Lecture 14:

Recap

While I was gone, you looked at aliasing and frequency-wavenumber spectra. There are lots of details, but the key concepts are important. Aliasing:

1. Variability beyond the Nyquist frequency will alias into your resolved frequencies. That means that if you find blue spectra, or if there's a spectral peak above the Nyquist frequency, it will show itself within your resolved frequencies. Thus rule #1 is if your spectra are white or blue near the high-wavenumber limit, you should be very skeptical of your computed high-wavenumber spectra, on the assumption that the spectral slopes are corrupted (at least a bit) by aliased, unresolved variability.

2. Aliasing can also be used strategically, for example if you want to detect tidal variability from a satellite that samples infrequently compared with the tidal period.

Frequency-wavenumber spectra:

1. If you can Fourier transform in time or space, then you can definitely do both together.

2. While we usually look at only positive frequencies or positive wavenumbers when we plot spectra, in the frequency-wavenumber two-dimensional plane, it usually makes sense to allow for positive and negative propagation, so we look at positive and negative wavenumbers with positive frequencies, or vice versa.

3. When we plot frequency-wavenumber spectra, we want to center around zero, so we have to shift the information from our fft around. The function "fftshift" will do this in Matlab, or you can do it manually.

Next up, what happens when we want to look at the relationship between two signals?

Extracting phase information from the Fourier coefficients

Before we get into frequency-wavenumber digression, let's make a quick detour to look at the phase information that we've been ignoring for most of the quarter. After all this effort to square Fourier coefficients, you might wonder what the real and imaginary parts are really good for. They are useful for sorting out the phasing of your sinusoidal oscillations. When is the amplitude at a maximum? To do this you can keep in mind that

$$A\cos(\sigma t+\phi) = a\cos(\sigma t) + b\sin(\sigma t). \tag{1}$$

This can be rewritten:

$$\cos(\sigma t)\cos(\phi) - \sin(\sigma t)\sin(\phi) = \frac{a}{A}\cos(\sigma t) + \frac{b}{A}\sin(\sigma t), \tag{2}$$

which means that

$$\frac{a}{A} = \cos(\phi) \tag{3}$$

$$\frac{b}{A} = -\sin(\phi) \tag{4}$$

so

$$\phi = \text{atan}\left(-\frac{b}{a}\right). \tag{5}$$

Actually there's more information in the Fourier coefficients than this conveys, since you know the signs of both a and b, and not just their relative magnitudes. The arctangent function doesn't distinguish +45°from -135°, but we can. In some numerical implementations, you can address this using a function called atan2.

```
phi = atan2(-b,a);
```

Covariance

Early in the quarter we discussed the variance, and we left for later the concept of correlation or covariance. If we want to compare two time series, we can compute the variance of one record relative to the other. Formally we can write:

$$\text{cov}x, y = \langle x(t)y(t)\rangle. \tag{6}$$

or in discrete terms

$$\text{cov}x, y = \frac{1}{N}\sum_{i=1}^{N} x_i y_i. \tag{7}$$

For comparison purposes, we often normalize this to produce a correlation coefficient, which is normalized by the variance:

$$r = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 \frac{1}{N} y_i^2}}. \tag{8}$$

(You might wonder how to judge whether a correlation coefficient is statistically significant. Correlation coefficients should have a Gaussian distribution, which means that cumulative distribution function will be an error function. We can use this to determine the correlation coefficient that we might expect from an equivalent number of random white noise variables:

$$\delta r = \text{erf}^{-1}(p)\sqrt{\frac{2}{N}} \tag{9}$$

where p is the significance level we want to consider, typically 0.95, and N is the effective number of degrees of freedom.)

Coherence

Coherence provides information that is analagous to a correlation coefficient for Fourier transforms. It tells us whether two series are statistically linked at any specific frequency. This can be important if we think that the records are noisy or otherwise uncorrelated at some frequencies, but that they also contain statistically correlated signals.

To compute coherence, first we need a cross-spectrum. (We looked at this in passing when we considered Parseval's theorem, but at that stage, I quickly set my different variables equal to each other.) Consider two time series $x(t)$ and $y(t)$:

$$x(t) = \sum_{n=-\infty}^{\infty} X_n e^{i\sigma_n t} \tag{10}$$

$$y(t) = \sum_{n=-\infty}^{\infty} Y_n e^{i\sigma_n t} \tag{11}$$

The the cross spectrum is computed in analogy with the spectrum:

$$\hat{C}_{XY}(\sigma_m) = \frac{\langle X_m^* Y_m \rangle}{\Delta\sigma} \tag{12}$$

The relationship between the cross-spectrum and the covariance is analogous to the relationship between the spectrum and the variance. There are some important details to notice.

1. The cross spectrum is complex, while the spectrum was real.

2. The cross spectrum is computed as an average of multiple spectral segments.

3. In our discrete Fourier transform, we should be normalizing by N, as always, but we're mostly concerned with relative values.

The cross-spectrum is complex, and when we use it we distinguish between the real and imaginary parts. The real part is called the "co-spectrum":

$$c(\omega_k) = \Re \sum_{n=1}^{N} (X_k Y_k^*) \tag{13}$$

and the imaginary part is called the "quadrature spectrum"

$$q(\omega_k) = \Im \sum_{n=1}^{N} (X_k Y_k^*). \tag{14}$$

To determine the frequency-space relationship between two data sets x_n and y_n, we first divide them into segments and Fourier transform 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

```
sum(X.*conj(Y),2);
```

The corresponding amplitude is $\sqrt{c^2(\omega_k) + q^2(\omega_k)}$. For comparison the spectra for X was:

$$f_x(\omega_k) = \sum_{n=1}^{N} X_k X_k^*, \tag{15}$$

and it was always real.

The coherence resembles a correlation coefficient. It's the amplitude squared divided by the power spectral amplitudes for each of the two components:

$$C^2(\omega_k) = \frac{c^2(\omega_k) + q^2(\omega_k)}{f_x(\omega_k) f_y(\omega_k)} \tag{16}$$

It's really important that your spectra are based on more than one segment, that is that N exceeds 1. If that weren't the case, you'd just have a single realization of each spectra, and the resulting squared coherence would be

$$C^2(\omega_k) = \frac{X(\omega_k) Y^*(\omega_k) X^*(\omega_k) Y(\omega_k)}{X(\omega_k) X^*(\omega_k) Y(\omega_k) Y^*(\omega_k)} = 1, \tag{17}$$

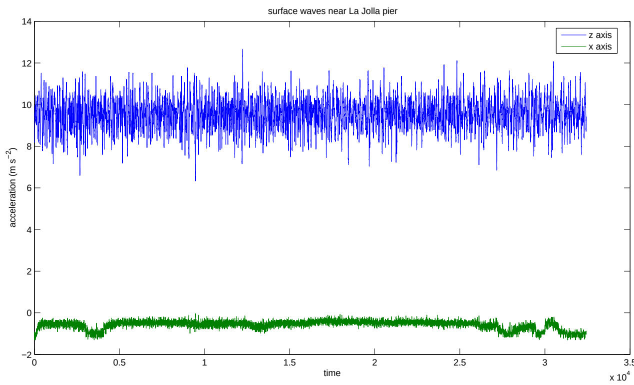

Figure 1: Time series of vertical acceleration and x-axis accleration for free-floating accelerometer near Scripps pier.

which is not a terribly informative result. When it's done properly, coherence measures how well different segments of x and y show the same type of relationship at a given frequency.

The phase $\phi(\omega_k) = \tan^{-1}(-q(\omega_k)/c(\omega_k))$ tells us the timing difference between the two time series. If $\phi = 0$, changes in x and y happen at the same time. If $\phi = \pi$, then x is at a peak when y is at a trough. And a value of $\phi = \pi/2$ or $\phi = -\pi/2$ tells us that the records are a quarter cycle different.

How much confidence do we have in our results? For the coherence, we require that the squared coherence exceed:

$$\beta = 1 - \alpha^{1/(n_d - 1)} \tag{18}$$

where α is a measure of the significance level. If $\alpha = 0.05$ that means that there is less than a 5% chance that random noise could have produced a coherence as high as the observed value. The number of data segments used is n_d.

The phase error can seem a little murky. Formally, the uncertainty in the phase is

$$\delta_\phi = \sin^{-1}\left[t_{\alpha,2n_d}\sqrt{\frac{1 - C_{xy}^2}{2n_d C_{xy}^2}}\right] \tag{19}$$

where $t_{\alpha,2n_d}$ is the "Student t distribution".

Example: Coherence and Wave Spectra

So let's see whether surfboard acceleration measurements show any signs of coherence. We'll start by comparing vertical and horizontal accelerations of the free floating accelerometer, as shown in Figure 1. These two records have rather different spectra as shown in Figure 2. The two records are coherent, as shown in Figure 3 with a phase difference of roughly π radians, implying that they are 180° out of phase, at least at the frequencies at which they are actually coherent. In contrast, the vertical acceleration for the free floating accelerometer is not coherent with vertical acceleration from the shortboard.

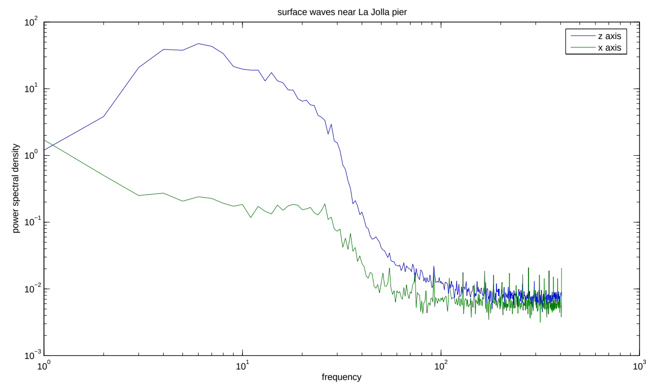

Figure 2: Spectra for vertical and x acceleration of free-floating accelerometer near Scripps pier.

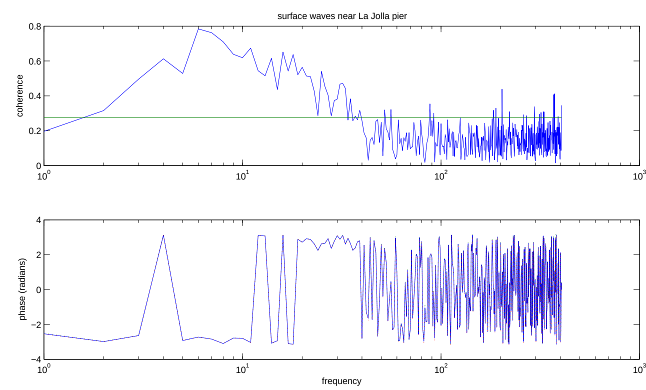

Figure 3: (top) Coherence of vertical and x acceleration of free-floating accelerometer near Scripps pier. (bottom) Phase difference between vertical and x acceleration components.

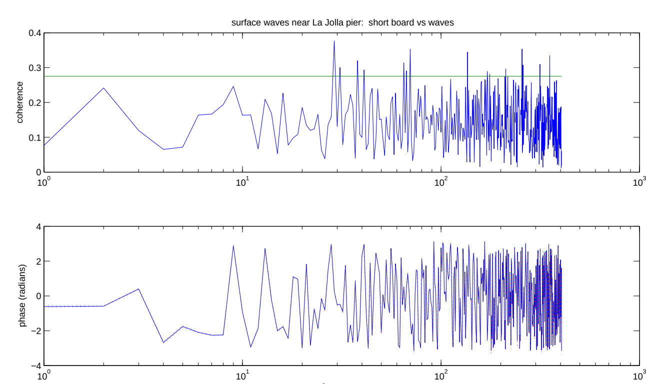

Figure 4: (top) Coherence of vertical acceleration of free-floating accelerometer versus shortboard accelerometer near Scripps pier. (bottom) Phase difference.