The Complex Fourier Transform:

In previous sections we presented the Fourier Transform in real arithmetic using sine and cosine functions. It is much more compact and efficient to write the Fourier Transform and its associated manipulations in complex arithmetic. In a domain of continuous time and frequency, we can write the Fourier Transform Pair as integrals:

$$f(t) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega$$

$$F(\omega) = \int\limits_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

Here $f(t)$ is some real time series in the independent variable t, and $F(\omega)$ is the Fourier Transform of f(t), and is generally a complex number with a real and imaginary part. ω is the frequency in radians per unit time. If the period is T, then the radial frequency is $2\pi/T$, and the frequency in cycles per unit time is 1/T.

_____

Parseval's Theorem:

The following theorem by Parseval is important in spectral analysis and filtering theory. It states that for two functions $f_1(t)$ and $f_2(t)$ with Fourier Transforms $F_1(\omega)$ and $F_2(\omega)$ we may perform the following manipulation.

$$\int\limits_{-\infty}^{\infty} f_1(t) f_2(t) dt = \int\limits_{-\infty}^{\infty} f_1(t) \left[\frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F_2(\omega) e^{i\omega t} d\omega \right] dt$$

$$\text{where} \ \ f_2(t) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F_2(\omega) e^{i\omega t} d\omega$$

$$= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F_2(\omega) \int\limits_{-\infty}^{\infty} f_1(t) e^{i\omega t} dt \, d\omega$$

$$= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F_2(\omega) F_1(-\omega) d\omega$$

So that we may show,

$$\int_{-\infty}^{\infty} f_1(t) f_2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_2(\omega) F_1^*(\omega) d(\omega) \tag{6.30}$$

Here the asterisk indicates a complex conjugate. Since the original time series are real, we must have $F(-\omega) = F^*(\omega)$. In the special case where $f_2(t) = f_1(t) = f(t)$, we get, since $f(t)$ is real and $F(\omega)$ Hermitian

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{\pi} \int_0^{\infty} |F(\omega)|^2 d\omega \tag{6.31}$$

Thus the square of the time series integrated over time is equal to the square (inner product) of the Fourier transform integrated over frequency. This just shows that the integrated variance in time is equal to the power spectrum integrated over frequency. I'll argue later that we should plot the power spectrum so that the area under the curve represents variance.

_____

Fourier Transform with Discrete Frequency and Time.

In real applications we generally work with a finite sample of time with data given at discrete intervals of time, Δt, and represent the Fourier transform with a discrete set of frequencies determined to be harmonics of the chunk length of the FFT.

Derivation of the Lag Correlation method of computing power spectra:

The continuous form of the definition of the lag correlation function is,

$$r(\tau) = \int_{-\infty}^{\infty} f_1(t) f_1(t + \tau) dt \tag{6.32}$$

We can use Parseval's theorem to write,

$$\int_{-\infty}^{\infty} f_1(t) f_2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_2(\omega) F_1^*(\omega) d\omega \tag{6.33}$$

where

$$f_{1,2}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{1,2}(\omega) e^{i\omega t} d\omega \tag{6.34}$$

In addition we will need the time-shifting theorem,

$$f(t \pm a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega(t \pm a)} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{\pm i\omega a} e^{i\omega t} d\omega \qquad (6.35)$$

Now we may define

$$\int_{-\infty}^{\infty} f_1(t + \tau) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_2(\omega) e^{i\omega\tau} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(\omega) e^{i\omega\tau} e^{i\omega t} d\omega \qquad (6.36)$$

Substituting this (6.36) back into the definition of the lag correlation(6.32) and using Parseval's Theorem we obtain the result we want.

$$r_{11}(\tau) = \int_{-\infty}^{\infty} f_1(t) f_1(t + \tau) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(\omega) F_1^*(\omega) e^{i\omega\tau} d\omega$$

$$\qquad (6.37)$$

$$r(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{11}(\omega) e^{i\omega\tau} d\omega$$

and conversely (since the above is just a Fourier Transform)

$$\Phi_{11}(\omega) = \int_{-\infty}^{\infty} r_{11}(\tau) e^{-i\omega\tau} d\tau \qquad (6.38)$$

Thus the auto correlation coefficient, $r_{11}(\tau)$, and the power spectrum, $\Phi_{11}(\omega)$ form a Fourier transform pair.

We can apply these formulas to finite time and frequency ranges, and to the calculation of cospectra and quadrature spectra.

$$r_{12}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{12}(\omega) e^{i\omega\tau} d\omega$$

$$\qquad (6.39)$$

$$\Phi_{12}(\omega) = \int_{-\infty}^{\infty} r_{12}(\tau) e^{-i\omega\tau} d\tau$$

The cross power $\Phi_{12}(\omega)$ is generally complex and can be divided into its real (cospectra) and imaginary (quadrature spectra) parts.

Example: What is the power spectrum of Red Noise?

The autocorrelation function for red noise is

$$r(\tau) = \exp\left\{\frac{-\tau}{T}\right\} \tag{6.40}$$

So that the power spectrum

$$\Phi(\omega) = \int\limits_{-\infty}^{\infty} \exp\left[\frac{-\tau}{T}\right] e^{-i\omega\tau}\, d\tau$$

$$= \frac{-1}{\frac{1}{T} + i\omega} \exp\left\{-\tau\left(\frac{1}{T} + i\omega\right)\right\}\ \Big|_{-\infty}^{\infty} \tag{6.40}$$

This expression has an irreducible singularity at $\tau = -\infty$. Contour integration to evaluate the integral yields

$$\Phi(\omega) = \frac{\frac{2}{T}}{\frac{1}{T^2} + \omega^2} = \frac{2T}{1 + \omega^2 T^2} \tag{6.41}$$

This function, the power spectrum of red noise, has its largest value at $\omega = 0$ and falls off rapidly with increasing ω, for small ω it falls off more rapidly with large T, as is illustrated in the following figure.
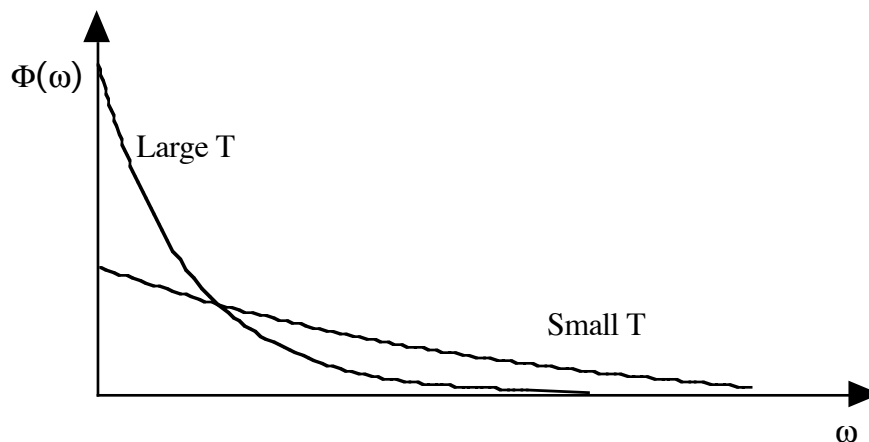
Figure 6.8 Schematic of a red noise power spectrum with a large e-folding time for autocorrelation T, and one with a small e-folding time for autocorrelation.

<u>Example 2</u>: Power spectrum of white noise

The autocorrelation of white noise is a delta function of the lag.
$$r(\tau) = \delta(\tau)$$

The power spectrum is therefore,

$$\Phi(\omega) = \int_{-\infty}^{\infty} \delta(\tau) e^{-i\omega\tau} d\tau$$
$$= 1$$

(6.42)

White noise has an equal amount of variance contained in all frequencies.

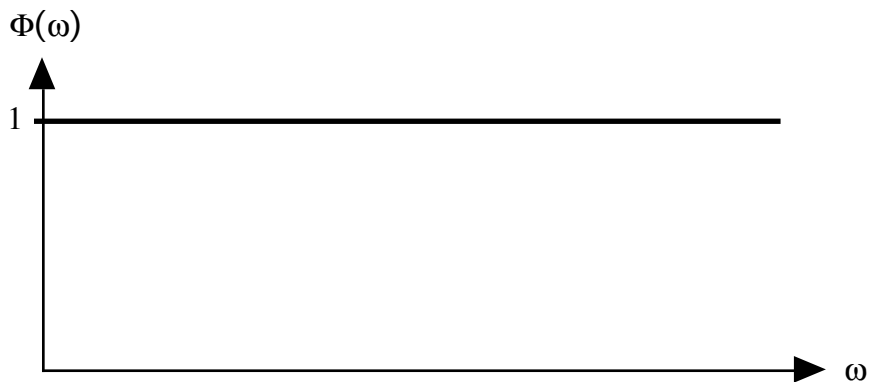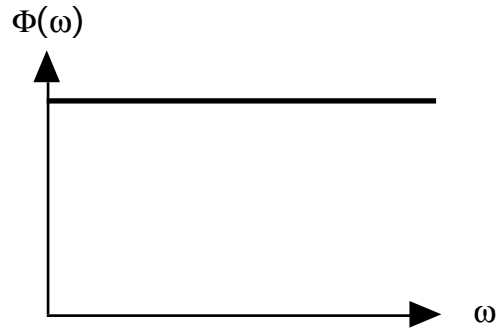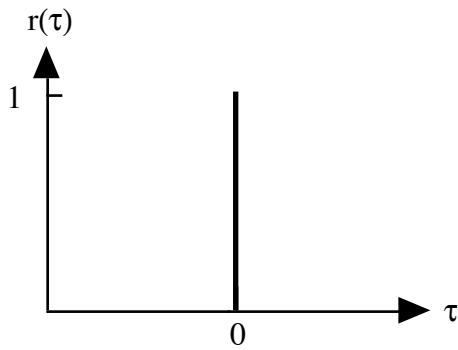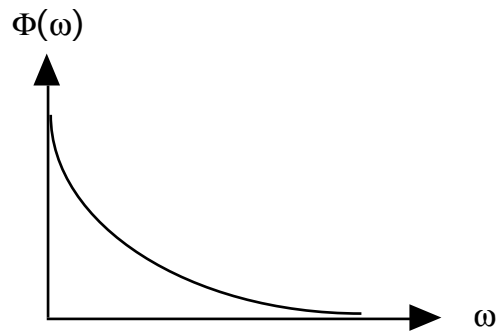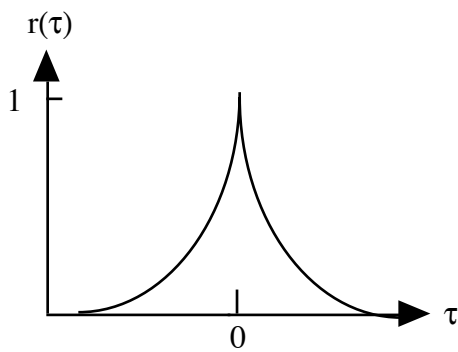*Figure 6.9 Schematic of a white noise power spectrum with zero autocorrelation time.*

Some Examples of lag correlation - Power spectrum pairs

a)   white noise $r(\tau) = \delta(\tau)$

r(τ)   Φ(ω)

1

0   τ   ω

b)   red noise $r = \exp(-\tau/T)$

r(τ)   Φ(ω)

1

0   τ   ω

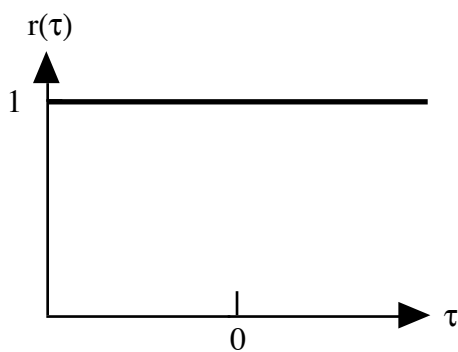c)   linear trend $r(\tau) = 1$ $\Phi = \delta(\omega)$
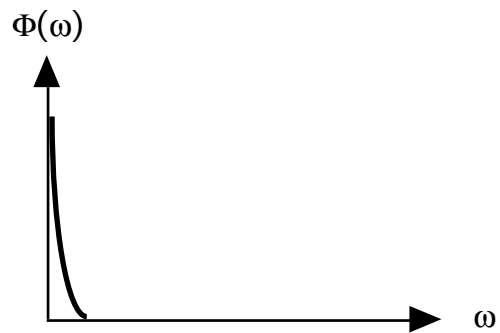
r(τ)   Φ(ω)
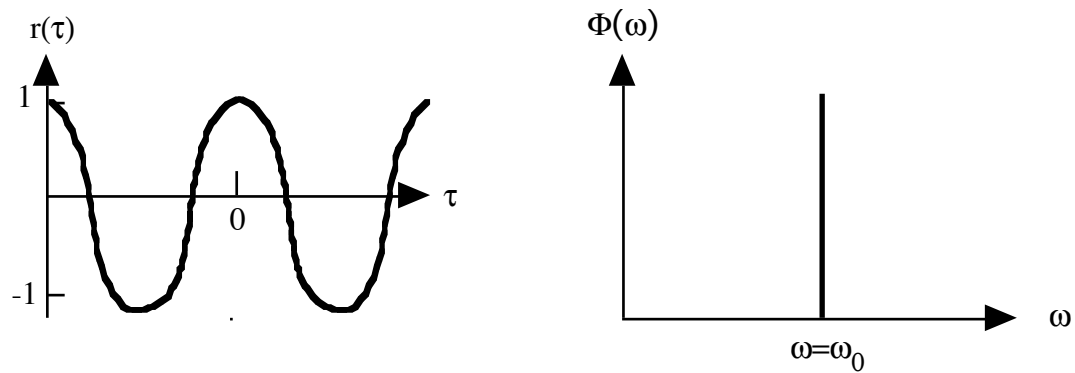
1

0   τ   ω

d)   pure periodicity

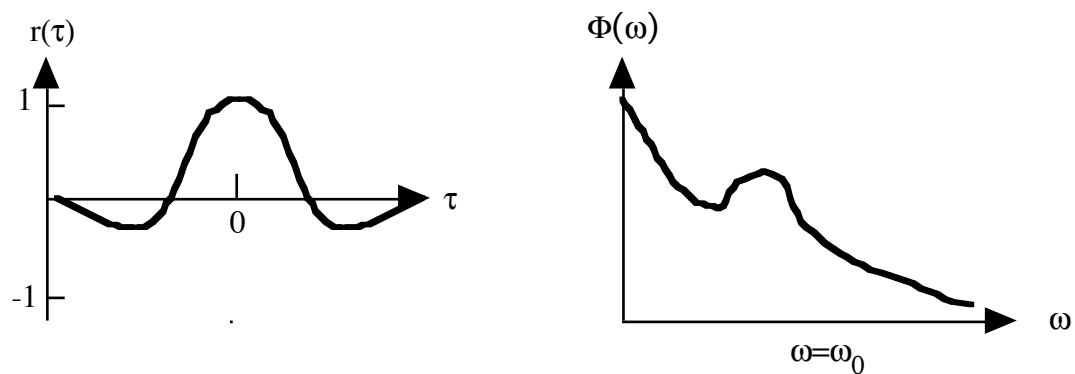e)   quasi-periodicity with trend or other low-frequency noise.

Fig. 6.10 *Examples of autocorrelation function – power spectrum pairs.*

6.2.4 Plotting Power Spectra

Two basic schemes are in common use for plotting power spectral densities as a function of frequency; the linear scale and the logarithmic scale. These both have the property that the area under the curve is proportional to the total variance. In many applications with large ranges of frequency and variance, you will see log scales on both the power and the frequency axes. You have to be careful with these. If you use a logarithmic power scale and a linear frequency scale, then the weak variances look more important than they are. If you use a logarithmic frequency scale and linear power scale, then the high frequencies are much more important than they look. If you use a log-log scale, then both illusions occur.

Linear Scale:

In the linear scale power spectral density is plotted against frequency and the total variance is then simply the area under the curve.
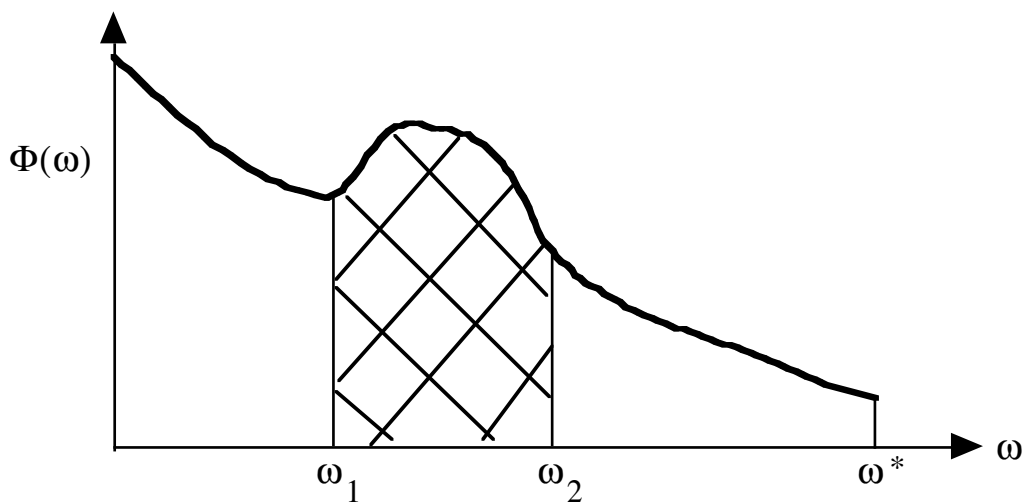

Fig. 6.11 Schematic of a continuous power versus frequency spectrum for which the power or variance between two frequencies is proportional to the integral under the curve..

$$\text{variance} = \int_{\omega_1}^{\omega_2} \Phi(\omega) \, d\omega$$

Logarithmic scale:

When the frequency range is very large we may wish to plot the power spectral density times the frequency versus the natural logarithm of the frequency. In this case the area under the curve is again equal to the total variance, but the low frequency end of the spectrum has been stretched out and the high frequency end contracted. This rescaling of the abscissa implies that the relative height of the power spectral density as it appears on

the plot will also change. This will tend to make small peaks in a generally red spectrum appear more significant, as the low frequencies are stretched over a larger horizontal extent in the graph.
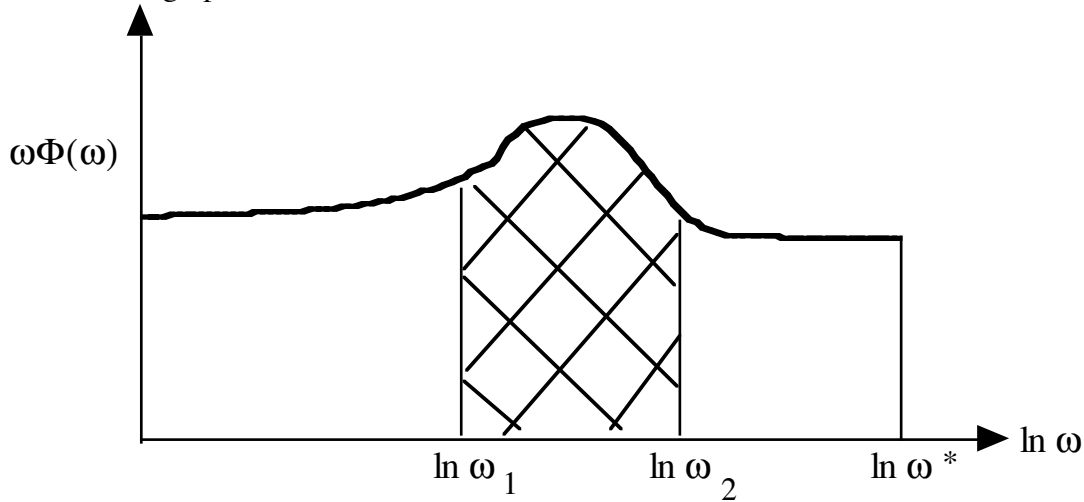

Fig. 6.12 *Schematic of a continuous frequency times power versus log of frequency spectrum for which the power or variance between two frequencies is proportional to the integral under the curve.*

In this case we see that the effect of the rescaling of the ordinate and abscissa are such that the area under the curve between two frequencies is still equal to the variance contained in that interval, so each representation is a proper representation of power spectral density.

$$\int_{\omega_1}^{\omega_2} \Phi(\omega)\, d\omega = \int_{\ln\omega_1}^{\ln\omega_2} \omega \Phi(\omega)\, d\ln\omega \qquad (6.43)$$

6.2.5 Data windows and Window Carpentry

In the analytic case we presume an infinite domain so that the true spectrum can be calculated exactly, provided the analytic function satisfies certain conditions:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega\tau}\, dt \qquad \Phi(\omega) = F(\omega)F^*(\omega) \qquad (6.44)$$

In real cases, however, where we cannot observe *f(t)* on the interval $-\infty < t < \infty$, we are looking at the function through a "window." The window can be represented by a function *w(t)*. The window function for the ideal analytic case where we know the

function on the interval ($-\infty$, ∞), and the more realistic case where we know the function only on the interval $(0,T)$, are shown below.

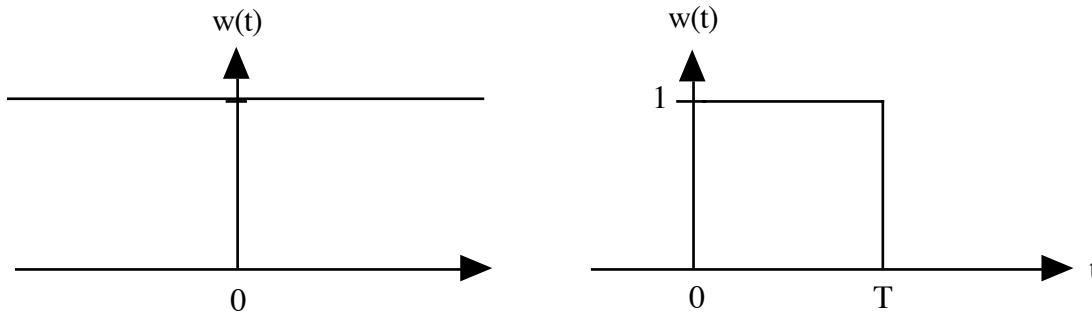a.) analytic case infinite data   b.) sample of length T

Fig. 6.13 *Schematic comparing the window function of an infinite time series, with that of a finite time series..*

a) **Infinite Data**: In the case of infinite data, the window function is one for all t and we get an exact representation of $f(t)$,

$$\omega(t)\,f(t)=1\cdot f(t)\Leftrightarrow F(\omega) \qquad (6.45)$$

and we get the true spectrum.

b) **Finite Sample of Data:** In the case where we view the time series $f(t)$ through a finite window, we do not get an exact measure of the spectrum, but something that is affected by the geometry of the window we are looking through.

$$F(\omega)=\int\limits_{-\infty}^{\infty} f(t)\,w(t)e^{-i\omega t}\,dt \qquad (6.46)$$

$$=\int\limits_{-\infty}^{\infty} f(t)\Pi\!\left(\frac{t}{T}\right)e^{-i\omega t}\,dt$$

where
$$\Pi(x)=\begin{cases} 1 & 0\le x\le 1 \\ 0 & \text{elsewhere} \end{cases}$$

by the convolution theorem

$$\int\limits_{-\infty}^{\infty} f_1(t) f_2(t) e^{-i\omega t} dt = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F_1(\overline{\omega}) F_2(\omega - \overline{\omega}) d\overline{\omega} \qquad (6.47)$$

Recall the formula for a continuous Fourier Transform,

$$\int\limits_{-\infty}^{\infty} f_1(t) e^{-i\omega t} dt = F(\omega) \qquad (6.45)$$

with the window included we get

$$F_?(\omega) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} F(\omega) \frac{\sin\left((\omega - \overline{\omega})\dfrac{T}{2}\right)}{(\omega - \overline{\omega})\dfrac{T}{2}} d\overline{\omega} \neq F(\omega) \qquad (6.46)$$

where we have used the known Fourier transform pair for the Boxcar function, which we will call the *sinch* function.

$$\Pi(t) \Leftrightarrow \frac{\sin\dfrac{\omega}{2}}{\dfrac{\omega}{2}} \qquad (6.47)$$

In this case <u>we do not obtain the true spectrum $F(\omega)$ but a convolution of the true spectrum with a weighting window</u>

$$W(\omega) = \frac{2}{\pi} \int\limits_{-\infty}^{\infty} \delta(\omega - \overline{\omega}) \sin(\omega - \overline{\omega}) \frac{T}{2} \qquad (6.48)$$

This acts as a smoothing on the true spectrum. The degree of smoothing depends on the length of the data window T. The shorter that T is, the stronger the smoothing/spread will be. In addition to the smoothing effect, the side lobes of the frequency window lead to spectral leakage. Therefore we can see that we must carefully design the window through which we view the data prior to spectral analysis, if we want to obtain the best results. The response function for the rectangular window is shown below. Note that while the response does peak strongly at the central frequency, significant negative side lobes are present. This means that our spectral analysis will introduce spurious oscillations at higher and lower frequencies that are out of phase with the actual oscillation.

Boxcar Window Sinch Frequency Response
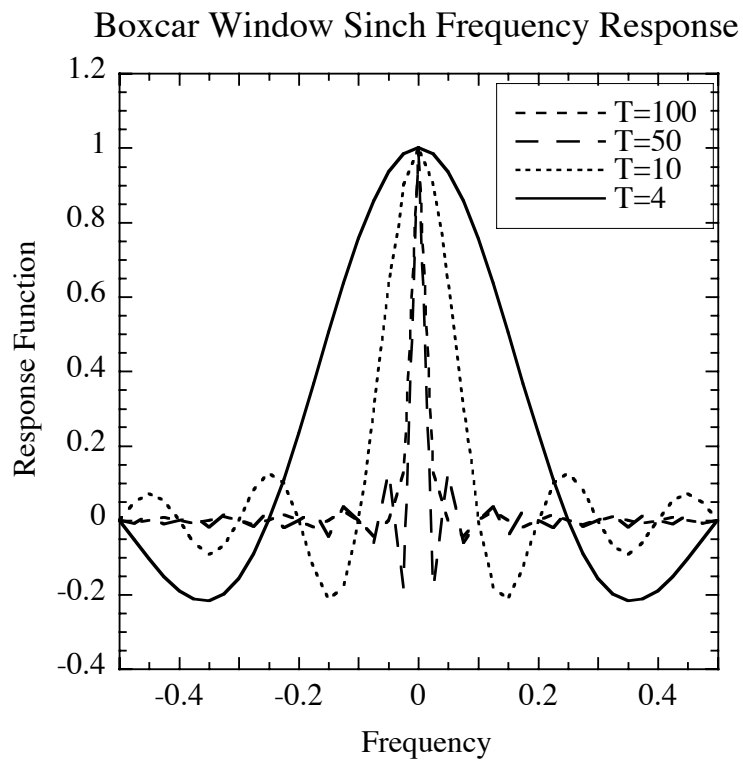

Fig. 6.14 *The sinch function plotted on the Nyquist interval -0.5 < f < 0.5 cycles per time step. Notice how the bandwidth and spread of the response function is narrowed as the total length of the Fourier Transform, T, is made longer. Notice the large negative side lobes and their rather slow decay.*

It is very important to understand the implications of Fig. 6.14.

Square Window

$w(t)$

$W(\omega)$

Tapered Window $w(t)$

$W(\omega)$
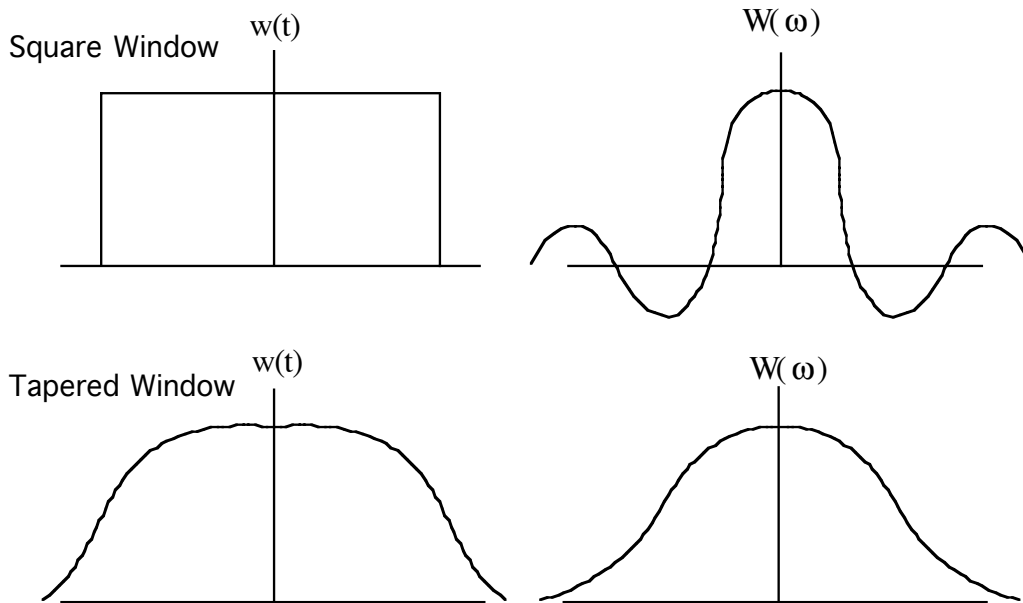
Fig. 6.15 *Schematic of square window w(t) and spectral response W(ω) and tapered window and its spectral response.*

In general, the ideal window response function, $W(\omega)$, would have a narrow central lobe and insignificant side lobes. We can improve on the naive rectangular window function and the rather unsatisfactory *sinch* function window response through modifications of the window function $w(t)$.

$W(\omega - \omega_0)$

"Ideal" window response

$(\omega - \omega_0)$

0

Fig. 6.16 *The ideal window response is a narrow peak with a value of 1.0, and no side lobes.*

The influence of the length of the record, *T*, on the spectra can be seen in the window response function for a rectangular window. The choice of the proper *T* for the problem in question is affected by several considerations.
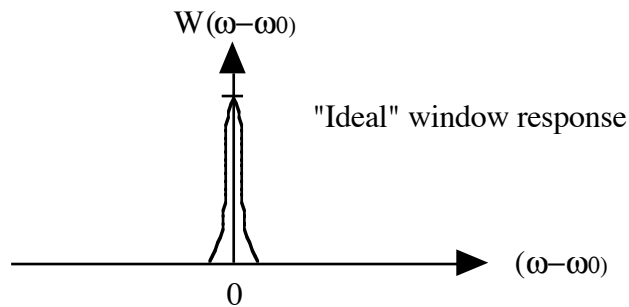
1. Large $T \Rightarrow$ better resolution in frequency
2. Small $T \Rightarrow$ better stability if we use many realizations: Bandwidth $1/T$
3. Choose *T* to encompass some event - seismology
                                                          - intermittency in turbulence

The particular choice made depends on the nature of the phenomena of interest and the frequency resolution required. Often our options are limited by the availability of data.

Window Carpentry:

The characteristics that we want of frequency window are:

1. A high concentration in the central lobe; requires a large *T*.

2. Small or insignificant side lobes; requires a smooth time window without sharp corners.

A rectangular or Boxcar window leaves the time series undistorted, but can seriously distort the frequency spectrum.   A tapered window distorts the time series but may yield a more representative frequency spectrum.
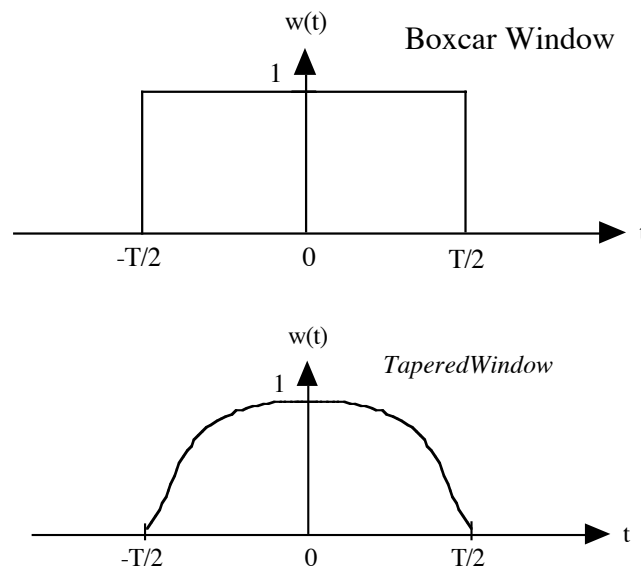

Fig. 6.17 *A rectangular or Boxcar window is rectangular, a tapered window has rounded edges.*

Some windows (commonly used):

Boxcar _(square or rectangular)_; also what you use if you are naive and don't apply an explicit window function.

$$W(t) = \begin{cases} 1 & 0 \leq |t| \leq T \\ 0 & |t| > T \end{cases} \tag{6.49}$$

has the Fourier Transform that is the _sinch_ function,

$$W(\omega) = \frac{2\sin\left(\frac{\omega T}{2}\right)}{\frac{\omega T}{2}} = 2\operatorname{sinc}\left(\frac{\omega T}{2\pi}\right) \tag{6.50}$$

     The rectangular window is a commonly used window; indeed this is the window people use when they ignore the fact that they are using a window, or when they are ignorant of it. The rectangular window has a serious side-lobe problem. Frequencies that are outside the range of frequencies actually resolved can have too strong an influence on the power spectra at the frequencies resolved. This is another form of aliasing.

Hanning: (Cosine Bell; Tukey Window)

$$w(t) = \begin{cases} \dfrac{1}{2} \cdot \left(1 - \cos\dfrac{2\pi t}{T}\right) & ; 0 \leq t \leq T \\ 0 & ; t > T \ or \ t < 0 \end{cases} \tag{6.51}$$

$$= \frac{2}{T}\cos^2\frac{\pi t}{T} \ ; 0 \leq t \leq T$$

The cosine bell window is perhaps the most frequently used window in meteorological applications. It is automatically applied in the Matlab routine "spectrum" or "spectrum.welch".

$$W(\omega) = \underbrace{\operatorname{sinc}\ \left(\frac{\omega T}{\pi}\right)}_{\substack{\text{Same as} \\ \text{Bartlett}}} + \underbrace{\frac{1}{2}\left[\operatorname{sinc}\ \left(\frac{\omega T}{\pi} + 1\right) \ + \operatorname{sinc}\ \left(\frac{\omega T}{\pi} - 1\right)\right]}_{\text{CANCELS 1st TWO SIDE LOBES}} \tag{6.52}$$

We can see that the first part of the response function for the cosine bell window is a _sinch_ function exactly like that of the rectangular window. In addition, however, we have two additional _sinch_ functions of reduced amplitude that maximize one unit on either side

of the central lobe. The effect of these is to partially cancel the negative side lobes of the rectangular response function and to slightly broaden the central lobe. We like the fact that the side lobes are now smaller, but the broadened central lobe means that the spectrum will be slightly smoothed compared to a rectangular window. This smoothing is not a disaster in most applications, since we often end up doing some smoothing anyway, and the smoothing effect of the cosine bell window allows us to claim a small increase in the number of degrees of freedom per spectral estimate.

Hamming: Slight modification of Hanning

$$\omega(t) = \begin{cases} \left(0.54 + 0.46 \cos \pi \dfrac{t}{T} \right) & 0 \le |t| \le T \\ 0 & |t| > T \end{cases} \tag{6.53}$$

$$W(\omega) = 1.08 \, \text{sinc} \, \dfrac{\omega T}{\pi} + 0.46 \left[\text{sinc} \left(\dfrac{\omega T}{\pi} + 1 \right) + \text{sinc} \left(\dfrac{\omega T}{\pi} - 1 \right) \right] \tag{6.54}$$

The Hamming window provides a slightly more optimal reduction in the side lobes and slightly more smoothing of the central lobe.

Parzen: Power window

$$w(t) = \begin{cases} \left[1 - \dfrac{|t|^m}{T} \right] & 0 \le |t| \le T \\ 0 & |t| > T \end{cases} \tag{6.55}$$

Special (common) form

$$w(t) = \begin{cases} \left[1 - 6 \left(\dfrac{t}{T} \right)^2 + 6 \left(\dfrac{|t|}{T} \right)^3 \right] & 0 \le |t| \le \dfrac{T}{2} \\ 2 \left[1 - \dfrac{|t|}{T} \right]^3 & \dfrac{T}{2} \le |t| \le T \\ 0 & |t| > T \end{cases} \tag{6.56}$$

The response function for the common Parzen window is proportional to the fourth power of the central frequency divided by the frequency. Thus there are no negative side lobes. The response function is rather broad.

$$W(t) = \frac{P}{\ } \left(\frac{\sin \omega}{\omega} \right)^4 \qquad (6.57)$$

Using the Fast Fourier Transform Method: WOSA

One of the most common methods of spectral analysis used is the Fast Fourier Transform method. In this method a direct Fourier Transform is made of the data using an efficient algorithm that makes use of the fact that the length of the time series has been chosen to be an integer power of two $M_{ch}=2^n$. Mixed radix FFT's are also available for which $M_{ch} = 2^n \, 3^m \, 5^j$. In applying these methods the total time series of length $N\Delta t$ is broken up into a series of smaller chunks of length M_{ch}. The spectra and cross-spectra for these smaller chunks can be averaged into a grand spectrum that has some degree of statistical reliability. This is called Welch's Overlapping Segment Analysis, or WOSA. Because a tapered window is normally applied, the data near the ends of the chunks of length M_{ch} are not weighted as heavily as the data in the center of the chunks. So that all the data are weighted equally, it is desirable to have the chunks overlap somewhat. In the Matlab spectral analysis routines, this is handled by the 'noverlap' parameter. If you overlap the windows by exactly one half of the chunk length, and you are using the Hanning window, then each data point gets exactly the same weight in the resulting averaged spectrum, except the first and last $M_{ch}/2$ points in the total data set.

6.2.6 Statistical Significance of Spectral Peaks

The statistical significance of peaks in a power spectrum are assessed as in any case by stating the significance level desired, and then stating the null hypothesis and its alternative. The null hypothesis is usually that the time series is not periodic in the region of interest, but simply noise. We thus compare amplitude of a spectral peak to a background value determined by a red noise fit to the spectrum, or at least this is one way to do it.
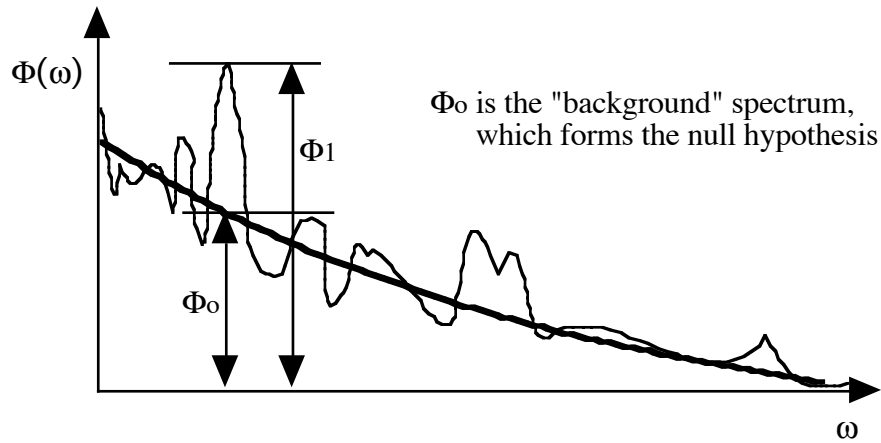
Figure. 6.18 *Schematic of a spectrum, its null hypothesis spectrum and the ratio of the variances at a frequency of interest.*

One way to evaluate the significance of a spectral peak is to compute the ratio Φ/Φ_O and compare this value with a "Chi-Squared" test with the corresponding number of degrees of freedom

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}; \quad v = n-1 \tag{6.58}$$

Number of degrees of freedom, *n*

$$n = \frac{N}{M_{sp}} \bullet f_\omega \tag{6.59}$$

N = total sample size
M_{sp} = number of spectral estimates
f_ω = factor to compensate for smoothing done by the window.

N.B. The factor f_ω is generally between 1 and 1.5, depending on how much smoothing that the window does. We will set it equal to 1 in our work here, which is conservative. For the Hanning Window, f_ω is equal to 1.2.

A more convenient way to apply this test is to use the *F*-test, based on the *F* distribution.

Theorem: If s_1^2 and s_2^2 are the variances of independent random samples of size n_1, and n_2, respectively, taken from two normal populations having the same variance, then

$$F = \frac{s_1^2}{s_2^2} \tag{6.60}$$

is a value of a random variable having the *F* distribution with the parameters $v_1 = n_1-1$ and $v_2 = n_2-1$.

The *F* distribution can be used to determine whether two sample variances are different in a statistical sense. Tables of the *F* distribution are included on page 16. You should not confuse this F with F(z), the cumulative distribution of the Normal distribution.

In assigning a confidence level and interpreting statistical tests of significance one must be concerned very much with the distinction between *a priori* and *a posteriori* statistics.

a priori

If we have stated in advance that we expect a peak at a particular frequency (and given a good reason beforehand), then we can simply test the significance of the spectral peak above the background using the normal confidence limits set forth for the chi-squared or *F* statistics.

a posteriori

If we have not stated at which frequency we expect the peak, then we must determine the probability that one frequency out of the M_{sp} we have computed should show a significant peak. The usual way to do this would be to take the probability of a type II error (accepting a false hypothesis as true) and multiply this by the number of chances we have given the spectrum to exceed the required level.

Example: Suppose we have a spectral peak which exceeds the background significantly at the 99.9% probability level.

1. If we had predicted this frequency, then we can use the 99.9% probability level and infer that only a 0.1% possibility exists that this spectral peak could have occurred by chance. We have an *a priori* reason for expecting this peak and we can therefore use *a priori* statistics.

2. If we had not predicted the frequency of the peak, then we must test the probability that one frequency out of our sample of M_{sp}/f_ω independent estimates should show a significant peak. M_{sp} is the number of frequencies retained in our spectrum and f_ω is the factor indicating the degree of smoothing by the window. If $M_{sp}/f_\omega = 50$ then our

chance of getting a spectrum with no significant peaks is $(0.999)^{50} = 0.95$ and we have only 95% confidence that this peak at <u>this</u> frequency is significant. If we had started with a 95% confidence limit the significance would be $(0.95)^{50} = 7.7\%$. It is almost guaranteed that at least one frequency would show a 95% peak. See Madden and Julian, 1971, *J. Atmos. Sci.,* p. 702 for a discussion of *a priori* and *a posteriori* confidence limits.

Examples:

1. Suppose we have 100 days of 4-times-daily observations. We have good reason to expect a peak in the 4-6 day period range. What value of M_{sp} should we shoot for and what does this mean about the chunk length we should use?

If we want to distinguish a peak in the 4-6 day range from adjacent frequencies, then we may use the following calculation.

$$4 - 6 \text{ day periods} \implies 0.25 \text{ - } 0.15 \text{ cycles per day} \implies \Delta f \sim 0.1 \text{ cpd}$$
$$f_i \; = \; 0, \; 0.1, \; 0.2, \; 0.3$$
$$\uparrow$$
$$\text{this one}$$

To distinguish 4-day and 6-day periods, we need a bandwidth, Δf , that is about 0.1 cycles per day. To get this bandwidth we need a chunk length that is about 10 days long, since the bandwidth is one over the chunk length (the chunk length M_{ch} is the time interval to which the FFT is applied.) This gives us rather minimal frequency resolution between the zero frequency and the frequency of interest, plus there will be some smoothing of the spectrum, so let's choose a chunk length of 20 days. This is 80 time points, since we have 4 observations per day. If we want the chunk to be a power of two we must choose 64 or 128, or chunks of 16 or 32 days.

If we choose 32we will only have 6 or 8degrees of freedom per spectral estimate, so let's choose 16-day chunks, consisting of 64 4xdaily observations from which we will obtain 32 spectral estimates ($M_{sp}=M_{ch}$ /2). Our degrees of freedom will be about 400/32 = 12.5, forgetting for the moment about the factor associated with the smoothing due to the window.

$$\Delta f = \frac{1}{M_{ch}\Delta t} = \frac{1}{T} \text{ , where T is the total time in the chunk.}$$

$d.o.f \cong N / M_{sp}$, where N is the total number of data points and M_{sp} is the number of
spectral estimates. We could multiply this by $f_\omega = 1.2$ to take into account the

spectral smoothing by the Hanning window if we get in a pinch, and we have used the Hanning window.

The frequencies resolved go from 1/T to 2/Δt, as n/T, n=0,1, 2, 3, . . M_{sp}. Half of these frequencies are greater than 4/Δt, which is our case is 1 cycle per day. Thus, more than half of the frequencies we are resolving are higher than our level of interest. It would have been better to start with daily observations. In general one should sample only frequently enough to resolve the wave of interest, which requires only 4-6 points. So in our case daily data are adequate. Fig. 6.19 shows power spectra with red noise plus a periodicity at 4 days period, frequency 0.25 cycles per day, computed with 4xdaily and daily data. It is not necessary to use data taken more frequently than once a day.
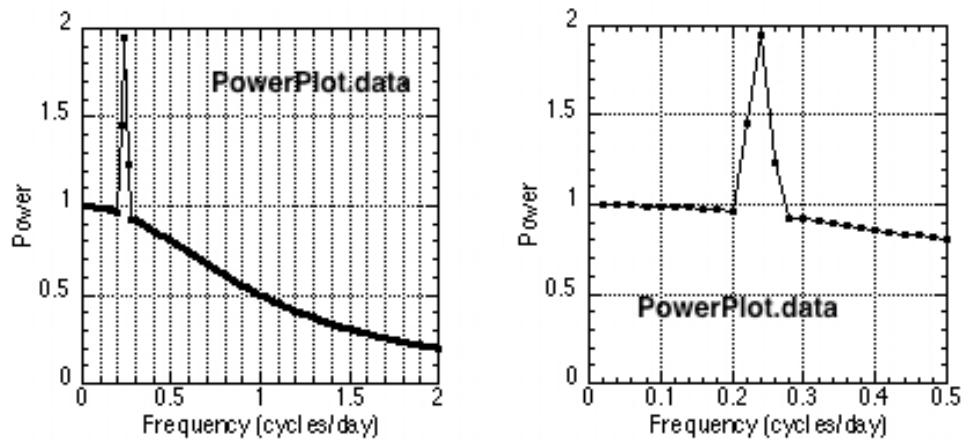

Figure 6.19 Left: Example of power spectrum in which the sampling rate is 4xdaily, so that the Nyquist frequency is two cycles per day. Note that half of the frequencies computed are between 1 cycle per day and 2 cycles per day. If the interest is around 0.25 cycles per day (period of 4 days), then many frequencies are being calculated for nothing. Right: If daily data were used the Nyquist period would be 0.5 cycles per day, and the period of interest appears near the center of the Nyquist interval.

When we sample more frequently than necessary to resolve the frequencies of interest, a lot of effort and information goes into frequencies that we aren't interested in (Fig. 6.19), plus the problems with *a priori* versus *a posteriori* confidence are worsened..

What if we had not pre-specified the desired frequency?

A posteriori possibility of type II error. OK, if we have not guessed the frequency, we have given the spectrum many chances to pass the significance level. The number of independent frequencies is the number of frequencies computed, divided b the smoothing factor. The number of frequencies in Fourier spectral analysis, M_{sp}, is half the blocklength M_{ch}.

a.) $4 \times$ daily: number of independent frequencies, and hence the number of chances we have given the spectrum to exceed our statistical criterion is

$$\frac{M_{sp}}{f_\omega} = \frac{32}{f_\omega}$$

b.) $1 \times$ daily

$$\frac{M_{sp}}{f_\omega} = \frac{8}{f_\omega}$$

smaller multiplication of probability to type II error.

a)
$$0.01 error >= (0.99)^{\frac{32}{f_\omega}} \approx 72\% \text{ confidence}$$

b)
$$0.01 error >= (0.99)^{\frac{8}{f_\omega}} \approx 92\% \text{ confidence}$$

Here it is plain to see that by including unnecessary resolution in the input data we have exaggerated the difference between *a priori* and *a posteriori* confidence estimates. It is a good idea to not use more frequent sampling than is needed to measure the phenomena of interest.

Example: Suppose we wish to examine climatic fluctuations in the 10-1000 year range. How long a time series is required and what should be the spacing of the data values?

If we want a minimum number of about 10 degrees of freedom before we take into account the window, then we should have a data record which is at least 10 times as long as the lowest frequency we wish to resolve (i.e., 10 realizations of the signal).

Then $T = 10 \times 1000$ years = 10,000 years

In order to resolve a sinusoid of 10 years duration we need at least 4 data points per cycle so $\Delta t = 1$ or 2 years. 2 is adequate, 1 is plenty.

One might wish to analyze the time series separately for the long and short period ends of this long time series.

a.) High frequency analysis = periods < 50 years

$$2M_{sp} = 200 \, day \, blocks$$

$$\Delta f = \frac{1}{200} = 0.005 \, cpy$$

$$f_i = \frac{i}{200 \, yr} = 0, 0.005, 0.01, 0.015, 0.02, ...0.5 \, cpy$$

$$periods = \infty, 200, 100, 67, 50, ...2 \, years$$

$$\text{d.o.f.} = \frac{N}{M_{sp}} \cdot f_\omega = \frac{10000}{100} \cdot f_\omega = 100 \, f_\omega \quad \text{d.o.f.} \, (great)$$

If we are using a power of 2 FFT, then we would choose a blocklength of M_{ch}=256 and we would get at least 10,000/128 = 78 degrees of freedom (still great).

b. Low frequency analysis

$$2M_{sp} = 2000 \, day \, blocks$$

$$\Delta f = \frac{1}{2000} = 0.0005 \, cpy$$

$$f_i = \frac{i}{2000 \, yr} = 0, 0.0005, 0.001, 0.0015, 0.002, ...0.5 \, cpy$$

$$periods = \infty, 2000, 1000, 670, 500, ...2 \, years$$

If we are using a power of 2 FFT, then we would choose a block length of M_{ch} =2048 and we would get at least 10,000/1024 = 9.8 degrees of freedom (not good). Also, the frequency of interest is very close to the low end. We should really choose the block length of 4000 years.

6.2.7 Experimental Red Noise Spectra

Gilman et al. (1963) have provided a formula for calculating the shape of experimental red noise spectra to be used in constructing null hypotheses for significance testing. Using the lag correlation function for a first-order linear Markov process,

$$r(k\Delta t) = \rho^k; \quad \text{where} \quad r(\Delta t) = \rho \tag{6.61}$$

ρ is the one-lag autocorrelation, or the autocorrelation between time points separated by one unit of time, Δt. Using the finite Fourier transform of this lag correlation function

expressed with a maximum number of lags, M_{sp}, they present an approximate shape function for the spectrum associated with this autocorrelation function.

$$P\left(h,\rho,M_{sp}\right)=\frac{1-\rho^2}{1-2\rho\,\cos\left(\dfrac{h\pi}{M_{sp}}\right)+\rho^2} \tag{6.62}$$

In (6.62) h represents the frequency, but is given in peculiar units. It varies from zero to M_{sp}, corresponding to the M_{sp} +1 frequencies provided by the Fourier transform of a lag-correlation function with a maximum lag of M_{sp} or an FFT based on a block length of $2M_{sp} = M_{ch}$. To fit the shape function (6.62) to a real spectrum we would normally obtain the parameter ρ from the original time series as the average of the one-lag autocorrelation and the square root of the two-lag autocorrelation. We then multiply the shape for that value of ρ and the value of M_{sp} by a factor that will make the variance equal to that of the time series in question. One simple way to do this is to match the total variance. The total variance is the sum over all non-zero frequencies. So sum the power in the observed spectrum and the power in the idealized red noise spectrum. Then multiply the idealized red noise spectrum by this ratio, so that the red noise spectrum has the same total variance as the observed spectrum..

In section 6.2.3 we derived the power spectrum for red noise using continuous, infinite lag-correlation functions and time series. The functional form obtained is not identical to the discrete, finite form (6.62). The continuous red noise spectrum is,

$$\Phi(\omega) = \frac{2T}{1+\omega^2 T^2} \tag{6.63}$$

We can translate this into an analogy with (6.62) by converting from e-folding time, T, to ρ, and from frequency ω to frequency index h. First note the relationship between T and ρ.

$$T = \frac{-\Delta t}{\ln(\rho)} \tag{6.64}$$

Substituting (6.64) into (6.63), we obtain,

$$\Phi(\omega) = \frac{-2\Delta t \ln(\rho)}{\ln(\rho)^2 + \omega^2 \Delta t^2} \tag{6.65}$$

Given in units of the parameter h, the frequency is,

$$\omega = \frac{2\pi h}{2M_{sp}\,\Delta t} = \frac{\pi h}{M_{sp}\,\Delta t} \qquad (6.66)$$

So that (6.65) becomes,

$$\Phi(\omega) = \frac{-2\Delta t \ln(\rho)}{\ln(\rho)^2 + \left(\dfrac{\pi h}{M_{sp}}\right)^2} \qquad (6.67)$$

Now we must remember that the theoretical spectrum applies to a hypothetical infinite data set that would allow all possible lags and all possible frequencies to be represented, whereas the sample red–noise spectrum is derived for a sample of finite length. This also means that the norms for the two spectra cannot be the same, since (6.63) is meant to be integrated over all frequencies $0 \rightarrow \infty$, while (6.62) is integrated only over a finite range of frequencies. If we integrate (6.63) over the interval from $-\infty$ to ∞ we get a norm of 2π. If we integrate (6.67) over the range of 0 to $\pi/\Delta t$, we get a norm of,

$$\text{Norm} = -\,2 \tan^{-1}\!\left(\frac{\pi}{\ln \rho}\right) \qquad (6.68)$$

which makes (6.67) take the form,

$$\Phi'(\omega) = \frac{\Delta t \ln(\rho)}{\tan^{-1}\!\left(\dfrac{\pi}{\ln \rho}\right)} \left(\frac{1}{\ln(\rho)^2 + \left(\dfrac{\pi h}{M_{sp}}\right)^2} \right) \qquad (6.69)$$

(6.62) is normalized such that all frequencies have unit variance if the spectrum is white. Equation (6.69) is normalized such that the integral under the curve is unity. To bring the two normalizations together we can multiply (6.69) by the total range of frequency, which is the Nyquist frequency, $\pi/\Delta t$. Multiplying (6.69) by this factor will make all the frequencies have unit spectrum when $\rho = 0$, in agreement with the normalization of (6.62). The final result we get for a properly normalized theoretical spectrum is,

$$\Phi'(\omega) = \frac{\pi \ln(\rho)}{\tan^{-1}\left(\dfrac{\pi}{\ln \rho}\right)}\left(\frac{1}{\ln(\rho)^2 + \left(\dfrac{\pi h}{M_{sp}}\right)^2}\right)$$ (6.70)

In the following figure we plot the red noise spectra according to (6.62) and (6.70) for one-lag autocorrelations of 0.1, 0.5, and 0.7. The unit of frequency is radians per time step. You will note that the theoretical value is a little more steep than that of Gilman et al. (1963). This is probably associated with the smoothing and aliasing produced by the finite window size used to estimate the spectra. In practice, you should always use (6.62), since that is what you expect to see through a finite window, but it is interesting to know that the experimental value and the theoretical value are not too different.
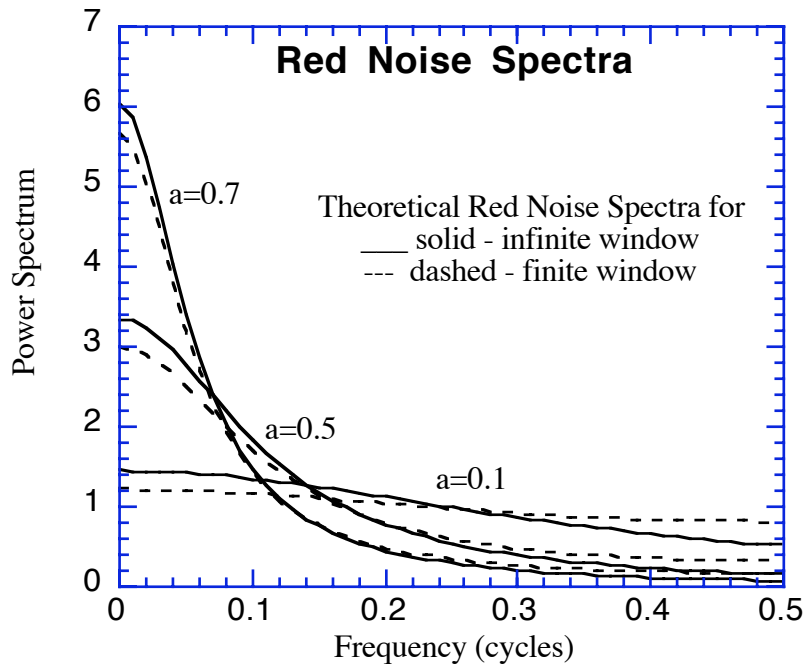

Fig. 6.20 *Theoretical red noise spectra for various one-lag autocorrelations, using the theoretical red noise spectrum and the one derived by Gilman et al for a finite window.*

Hard Lessons with Small Samples:

Suppose we take a hypothetical red noise time series constructed from a first-order linear Markov process.

$$x_t = a\,x_{t-\Delta t} + \left(1-a^2\right)^{1/2}\varepsilon(t)$$ (6.71)

Here $\varepsilon(t)$ is a standardized white noise process.

       If we take one sample of such a time series, say 256 daily values, and compute a spectrum, we can get a variety of things (See Fig. 6.21). Only after averaging over many such realizations will we begin to converge on the red noise spectra described above. To demonstrate this we show several examples of spectra obtained from 256 day segments of the time series (6.16), calculated with $a = 0.5$. In the first example (Fig. 6.21a) there appears to be a dominant frequency around 0.04 cpd, or a period of 25 days. In the next example (Fig. 6.12b) spikes appear in the spectrum again, but at different frequencies. After averaging together 10 such spectra (Fig. 6.21e), the averaged spectrum still shows considerable spikiness. Even with 40 realizations (about 80 degrees of freedom) the spectrum is jagged. The sampling uncertainty is only reduced as the square root of the sample size.
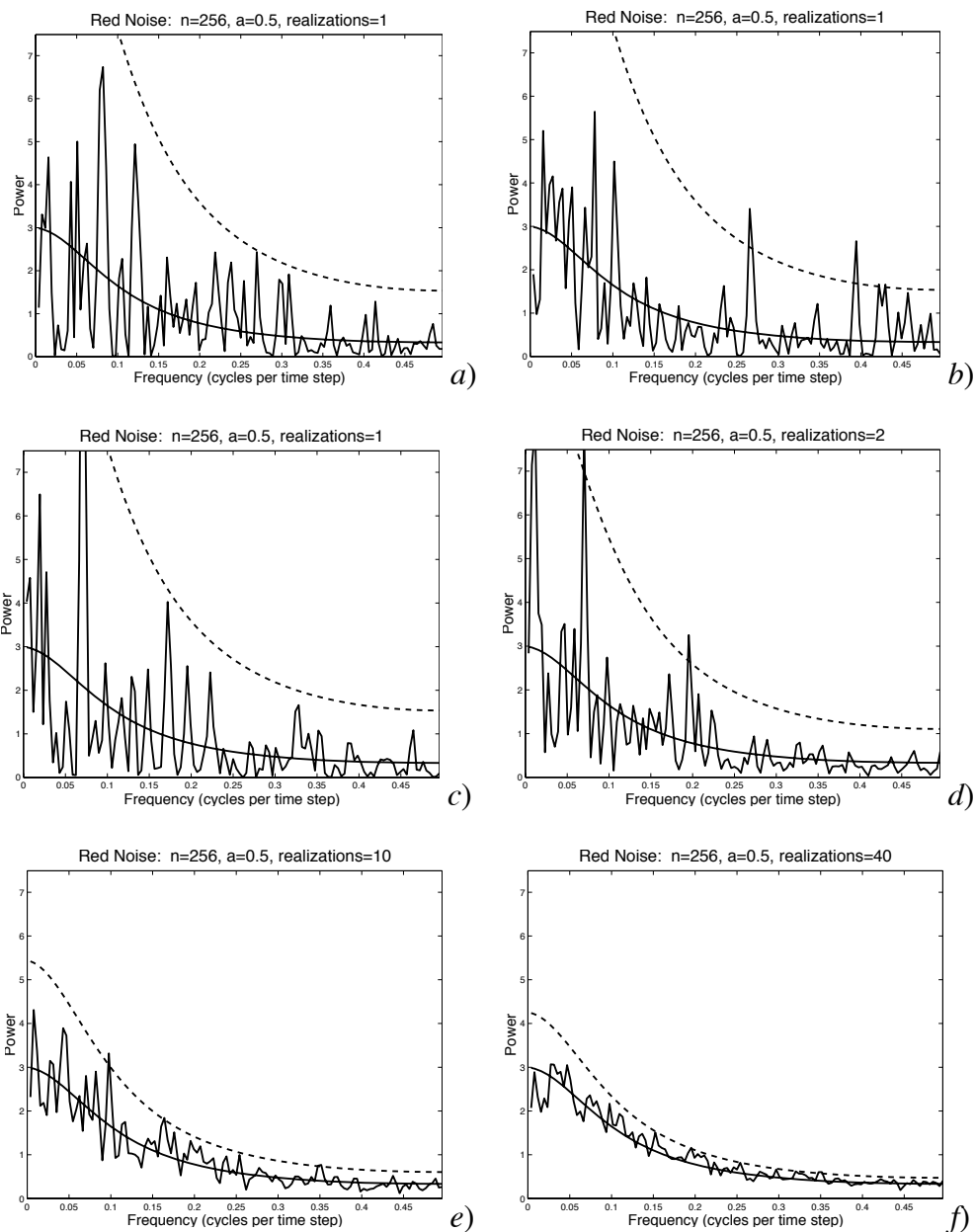
Fig. 6.21: *a) and b) are two single realizations of spectra based on 256 times steps drawn from a red noise process with a one-lag autocorrelation of 0.5. c) is another single realization, d) is the average of c) and another independent realization. e) and f) are averages over 10 and 40 independent realizations, respectively. Notice the enormous peak in c) that persists into d), but then disappears after more realizations are added. Huge peaks can appear by chance in short records of data that are in reality only autocorrelated noise. The smooth solid curve is the Gilman et al estimate of the true spectrum of red noise observed through a Rectangular window. The dashed curve is the 99% confidence level needed to reject a null hypothesis of red noise on an a priori basis.*