Spectral Analysis of Univariate and Bivariate Time Series

Donald B. Percival
Senior Mathematician
Applied Physics Laboratory
HN–10
University of Washington
Seattle, WA 98195

October 9, 1993

11.1 Introduction

The spectral analysis of time series is one of the most commonly used data analysis techniques in the physical sciences. The basis for this analysis is a representation for a time series in terms of a linear combination of sinusoids with different frequencies and amplitudes. This type of representation is called a *Fourier representation*. If the time series is sampled at instances in time spaced Δt units apart and if the series is a realization of one portion of a real-valued stationary process $\{X_t\}$ with zero mean, then we have the representation (due to Cramér [1])

$$X_t = \int_{-f_{(N)}}^{f_{(N)}} e^{i2\pi f t\,\Delta t}\, dZ(f), \qquad t = 0, \pm 1, \pm 2, \ldots, \tag{11.1}$$

where $f_{(N)} = 1/(2\,\Delta t)$ is the Nyquist frequency (if the units of Δt are measured in, say, seconds, then $f_{(N)}$ is measured in Hertz (Hz), i.e., cycles per second); $i = \sqrt{-1}$; and $\{Z(f)\}$ is an orthogonal process (a complex-valued stochastic process with quite special properties). This representation is rather formidable at first glance, but the main idea is simple: since, by definition, $e^{i2\pi f t\,\Delta t} = \cos(2\pi f t\,\Delta t) + i \sin(2\pi f t\,\Delta t)$, Equation 11.1 says that we can express X_t as a linear combination of sinusoids at different frequencies f, with the sinusoids at frequency f receiving a random amplitude generated by the increment $dZ(f) = Z(f + df) - Z(f)$ (here df is a small positive increment in frequency). The expected value of the squared magnitude of this random amplitude defines the *spectral density function* $S_X(\cdot)$ for the stationary process $\{X_t\}$ in the following way:

$$E[|dZ(f)|^2] = S_X(f)\, df, \qquad -f_{(N)} \le f \le f_{(N)}$$

(the notation '$E(X)$' refers to the expected value (mean) of the random variable (rv) X). Because $|dZ(f)|^2$ is a nonnegative rv, its expectation must be nonnegative, and hence the sdf $S_X(\cdot)$ is a nonnegative function of frequency. Large values of the sdf tell us which frequencies in Equation 11.1 contribute the most in constructing the process $\{X_t\}$. (We have glossed over many details here, including the fact that a 'proper' sdf does not exist for some stationary processes unless we allow use of the Dirac delta function. See Koopmans [2] or Priestley [3] for a precise statement and proof of Cramér's spectral representation theorem or Section 4.1 of Percival and Walden [4] for a heuristic development.)

Because $\{X_t\}$ is a real-valued process, the sdf is an even function; i.e., $S_X(-f) = S_X(f)$. Our definition for the sdf is 'two-sided' because it uses both positive and negative frequencies, the latter being a nonphysical – but mathematically convenient – concept. Some branches of the physical

sciences routinely use a 'one-sided' sdf that, in terms of our definition, is equal to $2S_X(f)$ over the interval $[0, f_{(N)}]$.

Let us denote the τth component of the *autocovariance sequence* (acvs) for $\{X_t\}$ as $C_{\tau,X}$; i.e.,

$$C_{\tau,X} = Cov(X_t, X_{t+\tau}) = E(X_t X_{t+\tau})$$

(the notation '$Cov(X, Y)$' refers to the covariance between the rv's X and Y). The spectral representation in Equation 11.1 can be used to derive the important relationship

$$C_{\tau,X} = \int_{-f_{(N)}}^{f_{(N)}} S_X(f) e^{i2\pi f \tau \, \Delta t} \, df \tag{11.2}$$

(for details, see [4], Section 4.1). In words, $S_X(f)$ is the (nonrandom) amplitude associated with the frequency f in the above Fourier representation for the acvs $\{C_{\tau,X}\}$. If we recall that $C_{0,X}$ is just the process variance, we obtain (by setting $\tau = 0$ in the above equation)

$$V(X_t) = C_{0,X} = \int_{-f_{(N)}}^{f_{(N)}} S_X(f) \, df$$

(the notation '$V(X)$' refers the variance of the rv X). The sdf thus represents a decomposition of the process variance into components attributable to different frequencies. In particular, if we were to run the process $\{X_t\}$ through a narrow-band filter with bandwidth df centered at frequencies $\pm f$, the variance of the process coming out of the filter would be approximately given by $2S_X(f) \, df$ (the factor of 2 arises because $S_X(\cdot)$ is a two-sided sdf). Spectral analysis is an analysis of variance technique in which we portion out contributions to $V(X_t)$ across different frequencies. Because variance is closely related to the concept of power, $S_X(\cdot)$ is sometimes referred to as a *power spectral density function*.

In this chapter we discuss estimation of the sdf $S_X(\cdot)$ based upon a time series that can be regarded as a realization of a portion $X_1, \ldots, X_n$ of a stationary process. The problem of estimating $S_X(\cdot)$ in general is quite complicated, due both to the wide variety of sdf's that arise in physical applications and also to the large number of specific uses for spectral analysis. To focus our discussion, we use as examples two time series that are fairly representative of many in the physical sciences; however, there are some important issues that these series do not address and others that we must gloss over due to space (for a more detailed exposition of spectral analysis with a physical science orientation, see [4]). The two series are shown in Figure 11.1 and are a record of the height of ocean waves as a

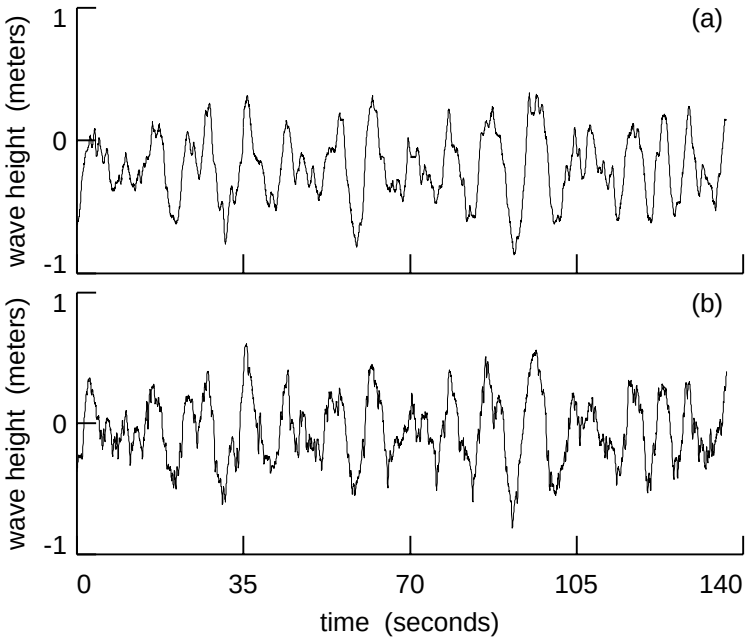

Figure 11.1: Plot of height of ocean waves versus time as measured by a wire wave gauge (a) and an infrared wave gauge (b). Both series were collected at a rate of 30 samples per second. There are $n = 4096$ data values in each series. (These series were supplied through the courtesy of A. T. Jessup, Applied Physics Laboratory, University of Washington. As of 1993, they could be obtained via electronic mail by sending a message with the single line '`send saubts from datasets`' to the Internet address `statlib@lib.stat.cmu.edu` – this is the address for StatLib, a statistical archive maintained by Carnegie Mellon University).

function of time as measured by two instruments of quite different design. Both instruments were mounted 6 meters apart on the same platform off Cape Henry near Virginia Beach, Virginia. One instrument was a wire wave gauge, while the other was an infrared wave gauge. The sampling frequency for both instruments was 30 Hz (30 samples per second) so the sampling period is $\Delta t = 1/30$ second and the Nyquist frequency is $f_{(N)} = 15$ Hz. The series were collected mainly to study the sdf of ocean waves for frequencies from 0.4 to 4 Hz. The frequency responses of the instruments are similar only over certain frequency ranges. As we shall see, the infrared wave gauge inadvertently increases the power in the measured spectra by an order of magnitude at frequencies 0.8 to 4 Hz. The power spectra for

the time series have a relatively large dynamic range (greater than 50 dB), as is often true in the physical sciences. Because the two instruments were 6 meters apart and because of the prevalent direction of the ocean waves, there is a lead/lag relationship between the two series. (For more details, see Jessup *et al.* [5] and references therein.)

11.2 Univariate Time Series

11.2.1 The Periodogram

Suppose we have a time series of length n that is a realization of a portion $X_1, X_2, \ldots, X_n$ of a zero mean real-valued stationary process with sdf $S_X(\cdot)$ and acvs $\{C_{\tau,X}\}$ (note that, if $E(X_t)$ is unknown and hence cannot be assumed to be zero, the common practice is to replace X_t with $X_t - \bar{X}$ prior to all other computations, where $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$ is the sample mean). Under a mild regularity condition (such as $S_X(\cdot)$ having a finite derivative at all frequencies), we can then write

$$S_X(f) = \Delta t \sum_{\tau=-\infty}^{\infty} C_{\tau,X} e^{-i2\pi f \tau \Delta t}. \tag{11.3}$$

Our task is to estimate the sdf $S_X(\cdot)$ based upon $X_1, \ldots, X_n$. Equation 11.3 suggests the following 'natural' estimator. Suppose that, for $|\tau| \le n-1$, we estimate $C_{\tau,X}$ via

$$\hat{C}^{(p)}_{\tau,X} = \frac{1}{n} \sum_{t=1}^{n-|\tau|} X_t X_{t+|\tau|}$$

(the rationale for the superscript '(p)' is explained below). The estimator $\hat{C}^{(p)}_{\tau,X}$ is known in the literature as the *biased estimator* of $C_{\tau,X}$ since its expected value is

$$E(\hat{C}^{(p)}_{\tau,X}) = \frac{1}{n} \sum_{t=1}^{n-|\tau|} E(X_t X_{t+|\tau|}) = \left(1 - \frac{|\tau|}{n}\right) C_{\tau,X}, \tag{11.4}$$

and hence $E(\hat{C}^{(p)}_{\tau,X}) \neq C_{\tau,X}$ in general. If we now decree that $\hat{C}^{(p)}_{\tau,X} = 0$ for $|\tau| \ge n$ and substitute the $\hat{C}^{(p)}_{\tau,X}$'s for the $C_{\tau,X}$'s in Equation 11.3, we obtain the spectral estimator

$$\hat{S}^{(p)}_X(f) = \Delta t \sum_{\tau=-(n-1)}^{n-1} \hat{C}^{(p)}_{\tau,X} e^{-i2\pi f \tau \Delta t}. \tag{11.5}$$

This estimator is known in the literature as the *periodogram* – hence the superscript '(p)' – even though it is more natural to regard it as a function of frequency f than of period $1/f$. By substituting the definition for $\hat{C}_{\tau,X}^{(p)}$ into the above equation and making a change of variables, we find also that

$$\hat{S}_X^{(p)}(f) = \frac{\Delta t}{n} \left| \sum_{t=1}^{n} X_t e^{-i2\pi f t\, \Delta t} \right|^2 . \tag{11.6}$$

Hence we can interpret the periodogram in two ways: it is the Fourier transform of the biased estimator of the acvs (with $\hat{C}_{\tau,X}^{(p)}$ defined to be zero for $|\tau| \geq n$), and it is – to within a scaling factor – the squared modulus of the Fourier transform of $X_1, \ldots, X_n$.

Let us now consider the statistical properties of the periodogram. Ideally, we might like the following to be true:

1. $E[\hat{S}_X^{(p)}(f)] \approx S_X(f)$ (approximately unbiased);
2. $V[\hat{S}_X^{(p)}(f)] \to 0$ as $n \to \infty$ (consistent); and
3. $Cov[\hat{S}_X^{(p)}(f), \hat{S}_X^{(p)}(f')] \approx 0$ for $f \neq f'$ (approximately uncorrelated).

The 'tragedy of the periodogram' is that in fact

1. $\hat{S}_X^{(p)}(f)$ can be a badly biased estimator of $S_X(f)$ even for large sample sizes (Thomson [6] reports an example in which the periodogram is severely biased for $n = 1.2$ *million* data points) and
2. $V[\hat{S}_X^{(p)}(f)]$ does *not* decrease to 0 as $n \to \infty$ (unless $S_X(f) = 0$, a case of little practical interest).

As a consolation, however, we do have that $\hat{S}_X^{(p)}(f)$ and $\hat{S}_X^{(p)}(f')$ are approximately uncorrelated under certain conditions (see below).

We can gain considerable insight into the nature of the bias in the periodogram by studying the following expression for its expected value:

$$E[\hat{S}_X^{(p)}(f)] = \int_{-f_{(N)}}^{f_{(N)}} \mathcal{F}(f - f') S_X(f')\, df', \quad \text{with } \mathcal{F}(f) = \frac{\Delta t \sin^2(n\pi f\, \Delta t)}{n \sin^2(\pi f\, \Delta t)} \tag{11.7}$$

(for details, see [4], Section 6.3). The function $\mathcal{F}(\cdot)$ is known as *Fejér's kernel*. We also call it the *spectral window* for the periodogram. Figure 11.2(a) shows $\mathcal{F}(f)$ versus f with $-f_{(N)} \leq f \leq f_{(N)}$ for the case $n = 32$ with $\Delta t = 1$ so that $f_{(N)} = 1/2$ (note that $\mathcal{F}(-f) = \mathcal{F}(f)$; i.e., Fejér's kernel is an even function). Figure 11.2(b) plots $10 \cdot \log_{10}(\mathcal{F}(f))$ versus f for $0 \leq f \leq 1/2$

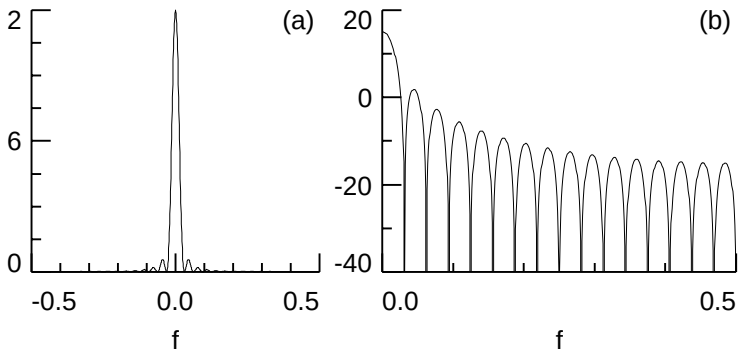

Figure 11.2: Fejér's kernel for sample size $n = 16$ with $f_{(N)} = 1/2$.

(i.e., $\mathcal{F}(\cdot)$ on a decibel scale). The numerator of Equation 11.7 tells us that $\mathcal{F}(f) = 0$ when the product $nf\,\Delta t$ is equal to a nonzero integer – there are 16 of these nulls evident in 11.2(b). The nulls closest to zero frequency occur at $f = \pm 1/(n\,\Delta t) = \pm 1/32$. Figure 11.2(a) indicates that $\mathcal{F}(\cdot)$ is concentrated mainly in the interval of frequencies between these two nulls, the region of the 'central lobe' of Fejér's kernel. A convenient measure of this concentration is the ratio $\int_{-1/(N\Delta t)}^{1/(N\Delta t)} \mathcal{F}(f)\,df \Big/ \int_{-f_{(N)}}^{f_{(N)}} \mathcal{F}(f)\,df$. An easy exercise shows that the denominator is unity for all n, while – to two decimal places – the numerator is equal to 0.90 for all $n \geq 13$. As $n \to \infty$, the length of the interval over which 90% of $\mathcal{F}(\cdot)$ is concentrated shrinks to zero, so in the limit Fejér's kernel acts like a Dirac delta function. If $S_X(\cdot)$ is continuous at f, Equation 11.7 tells us that $\lim_{n\to\infty} E[\hat{S}_X^{(p)}(f)] = S_X(f)$; i.e., the periodogram is asymptotically unbiased.

While this asymptotic result is of some interest, for practical applications we are much more concerned about possible biases in the periodogram for finite sample sizes n. Equation 11.7 tells us that the expected value of the periodogram is given by the convolution of the true sdf with Fejér's kernel. Convolution is often regarded as a smoothing operation. From this viewpoint, $E[\hat{S}_X^{(p)}(\cdot)]$ should be a smoothed version of $S_X(\cdot)$ – hence, if $S_X(\cdot)$ is itself sufficiently smooth, $E[\hat{S}_X^{(p)}(\cdot)]$ should closely approximate $S_X(\cdot)$. An extreme example of a process with a smooth sdf is white noise. Its sdf is constant over all frequencies, and in fact $E[\hat{S}_X^{(p)}(f)] = S_X(f)$ for a white noise process.

For sdf's with more structure than white noise, we can identify two sources of bias in the periodogram. The first source is often called a *loss of resolution* and is due to the fact that the central lobe of Fejér's kernel

will tend to smooth out spectral features with widths less than $1/(n\,\Delta t)$. Unfortunately, unless *a priori* information is available (or we are willing to make a modeling assumption), the cure for this bias is to increase the sample size n, i.e., to collect a longer time series, the prospect of which might be costly or – in the case of certain geophysical time series spanning thousands of years – impossible within our lifetimes.

The second source of bias is called *leakage* and is attributable to the sidelobes in Fejér's kernel. These sidelobes are prominently displayed in Figure 11.2(b). Figure 11.3 illustrates how these sidelobes can induce bias in the periodogram. The thick curve in Figure 11.3(a) shows an sdf plotted on a decibel scale from $f = -f_{(N)}$ to $f = f_{(N)}$ with $f_{(N)} = 1/2$ (recall that the sdf is symmetric about 0 so that $S_X(-f) = S_X(f)$). The thin bumpy curve is Fejér's kernel for $n = 32$, shifted so that its central lobe is at $f = 0.2$. The product of this shifted kernel and the sdf is shown in Figure 11.3(b) (again on a decibel scale). Equation 11.7 says that $E[\hat{S}_X^{(p)}(0.2)]$ is the integral of this product. The plot shows this integral to be mainly determined by values close to $f = 0.2$; i.e., $E[\hat{S}_X^{(p)}(0.2)]$ is largely due to the sdf at values close to this frequency, a result that is quite reasonable. Figures 11.3(c) and (d) show the corresponding plots for $f = 0.4$. Note that $E[\hat{S}_X^{(p)}(0.4)]$ is substantially influenced by values of $S_X(\cdot)$ away from $f = 0.4$. The problem is that the sidelobes of Fejér's kernel are interacting with portions of the sdf that are the dominant contributors to the variance of the process so that $E[\hat{S}_X^{(p)}(0.4)]$ is biased upwards. Figure 11.3(e) shows a plot of $E[\hat{S}_X^{(p)}(f)]$ versus f (the thin curve), along with the true sdf $S_X(\cdot)$ (the thick curve). While the periodogram is essentially unbiased for frequencies satisfying $0.1 \le |f| \le 0.35$, there is substantial bias due to leakage at frequencies close to $f = 0$ and $f = \pm 1/2$ (in the latter case, the bias is almost 40 dB, i.e., 4 orders of magnitude).

While it is important to know that the periodogram can be severely biased for certain processes, it is also true that, if the true sdf is sufficiently lacking in structure (i.e., 'close to white noise'), then $S_X(\cdot)$ and $E[\hat{S}_X^{(p)}(\cdot)]$ can be close enough to each other so that the periodogram is essentially bias free. Furthermore, even if leakage is present, it might not be of importance in certain practical applications. If, for example, we were performing a spectral analysis to determine the height and structure of the sdf in Figure 11.3 near $f = 0.2$, then the bias due to leakage at other frequencies is of little concern.

If the portions of the sdf affected by leakage are in fact of interest or if we are carrying out a spectral analysis on a time series for which little is known *a prior* about its sdf, we need to find ways to recognize when leakage is a problem and, if it is present, to minimize it. As is the case for loss of

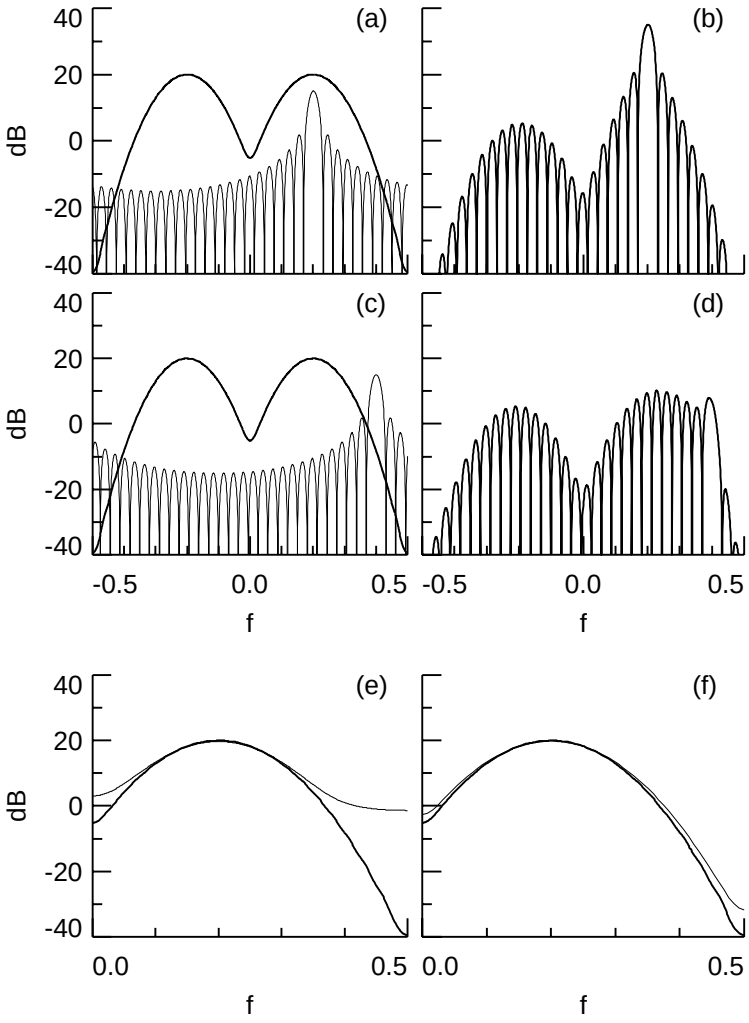

Figure 11.3: Illustration of leakage (plots (a) to (e)). The Nyquist frequency $f_{(N)}$ here is taken to be 1/2. Plot (f) shows the alleviation of leakage via tapering and is discussed in Section 11.2.2.

resolution, we can decrease leakage by increasing the sample size (more data can solve many problems!), but unfortunately a rather substantial increase might be required to obtain a periodogram that is essentially free of leakage. Consider again the sdf used as an example in Figure 11.3. Even with a 32 fold increase in the sample size from $n = 32$ to 1024, there is still more than a 20 dB difference between $E[\hat{S}_X^{(p)}(0.5)]$ and $S_X(0.5)$.

If we regard the sample size as fixed, there are two well-known ways of decreasing leakage, namely, *data tapering* and *prewhitening*. Both of these techniques have a simple interpretation in terms of the integral in Equation 11.7. On the one hand, tapering essentially replaces Fejér's kernel $\mathcal{F}(\cdot)$ by a function with substantially reduced sidelobes; on the other hand, prewhitening effectively replaces the sdf $S_X(\cdot)$ with one that is closer to white noise. Both techniques are discussed in the next subsection.

11.2.2 Correcting for Bias

Tapering

For a given time series $X_1, X_2, \ldots, X_n$, a *data taper* is a finite sequence $h_1, h_2, \ldots, h_n$ of real-valued numbers. The product of this sequence and the time series, namely, $h_1X_1, h_2X_2, \ldots, h_nX_n$, is used to create a *direct spectral estimator* of $S_X(f)$, defined as

$$\hat{S}_X^{(d)}(f) = \Delta t \left| \sum_{t=1}^{n} h_t X_t e^{-i2\pi f t\, \Delta t} \right|^2 . \tag{11.8}$$

Note that, if we let $h_t = 1/\sqrt{n}$ for all t (the so-called rectangular data taper), a comparison of Equations 11.8 and 11.6 tells us that $\hat{S}_X^{(d)}(\cdot)$ reduces to the periodogram. The acvs estimator corresponding to $\hat{S}_X^{(d)}(\cdot)$ is just

$$\hat{C}_{\tau,X}^{(d)} = \sum_{t=1}^{n-|\tau|} h_t X_t h_{t+|\tau|} X_{t+|\tau|}, \text{ and so } E(\hat{C}_{\tau,X}^{(d)}) = C_{\tau,X} \sum_{t=1}^{n-|\tau|} h_t h_{t+|\tau|}. \tag{11.9}$$

If we insist that $\hat{C}_{0,X}^{(d)}$ be an unbiased estimator of the process variance $C_{0,X}$, then we obtain the normalization $\sum_{t=1}^{n} h_t^2 = 1$ (note that the rectangular data taper satisfies this constraint).

The rationale for tapering is to obtain a spectral estimator whose expected value is close to $S_X(\cdot)$. In analogy to Equation 11.7, we can express this expectation as

$$E[\hat{S}_X^{(d)}(f)] = \int_{-f_{(N)}}^{f_{(N)}} \mathcal{H}(f - f') S_X(f')\, df', \tag{11.10}$$

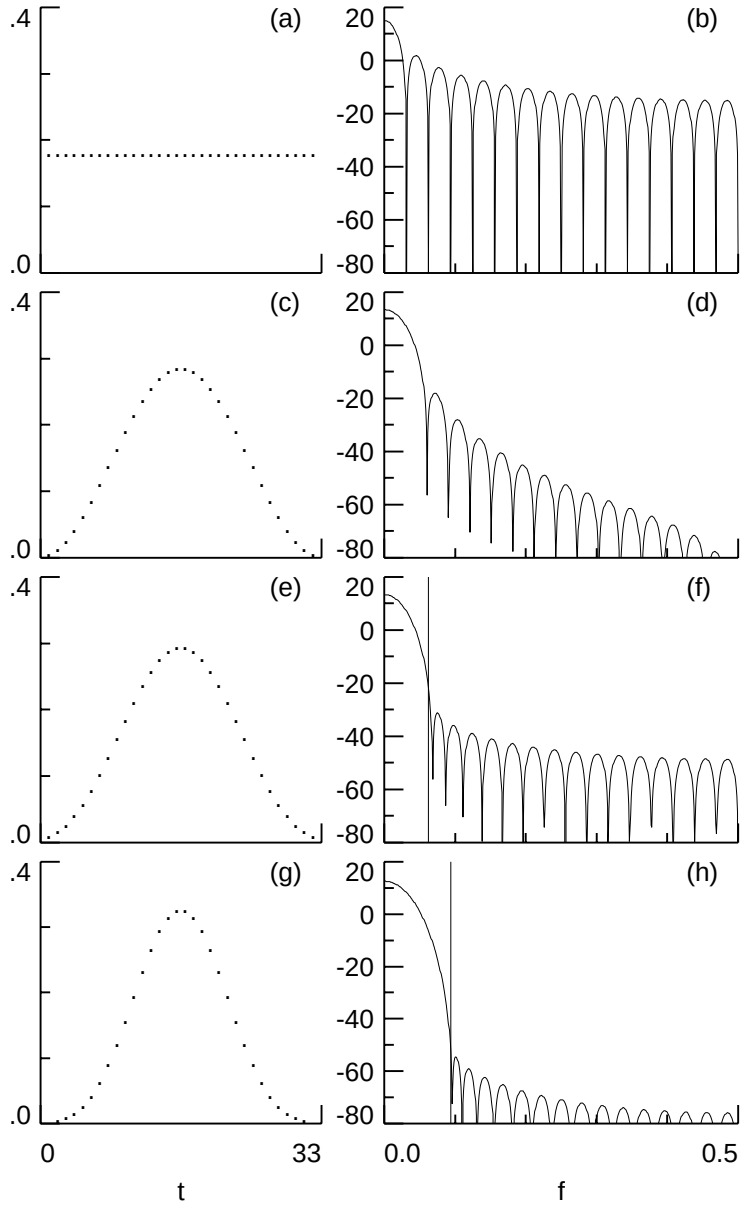

Figure 11.4: Four data tapers (left-hand column of plots) and their spectral windows (right-hand column) for a sample size of $n = 32$. From top to bottom, the tapers are the rectangular taper, the Hanning taper, the dpss data taper with $nW = 2/\Delta t$ and the dpss data taper with $nW = 3/\Delta t$ (here $\Delta t = 1$).

where $\mathcal{H}(\cdot)$ is proportional to the squared modulus of the Fourier transform of the h_t's and is called the *spectral window* for the direct spectral estimator $\hat{S}_X^{(d)}(\cdot)$ (just as we called Fejér's kernel the spectral window for the periodogram). The claim is that, with a proper choice of h_t's, we can produce a spectral window that offers better protection against leakage than Fejér's kernel. Figure 11.4 supports this claim. The left-hand column of plots shows four data tapers for sample size $n = 32$, while the right-hand plots show the corresponding spectral windows. For the sake of comparison, the data taper in 11.4(a) is just the rectangular data taper, so the corresponding spectral window is Fejér's kernel (cf. Figure 11.2(b)). The data taper in 11.4(c) is the well-known *Hanning data taper*, which we define to be $h_t = \sqrt{2/3(n+1)}\,[1 + \cos(2\pi t/(n+1))]$ for $t = 1, \ldots, n$ (there are other – slightly different – definitions in the literature). Note carefully the shape of the corresponding spectral window in 11.4(d): its sidelobes are considerably suppressed in comparison to those of Fejér's kernel, but the width of its central lobe is markedly larger. The convolutional representation for $E[\hat{S}_X^{(d)}(f)]$ given in Equation 11.10 tells us that an increase in central lobe width can result in a loss of resolution when the true sdf has spectral features with widths smaller than the central lobe width. This illustrates one of the tradeoffs in using a data taper, namely, that tapering typically decreases leakage at the expense of a potential loss of resolution.

A convenient family of data tapers that facilitates the tradeoff between sidelobe suppression and the width of the central lobe is the *discrete prolate spheroidal sequence* (dpss) tapers. These tapers arise as the solution to the following 'concentration' problem. Suppose we pick a number W that – roughly speaking – we think of as half the desired width of the central lobe of the resulting spectral window (typically $1/(n\,\Delta t) \le W \le 4/(n\,\Delta t)$ although larger values for W are sometimes useful). Under the requirement that the taper must satisfy the normalization $\sum_{t=1}^{n} h_t^2 = 1$, the dpss taper is, by definition, the taper whose corresponding spectral window is as concentrated as possible in the frequency interval $[-W, W]$ in the sense that the ratio $\int_{-W}^{W} \mathcal{H}(f)\,df \,/\, \int_{-f_{(N)}}^{f_{(N)}} \mathcal{H}(f)\,df$ is as large as possible (note that, if $\mathcal{H}(\cdot)$ were a Dirac delta function, this ratio would be unity). The quantity $2W$ is sometimes called the *resolution bandwidth*. For a fixed W, the dpss tapers have sidelobes that are suppressed as much as possible as measured by the concentration ratio. To a good approximation (Walden [7]), the dpss tapers can be calculated as $h_t = C \times I_0(\widetilde{W}\sqrt{1-(1-g_t)^2})/I_0(\widetilde{W})$ for $t = 1, \ldots, n$, where C is a scaling constant used to force the normalization $\sum h_t^2 = 1$; $\widetilde{W} = \pi W(n-1)\,\Delta t$; $g_t = (2t-1)/n$; and $I_0(\cdot)$ is the modified Bessel function of the first kind and zeroth order (this can be computed using the Fortran function `bessj0` in Section 6.5 of Press *et al.* [8]).

Figures 11.4(e) and (g) show dpss tapers for $n = 32$ and with W set such that $nW = 2/\Delta t$ and $nW = 3/\Delta t$, while plots (f) and (h) show the corresponding spectral windows $\mathcal{H}(\cdot)$ (the quantity $2nW$ is known as the duration-bandwidth product). The vertical lines in the latter two plots mark the locations of W. Note that in both cases the central lobe of $\mathcal{H}(\cdot)$ is approximately contained between $[-W, W]$. As expected, increasing W suppresses the sidelobes and hence offers increasing protection against leakage. A comparison of the spectral windows for the Hanning and $nW = 2/\Delta t$ dpss tapers (Figures 11.4(d) and (f)) shows that, whereas their central lobes are comparable, their sidelobe structures are quite different. While specific examples can be constructed in which one of the tapers offers better protection against leakage than the other, generally the two tapers are quite comparable in practical applications. The advantage of the dpss tapers is that, if, say, use of an $nW = 2/\Delta t$ dpss taper produces a direct spectral estimator that still suffers from leakage, we can easily obtain a greater degree of protection against leakage by merely increasing nW beyond $2/\Delta t$. The choice $nW = 1/\Delta t$ yields a spectral window with a central lobe closely resembling that of Fejér's kernel but with sidelobes about 10 dB smaller.

Let us now return to the example of Figure 11.3. Recall that the thin curve in 11.3(e) shows $E[\hat{S}_X^{(p)}(f)]$ versus f for a process with an sdf given by the thick curve. In 11.3(f) the thin curve now shows $E[\hat{S}_X^{(d)}(\cdot)]$ for a direct spectral estimator employing an $nW = 2/\Delta t$ dpss taper. Note that tapering has produced a spectral estimator that is overall much closer in expectation to $S_X(\cdot)$ than the periodogram is; however, mainly due to the small sample $n = 32$, $\hat{S}_X^{(d)}(\cdot)$ still suffers from leakage at some frequencies (about 10 dB at $f = 0.5$).

In practical situations, we can determine if leakage is present in the periodogram by carefully comparing it with a direct spectral estimate constructed using a dpss data taper with a fairly large value of W. As an example, Figure 11.5(a) shows the periodogram for the wire wave gauge time series shown in Figure 11.1. Since these data were collected mainly to investigate the rolloff rate of the sdf from 0.8 to 4 Hz, we have only plotted the low frequency portion of the periodogram. Figure 11.5(b) shows a direct spectral estimate for which we have used a dpss taper with $nW = 4/\Delta t$. Note that the direct spectral estimate is markedly lower than the periodogram at frequencies with a relatively small contribution to the overall variance – for example, the former is about 10 dB below the latter at frequencies close to 4 Hz. This pattern is consistent with what we would expect to see when there is leakage in the periodogram. Increasing nW beyond $4/\Delta t$ to, say, $6/\Delta t$ and comparing the resulting direct direct spectral estimate with that of Figure 11.5(b) indicates that the $nW = 4/\Delta t$

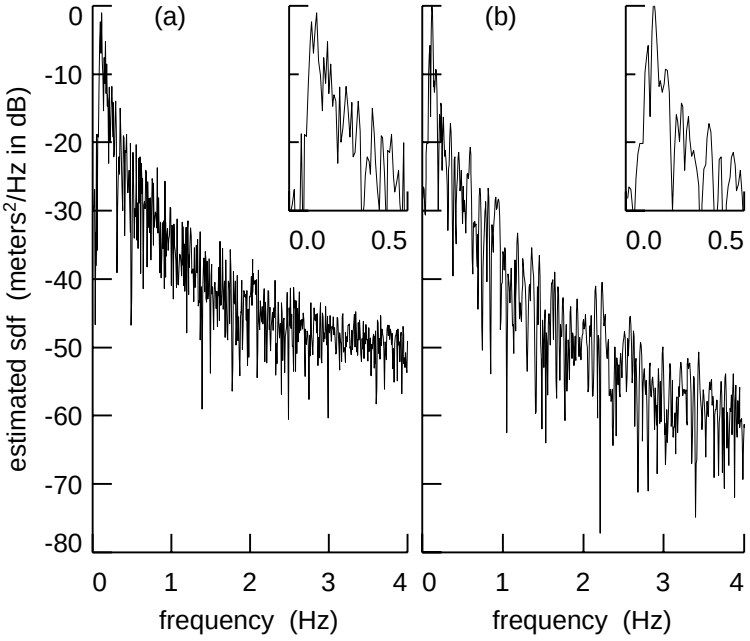

Figure 11.5: Periodogram (plot (a)) and a direct spectral estimate (plot (b)) using an $nW = 4/\Delta t$ dpss data taper for the wire gauge time series of Figure 1(a). Both estimates are plotted on a decibel scale. The small subplots in the upper right-hand corner of each plot give an expanded view of the estimators at $0 \leq f \leq 0.5$ Hz (the vertical scales of the subplots are the same as those of the main plots).

estimate is essentially leakage free; on the other hand, an examination of an $nW = 1/\Delta t$ direct spectral estimate indicates that it is also essentially leakage free for the wave gauge series. In general, we can determine an appropriate degree of tapering by carefully comparing the periodogram and direct spectral estimates corresponding to dpss tapers with different values of nW. If the periodogram proves to suffer from leakage (as it often does in the physical sciences), we then seek a leakage free direct spectral estimate formed using as small a value of W – and hence nW – as possible. A small W is desirable from 2 viewpoints: first, resolution typically decreases as W increases, and, second, the distance in frequency between approximately uncorrelated spectral estimates increases as W increases (this causes a loss in degrees of freedom when we subsequently smooth across frequencies – see Section 11.2.3 for details).

In checking for the presence of leakage, comparison between different

spectral estimates is best done graphically in the following ways:

1. by comparing spectral estimates side by side (as in Figure 11.5) or on top of each other using different colors;

2. by plotting the ratio of two spectral estimates on a decibel scale versus frequency and searching for frequency bands over which an average of this ratio is nonzero (presumably a statistical test could be devised here to assess the significance of departures from zero); and

3. by constructing scatter plots of, say, $\hat{S}_X^{(p)}(f)$ versus $\hat{S}_X^{(d)}(f)$ for f's in a selected band of frequencies and looking for clusters of points consistently above a line with unit slope and zero intercept.

The use of interactive graphical displays would clearly be helpful in determining the proper degree of tapering (Percival and Kerr [9]).

Prewhitening

The loss of resolution and of degrees of freedom inherent in tapering can be alleviated considerably if we can *prewhiten* our time series. To explain prewhitening, we need the following result from the theory of linear time-invariant filters. Given any set of $p+1$ real-valued numbers $a_0, a_1, \ldots, a_p$, we can filter $X_1, \ldots, X_n$ to obtain

$$W_t = \sum_{k=0}^{p} a_k X_{t-k}, \qquad t = p+1, \ldots, n. \tag{11.11}$$

The filtered process $\{W_t\}$ is a zero mean real-valued stationary process with an sdf $S_W(\cdot)$ related to $S_X(\cdot)$ via

$$S_W(f) = |A(f)|^2 \, S_X(f), \;\; \text{where} \;\; A(f) = \sum_{k=0}^{p} a_k e^{-i2\pi f k \, \Delta t}. \tag{11.12}$$

$A(\cdot)$ is the transfer function for the filter $\{a_k\}$. The idea behind prewhitening is to find a set of a_k's such that the sdf $S_W(\cdot)$ has substantially less structure than $S_X(\cdot)$ and ideally is as close to a white noise sdf as possible. If such a filter can be found, we can easily produce a leakage-free direct spectral estimate $\hat{S}_W^{(d)}(\cdot)$ of $S_W(\cdot)$ requiring little or no tapering. Equation 11.12 tells us that we can then estimate $S_X(\cdot)$ using $\hat{S}_X^{(pc)}(f) = \hat{S}_W^{(d)}(f) / |A(f)|^2$, where the superscript '(pc)' stands for *postcolored*.

The reader might note an apparent logical contradiction here: in order to pick a reasonable set of a_k's, we must know the shape of $S_X(\cdot)$, the very

function we are trying to estimate! There are two ways around this inherent difficulty. In many cases, there is enough prior knowledge concerning $S_X(\cdot)$ from, say, previous experiments so that, even though we don't know the exact shape of the sdf, we can still design an effective prewhitening filter. In other cases we can create a prewhitening filter based upon the data itself. A convenient way of doing this is by fitting *autoregressive models* of order p (hereafter AR(p)). In the present context, we postulate that, for some reasonably small p, we can model X_t as a linear combination of the p prior values $X_{t-1}, \ldots, X_{t-p}$ plus an error term; i.e,

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + W_t, \qquad t = p+1, \ldots, n, \tag{11.13}$$

where $\{W_t\}$ is an 'error' process with an sdf $S_W(\cdot)$ that has less structure than $S_X(\cdot)$. The stipulation that p be small is desirable because the prewhitened series $\{W_t\}$ will be shortened to length $n-p$. The above equation is a special case of Equation 11.11, as can be seen by letting $a_0 = 1$ and $a_k = -\phi_k$ for $k > 0$.

For a given p, we can obtain estimates of the ϕ_k's from our time series using a number of different methods (see [4], Chapter 9). One method that generally works well is Burg's algorithm, a computationally efficient way of producing estimates of the ϕ_k's that are guaranteed to correspond to a stationary process. The Fortran subroutine `memcof` in [8], Section 13.6, implements Burg's algorithm (the reader should note that the time series used as input to `memcof` is assumed to be already adjusted for any nonzero mean value).

Burg's algorithm (or any method that estimates the ϕ_k's in Equation 11.13) is sometimes used by itself to produce what is known as an *autoregressive spectral estimate* (this is one form of parametric spectral estimation). If we let $\bar{\phi}_k$ represent Burg's estimate of ϕ_k, then the corresponding autoregressive spectral estimate is given by

$$\bar{S}_X^{(ar)}(f) = \frac{\bar{\sigma}_p^2 \, \Delta t}{\left| 1 - \sum_{k=1}^{p} \bar{\phi}_k e^{-i 2\pi f k \, \Delta t} \right|^2}, \tag{11.14}$$

where $\bar{\sigma}_p^2$ is an estimate of the variance of $\{W_t\}$. The key difference between an autoregressive spectral estimate and prewhitening is in the use of Equation 11.13: the former assumes the process $\{W_t\}$ is *exactly* white noise, whereas the latter merely postulates $\{W_t\}$ to have an sdf with less structure than $\{X_t\}$. Autoregressive spectral estimation works well for many time series, but it depends heavily on a proper choice of the order p; moreover, simple approximations to the statistical properties of $\bar{S}_X^{(ar)}(\cdot)$

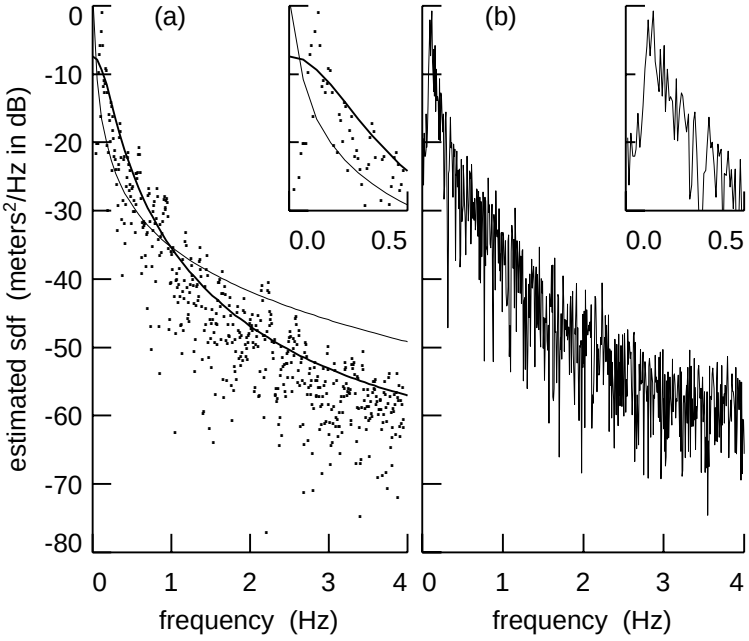

Figure 11.6: Determination of a prewhitening filter.

are currently lacking. (Autoregressive spectral estimation is sometimes misleadingly called *maximum entropy* spectral estimation – see [4] for a discussion about this misnomer.)

Figure 11.6 illustrates the construction of a prewhitening filter for the wire gauge series. The dots in plot (a) depict the same direct spectral estimate as shown in Figure 11.5(b). This estimate is essentially leakage free and hence serves as a 'pilot estimate' for comparing different prewhitening filters. Using Burg's algorithm, we computed autoregressive spectral estimates for various orders p via Equation 11.14. The thin and thick solid curves in plot (a) show, respectively, the AR(2) and AR(4) estimates. The AR(4) estimate is the estimate with lowest order p that reasonably captures the structure of the pilot estimate over the frequency range of main interest (namely, 0.4 to 4.0 Hz); moreover, use of these $\bar{\phi}_k$'s to form a prewhitening filter yields a prewhitened series W_t, $t = 5, \ldots, 4096$, with a sdf that has no regions of large power outside of 0.4 to 4.0 Hz. The periodogram of the W_t's appears to be leakage free for f between 0.4 and 4.0 Hz. The spectral estimate in plot (b) is $\hat{S}_X^{(pc)}(f)$, the result of postcoloring the periodogram $\hat{S}_W^{(p)}(\cdot)$ (i.e., dividing $\hat{S}_W^{(p)}(f)$ by $|1 - \sum_{k=1}^{4} \bar{\phi}_k e^{-i2\pi f k \, \Delta t}|^2$). Note

that $\hat{S}_X^{(pc)}(\cdot)$ generally agrees well with the leakage-free pilot estimate in plot (a).

We close this section with three comments. First, the prewhitening procedure we have outlined above might require a fairly large order p in order to work properly and can fail if $S_X(\cdot)$ contains narrow-band features (i.e., line components). Low order AR processes cannot capture narrow-band features adequately – their use can yield a prewhitened process with a more complicated sdf than that of the original process. Additionally, Burg's algorithm is known to be problematic for certain narrow-band processes for which it will sometimes incorrectly represent a line component as a double peak (this is called 'spontaneous line splitting'). The proper way to handle narrow-band processes is to separate out the narrow-band components and then deal with the resulting 'background' process. One attractive method for handling such processes uses multitapering (for details, see [6], Section XIII, or [4], Sections 10.11 and 10.13).

Second, in our example of prewhitening we actually *introduced* leakage into the very low frequency portion of the sdf. To see this fact, note that the spectral level of the prewhitened estimate of Figure 11.6(b) near $f = 0$ is about 10 dB *higher* than those of the two direct spectral estimates in Figure 11.5. This happened because we were mainly interested in creating a prewhitening filter to help estimate the sdf from 0.4 to 4.0 Hz. An AR(4) filter accomplishes this task, but it fails to represent the very low frequency portion of the sdf at all (see Figure 11.6(a)). As a result, the prewhitened series $\{W_t\}$ has an sdf that is evidently deficient in power near zero frequency and hence suffers from leakage at very low frequencies. Use of a much higher order AR filter (say, $p = 256$) can correct this problem.

Finally, we note that, if prewhitening is not used, we define W_t to be equal to X_t and set p equal to zero in what follows.

11.2.3 Variance Reduction

A cursory examination of the spectral estimates for the wire gauge series in Figures 11.5 and 11.6 reveals substantial variability across frequencies – so much so that it is difficult to discern the overall structure in the spectral estimates without a fair amount of study. All direct spectral estimators suffer from this inherent choppiness, which can be explained by considering the distributional properties of $\hat{S}_W^{(d)}(f)$. First, if f is not too close to 0 or $f_{(N)}$ and if $S_W(\cdot)$ satisfies a mild regularity condition, then $2\hat{S}_W^{(d)}(f)/S_W(f) \stackrel{\mathrm{d}}{=} \chi_2^2$; i.e., the rv $2\hat{S}_W^{(d)}(f)/S_W(f)$ is approximately equal in distribution to a chi-square rv with 2 degrees of freedom. If tapering is not used, f is considered 'not too close' to 0 or $f_{(N)}$ if $1/(n-p)\,\Delta t < f <$

$f_{(N)} - 1/(n-p)\,\Delta t$; if tapering is used, we must replace $1/(n-p)\,\Delta t$ by a larger term reflecting the increased width of the central lobe of the spectral window (for example, the term for the Hanning data taper is approximately $2/(n-p)\,\Delta t$ so f is 'not too close' if $2/(n-p)\,\Delta t < f < f_{(N)} - 2/(n-p)\,\Delta t$).

Since a chi-square rv χ^2_ν with ν degrees of freedom has a variance of 2ν, we have the approximation $V[\hat{S}_W^{(d)}(f)] = S_W^2(f)$. This result is independent of the number of W_t's we have: unlike statistics such as the sample mean of independent and identically distributed Gaussian rv's, the variance of $\hat{S}_W^{(d)}(f)$ does not decrease to 0 as the sample size $n-p$ gets larger (except in the uninteresting case $S_W(f) = 0$). This result explains the choppiness of the direct spectral estimates shown in Figures 11.5 and 11.6. In statistical terminology, $\hat{S}_W^{(d)}(f)$ is an inconsistent estimator of $S_W(f)$.

We now outline three approaches for obtaining a consistent estimator of $S_W(f)$. Each approach is based upon combining rv's that – under suitable assumptions – can be considered as approximately pairwise uncorrelated estimators of $S_W(f)$. Briefly, the three approaches are to:

1. smooth $\hat{S}_W^{(d)}(f)$ across frequencies, yielding what is known as a *lag window spectral estimator*;

2. break $\{X_t\}$ (or $\{W_t\}$) into a number of segments (some of which can overlap), compute a direct spectral estimate for each segment, and then average these estimates together, yielding Welch's *overlapped segment averaging* (WOSA) *spectral estimator*; and

3. compute a series of direct spectral estimates for $\{W_t\}$ using a set of orthogonal data tapers and then average these estimates together, yielding Thomson's *multitaper spectral estimator.*

Lag Window Spectral Estimators

A lag window spectral estimator of $S_W(\cdot)$ takes the form

$$\hat{S}_W^{(lw)}(f) = \int_{-f_{(N)}}^{f_{(N)}} W_m(f - f')\hat{S}_W^{(d)}(f')\,df', \tag{11.15}$$

where $W_m(\cdot)$ is a *smoothing window* whose smoothing properties are controlled by the smoothing parameter m. In words, the estimator $\hat{S}_W^{(lw)}(\cdot)$ is obtained by convolving a smoothing window with the direct spectral estimator $\hat{S}_W^{(d)}(\cdot)$. A typical smoothing window has much the same appearance as a spectral window. There is a central lobe with a width that can be adjusted by the smoothing parameter m: the wider this central lobe is, the

smoother $\hat{S}_W^{(lw)}(\cdot)$ will be. There can also be a set of annoying sidelobes that cause *smoothing window leakage*. The presence of smoothing window leakage is easily detected by overlaying plots of $\hat{S}_W^{(lw)}(\cdot)$ and $\hat{S}_W^{(d)}(\cdot)$ and looking for ranges of frequencies where the former does not appear to be a smoothed version of the latter.

If we have made use of an AR prewhitening filter, we can then postcolor $\hat{S}_W^{(lw)}(\cdot)$ to obtain an estimator of $S_X(\cdot)$, namely,

$$\hat{S}_X^{(pc)}(f) = \frac{\hat{S}_W^{(lw)}(f)}{\left|1 - \sum_{k=1}^{p} \bar{\phi}_k e^{-i2\pi f k\,\Delta t}\right|^2}.$$

The statistical properties of $\hat{S}_W^{(lw)}(\cdot)$ are tractable because of the following large sample result. If $\hat{S}_W^{(d)}(\cdot)$ is in fact the periodogram (i.e., we have not tapered the W_t's), the set of rv's $\hat{S}_W^{(d)}(j/(n-p)\,\Delta t)$, $j = 1, 2, \ldots, J$, are approximately pairwise uncorrelated, with each rv being proportional to a χ_2^2 rv (here J is the largest integer such that $J/(n-p) < 1/2$). If we have used tapering to form $\hat{S}_W^{(d)}(\cdot)$, a similar statement is true over a smaller set of rv's defined on a coarser grid of equally spaced frequencies – as the degree of tapering increases, the number of approximately uncorrelated rv's decreases. Under the assumptions that the sdf $S_W(\cdot)$ is slowly varying across frequencies (prewhitening helps to make this true) and that the central lobe of the smoothing window is sufficiently small compared to the variations in $S_W(\cdot)$, it follows that $\hat{S}_W^{(d)}(f)$ in Equation 11.15 can be approximated by a linear combination of uncorrelated χ_2^2 rv's. A standard 'equivalent degrees of freedom' argument can be then used to approximate the distribution of $\hat{S}_W^{(d)}(f)$ (see Equation 11.17 below).

There are two practical ways of computing $\hat{S}_W^{(lw)}(\cdot)$. The first way is to discretize Equation 11.15, yielding an estimator proportional to a convolution of the form $\sum_k W_m(f - f'_k)\hat{S}_W^{(d)}(f'_k)$, where the f'_k's are some set of equally spaced frequencies. The second way is to recall that 'convolution in one Fourier domain is equivalent to multiplication in the other' to rewrite Equation 11.15 as

$$\hat{S}_W^{(lw)}(f) = \sum_{\tau=-(n-p-1)}^{n-p-1} w_{\tau,m}\hat{C}_{\tau,W}^{(d)} e^{-i2\pi f\tau\,\Delta t}, \tag{11.16}$$

where $\hat{C}_{\tau,W}^{(d)}$ is the acvs estimator given in Equation 11.9 corresponding to $\hat{S}_W^{(d)}(\cdot)$, and $\{w_{\tau,m}\}$ is a *lag window* (this can be regarded as the inverse Fourier transform of the smoothing window $W_m(\cdot)$). In fact, because

$\hat{S}_W^{(d)}(\cdot)$ is a trigonometric polynomial, all discrete convolutions of the form $\sum_k W_m(f - f'_k)\hat{S}_W^{(d)}(f'_k)$ can also be computed via Equation 11.16 with an appropriate choice of $w_{\tau,m}$'s (for details, see [4], Section 6.7). Our two practical ways of computing $\hat{S}_W^{(lw)}(\cdot)$ thus yield equivalent estimators. Unless the discrete convolution is sufficiently short, Equation 11.16 is computationally faster to use.

Statistical theory suggests that, under reasonable assumptions,

$$\frac{\nu \hat{S}_W^{(lw)}(f)}{S_W(f)} \stackrel{\mathrm{d}}{=} \chi^2_\nu \tag{11.17}$$

to a good approximation, where ν is called the *equivalent degrees of freedom* for $\hat{S}_W^{(lw)}(f)$ and is given by $\nu = 2(n-p)B_W \, \Delta t / C_h$. Here B_W is a measure of the bandwidth of the smoothing window $W_m(\cdot)$ and can be computed via $B_W = 1/\Delta t \sum_{\tau=-(n-p-1)}^{n-p-1} w_{\tau,m}^2$; on the other hand, C_h depends only on the taper applied to the W_t's and can be computed via $C_h = (n-p)\sum_{t=p+1}^{n} h_t^4$. Note that, if we do not explicitly taper, then $h_t = 1/\sqrt{n-p}$ and hence $C_h = 1$; for a typical data taper, the Cauchy inequality tells us that $C_h > 1$ (for example, $C_h \approx 1.94$ for the Hanning data taper). The equivalent degrees of freedom for $\hat{S}_W^{(lw)}(f)$ thus increase as we increase the smoothing window bandwidth and decrease as we increase the degree of tapering. Equation 11.17 tells us that $E[\hat{S}_W^{(lw)}(f)] \approx S_W(f)$ and that $V[\hat{S}_W^{(lw)}(f)] \approx S_W^2(f)/\nu$, so increasing ν decreases $V[\hat{S}_W^{(lw)}(f)]$.

The approximation in Equation 11.17 can be used to construct a confidence interval for $S_W(f)$ in the following manner. Let $\eta_\nu(\alpha)$ denote the $\alpha \times 100\%$ percentage point of the χ^2_ν distribution; i.e., $\mathbf{P}[\chi^2_\nu \le \eta_\nu(\alpha)] = \alpha$. A $100(1-2\alpha)\%$ confidence interval for $S_W(f)$ is approximately given by

$$\left[\frac{\nu \hat{S}_W^{(lw)}(f)}{\eta_\nu(1-\alpha)}, \frac{\nu \hat{S}_W^{(lw)}(f)}{\eta_\nu(\alpha)} \right]. \tag{11.18}$$

The percentage points $\eta_\nu(\alpha)$ are tabulated in numerous textbooks or can be computed using an algorithm given by Best and Roberts [10].

The confidence interval of 11.18 is inconvenient in that its length is proportional to $\hat{S}_W^{(lw)}(f)$. On the other hand, the corresponding confidence interval for $10 \cdot \log_{10}(S_W(f))$ (i.e., $S_W(f)$ on a decibel scale) is just

$$\Big[10 \cdot \log_{10}\left(\nu/\eta_\nu(1-\alpha)\right) + 10 \cdot \log_{10}\left(\hat{S}_W^{(lw)}(f)\right), \\ 10 \cdot \log_{10}\left(\nu/\eta_\nu(\alpha)\right) + 10 \cdot \log_{10}\left(\hat{S}_W^{(lw)}(f)\right) \Big],$$

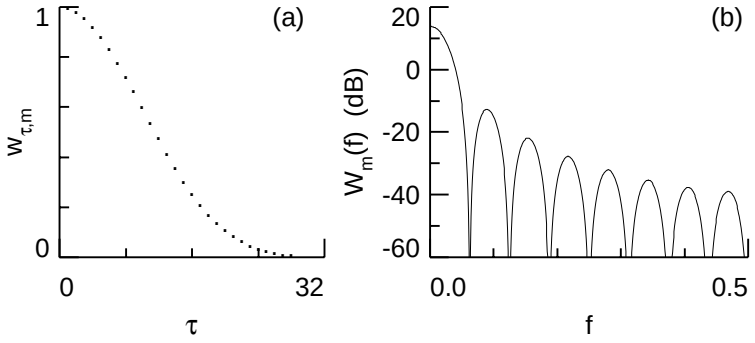

Figure 11.7: Parzen lag window (a) and the corresponding smoothing window (b) for $m = 32$. The smoothing window bandwidth is $B_W = 0.058$.

which has a width that is independent of $\hat{S}_W^{(lw)}(f)$. This is the rationale for plotting sdf estimates on a decibel (or logarithmic) scale.

A bewildering number of different lag windows has been discussed in the literature (see [3]). Here we give only one example, the well-known *Parzen lag window* (Parzen [11]):

$$w_{\tau,m} = \begin{cases} 1 - 6\tilde{\tau}^2 + 6|\tilde{\tau}|^3, & |\tau| \le m/2; \\ 2(1 - \tilde{\tau})^3, & m/2 < |\tau| \le m; \\ 0, & |\tau| > m, \end{cases}$$

where m is taken to be a positive integer and $\tilde{\tau} = \tau/m$. This lag window is easy to compute and has sidelobes whose envelope decays as f^{-4} so that smoothing window leakage is rarely a problem. To a good approximation, the smoothing window bandwidth for the Parzen lag window is given by $B_W = 1.85/(m\,\Delta t)$. As m increases, the smoothing window bandwidth decreases, and the resulting lag window estimator becomes less smooth in appearance. The associated equivalent degrees of freedom are given approximately by $\nu = 3.71(n - p)/(mC_h)$. The Parzen lag window for $m = 32$ and its associated smoothing window are shown in Figure 11.7.

As an example, Figure 11.8(a) shows a postcolored lag window estimator for the wire wave gauge data (the solid curve), along with the corresponding postcolored direct spectral estimator (the dots – these depict the same estimate as shown in Figure 11.6(b)). The Parzen lag window was used here with a value of $m = 237$ for the smoothing window parameter (the corresponding equivalent degrees of freedom ν is 64). This value was chosen after some experimentation and seems to produce a lag window estimator that captures all of the important spectral features indicated by the direct spectral estimator for frequencies between 0.4 and 4.0 Hz (note, however,

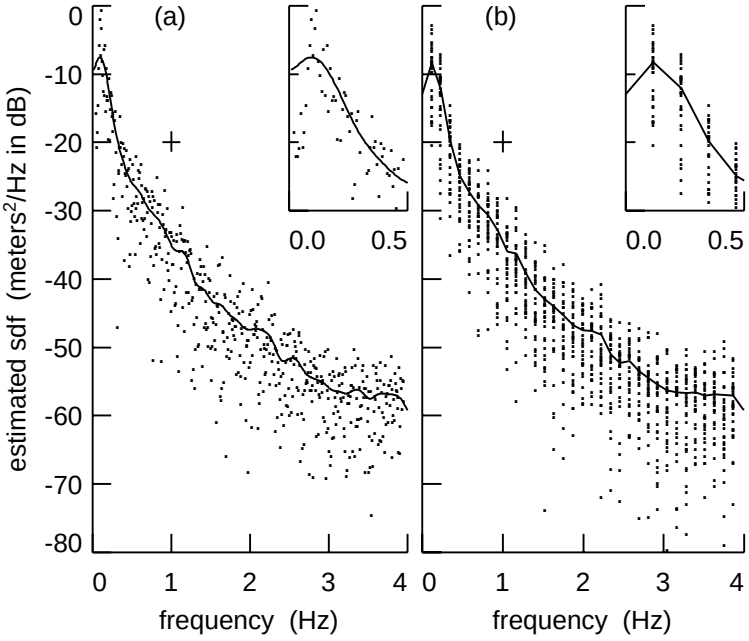

Figure 11.8: Postcolored Parzen lag window spectral estimate – solid curve on plot (a) – and WOSA spectral estimate – solid curve on (b) – for wire wave gauge time series. The smoothing window parameter for the Parzen lag window was $m = 237$, yielding $\nu = 64$ equivalent degrees of freedom. The WOSA spectral estimate was formed using a Hanning data taper on blocks with 256 data points, with adjacent blocks overlapping by 50%. The equivalent degrees of freedom for this estimate is $\nu = 59$.

that this estimator smears out the peak between 0.0 and 0.4 Hz rather badly). We have also plotted a crisscross whose vertical height represents the length of a 95% confidence interval for $10 \cdot \log_{10}(S_X(f))$ (based upon the postcolored lag window estimator) and whose horizontal width represents the smoothing window bandwidth B_W.

WOSA Spectral Estimators

Let us now consider the second common approach to variance reduction, namely, Welch's overlapped segment averaging (Welch [12]; Carter [13] and references therein). The basic idea is to break a time series into a number of blocks (i.e., segments), compute a direct spectral estimate for each block and then produce the WOSA spectral estimate by averaging these spectral

estimates together. In general, the blocks are allowed to overlap, with the degree of overlap being determined by the degree of tapering – the heavier the degree of tapering, the more the blocks should be overlapped (Thomson [14]). Thus, except at the very beginning and end of the time series, data values that are heavily tapered in one block are lightly tapered in another block, so intuitively we are recapturing 'information' lost due to tapering in one block from blocks overlapping it. Because it can be implemented in a computationally efficient fashion (using the fast Fourier transform algorithm) and because it can handle very long time series (or time series with a time varying spectrum), the WOSA estimation scheme is the basis for many of the commercial spectrum analyzers on the market.

To define the WOSA spectral estimator, let n_S represent a block size, and let $h_1, \ldots, h_{n_S}$ be a data taper. We define the direct spectral estimator of $S_X(f)$ for the block of n_S contiguous data values starting at index l as

$$\hat{S}^{(d)}_{l,X}(f) = \Delta t \left| \sum_{t=1}^{n_S} h_t X_{t+l-1} e^{-i2\pi f t\, \Delta t} \right|^2, \qquad 1 \le l \le n + 1 - n_S$$

(there is no reason why we cannot use a prewhitened series $\{W_t\}$ here rather than the X_t's, but prewhitening is rarely used in conjunction with WOSA, perhaps because block overlapping is regarded as an efficient way of compensating for the degrees of freedom lost due to tapering). The WOSA spectral estimator of $S_X(f)$ is defined to be

$$\hat{S}^{(wosa)}_X(f) = \frac{1}{n_B} \sum_{j=0}^{n_B-1} \hat{S}^{(d)}_{js+1,X}(f), \tag{11.19}$$

where n_B is the total number of blocks and s is an integer shift factor satisfying $0 < s \le n_S$ and $s(n_B - 1) = n - n_S$ (note that the block for $j = 0$ uses data values $X_1, \ldots, X_{n_S}$, while the block for $j = n_B - 1$ uses $X_{n-n_S+1}, \ldots, X_n$).

The large sample statistical properties of $\hat{S}^{(wosa)}_X(f)$ closely resemble those of lag window estimators. In particular, we have the approximation that $\nu \hat{S}^{(wosa)}_X(f)/S_X(f) \stackrel{\mathrm{d}}{=} \chi^2_\nu$, where the equivalent degrees of freedom ν are given by

$$\nu = \frac{2n_B}{1 + 2\sum_{m=1}^{n_B-1} \left(1 - \frac{m}{n_B}\right) \left|\sum_{t=1}^{n_S} h_t h_{t+ms}\right|^2}$$

(here $h_t = 0$ by definition for all $t > n_S$). If we specialize to the case of 50% block overlap (i.e., $s = n_S/2$) with a Hanning data taper (a common

recommendation in the engineering literature), the above can be approximated by the simple formula $\nu \approx 36n_B^2/(19n_B - 1)$. Thus, as the number of blocks n_B increases, the equivalent degrees of freedom increase also, yielding a spectral estimator with reduced variance. Unless $S_X(\cdot)$ has a relatively featureless sdf, we cannot, however, make n_B arbitrarily small without incurring severe bias in the individual direct spectral estimators mainly due to loss of resolution. (For details on the above results, see [4], Section 6.17.)

Figure 11.8(b) shows a WOSA spectral estimator for the wire wave gauge data (the solid curve). This series has $n = 4096$ data values. Some experimentation indicated that a block size of $n_S = 256$ and the Hanning data taper are reasonable choices for estimating the sdf between 0.4 and 4.0 Hz using WOSA. With a 50% block overlap, the shift factor is $s = n_S/2 = 128$; the total number of blocks is $n_B = \frac{1}{s}(n - n_S) + 1 = 31$; and ν, the equivalent degrees of freedom, is approximately 59. The 31 individual direct spectral estimates that were averaged together to form the WOSA estimate are shown as the dots in Figure 11.8(b).

We have also plotted a 'bandwidth/confidence interval' crisscross similar to that on Figure 11.8(a), but now the 'bandwidth' (i.e., the horizontal width) is the distance in frequency between approximately uncorrelated spectral estimates. This measure of bandwidth is a function of the block size n_S and the data taper used in WOSA. For the Hanning taper, the bandwidth is approximately $1.94/(n_S\,\Delta t)$. The crisscrosses in Figures 11.8(a) and (b) are quite similar, indicating that the statistical properties of the postcolored Parzen lag window and WOSA spectral estimates are comparable: indeed, the actual estimates agree closely, with the WOSA estimate being slightly smoother in appearance.

Multitaper Spectral Estimators

An interesting alternative to either lag window or WOSA spectral estimation is the multitaper approach due to Thomson [6]. Multitaper spectral estimation can be regarded as a way of producing a direct spectral estimator with more than just 2 equivalent degrees of freedom (typical values are 4 to 16). As such, the multitaper method is different in spirit from the other two estimators in that it does not seek to produce highly smoothed spectra. An increase in degrees of freedom from 2 to just 10 is enough, however, to shrink the width of a 95% confidence interval for the sdf by more than an order of magnitude and hence to reduce the variability in the spectral estimate to the point where the human eye can readily discern the overall structure. Detailed discussions on the multitaper approach are given in [6] and Chapter 7 of [4]. Here we merely sketch the main ideas.

Multitaper spectral estimation is based upon the use of a set of K data tapers $\{ h_{t,k} : t = 1, \ldots, n \}$, where k ranges from 0 to $K-1$. We assume that these tapers are orthonormal (i.e., $\sum_{t=1}^{n} h_{t,j} h_{t,k} = 1$ if $j = k$ and 0 if $j \neq k$). The simpliest multitaper estimator is defined by

$$\hat{S}_X^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_{k,X}^{(mt)}(f) \;\text{ with }\; \hat{S}_{k,X}^{(mt)}(f) = \Delta t \left| \sum_{t=1}^{n} h_{t,k} X_t e^{-i 2\pi f t \, \Delta t} \right|^2$$

(Thomson [6] advocates adaptively weighting the $\hat{S}_{k,X}^{(mt)}(f)$'s rather than simply averaging them together). A comparison of the above definition for $\hat{S}_{k,X}^{(mt)}(\cdot)$ with Equation 11.8 shows that $\hat{S}_{k,X}^{(mt)}(\cdot)$ is in fact just a direct spectral estimator, so the multitaper estimator is just an average of direct spectral estimators employing an orthonormal set of tapers. Under certain mild conditions, the orthonormality of the tapers translates into the frequency domain as approximate independence of the individual $\hat{S}_{k,X}^{(mt)}(f)$'s; i.e., $\hat{S}_{j,X}^{(mt)}(f)$ and $\hat{S}_{k,X}^{(mt)}(f)$ are approximately independent for $j \neq k$. Approximate independence in turn implies that $2K \hat{S}_X^{(mt)}(f) / S_X(f) \stackrel{\mathrm{d}}{=} \chi^2_{2K}$ approximately, so that the equivalent degrees of freedom for $\hat{S}_X^{(mt)}(f)$ is equal to twice the number of data taper employed.

The key trick then is to find a set of K orthonormal sequences, each one of which does a proper job of tapering. One appealing approach is to return to the concentration problem that gave us the dpss taper for a fixed resolution bandwidth $2W$. If we now refer to this taper as the 'zeroth-order' dpss taper and denote it by $\{h_{t,0}\}$, we can recursively construct the remaining $K-1$ 'higher-order' dpss tapers $\{h_{t,k}\}$ as follows. For $k = 1$, $\ldots$, $K-1$, we define the kth-order dpss taper as the set of n numbers $\{ h_{t,k} : t = 1, \ldots, n \}$ such that

1. $\{h_{t,k}\}$ is orthogonal to each of the k sequences $\{h_{t,0}\}$, $\ldots$, $\{h_{t,k-1}\}$ (i.e., $\sum_{t=1}^{n} h_{t,j} h_{t,k} = 0$ for $j = 0, \ldots, k-1$);
2. $\{h_{t,k}\}$ is normalized such that $\sum_{t=1}^{n} h_{t,k}^2 = 1$; and,
3. subject to conditions [1] and [2], the spectral window $\mathcal{H}_k(\cdot)$ corresponding to $\{h_{t,k}\}$ maximizes the concentration ratio

$$\int_{-W}^{W} \mathcal{H}_k(f) \, df \Big/ \int_{-f_{(N)}}^{f_{(N)}} \mathcal{H}_k(f) \, df = \lambda_k(n, W).$$

In words, subject to the constraint of being orthogonal to all lower-order dpss tapers, the kth-order dpss taper is 'optimal' in the restricted sense that

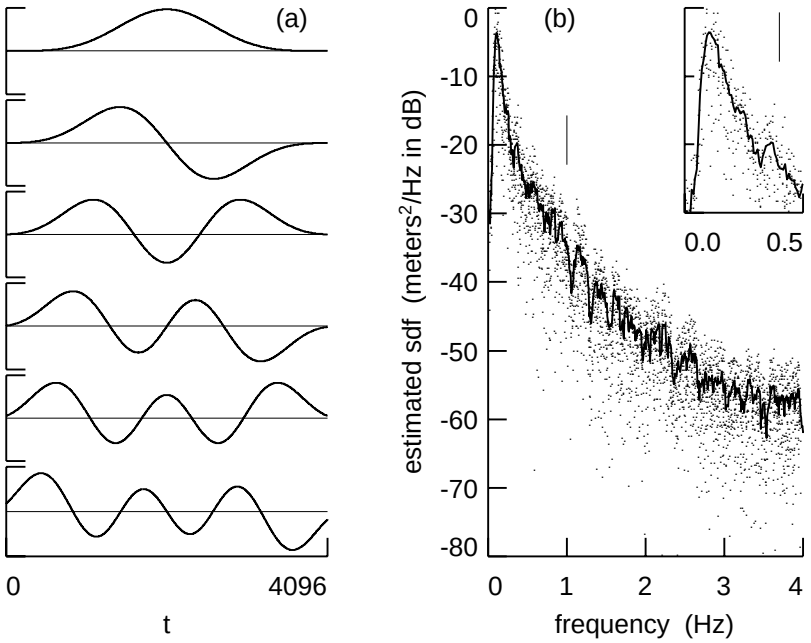

Figure 11.9: Multitaper spectral estimation.

the sidelobes of its spectral window are suppressed as much as possible as measured by the concentration ratio. Methods for calculating the dpss data tapers are discussed in [4], Chapter 8.

In a series of papers, Slepian [15] (and references therein) has extensively studied the nature of the dpss's. One important fact he discusses is that the concentration ratio $\lambda_k(n, W)$ strictly decreases as k increases in a manner such that $\lambda_k(n, W)$ is close to unity for $k < 2nW\,\Delta t$, after which it rapidly approaches zero with increasing k (the value $2nW\,\Delta t$ is sometimes called the *Shannon number*). Since $\lambda_k(n, W)$ must be close to unity for $\{h_{t,k}\}$ to be a decent data taper, multitaper spectral estimation is restricted to the use of at most – and, in practice, usually less than – $2nW\,\Delta t$ orthonormal dpss tapers.

An example of multitaper spectral estimation is shown in Figure 11.9. The left-hand column of plots shows the kth-order dpss data tapers for $n = 4096$, $nW = 4/\Delta t$ and k ranging from 0 (top plot) to $K - 1 = 5$ (bottom plot). The thin horizontal lines in each of these plots indicate the zero level, so, whereas the zeroth-order dpss is strictly positive everywhere (but quite close to zero near $t = 1$ and $t = n$), the higher-order tapers assume both positive and negative values. Note also that the zeroth-order taper

heavily downweights values of the time series close to $t = 1$ and $t = n$, but that these values are given successively more weight by the higher-order tapers (one interpretation of multitapering is that the higher-order tapers are recapturing information 'lost' when but a single data taper is used). The solid curve in Figure 11.9(b) shows a multitaper spectral estimate $\hat{S}_X^{(mt)}(\cdot)$ for the wire wave gauge data based upon these 6 dpss tapers, whereas the dots show the 6 individual direct spectral estimates $\hat{S}_{k,X}^{(mt)}(\cdot)$. Note that the number of tapers that we have used is below the Shannon number $2nW\,\Delta t = 8$ and that ν, the equivalent degrees of freedom, is here $2K = 12$. The multitaper spectral estimate is much choppier in appearance than either the lag window spectral estimate of Figure 11.8(a) or the WOSA estimate of 11.8(b), both of which have a markedly higher number of equivalent degrees of freedom ($\nu = 64$ and $\nu = 59$, respectively). Nonetheless, the variability in the multitaper spectral estimate is small enough so that the eye can readily detect the overall structure (cf. $\hat{S}_X^{(mt)}(\cdot)$ with the two spectral estimates in Figure 11.5), and, because it is not highly smoothed, the multitaper estimate does markedly better at capturing the spectral structure near $f = 0$.

Based upon performance bounds, Bronez [16] argues that the multitaper spectral estimator has statistical properties that are superior to WOSA for sdf's with very high dynamic ranges (more research is required, however, to verify that these bounds translate into an actual advantage in practice). In comparison to prewhitening, multitapering is useful in situations where leakage is a concern but it is not practical to carefully design prewhitening filters (this occurs in, for example, exploration geophysics due to the enormous volume of time series routinely collected). Finally, we note that Thomson and Chave [17] describe an appealing scheme in which multitapering is used in conjunction with WOSA.

11.2.4 Evaluating the Significance of Spectral Peaks

A common use for spectral analysis is the detection of a periodic signal in the presence of noise. In the simpliest case, we assume that the periodic signal is a sinusoid and that the noise is additive zero mean white noise $\{W_t\}$; i.e., we assume that our time series can be modeled as

$$X_t = D\cos(2\pi f_l t\,\Delta t + \phi) + W_t, \quad t = 1, \ldots, n,$$

where D, f_l and ϕ are constants. Suppose first that the frequency f_l of the sinusoid is known *a priori* to be equal to the Fourier frequency $l/(n\,\Delta t)$, where l is an integer such that $0 < l/n < 1/2$. We wish to test the null hypothesis that D is zero (i.e., that our time series in fact does *not* contain

a significant sinusoidal component at the proposed frequency f_l). Let J be the largest integer such that $J/n < 1/2$, and let $f_j = j/(n\,\Delta t)$ for $j = 1, \ldots, J$ be the set of all nonzero Fourier frequencies less than $f_{(N)}$. If the null hypothesis is true and if $\{W_t\}$ is a Gaussian process, then the set of periodogram ordinates $\hat{S}^{(p)}(f_1)$, $\hat{S}^{(p)}(f_2)$, $\ldots$, $\hat{S}^{(p)}(f_J)$ constitute a set of independent and identically distributed χ^2_2 rv's multiplied by the constant $V(W_t)/2$. The ratio

$$(J-1)\hat{S}^{(p)}(f_l) \Big/ \sum_{\substack{j=1,\ldots,J \\ j\neq l}} \hat{S}^{(p)}(f_j)$$

thus follows an $F_{2,2(J-1)}$ distribution under the null hypothesis. Under the alternative hypothesis $D \neq 0$, the periodogram ordinate $\hat{S}^{(p)}(f_l)$ will tend to be large, and hence the above ratio will also tend to be large. We thus reject the null hypothesis at level of significance α if the ratio exceeds the upper $(1-\alpha)\times 100\%$ percentage point of the $F_{2,2(J-1)}$ distribution. In general, this percentage point for an $F_{2,k}$ distribution is given by $k(1-\alpha^{2/k})/(2\alpha^{2/k})$.

The above test statistic can be easily modified to handle a periodic *nonsinusoidal* signal with known period $1/f_l$. In general, such periodic signals can be written as a linear combination of sinusoids with frequencies f_l, $2f_l$, $3f_l$, $\ldots$, where f_l is known as the *fundamental frequency* of the signal and $(k+1)f_l$ is called the kth *harmonic frequency*. For example, suppose that the signal can be assumed *a priori* to be represented by a fundamental frequency f_l and its first harmonic $2f_l$ with $2f_l < 1/(2\,\Delta t)$; i.e., because $2f_l = f_{2l}$, our assumed model is

$$X_t = D_1 \cos(2\pi f_l t\,\Delta t + \phi_1) + D_2 \cos(2\pi f_{2l} t\,\Delta t + \phi_2) + W_t.$$

We can then test the null hypothesis $D_1 = D_2 = 0$ using the ratio

$$(J-2)\left[\hat{S}^{(p)}(f_l) + \hat{S}^{(p)}(f_{2l})\right] \Big/ 2 \sum_{\substack{j=1,\ldots,J \\ j\neq l,\ j\neq 2l}} \hat{S}^{(p)}(f_j),.$$

Under the null hypothesis, this ratio follows an $F_{4,2(J-2)}$ distribution. We thus reject the null hypothesis with level of significance α if the above ratio exceeds the upper $(1-\alpha)\times 100\%$ percentage point of this distribution.

Suppose now that we do not know *a priori* the period of the potential signal in our time series, but that the periodogram for $X_1, X_2, \ldots, X_n$ has apparent large values at one or more Fourier frequencies. To assess whether these large values are ascribable to just random fluctuations, we can use

Fisher's g statistic, defined by

$$g = \max_{1 \le j \le J} \hat{S}^{(p)}(f_j) \Big/ \sum_{j=1}^{J} \hat{S}^{(p)}(f_j).$$

Under the null hypothesis that our time series is a portion of a realization of a Gaussian white noise process, Fisher [18] derived the exact distribution for g. In practice, a simple approximation to this distribution tells us to reject the null hypothesis at the (approximate) level of significance α if g exceeds the value $1 - (\alpha/J)^{1/(J-1)}$. Siegel [19] proposed a test statistic similar to Fisher's test, with the key modification that it considers *all* large values of the periodogram and hence is more powerful against an alternative hypothesis of multiple sinusoidal components plus Gaussian white noise (for details – including some simple approximations to the percentage points for Siegel's test – see [4], Section 10.9).

The three statistical tests we have briefly discussed so far can all give misleading results when the additive noise $\{W_t\}$ cannot reasonably be assumed to be white noise. An appealing approach in this case is to use Thomson's multitaper F-test, which is based the notion of frequency domain regression analysis. Details concerning – and examples of the use of – this test are given in [6] and Sections 10.11 and 10.13 of [4].

11.3 Bivariate Time Series

11.3.1 Basic Concepts

Let us now consider the bivariate real-valued stationary process $\{X_t, Y_t\}$ (see Chapter 3 of this book). We assume that the univariate real-valued stationary processes $\{X_t\}$ and $\{Y_t\}$ are both zero mean processes with sdf's given by, respectively, $S_X(\cdot)$ and $S_Y(\cdot)$. Under these conditions, the cross-spectral properties of $\{X_t, Y_t\}$ are given by the *cross spectrum* (sometimes called the *cross spectral density function*), which can be written as

$$S_{XY}(f) = \Delta t \sum_{\tau=-\infty}^{\infty} C_{\tau,XY} e^{-i2\pi f \tau \, \Delta t}, \qquad -f_{(N)} \le f \le f_{(N)} \qquad (11.20)$$

(cf. Equation 11.3), where $\{C_{\tau,XY}\}$ is the *cross covariance sequence* (ccvs) defined by

$$C_{\tau,XY} = Cov(X_t, Y_{t+\tau}) = E(X_t Y_{t+\tau}).$$

To learn what the cross spectrum can tell us, let us consider a specific example appropriate for the bivariate wave height time series described in

Section 11.1 and shown in Figure 11.1. Let $\{O_t\}$ be a zero mean stationary process (with sdf $S_O(\cdot)$) that represents the true height of the ocean waves at the wire wave gauge. Because of instrumentation noise, what the wire wave gauge actually records at index t is $X_t = O_t + U_t$, where we assume that $\{U_t\}$ is a zero mean stationary process (with sdf $S_U(\cdot)$) that represents the instrumentation noise and is pairwise uncorrelated with $\{O_t\}$ (i.e., $C_{\tau,OU} = E(O_t U_{t+\tau}) = 0$ for all τ). At the infrared wave gauge, we assume that we measure $\{Y_t\}$, which is the sum of a displaced version of $\{O_t\}$ and a separate instrumentation noise process $\{V_t\}$; i.e., $Y_t = O_{t+l} + V_t$, where l is a fixed integer, and $\{V_t\}$ is a zero mean stationary process (with sdf $S_V(\cdot)$) that is pairwise uncorrelated with both $\{O_t\}$ and the other instrumentation noise process $\{U_t\}$ (i.e., $C_{\tau,OV} = 0$ and $C_{\tau,UV} = 0$ for all τ). Note that, if $l < 0$, we can say that the process $\{Y_t\}$ *lags* the process $\{X_t\}$ by $-l\,\Delta t$ time units because both X_{t+l} and Y_t depend on O_{t+l}, and the index $t+l$ occurs before the index t. Conversely, if $l > 0$, we say that $\{Y_t\}$ *leads* the process $\{X_t\}$ by $l\,\Delta t$ time units.

If we denote the acvs's for $\{O_t\}$, $\{U_t\}$ and $\{V_t\}$ by, respectively, $\{C_{\tau,O}\}$, $\{C_{\tau,U}\}$ and $\{C_{\tau,V}\}$, then the ccvs for $\{X_t, Y_t\}$ is given by

$$C_{\tau,XY} = E(X_t Y_{t+\tau}) = E[(O_t + U_t)(O_{t+l+\tau} + V_{t+\tau})] = C_{l+\tau,O};$$

i.e., the ccvs is just a shifted version of the acvs for $\{O_t\}$, and the amount of the shift is just given by the lead/lag factor l. The cross-spectrum for $\{X_t, Y_t\}$ is just $S_{XY}(f) = e^{i2\pi f l\,\Delta t} S_O(f)$. Note that this cross-spectrum expresses the commonality between $\{X_t\}$ and $\{Y_t\}$; i.e., it does not depend on either $\{U_t\}$ (which is part of $\{X_t\}$ but not of $\{Y_t\}$) or $\{V_t\}$ (part of $\{Y_t\}$ but not $\{X_t\}$).

Because $S_{XY}(f) = e^{i2\pi f l\,\Delta t} S_O(f)$ in this simple example, we can deduce that, whereas the sdf is always real-valued, the cross-spectrum is in general complex-valued. This fact makes it difficult to interpret $S_{XY}(\cdot)$ and leads us to search for quantities related to it, but easier to deal with. If we express the conjugate of $S_{XY}(f)$ in terms of its real and imaginary components, i.e., $S_{XY}(f) = R_{XY}(f) - iI_{XY}(f)$, we can define the *co-spectrum* $R_{XY}(\cdot)$ and the *quadrature spectrum* $I_{XY}(\cdot)$; on the other hand, if we express $S_{XY}(f)$ as its absolute value times a complex exponential, i.e., $S_{XY}(f) = A_{XY}(f)e^{i\phi_{XY}(f)}$ with $A_{XY}(f) = |S_{XY}(f)|$, we can define the *cross amplitude spectrum* $A_{XY}(\cdot)$ and the *phase spectrum* $\phi_{XY}(f)$. Note that, if in fact $|A_{XY}(f)| = 0$, then $\phi_{XY}(f)$ is ill-defined; if $|A_{XY}(f)| > 0$, then $\phi_{XY}(f)$ is defined only up to an integer multiple of 2π. In contrast to $S_{XY}(\cdot)$, the derived functions $R_{XY}(\cdot)$, $I_{XY}(\cdot)$ and $A_{XY}(\cdot)$ are all real-valued and hence can at least be plotted easily. The function $\phi_{XY}(\cdot)$ is also real-valued, but the '2π' ambiguity in its definition makes plotting it somewhat problematic (see the discussion in the next section).

Let us now see what these derived functions are for our simple ocean wave example. Since then $S_{XY}(f) = e^{i2\pi f l\,\Delta t} S_O(f)$, we have $R_{XY}(f) = \cos(2\pi f l\,\Delta t)\, S_O(f)$, $I_{XY}(f) = -\sin(2\pi f l\,\Delta t)\, S_O(f)$, $A_{XY}(f) = S_O(f)$ and, assuming that $A_{XY}(f) > 0$, $\phi_{XY}(f) = 2\pi f l\,\Delta t$. For this example, the cross amplitude spectrum and the phase spectrum are most easily interpretable: the former is just the sdf for the ocean wave process $\{O_t\}$, while the latter is linear with a slope proportional to the lead/lag factor l. In general, if it exists, the function that by $-\frac{1}{2\pi} \cdot \frac{d\phi_{XY}(f)}{df}$ is called the *group delay* (or the *envelope delay* – cf. [3], p. 664). In our example, the group delay is equal to $-l\,\Delta t$ at all frequencies, but in general it depends upon the frequency f.

Whereas the co–, quadrature, cross amplitude and phase spectra and the group delay can all be derived solely from the cross spectrum, the *complex coherency*, defined as $w_{XY}(f) = S_{XY}(f)/\sqrt{S_X(f)S_Y(f)}$, depends on both the cross spectrum and the sdf's for $\{X_t\}$ and $\{Y_t\}$. If we write the spectral representations for $\{X_t\}$ and $\{Y_t\}$ as

$$X_t = \int_{-f_{(N)}}^{f_{(N)}} e^{i2\pi f t\,\Delta t}\, dZ_X(f) \text{ and } Y_t = \int_{-f_{(N)}}^{f_{(N)}} e^{i2\pi f t\,\Delta t}\, dZ_Y(f)$$

(cf. Equation 11.1), it can be shown that

$$w_{XY}(f) = \frac{Cov[dZ_X(f), dZ_Y(f)]}{\sqrt{V[dZ_X(f)]V[dZ_Y(f)]}}$$

(see [3], p. 661). The complex coherency is thus a complex-valued frequency domain correlation coefficient that measures the correlation in the random amplitudes assigned to the complex exponentials with frequency f in the spectral representations for $\{X_t\}$ and $\{Y_t\}$. Since it is a complex-valued correlation coefficient, we must have $0 \le |w_{XY}(f)|^2 \le 1$. The quantity $|w_{XY}(f)|^2$ is called the *magnitude squared coherence* (msc) at the frequency f. Note that

$$|w_{XY}(f)|^2 = \frac{|S_{XY}(f)|^2}{S_X(f)S_Y(f)} = \frac{A_{XY}^2(f)}{S_X(f)S_Y(f)};$$

i.e., the msc is a normalized version of the square of the cross amplitude spectrum. The msc essentially captures the 'amplitude' part of the cross spectrum, but completely ignores its phase, so the msc and the phase spectrum together are useful real-valued summaries of the 'information' in the complex-valued cross spectrum.

To derive the msc for our ocean wave example, we first note that, because $\{O_t\}$, $\{U_t\}$ and $\{V_t\}$ are pairwise uncorrelated processes, the sdf's for

$\{X_t\}$ and $\{Y_t\}$ can be expressed as, respectively, $S_X(f) = S_O(f) + S_U(f)$ and $S_Y(f) = S_O(f) + S_V(f)$. The msc for $\{X_t, Y_t\}$ is thus given by

$$|w_{XY}(f)|^2 = 1 \Big/ \left[1 + S_U(f)/S_O(f)\right] \left[1 + S_V(f)/S_O(f)\right],$$

which tells us that the msc is

1. zero if $S_O(f) = 0$ while either $S_U(f) > 0$ or $S_U(f) > 0$ (because then the variability in $\{X_t\}$ and $\{Y_t\}$ at frequency f is due to instrumentation noise, assumed to be uncorrelated between gauges);

2. close to zero if either $S_U(f)$ or $S_V(f)$ is large compared to $S_O(f)$ (because then the variability at f in at least one of the component processes of $\{X_t, Y_t\}$ is mainly due to instrumentation noise);

3. close to unity if both $S_U(f)$ and $S_V(f)$ are small compared to $S_O(f)$ (because then the variability in both $\{X_t\}$ and $\{Y_t\}$ at f is mainly due to the variability in $\{O_t\}$; and

4. unity if, for example, both $U_t = 0$ and $V_t = 0$ for all t (because then $\{Y_t\}$ is just a time-shifted version of $\{X_t\}$),

all of which lends credence to the interpretation of the msc as a measure of the correlation between $\{X_t\}$ and $\{Y_t\}$ at particular frequencies.

11.3.2 Bivariate Spectral Estimation

We now turn to the problem of estimating the cross spectrum and various functions derived from it. Our estimates are based upon a bivariate time series that can be regarded as a realization of a portion X_t, Y_t, $t = 1, \ldots, n$, of the bivariate stationary process $\{X_t, Y_t\}$ with cross spectrum $S_{XY}(\cdot)$ and sdf's $S_X(\cdot)$ and $S_Y(\cdot)$. In view of Equation 11.20, an 'obvious' estimator for $S_{XY}(f)$ is the *cross periodogram* given by

$$\hat{S}^{(p)}_{XY}(f) = \Delta t \sum_{\tau=-(n-1)}^{n-1} \hat{C}^{(p)}_{\tau,XY} e^{-i2\pi f \tau \, \Delta t}$$

(cf. the periodogram of Equation 11.5), where $\hat{C}^{(p)}_{\tau,XY} = \frac{1}{n} \sum_t X_t Y_{t+\tau}$ (here the summation ranges from $t = 1$ to $t = n - \tau$ for nonnegative τ's and from $t = 1 - \tau$ to $t = n$ for negative τ's). Note that this expression for $\hat{C}^{(p)}_{\tau,XY}$ implicitly assumes that both $\{X_t\}$ and $\{Y_t\}$ are known *a priori* to be zero mean processes – if this is not the case, the common practice is to use $X_t - \bar{X}$ and $Y_t - \bar{Y}$ in place of X_t and Y_t in all computational formulae,

where $\bar{X}$ and $\bar{Y}$ are the sample means. In analogy to Equation 11.6, the cross periodogram can also be written as

$$\hat{S}_{XY}^{(p)}(f) = \frac{\Delta t}{n} \left(\sum_{t=1}^{n} X_t e^{-i2\pi f t\, \Delta t} \right) \left(\sum_{t=1}^{n} Y_t e^{-i2\pi f t\, \Delta t} \right)^* , \qquad (11.21)$$

where the asterisk denotes complex conjugation.

Unfortunately, the bias and variance properties of the cross periodogram are as poor as those of the periodogram, so $\hat{S}_{XY}^{(p)}(f)$ must be used with great caution. For example, suppose we use the cross periodogram along with the individual periodograms to form the 'obvious' estimator for the msc, namely, $|\hat{S}_{XY}^{(p)}(f)|^2 \Big/ \hat{S}_X^{(p)}(f)\hat{S}_Y^{(p)}(f)$. Because of Equations 11.21 and 11.6, we have

$$|\hat{S}_{XY}^{(p)}(f)|^2 = \frac{(\Delta t)^2}{n^2} \left| \sum_{t=1}^{n} X_t e^{-i2\pi f t\, \Delta t} \right|^2 \left| \sum_{t=1}^{n} Y_t e^{-i2\pi f t\, \Delta t} \right|^2 = \hat{S}_X^{(p)}(f)\hat{S}_Y^{(p)}(f)$$

so that the 'obvious' estimator for the msc is in fact always *unity*! Priestley [3], p. 708, gives a reasonable explanation for this seemingly unreasonable result, namely, that the frequency domain correlation coefficient $|w_{XY}(f)|^2$ is essentially being estimated using just a single observation of a bivariate frequency domain process at frequency f. It is thus vital to reduce the inherent variability in the cross periodogram if the corresponding msc estimator is to make any sense. In principle, variance reduction can be achieved using straight-forward extensions of the approaches discussed in Section 11.2.3 for univariate sdf estimation, namely, smoothing across frequencies, WOSA and multitapering. In practice, however, WOSA and multitapering are the methods of choice because of subtle (but surmountable) problems that can arise when smoothing across frequencies (in addition, Walden [20] demonstrates an improvement in msc estimation using multitapering instead of smoothing across frequencies). For brevity, we consider just the multitaper estimator (see Carter [13] for details on WOSA; Thomson and Chave [17] discuss WOSA, multitapering and a combination of the two methods).

The simpliest multitaper estimator of the cross spectrum is given by

$$\hat{S}_{XY}^{(mt)}(f) = \frac{\Delta t}{K} \sum_{k=0}^{K-1} \left(\sum_{t=1}^{n} h_{t,k} X_t e^{-i2\pi f t\, \Delta t} \right) \left(\sum_{t=1}^{n} h_{t,k} Y_t e^{-i2\pi f t\, \Delta t} \right)^* ,$$

where, as before, $\{h_{t,k}\}$ is the kth-order dpss data taper for a sequence of length n and a fixed resolution bandwidth $2W$. As in the univariate

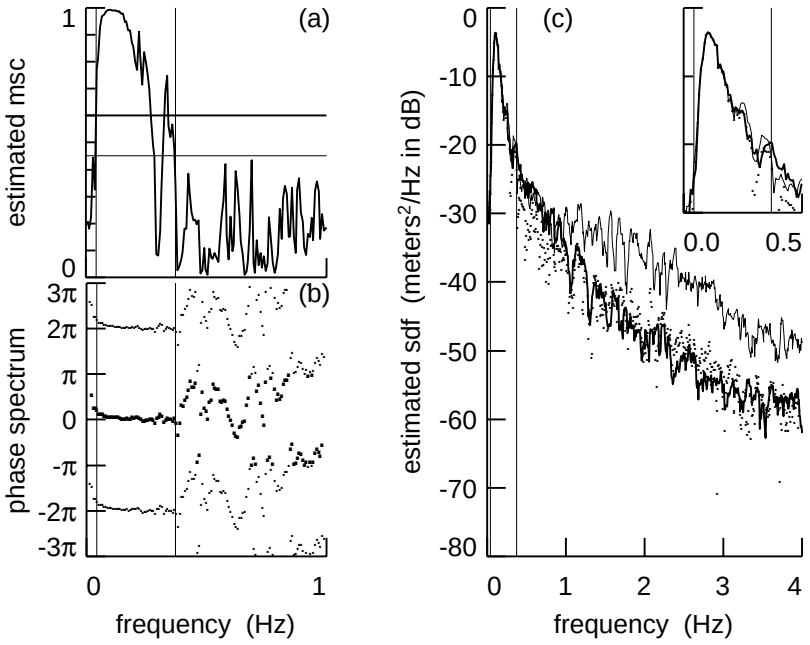

Figure 11.10: Bivariate spectral analysis (see text for details).

case, each data taper is designed to prevent leakage, whereas the use of multiple tapers yields an estimator of the cross spectrum having less variability than the cross periodogram. The number of equivalent degrees of freedom for $\hat{S}_{XY}^{(mt)}(f)$ is $\nu = 2K$. The corresponding multitaper estimators for the co-spectrum and related quantities follow directly from their definitions. For example, the estimators for the phase spectrum and the msc are given by $\hat{\phi}_{XY}^{(mt)}(f) = \arg\left(\hat{S}_{XY}^{(mt)}(f)\right)$ and $|\hat{w}_{XY}^{(mt)}(f)|^2 = |\hat{S}_{XY}^{(mt)}(f)|^2 \Big/ \hat{S}_X^{(mt)}(f)\hat{S}_Y^{(mt)}(f)$. In the expression for $\hat{\phi}_{XY}^{(mt)}(f)$, we can assume that the 'arg' function returns a value between $-\pi$ and π so that $\hat{\phi}_{XY}^{(mt)}(f)$ assumes values modulo 2π and hence can be discontinuous as phases pass over the $\pm\pi$ boundaries. These discontinuities can hinder ascertaining whether or not the phase spectrum varies linearly with frequency (an indication of a lead/lag relationship between $\{X_t\}$ and $\{Y_t\}$). To avoid these difficulties, Priestley [3], p. 709, advocates plotting $\hat{\phi}_{XY}^{(mt)}(f)$, $\hat{\phi}_{XY}^{(mt)}(f) + 2\pi$ and $\hat{\phi}_{XY}^{(mt)}(f) - 2\pi$ versus f all together.

Figure 11.10 summarizes a bivariate spectral analysis of the wire wave gauge time series $\{X_t\}$ of Figure 11.1(a) and the infrared wave gauge time

series $\{Y_t\}$ of Figure 11.1(b). The thick solid curve in 11.10(c) is a reproduction of the multitaper spectral estimate $\hat{S}_X^{(mt)}(f)$ shown previously in Figure 11.9(b). The thin solid curve is the corresponding estimate $\hat{S}_Y^{(mt)}(f)$ for the infrared wave gauge time series. The two spectral estimates agree almost perfectly between the frequencies $f = 0.04$ Hz and $f = 0.34$ Hz (these frequencies are marked by two thin vertical lines) and reasonably well between $f = 0.34$ and $f = 0.97$, beyond which the level of $\hat{S}_Y^{(mt)}(f)$ is consistently higher than that of $\hat{S}_X^{(mt)}(f)$. This finding is in agreement with the fact that the infrared wave gauge has substantially more measurement noise at high frequencies than does the wire wave gauge. The dots in Figure 11.10(c) show the multitaper estimate $\hat{A}_{XY}^{(mt)}(\cdot)$ of the cross amplitude spectrum. Recall that, in terms of the simple lead/lag model for this data, the cross amplitude spectrum is just the sdf for the common ocean wave process. Again there is very good agreement between $\hat{A}_{XY}^{(mt)}(\cdot)$ and either $\hat{S}_X^{(mt)}(\cdot)$ or $\hat{S}_Y^{(mt)}(\cdot)$ at the low frequencies delineated by the two thin vertical lines, after which $\hat{A}_{XY}^{(mt)}(\cdot)$ is in general agreement with $\hat{S}_X^{(mt)}(\cdot)$ but not $\hat{S}_Y^{(mt)}(\cdot)$. This finding is consistent with the measurement noise at high frequencies for the wire wave gauge being smaller than that of the other gauge. There are some small (less than 10 dB) differences between $\hat{S}_X^{(mt)}(\cdot)$ and $\hat{A}_{XY}^{(mt)}(\cdot)$, particularly between the frequencies $f = 0.34$ and $f = 1.0$, where $\hat{A}_{XY}^{(mt)}(f)$ is consistently lower than $\hat{S}_X^{(mt)}(\cdot)$. Our simple lead/lag model would suggest that the common ocean wave process $\{O_t\}$ is weak at these frequencies compared to local effects, all of which are lumped together with instrumentation noise in our simple model.

The solid curve in Figure 11.10(a) shows the corresponding estimated msc $|\hat{w}_{XY}^{(mt)}(f)|^2$ for $0 \leq f \leq 1$ Hz (the two thin vertical lines mark the same frequencies as the corresponding lines in 11.10(c)). We can test – at the level of significance α – the null hypothesis that the true msc is zero at frequency f by comparing $|\hat{w}_{XY}^{(mt)}(f)|^2$ to the value $1 - \alpha^{2/(\nu-2)}$ and rejecting the null hypothesis if $|\hat{w}_{XY}^{(mt)}(f)|^2$ exceeds this value (for details, see [2], p. 284). As before, ν represents the number of equivalent degrees of freedom associated with the spectral estimates. Because we have used $K = 6$ tapers in the multitaper estimates $\hat{S}_X^{(mt)}(\cdot)$, $\hat{S}_Y^{(mt)}(\cdot)$ and $\hat{S}_{XY}^{(mt)}(\cdot)$, we have $\nu = 2K = 12$ degrees of freedom. The thick and thin horizontal lines on Figure 11.10(a) define the rejection regions for levels of significance of, respectively, $\alpha = 0.01$ and $\alpha = 0.05$. We see that the region of nonzero msc is bounded by the thin vertical lines marking $f = 0.04$ Hz and $f = 0.34$ Hz, between which $\hat{S}_X^{(mt)}(\cdot)$, $\hat{S}_Y^{(mt)}(\cdot)$ and $\hat{A}_{XY}^{(mt)}(\cdot)$ are all in good agreement (for $1.0 < f < 4.0$, the estimated msc is such that, except at a few isolated

frequencies, we cannot reject the null hypothesis of zero msc). In terms of our simple lead/lag model, the small msc at frequencies higher than $f = 0.34$ Hz can be attributed to the large instrumentation noise associated with the infrared wave gauge.

Finally, the dots in Figure 11.10(b) show the estimated phase spectrum $\hat{\phi}_{XY}^{(mt)}(f)$ versus f for $0 \leq f \leq 1$ Hz, plotted per Priestley's recommendation (the dots for $\hat{\phi}_{XY}^{(mt)}(f) \pm 2\pi$ are smaller than the dots for $\hat{\phi}_{XY}^{(mt)}(f)$). Again, vertical lines delineate the frequencies of apparent nonzero msc, between which the estimated phase spectrum increases approximately linearly. A linear least squares fit to $\hat{\phi}_{XY}^{(mt)}(f)$ versus f at these frequencies yields an estimated slope of 1.02 radians per Hz. In terms of the lead/lag model, this slope translates into a lead/lag factor of $l = 4.9 \approx 5$ units or 1/6 second since one unit is equal to 1/30 second. Because l is positive, the prevailing ocean waves are arriving at the infrared wave gauge approximately 1/6 second ahead of the wire wave gauge; moreover, because the two gauges were 6 meters apart, the prevalent direction of the ocean waves is approximately perpendicular to a line drawn between the two gauges. Note that, in regions where the estimated msc is small, the estimated phase spectrum is erratic. This pattern is consistent with well-known approximations to the statistical properties of $\hat{\phi}_{XY}^{(mt)}(f)$, which indicate that it has large variability when the true msc is zero ([3], p. 703).

Detailed discussions on bivariate spectral analysis can be found in Koopmans [2] and Priestley [3]. Thomson and Chave [17] describe resampling schemes for assessing variability in bivariate spectral estimates.

Acknowledgments

R. Spindel and J. Harlett of the Applied Physics Laboratory, University of Washington, graciously provided discretionary funding to support writing this chapter. The author would also like to thank Andrew Walden, Christopher Chatfield and the editors for helpful critiques.

11.4 References Cited

1. H. Cramér, On Harmonic Analysis in Certain Functional Spaces, *Arkiv för Matematik, Astronomi och Fysik, 28B*, 1–7 (1942).

2. L. H. Koopmans, "The Spectral Analysis of Time Series," Academic Press, New York, 1974.

3. M. B. Priestley, "Spectral Analysis and Time Series," Academic Press, London, 1981.

4. D. B. Percival and A. T. Walden, "Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques," Cambridge University Press, Cambridge, England, 1993.

5. A. T. Jessup, W. K. Melville, and W. C. Keller, Breaking Waves Affecting Microwave Backscatter: 1. Detection and Verification, *Journal of Geophysical Research, 96*, 20,547–59 (1991).

6. D. J. Thomson, Spectrum Estimation and Harmonic Analysis, *Proceedings of the IEEE, 70*, 1055–96 (1982).

7. A. T. Walden, Accurate Approximation of a 0th Order Discrete Prolate Spheroidal Sequence for Filtering and Data Tapering, *Signal Processing, 18*, 341–8 (1989).

8. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Numerical Recipes: The Art of Scientific Computing (Second Edition)," Cambridge University Press, Cambridge, England, 1992.

9. D. B. Percival and R. K. Kerr, $\mathrm{P_ITSS_A}$– A Time Series Analysis System Embedded in LISP, *Proceedings of the 20th Symposium on the Interface of Computer Science and Statistics*, 321–30 (1986).

10. D. J. Best and D. E. Roberts, The Percentage Points of the χ^2 Distribution: Algorithm AS 91, *Applied Statistics, 24*, 385–8 (1975).

11. E. Parzen, Mathematical Considerations in the Estimation of Spectra, *Technometrics, 3*, 167–90 (1961).

12. P. D. Welch, The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms, *IEEE Transactions on Audio and Electroacoustics, 15*, 70–3, (1967).

13. G. C. Carter, Coherence and Time Delay Estimation, *Proceedings of the IEEE, 75*, 236–55 (1987).

14. D. J. Thomson, Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide – I, *Bell System Technical Journal, 56*, 1769–815 (1977).

15. D. Slepian, Some Comments on Fourier Analysis, Uncertainty and Modeling, *SIAM Review, 25*, 379–93 (1983).

16. T. P. Bronez, On the Performance Advantage of Multitaper Spectral Analysis, *IEEE Transactions on Signal Processing, 40*, 2941–6 (1992).

17. D. J. Thomson and A. D. Chave, Jackknifed Error Estimates for Spectra, Coherences, and Transfer Functions, in "Advances in Spectrum Analysis and Array Processing" (S. Haykin, editor), Volume I, 58–113, Prentice-Hall, Englewood Cliffs, New Jersey, 1991.

18. R. A. Fisher, Tests of Significance in Harmonic Analysis, *Proceedings of the Royal Society of London, Series A, 125*, 54–9 (1929).

19. A. F. Siegel, Testing for Periodicity in a Time Series, *Journal of the American Statistical Association, 75*, 345–8 (1980).

20. A. T. Walden, Wavelet Estimation Using the Multitaper Method, *Geophysical Prospecting, 39*, 625-42 (1991).