Principles of Oceanographic Instrument Systems: Sensors and Measurements

Spring 2004

Jim Irish

INTRODUCTION TO SAMPLING THEORY AND DATA ANALYSIS

These notes are meant to introduce the ocean scientist and engineer to the concepts associated with the sampling and analysis of oceanographic time series data, and the effects that the sensor, recorder, sampling plan and analysis can have on the results. In order to plan the optimum sampling and analysis plan, one needs to understand what information and analysis are required, and how all these factors will affect the final result. To get the most from these lecture notes, the student should do supplemental readings from the references listed below. Exercises utilizing the MATLAB software package will be assigned at the appropriate place in the lectures.

An outline of this section is given below, and covered in handouts.

1. Time Series and Analysis:
   - Properties of a random, stochastic processes
   - Statistical description: mean, variance, correlation/covariance, spectra
   - Fourier transforms, frequency domain/time domain description of a process
   - Digital filtering and filters: Convolution product, filters, and filter response
2. Sampling Theory:
   - Sampling process, sampling theorem, and sampling effects on statistics
   - Aliasing and the Nyquist frequency
   - Power density spectra, coherence, degrees of freedom, confidence limits
3. Environmental Sampling in the real world:
   - Calibrations: static, dynamic
   - Digitizing effects, prefiltering
   - Sensor frequency response effects
   - Sensor noise limitations

Suggested references and readings:

Jenkins, G.M. and D.G. Watts, Spectral Analysis and its Applications, Holden-Day, San Francisco, 1968.

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

Bendat, J.S. and A.G. Piersol, Random Data: Analysis and Measurement Procedures, Wiley-interscience, New York, Second Edition, 1986.

Daley, R., Atmospheric Data Analysis, Cambridge University Press, New York, 1991.

Cochran, W.T., et al, "What is the Fast Fourier Transform," *IEEE Trans. Audio and Electroacoustics,* AU-15(2), 45-55, 1967.

Glossary of Terms, from Blackman, R.B. and J.W. Tukey, The Measurement of Power Spectra, Dover, 1958.

Bingham, C., M.D. Godfrey and J.W. Tukey, “Modern Techniques of Power Spectrum Estimation,” *IEEE Trans. Audio and Electroacoustics*, AU-15(2), 56-66, 1967.

Welch, P.D., “The use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging of Short, Modified Periodograms,” *IEEE Trans. Audio and Electroacoustics,* AU-15(2), 70-73, 1967.

Carter G.C., C.H. Knapp and A.H. Nutall, “Estimation of the Magnitude-Squared Coherence Function Via Overlapped Fast Fourier Transform Processing,” *IEEE Trans. Audio and Electroacoustics*, AU-21(4), 337-344, 1967.

Background

Everyone has some idea vague of what is involved in making measurements of the environment. However, few people have the background to really know how to do it properly. This is an introduction to how to measure the environment and analyze the results to obtain information for scientific studies and management decisions. As an ocean scientist or engineer you desire to make and analyze observations that will give you certain statistics describing the environment. To get these statistics, you need to design an experiment, place sensors in the field, digitize and record the results, analyze them on a computer, and finally present them in a meaningful manner. In reality, all these processes that you must go through can be thought of as a filter, or that you are looking at the ocean through “colored” glasses. In order to know what your glasses are doing to your view of the ocean, you need to know how to design an experiment to get the data that you want, select the sensors which will properly measure the environment, use recorders that will satisfactorily record the data, and utilize analysis techniques which will give the desired results. What follows is a simple introduction to the background that you will need to know in order to sample the environment properly. To simplify the discussions, much of the statistical complexity has been removed, so in order to become really professionally involved in data analysis, further course work is required to fill in this statistical information.

Properties of Random variables

We make the assumption that the environmental data of interest is a stationary, random, stochastic process. If this is so, then the environmental process that we wish to study can be fully described by its statistics.

Random Variables - A deterministic variable is one whose value may be determined or estimated exactly. An example of a variable which can be predicted is the result from an explicit mathematical relationship, e.g. $y(x) = a + bx$ or $y(x,t) = \cos(kx - \omega t + \theta)$. A random variable is one in which perfect prediction of succeeding values is impossible. Examples of random variables are the time until the next alpha particles is emitted from a radioactive source, the next direction taken by a particle in Brownian motion, or the elevation of the sea surface at a specific latitude, longitude and time. A set of observations of a random variable represents only one of many possible realizations.

Stochastic Process - A stochastic process is a collection of random variables. One observes a stochastic process when he examines a process developing in time in a manner controlled by probabilistic laws. A single set of observations is called a "sample function" or "sample record." A random stochastic process is described by all its possible sample functions. Repeated observations will result in sample functions that are different, or are not the same function of time, but have the same statistics. Some examples of stochastic processes are the number of particles emitted from a radioactive source, the path of a particle in Brownian motion, or the sea surface elevation variations due to surface wind waves. One can not predict exactly any succeeding values, but one can describe succeeding values statistically.

Stationary processes - The assumption that a random process is stationary is the most important assumption made in time series analysis. Perhaps this assumption is bad, at best it is only approximately true. A process is stationary when its "statistics" remain constant with time. Examples would be the output of a white noise generator, or the path of a particle in Brownian motion. An economic time series of Gross National Product (GNP) of the U.S. tends to increase with time, so is non-stationary. Hence, a stationary process is in statistical equilibrium and contains no trends or ramps. In actual practice we see three kinds of processes, 1) stationary, such as the output of the white noise generator, 2) quasi- stationary over a short period, such as atmospheric turbulence over a few minutes, or ocean waves over a few tens of minutes, and 3) non-stationary, such as the GNP, where the properties (statistics) are obviously changing with time. An oceanic example of a non-stationarity in a time series would be the surface wave field at the WHOI dock measured for 3 hours. During that time the tide would change the mean sea level by an amount which would "look" like a trend, but is really just a large, low frequency signal which is not resolved by our short record length. The wave field also might be growing in response to wind forcing, so the wave height and wavelength are changing with time. Another example would be the short-term temperature fluctuations observed for a month during the spring which are superimposed on the yearly warming and cooling. Since most geophysical spectra are "red," or have more energy at low frequencies than at high frequencies, this may be a significant problem. Therefore, "beware" of unresolved low frequencies.

Time - Frequency Domain Description of a Process

Time Series - Everyone is familiar with a time series representation of some phenomena, that is a series of values at successive points in time. For example the tides or surface waves seen at the coast appear to vary somewhat like a sine wave as a function of time. Observations of oceanographic temperature or currents at a particular position as a function of time is a time series. Time is a continuous variable, and is represented as a function of time by x(t) as shown in Figure 1 below.

Although geophysical processes are continuous in time, practical considerations require that we sample environmental process at discrete intervals in time, δt, for a finite length of time, T. That is, at regular intervals such as once a second, one minute, two hours, etc., an observation is made of the continuous process for a finite length of points. In order to simplify the analysis, sampling should always be done at equally spaced intervals in time. (It should be noted that the monthly averages often tabulated are not evenly spaced in time, so should not be analyzed as such.) We will consider a process, x, which is sampled at equally spaced time intervals, δt.

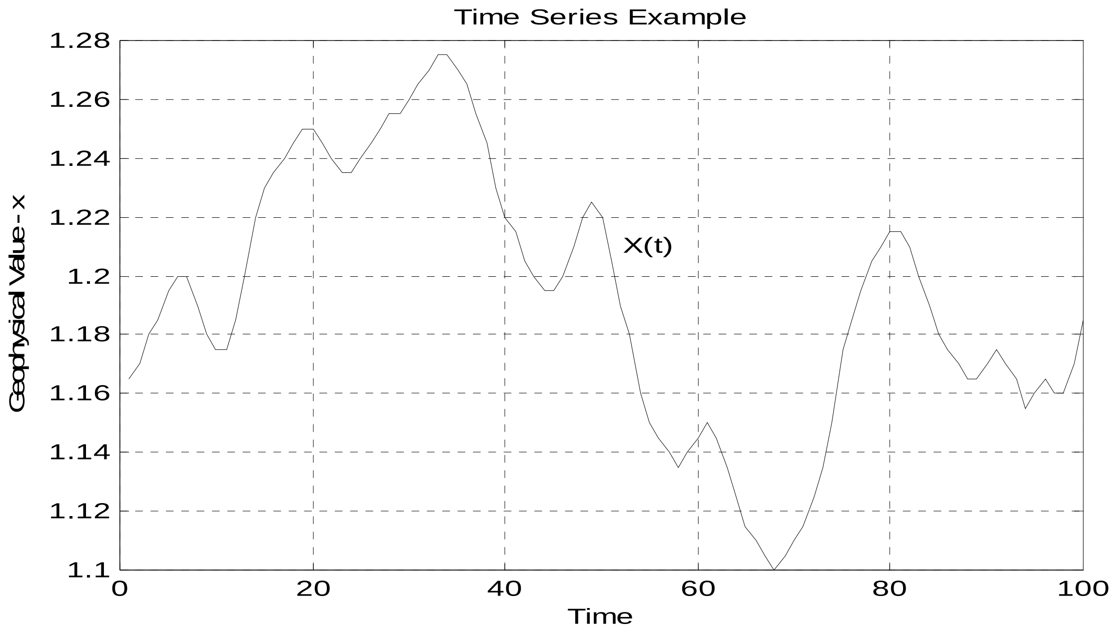

Figure 1. X is a continuous function of time, and is plotted with its value on the ordinate axis with value increasing upward and time increasing to the right on the abscissa.

If the number of samples or terms in the time series is n, then the length of the series (which was T) will be nδt. The relationship between the discrete sample set, x_t, (what we have to work with) and the original continuous function, x(t), will be covered below, but it is obvious that one needs to sample fast enough to be able to "see" the higher frequency fluctuations in the signal.

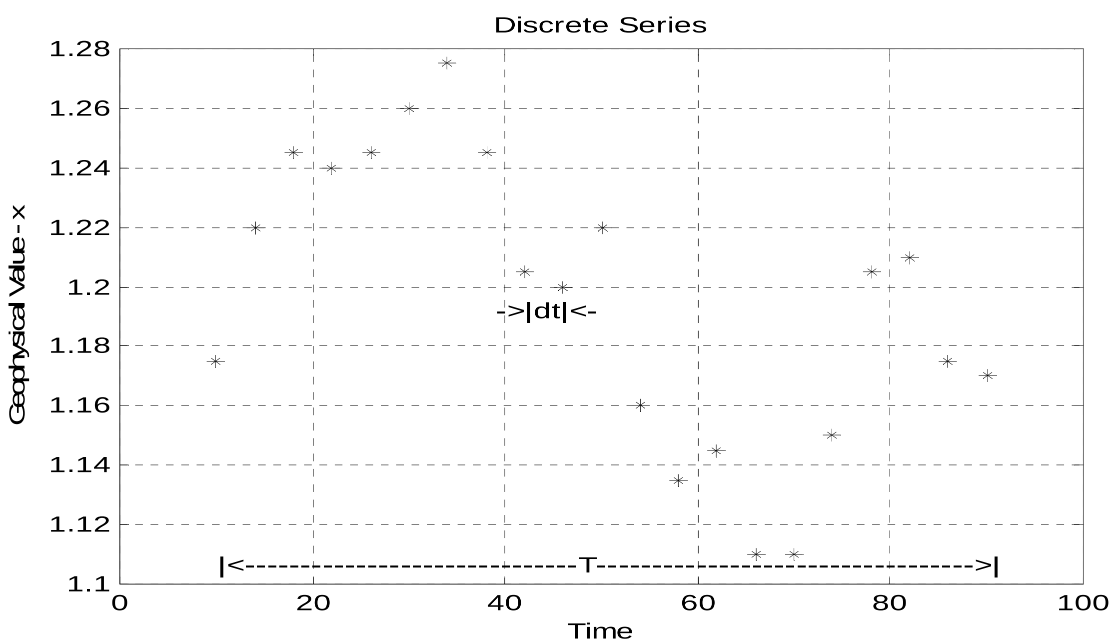

Figure 2. Discrete time series, x_t, of observations shown in Figure 1, but sampled every 4 units in time, or $\delta t = 4$.

Statistical descriptions - The second assumption (after stationarity) that we make is that the process may be adequately described by the lower moments of its probability density distributions. These include the mean, the variance, and the covariance, with its transform, the power spectrum. In our discussions below, we shall make a simplification by eliminating the probability density distribution function from our expected values. This would show up as multiplying the probability that x has a certain value in the integrals below. This makes the concepts involved in time series analysis a bit easier to follow in this introduction by eliminating some of the statistical considerations. However, in order to fully understand and use the power involved in proper data analysis procedures, one must go back and cover the concepts discussed below with a full statistical approach.

Mean. The mean is just the average or expected value for a process. Consider x(t) as a continuous realization of a sample function of a stochastic processes. Then the mean is just the "expected value" or most probable value,

$$\mu = \frac{\int_{-\infty}^{\infty} x(t)dt}{\int_{-\infty}^{\infty} dt} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt \qquad \text{Eq 1}$$

This is just the "static" component of the process. In practice we take the discrete time series, x_t, as the sample series which extends from t = 1 to m, at internals of time, δt. We can look at our statistics as being estimates of the actual statistics of the continuous process. As the number of terms in the series becomes large, then the estimated statistics converge to the actual statistics. The error in estimating the statistics is made up of several parts:

1. the sampling accuracy related to the length of the series, m, and sample interval, δt - that is how closely the sampled series, x_t, represents the true series, x(t), also of importance is the
2. digitizing accuracy used to create the actual numbers, x_t, and finally the
3. computer accuracy in doing the actual analysis calculations. On modern computers (workstations and PC's) this is generally not a problem.

The discrete representation of the mean is just the sum over all the terms, normalized by the number of terms

$$\mu \approx 1/m \sum_{i=1}^{m} x_i \text{ or } \mu \approx \mu' = 1/m \sum_{i=1}^{m} x_i \qquad \text{Eq 2}$$

as m, the number of terms, goes to infinity, this converges to the real mean.

Mean Square and Root Mean Square. The "intensity" of the series is given by the mean square value

$$\Psi^2 = \frac{\int_{-\infty}^{\infty} x(t)^2 dt}{\int_{-\infty}^{\infty} dt} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)^2 dt \qquad \text{Eq 3}$$

Its positive square root is called the rms or "root mean square" value.

Variance. The "dynamic component" of the series is given by the variance. This is the mean of the square of the differences from the mean,

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} [x(t)-\mu]^2 dt}{\int_{-\infty}^{\infty} dt} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [x(t)-\mu]^2 dt \qquad \text{Eq 4}$$

It is easy to expand the above equations to show that

$$\Psi^2 = \sigma^2 + \mu^2$$

The discrete representation of the variance is just

$$\sigma^2 \approx 1/m \sum_{i=1}^{m} [x_i-\mu]^2 \qquad \text{Eq 5}$$

The positive square root of the variance is called the **standard deviation**, σ.

Gaussian or Normal Distribution: One often hears about a normal or gaussian distribution, where the process is random, but the observations are grouped about the mean with a greater probability that they are nearer the mean, than farther away. This distribution is often found in geophysical processes and is often the statistics assumed by a processes when we calculate the statistics. When one samples the real environment and calculates the statistics of a process, the resultant is a good approximation of a Gaussian distribution. The probability density distribution function for a process x(t) with a mean μ_x and standard deviation σ_x is given by

$$p(x) = (\sigma_x \sqrt{2\pi})^{-1} e^{-(x-\mu x)^2/2\sigma x^2}$$

The normal or Gaussian distribution is plotted below in Figure 3 for $\sigma_x = 1$ and $\mu_x = 0$.

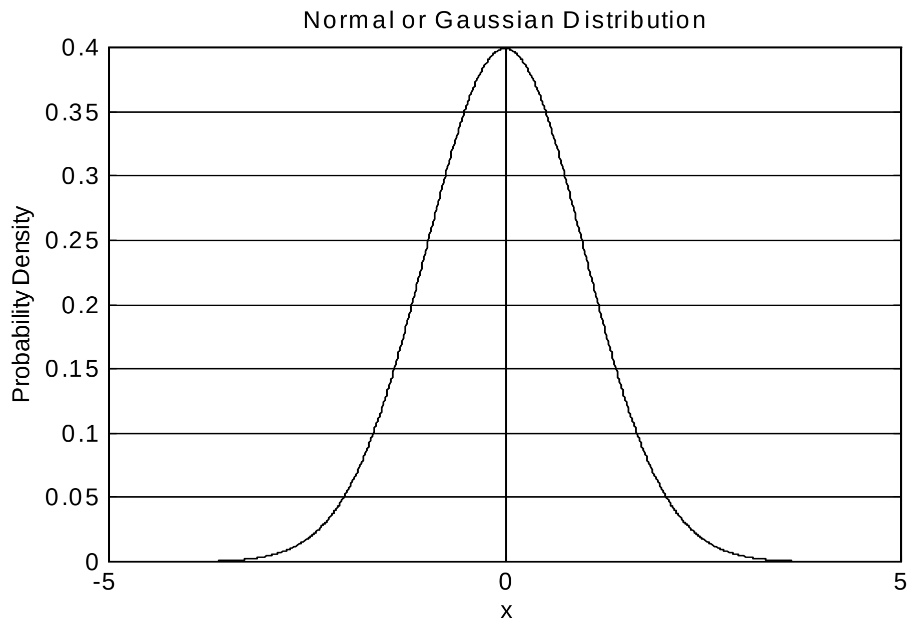

Figure 3. A Gaussian distribution of quantity x, giving the probability of the occurrence of a deviation in x from its mean value of 0.

*** Assignment #1 ***

A. Work with MATLAB until you are comfortable. Create a 2048-point sine wave with an amplitude of 2.0 and a period of 64. First you have to create an array of angles as the argument of the sine wave. Create the numbers from 1 to 2048 by i=1:2048 and multiply them by 2π and divide by 64 to get the argument. Take the sine of this and multiply by the amplitude, 2.0, to get your final series. You can change the initial phase of the sine wave by adding a constant to the angle series. Plot the resultant series and label the axes.

B. Calculate the statistics of the series (maximum, minimum, mean, mean square, root mean square (rms), variance and standard deviation). Some of these computations can be done with standard MATLAB functions, and others you will have to create. Do these results make sense with what we have discussed in class? Write your own MATLAB “stats” function for calculating these statistics and outputting the results.

C. The series you generated is deterministic, i.e. not a random series. Generate a 2048-point random series using MATLAB’s random number generator. If you use the RAND function, the series that has amplitude of 0 to 1.0 with uniform probability of any value between 0 and 1. If you use the RANDN function, the series is normally distributed about 0 – any value from –½ to ½. If you use the first, you will want to offset the series by 0.5 to bring the mean to zero. Then you can multiply the result by 0.1 to make a random series of amplitude 0.1, and add this to your sine wave. Calculate the statistics of this random "noise" series, and the sine wave with the random series added (this series is now random since you can not exactly predict the next value in time). Does this agree with what we have discussed in class?

D. Plot the normal and uniform distribution series and discuss of this simple exercise in terms of the statistics and your understanding of time series so far. Note: Discussion is important!

Covariance. The auto-covariance function describes the dependence of values of the sample function at one time, on those at another time. The auto-covariance for the continuous series is a function of time lag, τ.

$$R(\tau) = \frac{\int_{-\infty}^{\infty} [x(t)-\mu][x(t-\tau)-\mu]dt}{\int_{-\infty}^{\infty} dt}$$

Eq 6

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [x(t)-\mu][x(t-\tau)-\mu]dt$$

Note that $R(\tau)$ is an even function (symmetric about $\tau = 0$, or the same at $\tau = -\tau$) with a maximum at $\tau = 0$. The auto-covariance is useful in detecting a deterministic signal in the presence of random background noise. It is obvious that if $\tau = 0$, then the auto-covariance function, $R(0) = \sigma^2$, the variance.

Correlation. A normalized form of the covariance function is often used. The auto-correlation function or auto-correlation is

$$\rho(\tau) = \frac{R(\tau)}{R(0)} \qquad \text{Eq 7}$$

where $\rho(\tau)$ is a dimensionless number with $|\rho(\tau)| \leq 1$. If the sample function has a predominate periodicity, then at lag τ corresponding to that period, $R(\tau)$ will have a relative maximum or minimum since the series will be shifted over one period and line up.

Example. Consider the arbitrary function of time = A Cos(kx - ωt + ϕ) where A is an arbitrary amplitude of the sinusoid, k is the wavenumber (2π/wavelength), x is the horizontal coordinate in the direction of the sinusoid, ω is the frequency (2π/period), t is time and ϕ the initial phase. Then, since $R(\tau)$ is symmetric about t=0, we can take the integral over only positive times by doubling the right hand side, and holding x constant,

$$R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{0}^{T} A \,\mathrm{Cos}(kx-\omega t+\phi)\; A \,\mathrm{Cos}(kx-\omega(t-\tau)+\phi)dt$$

i. take t = 0, and changing the range of the integration from t=0 into blocks with t going from 0 to π/ω, and then multiplying by n where we let n go to infinity in the limit. Then T in the denominator becomes nπ/ω and

$$R(0) = \lim_{n\to\infty} \frac{n\omega}{n\pi} \int_0^{\pi/\omega} A^2 \, Cos^2\,(kx - \omega t + \phi)dt$$

and since Cos²(arg) where arg goes from 0 to π is ½,

$$R(0) = \tfrac{1}{2}A^2 \text{ and } \rho(0) = 1.0$$

ii. take t = 2π/ω and again change the range of the integration as above

$$R\left(\frac{2\pi}{\omega}\right) = \lim_{n\to\infty} \frac{n\omega}{n\pi} \int_0^{\pi/\omega} ACos(kx-\omega t+\phi)\; ACos(kx-\omega t+\phi-2\pi)dt$$

$$= \omega/\pi \int_0^{\pi/\omega} A^2 \, Cos^2(kx-\omega t+\phi)dt$$

$$R\left(\frac{2\pi}{\omega}\right) = \tfrac{1}{2} A^2 \text{ and } \quad \rho\left(\frac{2\pi}{\omega}\right) = 1.0$$

iii. take $\tau = \pi/\omega$

$$R\left(\frac{\pi}{\omega}\right) = \omega/\pi \int_0^{\pi/\omega} A \, Cos(kx-\omega t+\phi)\; A \, Cos(kx-\omega t+\phi-\pi)dt$$

$$= - \omega/\pi \int_0^{\pi/\omega} A^2 \, Cos^2(kx-\omega t+\phi)dt$$

$$R\left(\frac{\pi}{\omega}\right) = -\tfrac{1}{2} A^2 \text{ and } \quad \rho\left(\frac{\pi}{\omega}\right) = -1.0$$

iv. Finally take $\tau = \pi/2\omega$

$$R\left(\frac{\pi}{--}\right) = \omega/\pi \int^{\pi/\omega} A \, Cos(kx-\omega t+\phi)\; ACos(kx-\omega t+\phi-\pi/2)dt$$

$$2\omega \qquad \int_{0}$$

$$= \omega/\pi \int_{0}^{\pi/\omega} A^2 \, \mathrm{Cos}(kx-\omega t+\phi)\ \mathrm{Sin}(kx-\omega t+\phi)dt$$

$$R\left(\frac{\pi}{2\omega}\right) = 0 \quad \text{and} \quad \rho\left(\frac{\pi}{2\omega}\right) = 0$$

Again, it is clear that $R(0) = \sigma^2$, the variance.

Power Spectrum. If x(t) is a time series made up of a sum of "m" number of cosines each with its own amplitude, A_i, frequency, ω_i and phase θ_i, i.e.

$$x(y,t) = \sum_{i=1}^{m} A_i \, \mathrm{Cos}(k_i y - \omega_i t + \phi_i) \qquad \text{Eq 8}$$

where we have also included a wavenumber, k_i and distance, y. Then the variance is twice the integral from 0 to T in the limit as $T \to \infty$,

$$\sigma^2 = R(0) = \lim_{T\to\infty} \frac{2}{2T} \int_{0}^{T} \sum_{1}^{m} A_i^2 \mathrm{Cos}^2(k_i y - \omega_i t + \phi_i)dt$$

Now we can break the length T up in to n pieces, each π in length, and let the number of pieces go to infinity.

$$\sigma^2 = R(0) = \lim_{n\to\infty} \frac{n}{n\pi} \int_{0}^{\pi} \sum_{1}^{m} A_i^2 \, \mathrm{Cos}^2(k_i y - \omega_i t + \phi_i)dt$$

$$\sigma^2 = R(0) = \frac{1}{\pi} \int_{0}^{\pi} \sum_{1}^{m} A_i^2 \, \mathrm{Cos}^2(k_i y - \omega_i t + \phi_i)dt$$

$$= 1/\pi \sum_{i=1}^{m} A_i^2 \ \pi/2$$

$$\sigma^2 = R(0) = \sum_{i=1}^{m} \tfrac{1}{2} A_i^2 \qquad \text{Eq 9}$$

So if x(t) can be regarded as being made up of a sum of sinusoids, its variance can be decomposed into components of average power, $\frac{1}{2} A_i^2$, at the various frequencies, ω_i. Assuming a continuous distribution of frequencies, we obtain (without proof),

$$\sigma^2 = R(0) = \int_{-\infty}^{\infty} S(f)df \qquad \text{Eq 10}$$

where S(f) is called the power spectrum (or variance spectrum). Thus S(f) df is the measure of the average power or variance in the frequency band f - ½df to f + ½df – which really says that S is how the variance is distributed with frequency. It can further be shown that

$$S(f) = \int_{-\infty}^{\infty} R(\tau)\ e^{-2\pi i f\tau}\ d\tau \qquad \text{Eq 11}$$

This can be recognized as the Fourier transform of the covariance function (see below for definition and discussion of the Fourier transform). Then we must have

$$R(\tau) = \int_{-\infty}^{\infty} S(f)\ e^{2\pi i f\tau}\ df \qquad \text{Eq 12}$$

with $\tau = 0$, we again obtain equation 10.

Cross-Covariance and Cross-Correlation. Given two different sample functions, x and y, with means of μ_x and μ_y, the cross correlation and cross covariance function can be taken as discussed above. We had from equation 6,

$$R_x(\tau) = \frac{\int_{-\infty}^{\infty} [x(t)-\mu][x(t-\tau)-\mu]dt}{\int_{-\infty}^{\infty} dt}$$

for one series. For two series this just becomes

$$R_{xy}(\tau) = \frac{\int_{-\infty}^{\infty} [x(t)-\mu_x][y(t-\tau)-\mu_y]dt}{\int^{\infty}} \qquad \text{Eq 13}$$

$$\int_{-\infty} dt$$

Again, this can be normalized to give the cross-correlation,

$$\rho_{xy}(t) = \frac{R_{xy}(\tau)}{\sqrt{[R_x(0)\ R_y(0)]}} \qquad \text{Eq 14}$$

and again

$$|\rho_{xy}| \leq 1.0$$

The cross-correlation of two sets of data describes the dependence of the values of one set of data on those of the other set as a function of lag, τ. Note that now $R_{xy}(\tau)$ is not an even function and the maximum does not necessarily occur at $\tau = 0$. i.e. consider x = cosine and y = sine. When $\tau = \pi/(2\omega)$, they line up so you get a peak in ρ, so the maximum occurs at $\tau=\pi/(2\omega)$ and not $\tau=0$.

The discrete covariance functions are,

$$R_x\tau = 1/m \sum_{i=1}^{m} [x_i - \mu_x][x_{i-\tau} - \mu_x] \qquad \text{Eq 15}$$

$$R_{xy}\tau = 1/m \sum_{i=1}^{m} [x_i - \mu_x][y_{i-\tau} - \mu_y] \qquad \text{Eq 16}$$

The correlation is again the normalized covariance

$$\rho_x = R_x/\sigma^2 \qquad \text{Eq 17}$$

$$\rho_{xy} = R_{xy}/\sqrt{[R_x(0)\ R_y(0)]} \qquad \text{Eq 18}$$

*** Assignment #2 ***

A. Create two sine waves of the same frequency and but slightly different initial phase. Add in two different random noises (uniform and normal distribution), and calculate the Auto-covariance and Auto-correlation functions for one of these series. Hint: the MATLAB CONV (convolution) function does the multiplication and summing as required for the covariance and correlation. The MATLAB COV function returns the variance of the vector, not the covariance. If you use MATLAB's covariance function you will get a different result. Therefore, before using any of the MATLAB function, look at it and understand what it does. This is the power of MATLAB - the documentation of a function is always available as an "m" file. If you make your own, describe the results you get in terms of finite record length.

B. Calculate the Cross-covariance and Cross-correlation functions from these two series. Again describe the results. Do you understand what the results are telling you and how you might use this tool in your research?

C. Plot the original time series, the auto- and cross-correlation results. Are the results what you expected from our discussion in class?

Fourier Transforms

Time-frequency space - In the case of a sine wave, everyone recognizes that it can be expressed by an amplitude and phase at a specific frequency, and that this representation is more representative of the geophysical process than expressing the sine wave as a function of time. i.e. instead of x(t) now we have a separate representation of the process which we express as X(f) where f is the frequency. X(f) has an amplitude, A, and phase, θ, associated with each frequency, f. However, it should be obvious that expressing a sine wave as a function of time or as a function of frequency are really just two ways of expressing the same thing. Furthermore, it makes more physical sense to look at the sine wave as an amplitude and phase at a frequency than as a function of time. We can extend this to say that our continuous process x(t) can be represented by a sum of sinusoids with a given amplitude and phase at each frequency. This concept is now quite accepted, but in 1807 when Fourier first suggested that one could reconstruct exactly any function of time by an infinite sum of sinusoids, he astounded many contemporaries. Fourier's theorem states that any function can be reconstructed from a sum of sinusoids and that the Fourier transform is exactly that sum. Therefore, this says that we can express an observed process by a function of time, or by a Fourier transform of this as a function of frequency. For our discrete series, we have a time series in "time space" or a Fourier series in "frequency space."

Fourier Transform For our function x(t) we define its Fourier transform Z(f) as

$$Z(f) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi ift}\, dt \qquad \text{Eq 19}$$

Where Z is generally a complex number. It is made up of real and imaginary parts which are in reality the coefficients of the cosine and sine functions as

$$Z(F) = A(f) + i\,B(f)$$

$$x(t) = A(f)\,\mathrm{Cos}(2\pi ft) + B(f)\,\mathrm{Sin}(2\pi ft)$$

Therefore, a series x which is made up of a sum of cosine waves only will have only a real part or real coefficients, (only A coefficients) and be an even function symmetric around 0. Similarly a series x which is made of a sum of sine waves only, will have a Fourier transform which is only imaginary (only B coefficient) and be an odd function. Depending on the normalization of the discrete Fourier transform, the coefficients A and B are the amplitudes of the cosine and sine waves which go to make up the time series x.

The inverse Fourier transform of Z(f) is then defined by

$$x(t) = \int_{-\infty}^{\infty} Z(f)\, e^{2\pi ift}\, df \qquad \text{Eq 20}$$

These two expressions (equations 19 and 20) define a Fourier transform pair. It is obvious that

(1) Z(f) is a function of frequency,
(2) Z(f) is an exact mathematical representation of x(t) (which was our original series as a function of time), and
(3) Z(f) is defined by the Fourier transform of x(t).

Similarly x(t) is a representation of Z(f) as given by the inverse Fourier transform. The Fourier transform is the way of getting from a function of time to a function of frequency. Mathematically the Fourier transform of a function x(t) exists if

$$\int_{-\infty}^{\infty} |x(t)|dt$$

exists. Some useful functions do not have transforms.

$x(t)$ = a non-zero constant

$x(t) = A \, \mathrm{Sin}(2\pi ft)$

$x(t) = 1$ for $t > 0$

$= 0$ for $t < 0$

However, the problem exists because these series are infinite in length. A time series resulting from a geophysical process which is sampled for a finite period of time (or what we call "gated") always has a Fourier transform. Mathematically, we are saying that any gated function (a function which has a beginning and end and no infinite values) has a Fourier transform. Since our observations start and stop in a finite time interval, they have Fourier transforms.

Distance-wavenumber representation. Usually one represents processes as a function of time and uses the Fourier transform to obtain a representation as a function of frequency. Similarly one can define a Fourier transform pair which are functions of distance and wavenumber. So variations as a function of space (as for example vertical CTD profiles of temperature and salinity in the Moonakis river) can be represented as a space series, or a wavenumber series. The Fourier transform is the mechanism relating these two representations of the process. For example, if we have an observed series such as a surface wave field of the form

$$Y(x,t) = A \, \mathrm{Cos}(kx - \omega t)$$

then we have two Fourier transforms, first as a wavenumber spectrum at one point in time (constant t)

$$Z(k) = \int_{-\infty}^{\infty} Y(x)\, e^{-2\pi ikx}\, dx$$

or as a frequency spectrum at one point in space(constant x)

$$Z(f) = \int_{-\infty}^{\infty} Y(t)\, e^{-2\pi i f t}\, dt$$

Fourier transform relationships - Let x(t) and Z(f) be a Fourier transform pair, and be complex functions of real variables. We will use the shorthand

$$x(t) \supset Z(f) \qquad \text{Eq 21A}$$

and also

$$Z(f) = x(t) \qquad \text{Eq 21B}$$

then,

$$x(-t) \supset Z(-f) \qquad \text{Eq 22}$$

$$x^*(-t) \supset Z^*(-f) \qquad \text{Eq 23}$$

where * denotes the complex conjugate. Note that if we have a complex number defined as follows,

$$Z = R + i\, I,$$

where $i = \sqrt{-1}$

then $Z^* = R - i\, I$

So from Eq 19 and Eq 20, we can then say if

x(t) is even, $x(t) = x(-t) \supset Z(f) = Z(-f)$ which is even

x(t) is odd, $x(t) = -x(-t) \supset Z(f) = -Z(-f)$ which is odd

X(t) is real, $x(t) = x^*(-t) \supset Z(f) = Z^*(-f)$ which is Hermitian

X(t) is imag. $x(t) = -x^*(-t) \supset Z(f) = -Z^*(-f)$ which is anti-hermitian

Define X(f) and Y(f) as the Fourier transforms of x(t) and y(t) respectively, and let “a” be a scalar. Some useful theorems concerning Fourier transform pairs are,

$x(t) + y(t) \supset X(f) + Y(f)$ additive

$a\, x(t) \supset a\, X(f)$

$x(t-a) \supset e^{-2\pi i a f}\, X(f)$ shift

$$x(at) \supset 1/|a|\ X(f/a) \qquad \text{scale}$$

$$\partial x(t)/\partial t \supset 2\pi if\ X(f) \qquad \text{derivative}$$

$$\int_{-\infty}^{\infty} x(t)y^*(t)dt = \int_{-\infty}^{\infty} X(f)Y^*(f)df \qquad \text{power}$$

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df$$

We shall return to these relationships between transform pairs when we consider filtering.

*** Assignment #3 ***

A. Create a sinusoidal time series of amplitude 2.5 to which you have added a random component of amplitude 0.2. Note that the length of you series should be a power of two.

B. Calculate and plot the Fourier transform (FFT) of your original series and your noisy series. Does it agree with what you expect? Note that MATLAB normalizes the transform by the length of series. I like to normalize it so the FFT returns the amplitude of the sine or cosine.

C. Create your own MATLAB function (which we will eventually evolve into the power spectrum and coherence function) with proper normalization.

Convolution Product and Filters

Convolution Product: To discuss filters we need to define the convolution product which we will use to represent the filtering process. The convolution product of our series x(t) with some arbitrary set of weights w(t) is defined by the integral

$$Y(t) = \int_{-\infty}^{\infty} w(\tau)\ x(t-\tau)d\tau. \qquad \text{Eq 24}$$

We will abbreviate the convolution product by the symbol, $*$, and so

$$Y(t) = w(t) * x(t) = x(t) * w(t) \qquad \text{Eq 25}$$

Then we can show that this product is

$$g * h = h * g \qquad \text{communitive} \qquad \text{Eq 26}$$

$$f * (g * h) = (f * g) * h \qquad \text{associative} \qquad \text{Eq 27}$$

$$f * (g + h) = f * g + f * h \qquad \text{distributive} \qquad \text{Eq 28}$$

Convolution theorem If we denote the Fourier transform of x(t) by X(f) or $x \supset X$, as defined in Eq 21, then the convolution theorem states that (remember $f \supset F$ and $g \supset G$)

$$f * g \supset F \bullet G$$

$$f * g = f \bullet g$$

or Eq 29

$$f * g \supset f \bullet g.$$

This says that the Fourier transform of the convolution product of f and g is equal to the product of the Fourier transform of f and the Fourier transform of g. Conversely,

$$f \bullet g = f * g \qquad \text{Eq 30}$$

and we also have

$$f * g * h = f \bullet g \bullet h \qquad \text{Eq 31}$$

$$f * (f \bullet h) = f \bullet (g * h) \qquad \text{Eq 32}$$

A time series can be expressed either in the time domain or by its Fourier transform in the frequency domain. The convolution theorem tells us that we can substitute multiplication in one domain for convolution in the other. It is easier to visualize the product of two functions than the convolution, and it is often easier (quicker) to compute. The convolution theorem tells us that if we express the time series x as a Fourier transform, and multiply this Fourier transform by the Fourier transform of w, then inverse transform, we will obtain the same result as doing the convolution of the filter w and the series x. This is written

$$w * x = w \bullet x \qquad \text{Eq 33}$$

With the FFT (fast Fourier transform) algorithms available on digital computers, it is often faster (and hence cheaper) to transform, multiply, and inverse transform, rather than calculate the convolution. (For information on the introduction of the FFT, see the IEEE Trans. Audio and Electroacoustics, June 1967, Vol AU15, vol2 Pg 43-93.)

The convolution process is the filtering process where a filter, w, is applied to the series, x. If the time series x(t) is a discrete series (see section following on the sampling process for details on how the discrete series is created), x_t, t=0,1,2,…,n, and the filter, w_t, t=0,1,2,3,…,m, is defined over a different number of terms, m, but with the same sample interval, then the convolution product of the discrete series can be represented as the finite sum

$$Y_t = \sum_{\tau=0}^{m} w_\tau\, x_{t-\tau} \text{ for } t = 0,1,2,\ldots n+m$$

This will only have meaning if we define

$$x_t = 0 \ ; \ t=n+1,n+2,..,n+m$$

$$w_t = 0 \ ; \ t=m+1,m+2,..,n+m$$

For example, consider the filter to be a triangular filter represented by a set of 5 weights, w_t = {0, 0.25, 0.5, 0.25, 0}, and let the series be represented by a set of numbers, x_t = {1.2, 2.2, 1.6, 1.8, 1.2, 1.6, 1.4}. Note that the filter series weights sum to one. This is so that the values of the series will have the same value after the filter is applied. Then n=6 and m=4, and

$$Y_0 = 0.0 + 0.55 + 0.8 + 0.45 + 0.0 = 1.8$$

$$Y_1 = 0.0 + 0.4 + 0.9 + 0.3 + 0.0 = 1.6$$

$$Y_2 = 0.0 +.0.45 + 0.6 + 0.4 + 0.0 = 1.45$$

Note that the filtered series is shorter than the original series by m-1=3 terms so the filtered series is now n-(m-1) = 3 terms. It is obvious that $x * w = w * x$ since the sum is symmetric.

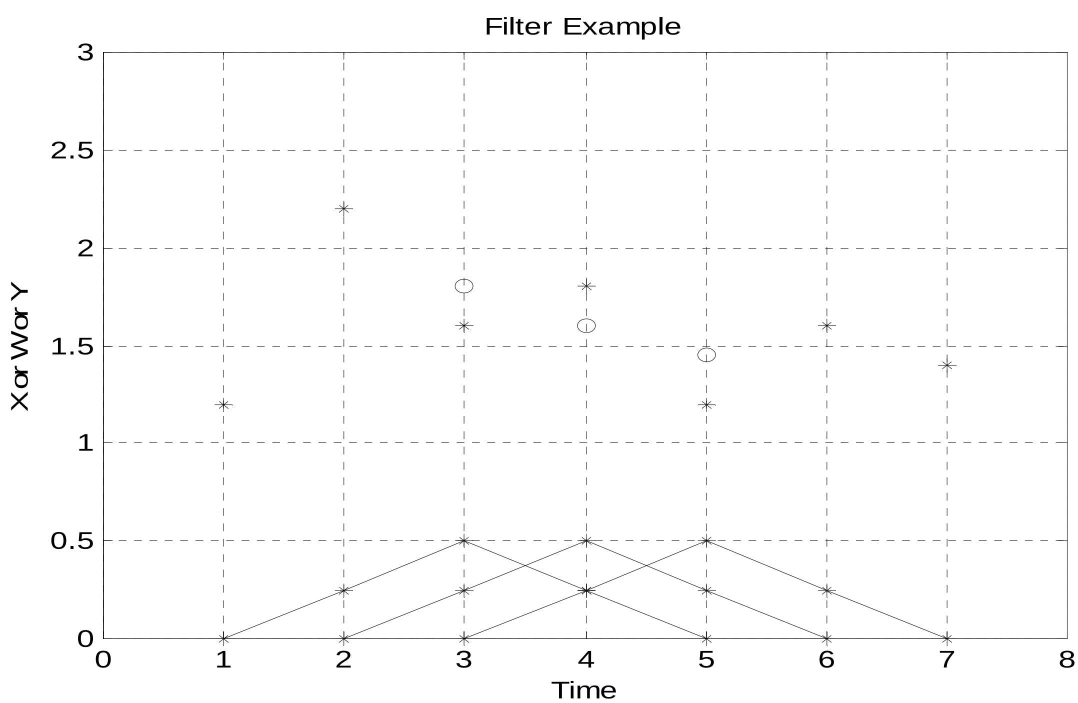

Figure 3. The example given in the text plotted.

Filters and filtering: Filters are selective devices which are used to discriminate and reduce time series. In a broad sense, the world, our experiment, the sensors, recorder, and data analysis are really filters which shift and reduce the data. We begin by considering some simple filters such as would be used in data reduction. Consider a filter as a box with an input x(t) and an output y(t)

$$x(t) \rightarrow \quad F \quad \rightarrow y(t)$$

We shall write the filtering operation as $x(t) \Rightarrow y(t)$.

Time Invariant Linear Filters. We want well behaved filters which will enhance some frequencies in a time series and suppress others. This type of filter is invariant in time, that is,

$$\text{if} \quad x(t) \Rightarrow y(t) \qquad \text{then} \quad x(t+\tau) \Rightarrow y(t+\tau) \text{ for all } \tau. \qquad \text{Eq 34}$$

and is linear, that is

$$\begin{aligned} &\text{if} \quad x_1(t) \Rightarrow y_1(t) \\ &\text{and} \quad x_2(t) \Rightarrow y_2(t) \qquad \text{Eq 35} \\ &\text{then} \quad x_1(t)+x_2(t) \Rightarrow y_1(t)+y_2(t) \end{aligned}$$

For any constant, a, it follows that

$$a\, x(t) \Rightarrow a\, y(t) \qquad \text{Eq 36}$$

A time invariant, linear filter has the important property that it does not confuse frequencies (no non-linearities). A sinusoidal input produces a sinusoidal output of the same frequency, that is

$$A \operatorname{Cos}(kx-\omega t+\phi) \Rightarrow B \operatorname{Cos}(kx-\omega t+\theta)$$

where A, B, ω and θ are constant in time. The filter has altered the amplitude from A to B (a gain of B/A) and altered the phase by $\theta - \phi$, but the frequency of the sinusoid is unchanged. In order to describe the response of a filter, let us observe what it does to an ideal input. We want to consider a complex filter so, consider two parallel filters

$$A_1 \operatorname{Cos}(\omega t) \Rightarrow y_1(t)$$

$$A_2 \operatorname{Sin}(\omega t) \Rightarrow y_2(t)$$

and regard the first as the real and the second as the imaginary part of this complex filter, then we can rewrite the two inputs as one and linearity gives

$$A_1 \operatorname{Cos}(\omega t) + i\, A_2 \operatorname{Sin}(\omega t) \Rightarrow y_1(t) + i\, y_2(t)$$

This can be written,

$$A\, e^{i\omega(t+\tau)} \Rightarrow y(t+\tau) \qquad \text{Eq 37}$$

where the phase is written as a frequency, ω, times a constant, τ. This can be broken up into the sinusoidal part and a phase part as

$$A\ e^{i\omega(t+\tau)} = A\ e^{i\omega t}\ e^{i\omega\tau} \Rightarrow y(\tau)\ e^{i\omega t}$$

where exp(iωt) is regarded as a multiplier. For $\tau = 0$, or zero phase shift, the above relationship gives

$$y(t) = y(0)\ e^{i\omega t} \qquad \text{Eq 38}$$

Where we now have the output as an initial amplitude which is expressed as y(0) and the sinusoidal time varying part expressed as the exponential, so that

$$A\ e^{i(\omega t)} \Rightarrow y(o)\ e^{i(\omega t+\theta)} \qquad \text{Eq 39}$$

The real part of Eq 39 is

$$A\ Cos(\omega t+\theta) \Rightarrow |y(0)|\ Cos(\omega t+\theta+\phi) \qquad \text{Eq 40}$$

$$B\ Cos(\omega t+\psi)$$

where $B = |y(0)|,\ \psi = \theta + \phi,\ y(0) = |y(0)|exp(i\phi)$.

We characterize the response of a filter (as above) by its transfer function, L, by use of Eq 38

$$L(\omega) = y(0)/A \qquad \text{Eq 41}$$

which is the ratio of the output to the input at a given frequency ω . L(ω) is generally complex, it has a real and imaginary part, which can also be written as an amplitude gain and a phase shift. In terms of the real input Eq 40

$$L(\omega) = B/A\ e^{i\phi} = B(\omega)/A(\omega)\ e^{i\phi(\omega)}$$

so that L(ω) is a complex number whose modulus is the amplitude gain (B/A) and whose argument is the phase shift (ϕ) produced by the real input at each frequency ω. A property of L(ω) is that

$$L(\omega) = L(-\omega)^* \qquad \text{Eq 42}$$

This transfer function, L(ω), characterizes how a filter responds to a sharp spike in frequency (a sinusoid). This is done in the laboratory, but applying a sine wave and observing the output voltage amplitude and phase change. An alternate description can be made by specifying its response to a spike or impulse in time. Let $\delta(t)$ represent a unit pulse in time. This pulse is narrow compared with the resolving power of our sensors. For example take the gate function

$$\Pi = 1 \quad \text{for} \quad |t| \le ½$$
$$\Pi = 0 \quad \text{for} \quad |t| > ½ \qquad \text{Eq 43}$$

This represents an impulse when written as

$$1/\tau \ \Pi(t/\tau) = \delta_\tau(t)$$

as τ goes to 0, $\delta_\tau(t)$ becomes arbitrary sharp (centered near t=0). We have

$$\int_{-\infty}^{\infty} \delta_\tau dt = \lim_{\tau\to 0} \int_{-\infty}^{\infty} (1/\tau)\Pi\ (t/\tau)dt = 1 \quad \text{Eq 44}$$

and as Π is non zero between $-½\tau$ and $½\tau$,

$$= \lim_{\tau\to 0} \int_{-½\tau}^{½\tau} (1/\tau)\Pi(t/\tau)dt = 1 \quad \text{Eq 44}$$

It can be shown by the limiting process that

$$\int_{-\infty}^{\infty} \delta(t)\ F(t)\ dt = F(0)$$

and by the shifting property of $\delta(t)$

$$\int_{-\infty}^{\infty} \delta(t-a)\ F(t)\ dt = F(a).$$

If $\delta(t) \Rightarrow l(t)$, we say that the impulse $\delta(t)$ produces an impulse response, $l(t)$.

An input which can be expressed as a linear combination of impulses, has an output which is described by the impulse response function.

$$\sum_{i=1}^{m} C_i\delta(t-t_i) \Rightarrow \sum_{i=1}^{m} C_i l(t-t_i)$$

Assuming that our impulses are infinitely close, we can go to the integral representation, and make the input exactly any input function of time.

Hence, we can express a filter by either its response to a sum of sinusoids (in the frequency domain, $L(\omega)$) or by its response to a sum of impulse (in the time domain, $l(t)$). This can be expressed,

if $\quad e^{i\omega t} \Rightarrow L(\omega)e^{i\omega t}$

then Eq 45

$$x(t) = \int_{-\infty}^{\infty} g(\omega)e^{i\omega t}d\omega \quad > \quad \int_{-\infty}^{\infty} g(\omega)L(\omega)e^{i\omega t}d\omega = y(t)$$

where $g(\omega)$ is the Fourier transform of $x(t)$, so that $x(t)$ is the inverse Fourier transform. This results is an integral representation of the transform, which is the amplitude and phase of the sinusoids at each frequency which add up to the initial function of time, which are now shifted in amplitude and phase by the complex response function, $L(\omega)$. Now we also can write,

$$\delta(t) \Rightarrow l(t)$$

Eq 46

$$\text{then } x(t) = \int_{-\infty}^{\infty} x(t)\delta(t-\tau)d\tau \Rightarrow \int_{-\infty}^{\infty} x(\tau)l(t-\tau)d\tau = y(t)$$

Therefore there are two ways of describing the response of a filter, as its frequency response function, $L(\omega)$ or as its impulse response function, $l(t)$. It should be obvious that these are just Fourier transforms of each other,

$$L(\omega) \supset l(t)$$

and

$$l(t) \supset L(\omega).$$

Examples of filters

1. The do-nothing filter, $x(t) \Rightarrow x(t)$

$L(\omega) = 1$ (e.g. amplitude multiplied by 1 & phase shifted by 0^o)
and $l(t) = \delta(t)$ (e.g. you get out what you put in)

2. The lag filter, $x(t) \Rightarrow x(t-\tau)$. The output is simply a delayed version of the input.

$$L(\omega) = e^{i\omega\tau}$$
$$l(t) = \delta(t-\tau)$$

3. Consider $x(t) \Rightarrow y(t) = x^2(t)$. An input consisting of two different frequencies produces sums, and difference frequencies of the inputs. These are called distortion (harmonic, intermodulation) and are the result of squaring a linear system. This is a bad filter to have to describe based on the output.

4. A time variant filter. If $x(t) \Rightarrow g(t)x(t)$ where $g(t)$ is any non-constant function of time. Such filters are commonly used on time series. Consider the gate function or “Boxcar,”

g(t) = Π(t). Since observations must be started and stopped, multiplication by Π(t) is unavoidable. We generally scale so the filter is Π(t/T) where the record length = T, so the resulting frequency confusion does not negate the experiment.

Let us more fully explore the gate function, II(t), as a filter. It has a Fourier transform

$$\Pi(t) \supset \int_{-\infty}^{\infty} \Pi(t)\,\mathrm{Cos}(2\pi ft)dt$$

where we need only to use the Cosine part of the transform since the gate function is an even function. Since the gate function is zero outside the interval |1/2|, the limits of the integral become plus and minus 1/2. We can further simplify this by noting that an even function is twice its value from 0 to 1/2. Therefore,

$$\Pi(t) \supset \frac{2}{2\pi f}\,\mathrm{Sin}(2\pi ft)\Big|_{0}^{½}$$

$$\Pi(t) \supset \frac{\mathrm{Sin}(\pi f)}{\pi f} \equiv \mathrm{Sinc}(f)$$

Applying this filter by multiplication in the time domain (our gating process) is the same as convolving the Fourier transform of your observations with the sinc function.

$$X(t) \bullet \Pi(t) \supset X(f) * \mathrm{Sinc}(f)$$

This often produces problems and must be dealt with properly in the design and sampling of the experiment. Therefore, filters are actually applied to the data in the process of sampling, and their effects must be known. To study these effects we must examine the sampling process.

*** Assignment #4 ***

A. Create a sine wave with noise as we did in assignment 3 and calculate the Fourier coefficients for your sine wave. Is the energy at the frequency it is supposed to be? Is your amplitude and phase properly represented in the Fourier coefficients?

C. Create a 5-point triangular filter of weights 0.0, 0.25, 0.50, 0.25, 0.0. To be normalized as a low pass filter, the weights must sum to one. Apply your triangular filter to your sine wave with noise.

D. Plot the initial and filtered series. Does the filter remove some of the random noise?

E. Using FFT, transform the filter weights (cosine transform or real part since the filter weights are an even function) to get the filter gain (filter frequency response function). Make a log-log plot of your filter's frequency response function. Does it look reasonable to you? Discuss.