Lecture 18:

Reading: Bendat and Piersol, Ch. 6.1

Recap

Last time we worked through a coherence example in detail and looked at links between coherence and co-variance. We talked about definitions of decibels. And we took a fleeting look at transfer functions. Now we just have a few more details to wrap up.

Transfer functions (or gain functions): a proper example

Last time I gave you a very fleeting and inadequate view of a transfer function example, so I want to work through this more carefully today, before taking a more holistic view at a case study.

As we noted last time, if we want to look at relative sizes, we can look at the transfer function (also known as the gain function):

$$\hat{H}_{xy}(f) = \frac{\hat{G}_{xy}(f)}{\hat{G}_{xx}(f)}, \tag{1}$$

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

Formally, we talk about the transfer function when we think about constructing a linear system:

$$\mathcal{L}(y(t)) = x(t) \tag{2}$$

If $\mathcal{L}$ is a linear operator, then we could think of this relationship as a convolution:

$$y_t = \int_{-\infty}^{\infty} h(u)x(t-u)\,du \tag{3}$$

or if we Fourier transform, this would state:

$$Y(f) = H(f)X(f). \tag{4}$$

Consider it this way. Suppose

$$x(t) = \frac{d^2y}{dt^2} + \alpha\frac{dy}{dt} + \beta y \tag{5}$$

Then by Fourier transforming, we have:

$$\begin{aligned} X(f) &= -f^2Y(f) + i\alpha fY(f) + \beta Y(f) && (6)\\ &= Y(f)\left[\beta - f^2 + i\alpha f\right] && (7)\end{aligned}$$

so

$$Y(f) = \frac{1}{[\beta - f^2 + i\alpha f]}X(f) \tag{8}$$

and

$$H(f) = \frac{1}{[\beta - f^2 + i\alpha f]} \tag{9}$$

This is a nice framework for solving differential equations, but can we use it to gain insights into our data as well? First some rules:

1. Linearity: If a given linear system has an input $x_1(t)$ which corresponds to an output $y_1(t)$, and input $x_2(t)$ corresponds to output $y_2(t)$m then a summed input $x(t) = \alpha x_1(t) + \beta x_2(t)$, will produce an output $y(t) = \alpha y_1(t) + \beta y_2(t)$.

2. Time invariance: If an input is delayed in time by τ, then the output is as well: If $x(t) \rightarrow x(t+\tau)$, then $y(t) \rightarrow y(t+\tau)$.

3. Causality: If h(t) represents an impulse, then it should be zero for $t < 0$. A response cannot occur before the forcing.

4. Sequential application: If the output of one linear system is an input to a second system, then the frequency response is
$$H_{12}(f) = H_1(f) \cdot H_2(f) \tag{10}$$

So suppose we measure $y(t)$ and $x(t)$. Can we determine h or H? We know that
$$Y(f) = H(f)X(f) \tag{11}$$

Let's multiply both sides of the equation by the complex conjugate of X to form the cross-spectrum:
$$\frac{Y(f)X^*(f)}{T} = H(f)\frac{X(f)X^*(f)}{T} \tag{12}$$

This becomes
$$G_{xy}(f) = H(f)G_{xx}(f) \tag{13}$$

so
$$H(f) = \frac{G_{xy}(f)}{G_{xx}(f)} \tag{14}$$

Salinity spiking examples

The slides show some examples of 'salinity spiking', which results from the temperature and conductivity sensors having different response times. If we assume that simultaneous measurements represent the same water sample, when in reality they don't, we end up computing:
$$S = S(T(t), C(t - \Delta t)). \tag{15}$$

If temperature and salinity are constant, this doesn't pose a problem, but when T or C change abruptly, this can lead to very odd results. How can we use the transfer function (or coherence) approach to examine this (even if we don't know the actual response characteristics of the sensor). Conductivity is strongly dependent on temperature, and we have to remove the temperature effect to determine salinity. But the faster response times of conductivity sensors relative to temperature sensors are a source of confusion. Here's a basic procedure.

1. Identify a segment of the water column in which temperature and conductivity should be well behaved, with fluctuations due to temperature onl. (This isn't essential, but it will give us a good shot at unraveling the sensor response time issues that lead to salinity or density spiking.) We'll identify the true values as T and C, and the measured values as $\hat{T}$ and $\hat{C}$.

2. Collect a lot of profiles of data.

3. Now treat this as a linear system:

$$\begin{aligned} \hat{T}(k) &= H_T(k)T(k) \quad (16) \\ \hat{C}(k) &= H_C(k)T(k) \quad (17) \\ & \quad (18) \end{aligned}$$

Here $H_T(k)$ is the spatial/frequency response of the temperature sensor, and $H_C(k)$ is the spatial/frequency response of the conductivity sensor. The use of $T(k)$ in the conductivity equation might seem a little crazy, but it's really important, since we're asserting that salinity is unimportant in our (hypothetical) study region.

4. Compute cross spectra:

$$\hat{G}_{TT}(k) = \frac{\langle \hat{T}^*(k)\hat{T}(k)\rangle}{N\Delta t} \tag{19}$$

$$\hat{G}_{CC}(k) = \frac{\langle \hat{C}^*(k)\hat{C}(k)\rangle}{N\Delta t} \tag{20}$$

$$\hat{G}_{TC}(k) = \frac{\langle \hat{T}^*(k)\hat{C}(k)\rangle}{N\Delta t} \tag{21}$$

$$\tag{22}$$

Then if we substitute in the expessions linking the observed values to the true values we obtain:

$$\frac{\langle \hat{T}^*(k)\hat{C}(k)\rangle}{N\Delta t} = \frac{\langle (H_T(k)T(k))^* H_C(k)T(k)\rangle}{N\Delta t} \tag{23}$$

$$= [H_T^*(k)H_C(k)]\frac{\langle T^*(k)T(k)\rangle}{N\Delta t} \tag{24}$$

$$\hat{G}_{TC}(k) = [H_T^*(k)H_C(k)]G_{TT}(k) \tag{25}$$

The same applies for the temperature spectrum:

$$\hat{G}_{TT}(k) = \frac{\langle \hat{T}^*(k)\hat{T}(k)\rangle}{N\Delta t} \tag{26}$$

$$= [H_T^*(k)H_T(k)]G_{TT}(k) \tag{27}$$

So the ratio of these becomes:

$$\frac{\hat{G}_{TC}(k)}{\hat{G}_{TT}(k)} = \frac{H_T^*(k)H_C(k)}{|H_T|^2} = \frac{H_C(k)}{H_T(k)} \tag{28}$$

This means that even without knowing the response function H, we can compute the ratio of the response functions from the transfer function, the ratio of the cross-spectrum to the spectrum.

An analogous relationship also holds:

$$\frac{\hat{G}_{TC}(k)}{\hat{G}_{CC}(k)} = \frac{H_T^*(k)H_C(k)}{|H_C|^2} = \frac{H_T(k)}{H_C(k)} \tag{29}$$

5. Now use this information to correct the conductivity sensor to have the same response as the temperature sensor. Here we'll define our corrected conductivity as $\hat{\hat{C}}(k)$, and we want to understand its relationship with the observed temperature $\hat{T}(k)$ and the true temperature $T(k)$.

$$\hat{\hat{C}}(k) = \alpha\hat{T}(k) = \alpha H_T(k)T(k) \tag{30}$$

This means we need a correction of the form:

$$\hat{\hat{C}}(k) = \hat{C}(k) \cdot P(k) = \alpha \hat{T}(k), \tag{31}$$

and our task is to figure out $P(k)$. We can also write:

$$\hat{C}(k) \cdot P(k) = H_C(k) T(k) P(k) \tag{32}$$

so putting this together:

$$\alpha H_T(k) T(k) = H_C(k) T(k) P(k) \tag{33}$$

Thus

$$P(k) = \frac{\alpha H_T(k)}{H_C(k)} = \alpha \frac{\hat{G}_{TC}(k)}{\hat{G}_{CC}(k)} \tag{34}$$

where α is real. So we do a bit of curve fitting to optimize our correction.

A typical correction might allow for errors both in the response time and a direct time lag:

$$\tau \frac{d\hat{T}(t)}{dt} + \hat{T}(t) = T(t - L) \tag{35}$$

(from Giles and McDougall, Deep-Sea Research, 1986) and this suggests corrections both in the frequency and time domain, either by minimizing phase differences or by maximizing correlation. We can Fourier transform this to find:

$$-i2\pi f \tau \mathcal{F}(\hat{\mathcal{T}}) + \mathcal{F}(\hat{\mathcal{T}}) = \mathcal{F}(\mathcal{T})\rceil^{-\rangle\in\pi\{\mathcal{L}}, \tag{36}$$

implying that

$$(1 - i2\pi f\tau) e^{i2\pi f L} \mathcal{F}(\hat{\mathcal{T}}) = \mathcal{F}(\mathcal{T}). \tag{37}$$

Since conductivity has the fast response, one strategy is to treat $\hat{C}$ as behaving like the true temperature T. So hypothetically:

$$\begin{aligned} \hat{T}(k) &= H_T(k) T(k) = \frac{1}{(1 - i2\pi f\tau) e^{i2\pi f L}} T(k) && (38) \\ \hat{C}(k) &= H_C(k) T(k) = T(k) && (39) \\ & && (40) \end{aligned}$$

implying a correction P(k) of the form:

$$P(k) = \frac{\alpha H_T(k)}{H_C(k)} = \alpha \frac{1}{(1 - i2\pi f\tau) e^{i2\pi f L}}. \tag{41}$$

Noise and the transfer function

Last week we looked at noise in coherence, but let's take a quick look at what noise does to the transfer function. Suppose:

$$Y(f) = X(f) H(f) + N(f) \tag{42}$$

If we multiply through by X^*:

$$\langle X^* Y \rangle = \langle X^* X \rangle H(f) + \langle X^* N \rangle \tag{43}$$

Since the signal is uncorrelated with noise, this still gives us

$$H(f) = \frac{G_{xy}(f)}{G_{xx}(f)}, \tag{44}$$

so noise appears to have no impact on the results, but is it all so rosy?

Alternatively, we might imagine that our forcing x is noisy, so that

$$Y(f) = [X(f) + N(f)]\, H(f) \tag{45}$$

In doing this, we assume that the noise associated with X is uncorrelated with the signal Y—in other wods, we assume that the response $y(t)$ should be responding to $x(t)$, but we've mismeasured the forcing as $x(t) + n(t)$. Then:

$$\langle (X+N)^* Y \rangle = \langle (X+N)^*(X+N) \rangle H(f) \tag{46}$$

And since y and n are uncorrelated in this formulation,

$$\hat{H}(f) = \frac{G_{xy}(f)}{G_{xx}(f) + G_{nn}(f)}, \tag{47}$$

which is biased low relative to the true response. (But the phase is unbiased.) This is analogous to the noise-related formulation that we looked at for coherence.

Putting it all together

In this course, we've looked at a broad range of strategies for analyzing time series. Can we make some decisions about how we might plan an experiment and analyze our data?

Let's consider a central problem of physical oceanography. Can we evaluate the ocean response to wind:

$$\begin{aligned} \frac{\partial u}{\partial t} + u \cdot \nabla \mathbf{u} - fv &= -\frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{1}{\rho}\frac{\partial \tau^x}{\partial z} && (48) \\ \frac{\partial v}{\partial t} + v \cdot \nabla \mathbf{u} + fu &= -\frac{1}{\rho}\frac{\partial p}{\partial y} + \frac{1}{\rho}\frac{\partial \tau^y}{\partial z} && (49) \end{aligned}$$

Suppose we cross out a few terms. What do we need to measure to evaluate whether we've crossed out the right terms? Can we show that one of these is a reasonable approximation?

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{\rho}\frac{\partial \tau^x}{\partial z} && (50) \\ -fv &= \frac{1}{\rho}\frac{\partial \tau^x}{\partial z} && (51) \\ -fv + -\frac{1}{\rho}\frac{\partial p}{\partial x} &= \frac{1}{\rho}\frac{\partial \tau^x}{\partial z} && (52) \\ & && (53) \end{aligned}$$

Methods we've explored in this class include:

1. Basic statistics: means, standard deviation, variance, standard error
2. Probability density function

3. Least-squares fitting
4. One-dimensional spectra (with windowing and uncertainties)
5. Two-dimensional spectra
6. Monte Carlo methods for evaluating confidence limits
7. Coherence

Intangibles

Besides the formal topics for which you've done problem sets, this class has aimed to start you thinking more like a data analyst. This has thrown you into the thorny world of real data problems, and I'm immensely grateful to you for your persistence. Some life lessons from this class:

1. In science, we favor evidence-based decision making over shoot-from-the-hip opinons. Data analysis gives you a set of tools for this.
2. Your good judgement matters in deciding how to approach a data analysis problem. You should always ask yourself how your understanding of the physics can inform your approach.
3. Even the rigors of the peer review process cannot guarantee the fidelity of published sources of information. Be skeptical and inquisitive.
4. You have the tools at your disposal to address your skepticism. Fake data and Monte Carlo methods are always an option.
5. Many questions have not been answered carefully, and there is room for you to make significant contributions. (I think the pier station staff are still looking for the perfect means to validate the methodology used for the pier samples.)
6. Just because something appears to be significant at the 95% level doesn't guarantee that it is a robust signal.
7. If your results are wildly dependent on the details of your methodology, that might mean that paying attention to methodology matters, but it also could be a warning sign that you're trying to identify a signal that is more wishful thinking than real signal.
8. The methods that we apply to real data can be exactly analogous to problems that we've done in this class, or they can be surprisingly divergent.