Lecture 17:

Recap

We've now spent some time looking closely at uncertainties for coherence. Can we think about coherence in a different way?

Noise and signals:

The signal that we started with had two related signals:

$$y(t) = c_1 x(t) + n(t), \tag{1}$$

where $n(t)$ is noise. Can we determine c_1 and $n(t)$ from the cross-spectrum? The cross spectrum is:

$$\begin{aligned} S_{XY}(f) &= \frac{\langle X^*(f)Y(f)\rangle}{T} & (2)\\ &= \frac{c_1\langle X^*(f)X(f)\rangle + \langle X^*(f)N(f)\rangle}{T} & (3)\\ &= c_1 S_{XX} & (4) \end{aligned}$$

The squared coherence is therefore:

$$\text{Coh}(f)^2 = \gamma^2_{XY} = \frac{|S_{XY}|^2}{S_{XX}S_{YY}} \tag{5}$$

So what is S_{YY}? We can compute the spectrum of y:

$$\begin{aligned} S_{YY} &= \frac{[c_1^2\langle X^*X\rangle + c_1(\langle X^*N\rangle + \langle XN^*\rangle + \langle N^*N\rangle]}{T} & (6)\\ &= c_1^2 S_{XX} + S_{NN}, & (7) \end{aligned}$$

where X and N are assumed to be uncorrelated. Then the squared coherence is:

$$\begin{aligned} \gamma^2_{XY}(f) &= \frac{c_1^2 S_{XX}(f)^2}{S_{XX}(c_1^2 S_{XX} + S_{NN})} & (8)\\ &= \frac{1}{\left(1 + \frac{S_{NN}(f)}{c_1^2 S_{XX}(f)}\right)} & (9) \end{aligned}$$

The final term in the denominator is a measure of the noise-to-signal ratio. (In our example, we imposed it from the beginning.) So if we knew a lot about the causal relations between our records, we could use the coherence to extract a measure of the noise-to-signal ratio.

What if there's a time lag? For example, if temperature today determines ice cover tomorrow. Or if rainfall today determines crop productivity next season? Then our initial model might be of the form:

$$y(t) = c_1 x(t+\tau) + n(t), \tag{10}$$

Then we have to build in a phase shift:

$$Y(f) = c_1 X(f) e^{i2\pi f\tau} + N(f) \tag{11}$$

and the cross-spectrum becomes:

$$S_{XY}(f) = \frac{\langle X^*(f)Y(f)\rangle}{T} \quad (12)$$
$$= \frac{c_1\langle X^*(f)X(f)\rangle e^{i2\pi f\tau} + \langle X^*(f)N(f)\rangle}{T} \quad (13)$$
$$= c_1 S_{XX} e^{i2\pi f\tau} \quad (14)$$

In this case, the phase tells us the time lag. In this case, if we solve for phase,

$$\tan(\phi) = \sin(2\pi f\tau)/\cos(2\pi f\tau), \quad (15)$$

which means that $\phi = 2\pi f\tau$, so phase changes linearly with frequency.

In this case, the squared coherence is:

$$\gamma^2_{XY}(f) = \frac{c_1 S_{XX}(f)e^{i\phi}\, c_1 S_{XX}(f)e^{-i\phi}}{S_{XX}(c_1^2 S_{XX} + S_{NN})} \quad (16)$$
$$= \frac{c_1^2 S_{XX}(f)}{c_1^2 S_{XX} + S_{NN}} \quad (17)$$
$$= \frac{1}{\left(1 + \frac{S_{NN}(f)}{c_1^2 S_{XX}(f)}\right)} \quad (18)$$

so the time lag influences the phase but not the coherence.

Now can we put this to work for a real time series? For the midterm problem set you looked at SST and air temperature at a mooring. Here let's consider a different mooring right on the Equator, at 165°E. What sorts of causal relationships could you hypothesize between those two records?

```
% read in the data:
time=ncread('met0n165e_10m.cdf','time');
Ta=ncread('met0n165e_10m.cdf','AT_21');
SST=ncread('met0n165e_10m.cdf','T_25');
QTa=ncread('met0n165e_10m.cdf','QAT_5021');
QSST=ncread('met0n165e_10m.cdf','QT_5025');
Ta(find(Ta>1.e34 | QTa==0 | QTa>=4 | Ta>36))=NaN;
SST(find(SST>1.e34 | QSST==0 | QSST>=4))=NaN;

% reformat to allow segments
yy=345000:525000;
Tause=Ta(yy);
Tause(find(isnan(Tause)))=nanmean(Tause);
SSTuse=SST(yy);
SSTuse(find(isnan(SSTuse)))=nanmean(SSTuse);
Ta_matrix=[reshape(Tause(1:180000),10000,18) ...
   reshape(Tause(5001:175000),10000,17)];
SST_matrix=[reshape(SSTuse(1:180000),10000,18) ...
    reshape(SSTuse(5001:175000),10000,17)];

% compute Fourier transform
```

```
fTa_matrix=fft(detrend(Ta_matrix).*(hanning(10000)*ones(1,35)));
fSST_matrix=fft(detrend(SST_matrix).*(hanning(10000)*ones(1,35)));
amp_Ta=abs(fTa_matrix(1:5001,:)).^2/10000;
amp_Ta(2:4999,:)=2*amp_Ta(2:4999,:);
amp_SST=abs(fSST_matrix(1:5001,:)).^2/10000;
amp_SST(2:4999,:)=2*amp_SST(2:4999,:);

C_TT=conj(fTa_matrix(1:5001,:)).*fSST_matrix(1:5001,:)/10000;
C_TT(2:4999,:)=2*C_TT(2:4999,:);
% here's the correct formulation
cc=abs(mean(C_TT,2))./sqrt((mean(amp_SST,2).*mean(amp_Ta,2)));
% you might imagine reversing the order of operations---cc2 is wrong
cc2=mean(abs(C_TT),2)./sqrt(mean(amp_SST,2).*mean(amp_Ta,2));

threshold = sqrt(1-.05^(1/34));

semilogx((0:5000)/10000*(6*24),cc,...
   [1/10000*6*24 5000/10000*6*24],[threshold threshold])

nd=35;
phase=atan2(-imag(mean(C_TT,2)),real(mean(C_TT,2)));
delta_phase = sqrt((1-cc.^2)./(abs(cc).^2*2*nd));
figure(2)
semilogx((0:5000)/10000*(6*24),[phase phase+delta_phase phase-delta_phase])
axis([1/10000*6*24 .5*6*24 -pi pi])
```

Coherence: The autocovariance perspective

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.

Last time we wrote the cross-spectrum for x and y:

$$\hat{S}_{XY}(f_m) = \frac{\langle X_m^* Y_m \rangle}{T} \tag{19}$$

Just as we considered spectra as the Fourier transform of the autocovariance, we can now think about the Fourier transform of the lagged co-variance.

$$R_{xy}(\tau) = \frac{1}{2T} \int_{-T}^{T} x^*(t) y(t+\tau)\, dt. \tag{20}$$

We can rewrite this:

$$
\begin{aligned}
R_{xy}(\tau) &= \frac{1}{2T}\int_{-T}^{T}\sum_{n=-\infty}^{\infty} X_n^* e^{-i2\pi f_n t}\sum_{m=-\infty}^{\infty} Y_m e^{i2\pi f_m(t+\tau)}\,dt && (21)\\
&= \frac{1}{2T}\sum_{n=-\infty}^{\infty} X_n^* \sum_{m=-\infty}^{\infty} Y_m e^{i2\pi f_m\tau}\int_{-T}^{T} e^{i2\pi(f_m-f_n)t}\,dt && (22)\\
&= \sum_{n=-\infty}^{\infty} X_n^* \sum_{m=-\infty}^{\infty} Y_m e^{i2\pi f_m\tau}\delta_{nm} && (23)\\
&= \sum_{n=-\infty}^{\infty} X_n^* Y_n e^{i2\pi f_n\tau} && (24)\\
&= T\sum_{n=-\infty}^{\infty} S_{XY} e^{i2\pi f_n\tau}, && (25)
\end{aligned}
$$

where we used the Kronecker delta δ_{nm} to extract only frequencies for which $n = m$, since all other modes are orthogonal. The result tells us that the lagged covariance is the inverse Fourier transform of the cross spectrum. In other words,

$$S_{XY}(f_n) = \int_{-T}^{T} R_{xy}(\tau)e^{-i2\pi f_n\tau}\,d\tau = \frac{X_n^* Y_n}{T} \tag{26}$$

Thus we could determine the cross-spectrum from the lagged covariance.

We can even extend this to consider frequency/wavenumber spectra. If we have a cross-covariance:

$$R(\mathbf{r},\tau) = \langle \mathbf{a}^*(\mathbf{x},\mathbf{t})\mathbf{a}(\mathbf{x}+\mathbf{r},\mathbf{t}+\tau)\rangle. \tag{27}$$

If we've deployed a set of instruments, an array of current meters for example, our sampling density in time might be wildly different than our sampling density in space. We would like to estimate the spectrum:

$$S_{AA}(\mathbf{k},f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(\mathbf{r},\tau)\mathbf{e}^{-\mathbf{ik\cdot r}-\mathbf{i2\pi f}\tau}\mathbf{drd}\tau \tag{28}$$

With sparse data, it's hard to justify integrating over the space domain. Suppose we could integrate only over the time domain to obtain the cross-spectrum in the frequency domain:

$$C(\mathbf{r},f) = \int_{-\infty}^{\infty} R(\mathbf{r},\tau)\mathbf{e}^{-\mathbf{i2\pi f}\tau}\mathbf{d}\tau \tag{29}$$

This leaves us sitting on the fence. We don't know the frequency wavenumber spectrum (S_{AA}), but we know the cross-spectrum in the space-frequency domain, so we might be able to infer S_{AA} with some inferences about variability in space/wavenumber. For example we could discretize in space and compute the Riemann sum.

$$\hat{S}_{AA}(\mathbf{k},f) = \mathcal{R}\sum_{lags} C(\mathbf{r}_{ij},f)e^{-i\mathbf{k}\cdot\mathbf{r}}\Delta\mathbf{r} \tag{30}$$

Rewriting to represent C as the inverse Fourier transform of the spectrum:

$$\hat{S}_{AA}(\mathbf{k},f) = \mathcal{R}\sum_{lags}\left[\int_{-\infty}^{\infty} S_{AA}(\mathbf{l},\mathbf{f})\mathbf{e}^{\mathbf{il\cdot r}}\,\mathbf{dl}\right]e^{-i\mathbf{k}\cdot\mathbf{r}}\Delta\mathbf{r} \tag{31}$$

Switching the order of integration, this becomes:

$$\hat{S}_{AA}(\mathbf{k}, f) = \left[\int_l S_{AA}(\mathbf{l}, \mathbf{f})\mathbf{W}(\mathbf{k}-\mathbf{l})\,\mathbf{dl}\right]. \tag{32}$$

where W is the spectral window associated with the array design. It's

$$W(\mathbf{k}-\mathbf{l}) = \mathcal{R}\sum_{\mathbf{lags}} \mathbf{e}^{\mathbf{i}(\mathbf{k}-\mathbf{l})\cdot\mathbf{r}}\boldsymbol{\Delta}\mathbf{r}. \tag{33}$$

If we knew the cross-spectrum for all lags, then $W = \delta(\mathbf{k}-\mathbf{l})$.

Decibels and Powers of 10

Decibels ("db", not to be confused with decibars) quantify power or variance. We often focus on orders of magnitude. But power in decibels is on a $\log_{10}$ scale, with a factor of ten normalization.

$$P_{db} = 10\log_{10}(P/P_0), \tag{34}$$

where P_0 is a reference level of power: Variance is a squared quantity so

P/P_0	P relative to P_0 (db)
1000	30
10	10
2	3
0.5	-3
0.1	-10

$$P_{db} = 20\log_{10}(V/V_0). \tag{35}$$

By convention dB is used for sound pressure and db for everything else.

Transfer function: If we want to look at relative sizes, we can look at the transfer function:

$$\hat{H}_{xy}(f) = \frac{\hat{G}_{xy}(f)}{\hat{G}_{xx}(f)}, \tag{36}$$

which provides a (complex-numbered) recipe for mapping from x to y.