Lecture 16:

Recap

We've looked at a bunch of strategies recently: frequency/wavenumber spectra, autocovariance methods, variance preserving spectra, coherence. Today's a day to stop and take stock of what we've learned, answer some questions, and think about how to interpret results.

The first question is addresses a trait that you might have noticed in frequency-wavenumber spectra for long records, for example the record in problem set 7. I usually emphasize that error bars should be consistent with high-frequency (or high-wavenumber) point-to-point variability. But sometimes we see that spectra appear to broaden, to be wider than the error bars at high wavenumbers or high frequencies. Plus we know that real data have some traits that aren't necessarily characteristic of Gaussian noise. Let's test some of the traits of real data:

Real data are reported with a finite number of digits.

No one wants to pay to store or transmit data with 15 digits accuracy if the sensor is only accurate to a couple of decimal places. For example, temperature might vary continuously, but the archives might report it to only a couple of decimal places. What does this truncation do to spectra? To test this, we can create a continuous synthetic red data set, and compare the spectrum from the continuous data set with the spectrum computed from the same data set, rounded to the nearest integer:

```
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*cumsum(randn(5000,1))+cos(2*pi/lambda*V*time);

% start with a segmented data approach:
xx=[reshape(x,500,10) reshape(x(251:4750),500,9)];
fxx=fft(detrend(xx).*(hanning(500)*ones(1,19)));
sxx=abs(fxx(1:251,:)).^2 /500;

xx2=round(xx);
fxx2=fft(detrend(xx2).*(hanning(500)*ones(1,19)));
sxx2=abs(fxx2(1:251,:)).^2 /500;
loglog(0:250,2*mean(sxx,2),0:250,2*mean(sxx2,2))

set(gca,'FontSize',14)
xlabel('frequency','FontSize',14)
ylabel('spectral density (m^2 s^{-2}/[unit frequency])','FontSize',14)
legend('baseline red spectrum','data with truncated digits')
```

You'll see that the baseline spectrum is red, while the truncated data set is white at high frequencies. The rounding process effectively adds noise to the high-frequency part of the spectrum.

What if we add noisy outliers to the data?

Real data, such as the temperature record that you used for problem set #7 can have lots of spurious outliers, implying a non-Gaussian distribution. What do those do to the overall spectrum. To test this, we work from the same data set as in the first example, but add larger amplitude noise to about 10% of the points.

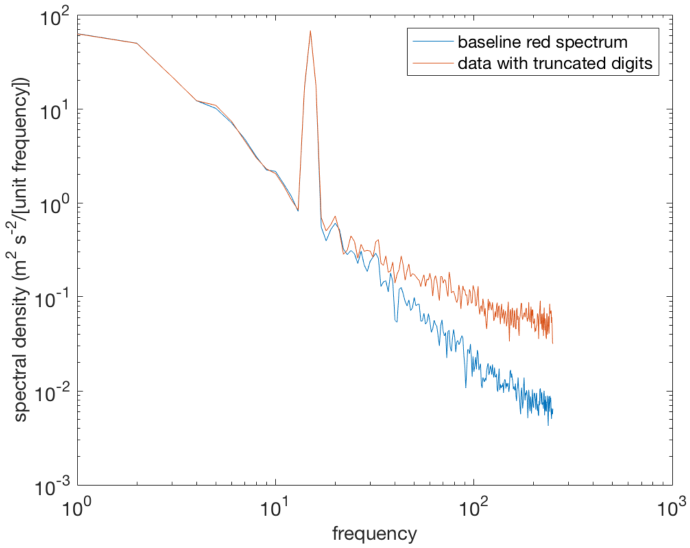

Figure 1: Spectra from auto-regressive red spectrum with cosine, and from the same signal truncated to integer values only. The loss of significant digits effectively adds a noise floor to the signal.

```
z0=ceil(rand(500,1)*5000);
y=x;
y(z0)=y(z0)+5*randn(500,1);

yy=[reshape(y,500,10) reshape(y(251:4750),500,9)];
fyy=fft(detrend(yy).*(hanning(500)*ones(1,19)));
syy=abs(fyy(1:251,:)).^2 /500;

figure(1)
plot(1:5000,y,1:5000,x)
shg

figure(2)
loglog(0:250,2*mean(sxx,2),0:250,2*mean(syy,2))
set(gca,'FontSize',14)
xlabel('frequency','FontSize',14)
ylabel('spectral density (m^2 s^{-2}/[unit frequency])','FontSize',14)
legend('baseline red spectrum','data with outliers added')
shg
```

You'll see that the addition of a few white noise outliers whitens the high-frequency portion of the spectrum. Neither of these examples explains the broadening that we sometimes notice in large spectral calculations, though they provide good warnings about traits in data that might influence the high-wavenumber or high-frequency portion of a spectrum.

What if we extend the record?

Finally let's consider what happens if we have a very long record to Fourier transform. We'll

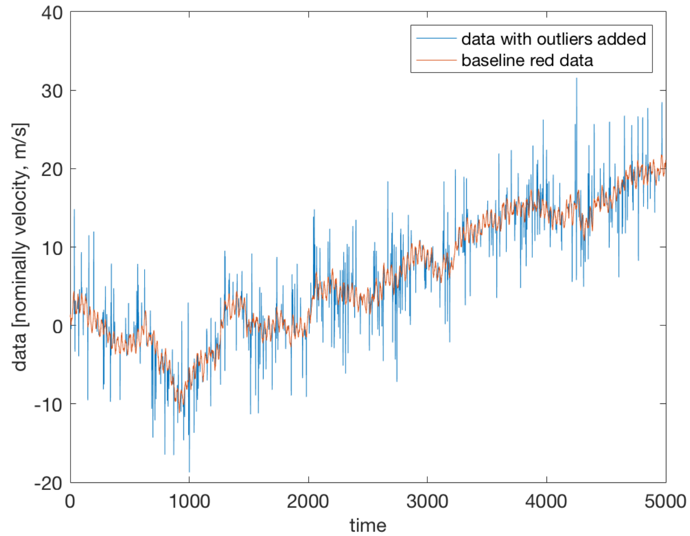

Figure 2: Auto-regressive red data (red) with one sinusoidal peak, and same signal with large amplitude noise (outliers) added to 10% of the data points (blue).

use the same structure for he data that we used before but make the record a couple of orders of magnitude larger.

With a long record, we have choices in how we determine the number of segments to use. Here we'll consider 19 segments and 199 segments:

```
x=n2s*cumsum(randn(500000,1));
xx=[reshape(x,50000,10) reshape(x(25001:475000),50000,9)];
fxx=fft(detrend(xx).*(hanning(50000)*ones(1,19)));
sxx=abs(fxx(1:25001,:)).^2;

xx2=[reshape(x,5000,100) reshape(x(2501:497500),5000,99)];
fxx2=fft(detrend(xx2).*(hanning(5000)*ones(1,199)));
sxx2=abs(fxx2(1:2501,:)).^2;

loglog((0:25000)/50000,2*mean(sxx,2)/50000,(0:2500)/5000,2*mean(sxx2,2)/5000
```

These spectra show a broadening at high frequencies that is smaller in amplitude, but still present when we average more segments. What does this mean. You can stretch the spectrum out in linear space in frequency:

```
semilogy((0:25000)/50000,2*mean(sxx,2)/50000,(0:2500)/5000,2*mean(sxx2,2)/50
```

In this case, you'll see that the spectrum appears to be consistent at high frequencies. In essence the broadening occurs because high frequencies are so tightly packed together that our eye sees the outliers. The point-to-point jitter that we're able to see in the spectral estimate is the 99% confidence limit, or perhaps the 99.9% confidence limit instead of the 95% confidence limit.

Coherence Examples

In the slides for this lecture are 5 or 6 coherence examples to class today to pass around and discuss. Your challenge within groups of 2-3 is to unravel the basics of these.

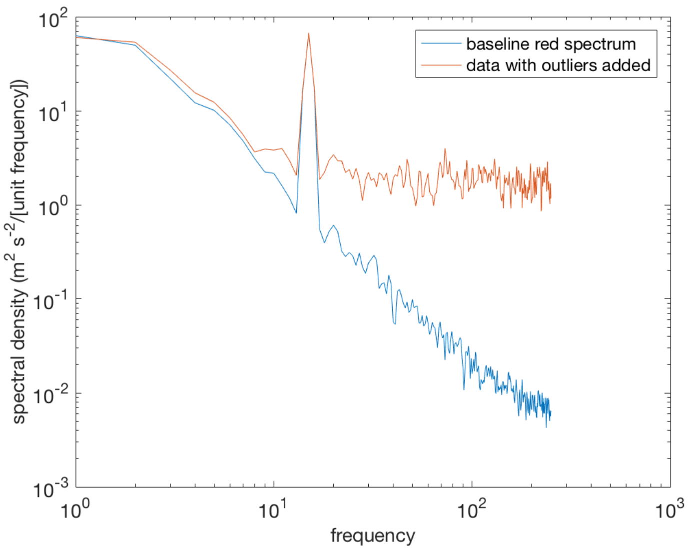

Figure 3: Spectra from auto-regressive red spectrum with one sinusoidal peak, and from the same signal with large amplitude noise (outliers) added to 10% of the data points. The addition of white noise to some of the data points also effectively adds a noise floor to the signal.

More generally, we should talk about what makes a good final project focused on coherence. Please check with me if you have questions.

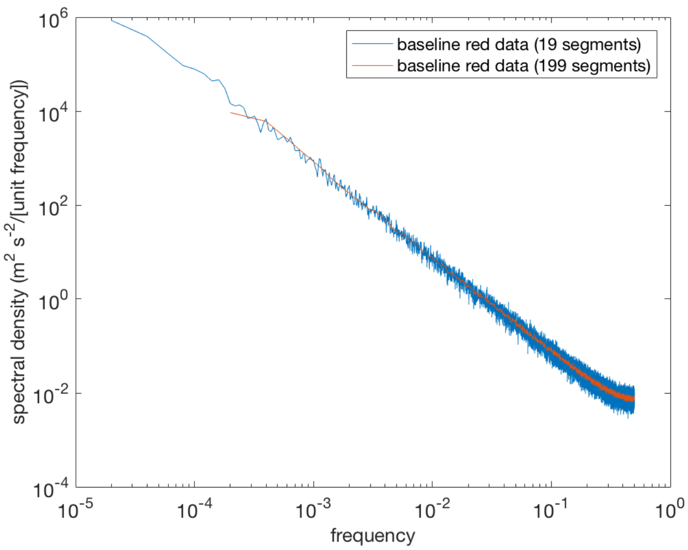

Figure 4: Spectra from auto-regressive red spectrum with 19 segments, and from the same signal cut into shorter segments to produce 199 segments. In both cases, the record is long enough that the spectra appear to show higher noise levels at high frequencies. This is an artifact of the high-density frequencies in log-log space.