Lecture 15:

Reading: Bendat and Piersol, Ch. 5.2.5, 5.2.6, 9.2

Recap

Last time we took a general look at correlation (and correlation coefficients) and their analog in spectral space: coherence. Coherence tells us how effectively two time series resemble each other at any given frequency.

Coherence: Examples

The power of coherence comes because it gives us a means to compare two different variables. With spectra we can ask, is there energy at a given frequency? With coherence we can ask whether wind energy at a given frequency drives an ocean response at a given frequency. Does the ocean respond to buoyancy forcing? Does momentum vary with wind? Does one geographic location vary with another location? Coherence is our window into the underlying physics of the system.

Let's put this to work, starting with an idealized case: Suppose we want to estimate currents entering and leaving the Mission Bay Channel. How do waves travel through the channel? We can represent this with a dispersion relationship describing the dominant propagation in frequency-wavenumber space: $k = K(\sigma)$.

You could imagine measuring Mission Bay by installing one current meter (with a cost of \$10-\$20,000), but another approach is to install a couple of pressure recorders along the axis of the channel (at a cost of \$1000 each). Let's assume all waves come from the ocean, and travel along the channel axis at speed $V = c + U_{current}$, where c is the wave speed and $U_{current}$ the background current speed. If the waves are surface gravity waves, $c = \sqrt{gD}$. The sensors measure time series of pressure only, so provide frequency information σ. How does σ relate to velocity? If we have a wavenumber $k = 2\pi/\lambda$, what does the pressure sensor see? It will detect frequencies $\sigma = kV = k(c + U_{current})$. So we can compute the cross spectrum between our two records.

Let's test this out. We'll define a hypothetical data set:

```
lambda=10; % 10 m wavelength
V=0.3; % 0.3 m/s propagation
n2s=0.2; % noise-to-signal ratio
time=(1:5000)';
x=n2s*randn(5000,1)+cos(2*pi/lambda*V*time);
y=n2s*randn(5000,1)+cos(2*pi/lambda*V*(time)+pi/2);
```

What happens if you Fourier transform without bothering to segment? Then the data end up being unrevealing. We can demonstrate this:

```
fx=fft(x);
fy=fft(y);
sx=abs(fx(1:end/2)).^2; sx(2:end)=2*sx(2:end);
sy=abs(fy(1:end/2)).^2; sy(2:end)=2*sy(2:end);
cxy=conj(fx(1:end/2)).*fy(1:end/2); cxy(2:end)=2*cxy(2:end);
C=abs(cxy)./sqrt(sx.*sy);
plot(C)
```

In this case, the coherence is 1 everywhere. Why is that? Because without averaging, we're merely

computing:

$$C_{xy}^2 = \frac{(X^*Y)^*(X^*Y)}{X^*XY^*Y} \quad (1)$$

$$= \frac{XY^*X^*Y}{X^*XY^*Y} \quad (2)$$

$$= \frac{X^*XY^*Y}{X^*XY^*Y} \quad (3)$$

$$= 1 \quad (4)$$

We need the averaging to find out if the phase relationship between x and y is repeatable. With only one segment, both x and y are guaranteed to have information at each frequency with a definable phase relationship between x and y. The multiple segments allow us to test whether this phase relationship is relatively stable in time: does x always lead y by about the same fraction of a cycle?

To do the coherence caluclaiton more constructively, we determine the frequency-space relationship between two data sets x_n and y_n, by first dividing them into segments and then Fourier transforming them, so that we have a set of X_k's and a set of Y_k's. When we computed spectra, we found the amplitude of each X_k and then summed over all our segments. Now we're going to do something slightly different. For each segment pair, we'll compute the product of X times the complex conjugate of Y: $X_k Y_k^*$. Then we'll sum over all the segments. In Matlab this becomes:

```
segment_length=500;
N=length(x);
M=segment_length/2; % define this value
Nseg=N/segment_length;
x_use=[reshape(x,segment_length,Nseg) ...
   reshape(x(M+1:end-M),segment_length,Nseg-1)];
y_use=[reshape(y,segment_length,Nseg) ...
  reshape(y(M+1:end-M),segment_length,Nseg-1)];
fx=fft(x_use); % should window and detrend here, but we're
               % skipping that for now
fy=fft(y_use);
sx=sum(abs(fx(1:M+1,:)).^2,2)/N; % sum over all spectra
                                 % (sum over 2nd index)
sx(2:end)=sx(2:end)*2;
sy=sum(abs(fy(1:M+1,:)).^2,2)/N; % sum over all spectra
                                 % (sum over 2nd index)
sy(2:end)=sy(2:end)*2;
cxy=sum(fx(1:M+1,:).*conj(fy(1:M+1,:)),2)/N;
cxy(2:end)=cxy(2:end)*2;  % since we multiplied the spectra by 2,
                     % we also need to multiply the cospectrum by 2

nd=size(x_use,2);
C=abs(cxy)./sqrt(sx.*sy);
delta_C = sqrt(1-alpha^(1/(nd-1)));

phase_C = atan2(-imag(cxy),real(cxy));
delta_phase = asin((1-betainc(2*nd/(2*nd+alpha^2),nd,.5))*...
         sqrt((1-C.^2)./(2*nd*C.^2)));
```

```
delta_phase2 = sqrt(1-C.^2)./(abs(C)*sqrt(2*nd));
```

Uncertainty for phase is often reported with the formula I provided last time

$$\delta_\phi = \sin^{-1}\left[t_{\alpha,2n_d}\sqrt{\frac{1-C_{xy}^2}{2n_d C_{xy}^2}}\right] \tag{5}$$

where $t_{\alpha,2n_d}$ is the "Student t distribution". But when we plot this up, it doesn't look very plausible. Bendat and Piersol provide a different formulation, which seems more plausible:

$$\text{std}\left[\phi_{xy}(f)\right] \approx \frac{\left[1-\gamma_{xy}^2(f)\right]}{|\gamma_{xy}(f)|\sqrt{2n_d}} \tag{6}$$

The phase difference that emerges from this is only relevant at the phase where there is coherence energy (15 cycles/1000 points in the example above), and in that case the phase is a quarter cycle different. If we reverse the order of x and y, we'll find negative phase, so a lead will turn into a lag.