analogy to (4.48), we will define the bandwidth in this case to be

$$B_w = \frac{L_h}{n}. \tag{4.59}$$

Using the approximation (4.58) we obtain an approximate $100(1-\alpha)\%$ confidence interval of the form

[13] The approximation proceeds as follows: If $\widehat{f} \dot{\sim} c\chi^2_\nu$, where c is a constant, then $E\widehat{f} \approx c\nu$ and $\operatorname{var}\widehat{f} \approx f^2 \sum_k h_k^2 \approx c^2 2\nu$. Solving, $c \approx f \sum_k h_k^2/2 = f/2L_h$ and $\nu \approx 2\left(\sum_k h_k^2\right)^{-1} = 2L_h$.

$$\frac{2L_h\widehat{f}(\omega)}{\chi^2_{2L_h}(1-\alpha/2)} \le f(\omega) \le \frac{2L_h\widehat{f}(\omega)}{\chi^2_{2L_h}(\alpha/2)} \tag{4.60}$$

for the true spectrum, $f(\omega)$. If the data are padded to n', then replace $2L_h$ in (4.60) with $df = 2L_h n/n'$ as in (4.52).

An easy way to generate the weights in R is by repeated use of the Daniell kernel. For example, with $m = 1$ and $L = 2m + 1 = 3$, the Daniell kernel has weights $\{h_k\} = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$; applying this kernel to a sequence of numbers, $\{u_t\}$, produces

$$\widehat{u}_t = \tfrac{1}{3}u_{t-1} + \tfrac{1}{3}u_t + \tfrac{1}{3}u_{t+1}.$$

We can apply the same kernel again to the $\widehat{u}_t$,

$$\widehat{\widehat{u}}_t = \tfrac{1}{3}\widehat{u}_{t-1} + \tfrac{1}{3}\widehat{u}_t + \tfrac{1}{3}\widehat{u}_{t+1},$$

which simplifies to

$$\widehat{\widehat{u}}_t = \tfrac{1}{9}u_{t-2} + \tfrac{2}{9}u_{t-1} + \tfrac{3}{9}u_t + \tfrac{2}{9}u_{t+1} + \tfrac{1}{9}u_{t+2}.$$

The modified Daniell kernel puts half weights at the end points, so with $m = 1$ the weights are $\{h_k\} = \{\frac{1}{4}, \frac{2}{4}, \frac{1}{4}\}$ and

$$\widehat{u}_t = \tfrac{1}{4}u_{t-1} + \tfrac{1}{2}u_t + \tfrac{1}{4}u_{t+1}.$$

Applying the same kernel again to $\widehat{u}_t$ yields

$$\widehat{\widehat{u}}_t = \tfrac{1}{16}u_{t-2} + \tfrac{4}{16}u_{t-1} + \tfrac{6}{16}u_t + \tfrac{4}{16}u_{t+1} + \tfrac{1}{16}u_{t+2}.$$

These coefficients can be obtained in R by issuing the `kernel` command. For example, `kernel("modified.daniell", c(1,1))` would produce the coefficients of the last example. It is also possible to use different values of m, e.g., try `kernel("modified.daniell", c(1,2))` or `kernel("daniell", c(5,3))`. The other kernels that are currently available in R are the Dirichlet kernel and the Fejér kernel, which we will discuss shortly.

Example 4.13 Smoothed Periodogram for SOI and Recruitment

In this example, we estimate the spectra of the SOI and Recruitment series using the smoothed periodogram estimate in (4.56). We used a modified Daniell kernel twice, with $m = 3$ both times. This yields $L_h = 1/\sum_{k=-m}^{m} h_k^2 = 9.232$, which is close to the value of $L = 9$ used in Example 4.11. In this case, the bandwidth is $B_w = 9.232/480 = .019$ and the modified degrees of freedom is $df = 2L_h 453/480 = 17.43$. The weights, h_k, can be obtained and graphed in R as follows:

```
kernel("modified.daniell", c(3,3))

   coef[-6] = 0.006944 = coef[ 6]
   coef[-5] = 0.027778 = coef[ 5]
   coef[-4] = 0.055556 = coef[ 4]
```

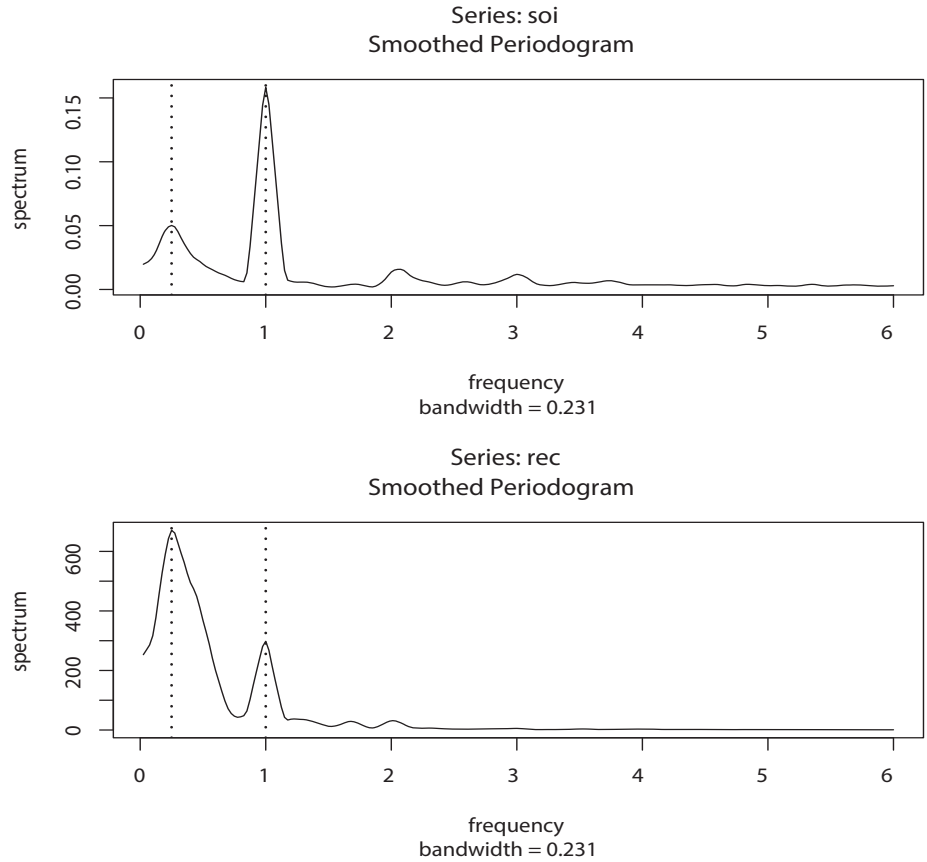

Fig. 4.8. Smoothed spectral estimates of the SOI and Recruitment series; see Example 4.13 for details.

```
coef[-3] = 0.083333 = coef[ 3]
coef[-2] = 0.111111 = coef[ 2]
coef[-1] = 0.138889 = coef[ 1]
coef[ 0] = 0.152778
```

```
plot(kernel("modified.daniell", c(3,3)))  # not shown
```

The resulting spectral estimates can be viewed in Figure 4.8 and we notice that the estimates more appealing than those in Figure 4.5. Figure 4.8 was generated in R as follows; we also show how to obtain df and B_w.

```
par(mfrow=c(2,1))
k = kernel("modified.daniell", c(3,3))
soi.smo = mvspec(soi, k, log="no")
abline(v=1, lty="dotted");  abline(v=1/4, lty="dotted")
## Repeat above lines with rec replacing soi in line 3
df = soi.smo$df             # df = 17.42618
soi.smo$bandwidth           # Bw = 0.2308103
```

Reissuing the `mvspec` commands with `log="no"` removed will result in a figure similar to Figure 4.6. Finally, we mention that the modified Daniell

kernel is used by default. For example, an easier way to obtain `soi.smo` is to issue the command:

```
soi.smo = mvspec(soi, spans=c(7,7))
```

Notice that `spans` is a vector of odd integers, given in terms of $L = 2m + 1$ instead of m. These values give the widths of the modified Daniell smoother to be used to smooth the periodogram.

We are now ready to briefly introduce the concept of *tapering*; a more detailed discussion may be found in Bloomfield (2000, §9.5). Suppose x_t is a mean-zero, stationary process with spectral density $f_x(\omega)$. If we replace the original series by the tapered series

$$y_t = h_t x_t, \tag{4.61}$$

for $t = 1, 2, \ldots, n$, use the modified DFT

$$d_y(\omega_j) = n^{-1/2} \sum_{t=1}^{n} h_t x_t \mathrm{e}^{-2\pi i \omega_j t}, \tag{4.62}$$

and let $I_y(\omega_j) = |d_y(\omega_j)|^2$, we obtain (see Problem 4.15)

$$E[I_y(\omega_j)] = \int_{-1/2}^{1/2} W_n(\omega_j - \omega)\ f_x(\omega)\ d\omega \tag{4.63}$$

where

$$W_n(\omega) = |H_n(\omega)|^2 \tag{4.64}$$

and

$$H_n(\omega) = n^{-1/2} \sum_{t=1}^{n} h_t \mathrm{e}^{-2\pi i \omega t}. \tag{4.65}$$

The value $W_n(\omega)$ is called a spectral window because, in view of (4.63), it is determining which part of the spectral density $f_x(\omega)$ is being "seen" by the estimator $I_y(\omega_j)$ on average. In the case that $h_t = 1$ for all t, $I_y(\omega_j) = I_x(\omega_j)$ is simply the periodogram of the data and the window is

$$W_n(\omega) = \frac{\sin^2(n\pi\omega)}{n \sin^2(\pi\omega)} \tag{4.66}$$

with $W_n(0) = n$, which is known as the Fejér or modified Bartlett kernel. If we consider the averaged periodogram in (4.46), namely

$$\bar{f}_x(\omega) = \frac{1}{L} \sum_{k=-m}^{m} I_x(\omega_j + k/n),$$

the window, $W_n(\omega)$, in (4.63) will take the form

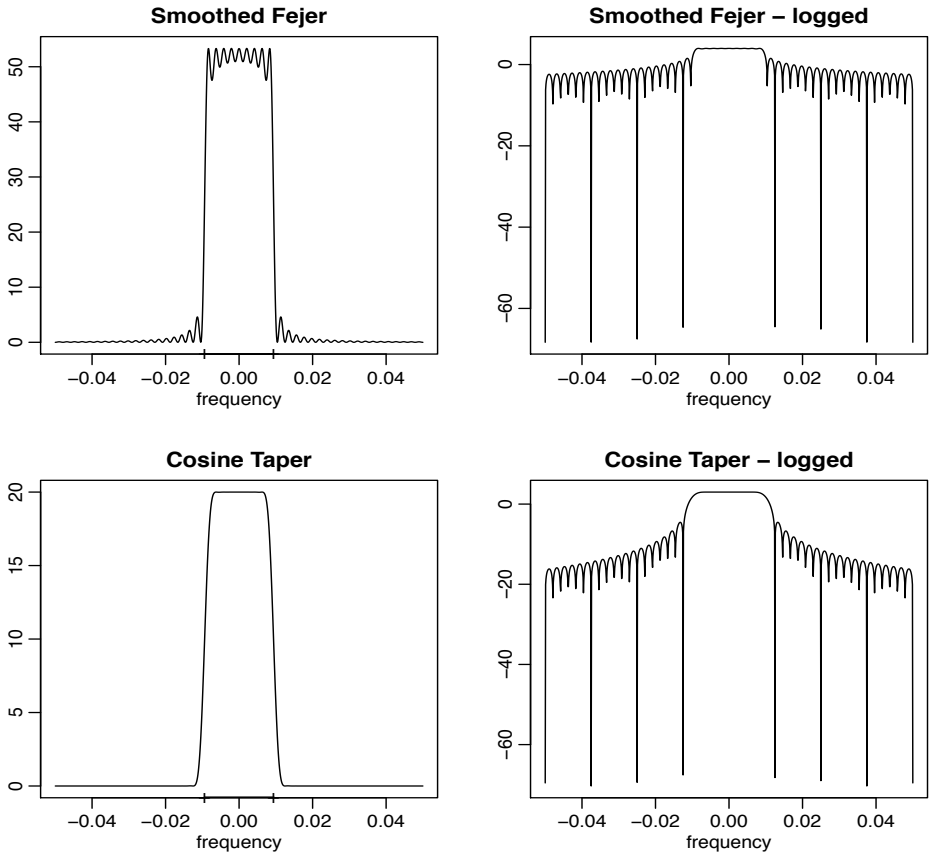

Fig. 4.9. Averaged Fejér window (top row) and the corresponding cosine taper window (bottom row) for $L = 9$, $n = 480$. The extra tic marks on the horizontal axis of the left-hand plots exhibit the predicted bandwidth, $B_w = 9/480 = .01875$.

$$W_n(\omega) = \frac{1}{nL} \sum_{k=-m}^{m} \frac{\sin^2[n\pi(\omega + k/n)]}{\sin^2[\pi(\omega + k/n)]} \,. \tag{4.67}$$

Tapers generally have a shape that enhances the center of the data relative to the extremities, such as a cosine bell of the form

$$h_t = .5\left[1 + \cos\left(\frac{2\pi(t - \bar{t})}{n}\right)\right], \tag{4.68}$$

where $\bar{t} = (n + 1)/2$, favored by Blackman and Tukey (1959). In Figure 4.9, we have plotted the shapes of two windows, $W_n(\omega)$, for $n = 480$ and $L = 9$, when (i) $h_t \equiv 1$, in which case, (4.67) applies, and (ii) h_t is the cosine taper in (4.68). In both cases the predicted bandwidth should be $B_w = 9/480 = .01875$ cycles per point, which corresponds to the "width" of the windows shown in Figure 4.9. Both windows produce an integrated average spectrum over this band but the untapered window in the top panels shows considerable ripples

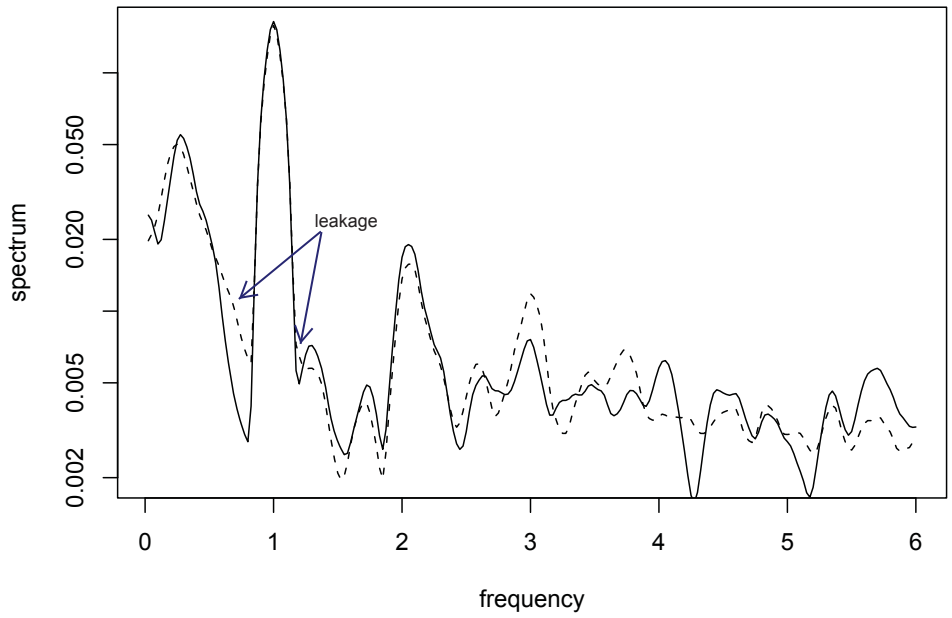

Fig. 4.10. Smoothed spectral estimates of the SOI without tapering (dashed line) and with full tapering (solid line); see Example 4.14 for details.

over the band and outside the band. The ripples outside the band are called sidelobes and tend to introduce frequencies from outside the interval that may contaminate the desired spectral estimate within the band. For example, a large dynamic range for the values in the spectrum introduces spectra in contiguous frequency intervals several orders of magnitude greater than the value in the interval of interest. This effect is sometimes called leakage. Figure 4.9 emphasizes the suppression of the sidelobes in the Fejér kernel when a cosine taper is used.

Example 4.14 The Effect of Tapering the SOI Series

In this example, we examine the effect of tapering on the estimate of the spectrum of the SOI series. The results for the Recruitment series are similar. Figure 4.10 shows two spectral estimates plotted on a log scale. The degree of smoothing here is the same as in Example 4.13. The dashed line in Figure 4.10 shows the estimate without any tapering and hence it is the same as the estimated spectrum displayed in the top of Figure 4.8. The solid line shows the result with full tapering. Notice that the tapered spectrum does a better job in separating the yearly cycle ($\omega = 1$) and the El Niño cycle ($\omega = 1/4$).

The following R session was used to generate Figure 4.10. We note that, by default, `mvspec` does not taper. For full tapering, we use the argument `taper=.5` to instruct `mvspec` to taper 50% of each end of the data; any value between 0 and .5 is acceptable.

```
s0 = mvspec(soi, spans=c(7,7), plot=FALSE)              # no taper
s50 = mvspec(soi, spans=c(7,7), taper=.5, plot=FALSE)   # full taper
```

```
plot(s0$freq, s0$spec, log="y", type="l", lty=2, ylab="spectrum",
    xlab="frequency")     # dashed line
lines(s50$freq, s50$spec)  # solid line
```

We close this section with a brief discussion of lag window estimators. First, consider the periodogram, $I(\omega_j)$, which was shown in (4.22) to be

$$I(\omega_j) = \sum_{|h|<n} \widehat{\gamma}(h) e^{-2\pi i \omega_j h}.$$

Thus, (4.56) can be written as

$$\begin{aligned} \widehat{f}(\omega) &= \sum_{|k|\le m} h_k \, I(\omega_j + k/n) \\ &= \sum_{|k|\le m} h_k \sum_{|h|<n} \widehat{\gamma}(h) e^{-2\pi i(\omega_j + k/n)h} \\ &= \sum_{|h|<n} g(h/n)\, \widehat{\gamma}(h) e^{-2\pi i \omega_j h}. \end{aligned} \tag{4.69}$$

where $g(h/n) = \sum_{|k|\le m} h_k \exp(-2\pi i k h/n)$. Equation (4.69) suggests estimators of the form

$$\widetilde{f}(\omega) = \sum_{|h|\le r} w(h/r)\, \widehat{\gamma}(h) e^{-2\pi i \omega h} \tag{4.70}$$

where $w(\cdot)$ is a weight function, called the lag window, that satisfies

(i) $w(0) = 1$
(ii) $|w(x)| \le 1$ and $w(x) = 0$ for $|x| > 1$,
(iii) $w(x) = w(-x)$.

Note that if $w(x) = 1$ for $|x| < 1$ and $r = n$, then $\widetilde{f}(\omega_j) = I(\omega_j)$, the periodogram. This result indicates the problem with the periodogram as an estimator of the spectral density is that it gives too much weight to the values of $\widehat{\gamma}(h)$ when h is large, and hence is unreliable [e.g, there is only one pair of observations used in the estimate $\widehat{\gamma}(n-1)$, and so on]. The smoothing window is defined to be

$$W(\omega) = \sum_{h=-r}^{r} w(h/r) e^{-2\pi i \omega h}, \tag{4.71}$$

and it determines which part of the periodogram will be used to form the estimate of $f(\omega)$. The asymptotic theory for $\widehat{f}(\omega)$ holds for $\widetilde{f}(\omega)$ under the same conditions and provided $r \to \infty$ as $n \to \infty$ but with $r/n \to 0$. We have

$$E\{\widetilde{f}(\omega)\} \to f(\omega), \tag{4.72}$$

$$\frac{n}{r} \mathrm{cov}\left(\widetilde{f}(\omega), \widetilde{f}(\lambda)\right) \to f^2(\omega) \int_{-1}^{1} w^2(x) dx \qquad \omega = \lambda \ne 0, 1/2. \tag{4.73}$$

In (4.73), replace $f^2(\omega)$ by 0 if $\omega \neq \lambda$ and by $2f^2(\omega)$ if $\omega = \lambda = 0$ or $1/2$.

Many authors have developed various windows and Brillinger (2001, Ch 3) and Brockwell and Davis (1991, Ch 10) are good sources of detailed information on this topic. We mention a few.

The rectangular lag window, which gives uniform weight in (4.70),

$$w(x) = 1, \quad |x| \leq 1,$$

corresponds to the Dirichlet smoothing window given by

$$W(\omega) = \frac{\sin(2\pi r + \pi)\omega}{\sin(\pi\omega)}. \tag{4.74}$$

This smoothing window takes on negative values, which may lead to estimates of the spectral density that are negative a various frequencies. Using (4.73) in this case, for large n we have

$$\text{var}\{\widetilde{f}(\omega)\} \approx \frac{2r}{n} f^2(\omega).$$

The Parzen lag window is defined to be

$$w(x) = \begin{cases} 1 - 6x + 6|x|^3 & |x| < 1/2, \\ 2(1 - |x|)^3 & 1/2 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

This leads to an approximate smoothing window of

$$W(\omega) = \frac{6}{\pi r^3} \frac{\sin^4(r\omega/4)}{\sin^4(\omega/2)}.$$

For large n, the variance of the estimator is approximately

$$\text{var}\{\widetilde{f}(\omega)\} \approx .539 f^2(\omega)/n.$$

The Tukey-Hanning lag window has the form

$$w(x) = \frac{1}{2}(1 + \cos(x)), \quad |x| \leq 1$$

which leads to the smoothing window

$$W(\omega) = \frac{1}{4} D_r(2\pi\omega - \pi/r) + \frac{1}{2} D_r(2\pi\omega) + \frac{1}{4} D_r(2\pi\omega + \pi/r)$$

where $D_r(\omega)$ is the Dirichlet kernel in (4.74). The approximate large sample variance of the estimator is

$$\text{var}\{\widetilde{f}(\omega)\} \approx \frac{3r}{4n} f^2(\omega).$$

The triangular lag window, also known as the Bartlett or Fejér window, given by

$$w(x) = 1 - |x|, \quad |x| \leq 1$$

leads to the Fejér smoothing window:

$$W(\omega) = \frac{\sin^2(\pi r\omega)}{r\sin^2(\pi\omega)}.$$

In this case, (4.73) yields

$$\mathrm{var}\{\widetilde{f}(\omega)\} \approx \frac{2r}{3n} f^2(\omega).$$

The idealized rectangular smoothing window, also called the Daniell window, is given by

$$W(\omega) = \begin{cases} r & |\omega| \leq 1/2r, \\ 0 & \text{otherwise}, \end{cases}$$

and leads to the sinc lag window, namely

$$w(x) = \frac{\sin(\pi x)}{\pi x}, \quad |x| \leq 1.$$

From (4.73) we have

$$\mathrm{var}\{\widetilde{f}(\omega)\} \approx \frac{r}{n} f^2(\omega).$$

For lag window estimators, the width of the idealized rectangular window that leads to the same asymptotic variance as a given lag window estimator is sometimes called the equivalent bandwidth. For example, the bandwidth of the idealized rectangular window is $b_r = 1/r$ and the asymptotic variance is $\frac{1}{nb_r} f^2$. The asymptotic variance of the triangular window is $\frac{2r}{3n} f^2$, so setting $\frac{1}{nb_r} f^2 = \frac{2r}{3n} f^2$ and solving we get $b_r = 3/2r$ as the equivalent bandwidth.

4.6 Parametric Spectral Estimation

The methods of §4.5 lead to estimators generally referred to as nonparametric spectra because no assumption is made about the parametric form of the spectral density. In Property 4.3, we exhibited the spectrum of an ARMA process and we might consider basing a spectral estimator on this function, substituting the parameter estimates from an ARMA(p, q) fit on the data into the formula for the spectral density $f_x(\omega)$ given in (4.15). Such an estimator is called a parametric spectral estimator. For convenience, a parametric spectral estimator is obtained by fitting an AR(p) to the data, where the order p is determined by one of the model selection criteria, such as AIC, AICc, and BIC, defined in (2.19)-(2.21). Parametric autoregressive spectral estimators will often have superior resolution in problems when several closely spaced narrow

spectral peaks are present and are preferred by engineers for a broad variety of problems (see Kay, 1988). The development of autoregressive spectral estimators has been summarized by Parzen (1983).

If $\widehat{\phi}_1, \widehat{\phi}_2, \ldots, \widehat{\phi}_p$ and $\widehat{\sigma}_w^2$ are the estimates from an AR(p) fit to x_t, then based on Property 4.3, a parametric spectral estimate of $f_x(\omega)$ is attained by substituting these estimates into (4.15), that is,

$$\widehat{f}_x(\omega) = \frac{\widehat{\sigma}_w^2}{|\widehat{\phi}(\mathrm{e}^{-2\pi i\omega})|^2}, \tag{4.75}$$

where

$$\widehat{\phi}(z) = 1 - \widehat{\phi}_1 z - \widehat{\phi}_2 z^2 - \cdots - \widehat{\phi}_p z^p. \tag{4.76}$$

The asymptotic distribution of the autoregressive spectral estimator has been obtained by Berk (1974) under the conditions $p \to \infty$, $p^3/n \to 0$ as $p,\ n \to \infty$, which may be too severe for most applications. The limiting results imply a confidence interval of the form

$$\frac{\widehat{f}_x(\omega)}{(1 + Cz_{\alpha/2})} \le f_x(\omega) \le \frac{\widehat{f}_x(\omega)}{(1 - Cz_{\alpha/2})}, \tag{4.77}$$

where $C = \sqrt{2p/n}$ and $z_{\alpha/2}$ is the ordinate corresponding to the upper $\alpha/2$ probability of the standard normal distribution. If the sampling distribution is to be checked, we suggest applying the bootstrap estimator to get the sampling distribution of $\widehat{f}_x(\omega)$ using a procedure similar to the one used for $p = 1$ in Example 3.35. An alternative for higher order autoregressive series is to put the AR(p) in state-space form and use the bootstrap procedure discussed in §6.7.

An interesting fact about rational spectra of the form (4.15) is that any spectral density can be approximated, arbitrarily close, by the spectrum of an AR process.

Property 4.5 AR Spectral Approximation

Let $g(\omega)$ be the spectral density of a stationary process. Then, given $\epsilon > 0$, there is a time series with the representation

$$x_t = \sum_{k=1}^{p} \phi_k x_{t-k} + w_t$$

where w_t is white noise with variance σ_w^2, such that

$$|f_x(\omega) - g(\omega)| < \epsilon \quad \text{forall } \omega \in [-1/2, 1/2].$$

Moreover, p is finite and the roots of $\phi(z) = 1 - \sum_{k=1}^{p} \phi_k z^k$ are outside the unit circle.

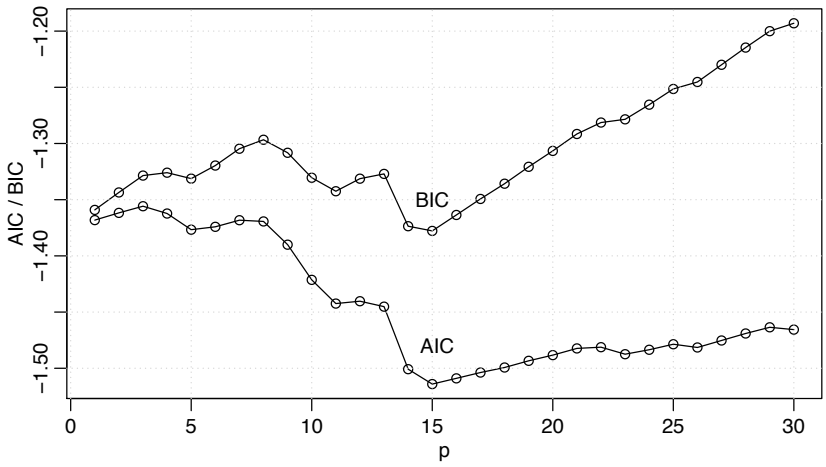

Fig. 4.11. Model selection criteria AIC and BIC as a function of order p for autoregressive models fitted to the SOI series.

One drawback of the property is that it does not tell us how large p must be before the approximation is reasonable; in some situations p may be extremely large. Property 4.5 also holds for MA and for ARMA processes in general, and a proof of the result may be found in Fuller (1996, Ch 4). We demonstrate the technique in the following example.

Example 4.15 Autoregressive Spectral Estimator for SOI

Consider obtaining results comparable to the nonparametric estimators shown in Figure 4.5 for the SOI series. Fitting successively higher order AR(p) models for $p = 1, 2, \ldots, 30$ yields a minimum BIC and a minimum AIC at $p = 15$, as shown in Figure 4.11. We can see from Figure 4.11 that BIC is very definite about which model it chooses; that is, the minimum BIC is very distinct. On the other hand, it is not clear what is going to happen with AIC; that is, the minimum is not so clear, and there is some concern that AIC will start decreasing after $p = 30$. Minimum AICc selects the $p = 15$ model, but suffers from the same uncertainty as AIC. The spectrum is shown in Figure 4.12, and we note the strong peaks at 52 months and 12 months corresponding to the nonparametric estimators obtained in §4.5. In addition, the harmonics of the yearly period are evident in the estimated spectrum.

To perform a similar analysis in R, the command `spec.ar` can be used to fit the best model via AIC and plot the resulting spectrum. A quick way to obtain the AIC values is to run the `ar` command as follows.

```
spaic = spec.ar(soi, log="no")     # min AIC spec
abline(v=frequency(soi)*1/52, lty="dotted")  # El Nino Cycle
(soi.ar = ar(soi, order.max=30))  # estimates and AICs
dev.new()
plot(1:30, soi.ar$aic[-1], type="o")   # plot AICs
```

R works only with the AIC in this case. To generate Figure 4.11 we used the following code to obtain AIC, AICc, and BIC. Because AIC and AICc

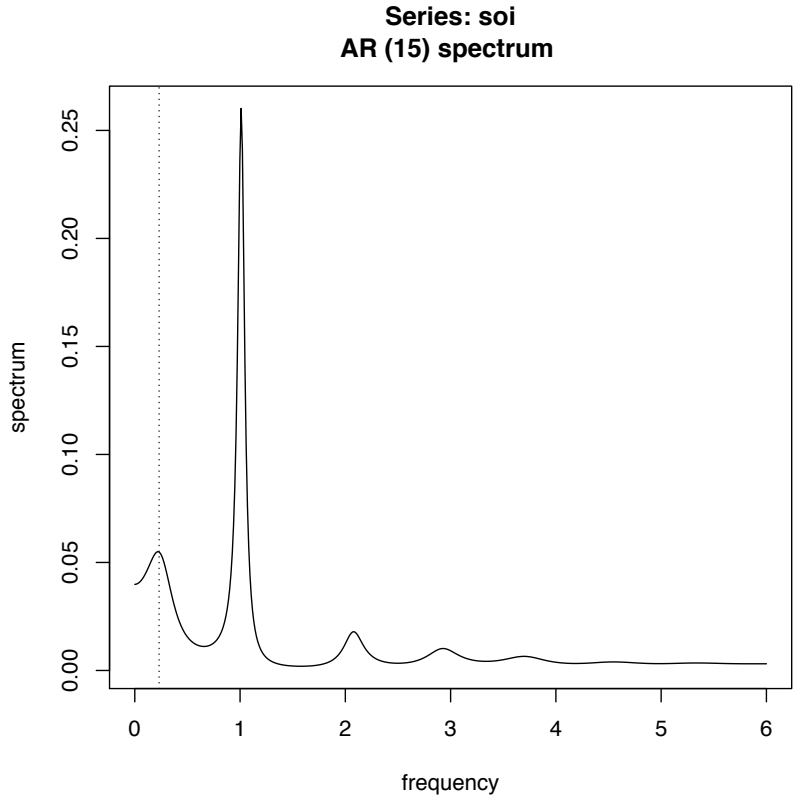

Fig. 4.12. Autoregressive spectral estimator for the SOI series using the AR(15) model selected by AIC, AICc, and BIC. The first peak (marked by a vertical dotted line) corresponds to the El Niño period of 52 months.

are nearly identical in this example, we only graphed AIC and BIC+1; we added 1 to the BIC to reduce white space in the graphic.

```
n = length(soi)
AIC = rep(0, 30) -> AICc -> BIC
for (k in 1:30){
 sigma2 = ar(soi, order=k, aic=FALSE)$var.pred
 BIC[k] = log(sigma2) + (k*log(n)/n)
 AICc[k] = log(sigma2) + ((n+k)/(n-k-2))
 AIC[k] = log(sigma2) + ((n+2*k)/n)
}
IC = cbind(AIC, BIC+1)
ts.plot(IC, type="o", xlab="p", ylab="AIC / BIC")
```

Finally, it should be mentioned that any parametric spectrum, say $f(\omega;\boldsymbol{\theta})$, depending on the vector parameter $\boldsymbol{\theta}$ can be estimated via the Whittle likelihood (Whittle, 1961), using the approximate properties of the discrete Fourier transform derived in Appendix C. We have that the DFTs, $d(\omega_j)$, are approximately complex normally distributed with mean zero and variance $f(\omega_j;\boldsymbol{\theta})$ and are approximately independent for $\omega_j \neq \omega_k$. This implies that an approximate log likelihood can be written in the form

$$\ln L(\boldsymbol{x};\boldsymbol{\theta}) \approx -\sum_{0<\omega_j<1/2}\left(\ln f_x(\omega_j;\boldsymbol{\theta}) + \frac{|d(\omega_j)|^2}{f_x(\omega_j;\boldsymbol{\theta})}\right), \tag{4.78}$$

where the sum is sometimes expanded to include the frequencies $\omega_j = 0, 1/2$. If the form with the two additional frequencies is used, the multiplier of the sum will be unity, except for the purely real points at $\omega_j = 0, 1/2$ for which the multiplier is $1/2$. For a discussion of applying the Whittle approximation to the problem of estimating parameters in an ARMA spectrum, see Anderson (1978). The Whittle likelihood is especially useful for fitting long memory models that will be discussed in Chapter 5.

4.7 Multiple Series and Cross-Spectra

The notion of analyzing frequency fluctuations using classical statistical ideas extends to the case in which there are several jointly stationary series, for example, x_t and y_t. In this case, we can introduce the idea of a correlation indexed by frequency, called the coherence. The results in Appendix C, §C.2, imply the covariance function

$$\gamma_{xy}(h) = E[(x_{t+h} - \mu_x)(y_t - \mu_y)]$$

has the representation

$$\gamma_{xy}(h) = \int_{-1/2}^{1/2} f_{xy}(\omega)\mathrm{e}^{2\pi i\omega h}\, d\omega \quad h = 0, \pm 1, \pm 2, ..., \tag{4.79}$$

where the cross-spectrum is defined as the Fourier transform

$$f_{xy}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h)\ \mathrm{e}^{-2\pi i\omega h} \quad -1/2 \le \omega \le 1/2, \tag{4.80}$$

assuming that the cross-covariance function is absolutely summable, as was the case for the autocovariance. The cross-spectrum is generally a complex-valued function, and it is often written as[14]

$$f_{xy}(\omega) = c_{xy}(\omega) - iq_{xy}(\omega), \tag{4.81}$$

where

$$c_{xy}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h)\ \cos(2\pi\omega h) \tag{4.82}$$

and

[14] For this section, it will be useful to recall the facts $e^{-i\alpha} = \cos(\alpha) - i\sin(\alpha)$ and if $z = a + ib$, then $\overline{z} = a - ib$.

$$q_{xy}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h)\ \sin(2\pi\omega h) \tag{4.83}$$

are defined as the cospectrum and quadspectrum, respectively. Because of the relationship $\gamma_{yx}(h) = \gamma_{xy}(-h)$, it follows, by substituting into (4.80) and rearranging, that

$$f_{yx}(\omega) = \overline{f_{xy}(\omega)}. \tag{4.84}$$

This result, in turn, implies that the cospectrum and quadspectrum satisfy

$$c_{yx}(\omega) = c_{xy}(\omega) \tag{4.85}$$

and

$$q_{yx}(\omega) = -q_{xy}(\omega). \tag{4.86}$$

An important example of the application of the cross-spectrum is to the problem of predicting an output series y_t from some input series x_t through a linear filter relation such as the three-point moving average considered below. A measure of the strength of such a relation is the squared coherence function, defined as

$$\rho^2_{y\cdot x}(\omega) = \frac{|f_{yx}(\omega)|^2}{f_{xx}(\omega) f_{yy}(\omega)}, \tag{4.87}$$

where $f_{xx}(\omega)$ and $f_{yy}(\omega)$ are the individual spectra of the x_t and y_t series, respectively. Although we consider a more general form of this that applies to multiple inputs later, it is instructive to display the single input case as (4.87) to emphasize the analogy with conventional squared correlation, which takes the form

$$\rho^2_{yx} = \frac{\sigma^2_{yx}}{\sigma^2_x \sigma^2_y},$$

for random variables with variances σ^2_x and σ^2_y and covariance $\sigma_{yx} = \sigma_{xy}$. This motivates the interpretation of squared coherence and the squared correlation between two time series at frequency ω.

Example 4.16 Three-Point Moving Average

As a simple example, we compute the cross-spectrum between x_t and the three-point moving average $y_t = (x_{t-1}+x_t+x_{t+1})/3$, where x_t is a stationary input process with spectral density $f_{xx}(\omega)$. First,

$$\begin{aligned}
\gamma_{xy}(h) &= \operatorname{cov}(x_{t+h}, y_t) = \tfrac{1}{3}\operatorname{cov}(x_{t+h}, x_{t-1} + x_t + x_{t+1}) \\
&= \frac{1}{3}\big(\gamma_{xx}(h+1) + \gamma_{xx}(h) + \gamma_{xx}(h-1)\big) \\
&= \frac{1}{3}\int_{-1/2}^{1/2} \big(\mathrm{e}^{2\pi i\omega} + 1 + \mathrm{e}^{-2\pi i\omega}\big)\, \mathrm{e}^{2\pi i\omega h} f_{xx}(\omega)\, d\omega \\
&= \frac{1}{3}\int_{-1/2}^{1/2} [1 + 2\cos(2\pi\omega)] f_{xx}(\omega) \mathrm{e}^{2\pi i\omega h}\, d\omega,
\end{aligned}$$

where we have use (4.11). Using the uniqueness of the Fourier transform, we argue from the spectral representation (4.79) that

$$f_{xy}(\omega) = \tfrac{1}{3}\left[1 + 2\cos(2\pi\omega)\right] f_{xx}(\omega)$$

so that the cross-spectrum is real in this case. From Example 4.5, the spectral density of y_t is

$$\begin{aligned} f_{yy}(\omega) &= \tfrac{1}{9}[3 + 4\cos(2\pi\omega) + 2\cos(4\pi\omega)] f_{xx}(\omega) \\ &= \tfrac{1}{9}\left[1 + 2\cos(2\pi\omega)\right]^2 f_{xx}(\omega), \end{aligned}$$

using the identity $\cos(2\alpha) = 2\cos^2(\alpha) - 1$ in the last step. Substituting into (4.87) yields the squared coherence between x_t and y_t as unity over all frequencies. This is a characteristic inherited by more general linear filters, as will be shown in Problem 4.23. However, if some noise is added to the three-point moving average, the coherence is not unity; these kinds of models will be considered in detail later.

Property 4.6 Spectral Representation of a Vector Stationary Process

If the elements of the $p \times p$ autocovariance function matrix

$$\Gamma(h) = E[(\boldsymbol{x}_{t+h} - \boldsymbol{\mu})(\boldsymbol{x}_t - \boldsymbol{\mu})']$$

of a p-dimensional stationary time series, $\boldsymbol{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tp})'$, has elements satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma_{jk}(h)| < \infty \tag{4.88}$$

for all $j, k = 1, \ldots, p$, then $\Gamma(h)$ has the representation

$$\Gamma(h) = \int_{-1/2}^{1/2} \mathrm{e}^{2\pi i\omega h} f(\omega)\, d\omega \quad h = 0, \pm 1, \pm 2, ..., \tag{4.89}$$

as the inverse transform of the spectral density matrix, $f(\omega) = \{f_{jk}(\omega)\}$, for $j, k = 1, \ldots, p$, with elements equal to the cross-spectral components. The matrix $f(\omega)$ has the representation

$$f(\omega) = \sum_{h=-\infty}^{\infty} \Gamma(h)\mathrm{e}^{-2\pi i\omega h} \quad -1/2 \le \omega \le 1/2. \tag{4.90}$$

Example 4.17 Spectral Matrix of a Bivariate Process

Consider a jointly stationary bivariate process (x_t, y_t). We arrange the autocovariances in the matrix

$$\Gamma(h) = \begin{pmatrix} \gamma_{xx}(h) & \gamma_{xy}(h) \\ \gamma_{yx}(h) & \gamma_{yy}(h) \end{pmatrix}.$$

The spectral matrix would be given by

$$f(\omega) = \begin{pmatrix} f_{xx}(\omega) & f_{xy}(\omega) \\ f_{yx}(\omega) & f_{yy}(\omega) \end{pmatrix},$$

where the Fourier transform (4.89) and (4.90) relate the autocovariance and spectral matrices.

The extension of spectral estimation to vector series is fairly obvious. For the vector series $\boldsymbol{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tp})'$, we may use the vector of DFTs, say $\boldsymbol{d}(\omega_j) = (d_1(\omega_j), d_2(\omega_j), \ldots, d_p(\omega_j))'$, and estimate the spectral matrix by

$$\bar{f}(\omega) = L^{-1} \sum_{k=-m}^{m} I(\omega_j + k/n) \tag{4.91}$$

where now

$$I(\omega_j) = \boldsymbol{d}(\omega_j)\, \boldsymbol{d}^*(\omega_j) \tag{4.92}$$

is a $p \times p$ complex matrix.[15]

Again, the series may be tapered before the DFT is taken in (4.91) and we can use weighted estimation,

$$\widehat{f}(\omega) = \sum_{k=-m}^{m} h_k\, I(\omega_j + k/n) \tag{4.93}$$

where $\{h_k\}$ are weights as defined in (4.56). The estimate of squared coherence between two series, y_t and x_t is

$$\widehat{\rho}^2_{y\cdot x}(\omega) = \frac{|\widehat{f}_{yx}(\omega)|^2}{\widehat{f}_{xx}(\omega)\widehat{f}_{yy}(\omega)}. \tag{4.94}$$

If the spectral estimates in (4.94) are obtained using equal weights, we will write $\bar{\rho}^2_{y\cdot x}(\omega)$ for the estimate.

Under general conditions, if $\rho^2_{y\cdot x}(\omega) > 0$ then

$$|\widehat{\rho}_{y\cdot x}(\omega)| \sim AN\left(|\rho_{y\cdot x}(\omega)|, \left(1 - \rho^2_{y\cdot x}(\omega)\right)^2 / 2L_h\right) \tag{4.95}$$

where L_h is defined in (4.57); the details of this result may be found in Brockwell and Davis (1991, Ch 11). We may use (4.95) to obtain approximate confidence intervals for the squared coherency $\rho^2_{y\cdot x}(\omega)$.

We can test the hypothesis that $\rho^2_{y\cdot x}(\omega) = 0$ if we use $\bar{\rho}^2_{y\cdot x}(\omega)$ for the estimate with $L > 1$,[16] that is,

[15] If Z is a complex matrix, then $Z^* = \overline{Z}'$ denotes the conjugate transpose operation. That is, Z^* is the result of replacing each element of Z by its complex conjugate and transposing the resulting matrix.

[16] If $L = 1$ then $\bar{\rho}^2_{y\cdot x}(\omega) \equiv 1$.

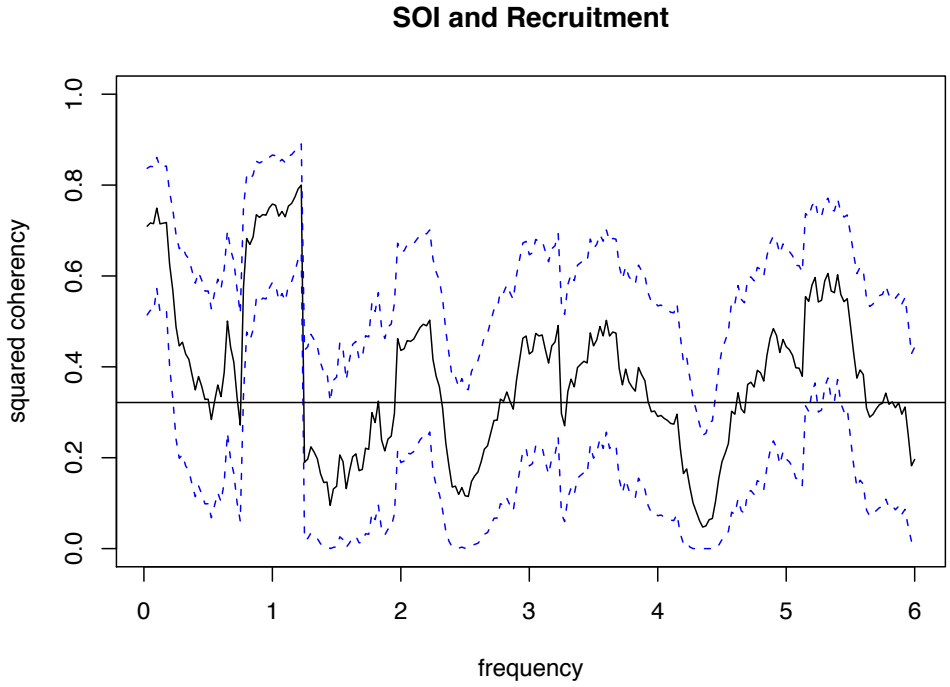

Fig. 4.13. Squared coherency between the SOI and Recruitment series; $L = 19$, $n = 453$, $n' = 480$, and $\alpha = .001$. The horizontal line is $C_{.001}$.

$$\bar{\rho}^2_{y\cdot x}(\omega) = \frac{|\bar{f}_{yx}(\omega)|^2}{\bar{f}_{xx}(\omega)\bar{f}_{yy}(\omega)}. \tag{4.96}$$

In this case, under the null hypothesis, the statistic

$$F = \frac{\bar{\rho}^2_{y\cdot x}(\omega)}{(1-\bar{\rho}^2_{y\cdot x}(\omega))}(L-1) \tag{4.97}$$

has an approximate F-distribution with 2 and $2L-2$ degrees of freedom. When the series have been extended to length n', we replace $2L-2$ by $df-2$, where df is defined in (4.52). Solving (4.97) for a particular significance level α leads to

$$C_\alpha = \frac{F_{2,2L-2}(\alpha)}{L-1+F_{2,2L-2}(\alpha)} \tag{4.98}$$

as the approximate value that must be exceeded for the original squared coherence to be able to reject $\rho^2_{y\cdot x}(\omega) = 0$ at an a priori specified frequency.

Example 4.18 Coherence Between SOI and Recruitment

Figure 4.13 shows the squared coherence between the SOI and Recruitment series over a wider band than was used for the spectrum. In this case, we used $L = 19$, $df = 2(19)(453/480) \approx 36$ and $F_{2,df-2}(.001) \approx 8.53$ at the significance level $\alpha = .001$. Hence, we may reject the hypothesis of no coherence for values of $\bar{\rho}^2_{y\cdot x}(\omega)$ that exceed $C_{.001} = .32$. We emphasize that this method

is crude because, in addition to the fact that the F-statistic is approximate, we are examining the squared coherence across all frequencies with the Bonferroni inequality, (4.55), in mind. Figure 4.13 also exhibits confidence bands as part of the R plotting routine. We emphasize that these bands are only valid for ω where $\rho^2_{y\cdot x}(\omega) > 0$.

In this case, the seasonal frequency and the El Niño frequencies ranging between about 3 and 7 year periods are strongly coherent. Other frequencies are also strongly coherent, although the strong coherence is less impressive because the underlying power spectrum at these higher frequencies is fairly small. Finally, we note that the coherence is persistent at the seasonal harmonic frequencies.

This example may be reproduced using the following R commands.

```
sr=spec.pgram(cbind(soi,rec),kernel("daniell",9),taper=0,plot=FALSE)
sr$df                    # df = 35.8625
f = qf(.999, 2, sr$df-2)  # = 8.529792
C = f/(18+f)              # = 0.318878
plot(sr, plot.type = "coh", ci.lty = 2)
abline(h = C)
```

4.8 Linear Filters

Some of the examples of the previous sections have hinted at the possibility the distribution of power or variance in a time series can be modified by making a linear transformation. In this section, we explore that notion further by defining a linear filter and showing how it can be used to extract signals from a time series. The linear filter modifies the spectral characteristics of a time series in a predictable way, and the systematic development of methods for taking advantage of the special properties of linear filters is an important topic in time series analysis.

A linear filter uses a set of specified coefficients a_j, for $j = 0, \pm1, \pm2, \ldots$, to transform an input series, x_t, producing an output series, y_t, of the form

$$y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |a_j| < \infty. \tag{4.99}$$

The form (4.99) is also called a convolution in some statistical contexts. The coefficients, collectively called the *impulse response function*, are required to satisfy absolute summability so y_t in (4.99) exists as a limit in mean square and the infinite Fourier transform

$$A_{yx}(\omega) = \sum_{j=-\infty}^{\infty} a_j \, \mathrm{e}^{-2\pi i \omega j}, \tag{4.100}$$

called the *frequency response function*, is well defined. We have already encountered several linear filters, for example, the simple three-point moving

average in Example 4.16, which can be put into the form of (4.99) by letting $a_{-1} = a_0 = a_1 = 1/3$ and taking $a_t = 0$ for $|j| \geq 2$.

The importance of the linear filter stems from its ability to enhance certain parts of the spectrum of the input series. To see this, assuming that x_t is stationary with spectral density $f_{xx}(\omega)$, the autocovariance function of the filtered output y_t in (4.99) can be derived as

$$\begin{aligned}
\gamma_{yy}(h) &= \operatorname{cov}(y_{t+h}, y_t) \\
&= \operatorname{cov}\left(\sum_r a_r x_{t+h-r}, \sum_s a_s x_{t-s}\right) \\
&= \sum_r \sum_s a_r \gamma_{xx}(h-r+s) a_s \\
&= \sum_r \sum_s a_r \left[\int_{-1/2}^{1/2} \mathrm{e}^{2\pi i\omega(h-r+s)} f_{xx}(\omega) d\omega\right] a_s \\
&= \int_{-1/2}^{1/2} \left(\sum_r a_r \mathrm{e}^{-2\pi i\omega r}\right)\left(\sum_s a_s \mathrm{e}^{2\pi i\omega s}\right) \mathrm{e}^{2\pi i\omega h} f_{xx}(\omega)\ d\omega \\
&= \int_{-1/2}^{1/2} \mathrm{e}^{2\pi i\omega h} |A_{yx}(\omega)|^2 f_{xx}(\omega)\ d\omega,
\end{aligned}$$

where we have first replaced $\gamma_{xx}(\cdot)$ by its representation (4.11) and then substituted $A_{yx}(\omega)$ from (4.100). The computation is one we do repeatedly, exploiting the uniqueness of the Fourier transform. Now, because the left-hand side is the Fourier transform of the spectral density of the output, say, $f_{yy}(\omega)$, we get the important filtering property as follows.

Property 4.7 Output Spectrum of a Filtered Stationary Series

The spectrum of the filtered output y_t in (4.99) is related to the spectrum of the input x_t by

$$f_{yy}(\omega) = |A_{yx}(\omega)|^2\ f_{xx}(\omega), \tag{4.101}$$

where the frequency response function $A_{yx}(\omega)$ is defined in (4.100).

The result (4.101) enables us to calculate the exact effect on the spectrum of any given filtering operation. This important property shows the spectrum of the input series is changed by filtering and the effect of the change can be characterized as a frequency-by-frequency multiplication by the squared magnitude of the frequency response function. Again, an obvious analogy to a property of the variance in classical statistics holds, namely, if x is a random variable with variance σ_x^2, then $y = ax$ will have variance $\sigma_y^2 = a^2\sigma_x^2$, so the variance of the linearly transformed random variable is changed by multiplication by a^2 in much the same way as the linearly filtered spectrum is changed in (4.101).

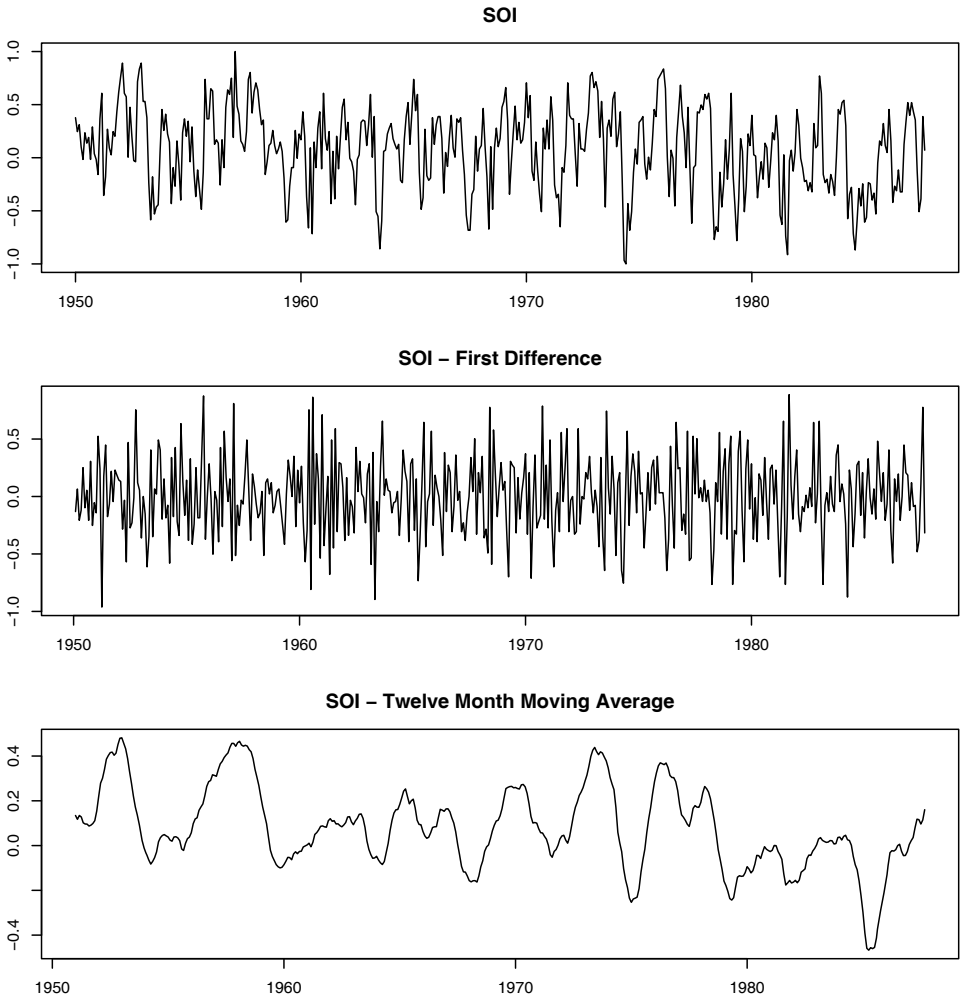

Fig. 4.14. SOI series (top) compared with the differenced SOI (middle) and a centered 12-month moving average (bottom).

Finally, we mention that Property 4.3, which was used to get the spectrum of an ARMA process, is just a special case of Property 4.7 where in (4.99), $x_t = w_t$ is white noise, in which case $f_{xx}(\omega) = \sigma_w^2$, and $a_j = \psi_j$, in which case

$$A_{yx}(\omega) = \psi(\mathrm{e}^{-2\pi i\omega}) = \theta(\mathrm{e}^{-2\pi i\omega}) \,/\, \phi(\mathrm{e}^{-2\pi i\omega}).$$

Example 4.19 First Difference and Moving Average Filters

We illustrate the effect of filtering with two common examples, the first difference filter

$$y_t = \nabla x_t = x_t - x_{t-1}$$

and the symmetric moving average filter

$$y_t = \tfrac{1}{24}\big(x_{t-6} + x_{t+6}\big) + \tfrac{1}{12}\sum_{r=-5}^{5} x_{t-r},$$

Fig. 4.15. Spectral analysis of SOI after applying a 12-month moving average filter. The vertical line corresponds to the 52-month cycle.

which is a modified Daniell kernel with $m = 6$. The results of filtering the SOI series using the two filters are shown in the middle and bottom panels of Figure 4.14. Notice that the effect of differencing is to roughen the series because it tends to retain the higher or faster frequencies. The centered moving average smoothes the series because it retains the lower frequencies and tends to attenuate the higher frequencies. In general, differencing is an example of a *high-pass filter* because it retains or passes the higher frequencies, whereas the moving average is a *low-pass filter* because it passes the lower or slower frequencies.

Notice that the slower periods are enhanced in the symmetric moving average and the seasonal or yearly frequencies are attenuated. The filtered series makes about 9 cycles in the length of the data (about one cycle every 52 months) and the moving average filter tends to enhance or extract the signal that is associated with El Niño. Moreover, by the low-pass filtering of the data, we get a better sense of the El Niño effect and its irregularity. Figure 4.15 shows the results of a spectral analysis on the low-pass filtered SOI series. It is clear that all high frequency behavior has been removed and the El Niño cycle is accentuated; the dotted vertical line in the figure corresponds to the 52 months cycle.

Now, having done the filtering, it is essential to determine the exact way in which the filters change the input spectrum. We shall use (4.100) and (4.101) for this purpose. The first difference filter can be written in the form (4.99) by letting $a_0 = 1, a_1 = -1$, and $a_r = 0$ otherwise. This implies that

$$A_{yx}(\omega) = 1 - \mathrm{e}^{-2\pi i\omega},$$

and the squared frequency response becomes

$$|A_{yx}(\omega)|^2 = (1 - \mathrm{e}^{-2\pi i\omega})(1 - \mathrm{e}^{2\pi i\omega}) = 2[1 - \cos(2\pi\omega)]. \tag{4.102}$$

The top panel of Figure 4.16 shows that the first difference filter will attenuate the lower frequencies and enhance the higher frequencies because the multiplier of the spectrum, $|A_{yx}(\omega)|^2$, is large for the higher frequencies and small for the lower frequencies. Generally, the slow rise of this kind of filter does not particularly recommend it as a procedure for retaining only the high frequencies.

For the centered 12-month moving average, we can take $a_{-6} = a_6 = 1/24$, $a_k = 1/12$ for $-5 \le k \le 5$ and $a_k = 0$ elsewhere. Substituting and recognizing the cosine terms gives

$$A_{yx}(\omega) = \tfrac{1}{12}\Big[1 + \cos(12\pi\omega) + 2\sum_{k=1}^{5}\cos(2\pi\omega k)\Big]. \tag{4.103}$$

Plotting the squared frequency response of this function as in Figure 4.16 shows that we can expect this filter to cut most of the frequency content above .05 cycles per point. This corresponds to eliminating periods shorter than $T = 1/.05 = 20$ points. In particular, this drives down the yearly components with periods of $T = 12$ months and enhances the El Niño frequency, which is somewhat lower. The filter is not completely efficient at attenuating high frequencies; some power contributions are left at higher frequencies, as shown in the function $|A_{yx}(\omega)|^2$ and in the spectrum of the moving average shown in Figure 4.3.

The following R session shows how to filter the data, perform the spectral analysis of this example, and plot the squared frequency response curve of the difference filter.

```
par(mfrow=c(3,1))
plot(soi)  # plot data
plot(diff(soi))  # plot first difference
k = kernel("modified.daniell", 6)  # filter weights
plot(soif <- kernapply(soi, k))    # plot 12 month filter
dev.new()
spectrum(soif, spans=9, log="no")  # spectral analysis
abline(v=12/52, lty="dashed")
dev.new()
w = seq(0, .5, length=500)  # frequency response
FR = abs(1-exp(2i*pi*w))^2
plot(w, FR, type="l")
```

The two filters discussed in the previous example were different in that the frequency response function of the first difference was complex-valued, whereas the frequency response of the moving average was purely real. A

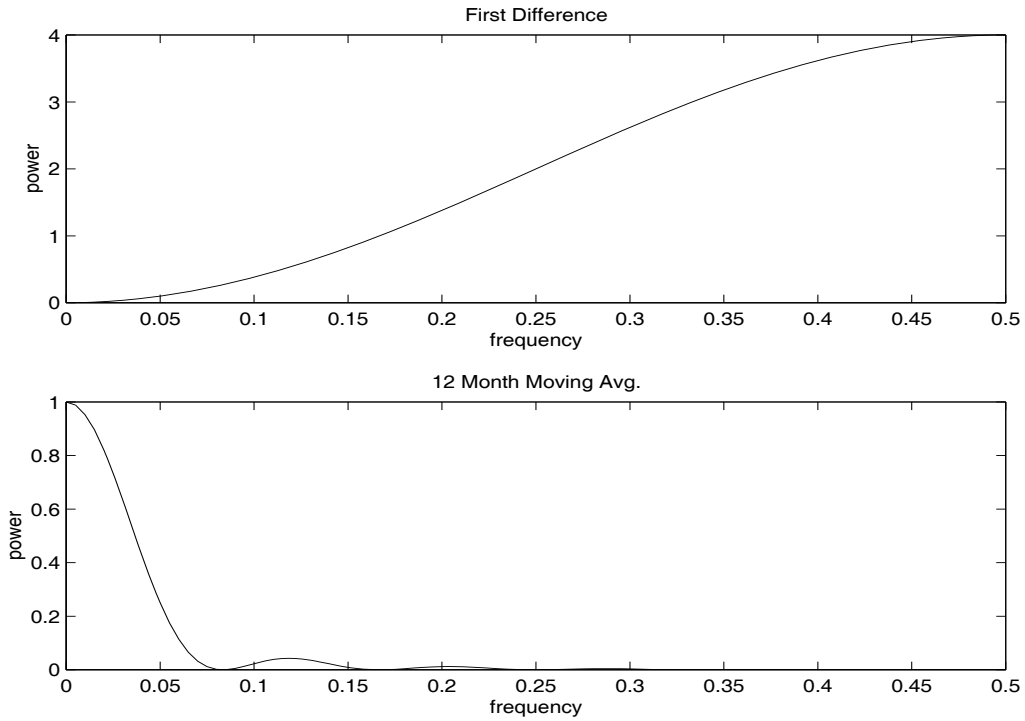

Fig. 4.16. Squared frequency response functions of the first difference and 12-month moving average filters.

short derivation similar to that used to verify (4.101) shows, when x_t and y_t are related by the linear filter relation (4.99), the cross-spectrum satisfies

$$f_{yx}(\omega) = A_{yx}(\omega) f_{xx}(\omega),$$

so the frequency response is of the form

$$A_{yx}(\omega) = \frac{f_{yx}(\omega)}{f_{xx}(\omega)} \tag{4.104}$$

$$= \frac{c_{yx}(\omega)}{f_{xx}(\omega)} - i\frac{q_{yx}(\omega)}{f_{xx}(\omega)}, \tag{4.105}$$

where we have used (4.81) to get the last form. Then, we may write (4.105) in polar coordinates as

$$A_{yx}(\omega) = |A_{yx}(\omega)| \ \exp\{-i \ \phi_{yx}(\omega)\}, \tag{4.106}$$

where the amplitude and phase of the filter are defined by

$$|A_{yx}(\omega)| = \frac{\sqrt{c_{yx}^2(\omega) + q_{yx}^2(\omega)}}{f_{xx}(\omega)} \tag{4.107}$$

and

$$\phi_{yx}(\omega) = \tan^{-1}\left(-\frac{q_{yx}(\omega)}{c_{yx}(\omega)}\right). \tag{4.108}$$

A simple interpretation of the phase of a linear filter is that it exhibits time delays as a function of frequency in the same way as the spectrum represents the variance as a function of frequency. Additional insight can be gained by considering the simple delaying filter

$$y_t = Ax_{t-D},$$

where the series gets replaced by a version, amplified by multiplying by A and delayed by D points. For this case,

$$f_{yx}(\omega) = A\mathrm{e}^{-2\pi i\omega D} f_{xx}(\omega),$$

and the amplitude is $|A|$, and the phase is

$$\phi_{yx}(\omega) = -2\pi\omega D,$$

or just a linear function of frequency ω. For this case, applying a simple time delay causes phase delays that depend on the frequency of the periodic component being delayed. Interpretation is further enhanced by setting

$$x_t = \cos(2\pi\omega t),$$

in which case

$$y_t = A\cos(2\pi\omega t - 2\pi\omega D).$$

Thus, the output series, y_t, has the same period as the input series, x_t, but the amplitude of the output has increased by a factor of $|A|$ and the phase has been changed by a factor of $-2\pi\omega D$.

Example 4.20 Difference and Moving Average Filters

We consider calculating the amplitude and phase of the two filters discussed in Example 4.19. The case for the moving average is easy because $A_{yx}(\omega)$ given in (4.103) is purely real. So, the amplitude is just $|A_{yx}(\omega)|$ and the phase is $\phi_{yx}(\omega) = 0$. In general, symmetric ($a_j = a_{-j}$) filters have zero phase. The first difference, however, changes this, as we might expect from the example above involving the time delay filter. In this case, the squared amplitude is given in (4.102). To compute the phase, we write

$$\begin{aligned} A_{yx}(\omega) &= 1 - e^{-2\pi i\omega} = e^{-i\pi\omega}(e^{i\pi\omega} - e^{-i\pi\omega}) \\ &= 2ie^{-i\pi\omega}\sin(\pi\omega) = 2\sin^2(\pi\omega) + 2i\cos(\pi\omega)\sin(\pi\omega) \\ &= \frac{c_{yx}(\omega)}{f_{xx}(\omega)} - i\frac{q_{yx}(\omega)}{f_{xx}(\omega)}, \end{aligned}$$

so

$$\phi_{yx}(\omega) = \tan^{-1}\left(-\frac{q_{yx}(\omega)}{c_{yx}(\omega)}\right) = \tan^{-1}\left(\frac{\cos(\pi\omega)}{\sin(\pi\omega)}\right).$$

Noting that

$$\cos(\pi\omega) = \sin(-\pi\omega + \pi/2)$$

and that

$$\sin(\pi\omega) = \cos(-\pi\omega + \pi/2),$$

we get

$$\phi_{yx}(\omega) = -\pi\omega + \pi/2,$$

and the phase is again a linear function of frequency.

The above tendency of the frequencies to arrive at different times in the filtered version of the series remains as one of two annoying features of the difference type filters. The other weakness is the gentle increase in the frequency response function. If low frequencies are really unimportant and high frequencies are to be preserved, we would like to have a somewhat sharper response than is obvious in Figure 4.16. Similarly, if low frequencies are important and high frequencies are not, the moving average filters are also not very efficient at passing the low frequencies and attenuating the high frequencies. Improvement is possible by designing better and longer filters, but we do not discuss this here.

We will occasionally use results for multivariate series $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tp})'$ that are comparable to the simple property shown in (4.101). Consider the matrix filter

$$\boldsymbol{y}_t = \sum_{j=-\infty}^{\infty} A_j \boldsymbol{x}_{t-j}, \tag{4.109}$$

where $\{A_j\}$ denotes a sequence of $q \times p$ matrices such that $\sum_{j=-\infty}^{\infty} \|A_j\| < \infty$ and $\|\cdot\|$ denotes any matrix norm, $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tp})'$ is a $p \times 1$ stationary vector process with mean vector $\boldsymbol{\mu}_x$ and $p \times p$, matrix covariance function $\Gamma_{xx}(h)$ and spectral matrix $f_{xx}(\omega)$, and $\boldsymbol{y}_t$ is the $q \times 1$ vector output process. Then, we can obtain the following property.

Property 4.8 Output Spectral Matrix of a Linearly Filtered Stationary Vector Series

The spectral matrix of the filtered output $\boldsymbol{y}_t$ *in* (4.109) *is related to the spectrum of the input* $\boldsymbol{x}_t$ *by*

$$f_{yy}(\omega) = \mathcal{A}(\omega) f_{xx}(\omega) \mathcal{A}^*(\omega), \tag{4.110}$$

where the matrix frequency response function $\mathcal{A}(\omega)$ *is defined by*

$$\mathcal{A}(\omega) = \sum_{j=-\infty}^{\infty} A_j \exp(-2\pi i \omega j). \tag{4.111}$$

4.9 Dynamic Fourier Analysis and Wavelets

If a time series, x_t, is stationary, its second-order behavior remains the same, regardless of the time t. It makes sense to match a stationary time series with sines and cosines because they, too, behave the same forever. Indeed, based on the Spectral Representation Theorem (Appendix C, §C.1), we may regard a stationary series as the superposition of sines and cosines that oscillate at various frequencies. As seen in this text, however, many time series are not stationary. Typically, the data are coerced into stationarity via transformations, or we restrict attention to parts of the data where stationarity appears to adhere. In some cases, the nonstationarity of a time series is of interest. That is to say, it is the local behavior of the process, and not the global behavior of the process, that is of concern to the investigator. As a case in point, we mention the explosion and earthquake series first presented in Example 1.7 (see Figure 1.7). The following example emphasizes the importance of dynamic (or time-frequency) Fourier analysis.

Example 4.21 Dynamic Spectral Analysis of Seismic Traces

Consider the earthquake and explosion series displayed in Figure 1.7; it should be apparent that the dynamics of the series are changing with time. The goal of this analysis is to summarize the spectral behavior of the signal as it evolves over time.

First, a spectral analysis is performed on a short section of the data. Then, the section is shifted, and a spectral analysis is performed on the new section. This process is repeated until the end of the data, and the results are shown an image in Figure 4.17 and Figure 4.18; in the images, darker areas correspond to higher power. Specifically, in this example, let x_t, for $t = 1, \ldots, 2048$, represent the series of interest. Then, the sections of the data that were analyzed were $\{x_{t_k+1}, \ldots, x_{t_k+256}\}$, for $t_k = 128k$, and $k = 0, 1, \ldots, 14$; e.g., the first section analyzed is $\{x_1, \ldots, x_{256}\}$, the second section analyzed is $\{x_{129}, \ldots, x_{384}\}$, and so on. Each section of 256 observations was tapered using a cosine bell, and spectral estimation was performed using a repeated Daniell kernel with weights $\frac{1}{9}\{1, 2, 3, 2, 1\}$; see page 204. The sections overlap each other, however, this practice is not necessary and sometimes not desirable.[17]

The results of the dynamic analysis are shown as the estimated spectra for frequencies up to 10 Hz (the folding frequency is 20 Hz) for each starting location (time), $t_k = 128k$, with $k = 0, 1, \ldots, 14$. The S component for the

[17] A number of technical problems exist in this setting because the process of interest is nonstationary and we have not specified the nature of the nonstationarity. In addition, overlapping intervals complicate matters by introducing another layer of dependencies among the spectra. Consequently, the spectral estimates of contiguous sections are dependent in a non-trivial way that we have not specified. Nevertheless, as seen from this example, dynamic spectral analysis can be a helpful tool in summarizing the local behavior of a time series.

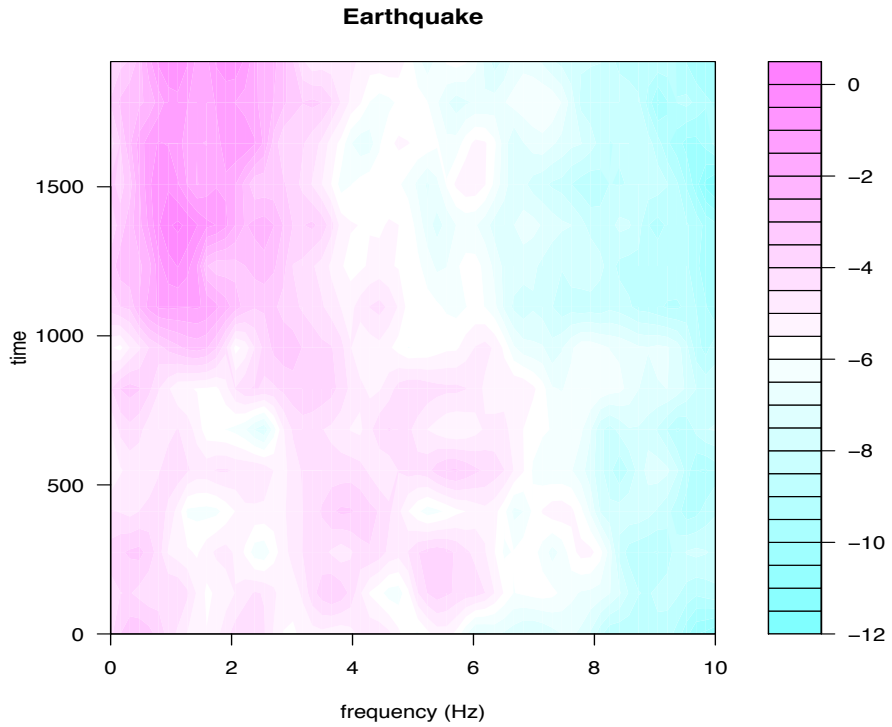

Fig. 4.17. Time-frequency image for the dynamic Fourier analysis of the earthquake series shown in Figure 1.7.

earthquake shows power at the low frequencies only, and the power remains strong for a long time. In contrast, the explosion shows power at higher frequencies than the earthquake, and the power of the signals (P and S waves) does not last as long as in the case of the earthquake.

The following is an R session that corresponds to the analysis of the explosion series. The images are generated using `filled.contour()` on the log of the power; ~~this, as well as using a gray scale and limiting the number of levels was done to produce a decent black-and-white graphic. The images look better in color, so we advise removing the `nlevels=...` and the `col=gray(...)` parts of the code.~~ We also include the code for obtaining a three-dimensional graphic to display the information, however, the graphic is not exhibited in the text.

```
nobs = length(EXP6)  # number of observations
wsize = 256   # window size
overlap = 128  # overlap
ovr = wsize-overlap
nseg = floor(nobs/ovr)-1;  # number of segments
krnl = kernel("daniell", c(1,1))  # kernel
ex.spec = matrix(0, wsize/2, nseg)
 for (k in 1:nseg) {
  a = ovr*(k-1)+1
```

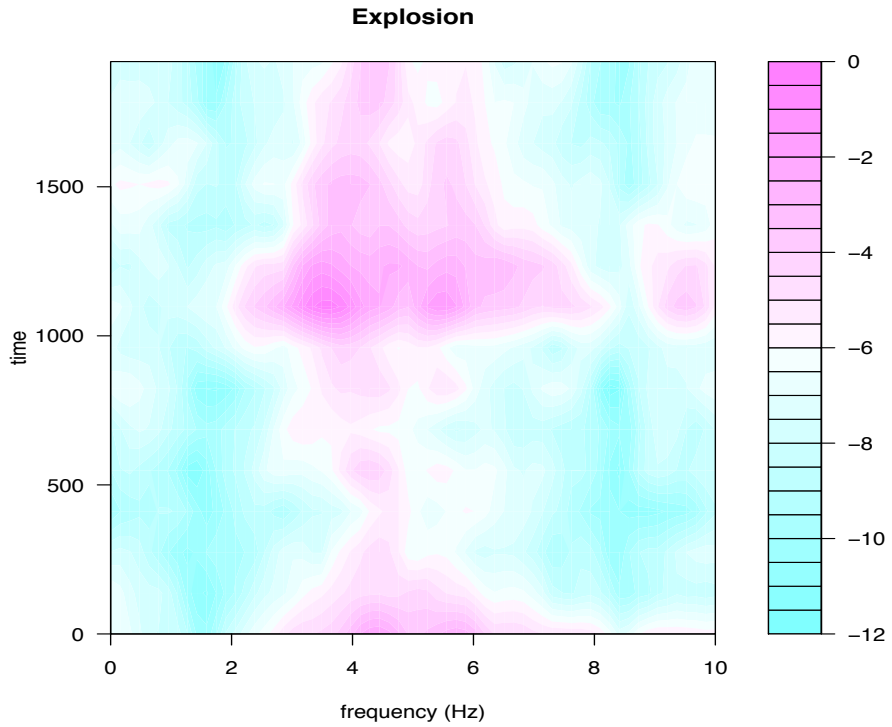

Fig. 4.18. Time-frequency image for the dynamic Fourier analysis of the explosion series shown in Figure 1.7.

```
  b = wsize+ovr*(k-1)
  ex.spec[,k] = spectrum(EXP6[a:b], krnl, taper=.5, plot=F)$spec }
x = seq(0, 10,  len = nrow(ex.spec)/2)
y = seq(0, ovr*nseg, len = ncol(ex.spec))
z = ex.spec[1:(nrow(ex.spec)/2),]
filled.contour(x, y, log(z), ylab="time", xlab="frequency (Hz)",
    main="Explosion")
persp(x, y, z, zlab="Power", xlab="frequency (Hz)", ylab="time",
    ticktype="detailed", theta=25,d=2, main="Explosion")  # not shown
```

One way to view the time-frequency analysis of Example 4.21 is to consider it as being based on local transforms of the data x_t of the form

$$d_{j,k} = n^{-1/2} \sum_{t=1}^{n} x_t \psi_{j,k}(t), \tag{4.112}$$

where

$$\psi_{j,k}(t) = \begin{cases} (n/m)^{1/2} h_t \, e^{-2\pi i t j/m} & t \in [t_k + 1, t_k + m], \\ 0 & \text{otherwise}, \end{cases} \tag{4.113}$$

where h_t is a taper and m is some fraction of n. In Example 4.21, $n = 2048$, $m = 256$, $t_k = 128k$, for $k = 0, 1, \ldots, 14$, and h_t was a cosine bell taper

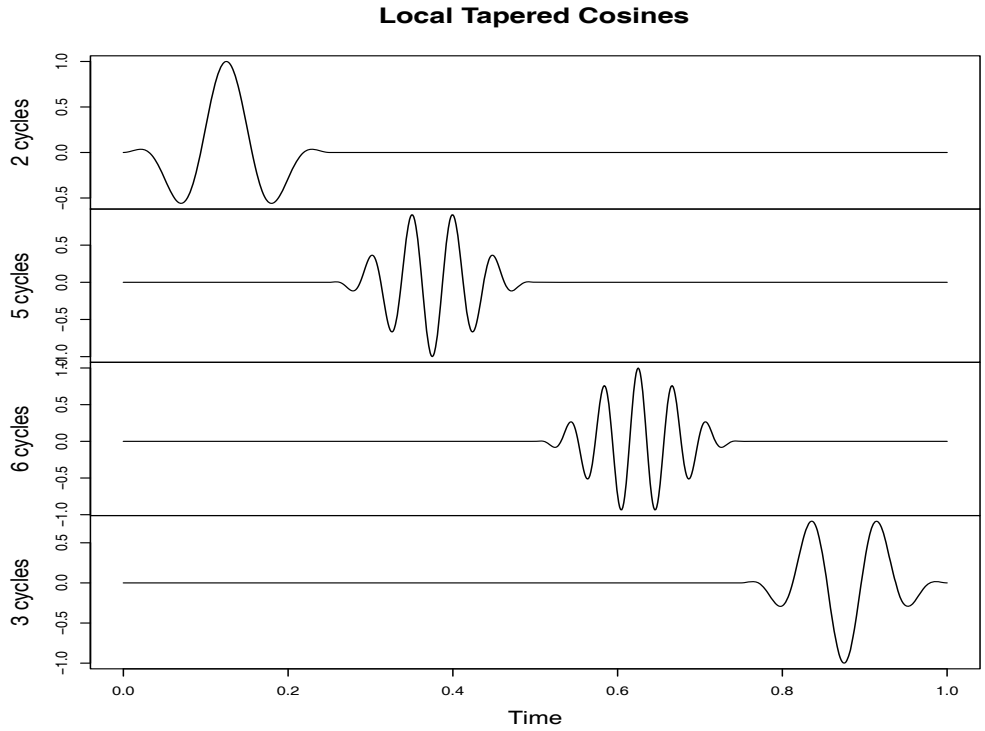

Fig. 4.19. Local, tapered cosines at various frequencies.

over 256 points. In (4.112) and (4.113), j indexes frequency, $\omega_j = j/m$, for $j = 1, 2, \ldots, [m/2]$, and k indexes the location, or time shift, of the transform. In this case, the transforms are based on tapered cosines and sines that have been zeroed out over various regions in time. The key point here is that the transforms are based on *local* sinusoids. Figure 4.19 shows an example of four local, tapered cosine functions at various frequencies. In that figure, the length of the data is considered to be one, and the cosines are localized to a fourth of the data length.

In addition to dynamic Fourier analysis as a method to overcome the restriction of stationarity, researchers have sought various alternative methods. A recent, and successful, alternative is wavelet analysis. The website `http://www.wavelet.org` is devoted to wavelets, which includes information about books, technical papers, software, and links to other sites. In addition, we mention the monograph on wavelets by Daubechies (1992), the text by Percival and Walden (2000), and we note that many statistical software manufacturers have wavelet modules that sit on top of their base package. In this section, we rely primarily on the S-PLUS wavelets module (with a manual written by Bruce and Gao, 1996), however, we will present some R code where possible. The basic idea of wavelet analysis is to imitate dynamic Fourier analysis, but with functions (wavelets) that may be better suited to capture the local behavior of nonstationary time series.

Wavelets come in families generated by a father wavelet, ϕ, and a mother wavelet, ψ. The father wavelets are used to capture the smooth, low-frequency nature of the data, whereas the mother wavelets are used to capture the detailed, and high-frequency nature of the data. The father wavelet integrates to one, and the mother wavelet integrates to zero

$$\int \phi(t)dt = 1 \quad \text{and} \quad \int \psi(t)dt = 0. \tag{4.114}$$

For a simple example, consider the Haar function,

$$\psi(t) = \begin{cases} 1, & 0 \le t < 1/2, \\ -1, & 1/2 \le t < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.115}$$

The father in this case is $\phi(t) = 1$ for $t \in [0, 1)$ and zero otherwise. The Haar functions are useful for demonstrating properties of wavelets, but they do not have good time-frequency localization properties. Figure 4.20 displays two of the more commonly used wavelets that are available with the S-PLUS wavelets module, the *daublet4* and *symmlet8* wavelets, which are described in detail in Daubechies (1992). The number after the name refers to the width and smoothness of the wavelet; for example, the symmlet10 wavelet is wider and smoother than the symmlet8 wavelet. Daublets are one of the first type of continuous orthogonal wavelets with compact support, and symmlets were constructed to be closer to symmetry than daublets. In general, wavelets do not have an analytical form, but instead they are generated using numerical methods.

Figure 4.20 was generated in S-PLUS using the wavelet module as follows:[18]

```
d4f <- wavelet("d4", mother=F)
d4m <- wavelet("d4")
s8f <- wavelet("s8", mother=F)
s8m <- wavelet("s8")
par(mfrow=c(2,2))
plot(d4f);  plot(d4m)
plot(s8f);  plot(s8m)
```

It is possible to draw some wavelets in R using the `wavethresh` package. In that package, daublets are called DaubExPhase and symmlets are called DaubLeAsymm. The following R session displays some of the available wavelets (this will produce a figure similar to Figure 4.20) and it assumes the `wavethresh` package has been downloaded and installed (see Appendix R, §R.2, for details on installing packages). The `filter.number` determines the width and smoothness of the wavelet.

[18] At this time, the R packages available for wavelet analysis are not extensive enough for our purposes, hence we will rely on S-PLUS for some of the demonstrations. We will provide R code when possible, and that will be based on the `wavethresh` package (version 4.2-1) that accompanies Nason (2008).

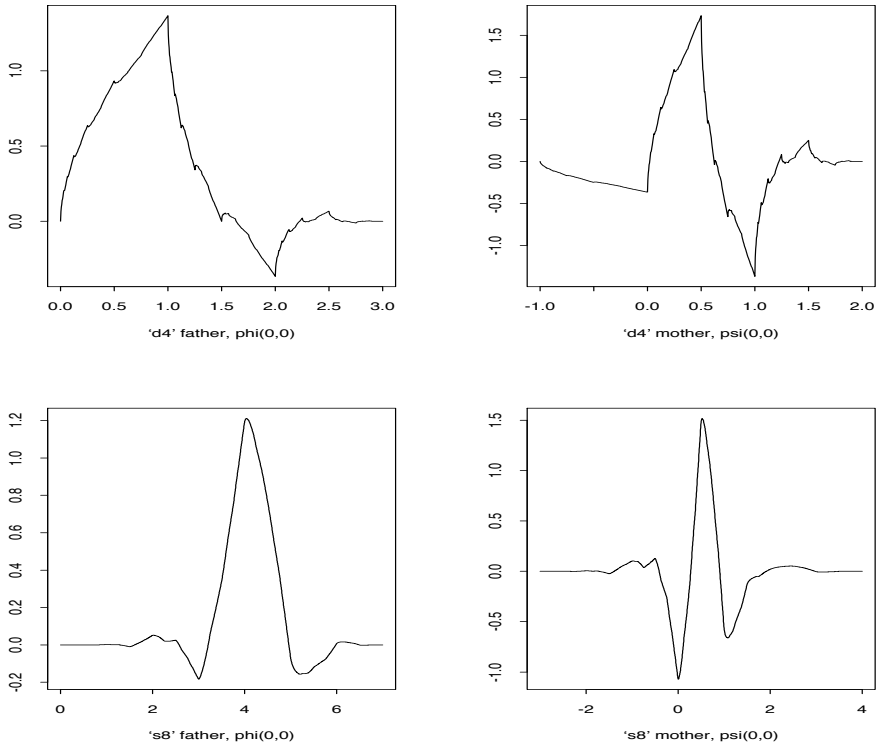

Fig. 4.20. Father and mother daublet4 wavelets (top row); father and mother symmlet8 wavelets (bottom row).

```
library(wavethresh)
par(mfrow=c(2,2))
draw(filter.number=4, family="DaubExPhase", enhance=FALSE, main="")
draw(filter.number=8, family="DaubExPhase", enhance=FALSE, main="")
draw(filter.number=4, family="DaubLeAsymm", enhance=FALSE, main="")
draw(filter.number=8, family="DaubLeAsymm", enhance=FALSE, main="")
```

When we depart from periodic functions, such as sines and cosines, the precise meaning of frequency, or cycles per unit time, is lost. When using wavelets, we typically refer to scale rather than frequency. The orthogonal wavelet decomposition of a time series, x_t, for $t = 1, \ldots, n$ is

$$\begin{aligned} x_t = &\sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) \\ &+ \sum_k d_{J-1,k}\psi_{J-1,k}(t) + \cdots + \sum_k d_{1,k}\psi_{1,k}(t), \end{aligned} \tag{4.116}$$

where J is the number of scales, and k ranges from one to the number of coefficients associated with the specified component (see Example 4.22). In (4.116), the wavelet functions $\phi_{J,k}(t), \psi_{J,k}(t), \psi_{J-1,k}(t), \ldots, \psi_{1,k}(t)$ are generated from the father wavelet, $\phi(t)$, and the mother wavelet, $\psi(t)$, by translation (shift) and scaling:

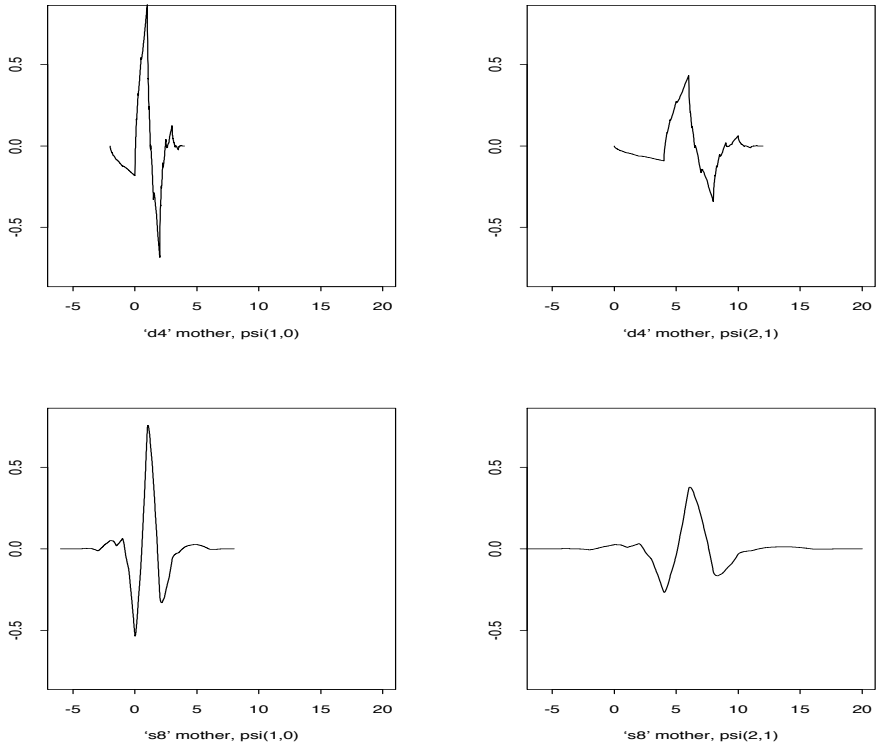

Fig. 4.21. Scaled and translated daublet4 wavelets, $\psi_{1,0}(t)$ and $\psi_{2,1}(t)$ (top row); scaled and translated symmlet8 wavelets, $\psi_{1,0}(t)$ and $\psi_{2,1}(t)$ (bottom row).

$$\phi_{J,k}(t) = 2^{-J/2}\phi\left(\frac{t - 2^J k}{2^J}\right), \tag{4.117}$$

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(\frac{t - 2^j k}{2^j}\right), \quad j = 1, \ldots, J. \tag{4.118}$$

The choice of dyadic shifts and scales is arbitrary but convenient. The shift or translation parameter is $2^j k$, and scale parameter is 2^j. The wavelet functions are spread out and shorter for larger values of j (or scale parameter 2^j) and tall and narrow for small values of the scale. Figure 4.21 shows $\psi_{1,0}(t)$ and $\psi_{2,1}(t)$ generated from the daublet4 (top row), and the symmlet8 (bottom row) mother wavelets. We may think of $1/2^j$ (or 1/scale) in wavelet analysis as being the analogue of frequency ($\omega_j = j/n$) in Fourier analysis. For example, when $j = 1$, the scale parameter of 2 is akin to the Nyquist frequency of 1/2, and when $j = 6$, the scale parameter of 2^6 is akin to a low frequency ($1/2^6 \approx 0.016$). In other words, larger values of the scale refer to slower, smoother (or coarser) movements of the signal, and smaller values of the scale refer to faster, choppier (or finer) movements of the signal. Figure 4.21 was generated in S-PLUS using the wavelet module as follows:

```
d4.1 <- wavelet("d4", level=1, shift=0)
d4.2 <- wavelet("d4", level=2, shift=1)
```

```
s8.1 <- wavelet("s8", level=1, shift=0)
s8.2 <- wavelet("s8", level=2, shift=1)
par(mfrow=c(2,2))
plot(d4.1, ylim=c(-.8,.8), xlim=c(-6,20))
plot(d4.2, ylim=c(-.8,.8), xlim=c(-6,20))
plot(s8.1, ylim=c(-.8,.8), xlim=c(-6,20))
plot(s8.2, ylim=c(-.8,.8), xlim=c(-6,20))
```

The discrete wavelet transform (DWT) of the data x_t are the coefficients $s_{J,k}$ and $d_{j,k}$ for $j = J, J-1, \ldots, 1$, in (4.116). To some degree of approximation, they are given by[19]

$$s_{J,k} = n^{-1/2} \sum_{t=1}^{n} x_t \phi_{J,k}(t), \tag{4.119}$$

$$d_{j,k} = n^{-1/2} \sum_{t=1}^{n} x_t \psi_{j,k}(t) \quad j = J, J-1, \ldots, 1. \tag{4.120}$$

It is the magnitudes of the coefficients that measure the importance of the corresponding wavelet term in describing the behavior of x_t. As in Fourier analysis, the DWT is not computed as shown but is calculated using a fast algorithm. The $s_{J,k}$ are called the smooth coefficients because they represent the smooth behavior of the data. The $d_{j,k}$ are called the detail coefficients because they tend to represent the finer, more high-frequency nature, of the data.

Example 4.22 Wavelet Analysis of Earthquake and Explosion

Figure 4.22 and Figure 4.23 show the DWTs, based on the symmlet8 wavelet basis, for the earthquake and explosion series, respectively. Each series is of length $n = 2^{11} = 2048$, and in this example, the DWTs are calculated using $J = 6$ levels. In this case, $n/2 = 2^{10} = 1024$ values are in $d1 = \{d_{1,k};\ k = 1, \ldots, 2^{10}\}$, $n/2^2 = 2^9 = 512$ values are in $d2 = \{d_{2,k};\ k = 1, \ldots, 2^9\}$, and so on, until finally, $n/2^6 = 2^5 = 32$ values are in $d6$ and in $s6$. The detail values $d_{1,k}, \ldots, d_{6,k}$ are plotted at the same scale, and hence, the relative importance of each value can be seen from the graph. The smooth values $s_{6,k}$ are typically larger than the detail values and plotted on a different scale. The top of Figure 4.22 and Figure 4.23 show the inverse DWT (IDWT) computed from all of the coefficients. The displayed IDWT is a reconstruction of the data, and it reproduces the data except for round-off error.

Comparing the DWTs, the earthquake is best represented by wavelets with larger scale than the explosion. One way to measure the importance of each level, $d1, d2, \ldots, d6, s6$, is to evaluate the proportion of the total power (or energy) explained by each. The total power of a time series x_t, for

[19] The actual DWT coefficients are defined via a set of filters whose coefficients are close to what you would get by sampling the father and mother wavelets, but not exactly so; see the discussion surrounding Figures 471 and 478 in Percival and Walden (2000).

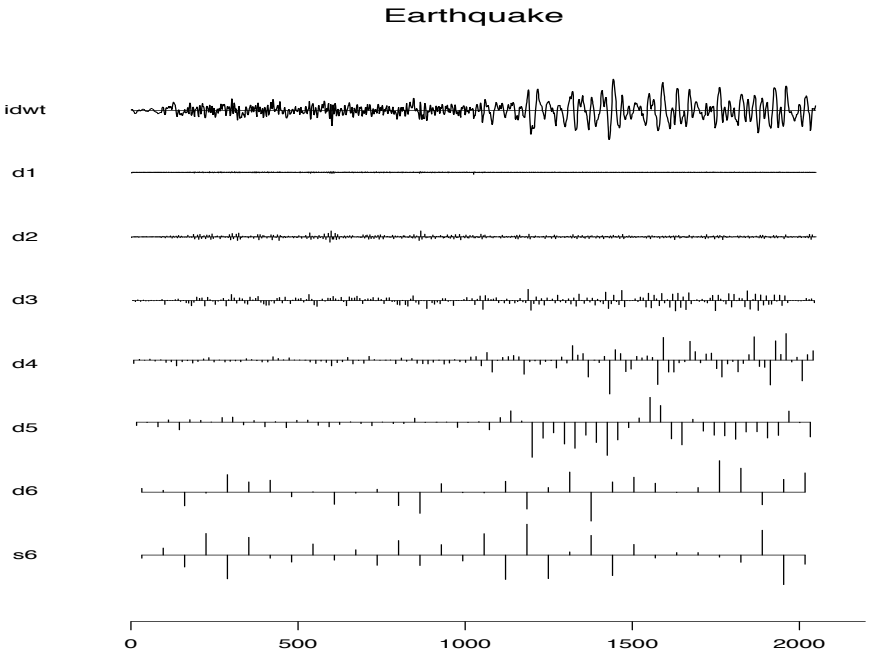

Fig. 4.22. Discrete wavelet transform of the earthquake series using the symmlet8 wavelets, and $J = 6$ levels of scale.

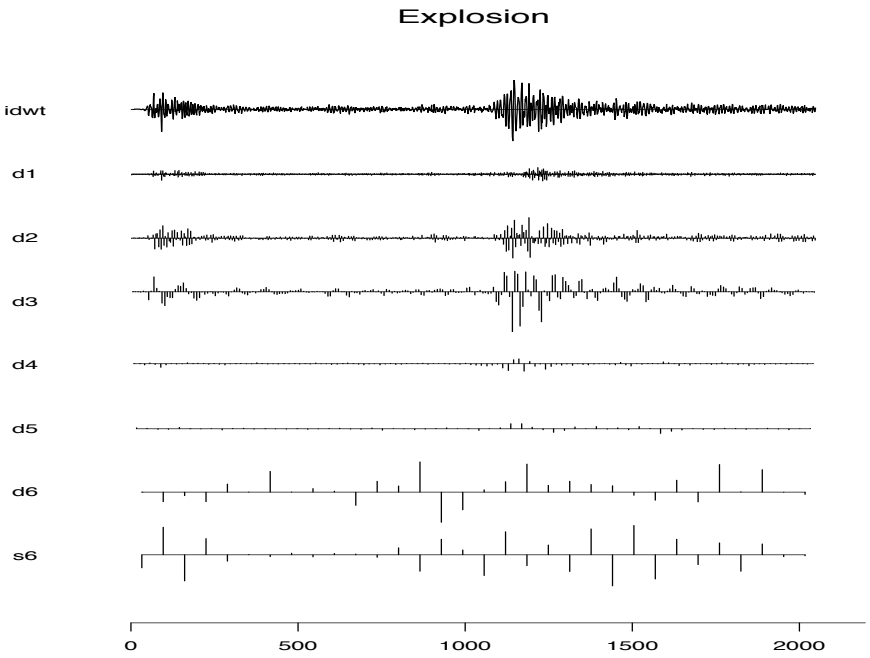

Fig. 4.23. Discrete wavelet transform of the explosion series using the symmlet8 wavelets and $J = 6$ levels of scale.

Table 4.2. Fraction of Total Power

Component	Earthquake	Explosion
s6	0.009	0.002
d6	0.043	0.002
d5	0.377	0.007
d4	0.367	0.015
d3	0.160	0.559
d2	0.040	0.349
d1	0.003	0.066

$t = 1, \ldots, n$, is $TP = \sum_{t=1}^{n} x_t^2$. The total power associated with each level of scale is (recall $n = 2^{11}$),

$$TP_6^s = \sum_{k=1}^{n/2^6} s_{6,k}^2 \quad \text{and} \quad TP_j^d = \sum_{k=1}^{n/2^j} d_{j,k}^2, \quad j = 1, \ldots, 6.$$

Because we are working with an orthogonal basis, we have

$$TP = TP_6^s + \sum_{j=1}^{6} TP_j^d,$$

and the proportion of the total power explained by each level of detail would be the ratios TP_j^d/TP for $j = 1, \ldots, 6$, and for the smooth level, it would be TP_6^s/TP. These values are listed in Table 4.2. From that table nearly 80% of the total power of the earthquake series is explained by the higher scale details $d4$ and $d5$, whereas 90% of the total power is explained by the smaller scale details $d2$ and $d3$ for the explosion.

Figure 4.24 and Figure 4.25 show the time-scale plots (or scalograms) based on the DWT of the earthquake series and the explosion series, respectively. These figures are the wavelet analog of the time-frequency plots shown in Figure 4.17 and Figure 4.18. The power axis represents the magnitude of each value d_{jk} or $s_{6,k}$. The time axis matches the time axis in the DWTs shown in Figure 4.22 and Figure 4.23, and the scale axis is plotted as 1/scale, listed from the coarsest scale to the finest scale. On the 1/scale axis, the coarsest scale values, represented by the smooth coefficients $s6$, are plotted over the range $[0, 2^{-6})$, the coarsest detail values, $d6$, are plotted over $[2^{-6}, 2^{-5})$, and so on. In these figures, we did not plot the finest scale values, $d1$, so the finest scale values exhibited in Figure 4.24 and Figure 4.25 are in $d2$, which are plotted over the range $[2^{-2}, 2^{-1})$.

The conclusions drawn from these plots are the same as those drawn from Figures Figure 4.17 and Figure 4.18. That is, the S wave for the earthquake shows power at the high scales (or low 1/scale) only, and the power remains strong for a long time. In contrast, the explosion shows power at smaller

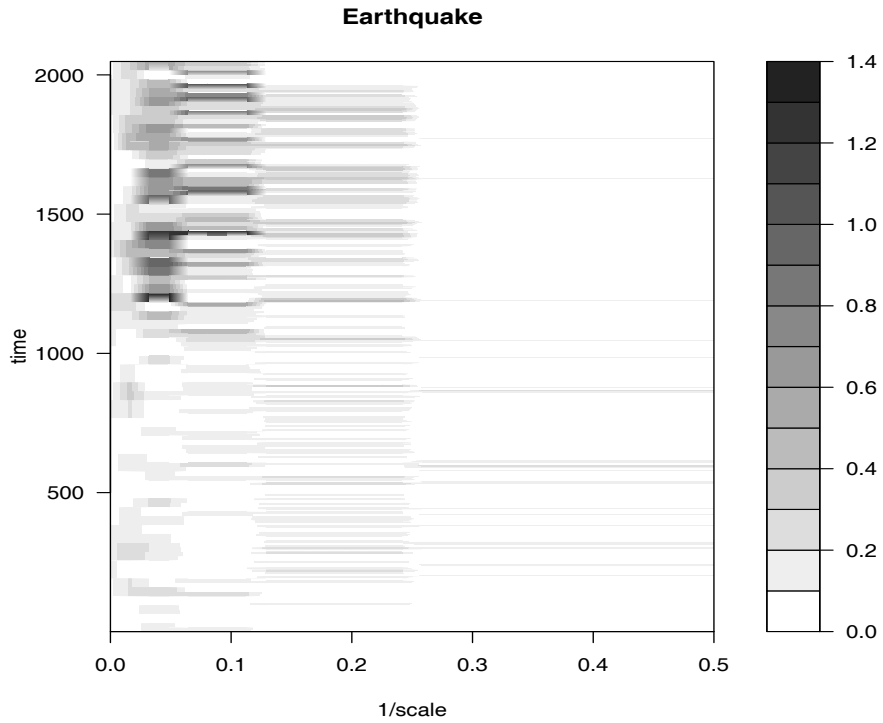

Fig. 4.24. Time-scale image (scalogram) of the earthquake series.

scales (or higher 1/scale) than the earthquake, and the power of the signals (P and S waves) do not last as long as in the case of the earthquake.

Assuming the data files EQ5 and EXP6 have been read into S-PLUS, the analyses of this example can performed using the S-PLUS wavelets module (which must be loaded prior to the analyses) as follows:

```
eq <- scale(EQ5)
ex <- scale(EXP6)
eq.dwt <- dwt(eq)
ex.dwt <- dwt(ex)
plot(eq.dwt)
plot(ex.dwt)
# energy distributions (Table 4.2)
dotchart(eq.dwt)  # a graphic
summary(eq.dwt)  # numerical details
dotchart(ex.dwt)
summary(ex.dwt)
# time scale plots
time.scale.plot(eq.dwt)
time.scale.plot(ex.dwt)
```

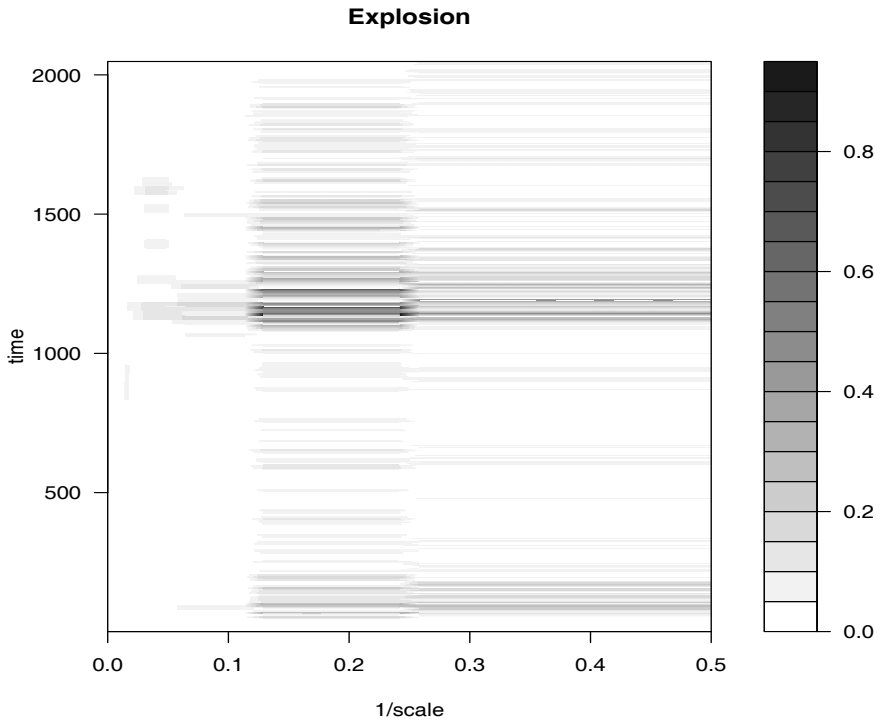

Fig. 4.25. Time-scale image (scalogram) of the explosion series.

Similar analyses may be performed in R using the `wavelets`, `wavethresh`, or `waveslim` packages. We exhibit the analysis for the earthquake series using `wavesthresh`, assuming it has been downloaded and installed.[20]

```
library(wavethresh)
eq = scale(EQ5)  # standardize the series
ex = scale(EXP6)
eq.dwt = wd(eq, filter.number=8)
ex.dwt = wd(ex, filter.number=8)
# plot the wavelet transforms
par(mfrow = c(1,2))
plot(eq.dwt, main="Earthquake")
plot(ex.dwt, main="Explosion")
# total power
TPe = rep(NA,11)  # for the earthquake series
for (i in 0:10){TPe[i+1] = sum(accessD(eq.dwt, level=i)^2)}
TotEq = sum(TPe)  # check with sum(eq^2)
TPx = rep(NA,11)  # for the explosion series
for (i in 0:10){TPx[i+1] = sum(accessD(ex.dwt, level=i)^2)}
TotEx = sum(TPx)  # check with sum(ex^2)
# make a nice table
```

[20] In `wavethresh`, the transforms are denoted by the resolution rather than the scale. If the series is of length $n = 2^p$, then resolution $p - i$ corresponds to level i for $i = 1, \ldots, p$.

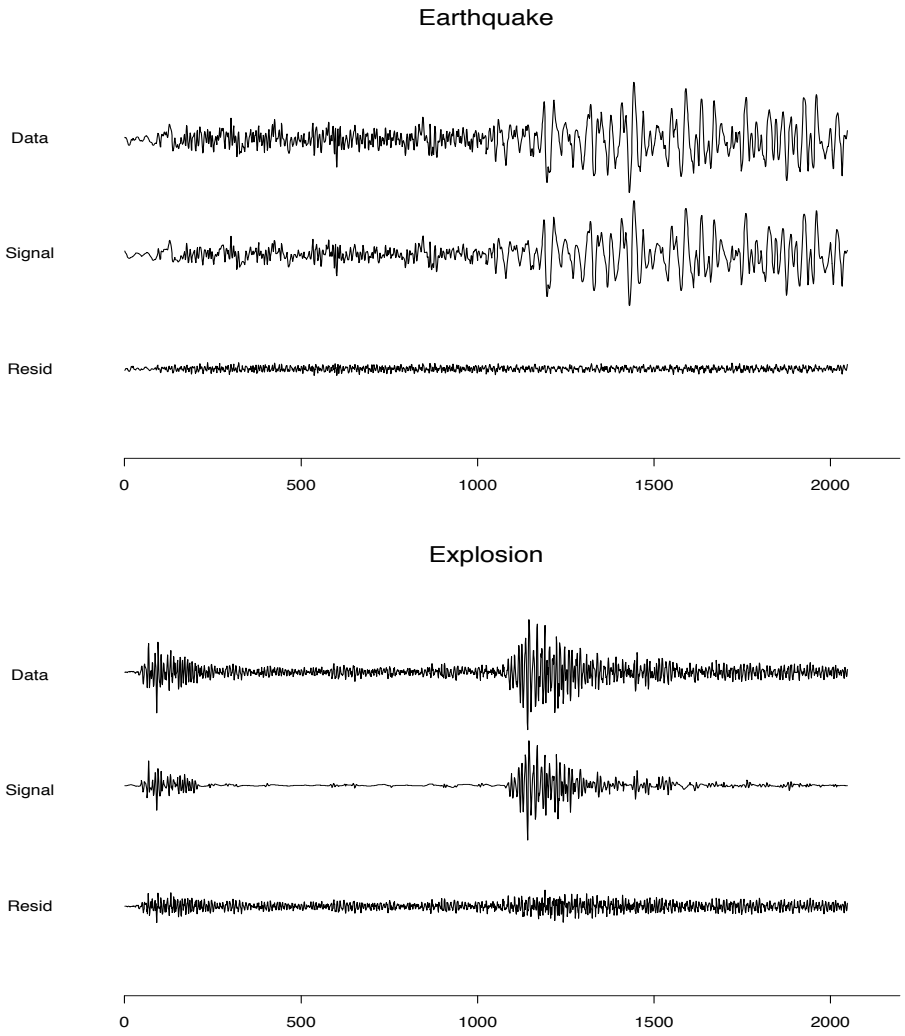

Fig. 4.26. Waveshrink estimates of the earthquake and explosion signals.

```
Power = round(cbind(11:1, 100*TPe/TotEq, 100*TPx/TotEx), digits=3)
colnames(Power) = c("Level", "EQ(%)", "EXP(%)")
Power
```

Wavelets can be used to perform nonparametric smoothing along the lines first discussed in §2.4, but with an emphasis on localized behavior. Although a considerable amount of literature exists on this topic, we will present the basic ideas. For further information, we refer the reader to Donoho and Johnstone (1994, 1995). As in §2.4, we suppose the data x_t can be written in terms of a signal plus noise model as

$$x_t = s_t + \epsilon_t. \tag{4.121}$$

The goal here is to remove the noise from the data, and obtain an estimate of the signal, s_t, without having to specify a parametric form of the signal. The technique based on wavelets is referred to as waveshrink.

The basic idea behind waveshrink is to shrink the wavelet coefficients in the DWT of x_t toward zero in an attempt to denoise the data and then to estimate the signal via (4.116) with the new coefficients. One obvious way to shrink the coefficients toward zero is to simply zero out any coefficient smaller in magnitude than some predetermined value, λ. Such a shrinkage rule is discontinuous and sometimes it is preferable to use a continuous shrinkage function. One such method, termed soft shrinkage, proceeds as follows. If the value of a coefficient is a, we set that coefficient to zero if $|a| \leq \lambda$, and to $\text{sign}(a)(|a| - \lambda)$ if $|a| > \lambda$. The choice of a shrinkage method is based on the goal of the signal extraction. This process entails choosing a value for the shrinkage threshold, λ, and we may wish to use a different threshold value, say, λ_j, for each level of scale $j = 1, \ldots, J$. One particular method that works well if we are interested in a relatively high degree of smoothness in the estimate is to choose $\lambda = \widehat{\sigma}_\epsilon \sqrt{2 \log n}$ for all scale levels, where $\widehat{\sigma}_\epsilon$ is an estimate of the scale of the noise, σ_ϵ. Typically a robust estimate of σ_ϵ is used, e.g., the median of the absolute deviations of the data from the median (MAD). For other thresholding techniques or for a better understanding of waveshrink, see Donoho and Johnstone (1994, 1995), or the S-PLUS wavelets module manual (Bruce and Gao, 1996, Ch 6).

Example 4.23 Waveshrink Analysis of Earthquake and Explosion

Figure 4.26 shows the results of a waveshrink analysis on the earthquake and explosion series. In this example, soft shrinkage was used with a universal threshold of $\lambda = \widehat{\sigma}_\epsilon \sqrt{2 \log n}$ where $\widehat{\sigma}_\epsilon$ is the MAD. Figure 4.26 displays the data x_t, the estimated signal $\widehat{s}_t$, as well as the residuals $x_t - \widehat{s}_t$. According to this analysis, the earthquake is mostly signal and characterized by prolonged energy, whereas the explosion is comprised of short bursts of energy.

Figure 4.26 was generated in S-PLUS using the wavelets module. For example, the analysis of the earthquake series was performed as follows.

```
eq.dwt <- dwt(eq)
eq.shrink <- waveshrink(eq.dwt, shrink.rule="universal",
    shrink.fun="soft")
```

In R, using the `wavethresh` package, use the following commands for the earthquake series.

```
library(wavethresh)
eq = scale(EQ5)
par(mfrow=c(3,1))
eq.dwt = wd(eq, filter.number=8)
eq.smo = wr(threshold(eq.dwt, levels=5:10))
ts.plot(eq, main="Earthquake", ylab="Data")
ts.plot(eq.smo, ylab="Signal")
ts.plot(eq-eq.smo, ylab="Resid")
```

4.10 Lagged Regression Models

One of the intriguing possibilities offered by the coherence analysis of the relation between the SOI and Recruitment series discussed in Example 4.18 would be extending classical regression to the analysis of lagged regression models of the form

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t, \tag{4.122}$$

where v_t is a stationary noise process, x_t is the observed input series, and y_t is the observed output series. We are interested in estimating the filter coefficients β_r relating the adjacent lagged values of x_t to the output series y_t.

In the case of SOI and Recruitment series, we might identify the El Niño driving series, SOI, as the input, x_t, and y_t, the Recruitment series, as the output. In general, there will be more than a single possible input series and we may envision a $q \times 1$ vector of driving series. This multivariate input situation is covered in Chapter 7. The model given by (4.122) is useful under several different scenarios, corresponding to different assumptions that can be made about the components.

We assume that the inputs and outputs have zero means and are jointly stationary with the 2×1 vector process $(x_t, y_t)'$ having a spectral matrix of the form

$$f(\omega) = \begin{pmatrix} f_{xx}(\omega) & f_{xy}(\omega) \\ f_{yx}(\omega) & f_{yy}(\omega) \end{pmatrix}. \tag{4.123}$$

Here, $f_{xy}(\omega)$ is the cross-spectrum relating the input x_t to the output y_t, and $f_{xx}(\omega)$ and $f_{yy}(\omega)$ are the spectra of the input and output series, respectively. Generally, we observe two series, regarded as input and output and search for regression functions $\{\beta_t\}$ relating the inputs to the outputs. We assume all autocovariance functions satisfy the absolute summability conditions of the form (4.30).

Then, minimizing the mean squared error

$$MSE = E\left(y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r}\right)^2 \tag{4.124}$$

leads to the usual orthogonality conditions

$$E\left[\left(y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r}\right) x_{t-s}\right] = 0 \tag{4.125}$$

for all $s = 0, \pm 1, \pm 2, \ldots$. Taking the expectations inside leads to the normal equations

$$\sum_{r=-\infty}^{\infty} \beta_r \, \gamma_{xx}(s-r) = \gamma_{yx}(s) \tag{4.126}$$

for $s = 0, \pm 1, \pm 2, \ldots$. These equations might be solved, with some effort, if the covariance functions were known exactly. If data (x_t, y_t) for $t = 1, ..., n$ are available, we might use a finite approximation to the above equations with $\widehat{\gamma}_{xx}(h)$ and $\widehat{\gamma}_{yx}(h)$ substituted into (4.126). If the regression vectors are essentially zero for $|s| \geq M/2$, and $M < n$, the system (4.126) would be of full rank and the solution would involve inverting an $(M-1) \times (M-1)$ matrix.

A frequency domain approximate solution is easier in this case for two reasons. First, the computations depend on spectra and cross-spectra that can be estimated from sample data using the techniques of §4.6. In addition, no matrices will have to be inverted, although the frequency domain ratio will have to be computed for each frequency. In order to develop the frequency domain solution, substitute the representation (4.89) into the normal equations, using the convention defined in (4.123). The left side of (4.126) can then be written in the form

$$\int_{-1/2}^{1/2} \sum_{r=-\infty}^{\infty} \beta_r \, \mathrm{e}^{2\pi i\omega(s-r)} \, f_{xx}(\omega) \, d\omega = \int_{-1/2}^{1/2} \mathrm{e}^{2\pi i\omega s} B(\omega) f_{xx}(\omega) \, d\omega,$$

where

$$B(\omega) = \sum_{r=-\infty}^{\infty} \beta_r \, \mathrm{e}^{-2\pi i\omega r} \tag{4.127}$$

is the Fourier transform of the regression coefficients β_t. Now, because $\gamma_{yx}(s)$ is the inverse transform of the cross-spectrum $f_{yx}(\omega)$, we might write the system of equations in the frequency domain, using the uniqueness of the Fourier transform, as

$$B(\omega) f_{xx}(\omega) = f_{yx}(\omega), \tag{4.128}$$

which then become the analogs of the usual normal equations. Then, we may take

$$\widehat{B}(\omega_k) = \frac{\widehat{f}_{yx}(\omega_k)}{\widehat{f}_{xx}(\omega_k)} \tag{4.129}$$

as the estimator for the Fourier transform of the regression coefficients, evaluated at some subset of fundamental frequencies $\omega_k = k/M$ with $M << n$. Generally, we assume smoothness of $B(\cdot)$ over intervals of the form $\{\omega_k + \ell/n;\ \ell = -(L-1)/2, \ldots, (L-1)/2\}$. The inverse transform of the function $\widehat{B}(\omega)$ would give $\widehat{\beta}_t$, and we note that the discrete time approximation can be taken as

$$\widehat{\beta}_t = M^{-1} \sum_{k=0}^{M-1} \widehat{B}(\omega_k) \mathrm{e}^{2\pi i\omega_k t} \tag{4.130}$$

for $t = 0, \pm 1, \pm 2, \ldots, \pm(M/2 - 1)$. If we were to use (4.130) to define $\widehat{\beta}_t$ for $|t| \geq M/2$, we would end up with a sequence of coefficients that is periodic with a period of M. In practice we define $\widehat{\beta}_t = 0$ for $|t| \geq M/2$ instead. Problem 4.32 explores the error resulting from this approximation.

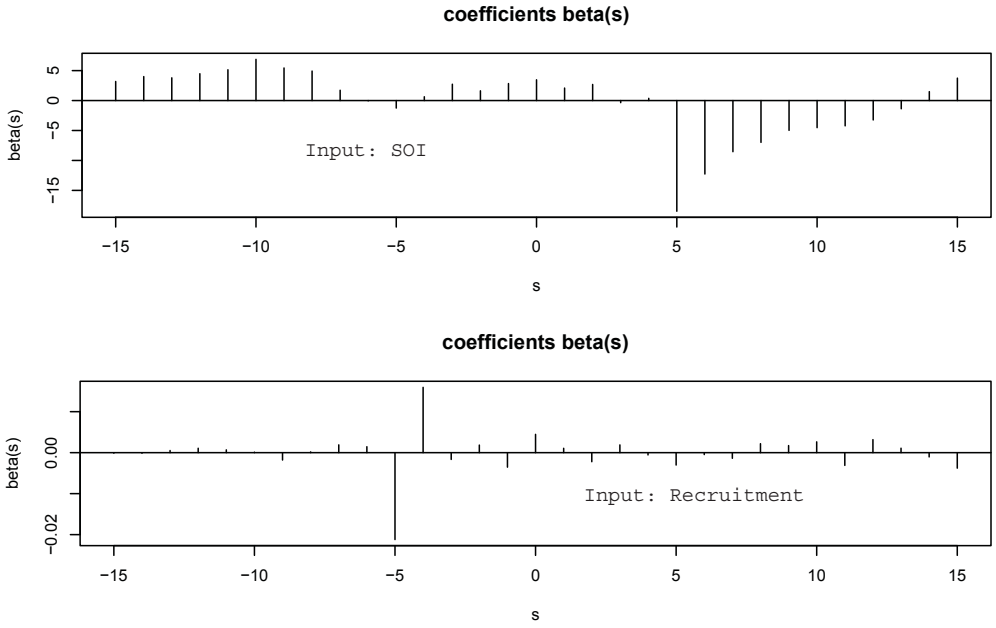

Fig. 4.27. Estimated impulse response functions relating SOI to Recruitment (top) and Recruitment to SOI (bottom) $L = 15, M = 32$.

Example 4.24 Lagged Regression for SOI and Recruitment

The high coherence between the SOI and Recruitment series noted in Example 4.18 suggests a lagged regression relation between the two series. A natural direction for the implication in this situation is implied because we feel that the sea surface temperature or SOI should be the input and the Recruitment series should be the output. With this in mind, let x_t be the SOI series and y_t the Recruitment series.

Although we think naturally of the SOI as the input and the Recruitment as the output, two input-output configurations are of interest. With SOI as the input, the model is

$$y_t = \sum_{r=-\infty}^{\infty} a_r x_{t-r} + w_t$$

whereas a model that reverses the two roles would be

$$x_t = \sum_{r=-\infty}^{\infty} b_r y_{t-r} + v_t,$$

where w_t and v_t are white noise processes. Even though there is no plausible environmental explanation for the second of these two models, displaying both possibilities helps to settle on a parsimonious transfer function model.

Based on the script `LagReg` (see Appendix R, §R.1), the estimated regression or impulse response function for SOI, with $M = 32$ and $L = 15$ is

```
LagReg(soi, rec, L=15, M=32, threshold=6)
```

```
           lag s    beta(s)
     [1,]      5 -18.479306
     [2,]      6 -12.263296
     [3,]      7  -8.539368
     [4,]      8  -6.984553
The prediction equation is
rec(t) = alpha + sum_s[ beta(s)*soi(t-s) ], where alpha = 65.97
MSE = 414.08
```

Note the negative peak at a lag of five points in the top of Figure 4.27; in this case, SOI is the input series. The fall-off after lag five seems to be approximately exponential and a possible model is

$$y_t = 66 - 18.5x_{t-5} - 12.3x_{t-6} - 8.5x_{t-7} - 7x_{t-8} + w_t.$$

If we examine the inverse relation, namely, a regression model with the Recruitment series y_t as the input, the bottom of Figure 4.27 implies a much simpler model,

```
LagReg(rec, soi, L=15, M=32, inverse=TRUE, threshold=.01)
```

```
          lag s     beta(s)
     [1,]     4  0.01593167
     [2,]     5 -0.02120013
The prediction equation is
soi(t) = alpha + sum_s[ beta(s)*rec(t+s) ], where alpha = 0.41
MSE = 0.07
```

depending on only two coefficients, namely,

$$x_t = .41 + .016y_{t+4} - .02y_{t+5} + v_t.$$

Multiplying both sides by $50B^5$ and rearranging, we have

$$(1 - .8B)y_t = 20.5 - 50B^5x_t + \epsilon_t,$$

where ϵ_t is white noise, as our final, parsimonious model.

The example shows we can get a clean estimator for the transfer functions relating the two series if the coherence $\widehat{\rho}^2_{xy}(\omega)$ is large. The reason is that we can write the minimized mean squared error (4.124) as

$$MSE = E\Big[\big(y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r}\big)y_t\Big] = \gamma_{yy}(0) - \sum_{r=-\infty}^{\infty} \beta_r \gamma_{xy}(-r),$$

using the result about the orthogonality of the data and error term in the Projection theorem. Then, substituting the spectral representations of the autocovariance and cross-covariance functions and identifying the Fourier transform (4.127) in the result leads to

$$MSE = \int_{-1/2}^{1/2} [f_{yy}(\omega) - B(\omega) f_{xy}(\omega)] \, d\omega$$
$$= \int_{-1/2}^{1/2} f_{yy}(\omega)[1 - \rho_{yx}^2(\omega)] d\omega, \tag{4.131}$$

where $\rho_{yx}^2(\omega)$ is just the squared coherence given by (4.87). The similarity of (4.131) to the usual mean square error that results from predicting y from x is obvious. In that case, we would have

$$E(y - \beta x)^2 = \sigma_y^2(1 - \rho_{xy}^2)$$

for jointly distributed random variables x and y with zero means, variances σ_x^2 and σ_y^2, and covariance $\sigma_{xy} = \rho_{xy}\sigma_x\sigma_y$. Because the mean squared error in (4.131) satisfies $MSE \geq 0$ with $f_{yy}(\omega)$ a non-negative function, it follows that the coherence satisfies

$$0 \leq \rho_{xy}^2(\omega) \leq 1$$

for all ω. Furthermore, Problem 4.33 shows the squared coherence is one when the output are linearly related by the filter relation (4.122), and there is no noise, i.e., $v_t = 0$. Hence, the multiple coherence gives a measure of the association or correlation between the input and output series as a function of frequency.

The matter of verifying that the F-distribution claimed for (4.97) will hold when the sample coherence values are substituted for theoretical values still remains. Again, the form of the F-statistic is exactly analogous to the usual t-test for no correlation in a regression context. We give an argument leading to this conclusion later using the results in Appendix C, §C.3. Another question that has not been resolved in this section is the extension to the case of multiple inputs $x_{t1}, x_{t2}, \ldots, x_{tq}$. Often, more than just a single input series is present that can possibly form a lagged predictor of the output series y_t. An example is the cardiovascular mortality series that depended on possibly a number of pollution series and temperature. We discuss this particular extension as a part of the multivariate time series techniques considered in Chapter 7.

4.11 Signal Extraction and Optimum Filtering

A model closely related to regression can be developed by assuming again that

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t, \tag{4.132}$$

but where the βs are known and x_t is some unknown random signal that is uncorrelated with the noise process v_t. In this case, we observe only y_t and are interested in an estimator for the signal x_t of the form

$$\widehat{x}_t = \sum_{r=-\infty}^{\infty} a_r y_{t-r}. \tag{4.133}$$

In the frequency domain, it is convenient to make the additional assumptions that the series x_t and v_t are both mean-zero stationary series with spectra $f_{xx}(\omega)$ and $f_{vv}(\omega)$, often referred to as the signal spectrum and noise spectrum, respectively. Often, the special case $\beta_t = \delta_t$, in which δ_t is the Kronecker delta, is of interest because (4.132) reduces to the simple signal plus noise model

$$y_t = x_t + v_t \tag{4.134}$$

in that case. In general, we seek the set of filter coefficients a_t that minimize the mean squared error of estimation, say,

$$MSE = E\left[\left(x_t - \sum_{r=-\infty}^{\infty} a_r y_{t-r}\right)^2\right]. \tag{4.135}$$

This problem was originally solved by Kolmogorov (1941) and by Wiener (1949), who derived the result in 1941 and published it in classified reports during World War II.

We can apply the orthogonality principle to write

$$E\left[\left(x_t - \sum_{r=-\infty}^{\infty} a_r y_{t-r}\right) y_{t-s}\right] = 0$$

for $s = 0, \pm 1, \pm 2, \ldots$, which leads to

$$\sum_{r=-\infty}^{\infty} a_r \gamma_{yy}(s-r) = \gamma_{xy}(s),$$

to be solved for the filter coefficients. Substituting the spectral representations for the autocovariance functions into the above and identifying the spectral densities through the uniqueness of the Fourier transform produces

$$A(\omega) f_{yy}(\omega) = f_{xy}(\omega), \tag{4.136}$$

where $A(\omega)$ and the optimal filter a_t are Fourier transform pairs for $B(\omega)$ and β_t. Now, a special consequence of the model is that (see Problem 4.23)

$$f_{xy}(\omega) = \overline{B(\omega)} f_{xx}(\omega) \tag{4.137}$$

and

$$f_{yy}(\omega) = |B(\omega)|^2 f_{xx}(\omega) + f_{vv}(\omega), \tag{4.138}$$

implying the optimal filter would be Fourier transform of

$$A(\omega) = \frac{\overline{B(\omega)}}{\left(|B(\omega)|^2 + \frac{f_{vv}(\omega)}{f_{xx}(\omega)}\right)}, \tag{4.139}$$

where the second term in the denominator is just the inverse of the signal to noise ratio, say,

$$\mathrm{SNR}(\omega) = \frac{f_{xx}(\omega)}{f_{vv}(\omega)}. \tag{4.140}$$

The result shows the optimum filters can be computed for this model if the signal and noise spectra are both known or if we can assume knowledge of the signal-to-noise ratio SNR(ω) as function of frequency. In Chapter 7, we show some methods for estimating these two parameters in conjunction with random effects analysis of variance models, but we assume here that it is possible to specify the signal-to-noise ratio a priori. If the signal-to-noise ratio is known, the optimal filter can be computed by the inverse transform of the function $A(\omega)$. It is more likely that the inverse transform will be intractable and a finite filter approximation like that used in the previous section can be applied to the data. In this case, we will have

$$a_t^M = M^{-1} \sum_{k=0}^{M-1} A(\omega_k) \mathrm{e}^{2\pi i \omega_k t} \tag{4.141}$$

as the estimated filter function. It will often be the case that the form of the specified frequency response will have some rather sharp transitions between regions where the signal-to-noise ratio is high and regions where there is little signal. In these cases, the shape of the frequency response function will have ripples that can introduce frequencies at different amplitudes. An aesthetic solution to this problem is to introduce tapering as was done with spectral estimation in (4.61)-(4.68). We use below the tapered filter $\tilde{a}_t = h_t a_t$ where h_t is the cosine taper given in (4.68). The squared frequency response of the resulting filter will be $|\tilde{A}(\omega)|^2$, where

$$\tilde{A}(\omega) = \sum_{t=-\infty}^{\infty} a_t h_t \mathrm{e}^{-2\pi i \omega t}. \tag{4.142}$$

The results are illustrated in the following example that extracts the El Niño component of the sea surface temperature series.

Example 4.25 Estimating the El Niño Signal via Optimal Filters
Figure 4.5 shows the spectrum of the SOI series, and we note that essentially two components have power, the El Niño frequency of about .02 cycles per month (the four-year cycle) and a yearly frequency of about .08 cycles per month (the annual cycle). We assume, for this example, that we wish to preserve the lower frequency as signal and to eliminate the higher order

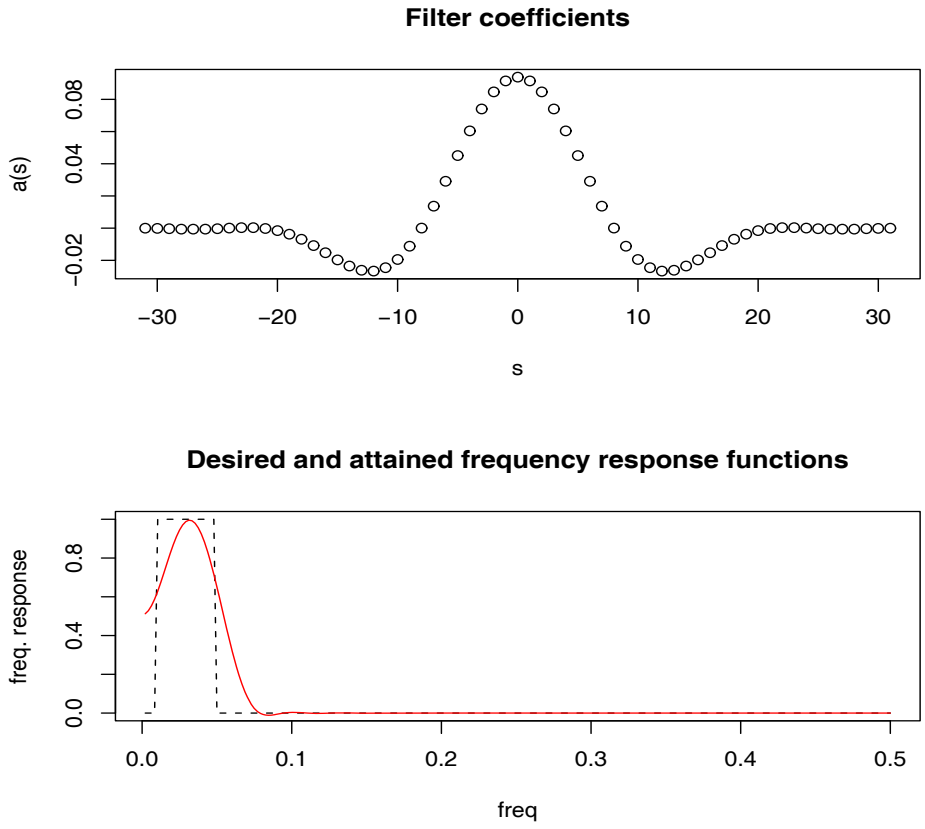

Fig. 4.28. Filter coefficients (top) and frequency response functions (bottom) for designed SOI filters.

frequencies, and in particular, the annual cycle. In this case, we assume the simple signal plus noise model

$$y_t = x_t + v_t,$$

so that there is no convolving function β_t. Furthermore, the signal-to-noise ratio is assumed to be high to about .06 cycles per month and zero thereafter. The optimal frequency response was assumed to be unity to .05 cycles per point and then to decay linearly to zero in several steps. Figure 4.28 shows the coefficients as specified by (4.141) with $M = 64$, as well as the frequency response function given by (4.142), of the cosine tapered coefficients; recall Figure 4.9, where we demonstrated the need for tapering to avoid severe ripples in the window. The constructed response function is compared to the ideal window in Figure 4.28.

Figure 4.29 shows the original and filtered SOI index, and we see a smooth extracted signal that conveys the essence of the underlying El Niño signal. The frequency response of the designed filter can be compared with that of the symmetric 12-month moving average applied to the same series in Example 4.19. The filtered series, shown in Figure 4.14, shows a good deal

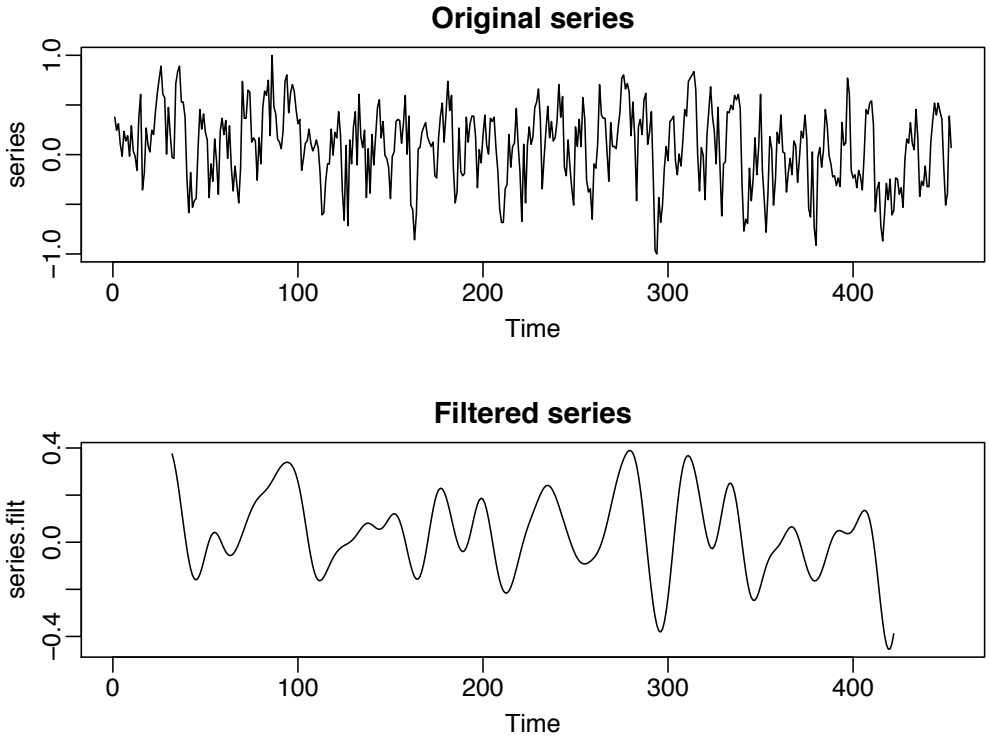

Fig. 4.29. Original SOI series (top) compared to filtered version showing the estimated El Niño temperature signal (bottom).

of higher frequency chatter riding on the smoothed version, which has been introduced by the higher frequencies that leak through in the squared frequency response, as in Figure 4.16.

The analysis can be replicated using the script `SigExtract`; see Appendix R, §R.1, for details.

```
SigExtract(soi, L=9, M=64, max.freq=.05)
```

The design of finite filters with a specified frequency response requires some experimentation with various target frequency response functions and we have only touched on the methodology here. The filter designed here, sometimes called a low-pass filter reduces the high frequencies and keeps or passes the low frequencies. Alternately, we could design a high-pass filter to keep high frequencies if that is where the signal is located. An example of a simple high-pass filter is the first difference with a frequency response that is shown in Figure 4.16. We can also design band-pass filters that keep frequencies in specified bands. For example, seasonal adjustment filters are often used in economics to reject seasonal frequencies while keeping both high frequencies, lower frequencies, and trend (see, for example, Grether and Nerlove, 1970).

The filters we have discussed here are all symmetric two-sided filters, because the designed frequency response functions were purely real. Alternatively, we may design recursive filters to produce a desired response. An ex-

ample of a recursive filter is one that replaces the input x_t by the filtered output

$$y_t = \sum_{k=1}^{p} \phi_k y_{t-k} + x_t - \sum_{k=1}^{q} \theta_k x_{t-k}. \tag{4.143}$$

Note the similarity between (4.143) and the ARIMA$(p, 1, q)$ model, in which the white noise component is replaced by the input. Transposing the terms involving y_t and using the basic linear filter result in Property 4.7 leads to

$$f_y(\omega) = \frac{|\theta(\mathrm{e}^{-2\pi i\omega})|^2}{|\phi(\mathrm{e}^{-2\pi i\omega})|^2} f_x(\omega), \tag{4.144}$$

where

$$\phi(\mathrm{e}^{-2\pi i\omega}) = 1 - \sum_{k=1}^{p} \phi_k \mathrm{e}^{-2\pi ik\omega}$$

and

$$\theta(\mathrm{e}^{-2\pi i\omega}) = 1 - \sum_{k=1}^{q} \theta_k \mathrm{e}^{-2\pi ik\omega}.$$

Recursive filters such as those given by (4.144) distort the phases of arriving frequencies, and we do not consider the problem of designing such filters in any detail.

4.12 Spectral Analysis of Multidimensional Series

Multidimensional series of the form $x_{\boldsymbol{s}}$, where $\boldsymbol{s} = (s_1, s_2, \ldots, s_r)'$ is an r-dimensional vector of spatial coordinates or a combination of space and time coordinates, were introduced in §1.7. The example given there, shown in Figure 1.15, was a collection of temperature measurements taking on a rectangular field. These data would form a two-dimensional process, indexed by row and column in space. In that section, the multidimensional autocovariance function of an r-dimensional stationary series was given as $\gamma_x(\boldsymbol{h}) = E[x_{\boldsymbol{s}+\boldsymbol{h}} x_{\boldsymbol{s}}]$, where the multidimensional lag vector is $\boldsymbol{h} = (h_1, h_2, \ldots, h_r)'$.

The multidimensional wavenumber spectrum is given as the Fourier transform of the autocovariance, namely,

$$f_x(\boldsymbol{\omega}) = \sum_{\boldsymbol{h}} \gamma_x(\boldsymbol{h}) \mathrm{e}^{-2\pi i \boldsymbol{\omega}'\boldsymbol{h}}. \tag{4.145}$$

Again, the inverse result

$$\gamma_x(\boldsymbol{h}) = \int_{-1/2}^{1/2} f_x(\boldsymbol{\omega}) e^{2\pi i \boldsymbol{\omega}'\boldsymbol{h}} d\boldsymbol{\omega} \tag{4.146}$$

holds, where the integral is over the multidimensional range of the vector $\boldsymbol{\omega}$. The wavenumber argument is exactly analogous to the frequency argument, and we have the corresponding intuitive interpretation as the cycling rate ω_i per distance traveled s_i in the i-th direction.

Two-dimensional processes occur often in practical applications, and the representations above reduce to

$$f_x(\omega_1, \omega_2) = \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} \gamma_x(h_1, h_2) \mathrm{e}^{-2\pi i(\omega_1 h_1 + \omega_2 h_2)} \tag{4.147}$$

and

$$\gamma_x(h_1, h_2) = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} f_x(\omega_1, \omega_2) \mathrm{e}^{2\pi i(\omega_1 h_1 + \omega_2 h_2)} d\omega_1 \; d\omega_2 \tag{4.148}$$

in the case $r = 2$. The notion of linear filtering generalizes easily to the two-dimensional case by defining the impulse response function a_{s_1,s_2} and the spatial filter output as

$$y_{s_1,s_2} = \sum_{u_1} \sum_{u_2} a_{u_1,u_2} x_{s_1-u_1,s_2-u_2}. \tag{4.149}$$

The spectrum of the output of this filter can be derived as

$$f_y(\omega_1, \omega_2) = |A(\omega_1, \omega_2)|^2 f_x(\omega_1, \omega_2), \tag{4.150}$$

where

$$A(\omega_1, \omega_2) = \sum_{u_1} \sum_{u_2} a_{u_1,u_2} \mathrm{e}^{-2\pi i(\omega_1 u_1 + \omega_2 u_2)}. \tag{4.151}$$

These results are analogous to those in the one-dimensional case, described by Property 4.7.

The multidimensional DFT is also a straightforward generalization of the univariate expression. In the two-dimensional case with data on a rectangular grid, $\{x_{s_1,s_2};\ s_1 = 1, ..., n_1,\ s_2 = 1, ..., n_2\}$, we will write, for $-1/2 \leq \omega_1, \omega_2 \leq 1/2$,

$$d(\omega_1, \omega_2) = (n_1 n_2)^{-1/2} \sum_{s_1=1}^{n_1} \sum_{s_2=1}^{n_2} x_{s_1,s_2} \mathrm{e}^{-2\pi i(\omega_1 s_1 + \omega_2 s_2)} \tag{4.152}$$

as the two-dimensional DFT, where the frequencies ω_1, ω_2 are evaluated at multiples of $(1/n_1, 1/n_2)$ on the spatial frequency scale. The two-dimensional wavenumber spectrum can be estimated by the smoothed sample wavenumber spectrum

$$\bar{f}_x(\omega_1, \omega_2) = (L_1 L_2)^{-1} \sum_{\ell_1,\ell_2} \left| d(\omega_1 + \ell_1/n_1, \omega_2 + \ell_2/n_2) \right|^2, \tag{4.153}$$

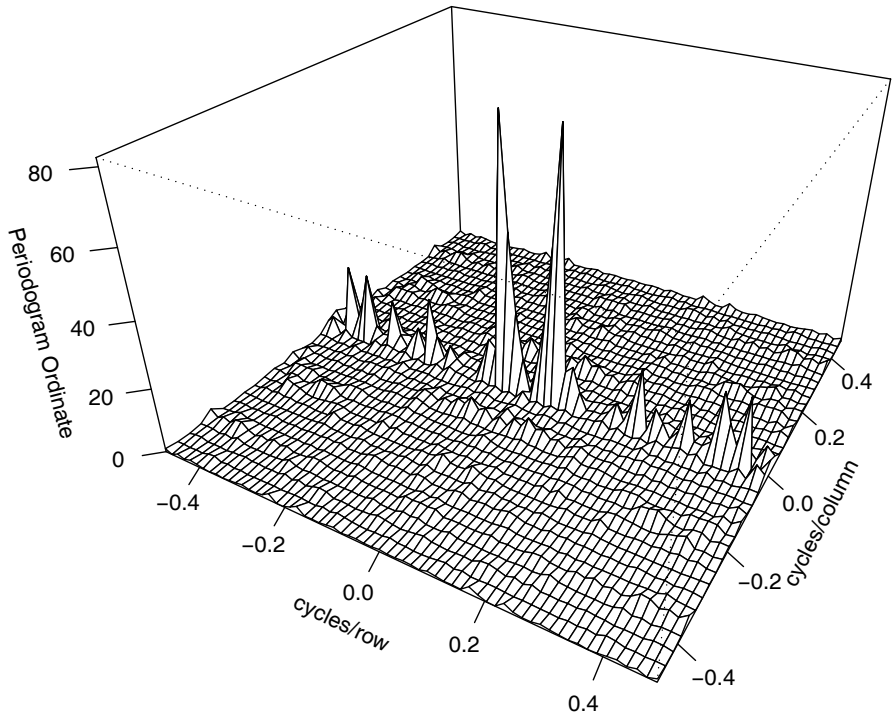

Fig. 4.30. Two-dimensional periodogram of soil temperature profile showing peak at .0625 cycles/row. The period is 16 rows, and this corresponds to 16×17 ft $= 272$ ft.

where the sum is taken over the grid $\{-m_j \le \ell_j \le m_j;\ j = 1, 2\}$, where $L_1 = 2m_1 + 1$ and $L_2 = 2m_2 + 1$. The statistic

$$\frac{2L_1L_2\bar{f}_x(\omega_1, \omega_2)}{f_x(\omega_1, \omega_2)} \sim \chi^2_{2L_1L_2} \tag{4.154}$$

can be used to set confidence intervals or make approximate tests against a fixed assumed spectrum $f_0(\omega_1, \omega_2)$. We may also extend this analysis to weighted estimation and window estimation as discussed in §4.5.

Example 4.26 Soil Surface Temperatures

As an example, consider the periodogram of the two-dimensional temperature series shown in Figure 1.15 and analyzed by Bazza et al. (1988). We recall the spatial coordinates in this case will be (s_1, s_2), which define the spatial coordinates rows and columns so that the frequencies in the two directions will be expressed as cycles per row and cycles per column. Figure 4.30 shows the periodogram of the two-dimensional temperature series, and we note the ridge of strong spectral peaks running over rows at a column frequency of zero. An obvious periodic component appears at frequencies of .0625 and $-.0625$ cycles per row, which corresponds to 16 rows or about 272 ft. On further investigation of previous irrigation patterns over this field,

treatment levels of salt varied periodically over columns. This analysis is extended in Problem 4.17, where we recover the salt treatment profile over rows and compare it to a signal, computed by averaging over columns.

Figure 4.30 may be reproduced in R as follows. In the code for this example, the periodogram is computed in one step as `per`; the rest of the code is simply manipulation to obtain a nice graphic.

```
per = abs(fft(soiltemp-mean(soiltemp))/sqrt(64*36))^2
per2 = cbind(per[1:32,18:2], per[1:32,1:18])
per3 = rbind(per2[32:2,],per2)
par(mar=c(1,2.5,0,0)+.1)
persp(-31:31/64, -17:17/36, per3, phi=30, theta=30, expand=.6,
    ticktype="detailed", xlab="cycles/row", ylab="cycles/column",
    zlab="Periodogram Ordinate")
```

Another application of two-dimensional spectral analysis of agricultural field trials is given in McBratney and Webster (1981), who used it to detect ridge and furrow patterns in yields. The requirement for regular, equally spaced samples on fairly large grids has tended to limit enthusiasm for strict two-dimensional spectral analysis. An exception is when a propagating signal from a given velocity and azimuth is present so predicting the wavenumber spectrum as a function of velocity and azimuth becomes feasible (see Shumway et al., 1999).

Problems

Section 4.2

4.1 Repeat the simulations and analyses in Example 4.1 and Example 4.2 with the following changes:

(a) Change the sample size to $n = 128$ and generate and plot the same series as in Example 4.1:

$$\begin{aligned} x_{t1} &= 2\cos(2\pi\,.06\,t) + 3\sin(2\pi\,.06\,t), \\ x_{t2} &= 4\cos(2\pi\,.10\,t) + 5\sin(2\pi\,.10\,t), \\ x_{t3} &= 6\cos(2\pi\,.40\,t) + 7\sin(2\pi\,.40\,t), \\ x_t &= x_{t1} + x_{t2} + x_{t3}. \end{aligned}$$

What is the major difference between these series and the series generated in Example 4.1? (Hint: The answer is *fundamental*. But if your answer is the series are longer, you may be punished severely.)

(b) As in Example 4.2, compute and plot the periodogram of the series, x_t, generated in (a) and comment.

(c) Repeat the analyses of (a) and (b) but with $n = 100$ (as in Example 4.1), and adding noise to x_t; that is

$$x_t = x_{t1} + x_{t2} + x_{t3} + w_t$$

where $w_t \sim$ iid N$(0, 25)$. That is, you should simulate and plot the data, and then plot the periodogram of x_t and comment.

4.2 With reference to equations (4.1) and (4.2), let $Z_1 = U_1$ and $Z_2 = -U_2$ be independent, standard normal variables. Consider the polar coordinates of the point (Z_1, Z_2), that is,

$$A^2 = Z_1^2 + Z_2^2 \quad \text{and} \quad \phi = \tan^{-1}(Z_2/Z_1).$$

(a) Find the joint density of A^2 and ϕ, and from the result, conclude that A^2 and ϕ are independent random variables, where A^2 is a chi-squared random variable with 2 df, and ϕ is uniformly distributed on $(-\pi, \pi)$.
(b) Going in reverse from polar coordinates to rectangular coordinates, suppose we assume that A^2 and ϕ are independent random variables, where A^2 is chi-squared with 2 df, and ϕ is uniformly distributed on $(-\pi, \pi)$. With $Z_1 = A\cos(\phi)$ and $Z_2 = A\sin(\phi)$, where A is the positive square root of A^2, show that Z_1 and Z_2 are independent, standard normal random variables.

4.3 Verify (4.4).

Section 4.3

4.4 A time series was generated by first drawing the white noise series w_t from a normal distribution with mean zero and variance one. The observed series x_t was generated from

$$x_t = w_t - \theta w_{t-1}, \quad t = 0, \pm 1, \pm 2, \ldots,$$

where θ is a parameter.

(a) Derive the theoretical mean value and autocovariance functions for the series x_t and w_t. Are the series x_t and w_t stationary? Give your reasons.
(b) Give a formula for the power spectrum of x_t, expressed in terms of θ and ω.

4.5 A first-order autoregressive model is generated from the white noise series w_t using the generating equations

$$x_t = \phi x_{t-1} + w_t,$$

where ϕ, for $|\phi| < 1$, is a parameter and the w_t are independent random variables with mean zero and variance σ_w^2.

(a) Show that the power spectrum of x_t is given by

$$f_x(\omega) = \frac{\sigma_w^2}{1 + \phi^2 - 2\phi\cos(2\pi\omega)}.$$

(b) Verify the autocovariance function of this process is

$$\gamma_x(h) = \frac{\sigma_w^2 \ \phi^{|h|}}{1-\phi^2},$$

$h = 0, \pm 1, \pm 2, \ldots$, by showing that the inverse transform of $\gamma_x(h)$ is the spectrum derived in part (a).

4.6 In applications, we will often observe series containing a signal that has been delayed by some unknown time D, i.e.,

$$x_t = s_t + As_{t-D} + n_t,$$

where s_t and n_t are stationary and independent with zero means and spectral densities $f_s(\omega)$ and $f_n(\omega)$, respectively. The delayed signal is multiplied by some unknown constant A.

(a) Prove

$$f_x(\omega) = [1 + A^2 + 2A\cos(2\pi\omega D)]f_s(\omega) + f_n(\omega).$$

(b) How could the periodicity expected in the spectrum derived in (a) be used to estimate the delay D? (Hint: Consider the case where $f_n(\omega) = 0$; i.e., there is no noise.)

4.7 Suppose x_t and y_t are stationary zero-mean time series with x_t independent of y_s for all s and t. Consider the product series

$$z_t = x_t y_t.$$

Prove the spectral density for z_t can be written as

$$f_z(\omega) = \int_{-1/2}^{1/2} f_x(\omega - \nu) f_y(\nu) \ d\nu.$$

Section 4.4

4.8 Figure 4.31 shows the biyearly smoothed (12-month moving average) number of sunspots from June 1749 to December 1978 with $n = 459$ points that were taken twice per year; the data are contained in `sunspotz`. With Example 4.10 as a guide, perform a periodogram analysis identifying the predominant periods and obtaining confidence intervals for the identified periods. Interpret your findings.

4.9 The levels of salt concentration known to have occurred over rows, corresponding to the average temperature levels for the soil science data considered in Figure 1.15 and Figure 1.16, are in `salt` and `saltemp`. Plot the series and then identify the dominant frequencies by performing separate spectral analyses on the two series. Include confidence intervals for the dominant frequencies and interpret your findings.

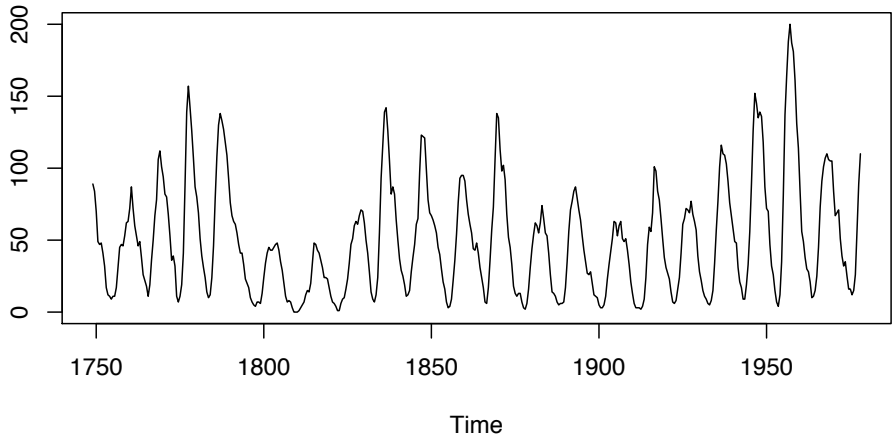

Fig. 4.31. Smoothed 12-month sunspot numbers (`sunspotz`) sampled twice per year.

4.10 Let the observed series x_t be composed of a periodic signal and noise so it can be written as

$$x_t = \beta_1 \cos(2\pi\omega_k t) + \beta_2 \sin(2\pi\omega_k t) + w_t,$$

where w_t is a white noise process with variance σ_w^2. The frequency ω_k is assumed to be known and of the form k/n in this problem. Suppose we consider estimating β_1, β_2 and σ_w^2 by least squares, or equivalently, by maximum likelihood if the w_t are assumed to be Gaussian.

(a) Prove, for a fixed ω_k, the minimum squared error is attained by

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = 2n^{-1/2} \begin{pmatrix} d_c(\omega_k) \\ d_s(\omega_k) \end{pmatrix},$$

where the cosine and sine transforms (4.23) and (4.24) appear on the right-hand side.

(b) Prove that the error sum of squares can be written as

$$SSE = \sum_{t=1}^{n} x_t^2 - 2I_x(\omega_k)$$

so that the value of ω_k that minimizes squared error is the same as the value that maximizes the periodogram $I_x(\omega_k)$ estimator (4.20).

(c) Under the Gaussian assumption and fixed ω_k, show that the F-test of no regression leads to an F-statistic that is a monotone function of $I_x(\omega_k)$.

4.11 Prove the convolution property of the DFT, namely,

$$\sum_{s=1}^{n} a_s x_{t-s} = \sum_{k=0}^{n-1} d_A(\omega_k) d_x(\omega_k) \exp\{2\pi\omega_k t\},$$

for $t = 1, 2, \ldots, n$, where $d_A(\omega_k)$ and $d_x(\omega_k)$ are the discrete Fourier transforms of a_t and x_t, respectively, and we assume that $x_t = x_{t+n}$ is periodic.

Section 4.5

4.12 Repeat Problem 4.8 using a nonparametric spectral estimation procedure. In addition to discussing your findings in detail, comment on your choice of a spectral estimate with regard to smoothing and tapering.

4.13 Repeat Problem 4.9 using a nonparametric spectral estimation procedure. In addition to discussing your findings in detail, comment on your choice of a spectral estimate with regard to smoothing and tapering.

4.14 The periodic behavior of a time series induced by echoes can also be observed in the spectrum of the series; this fact can be seen from the results stated in Problem 4.6(a). Using the notation of that problem, suppose we observe $x_t = s_t + As_{t-D} + n_t$, which implies the spectra satisfy $f_x(\omega) = [1 + A^2 + 2A\cos(2\pi\omega D)]f_s(\omega) + f_n(\omega)$. If the noise is negligible ($f_n(\omega) \approx 0$) then $\log f_x(\omega)$ is approximately the sum of a periodic component, $\log[1 + A^2 + 2A\cos(2\pi\omega D)]$, and $\log f_s(\omega)$. Bogart et al. (1962) proposed treating the detrended log spectrum as a pseudo time series and calculating its spectrum, or *cepstrum*, which should show a peak at a *quefrency* corresponding to $1/D$. The cepstrum can be plotted as a function of quefrency, from which the delaty D can be estimated.

For the speech series presented in Example 1.3, estimate the pitch period using cepstral analysis as follows. The data are in `speech`.

(a) Calculate and display the log-periodogram of the data. Is the periodogram periodic, as predicted?
(b) Perform a cepstral (spectral) analysis on the detrended logged periodogram, and use the results to estimate the delay D. How does your answer compare with the analysis of Example 1.24, which was based on the ACF?

4.15 Use Property 4.2 to verify (4.63). Then verify (4.66) and (4.67).

4.16 Consider two time series

$$x_t = w_t - w_{t-1},$$

$$y_t = \tfrac{1}{2}(w_t + w_{t-1}),$$

formed from the white noise series w_t with variance $\sigma_w^2 = 1$.

(a) Are x_t and y_t jointly stationary? Recall the cross-covariance function must also be a function only of the lag h and cannot depend on time.
(b) Compute the spectra $f_y(\omega)$ and $f_x(\omega)$, and comment on the difference between the two results.

(c) Suppose sample spectral estimators $\bar{f}_y(.10)$ are computed for the series using $L = 3$. Find a and b such that

$$P\left\{a \le \bar{f}_y(.10) \le b\right\} = .90.$$

This expression gives two points that will contain 90% of the sample spectral values. Put 5% of the area in each tail.

Section 4.6

4.17 Analyze the coherency between the temperature and salt data discussed in Problem 4.9. Discuss your findings.

4.18 Consider two processes

$$x_t = w_t \quad \text{and} \quad y_t = \phi x_{t-D} + v_t$$

where w_t and v_t are independent white noise processes with common variance σ^2, ϕ is a constant, and D is a fixed integer delay.

(a) Compute the coherency between x_t and y_t.
(b) Simulate $n = 1024$ normal observations from x_t and y_t for $\phi = .9$, $\sigma^2 = 1$, and $D = 0$. Then estimate and plot the coherency between the simulated series for the following values of L and comment:
(i) $L = 1$, (ii) $L = 3$, (iii) $L = 41$, and (iv) $L = 101$.

Section 4.7

4.19 For the processes in Problem 4.18:

(a) Compute the phase between x_t and y_t.
(b) Simulate $n = 1024$ observations from x_t and y_t for $\phi = .9$, $\sigma^2 = 1$, and $D = 1$. Then estimate and plot the phase between the simulated series for the following values of L and comment:
(i) $L = 1$, (ii) $L = 3$, (iii) $L = 41$, and (iv) $L = 101$.

4.20 Consider the bivariate time series records containing monthly U.S. production as measured by the Federal Reserve Board Production Index and monthly unemployment as given in Figure 3.22.

(a) Compute the spectrum and the log spectrum for each series, and identify statistically significant peaks. Explain what might be generating the peaks. Compute the coherence, and explain what is meant when a high coherence is observed at a particular frequency.

(b) What would be the effect of applying the filter

$$u_t = x_t - x_{t-1} \quad \text{followed by} \quad v_t = u_t - u_{t-12}$$

to the series given above? Plot the predicted frequency responses of the simple difference filter and of the seasonal difference of the first difference.

(c) Apply the filters successively to one of the two series and plot the output. Examine the output after taking a first difference and comment on whether stationarity is a reasonable assumption. Why or why not? Plot after taking the seasonal difference of the first difference. What can be noticed about the output that is consistent with what you have predicted from the frequency response? Verify by computing the spectrum of the output after filtering.

4.21 Determine the theoretical power spectrum of the series formed by combining the white noise series w_t to form

$$y_t = w_{t-2} + 4w_{t-1} + 6w_t + 4w_{t+1} + w_{t+2}.$$

Determine which frequencies are present by plotting the power spectrum.

4.22 Let $x_t = \cos(2\pi\omega t)$, and consider the output

$$y_t = \sum_{k=-\infty}^{\infty} a_k x_{t-k},$$

where $\sum_k |a_k| < \infty$. Show

$$y_t = |A(\omega)| \cos(2\pi\omega t + \phi(\omega)),$$

where $|A(\omega)|$ and $\phi(\omega)$ are the amplitude and phase of the filter, respectively. Interpret the result in terms of the relationship between the input series, x_t, and the output series, y_t.

4.23 Suppose x_t is a stationary series, and we apply two filtering operations in succession, say,

$$y_t = \sum_r a_r x_{t-r} \quad \text{then} \quad z_t = \sum_s b_s y_{t-s}.$$

(a) Show the spectrum of the output is

$$f_z(\omega) = |A(\omega)|^2 |B(\omega)|^2 f_x(\omega),$$

where $A(\omega)$ and $B(\omega)$ are the Fourier transforms of the filter sequences a_t and b_t, respectively.

(b) What would be the effect of applying the filter

$$u_t = x_t - x_{t-1} \quad \text{followed by} \quad v_t = u_t - u_{t-12}$$

to a time series?

(c) Plot the predicted frequency responses of the simple difference filter and of the seasonal difference of the first difference. Filters like these are called seasonal adjustment filters in economics because they tend to attenuate frequencies at multiples of the monthly periods. The difference filter tends to attenuate low-frequency trends.

4.24 Suppose we are given a stationary zero-mean series x_t with spectrum $f_x(\omega)$ and then construct the derived series

$$y_t = ay_{t-1} + x_t, \quad t = \pm 1, \pm 2, \ldots .$$

(a) Show how the theoretical $f_y(\omega)$ is related to $f_x(\omega)$.
(b) Plot the function that multiplies $f_x(\omega)$ in part (a) for $a = .1$ and for $a = .8$. This filter is called a recursive filter.

Section 4.8

4.25 Often, the periodicities in the sunspot series are investigated by fitting an autoregressive spectrum of sufficiently high order. The main periodicity is often stated to be in the neighborhood of 11 years. Fit an autoregressive spectral estimator to the sunspot data using a model selection method of your choice. Compare the result with a conventional nonparametric spectral estimator found in Problem 4.8.

4.26 Fit an autoregressive spectral estimator to the Recruitment series and compare it to the results of Example 4.13.

4.27 Suppose a sample time series with $n = 256$ points is available from the first-order autoregressive model. Furthermore, suppose a sample spectrum computed with $L = 3$ yields the estimated value $\bar{f}_x(1/8) = 2.25$. Is this sample value consistent with $\sigma_w^2 = 1, \phi = .5$? Repeat using $L = 11$ if we just happen to obtain the same sample value.

4.28 Suppose we wish to test the noise alone hypothesis $H_0 : x_t = n_t$ against the signal-plus-noise hypothesis $H_1 : x_t = s_t + n_t$, where s_t and n_t are uncorrelated zero-mean stationary processes with spectra $f_s(\omega)$ and $f_n(\omega)$. Suppose that we want the test over a band of $L = 2m + 1$ frequencies of the form $\omega_{j:n} + k/n$, for $k = 0, \pm 1, \pm 2, \ldots, \pm m$ near some fixed frequency ω. Assume that both the signal and noise spectra are approximately constant over the interval.

(a) Prove the approximate likelihood-based test statistic for testing H_0 against H_1 is proportional to

$$T = \sum_k |d_x(\omega_{j:n} + k/n)|^2 \left(\frac{1}{f_n(\omega)} - \frac{1}{f_s(\omega) + f_n(\omega)} \right).$$

(b) Find the approximate distributions of T under H_0 and H_1.
(c) Define the false alarm and signal detection probabilities as $P_F = P\{T > K|H_0\}$ and $P_d = P\{T > k|H_1\}$, respectively. Express these probabilities in terms of the signal-to-noise ratio $f_s(\omega)/f_n(\omega)$ and appropriate chi-squared integrals.

Section 4.9

4.29 Repeat the dynamic Fourier analysis of Example 4.21 on the remaining seven earthquakes and seven explosions in the data file `eqexp`. Do the conclusions about the difference between earthquakes and explosions stated in the example still seem valid?

4.30 Repeat the wavelet analyses of Example 4.22 and Example 4.23 on all earthquake and explosion series in the data file `eqexp`. Do the conclusions about the difference between earthquakes and explosions stated in Example 4.22 and Example 4.23 still seem valid?

4.31 Using Example 4.21–Example 4.23 as a guide, perform a dynamic Fourier analysis and wavelet analyses (dwt and waveshrink analysis) on the event of unknown origin that took place near the Russian nuclear test facility in Novaya Zemlya. State your conclusion about the nature of the event at Novaya Zemlya.

Section 4.10

4.32 Consider the problem of approximating the filter output

$$y_t = \sum_{k=-\infty}^{\infty} a_k x_{t-k}, \quad \sum_{-\infty}^{\infty} |a_k| < \infty,$$

by

$$y_t^M = \sum_{|k|<M/2} a_k^M x_{t-k}$$

for $t = M/2 - 1, M/2, \ldots, n - M/2$, where x_t is available for $t = 1, \ldots, n$ and

$$a_t^M = M^{-1} \sum_{k=0}^{M-1} A(\omega_k) \exp\{2\pi i \omega_k t\}$$

with $\omega_k = k/M$. Prove

$$E\{(y_t - y_t^M)^2\} \le 4\gamma_x(0) \Bigg(\sum_{|k|\ge M/2} |a_k| \Bigg)^2.$$

4.33 Prove the squared coherence $\rho^2_{y\cdot x}(\omega) = 1$ for all ω when

$$y_t = \sum_{r=-\infty}^{\infty} a_r x_{t-r},$$

that is, when x_t and y_t can be related exactly by a linear filter.

4.34 The data set `climhyd`, contains 454 months of measured values for six climatic variables: (i) air temperature [`Temp`], (ii) dew point [`DewPt`], (iii) cloud cover [`CldCvr`], (iv) wind speed [`WndSpd`], (v) precipitation [`Precip`], and (vi) inflow [`Inflow`], at Lake Shasta in California; the data are displayed in Figure 7.3. We would like to look at possible relations among the weather factors and between the weather factors and the inflow to Lake Shasta.

(a) First transform the inflow and precipitation series as follows: $I_t = \log i_t$, where i_t is inflow, and $P_t = \sqrt{p}_t$, where p_t is precipitation. Then, compute the square coherencies between all the weather variables and transformed inflow and argue that the strongest determinant of the inflow series is (transformed) precipitation. [*Tip:* If `x` contains multiple time series, then the easiest way to display all the squared coherencies is to first make an object of class *spec*; e.g., `u = spectrum(x, span=c(7,7), plot=FALSE)` and then plot the coherencies suppressing the confidence intervals, `plot(u, ci=-1, plot.type="coh")`.]

(b) Fit a lagged regression model of the form

$$I_t = \beta_0 + \sum_{j=0}^{\infty} \beta_j P_{t-j} + w_t,$$

using thresholding, and then comment of the predictive ability of precipitation for inflow.

Section 4.11

4.35 Consider the *signal plus noise* model

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t,$$

where the signal and noise series, x_t and v_t are both stationary with spectra $f_x(\omega)$ and $f_v(\omega)$, respectively. Assuming that x_t and v_t are independent of each other for all t, verify (4.137) and (4.138).

4.36 Consider the model

$$y_t = x_t + v_t,$$

where

$$x_t = \phi x_{t-1} + w_t,$$

such that v_t is Gaussian white noise and independent of x_t with $\text{var}(v_t) = \sigma_v^2$, and w_t is Gaussian white noise and independent of v_t, with $\text{var}(w_t) = \sigma_w^2$, and $|\phi| < 1$ and $Ex_0 = 0$. Prove that the spectrum of the observed series y_t is

$$f_y(\omega) = \frac{\sigma^2 |1 - \theta e^{-2\pi i \omega}|^2}{|1 - \phi e^{-2\pi i \omega}|^2},$$

where

$$\theta = \frac{c \pm \sqrt{c^2 - 4}}{2}, \quad \sigma^2 = \frac{\sigma_v^2 \phi}{\theta},$$

and

$$c = \frac{\sigma_w^2 + \sigma_v^2(1 + \phi^2)}{\sigma_v^2 \phi}.$$

4.37 Consider the same model as in the preceding problem.

(a) Prove the optimal smoothed estimator of the form

$$\widehat{x}_t = \sum_{s=-\infty}^{\infty} a_s y_{t-s}$$

has

$$a_s = \frac{\sigma_w^2}{\sigma^2} \frac{\theta^{|s|}}{1 - \theta^2}.$$

(b) Show the mean square error is given by

$$E\{(x_t - \widehat{x}_t)^2\} = \frac{\sigma_v^2 \sigma_w^2}{\sigma^2(1 - \theta^2)}.$$

(c) Compare mean square error of the estimator in part (b) with that of the optimal finite estimator of the form

$$\widehat{x}_t = a_1 y_{t-1} + a_2 y_{t-2}$$

when $\sigma_v^2 = .053, \sigma_w^2 = .172$, and $\phi_1 = .9$.

Section 4.12

4.38 Consider the two-dimensional linear filter given as the output (4.149).

(a) Express the two-dimensional autocovariance function of the output, say, $\gamma_y(h_1, h_2)$, in terms of an infinite sum involving the autocovariance function of $x_{\boldsymbol{s}}$ and the filter coefficients a_{s_1,s_2}.
(b) Use the expression derived in (a), combined with (4.148) and (4.151) to derive the spectrum of the filtered output (4.150).

The following problems require supplemental material from Appendix C

4.39 Let w_t be a Gaussian white noise series with variance σ_w^2. Prove that the results of Theorem C.4 hold without error for the DFT of w_t.

4.40 Show that condition (4.40) implies (C.19) by showing

$$n^{-1/2} \sum_{h \geq 0} h\,|\gamma(h)| \leq \sigma_w^2 \sum_{k \geq 0} |\psi_k| \sum_{j \geq 0} \sqrt{j}\,|\psi_j|.$$

4.41 Prove Lemma C.4.

4.42 Finish the proof of Theorem C.5.

4.43 For the zero-mean complex random vector $\boldsymbol{z} = \boldsymbol{x}_c - i\boldsymbol{x}_s$, with $\text{cov}(\boldsymbol{z}) = \Sigma = C - iQ$, with $\Sigma = \Sigma^*$, define

$$w = 2\text{Re}(\boldsymbol{a}^*\boldsymbol{z}),$$

where $\boldsymbol{a} = \boldsymbol{a}_c - i\boldsymbol{a}_s$ is an arbitrary non-zero complex vector. Prove

$$\text{cov}(w) = 2\boldsymbol{a}^*\Sigma\boldsymbol{a}.$$

Recall * denotes the complex conjugate transpose.