

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.

$$\begin{aligned}\int_{-\infty}^{\infty} f_1(t)f_2(t)dt &= \int_{-\infty}^{\infty} f_1(t) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} F_2(\omega)e^{i\omega t}d\omega \right] dt \\ \text{where } f_2(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F_2(\omega)e^{i\omega t}d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F_2(\omega) \int_{-\infty}^{\infty} f_1(t)e^{i\omega t}dt d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} F_2(\omega) F_1(-\omega) d\omega\end{aligned}$$

So that we may show,

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

For which, in the special case where $f_2(t) = f_1(t) = f(t)$, we get, since $f(t)$ is real and $F(\omega)$ Hermitian

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

Thus the square of the time series integrated over time is equal to the square (inner product) of the Fourier transform integrated over frequency.

Derivation of the Lag Correlation method of computing power spectra:

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

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

We can use Parseval's theorem to write,

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

where

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

In addition we will need the time-shifting theorem,

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

Now we may define

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

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.

$$\begin{aligned} r_{11}(\tau) &= \int_{-\infty}^{\infty} f_1(t) f_1(t + \tau) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(\omega) F_1^*(\omega) e^{i\omega\tau} d\omega \\ r(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{11}(\omega) e^{i\omega\tau} d\omega \end{aligned} \quad (6.37)$$

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

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

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.

$$\begin{aligned} r_{12}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{12}(\omega) e^{i\omega\tau} d\omega \\ \Phi_{12}(\omega) &= \int_{-\infty}^{\infty} r_{12}(\tau) e^{-i\omega\tau} d\tau \end{aligned} \quad (6.39)$$

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

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

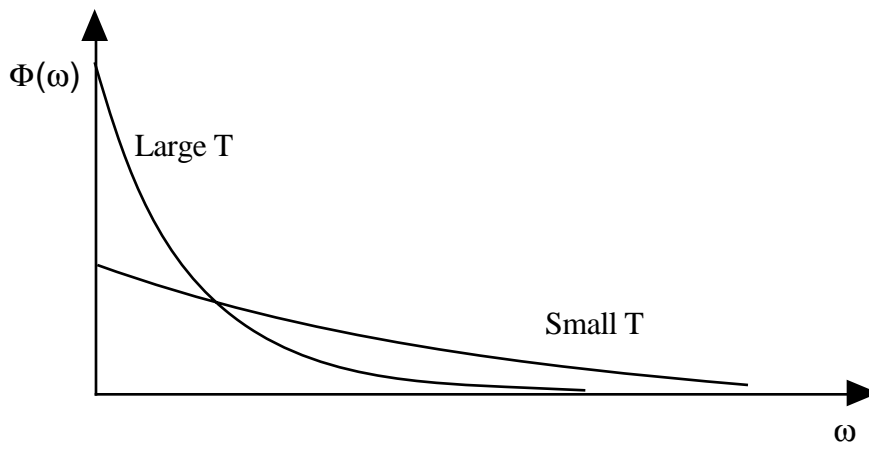
So that the power spectrum

$$\begin{aligned} \Phi(\omega) &= \int_{-\infty}^{\infty} \exp\left[\frac{-\tau}{T}\right] e^{-i\omega\tau} d\tau \\ &= \frac{-1}{\frac{1}{T} + i\omega} \exp\left\{-\tau\left(\frac{1}{T} + i\omega\right)\right\} \Big|_{-\infty}^{\infty} \end{aligned} \quad (6.40)$$

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

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

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.



Example 2: Power spectrum of white noise

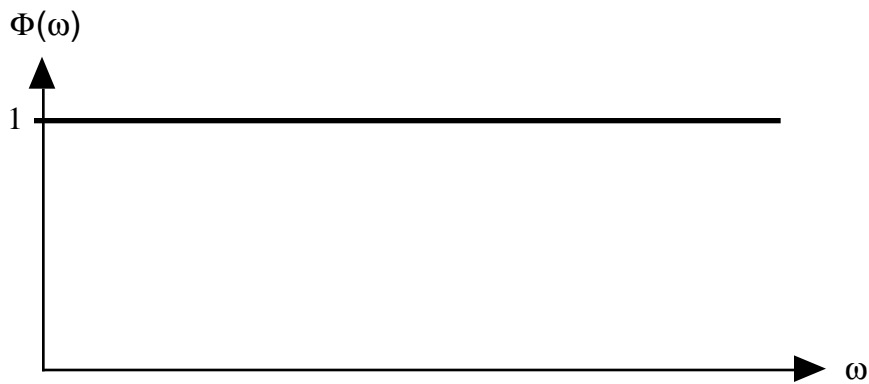
The autocorrelation of white noise is a delta function of the lag.

$$r(\tau) = \delta(\tau)$$

The power spectrum is therefore,

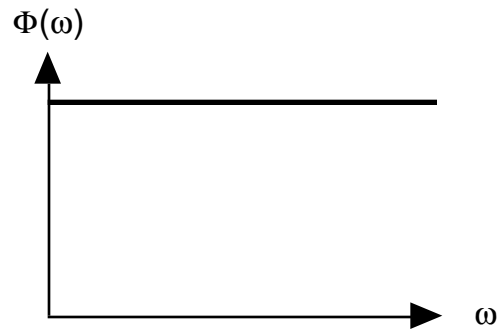
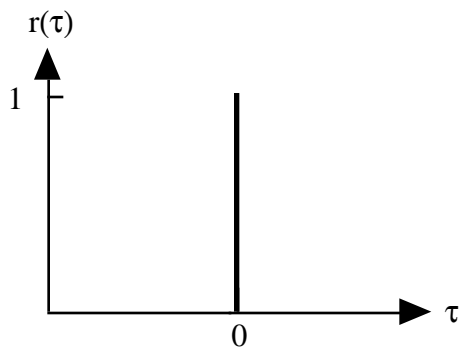
$$\begin{aligned} \Phi(\omega) &= \int_{-\infty}^{\infty} \delta(\tau) e^{-i\omega\tau} d\tau \\ &= 1 \end{aligned} \quad (6.42)$$

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

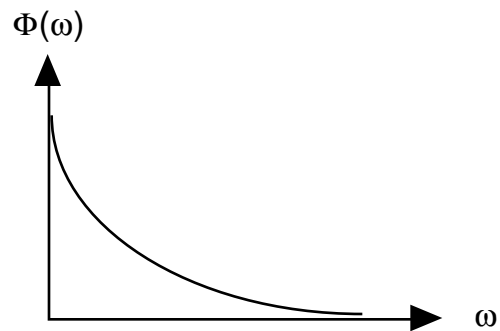
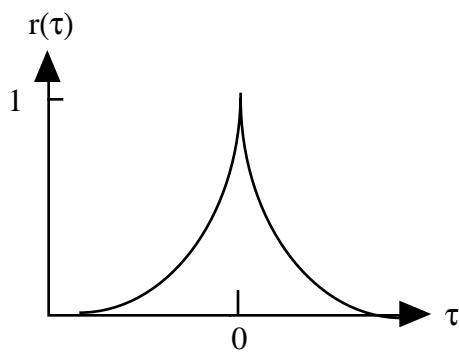


Some Examples of lag correlation - Power spectrum pairs

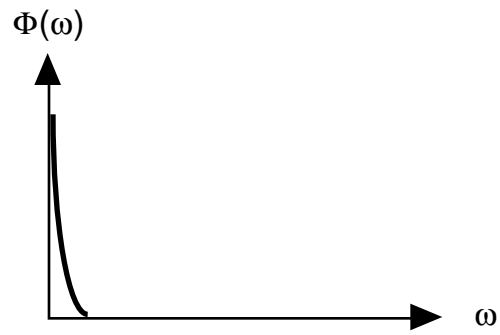
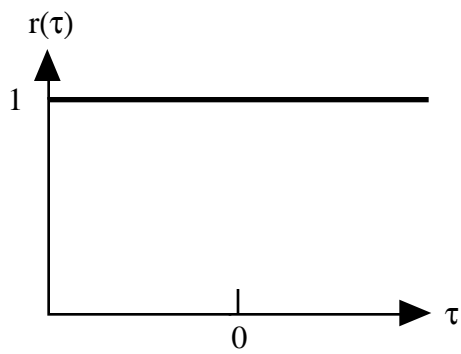
a) white noise $r(\tau) = \delta(\tau)$



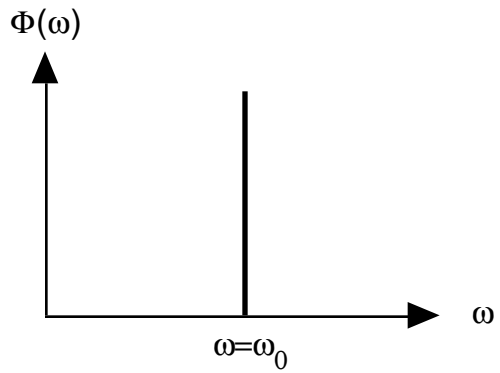
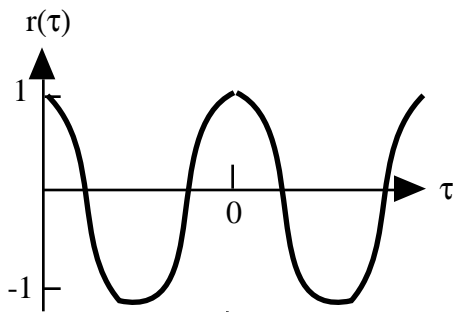
b) red noise $r = \exp(-\tau/T)$



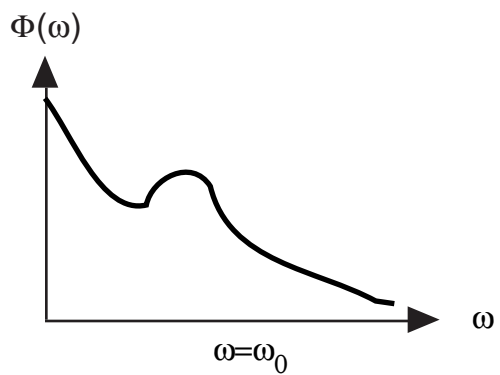
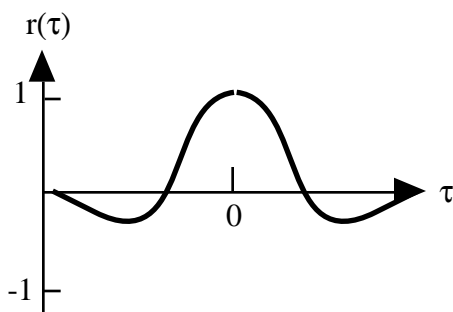
c) linear trend $r(\tau)=1$ $\Phi(\omega)$



d) pure periodicity



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

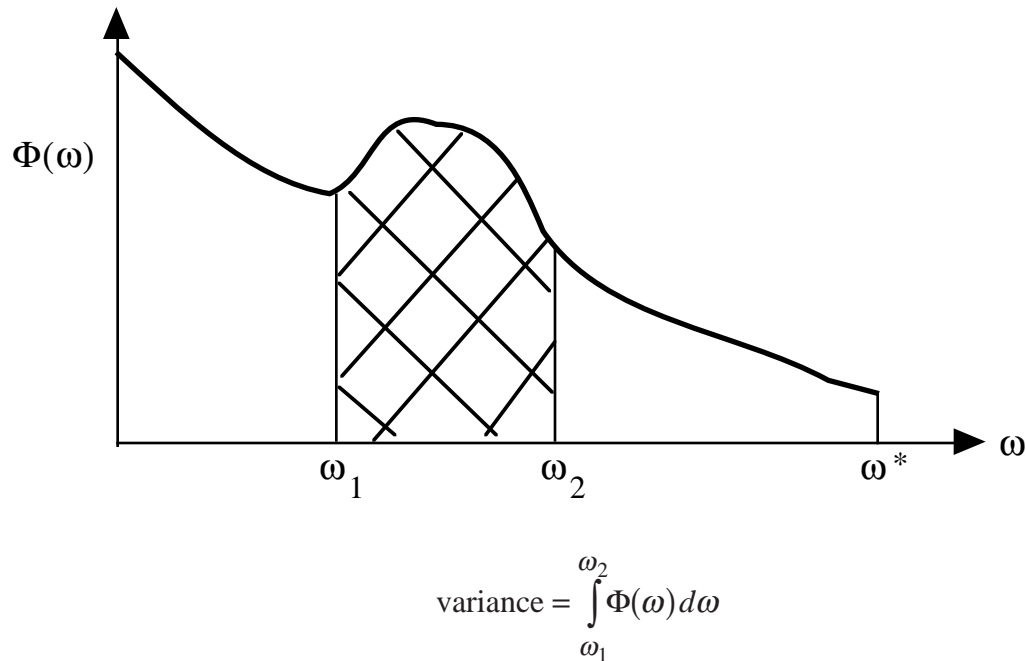


6.2.4 Plotting Power Spectra

Two basic schemes are in common use for plotting power spectral densities as a function of frequency; the linear scale and the logarithmic scale.

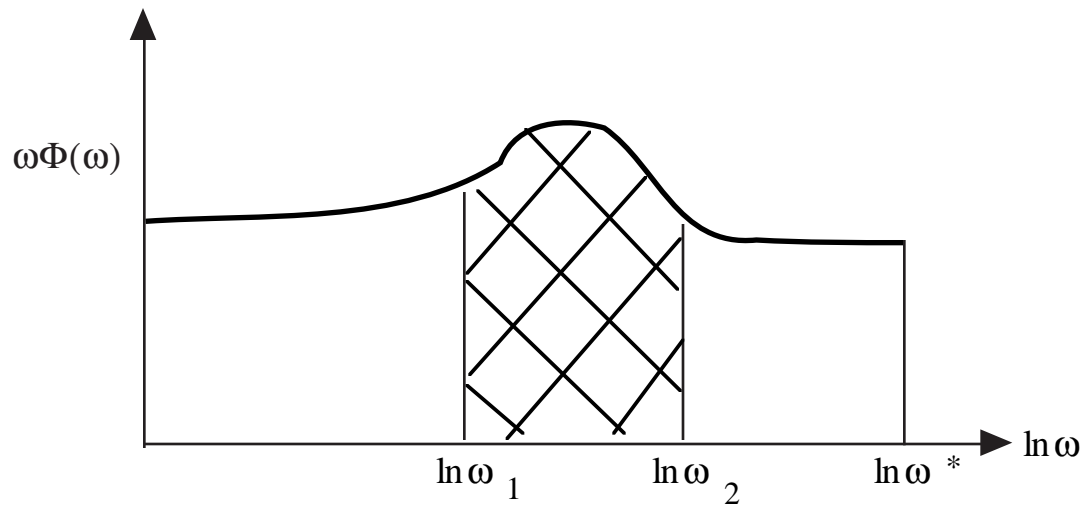
Linear Scale:

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



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.



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.

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

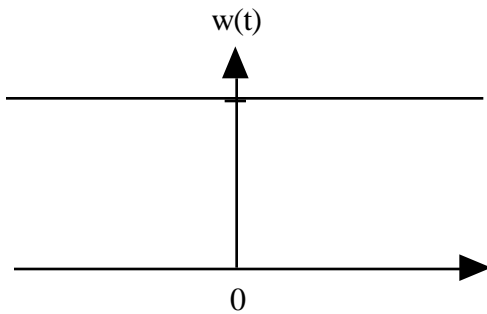
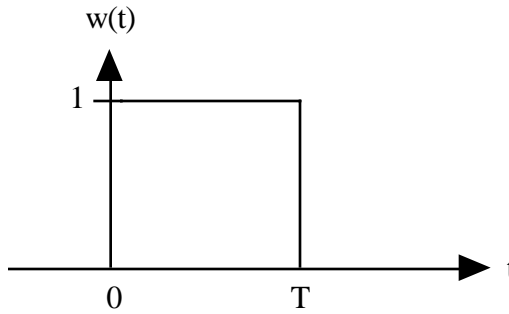
6.2.5 Data windows and Window Carpentry

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:

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

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 $(-\infty, \infty)$, 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 

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

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

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.

$$F(\omega) = \int_{-\infty}^{\infty} f(t) w(t) e^{-i\omega t} dt \quad (6.46)$$

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

where

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

by the convolution theorem

$$\int_{-\infty}^{\infty} f_1(t) f_2(t) e^{-i\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_1(\bar{\omega}) F_2(\omega - \bar{\omega}) d\bar{\omega} \quad (6.47)$$

Recall the formula for a continuous Fourier Transform,

$$\int_{-\infty}^{\infty} f_1(t) e^{-i\omega t} dt = F(\omega) \quad (6.45)$$

with the window included we get

$$F_{\hat{}}(\omega) = \frac{1}{2\pi} \int_{-eo}^{eo} F(\omega) \frac{\sin\left((\omega - \bar{\omega}) \frac{T}{2}\right)}{(\omega - \bar{\omega}) \frac{T}{2}} d\bar{\omega} \neq F(\omega) \quad (6.46)$$

where we have used the known Fourier transform pair for the Boxcar function,

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

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

$$W(\omega) = \frac{2}{\pi} \int_{-0}^0 \delta(\omega - \bar{\omega}) \sin(\omega - \bar{\omega}) \frac{T}{2} \quad (6.48)$$

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.

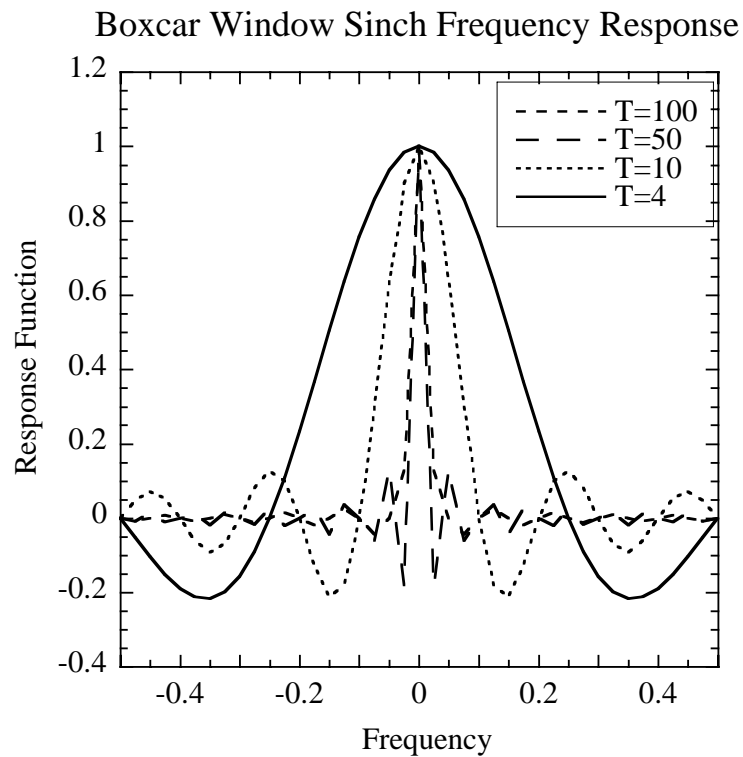
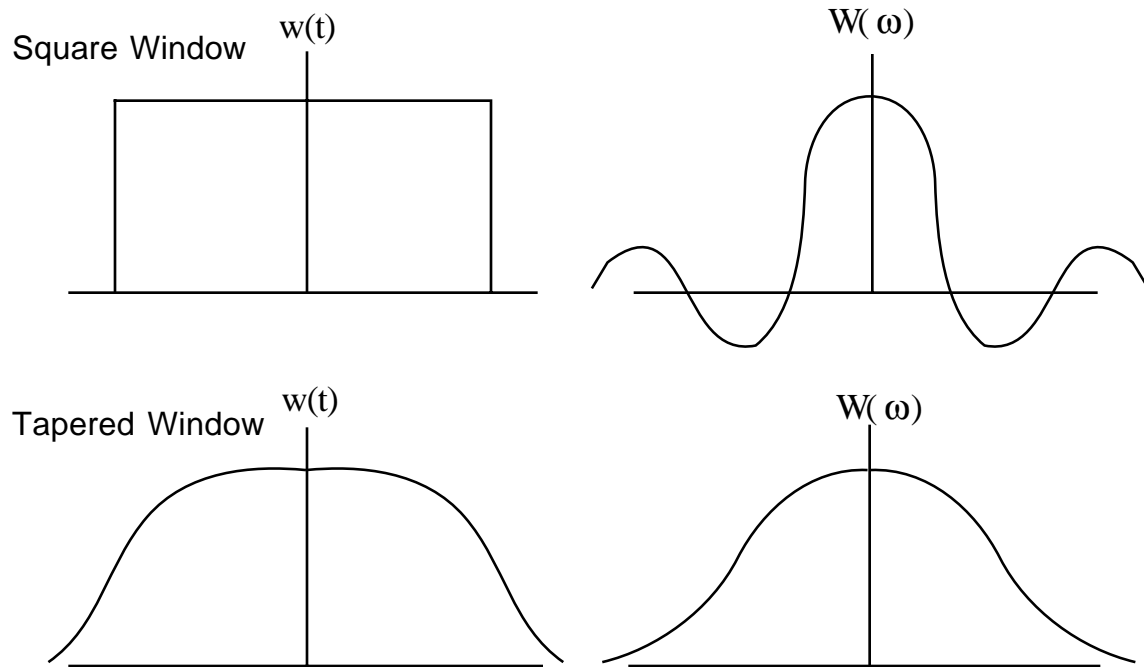
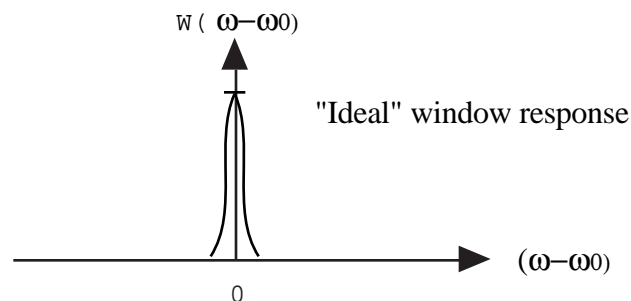


Figure. The sinch function plotted on the Nyquist interval $-0.5 < f < 0.5$ cycles per time step. Notice how the bandwidth and spread of the response function is narrowed as the total length of the Fourier Transform, T , is made longer. Notice the large negative side lobes and their rather slow decay.



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 sinc function window response through modifications of the window function $w(t)$.



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

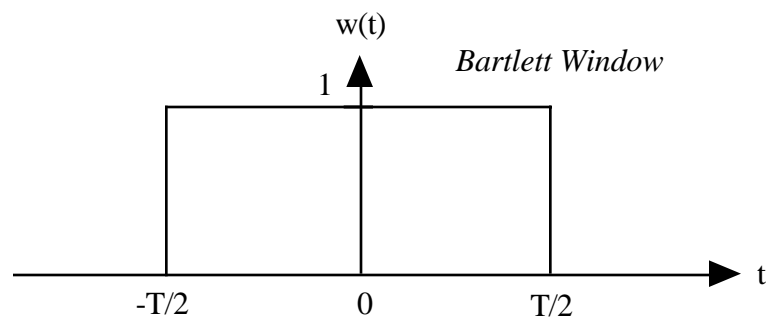
The particular choice made depends on the nature of the phenomena of interest and the frequency resolution required. Often our options are limited by the availability of data.

Window Carpentry:

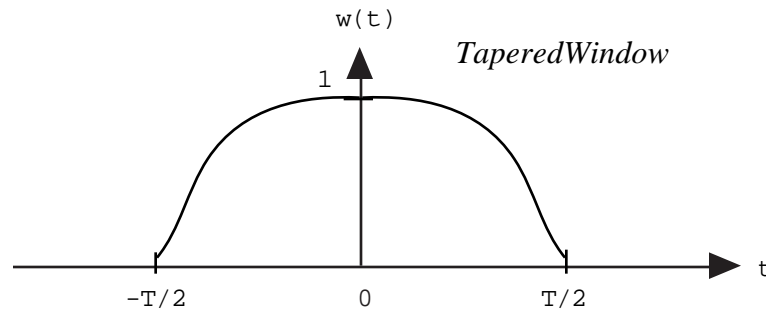
The characteristics that we want of frequency window are:

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 Bartlett 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.



Some windows (commonly used):Bartlett (square or rectangular); also no window

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

has the Fourier Transform,

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

The Bartlett 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 (something that happens frequently). The Bartlett window (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)

$$w(t) = \begin{cases} \frac{1}{T} \cdot \left(1 + \cos \frac{2\pi t}{T}\right) & ; 0 \leq |t| \leq T/2 \\ 0 & ; |t| > T/2 \end{cases} \quad (6.51)$$

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

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

$$W(\omega) = \underbrace{\operatorname{sinc}\left(\frac{\omega T}{\pi}\right)}_{\text{Same as Bartlett}} + \frac{1}{2} \left[\operatorname{sinc}\left(\frac{\omega T}{\pi} + 1\right) + \operatorname{sinc}\left(\frac{\omega T}{\pi} - 1\right) \right] \quad (6.52)$$

CANCELS 1st TWO SIDE LOBES

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

$$\omega(t) = \begin{cases} \frac{1}{T} \cdot \left(0.54 + 0.46 \cos \pi \frac{t}{T} \right) & 0 \leq |t| \leq T \\ 0 & |t| > T \end{cases} \quad (6.53)$$

$$W(\omega) = 1.08 \operatorname{sinc} \frac{\omega T}{\pi} + 0.46 \left[\operatorname{sinc} \left(\frac{\omega T}{\pi} + 1 \right) + \operatorname{sinc} \left(\frac{\omega T}{\pi} - 1 \right) \right] \quad (6.54)$$

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

Parzen: Power window

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

Special (common) form

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

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.

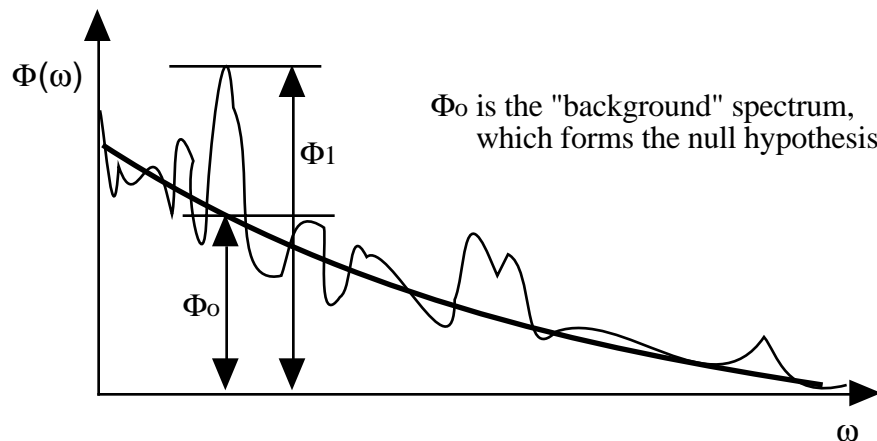
$$W(t) = \left(\frac{\sin \omega}{\omega} \right)^4 \quad (6.57)$$

Using the Fast Fourier Transform Method:

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 $M=2^n$. Mixed radix FFT's are also available for which $M=2^n 3^m 5^j$. 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 . The spectra and cross-spectra for these smaller chunks can be averaged into a grand spectrum that has some degree of statistical reliability. Because a tapered window is normally applied, the data near the ends of the chunks of length M 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 on 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 $M/2$ 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.



One way to evaluate the significance of a spectral peak is to compute the ratio Φ/Φ_0 and compare this value with a “Chi-Squared” test with the corresponding number of degrees of freedom

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

Number of degrees of freedom, n

$$n = \frac{N}{L} \cdot f_{\omega} \quad (6.59)$$

N = sample size

L = number of spectral estimates

f_{ω} = factor to compensate for smoothing done by the window.

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

$$F = \frac{s_1^2}{s_2^2} \quad (6.60)$$

is a value of a random variable having the F distribution with the parameters $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$.

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 11.

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

Now 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 L 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 L/f_ω independent estimates should show a significant peak. L is the number of frequencies retained in our spectrum and f_ω is the factor indicating the degree of smoothing by the window. If $L/f_\omega = 50$ then our chance of error is $0.1\% \times 50 = 5\%$ and we have only 95% confidence that this peak at this frequency is significant. 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 is the maximum lag to use? L ?

If you had applied a Fourier transform to the data and obtained a resolution of 0.01 cpd, a similarly smoothed spectrum could have been obtained by averaging 10 adjacent spectral estimates together, giving a bandwidth of 0.1cpd and $20 f_\omega$ degrees of freedom per averaged spectral estimate.

$$4 - 6 \text{ day periods} \Rightarrow 0.25 - 0.15 \text{ cycles per day} \Rightarrow \Delta f \sim 0.1 \text{ cpd}$$

$$f_i = 0, 0.1, 0.2, 0.3$$

↑

this one

$$\Delta f = \frac{1}{2 \cdot L \cdot \Delta t} = \frac{1}{2 \cdot 5 \text{ days}} \Rightarrow L = 20$$

$$\text{d.o.f.} = \frac{N}{L} \cdot f_\omega = \frac{400}{20} \cdot f_\omega = 20 \cdot f_\omega$$

What if you had used daily data?

$$\Delta f = \frac{1}{2L \cdot \Delta t} = \frac{1}{2 \cdot 5 \text{ days}} \Rightarrow L = 5$$

$$\text{d.o.f.} = \frac{N}{L} \cdot f_{\omega} = \frac{100}{5} \cdot f_{\omega} = 20 f_{\omega}$$

What has happened here? 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.

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 by the smoothing factor. The number of frequencies in Fourier spectral analysis is half the blocklength $L=M/2$.

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

$$\left(\frac{L}{f_{\omega}} \right) = \left(\frac{20}{f_{\omega}} \right)$$

b.) $1 \times$ daily

$$\left(\frac{L}{f_{\omega}} \right) = \left(\frac{5}{f_{\omega}} \right)$$

smaller multiplication of probability to type II error.

$$\text{a) } .01 \text{ error } \succ = (0.99)^{\frac{20}{f_{\omega}}} \cong 82\% \text{ confidence}$$

$$\text{b) } .01 \text{ error } \succ (0.99)^{\frac{5}{f_{\omega}}} \cong 95\% \text{ confidence}$$

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 $T = 10 \times 1000 \text{ years} = 10,000 \text{ years}$

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

$$L = 100 \Rightarrow L\Delta t = 100 \text{ years}$$

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

$$f = 0, 0.005 \text{ cpy}, 0.01 \text{ cpy}, 0.015, 0.02,$$

$$\text{period} = \infty, 200 \text{ years}, 100 \text{ years}, 67 \text{ years}, 50 \text{ years}.$$

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

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

b. Low frequency analysis

$$L = 1000$$

$$\Delta f = \frac{1}{2000} = 0.00050$$

$$\text{periods} = \infty, 2000 \text{ years}, 1000 \text{ years}, 667 \text{ years}, 500 \text{ years}$$

$$\text{d.o.f.} = \frac{10000}{1000} \cdot f_{\omega} = 10 \cdot f_{\omega}$$

If we are using a power of 2 FFT, then we would choose a block length of $M=2048$ and we would get at least $10,000/1024 = 9.8$ degrees of freedom (not good).

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,

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

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 , they present an approximate shape function for the spectrum associated with this autocorrelation function.

$$P(h, \rho, M) = \frac{1 - \rho^2}{1 - 2\rho \cos\left(\frac{h\pi}{M}\right) + \rho^2} \quad (6.62)$$

In (6.62) h represents the frequency, but is given in peculiar units. It varies from zero to M , corresponding to the $M+1$ frequencies provided by the Fourier transform of a lag-correlation function with a maximum lag of M . 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 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. When using direct Fourier transforms to compute spectra we can substitute $N/2$ for M , where N is the length of the time series transformed.

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,

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

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 .

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

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

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

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

$$\omega = \frac{2\pi h}{2M \Delta t} = \frac{\pi h}{M \Delta t} \quad (6.66)$$

So that (6.65) becomes,

$$\Phi(\omega) = \frac{-2\Delta t \ln(\rho)}{\ln(\rho)^2 + \left(\frac{\pi h}{M}\right)^2} \quad (6.67)$$

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,

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

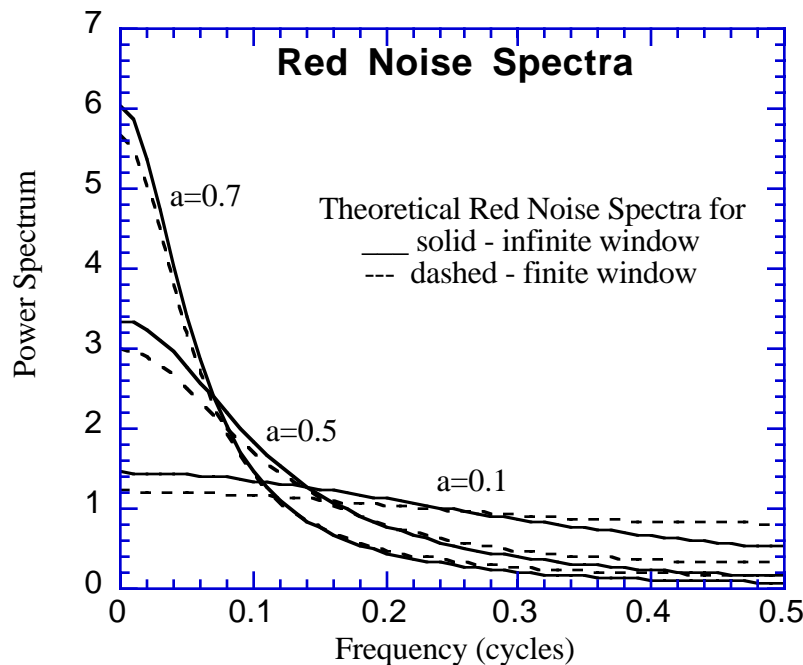
which makes (6.67) take the form,

$$\Phi'(\omega) = \frac{\Delta t \ln(\rho)}{\tan^{-1}\left(\frac{\pi}{\ln \rho}\right)} \left(\frac{1}{\ln(\rho)^2 + \left(\frac{\pi h}{M}\right)^2} \right) \quad (6.69)$$

(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,

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

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.



Hard Lessons with Small Samples:

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

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

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. 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 there appears to be a dominant frequency around 0.04 cpd, or a period of 25 days. In the next example spikes appear in the spectrum again, but at different frequencies. After averaging together 15 such spectra, the averaged spectrum still shows considerable spikiness. Even with 60 realizations (about 120 degrees of freedom) the spectrum is jagged. The sampling uncertainty is only reduced as the square root of the sample size.

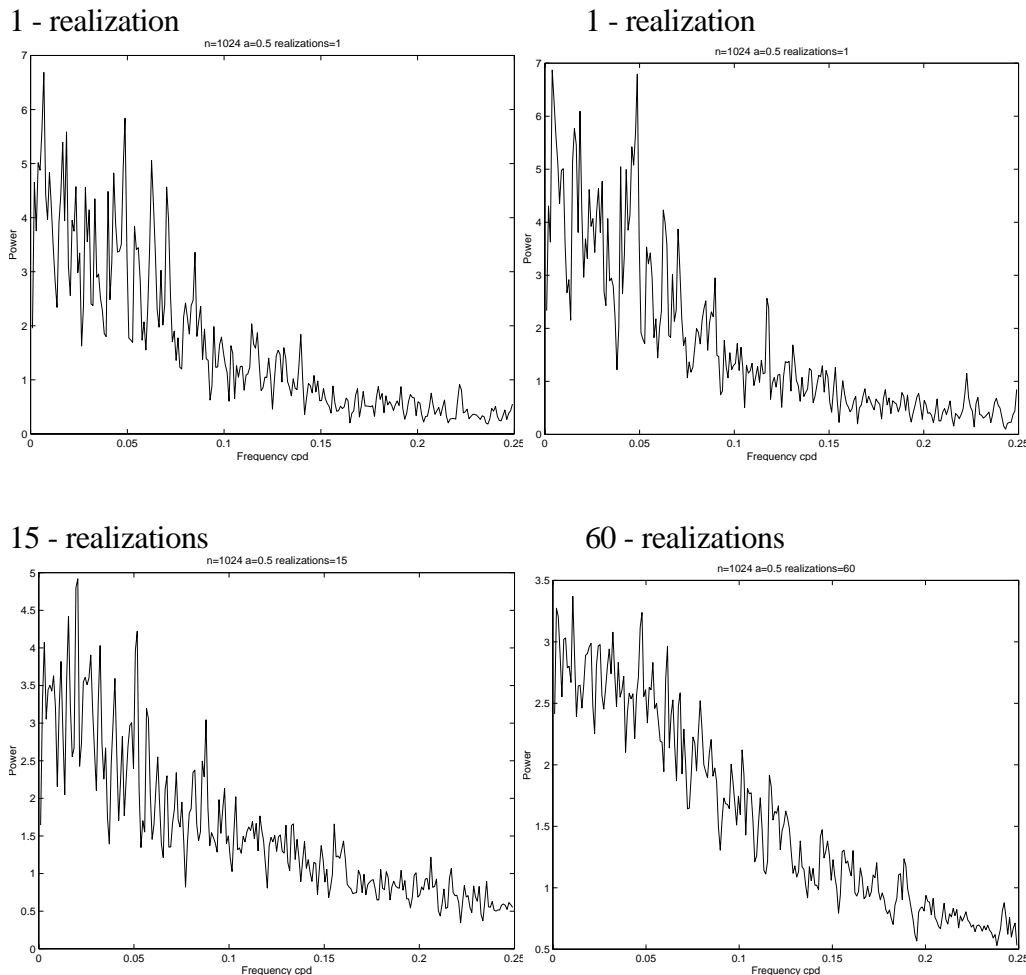


Figure: Sample spectra from a red noise process with a one-lag autocorrelation of 0.5. Top two single realizations of the lowest half of a spectrum of 1024 daily observations. The bottom two are averages of 15 and 60 independent realizations, respectively.

Single realizations can appear to have large peaks, as the upper right where an apparent huge peak just below a frequency of 0.05 cpd appears. The true spectrum of the process is a smooth red noise curve. Even after 60 realizations are averaged together, the spectrum is still spikey. The spectrum at the low right has about 120 degrees of freedom. The F statistic at 95% requires a ratio of 1.25, so that peaks must exceed a smooth line through the spectrum to be significant. Most of the wiggles in the spectrum do not satisfy this criterion. For the spectrum with 15 realizations (~30 dof) a ratio of about 1.5 is required to reject a null hypothesis of red noise, if you have picked the frequency before computing the spectrum.