Ch.4 Time series analysis [Book, Ch.3]

We have surveyed the linear multivariate techniques for extracting features or recognizing patterns in a dataset or in two datasets, without considering time as an explicit factor. In the real world, the variables may be governed by dynamical equations, and the manifested patterns evolve with time.

We will look at:
(1) Classical Fourier spectral analysis as a first step in analyzing the temporal behaviour in a time series.

(2) Windows and filters.

(3) Singular spectrum analysis (SSA).

4.1 Spectrum [Book, Sect. 3.1]

With time series data, an alternative way to view the data is via the frequency representation. Given a function $y(t)$ defined on the interval $[0, T]$, the Fourier series representation for $y(t)$ is

$$\hat{y}(t) = \frac{a_0}{2} + \sum_{m=1}^{\infty} [a_m \cos(\omega_m t) + b_m \sin(\omega_m t)], \tag{1}$$

with the (angular) frequency ω_m given by

$$\omega_m = \frac{2\pi m}{T}, \qquad m = 1, 2, \cdots \tag{2}$$

and the Fourier coefficients a_m and b_m given by

$$a_m = \frac{2}{T} \int_0^T y(t) \cos(\omega_m t) \, \mathrm{d}t, \quad m = 0, 1, 2, \cdots \tag{3}$$

$$b_m = \frac{2}{T}\int_0^T y(t)\sin(\omega_m t)\,\mathrm{d}t, \quad m = 1, 2, \cdots \tag{4}$$

With

$$a_0 = \frac{2}{T}\int_0^T y(t)\,\mathrm{d}t, \tag{5}$$

we see that

$$a_0/2 = \overline{y}, \tag{6}$$

the mean of y.

If $y(t)$ is a continuous function, then Eq.(1) has $\hat{y}(t) \to y(t)$.

If y is discontinuous at t, then $\hat{y}(t) \to [y(t+) + y(t-)]/2$.

For a discrete time series, $y(t)$ is replaced by
$y(t_n) \equiv y_n, \quad n = 1, \cdots, N$.

With a sampling interval $\Delta t = T/N$, the observations are made at time $t_n = n\Delta t$.

The discrete Fourier series representation is

$$y_n = \frac{a_0}{2} + \sum_{m=1}^{M} [a_m \cos(\omega_m t_n) + b_m \sin(\omega_m t_n)], \tag{7}$$

where M is the largest integer $\leq N/2$, with the Fourier coefficients:

$$a_m = \frac{2}{N} \sum_{n=1}^{N} y_n \cos(\omega_m t_n), \quad m = 0, 1, 2, \cdots, M, \tag{8}$$

$$b_m = \frac{2}{N} \sum_{n=1}^{N} y_n \sin(\omega_m t_n), \quad m = 1, 2, \cdots, M. \tag{9}$$

For N even, $b_M = 0$, so the number of nontrivial Fourier coefficients is N.

The cosine and sine functions have orthogonality properties:

$$\sum_{n=1}^{N} \cos(\omega_l t_n) \cos(\omega_m t_n) = \frac{N}{2} \delta_{lm} \,,$$

$$\sum_{n} \sin(\omega_l t_n) \sin(\omega_m t_n) = \frac{N}{2} \delta_{lm} \,,$$

$$\sum_{n} \cos(\omega_l t_n) \sin(\omega_m t_n) = 0 \,, \tag{10}$$

where δ_{lm} is the Kronecker delta function.

4.1.1 Autospectrum [Book, Sect. 3.1.1]

The variance of the time series y can be written as:

$$\begin{aligned} \mathrm{var}(y) &= \frac{1}{N}\sum_{n=1}^{N}(y_n - \overline{y})^2 = \frac{1}{N}\sum_{n}(y_n - \frac{a_0}{2})^2 \\ &= \frac{1}{N}\sum_{n}\left[\sum_{m}(a_m \cos(\omega_m t_n) + b_m \sin(\omega_m t_n))\right]^2 . \quad (11) \end{aligned}$$

Using (10), var(y) can be expressed in terms of the Fourier coefficients,

$$\mathrm{var}(y) = \frac{1}{2}\sum_{m}(a_m^2 + b_m^2). \quad (12)$$

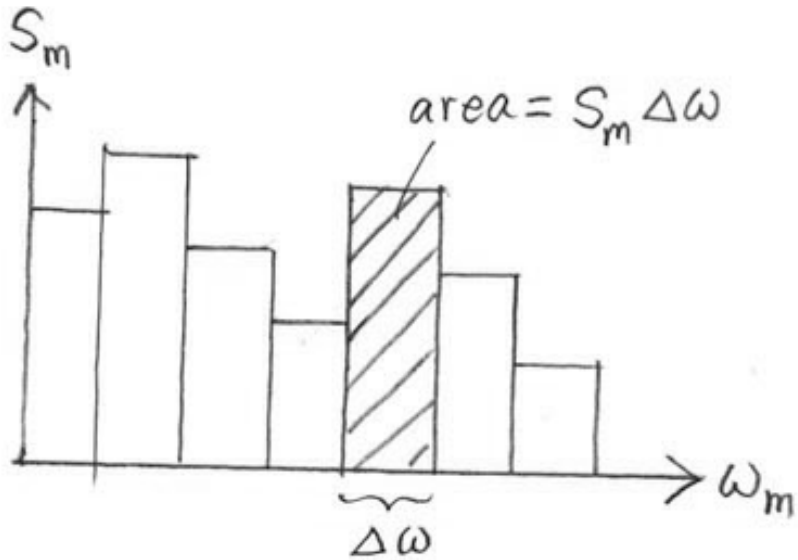
The *autospectrum*, also called the *spectrum*, the *power spectrum* or the *periodogram*, is defined as

$$S_m = \frac{N\Delta t}{4\pi}(a_m^2 + b_m^2). \tag{13}$$

Thus (12) can be expressed as

$$\mathrm{var}(y) = \sum_m \mathrm{S}_m \Delta\omega, \tag{14}$$

$$\Delta\omega = \frac{2\pi}{T} = \frac{2\pi}{N\Delta t}. \tag{15}$$



Hence, the spectrum S_m can be viewed as the variance or 'energy' in the ω_m frequency band (with bandwidth $\Delta\omega$), and the total variance var(y) can be computed by integrating S_m over all frequency bands.

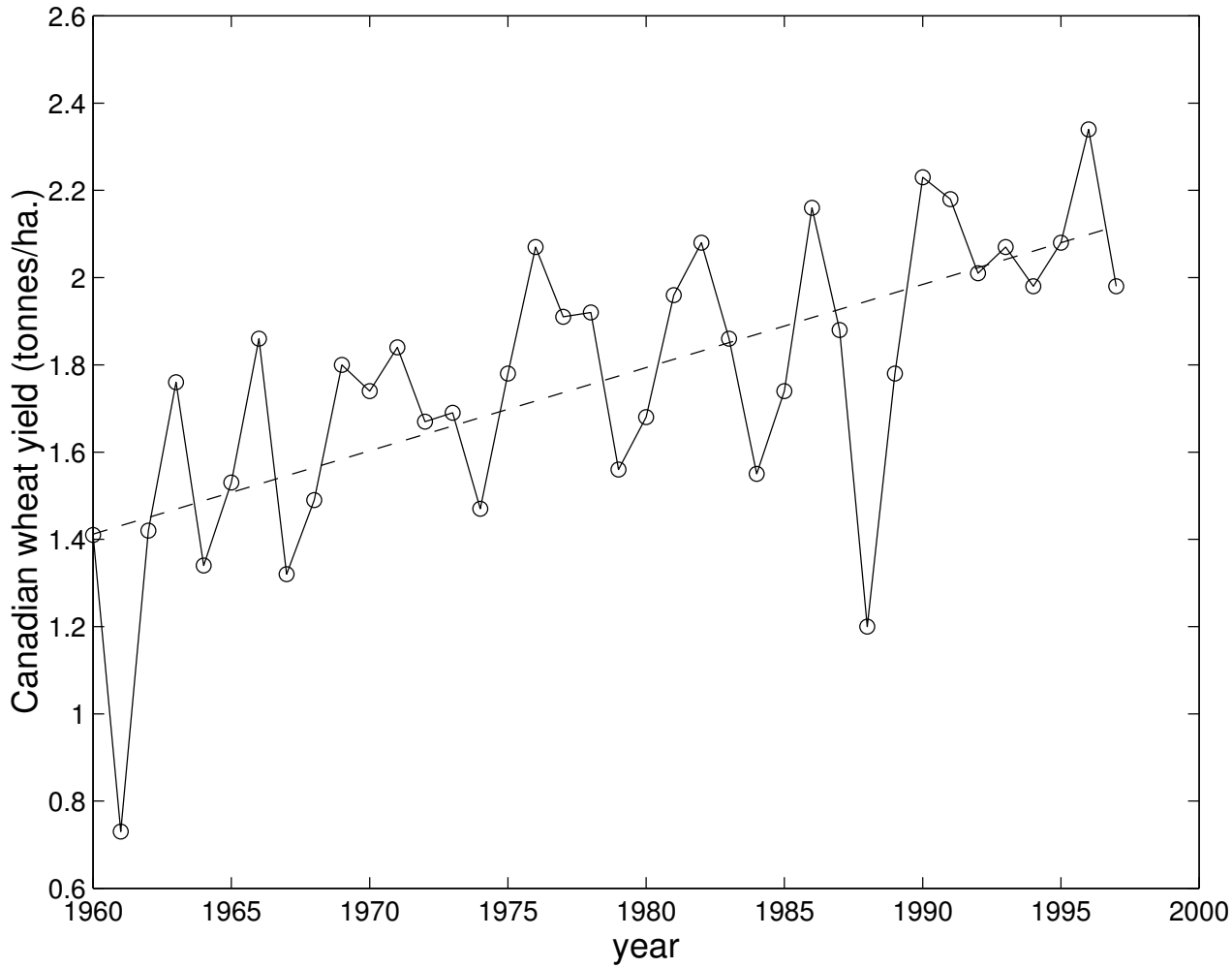
When S_m is plotted as a function of the frequency, peaks in S_m reveal the frequencies where the energy is relatively high.

The lowest frequency in the spectrum, known as the *fundamental frequency*, is

$$\omega_1 = \frac{2\pi}{T}, \quad \text{or} \quad f_1 = \frac{1}{T}. \tag{16}$$

Often a time series displays a trend, i.e. a positive or negative slope in the data over the time record.

E.g. Canadian prairie wheat yield shows a positive trend with time, largely due to the gradual improvement in agricultural technology.


Canadian wheat yield (tonnes/ha.)
year
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
1960
1965
1970
1975
1980
1985
1990
1995
2000

The frequency associated with a trend is lower than the fundamental frequency, thus energy from the trend will leak to other low frequency spectral bands, thereby distorting the low frequency part of the spectrum.

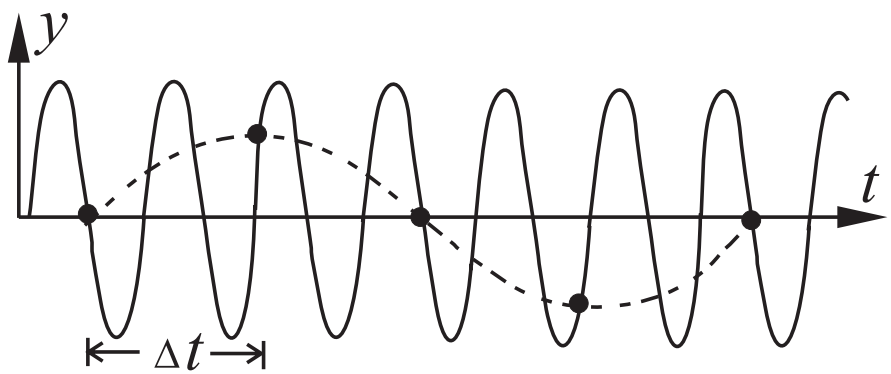
By subtracting the linear regression line from the data time series, trends can easily be removed prior to spectral analysis.

The highest resolvable frequency from (2) is $\omega = 2\pi M/T$, but with $M \approx N/2$, we have $M/T \approx 1/(2\Delta t)$. Hence the highest resolvable frequency, called the Nyquist frequency, is

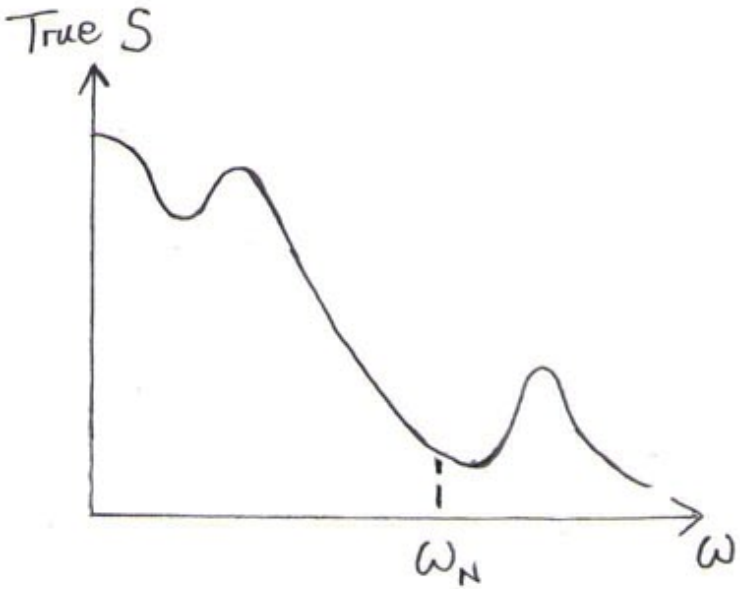
$$\omega_N = \frac{\pi}{\Delta t}, \quad \text{or} \quad f_N = \frac{1}{2\Delta t}. \tag{17}$$

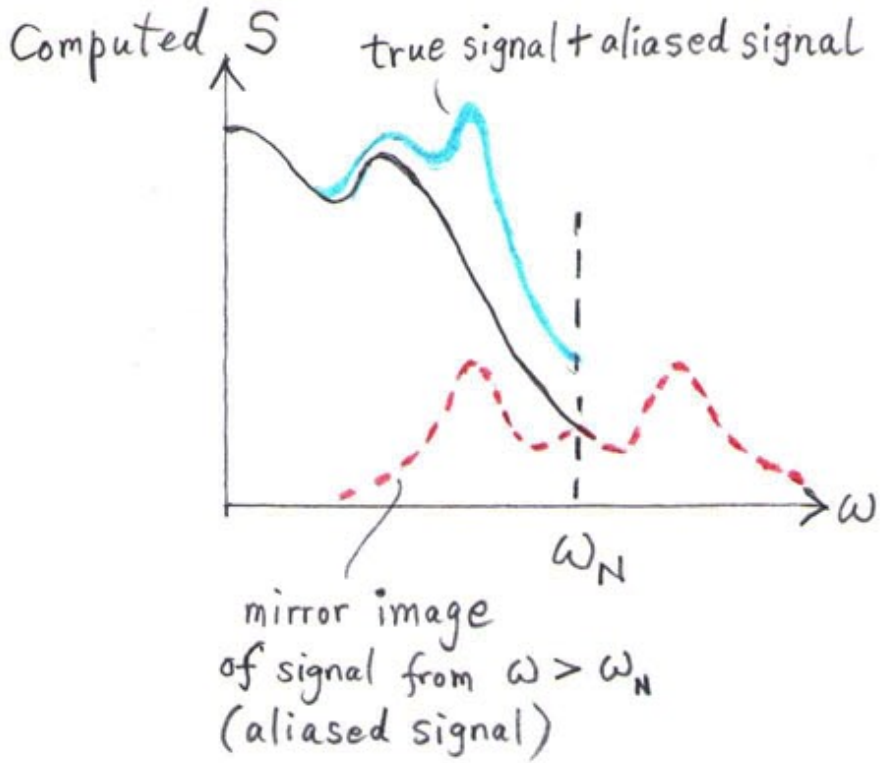
To resolve a wave of period τ, we need at least two data points to cover the period τ, i.e. $\tau = 2\Delta t = 1/f_N$.

Aliasing arises when the sampling time interval Δt is too large to resolve the highest frequency oscillations in the data. From the observations (dots), an incorrect signal (dashed curve) of much lower frequency is inferred by the observer.



In a spectrum, signals with frequency above the Nyquist frequency are reflected across the Nyquist frequency into the frequency bands below the Nyquist frequency— resulting in a distortion of the high frequency part of the spectrum.





From (15), the frequency $\Delta\omega$ between adjacent frequency bands is

$$\Delta\omega = \frac{2\pi}{T}. \tag{18}$$

The ability to resolve neighbouring spectral peaks is controlled by $\Delta\omega$, which is proportional to $1/T$.

A longer record T will yield sharper spectral peaks, allowing the resolution of two signals with close-by frequencies as distinct peaks in the spectrum.

Q1: If you have an instrument taking one measurement every second for 24 hours. What are (a) the fundamental frequency ω_1, (b) the Nyquist frequency ω_N and (c) $\Delta\omega$, the frequency between two adjacent frequency bands?

The raw spectrum S_m calculated from (13) is often very noisy in appearance. Two common methods for smoothing the spectrum:
(a) band-averaging (the Daniell estimator)
(b) ensemble averaging.

In (a), a moving average (or running mean) is applied to the raw spectrum

$$\tilde{S}_m = \frac{1}{(2K+1)} \sum_{k=-K}^{K} S_{m+k} \,, \tag{19}$$

where $\tilde{S}_m$ is the smoothed spectrum resulting from averaging the raw spectrum over $2K+1$ frequency bands.

In (b), data record is divided into J blocks of equal length $L = T/J$. Compute the periodogram for each block to get $S_m^{(j)}$ $(j = 1, \cdots, J)$.

The spectrum S_m is the ensemble average over the J periodograms:

$$S_m = \frac{1}{J} \sum_{j=1}^{J} S_m^{(j)} \,. \tag{20}$$

Method (b) has an advantage over (a) when there are data gaps—in (b), the data gaps do not pose a serious problem since the data record is to be chopped into J blocks anyway, whereas in (a), the data gaps may have to be filled with interpolated values or zeros.

The disadvantage of (b) is that the lowest resolvable frequency is $f_1 = 1/L = J/T$, hence there is a loss of low frequency information when the record is chopped up.

There is a trade-off between the variance of the spectrum S and the band width. Increasing the band width (by increasing K or J) leads

to a less noisy S, but spectral peaks are broadened, so that nearby spectral peaks may merge together, resulting in a loss of resolution.

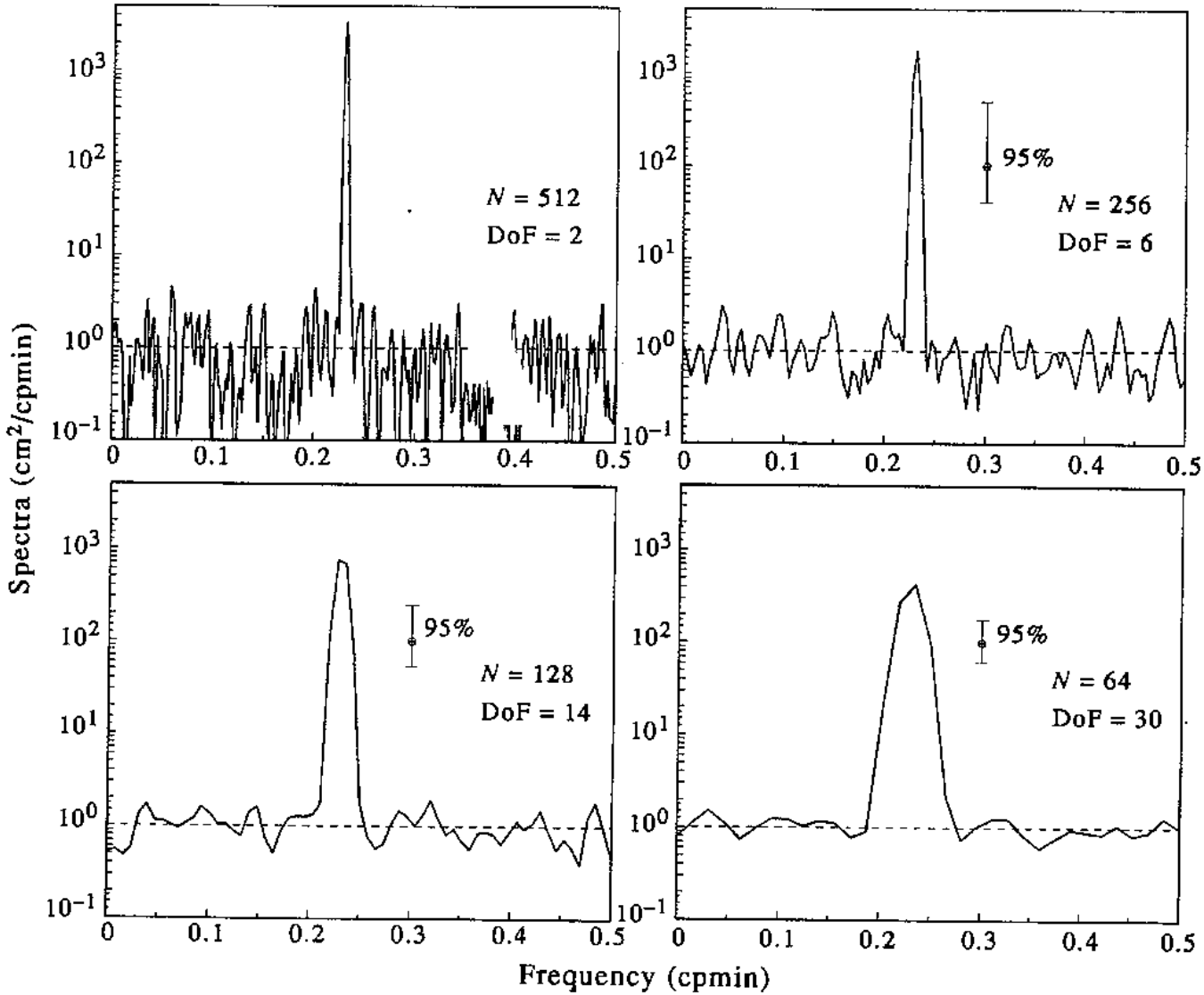


Figure 5.6.25. Periodogram power spectral estimates for a time series composed of Gaussian white noise and a single cosine constituent with a frequency of 0.23 cpmin and amplitude five times that of the white noise component. N = number of spectral bands and vertical lines are the 95% confidence intervals. (a) Raw (unsmoothed) periodogram, with DOF = 2; (b) smoothed periodogram, by averaging three adjacent spectral estimates such that DOF = 6; (c) as with (b) but for seven frequency bands, and DOF = 14; as with (b) but for 15 frequency bands, DOF = 30.

Figure : [From Emery and Thomson (1997)]

In complex notation, the Fourier transform

$$\hat{y}_m = \frac{2}{N}\sum_{n=1}^{N} y_n\, e^{-i\omega_m t_n}\,, \tag{21}$$

$$= a_m - ib_m\,, \tag{22}$$

where (8) and (9) have been invoked. Eq.(13) can be written as

$$S_m = \frac{N\Delta t}{4\pi}|\hat{y}_m|^2\,. \tag{23}$$

Optional material:

Prove that the spectrum is related to the auto-covariance function by a Fourier transform:

Assume $\{y_n\}$ is stationary and the mean $\overline{y}$ has been subtracted from the data, then

$$S_m = \frac{\Delta t}{N\pi}\left[\sum_n y_n\, e^{-i\omega_m t_n}\right]\left[\sum_j y_j\, e^{i\omega_m t_j}\right], \tag{24}$$

$$= \frac{\Delta t}{\pi}\sum_{l=-(N-1)}^{N-1}\left[\frac{1}{N}\left(\sum_{j-n=l} y_n y_j\right)\right] e^{i\omega_m t_l}\,. \tag{25}$$

In general, the *auto-covariance* function with lag l is defined as

$$C_l = \frac{1}{N}\sum_{n=1}^{N-l}(y_n - \overline{y})(y_{n+l} - \overline{y})\,. \tag{26}$$

Here (with $\overline{y} = 0$), we have the important relation

$$S_m = \frac{\Delta t}{\pi} \sum_{l=-(N-1)}^{N-1} C_l \, e^{i\omega_m t_l} \,, \tag{27}$$

i.e. the spectrum S_m is related to the auto-covariance function C_l by a Fourier transform.

4.1.2 Cross-spectrum [Book, Sect. 3.1.2]

Consider two time series, $x_1, \cdots, x_N$ and $y_1, \cdots, y_N$, with respective Fourier transforms $\hat{x}_m$ and $\hat{y}_m$ (which are in general complex numbers). The *cross-spectrum*

$$C_m = \frac{N\Delta t}{4\pi} \hat{x}_m \hat{y}_m^* \,, \tag{28}$$

where asterisk denotes complex conjugation, so C_m is in general complex. If $\hat{y}_m = \hat{x}_m$, C_m reduces to S_m, which is real.

C_m can be split into real and imaginary parts,

$$C_m = R_m + i\, I_m \,, \tag{29}$$

where R_m is the *co-spectrum* and I_m is the *quadrature spectrum*.

C_m can also be expressed in polar form,

$$C_m = A_m\, e^{i\theta_m} \,, \tag{30}$$

where A_m is the *amplitude spectrum* and θ_m, the *phase spectrum*, with

$$A_m = [\, R_m^2 + I_m^2 \,]^{\frac{1}{2}} \quad \text{and} \quad \theta_m = \tan^{-1}(I_m \,/\, R_m) \,. \tag{31}$$

A useful quantity is the *squared coherency spectrum* (where the word 'squared' is often omitted for brevity):

$$r_m^2 = \frac{A_m^2}{S_m^{(x)}\, S_m^{(y)}}\,, \tag{32}$$

where $S_m^{(x)}, S_m^{(y)}$ are the autospectrum for the x and y time series, respectively.

Can interpret r_m^2 as the square of the correlation between x and y in the mth frequency band. However, if one does not perform band averaging or ensemble averaging, then $r_m^2 = 1$, i.e. perfect correlation for all m!

To see this, let

$$\hat{x}_m = a_m\, e^{i\alpha_m} \quad \text{and} \quad \hat{y}_m = b_m\, e^{i\beta_m}\,. \tag{33}$$

Eq.(28) becomes

$$C_m = \frac{N\Delta t}{4\pi}\, a_m\, b_m\, e^{i(\alpha_m - \beta_m)}\,. \tag{34}$$

Thus

$$A_m = \frac{N\Delta t}{4\pi}\, a_m\, b_m \quad \text{and} \quad \theta_m = \alpha_m - \beta_m\,. \tag{35}$$

Also,

$$S_m^{(x)} = \frac{N\Delta t}{4\pi}\, a_m^2\,, \quad \text{and} \quad S_m^{(y)} = \frac{N\Delta t}{4\pi}\, b_m^2\,. \tag{36}$$

Substituting these into eq.(32) yields $r_m^2 = 1$. The reason is that in a single frequency band, the x and y signals are simply sinusoidals of the same frequency, which are perfectly correlated (other than a possible phase shift between the two).

Suppose there is no real relation between $\hat{x}_m$ and $\hat{y}_m$, then the phase $\alpha_m - \beta_m$ tends to be random. Consider ensemble averaging, with

$$C_m = \frac{1}{J} \sum_{j=1}^{J} C_m^{(j)} \,. \tag{37}$$

With random phase, the $C_m^{(j)}$ vectors are randomly oriented in the complex plane, so the summing of the $C_m^{(j)}$ vectors tends not to produce a C_m vector with large magnitude A_m. In general for large J, $A_m^2 \ll S_m^{(x)} \, S_m^{(y)}$, resulting in a small value for r_m^2, as desired.

Thus some form of ensemble averaging or band averaging is essential for computing the squared coherency spectrum— without the averaging, even random noise has r_m^2 equal to unity.

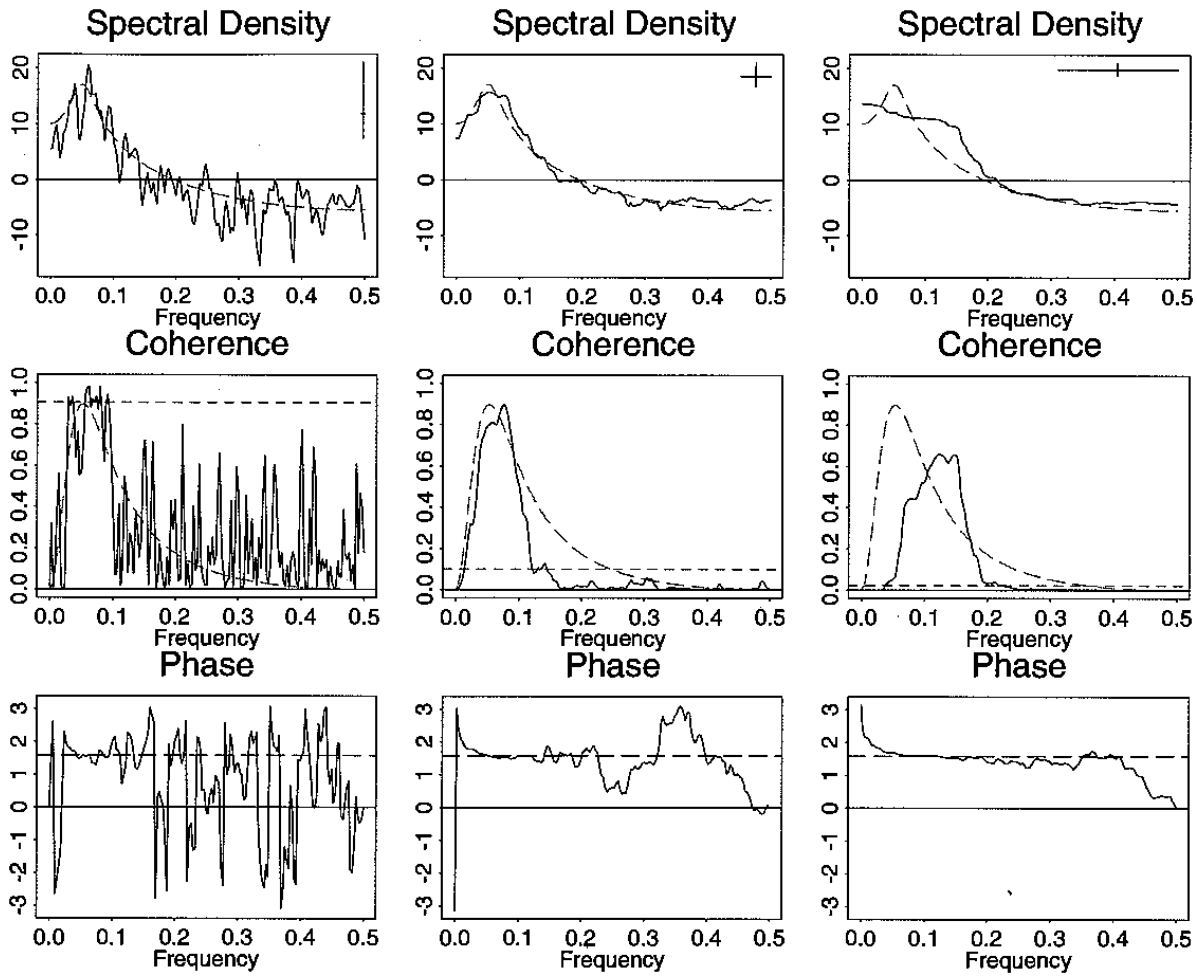


Figure 12.23: *Cross-spectral estimates computed from a time series of length* $T = 384$ *generated from a bivariate AR(1) process with a rotational parameter matrix (11.51) (cf. [11.3.8] and [11.4.8,9]). The columns contain Daniell estimates for* $n = 2$, *16, and* 64, *from the left.*
Top row: *The estimated spectrum of the* $\mathbf{X}_t$ *component of the process, in decibels. The dashed curve indicates the true spectrum. The cross indicates the bandwidth (horizontal) and width of the 95% confidence interval (vertical).*
Middle row: *The estimated coherency. The long-dashed curve indicates the true coherency spectrum. The short-dashes indicate the critical value for the 5% significance level test of zero coherency.*
Bottom row: *The estimated phase. The dashed line indicates the true phase.*

Figure : [From von Storch and Zwiers (1999)]

It can also be shown that the cross-spectrum is the Fourier transform of the *cross-covariance* γ, where

$$\gamma = \frac{1}{N} \sum_{n=1}^{N-l} (x_n - \overline{x})(y_{n+l} - \overline{y}) \,. \tag{38}$$

Spectral functions in Matlab

The autospectrum (periodogram) can be computed by using `www.mathworks.com/help/toolbox/signal/ref/periodogram.html`

[S, omega] = periodogram(y)
where S, the autospectrum, is a function of omega, the angular frequency, and y is the original time series.

Fast Fourier Transform (FFT) is particularly fast when the number of data points is in powers of 2, the default of the periodogram function tries to use this property.

[S, omega] = periodogram(y, [], nfft)

where nfft is the number of data points to be used in the FFT (e.g. nfft = total number of data points).

[S, omega] = periodogram(y, window, nfft)

allows user to specify a window (discussed later in this chapter).

Cross-spectral analysis can be done using Matlab functions mscohere and cpsd:

`http://www.mathworks.com/help/signal/ug/cross-spectrum-and-magnitude-squared-coherence.html`

Spectral and cross-spectral analysis can also be done using the FFT function provided by Matlab.

`www.mathworks.com/help/techdoc/ref/fft.html`
fx = fft(x)
where fx is the (discrete) Fourier transform of x computed by an FFT algorithm.

References:

Emery, W. J. and Thomson, R. E. (1997). *Data Analysis Methods in Physical Oceanography*. Pergamon, Oxford.

von Storch, H. and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge Univ. Pr., Cambridge.