Cross Spectrum Analysis

Basic idea: Determine the relationship between two time series as a function of frequency.

2 Questions

1) Are there periodicities that are related?
2) what is the phase relationship between related periodicities?

Recall how we can decompose the time series (or any other data) using Harmonic Analysis. Now we do it for 2 pieces of data series instead of just one.

$$x = \bar{x} + \sum_{k=1}^{N/2-1} \left(A_{xk} \cos\left(\frac{2\pi k t}{T}\right) + B_{xk} \sin\left(\frac{2\pi k t}{T}\right)\right)$$

$$+ A_{xN/2} \cos\left(\frac{\pi N t}{T}\right)$$

$$y = \bar{y} + \sum_{k=1}^{N/2-1} \left(A_{yk} \cos\left(\frac{2\pi k t}{T}\right) + B_{yk} \left(\frac{2\pi k t}{T}\right)\right)$$

$$+ A_{yN/2} \cos\left(\frac{\pi N t}{T}\right)$$

Where the same definitions hold as used before in the harmonic analysis discussion.

The covariance $(\overline{x'y'})$ is just the sum of the contributions from the various frequencies

Similar to the explained variance, for an individual k

$$Cov_k = \frac{A_{xk}A_{yk} + B_{xk}B_{yk}}{2} = Co(k)$$

→ Cospectrum of x and y.

Then the total covariance is:

$$\overline{x'y'} = \frac{1}{2}\sum_{k=1}^{N/2-1}(A_{xk}A_{yk} + B_{xk}B_{yk}) + A_{xN/2}A_{yN/2}$$

$$\overline{x'y'} = \sum_{k=1}^{N/2} Co(k)$$

The cospectrum answers the <u>first</u> question (related periodicity), but <u>does</u> <u>not</u> address the second (phase relationships) So we're not done quite yet!

From this point on, the most straightforward discussion is found in Stull's book on boundary layer meteorology (not Hartmann notes)

If we define the total spectral power in a given band $C_K^2/2$.

$$\frac{C_K^2}{2} = \frac{A_K^2 + B_K^2}{2}$$

This can be represented via complex number.

$$\frac{C_K^2}{2} = \frac{(A_K + iB_K)(A_K - iB_K)}{2}$$

Multiply $(A_K + iB_K)/2$ by it's complex conjugate.

Now, what if we do the same for the two time series (x & y)

$$\frac{C_{xy_K}^2}{2} = \frac{(A_{xK} + iB_{xK})(A_{yK} - iB_{yK})}{2}$$

$$\frac{C_{xyk}^2}{2} = \frac{A_{xk}A_{yk} + B_{xk}B_{yk} + iA_{yk}B_{xk} - iA_{xk}B_{yk}}{2}$$

This has 2 parts:

Cospectrum: Real part (Co)

$(A_{xk}A_{yk} + B_{xk}B_{yk})/2$

→ The in-phase signal

Quadrature spectrum: Imaginary part (Q)

$(A_{yk}B_{xk} - A_{xk}B_{yk})/2$

→ The out of phase signal

So the cospectrum is only part of what we need to consider when doing cospectral analysis. Completely miss the out of phase signal otherwise!

Spectral coherence (Coh^2)

Basic idea: A normalized amplitude, real number between 0 and 1, like a frequency dependent correlation coefficient.

To compute: Take the term C_{xyk}^2 and multiply by its complex conjugate; then divide by spectral power of x and ~~[illegible]~~ spectral power of y.

Simplier way to think about it: Sum of the square of cospectrum and sq. of quadrature spectrum divided by spectral powers of x & y.

$$coh^2 = \frac{Co^2(k) + Q^2(k)}{\Phi_x(k)\,\Phi_y(k)}$$

Note similarity to the correlation coefficient

$$r^2 = \frac{(\overline{x'y'})^2}{\overline{x'^2}\;\overline{y'^2}}$$

Coherence Significance Levels

Hartmann provides a table with significance levels for spectral coherence.

These values are derived from the formula:

$$c^2 = 1 - \alpha^{\frac{1}{(n-1)}}$$

(Goodman 1957)

$\alpha = 1 - p$

p = significance level (e.g. 0.90, 0.95, 0.99)

n = degrees of freedom

Similar results are obtained if a Monte Carlo approach is used:

- Draw 2 series generated by random number generator.
- Compute the spectral coherence
- Repeat a large number of times, like 1000
- Pick out 90, 95, 99 % levels (like in the field significance test)

Additional notes - Cross spectrum analysis
Another useful measure is the phase spectrum

$$\tan \phi = Q/CO$$

Gives the phase difference between two time series that yields the greatest correlation for the given frequency.

Examples: Would be useful for looking at mountain waves, where it would be expected temp. variations would be out of phase with vertical velocity.

if two time series are unrelated, the coherence will decrease rapidly with increasing degrees of freedom (i.e. more spectral estimates or smoothing)

If related, the coherence will drop much more slowly as d.o.f. $\uparrow$.

Average the real and imaginary parts of the cross spectrum seperately before calculating mean coherence from multiple spectral estimates.

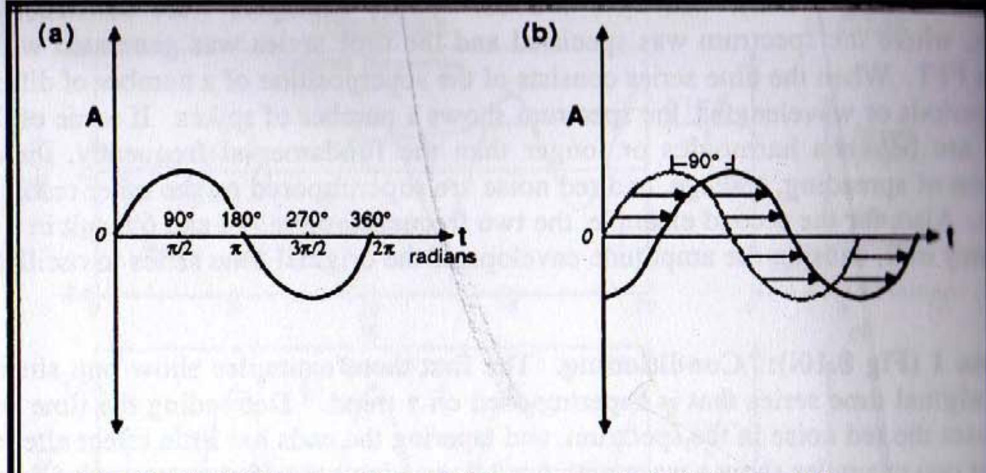

Fig. 8.11 (a) Phase angles and (b) example of a 90° phase shift.

$$A(k,n) \;=\; C_s(n)\cdot\sin\!\left(\frac{2\pi kn}{N}\right) \;+\; C_c(n)\cdot\cos\!\left(\frac{2\pi kn}{N}\right) \qquad (8.8.1b)$$

where $C_s = C\cdot\cos\Phi$ and $C_c = -C\cdot\sin\Phi$.

As shown in section 8.4.1 the Fourier transforms give the amplitudes of sine and cosine terms in the spectral decomposition of the original field. Thus, we can also interpret the spectra in terms of an amplitude and phase shift for waves of each frequency.

8.8.2 Cross Spectra

Define $G_A = |F_A(n)|^2$ as the unfolded spectral energy for variable A and frequency n. We can rewrite this definition as $G_A = F_A^*\cdot F_A$, where F_A^* is the complex conjugate of F_A, and where the dependence on n is still implied.

To demonstrate this last definition, let $F_A = F_{Ar} + i\cdot F_{Ai}$, where subscripts r and i denote real and imaginary parts respectively. Thus, the complex conjugate is simply $F_A^* = F_{Ar} - i\cdot F_{Ai}$. The expression for the spectral energy can now be written as:

$$\begin{aligned}
G_A &= F_A^*\cdot F_A \\
&= (F_{Ar} - i\,F_{Ai})\cdot(F_{Ar} + i\,F_{Ai}) \\
&= F_{Ar}^{\,2} + i\,F_{Ai}\,F_{Ar} - i\,F_{Ai}\,F_{Ar} - i^2\,F_{Ai}^{\,2} \\
&= F_{Ar}^{\,2} + F_{Ai}^{\,2} \\
&= |F_A(n)|^2
\end{aligned}$$

leaving the magnitude squared as a real number.

Similarly, define the spectral intensity $G_B = F_B^* \cdot F_B$, for a different variable B. We can now define the ***cross spectrum*** between A and B by

$$G_{AB} = F_A^* \cdot F_B \tag{8.8.2a}$$

$$= F_{Ar} F_{Br} + i F_{Ar} F_{Bi} - i F_{Ai} F_{Br} - i^2 F_{Ai} F_{Bi}$$

Upon collecting the real parts and the imaginary parts, the real part is defined as the ***cospectrum***, Co, and the imaginary part is called the ***quadrature spectrum***, Q:

$$G_{AB} = Co - i Q \tag{8.8.2b}$$

where

$$Co = F_{Ar} F_{Br} + F_{Ai} F_{Bi} \tag{8.8.2c}$$

and

$$Q = F_{Ai} F_{Br} - F_{Ar} F_{Bi} \tag{8.8.2d}$$

Although not explicitly written in the equations above, F_A and F_B are functions of n, making both the cospectrum and quadrature spectrum functions of n too: Co(n) and Q(n).

The cospectrum is frequently used in meteorology, because the sum over frequency of all cospectral amplitudes, Co, equals the covariance between A and B, (i.e., $\sum_n Co(n) = \overline{a'b'}$). Note that the cospectrum computed as above is NOT equal to the spectrum of the time series of the product a'b'.

The quadrature spectrum is usually not used directly, but it too has a physical interpretation. The quadrature spectrum is equal to the spectrum of the product of b' times a phase shifted a', where a' is phase shifted a quarter period of n. In other words, the amount of time lag applied to a' depends on the frequency, n, such that the phase shift is always 90° for each n.

Three additional spectra can be constructed from the quad and co-spectra. An ***amplitude spectrum***, Am, can be defined as

$$Am = G_{AB}^* \cdot G_{AB}$$

$$= Q^2 + Co^2 \tag{8.8.2e}$$

A large amplitude at any frequency n implies that A is very strongly correlated to B at that frequency, regardless of phase differences between A and B. In other words if both A and B have a strong amplitude component with frequency n = 5 even if A and B are out of phase, then Am will be large for n = 5. Also, if the amplitude is small for any frequency n, then coherence and phase spectra (described next) are not significant (i.e., unreliable) for that frequency.

The ***coherence spectrum***, Coh, is defined by:

$$\text{Coh}^2 = \frac{G^*_{AB}\, G_{AB}}{G_A\, G_B} = \frac{Q^2 + Co^2}{G_A\, G_B} \tag{8.8.2f}$$

This is essentially a normalized amplitude, and is a real number in the range 0 to 1. It acts very much like a frequency dependent correlation coefficient. Note that in some of the literature Coh^2 is defined as the coherence, rather than Coh. Like the amplitude spectrum, it is not a function of phase shift.

Finally, a ***phase spectrum***, Φ, can be defined as

$$\tan\Phi = Q / Co \tag{8.8.2g}$$

This can be interpreted as the phase difference between the two time series A and B that yielded the greatest correlation for any frequency, n. The phase spectrum can be used to infer the nature of the physical flow. For buoyancy waves, θ' is characteristically 90° out of phase with w'; while for turbulence, the two variables either in phase or 180° out of phase.

8.8.3 Example

Problem: Given the time series from section 8.4.2 for humidity, and the time series below for vertical velocity, w:

Index (k):	0	1	2	3	4	5	6	7
Time (UTC):	1200	1215	1230	1245	1300	1315	1330	1345
w (m/s):	0	-2	-1	1	-2	2	1	1

Find and plot:

a) the discrete Fourier transform and the spectrum for w

b) the cospectrum for w and q

c) the quadrature spectrum

d) the amplitude spectrum

e) the coherence spectrum

f) the phase spectrum.

Also find the discrete Fourier transform and the spectrum for the product w'q'.

Solution: The original time series are listed in Table 8-2 as a reference, along with the deviations squared and the series w'q'. The Fourier transforms for both w and q are

Table 8-2. Spectra and cospectra data, computed with an FFT program, and then displayed here in spreadsheet form.

Timeseries:

k	w	q		w'2	q'2		w'q'
0	0	8		0	1		0
1	-2	9		4	4		-4
2	-1	9		1	4		-2
3	1	6		1	1		-1
4	-2	10		4	9		-6
5	2	3		4	16		-8
6	1	5		1	4		-2
7	1	6		1	1		-1
Sum:	0	56	Sum:	16	40	Sum:	-24
Mean:	0	7	Variance:	2	5	Covar:	-3

Simple Spectra:

n	Fw real	Fw imag	Gw	$Ew/\overline{w'^2}$	Fq real	Fq imag	Gq	$Eq/\overline{q'^2}$
0	0.000	0.000			7.000	0.000		
1	-0.104	0.604	0.375	0.375	0.280	-1.030	1.140	0.456
2	-0.250	0.250	0.125	0.125	0.500	0.000	0.250	0.100
3	0.604	0.104	0.375	0.375	-0.780	-0.030	0.610	0.244
4	-0.500	0.000	0.250	0.125	1.000	0.000	1.000	0.200
5	0.604	-0.104	0.375		-0.780	0.030	0.610	
6	-0.250	-0.250	0.125		0.500	0.000	0.250	
7	-0.104	-0.604	0.375		0.280	1.030	1.140	
Sum:			2.000	1.000			5.000	1.000

Cross-spectra (based on F & G values above):

Simple Spectrum of w'q' timeseries:

n	Gwq Co	Gwq Q	Am	Coh2	Phase(°)	Fwq real	Fwq imag	Gwq	$Ewq/\overline{(w'q')^2}$
0						-3.000	0.000		
1	-0.651	0.062	0.428	1.000	174.52	1.104	-0.354	1.343	0.398
2	-0.125	0.125	0.031	1.000	135.00	-0.250	1.250	1.625	0.481
3	-0.474	-0.063	0.229	1.000	187.52	0.396	-0.354	0.282	0.084
4	-0.500	0.000	0.250	1.000	180.00	0.500	0.000	0.250	0.037
5	-0.474	0.063	0.229	1.000	172.48	0.396	0.354	0.282	
6	-0.125	-0.125	0.031	1.000	225.00	-0.250	-1.250	1.625	
7	-0.651	-0.062	0.428	1.000	185.48	1.104	0.354	1.343	
Sum:	-3.000	0.000						6.750	1.000

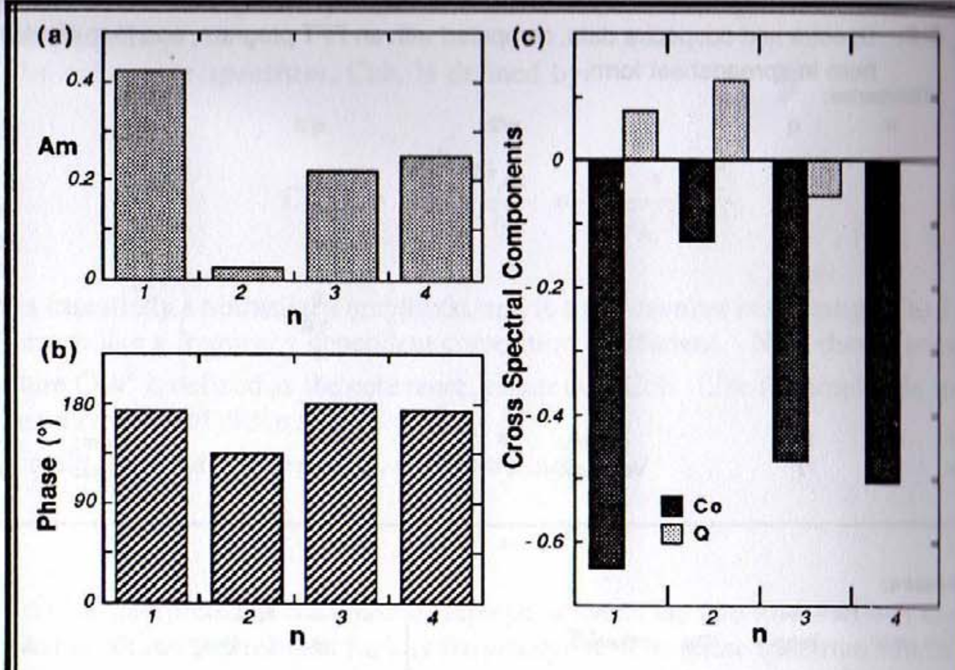

Fig. 8.12 (a) Amplitude spectrum, (b) phase spectrum, and (c) the co-spectral and quadrature components of the cross spectrum for example 8.8.3.

then found using an FFT program, and are listed in Table 8-2 along with their corresponding unfolded spectral intensities, G_w and G_q, and the fraction of variance explained, E_w / s_w^2 and E_q / s_q^2, where s^2 represents the variance.

Also listed is a subtable with co- and quad- spectral components of G_{wq}, the resulting values of Am, Coh2, and the phase angles in degrees. These are plotted in Fig 8.12. Finally, the simple spectrum of the w'q' time series is listed.

Discussion: The biased variances of the w and q time series are 2.0 and 5.0, respectively. From Table 8-2, we see that the sum of the G_w and G_q spectral components equals their respective variances. This is always a good check to do with the analysis. The associated normalized spectral components, E_w / s_w^2 and E_q / s_q^2, sum to unity as desired. Also, the covariance $\overline{w'q'} = -3.0$, which agrees with the sum of the Co cospectral components.

Looking at the original time series, we see that w' is usually positive when q' is negative, as confirmed by the negative covariance. Thus, we anticipate that w' and q' are 180° out of phase. The phase spectrum supports this. In fact, the only phase values which are substantially different from 180° are those for which the amplitude (Am) values are small, suggesting that these phase values can't be trusted.

It is surprising to find that the coherence is 1.0 for all frequencies. This indicates that there is a very close relationship between w and q for all frequencies or wavelengths, for this contrived example. For real turbulence data the coherence would not equal 1.0 for all frequencies.

Next, look at the individual q series. There is an obvious oscillation with three cycles within the whole period of record. In addition there is a background low frequency change of the time series. Looking at the simple spectrum for q, the spectral intensity is indeed large for $n = 3$ and $n = 1$. A similar conclusion can be reached for w. For both of these series, there is a distinct spectral minimum at $n = 2$.

This minimum shows up in the cospectrum at $n = 2$. Thus, waves with two cycles per period contribute little to the total covariance $\overline{w'q'}$. This is in sharp contrast with the w'q' time series itself, which shows a very definite $n = 2$ wave. The simple spectrum analysis of w'q' also yields the largest spectral component at $n = 2$. This tells us that the variance (not covariance) of the w'q' time series has a large contribution at $n = 2$, even though the covariance itself, $\overline{w'q'}$, has a minimum at $n = 2$.

In the discussion presented above, it was easy to compare the spectra with features in the original time series, because the series were so short. For real turbulence data consisting of thousands of data points, it is not so easy to pick out features by eye. For these situations, spectral analysis is particularly valuable.

8.9 Periodogram

The periodogram is just a least squares best fit of sine and cosine waves to the original signal (i.e., to the time series). Because the original time series need not consist of evenly spaced data points for the periodogram to work, it has a very distinct advantage over the discrete Fourier transform. In fact, for some data sets with data gaps or missing data, it is the only method to calculate spectral information short of making up bogus data to fill the gaps. The prime disadvantage of the periodogram is that it takers longer to compute than an FFT.

First, the mean of the original time series of variable A is subtracted from each A(k) data point to yield a modified time series for A'(k). For each frequency (n) a wave of the following form is fitted to the data:

$$A' = a_1 \cos\left[\frac{2\pi kn}{N}\right] + a_2 \sin\left[\frac{2\pi kn}{N}\right] \tag{8.9a}$$

where A' is the deviation of A from the mean, and where a_1 and a_2 are the best-fit coefficients to be determined. Solving for a_1 and a_2 (both a function of n) in the least-squares sense gives: