**GEOPHYSICAL DATA ANALYSIS**

**Class Notes by Bob Parker**

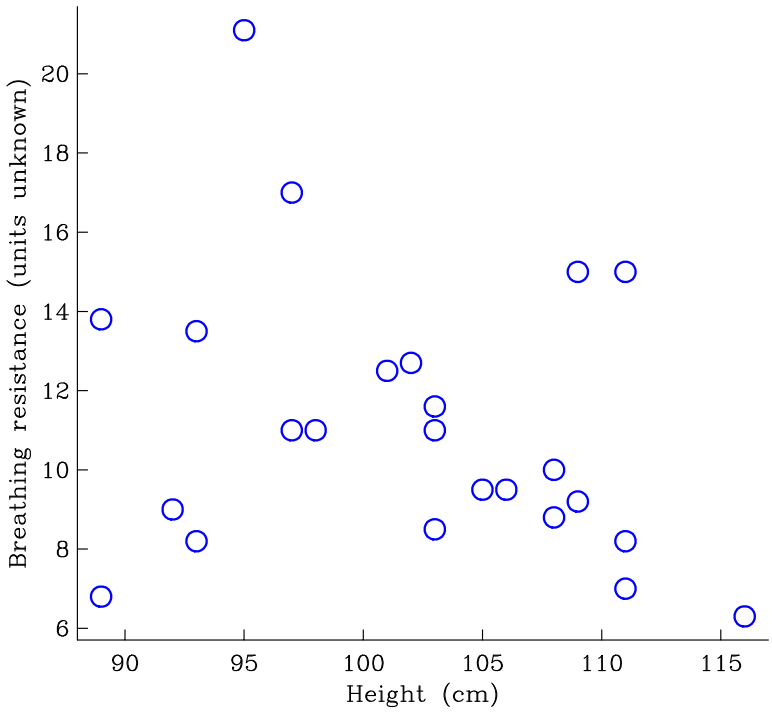
**CHAPTER 4: MULTIVARIATE and MULTIDIMENSIONAL SPECTRA**

## 1. Random Data Pairs

In geophysics and every other branch of physical science we encounter time series in pairs that are related to one another. Examples: the vertical ground motion, and the output voltage of a seismometer; two components of a vector field (say the vertical and the horizontal) as the instrument measures values along a path in space; the topographic height and the observed strength of gravity as seen along a profile (here time has been replaced by distance); the north electric field and the east magnetic field at a site. You can easily think of dozens more yourself. One of the characteristics of the apparently random geophysical list above is that in every case one can discover a theory capable of predicting the relationship between the signal pairs, and a practical need for a data analysis technique to estimate the quantitative behavior. We will consider both these aspects

First, however, let us briefly consider how pairs of variables are traditionally analyzed when there is no independent variable, like time, against which the system is evolving. A first example (from John Rice's book *Mathematical Statistics and Data Analysis*, 2nd ed, Duxbury Press, 1995) concerns breathing resistance in 24 children with cystic fibrosis tabulated as a function of their heights; see Fig 1. We plot this pair of

**Figure 1:** Scatter plot of medical data.



parameters as a point on a graph, (called a **scatterplot**) and if we have enough data we may see a pattern emerging: very roughly, a taller child has a lower breathing resistance. In this case there is no theoretical model, just a statistical tendency. We don't expect a perfectly linear relationship, but it would be nice to have a measure of common tendency, and a test to say if a given sample really exhibits such a tendency or not. (Biologists and social scientists, who usually lack predictive theories, depend on this measure a great deal.)

In cases of relations like this we would like to measure the degree of association of one variable with another. The commonly used one is the **correlation coefficient** $\rho$, or its square. For a pair of random variables $X$ and $Y$, jointly distributed we can calculate:

$$\rho_{XY} = \frac{\mathcal{E}\left[(X - \mathcal{E}[X])(Y - \mathcal{E}[Y])\right]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[Y]}} \tag{1.1}$$

where $\mathcal{E}$ is the expectation, and var is the variance. That is something you can calculate for a theoretical distribution like a joint Gaussian. The numerator you will recognize as your old friend the **covariance** between $X$ and $Y$. A natural estimator of $\rho$ is:

$$\hat{\rho}_{XY} = \frac{\sum_j (x_j - \bar{x})(y_j - \bar{y})}{\left(\sum_j (x_j - \bar{x})^2 \sum_k (y_k - \bar{y})^2\right)^{1/2}} \tag{1.2}$$

where $x_j$ and $y_j$ are the $N$ data samples and $\bar{x}$ and $\bar{y}$ are the sample means. We can shown that, for both the parameter and its estimator, $\rho$ is a number lying between $-1$ and 1. In the $N$-dimensional space $\mathbb{R}^N$ containing the data sample, (1.2) is the inner product between the two sample vectors, divided by their Euclidean norms. Schwarz's inequality $(\mathbf{x}, \mathbf{y}) \le \|\mathbf{x}\|\ \|\mathbf{y}\|$, shows that $|\rho_{XY}| \le 1$. The inequality also shows that $\rho$ is $\pm 1$ only if the two variables are perfectly correlated and the scatterplot would be a perfect straight line; conversely zero corresponds to no statistical connection between the two.

In our example I calculate $\rho$ to be $-0.2603$. Is this significant, or would a random collection of data pairs commonly come up with a number as big as this in magnitude? To answer this question we will assume that $X$ and $Y$ are jointly distributed with a Gaussian law. The theoretical sampling distribution for $\hat{\rho}$ in general is very complicated; see Priestley, Chap 9 for references. But in one case it turns out to be relatively easy to compute – just the case we have: we test the null hypothesis that the true $\rho$ is zero. Then (Kendall and Stuart, *Advanced Theory of Statistics* Vol 2, p 316) it can be shown that the variable $t$ defined by

$$t^2 = \frac{(N-2)\,\hat{\rho}_{XY}^2}{1 - \hat{\rho}_{XY}^2} \tag{1.3}$$

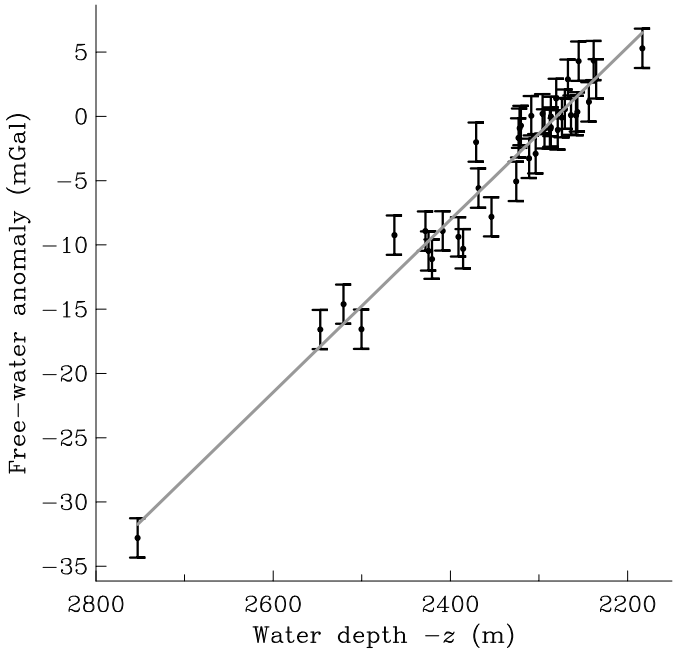
follows the **Student $t$-distribution** with $N-2$ degrees of freedom. Thus we can find the probability that even though the true correlation coefficient of the distribution is zero, the estimator, by chance gets bigger than $t$. Plugging in the numbers for Fig 1 we get $t = 1.264$, and from the $t$ tables, I find the probability of its being this big by chance is about 0.15; this implies an 85% confidence there is a relationship.

We continue to look at the case of a pair of variables, without an underlying evolutionary process in time. Often in the physical sciences we have a reason for believing in a linear relationship between two variables based on some theory, but that there is a random component affecting one (or both) of the observations. Fig 2 is a geophysical application from the PhD thesis of Mark Stevenson, who worked for Mark Zumberge. This plot (from Fig 3 of Stevenson et al., *JGR* 99, 4875-88, 1994) concerns gravity survey made on the Vance-Cleft Overlapping Rift Zone on the Juan de Fuca ridge. The two variables are: (y-axis) the observed gravity anomaly measured on the seafloor, corrected for the earth's vertical gravity gradient and the presence of seawater; (x-axis) the depth of the observation site. For smooth terrain we expect the two to be connected by:

$$\Delta g = 2\pi G\, \rho_c\, z + c \tag{1.4}$$

where $G$ is Newton's gravitational constant and $\rho_c$ is the density of the crust. (What is $c$ in this equation?) Finding the crustal density from the slope of the line, is known as **Nettleton's Method**. The gravity variable is much less accurately known than the water depth because of instrument shaking due to water currents and the factor that the seafloor density really isn't constant; so the fit is not exact (of course). Here the

**Figure 2:** Sea-floor gravity versus bathymetry.



appropriate statistical model is not one of a pair of random variables with joint two-dimensional Gaussian distribution, but rather:

$$Y_j = \alpha x_j + \beta + N_j \tag{1.5}$$

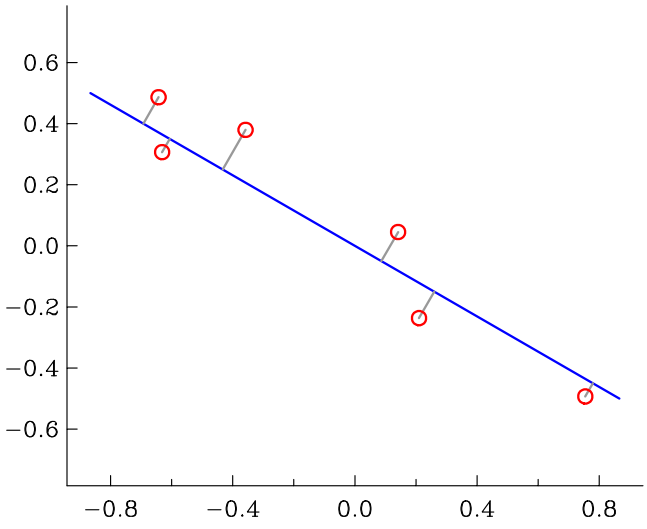
where $N_j$ are random variables, ideally distributed identically with zero mean, uncorrelated and maybe with known variance. (Why did I write $Y$ in upper case but $x$ lower case?) Though you can calculate the correlation coefficient here, the graph makes it clear there is a strong relationship and the real interest centers on estimation of the slope of the line, which leads to the constant $\rho_c$ in (1.4); also we would like a figure for the uncertainty of the estimate. You should recognize a good application of the **least squares** method. I won't go over that material here, since you must have seen it several times already. However, I want just to mention one result. The least-squares estimator of the slope is given by:

$$\hat{\alpha} = \frac{\sum_j (X_j - \bar{X})(Y_j - \bar{Y})}{\sum_j (X_j - \bar{X})^2} = \hat{\rho}_{XY} \, \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \, . \tag{1.6}$$

Remember, for uncorrelated random variables, (1.6) has the least variance of all unbiased linear estimators (Gauss-Markov Theorem). If the variables are Gaussian in addition, (1.6) is the best unbiased estimator of $\alpha$, period.

Notice that in (1.5) that **all** the noise is attributed to the $y$ variable, in the example, the gravity measurement. But what if both coordinates have random components? Then things get considerably more complicated, depending on what can be assumed about the noise. You will easily convince yourself that the estimate of the slope is different if one assumes all the noise is $x$ instead of $y$. The next simplest case is the one in which the noise in $x$ and $y$ is taken to be the same. But observe this really only makes any sense when $x$ and $y$ are measured in the same units; it is meaningless to say the water depth in meters has the same uncertainty

**Figure 2a:** Distances used in TLS.



as the gravity anomaly in $\mathrm{m\,s^{-2}}$ ! In this special case we perform what is called a **total least squares** estimate: we minimize the sum of squares of the Euclidean distances from the points in the plane to the line.

If the straight line is $y = \alpha x + \beta$, the TLS estimator for the intercept is

$$\hat{\beta} = \frac{1}{N}\sum_j Y_j - \frac{\hat{\alpha}}{N}\sum_j X_j = \hat{\hat{X}} - \hat{\alpha}\hat{\hat{Y}} \tag{1.7}$$

where $\hat{\alpha}$ is the slope estimate given by the solution of this quadratic equation:

$$0 = -c_2 + c_1\hat{\alpha} + c_2\hat{\alpha}^2 \tag{1.8}$$

$$\begin{aligned} c_2 &= \sum_j (X_j - \hat{\hat{X}})(Y_j - \hat{\hat{Y}}) \\ c_1 &= \sum_j [(X_j - \hat{\hat{X}})^2 - (Y_j - \hat{\hat{Y}})^2] \end{aligned} \quad . \tag{1.9}$$

A quadratic equation usually has two solutions of course; in thus case, one corresponding to the best fit, the other to the worst.

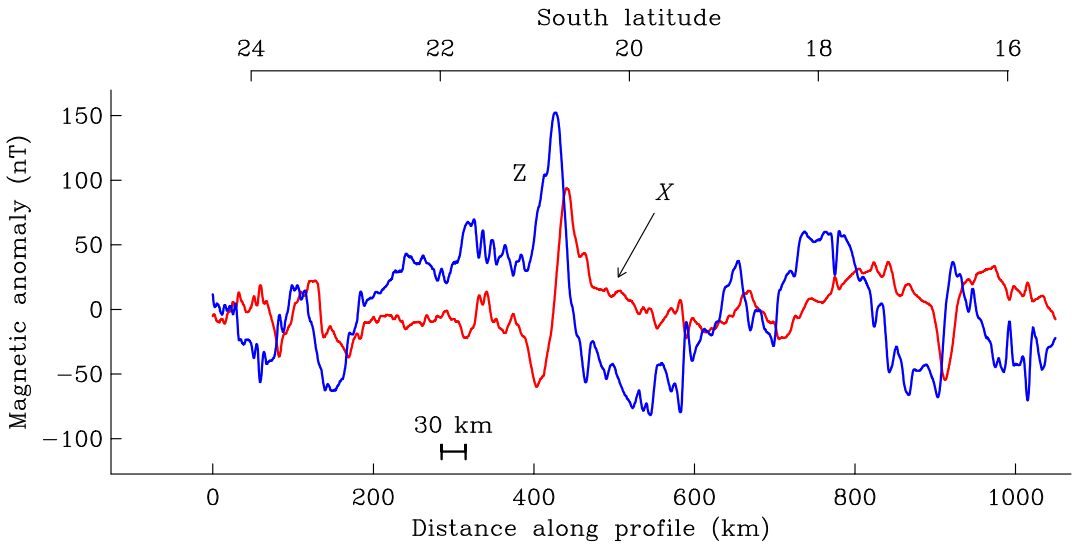
*Exercise*

In (1.5) suppose it is known that $\beta = 0$, so that the true straight lines runs through the origin. Solve the problem of estimating the slope $\alpha$ by the least-squares method. Give an explicit formula like (1.6) for the estimator.

## 2. Pairs of Stationary Signals

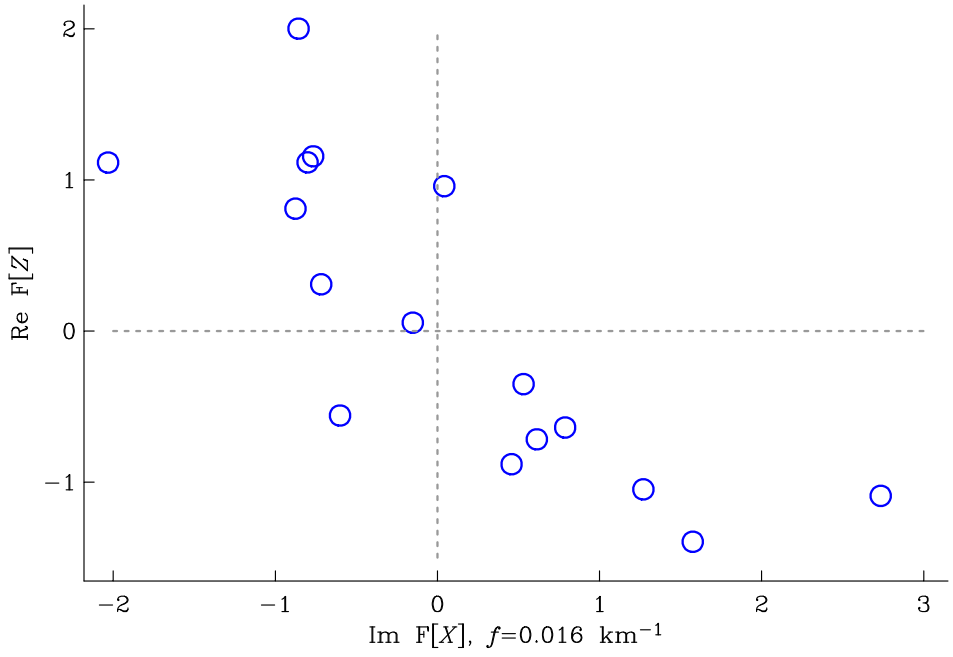
Suppose now we have a pair of signals, possibly related, both of them assumed to be stationary processes, for example the pair in Fig 3. These

**Figure 3:** Two components of an aeromagnetic signal.



are the track-parallel $X$, and vertical $Z$, magnetic anomaly components as measured on a high-flying aircraft traveling roughly northward in the southeastern Pacific over Nazca Plate. There are two ways of thinking about the power spectrum. The more intuitive one asks us to imagine filtering the time series through a very narrow band filter to isolate a given frequency component – then the variance of the random signal that results, normalized by the bandwidth of the filter is the power spectral density. The corresponding idea for handling pairs of signals is very similar: imagine taking the Fourier transform of an identical time sample of each series and selecting a single frequency; do this again and again with new realizations. Now form a *complex* scatterplot of the Fourier amplitude of series 1 against 2. Each point on the plot is one Fourier transform. If the series are related, the cloud of points will tend to be elongated. We can compute the correlation coefficient using the complex version of (1.2) and taking the absolute value; this is called the **coherency** at the selected frequency. You can also form the (complex) slope, and this is the **transfer function** between the two signals. To give you the flavor, I have plotted the imaginary part of $\mathcal{F}\,[X]$ versus the real part of $\mathcal{F}\,[Z]$ at one particular wavenumber: see Fig 4. I made separate realizations by dividing the original series into shorter pieces. You see the strong (negative) correlation between the two components; we will see why we should expect this behavior in a moment. To do a proper job of showing you the complete complex scatterplot I would have to plot every component against every other (six plots as we omit self-correlations, and the system is symmetric). But the frequency domain provides a basis for forming numerical estimates.

**Figure 4:** Scatter plot of two components of DFTs from Fig 3 at a fixed wavenumber.



Before discussing the estimation process, we develop in more detail the alternative of looking at spectra as Fourier transforms of covariance functions. Let two real stationary stochastic processes be $X_1$ and $X_2$ both with zero mean, and define the **cross-covariance** between them:

$$R_{21}(s) = \operatorname{cov}[X_1(t),\, X_2(t+s)] = \mathcal{E}\,[X_1(t)\, X_2(t+s)] \tag{2.1}$$

and in general the **covariance matrix** is

$$R_{kj}(s) = \mathcal{E}\,[X_j(t)\, X_k(t+s)], \quad j,\, k = 1, 2. \tag{2.2}$$

Notice the (somewhat illogical looking) reversal of the order of the indices in the function and under the expectation. Because the processes are stationary, $R_{jk}$ depends only on $s$ even though $t$ appears in the equation. Also some trivial algebra gives $R_{12}(s) = R_{21}(-s)$.

We have already seen that the **power spectral density** (PSD) is defined by taking the Fourier transform of the covariance function, in the present case choosing the diagonal elements of the matrix. Similarly, we define the **cross-spectrum** between $X_1$ and $X_2$ by

$$S_{12}(f) = \mathcal{F}\,[R_{12}](f) = \int_{-\infty}^{\infty} R_{12}(t)\, \mathrm{e}^{-2\pi \mathrm{i} f t}\, dt\ . \tag{2.3}$$

Unlike the PSD, which is always real and nonnegative, the cross-spectrum is usually a complex function of frequency. Traditionally the real and imaginary parts are given names:

$$S_{12} = C_{12} + \mathrm{i} Q_{12} \tag{2.4}$$

where $C_{12}$ is called the **cospectrum** and $Q$ the **quadrature** spectrum. Then the **coherency** spectrum is defined by

$$\gamma_{12} = \frac{|\, S_{12}\,|}{\sqrt{S_{11}\, S_{22}}} = \frac{\sqrt{C_{12}^2 + Q_{12}^2}}{\sqrt{S_{11}\, S_{22}}}\ . \tag{2.5}$$

I might add that I personally prefer to use the square of $\gamma_{12}$, which is called the **coherence**. Equation (2.5) gives the correlation coefficient between the two signals as a function of frequency. Its equivalence to the more intuitive definition I gave earlier certainly requires proof; we don't have time here for it. Unfortunately, I don't find Priestley's (1981) demonstration (p 661) particularly convincing, as it depends on the orthogonal increment representation of the processes.

In addition to the coherence, we define the **phase spectrum** in the straightforward way as:

$$\Phi_{12} = \tan^{-1}\left(\frac{-Q_{12}}{C_{12}}\right). \tag{2.6}$$

The definition (2.3) is not a good way to estimate the cross spectrum from data time series, but it is usually the best way to make theoretical calculations as we will illustrate in short while.

*Exercise*

Suppose $X_1$ is a stationary stochastic process of time, and $X_2$ is defined by $X_2(t) = X_1(t+t_0)$, where $t_0$ is a constant. Find the coherence and phase spectrum of these two signals. How do these spectra depend on the PSD of $X_1$?

### 3. Estimation of Cross Spectra (Briefly)

Recall the standard modern way of estimating the PSD. It is based on the idea that the PSD is the limit of a finite Fourier transform. Let us assume that $\Delta t = 1$, so that the Nyquist frequency $f_{Nyq} = ½$. The periodogram estimate for the PSD is

$$\hat{P}_X(m\Delta f) = \left| \sum_{n=0}^{N-1} X_n e^{-2\pi mn/N} \right|^2, \quad m = 0, 1, , \cdots ½N \tag{3.1}$$

where $\Delta f = 1/N$, and we have chosen the output frequency samples to take advantage of the FFT. As you should remember, this estimator has enormous variance, for white Gaussian noise the standard deviation as the same as the answer. And for non-white processes, the result can be strongly biased by spectral leakage. We ameliorate both these defects by using 3tapers. To improve the variance we must average over independent estimates. As you may recall we asserted that if the two tapers $u_n$ and $v_n$ are orthogonal (so that $\sum_n u_n v_n = 0$) then PSD estimates based on tapered versions of the time series $u_n X_n$ and $v_n X_n$ are independent estimates of $P_X$. So the modern strategy is to choose a set of mutually orthogonal tapers with good spectral leakage properties, and then to average the estimates together. That is **multitaper estimation**; see Percival and Walden, (1993), and Riedel and Sidorenko (1995).

Multitaper estimation can and should be used with cross spectra. The equivalent periodogram estimator for the cross spectrum between $X_n$ and $Y_n$ is just

$$\hat{P}_{XY}(m\Delta f) = \left( \sum_{n=0}^{N-1} X_n e^{-2\pi mn/N} \right) \left( \sum_{n=0}^{N-1} Y_n e^{-2\pi mn/N} \right)^* . \tag{3.2}$$

Tapered series can be formed and transformed with FFTs, but to preserve the independence the same taper set must be used for $X_n$ and $Y_n$ even though the two processes may be quite different, and the optimal set to suppress leakage would then be different too. When $K$ sets are used the variance is reduced by the factor $K^{-1}$. Estimates of coherence and phase are usually based on the definitions

$$\hat{\gamma}_{XY} = \frac{|\hat{P}_{XY}|}{\sqrt{\hat{P}_X \hat{P}_Y}} \tag{3.3}$$

$$\hat{\Phi}_{XY} = \tan^{-1}\left( \frac{-\mathrm{Im}(\hat{P}_{XY})}{\mathrm{Re}(\hat{P}_{XY})} \right). \tag{3.4}$$

These are **not** unbiased estimators, even when the various spectra and cross spectra are. If the same tapers are used to estimates $P_X$ and $P_Y$ then it can be shown that $0 \le \hat{\gamma}_{XY} \le 1$, which is desirable; if different sets are used, the coherence can wander above unity.

Uncertainty estimates can be formed based on Gaussian statistics, but the modern way is the Jack-knife estimator, which makes no such assumptions. One exception is the case when we want to know if the coherence (the analog of the correlation coefficient) differs significantly from zero. Priestley (p 706) shows that under the Gaussian assumption for $X_n$ and $Y_n$, if $K$ independent estimates are averaged to find $\hat{\gamma}_{XY}$ then the statistic $s$ defined by

$$s = \frac{2K\hat{\gamma}_{XY}}{1-\hat{\gamma}_{XY}} \tag{3.5}$$

has an $F_{2,4K}$ distribution. This allows a simple test for the hypothesis that $\gamma_{XY}$ is zero:

$$\Pr(\mid \hat{\gamma}_{XY} \mid \geq z) = (1-z^2)^{K-1} \tag{3.6}$$

which is often the most important thing one wishes to know.

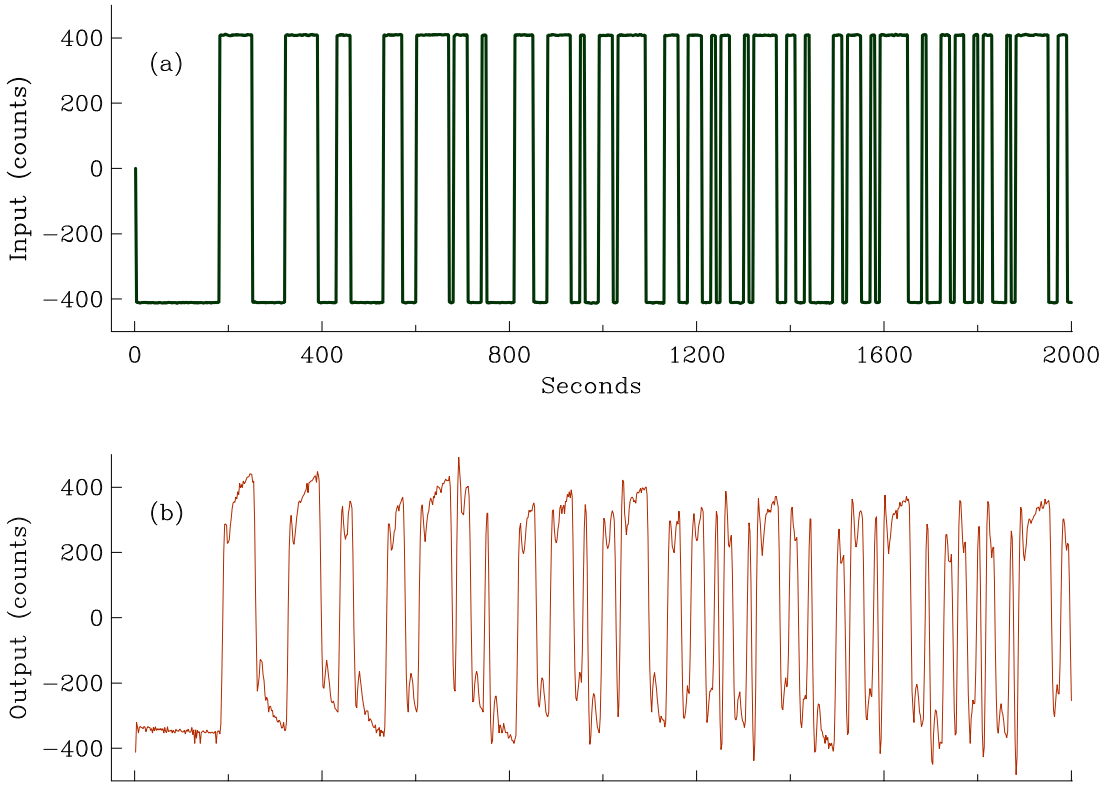
## 4. Example: Calibration — Convolution plus Noise

We return to time series – functions of a single independent variable, like time. Consider the calibration of a seismometer, by shaking it with a "random" signal generated by a computer

$$V = g * I + N \ . \tag{4.1}$$

Here $V$ is the output voltage, $g$ is the **impulse response** of the seismometer (it is the output for a delta function input), $I$ is the input signal, and $N$ is noise – it's the result of ordinary ground motion, from which one cannot isolate the system completely. (See the paper by Berger, et al., *Bull. Seism. Soc. Am., 69*, 271-88, 1979). Fig 5 shows the input (a), and output from a long-period seismometer, (b). The input is called a **random telegraph** signal – it switches randomly between one of two values. If the transition times were entirely uncorrelated $I$, would be a white noise but, because the transitions are allowed only on multiples of 5 data samples, this is not exactly true here; this input provides a fairly flat spectrum, without the large amplitudes; see Fig 6.

**Figure 5:** Input and output signals for a seismometer undergoing calibration.



If we possessed an infinite, noise-free record we could take the FT with time and obtain:

$$\hat{V} = \hat{g}\,\hat{I} \tag{4.2}$$

where $\hat{g}$ is called the **transfer function**, or simply the **frequency response** of the seismometer. At each frequency the output is just the input multiplied by a complex number, which varies with frequency. Our task here is to estimate $\hat{g}$ when $V$ and $I$ are stochastic processes which we have measured; the noise $N$ is unknown. At first you might think all you have to do is to take the FFT of the time series, and divide both sides of (4.2) by $\hat{I}$ but, just as a naive technique is poor way of estimating the power spectrum, this is not the best estimator. Notice that in (4.1) we have the same least-squares situation as the gravity problem at the beginning: *the noise is confined to only one of the signals*. For stationary stochastic processes the answer is the analog of (1.6) as you might expect:

$$\hat{g} = \frac{S_{VI}}{S_I} \tag{4.3}$$

the cross spectrum of $V$ with $I$ divided by the PSD of $I$. The magnitude $|\,\hat{g}\,|$ is often called the **gain** of the system; then we can define the transfer function in terms of its gain and, through (2.6), its phase.

Here is the proof of (4.3). To make any progress we must assume that $N$ in (4.1) is *uncorrelated* with the input $I$. Then we calculate the cross-covariance :

$$R_{VI}(s) = R_{IV}(-s) = \mathcal{E}\,[V(t)\,I(t-s)] = \mathcal{E}\,[(g * I + N)(t)\,I(t-s)] \tag{4.4}$$

$$= \mathcal{E}\,[(g * I)(t)\,I(t-s)] \tag{4.5}$$

$$= \mathcal{E}\,[\int dp\; g(p)\,I(t-p)\,I(t-s)] \tag{4.6}$$

$$= \int dp\; g(p)\,R_I(p-s) = \int dp\; g(p)\,R_I(s-p) \tag{4.7}$$

where in (4.7) we used the definition of $R_I$ to evaluate $\mathcal{E}\,[I(t-p)\,I(t-s)]$, having moved the expectation under the integral. Now take the Fourier transform of (4.7), which is by definition the cross spectrum between $V$ and $I$:

$$S_{VI}(f) = \int ds \int dp\; \mathrm{e}^{-2\pi \mathrm{i} s f}\; g(p)\,R_I(s-p) \tag{4.8}$$

$$= \int dp\,[\int ds\; \mathrm{e}^{-2\pi \mathrm{i}(s-p)f}\; R_I(s-p)]\; \mathrm{e}^{-2\pi \mathrm{i} p f}\; g(p) \tag{4.9}$$

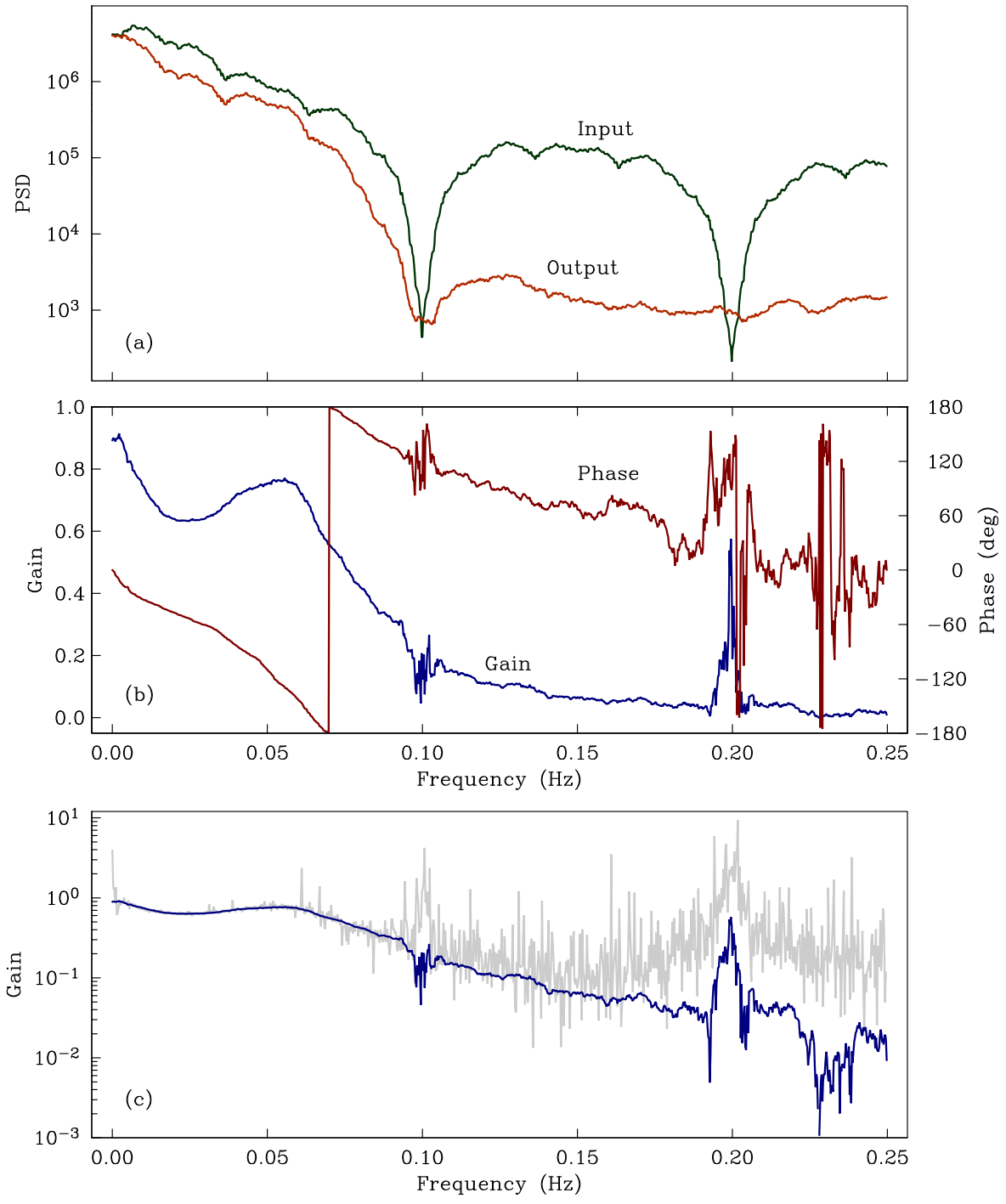
$$= S_I(f)\,\hat{g}(f)\;. \tag{4.10}$$

And this is just (4.3).

This is not of course a proper derivation of an *estimator*. What it shows is that if all the noise is in one of the stationary processes, the transfer function is given exactly by the ratio of the cross spectrum to the PSD of the input. It can be shown (see Priestley) that $\hat{g}$ is an unbiased estimator when estimates of $S_{VI}$ and $S_I$ are used rather than the exact functions.

Fig 6 shows in (a) the power spectra of input and output of our seismometer example. The random telegraph input has two very low power

**Figure 6:** Gain and phase characteristics of the seismometer.



"holes" in its PSD at 0.1 Hz, and again at 0.2 Hz. It is exactly at these frequencies the estimated phase and gain becomes unreliable, as can be seen in (b). The gain, here the magnitude of the seismometer frequency response, is in reality a very smooth function. The estimates are quite smooth, except in the neighborhood of the input power holes, and for $f > 0.17$Hz where ground motion is having an effect. In the bottom panel, which has a log scale for gain, I have plotted in grey the naive estimate $|\hat{V}/\hat{I}|$, a simple ratio of DFTs. Note how very rough that estimate is, and how strongly biased — the average values are much too large.

The geophysical application most commonly associated with cross spectral analysis is geomagnetic and magnetotelluric sounding, where time series of electric and magnetic field components are the basis for estimates for the impedance of the Earth as a function frequency, from which electrical conductivity structure can be inferred. One complication here is that both signals are subject to noise, so that the approach described for calibration, while often used, is not strictly valid, and alternatives to avoid these problems sometimes involve a special experimental setup called a remote reference station (Gamble, et al., (1979); Egbert and Booker (1986).

Another important topic in both Earth and planetary sciences is cross-spectral analysis between topography and gravity data. The transfer function, called the **isostatic response**, is a function wavenumber, not frequency, yields insight into crustal and lithospheric strength. In this case it is plausible to assert that the gravity data are far more prone to error than the topography, and so a transfer function estimate like the one we have just done will be appropriate. See Chapter 5 of Watts (2001) for many details. However, for this work we may need to work from data over an area, rather than profiles, and that brings us to our next subject.

## 5. Stationary Processes in the Plane

We take a brief look at **multidimensional spectra**, that is spectra of functions of several variables. Up to now we considered on a single independent variable (usually identified with time). Suppose there is a stationary processes in the plane (like bathymetry, or the gravity field from random sources). It is simple to define the autocovariance of $X$, a signal in the plane, by following the single-variable recipe:

$$R_X(\mathbf{r}) = \mathcal{E}\,[X(\mathbf{s})\,X(\mathbf{s}+\mathbf{r})] \tag{5.1}$$

for $\mathbf{r}, \mathbf{s} \in \mathbb{R}^2$. Then the two-dimensional power spectral density of $X$ is just the 2-D Fourier transform of $R_X$ as you would expect:

$$S_X(\mathbf{k}) = \mathcal{F}_2[R_X](\mathbf{k}) = \int_{\mathbb{R}^2} R_X(\mathbf{x})\, e^{-2\pi i \mathbf{k}\cdot\mathbf{x}}\, d^2\mathbf{x}\,. \tag{5.2}$$

Quite often we have measurements on a single profile, or a slice thorough a 2-dimensional random field. How do we collapse the 2-dimensional spectrum to get the observed profile spectrum? This is exactly what we have with the aeromagnetic data of Fig 3. We will prove the **Slice Theorem** for spectra and cross spectra. Then we'll apply it to magnetic data.

First we state without proof the Slice Theorem for ordinary Fourier transforms. Given the 2-dimensional Fourier transform of a function $f$:

$$\hat{f}(\mathbf{k}) = \mathcal{F}_2[f](\mathbf{k}) \tag{5.3}$$

we find the 1-dimensional FT of $f$ along the line $y = 0$ by:

$$\mathcal{F}_x[f(x,0)](k) = \int_{-\infty}^{\infty} dk_y\, \hat{f}(k_x, k_y)\,. \tag{5.4}$$

In words, we integrate the 2-dimensional Fourier transform in the perpendicular direction in the wavenumber domain to get the corresponding 1-dimensional FT.

Exactly the same result holds for power spectra and cross spectra! We prove it for the PSD. A one-dimensional stationary process $U$ is a sample on the line $x\hat{\mathbf{x}}$ of the process $W$ in the plane $\mathbb{R}^2$. The autocovariance of $W$ is $R_W$ and then the autocovariance of $U$ is

$$R_U(x) = R_W(x\hat{\mathbf{x}}) = R_W(x, 0)\,. \tag{5.5}$$

By definition, the PSD of $W$ is the 2-dimensional Fourier transform of the autocovariance $R_W$, that is

$$S_W = \mathcal{F}_2[R_W]\,. \tag{5.6}$$

Conversely:

$$R_W(\mathbf{x}) = \mathcal{F}_2^{-1}[S_W] = \int_{\mathbb{R}^2} d^2\mathbf{k}\, e^{2\pi i \mathbf{x}\cdot\mathbf{k}}\, S_W(\mathbf{k})\,. \tag{5.7}$$

Substituting (5.7) into (5.5) gives

$$R_U(x) = \int_{\mathbb{R}^2} d^2\mathbf{k}\, e^{2\pi i x \hat{\mathbf{x}}\cdot\mathbf{k}}\, S_W(\mathbf{k}) \tag{5.8}$$

$$= \int_{-\infty}^{\infty} dk_x\, e^{2\pi i k_x x} \int_{-\infty}^{\infty} dk_y\, S_W(k_x, k_y) \tag{5.9}$$

$$= \mathcal{F}_x^{-1}[\int_{-\infty}^{\infty} dk_y\, S_W(k_x, k_y)]\,. \tag{5.10}$$

And again the PSD of $U$ is the one-dimensional Fourier transform of $R_U$:

$$S_U(k) = \mathcal{F}_x[\mathcal{F}_x^{-1}[\int_{-\infty}^{\infty} dk_y\, S_W(k_x, k_y)]] \tag{5.11}$$

$$= \int_{-\infty}^{\infty} dk_y\, S_W(k_x, k_y)\,. \tag{5.12}$$

Equation (5.12) is the Slice Theorem for power spectra. The result is the same for cross spectra: simply integrate the 2-dimensional cross spectrum in the perpendicular direction in the wavenumber domain.

Suppose there are two stochastic processes $U$ and $V$ both derived from a third $X$ via convolution:

$$U = g * X, \quad V = h * X\,. \tag{5.13}$$

How is the cross spectrum $S_{UV}$ related to $S_X$, the PSD of $X$? The answer, which takes a few lines of simple algebra is:

$$S_{UV} = \hat{g}\cdot\hat{h^*}\, S_X \tag{5.14}$$

where the hat means Fourier transform. (When a hat decorates a function it denotes the Fourier transform; on a simple variable, it means a statistical estimate; I hope this convention does not cause confusion.) This result is valid for processes of a single variable or in the plane $\mathbb{R}^2$, or indeed higher dimensions. When $g = h$, so that $U = V$, we get the well-know result for the PSD of $U$:

$$S_U = |\hat{g}|^2\, S_X\,. \tag{5.15}$$

The generalization of these results to three dimensional space is quite straightforward and we will not spend time on it. Another, more interesting generalization is to the sphere. Clearly for geophysics it is only an approximation to say that the surface of the Earth is a plane; we can ask what happens if the region is so large that we must account for curvature, or even larger still, so that the whole surface of the Earth is the domain of the random process. We will discuss that question later.

A specially important kind of stochastic process in the plane is the **isotropic process**, which is one with the property that the autocovariance depends only on $|\mathbf{r}|$, the distance between the two points. Then let

$$R_X(\mathbf{r}) = \rho(|\mathbf{r}|). \tag{5.16}$$

The PSD is as always the Fourier transform of $R_X$ and so

$$S_X(\mathbf{k}) = \mathcal{F}_2[R_X] = \sigma(|\mathbf{k}|) \tag{5.17}$$

where you will recall that the 2-D FT of a circularly symmetric function like $\rho$ is another function with circular symmetry. (Here symbol $\sigma$ is nothing to do with stanadard error.) The relation between $\rho$ and $\sigma$ is a **Hankel transform**:

$$\sigma(k) = \mathcal{H}[\rho] = \int_0^\infty \rho(r)\, J_0(2\pi k r)\, 2\pi r\, dr \tag{5.18}$$

and $J_0(x)$ is the Bessel function. We have encountered the Hankel transform in Fourier theory. We note that the Hankel transform is its own inverse: $\mathcal{H}[f] = \mathcal{H}^{-1}[f]$. When a process in a plane is isotropic it is only necessary to gather data on a straight profile to construct the PSD, because the autocovariance along a straight line in any direction is just the function $\rho(x)$. This allows us to bring in the Slice Theorem again.

Suppose one has observations along a single straight line from which we can calculate the profile spectrum $P_X(k_x)$; notice that because the stochastic process is isotropic, any line across the field can be designated the $x$ axis. Then we can find the profile PSD in two ways: either as the one-dimensional FT of the autocovariance:

$$P_x(k_x) = \mathcal{F}_1[\rho] \tag{5.19}$$

or by the Slice Theorem:

$$P_X(k_x) = \int_{-\infty}^{\infty} S_X(k_x, k_y)\, dk_y = \int_{-\infty}^{\infty} \sigma(\sqrt{k_x^2 + k_y^2})\, dk_y \tag{5.20}$$

In practice, one may have observations on a profile and hence knowledge of $P_X$, which one would like to convert into the 2-D PSD function $\sigma$. The way is clear: we invert (5.19), then apply (5.18) to find $\sigma$. In fact the action of a 1-D FT followed by a Hankel transform can be combined into a new operation:

$$\sigma(k) = \mathcal{H}[\mathcal{F}_1^{-1}[P_X(k_x)]] = -\frac{1}{\pi} \int_k^\infty \frac{dP_X}{dk_x} \frac{dk_x}{\sqrt{k_x^2 - k^2}}. \tag{5.21}$$

I omit details of this derivation, which can be found in Bracewell, who calls (5.21) the **Abel transform**. Equation (5.21) takes us directly from a profile PSD to the 2-D PSD function. Equation (5.20) goes the opposite way.

## 6. Example: Magnetics over the Ocean

We can put all the 2-D material together for the aeromagnetic signals. The magnetic field components $X$ and $Z$ are related to $V$, the scalar magnetic potential in the plane at $z=0$. Since the potential $V$ is harmonic (it obeys Laplace's equation, $\nabla^2 V = 0$) above the sources, it can be upward continued from the plane $z = 0$. You should recall that for wavenumbers $\mathbf{k} \in \mathbb{R}^2$

$$\hat{V}(\mathbf{k}, z) = \hat{V}(\mathbf{k}, 0)\, \mathrm{e}^{-2\pi |\mathbf{k}| z} \tag{6.1}$$

where $\hat{V}$ is the FT of $V$ in a plane of constant $z$:

$$\hat{V}(\mathbf{k}, z) = \mathcal{F}_2[V(\mathbf{x}, z)] = \int_{IR^2} \mathrm{e}^{-2\pi i \mathbf{k}\cdot\mathbf{x}} V(\mathbf{x}, z)\, d^2\mathbf{x}\,. \tag{6.2}$$

Thus knowledge of $V$ on $z = 0$ is enough to determine $V$ anywhere above this level, provided the sources all lie below the plane $z = 0$. If we take the inverse 2-D FT of (6.1), and use the Convolution Theorem (backwards), we find

$$V(\mathbf{x}, z) = V_0 * G = G * V_0 \tag{6.3}$$

where $V_0 = V(\mathbf{x}, 0)$ and

$$G(\mathbf{x}) = \mathcal{F}_2^{-1}[\mathrm{e}^{-2\pi |\mathbf{k}| z}] = \frac{1}{2\pi} \frac{z}{(z^2 + |\mathbf{x}|^2)^{3/2}}\,. \tag{6.4}$$

We need, not the potential, but the vector $\mathbf{B}$, and $\mathbf{B} = -\nabla V$. From the definition of convolution it follows that

$$B_z = Z = -\frac{\partial V(\mathbf{x}, z)}{\partial z} = -\frac{\partial G}{\partial z} * V_0 = G_Z * V_0 \tag{6.5}$$

and similarly for the other components.

So far we have been considering ordinary functions for $V$ and $Z$. Now suppose that $V_0$ is a stationary stochastic process in the plane $z = 0$, a random field. Then from (6.5) so is the vertical field. Suppose the (two-dimensional) PSD of $V_0$ is $S_V(\mathbf{k})$. We calculate the (2-D) PSD of $Z$ in the standard way, equation (27):

$$S_Z = |\hat{G}_Z|^2 S_V \tag{6.6}$$

where $\hat{G}_Z$ is the 2-D FT of $G_Z$. But by (6.4) $G$ is the inverse FT of the exponential in (6.1), so it follows that

$$\hat{G}(\mathbf{k}) = \mathrm{e}^{-2\pi |\mathbf{k}| z} \tag{6.7}$$

$$\hat{G}_Z = -\frac{\partial \hat{G}}{\partial z} = 2\pi |\mathbf{k}|\, \mathrm{e}^{-2\pi |\mathbf{k}| z}\,. \tag{6.8}$$

In a very similar way we can calculate the spectrum of $X = B_x = -\partial V/\partial x$:

$$B_x = X = -\frac{\partial V(\mathbf{x}, z)}{\partial x} = -\frac{\partial G}{\partial x} * V_0 = G_X * V_0 \tag{6.9}$$

$$\hat{G}_X = 2\pi \mathrm{i} k_x \mathrm{e}^{-2\pi |\mathbf{k}| z} \tag{6.10}$$

$$S_X = |\hat{G}_X|^2 S_V \,. \tag{6.11}$$

And $Y$ follows in exactly the same way, with

$$\hat{G}_Y = 2\pi \mathrm{i} k_y \mathrm{e}^{-2\pi |\mathbf{k}| z} \,. \tag{6.12}$$

Notice that the cross spectrum, say between $X$ and $Z$, as given by (26) is

$$S_{XZ} = \hat{G}_X (\hat{G}_Z)^* S_V \,. \tag{6.13}$$

We have here a 2-D system in which the three processes $X$, $Y$, $Z$ (which happen to be three components of a random vector) are each given by different convolutions of a single process $V$. This means the 3 components of the random magnetic vector are closely related. For example, because $|\mathbf{k}|^2 = k_x^2 + k_y^2$, it follows that

$$S_Z = S_X + S_Y \,. \tag{6.14}$$

This result says that the vertical component of **B** of a field generated from random sources is always larger on average than either of the other two.

And now back to that long magnetic profile over Pacific Ocean shown in part in Fig 3. We have all the pieces to be able to say something interesting. To look at the power and cross spectrum of the profile data we simply apply the Slice Theorem:

$$P_X(k_x) = \int_{-\infty}^{\infty} dk_y \; |\hat{G}_X(\mathbf{k})|^2 \; S_V(\mathbf{k}) \tag{6.15}$$

$$P_Y(k_x) = \int_{-\infty}^{\infty} dk_y \; |\hat{G}_Y(\mathbf{k})|^2 \; S_V(\mathbf{k}) \tag{6.16}$$

$$P_Z(k_x) = \int_{-\infty}^{\infty} dk_y \; |\hat{G}_Z(\mathbf{k})|^2 \; S_V(\mathbf{k}) \tag{6.17}$$

$$P_{XZ}(k_x) = \int_{-\infty}^{\infty} dk_y \; \hat{G}_X(\mathbf{k}) \, \hat{G}_Z(\mathbf{k})^* \; S_V(\mathbf{k}) \,. \tag{6.18}$$

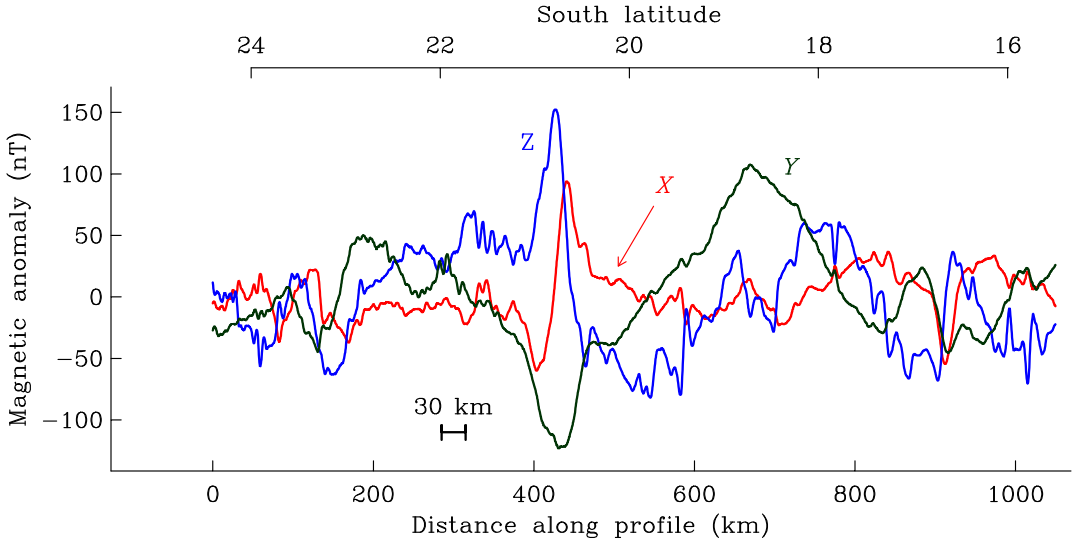
Observe that we have the same relationship between the profile PSDs of the components as in the 2-D system:

$$P_Z = P_X + P_Y \,. \tag{6.19}$$

This relationship has been called the **Power Sum Rule**. Another thing to notice in this example is that while $\hat{G}_Z$ is real, $\hat{G}_X$ (and $\hat{G}_Y$) is purely imaginary. Thus whatever the spectrum is for $S_V$, from (6.12) we see the phase spectrum between must be exactly $\pi/2$ and constant for all $k_x$. This is a very strong prediction. We'll return to this in a moment.

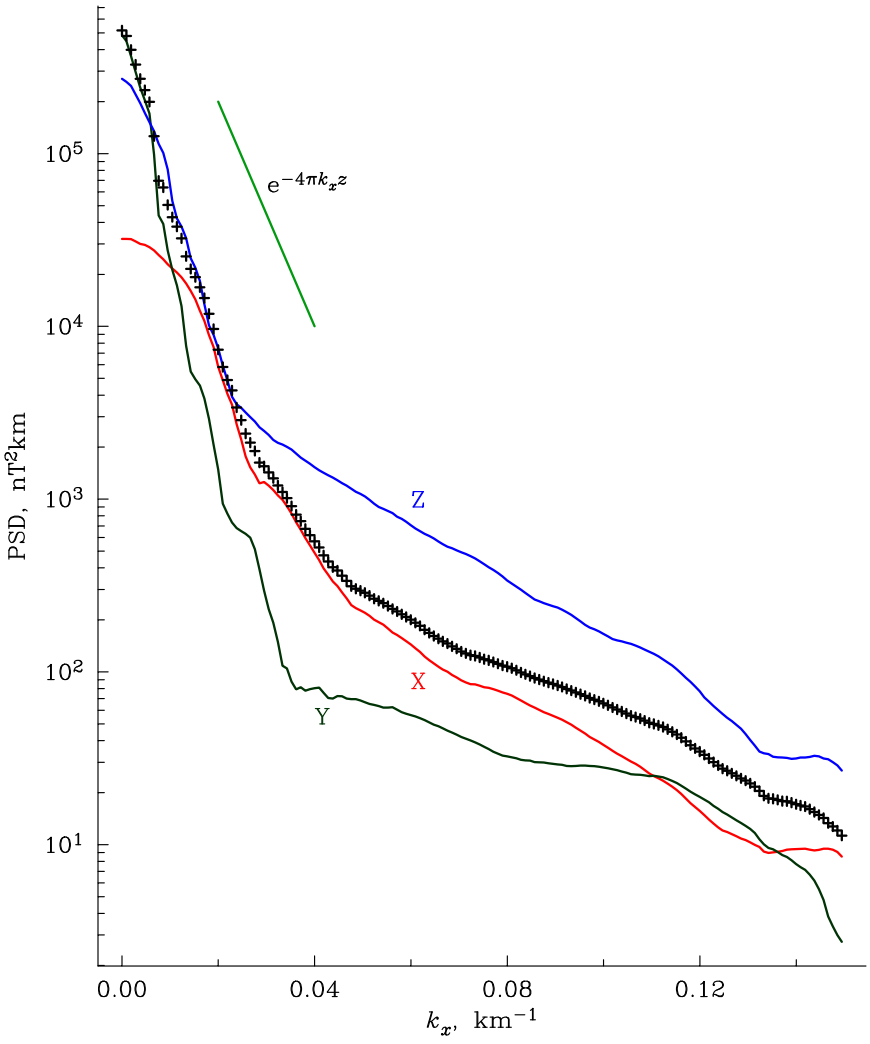
Looking at the one-dimensional spectra and cross spectra of measurements made through a random field of higher dimension is a very common practice in physical oceanography: the idea that you get a version that simply integrates over the unexplored directions is taken as obvious in that field, but this is not yet so in most of geophysics. Finally in this section, let's look at the estimated spectra for the data of Fig 7, which gives all three components (only $X$ and $Z$ were plotted before). This is part of great circle path flown from Easter Island to Bay St Louis.

**Figure 7:** Magnetic anomaly on a great circle path across the Pacific ocean.



This discussion is based closely on the paper of Parker and O'Brien *JGR*, 102, B11, pp 24815-24, 1997. The first interesting plot for us is Fig 8, the PSDs of all three components. Only the lowest tenth of the wavenumber spectrum are shown, since measurements of the field were made every 350 meters (What then is the Nyquist wavenumber?). All spectra fall off steeply, in fact exponentially to a first approximation, changing slope dramatically at about $k_x = 0.3$ km$^{-1}$. Approximate exponential decay like $\exp(-4\pi k_x z)$ is expected from (6.15)-(6.17) if $S_V$ is fairly flat. And the slope of the PSDs is about right initially: aircraft height plus water depth equals $7+4$ km, so $z = 11$ km. It is plausible that the field spectrum falls of faster since $S_V$ itself ought be a red spectrum, like most other geophysical PSDs. But clearly something happens when $k_x > 0.03$. Next look at the plus signs; these plot the value of $P_X + P_Y$. According to the Power Sum rule (6.19) the pluses should lie on top of the blue $P_Z$ line. Again things go as we might hope until 0.03 km$^{-1}$.

**Figure 8:** PSDs of the three components of the magnetic anomaly shown in Fig 7.

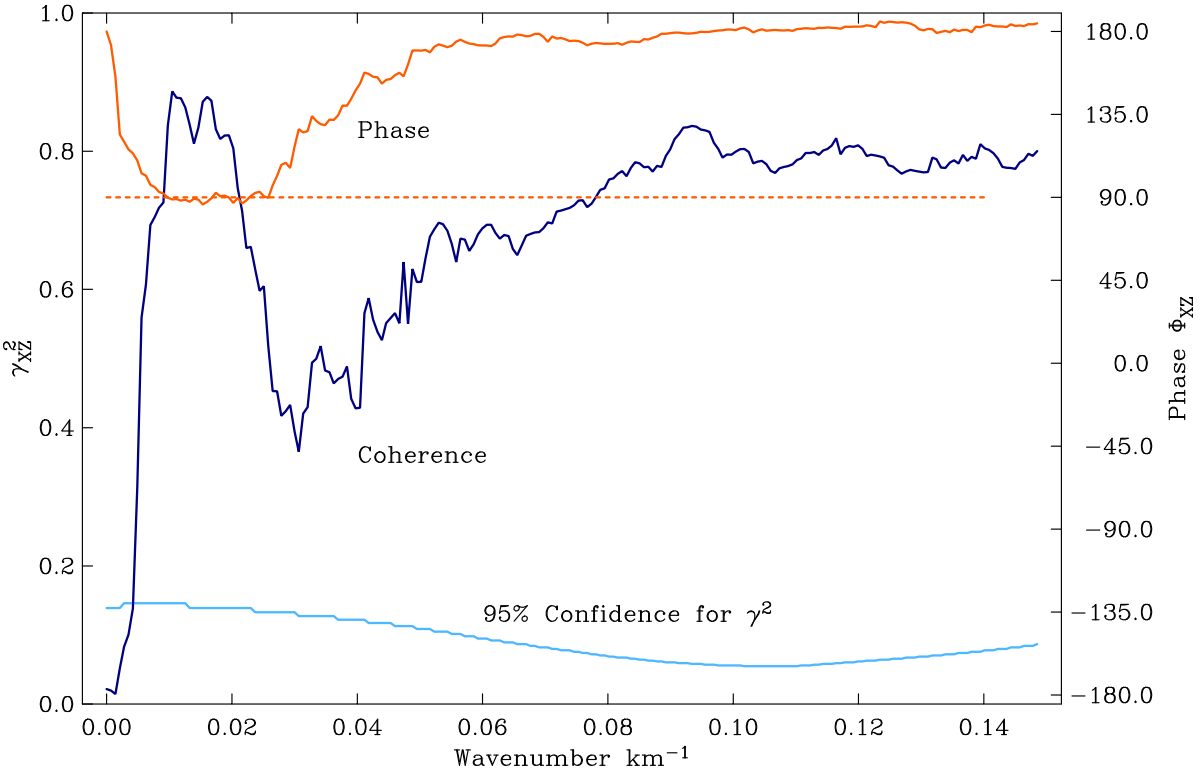


For further clues to solving this mystery we look at the cross spectra, here just at $X$ and $Z$. In Fig 9 we plot the phase and coherence spectra. The phase is about 90° for the longer wavelengths although it wanders off right near $k_x = 0$, which may be an estimation artifact, or a real issue. The theory we did early predicts the 90° phase for potential fields with sources below the plane. Like the PSDs this shows that only for $k < 0.03$ km$^{-1}$ (length wavelength > 33 km) are magnetic signals of crustal origin.

As O'Brien and I explain in the paper quoted, the cause of the noise is error in the orientation of the magnetometers. Fig 7 shows that the anomalies are about 100 nT a fraction of a percent of the main geomagnetic field, which is over 30,000 nT and it is roughly horizontal pointing north. When the gyro-stabilized platform shakes, a tiny fraction of this horizontal field appears on the $Z$ sensor and the other components are similarly corrupted, though less so; this is indicated clearly in Fig 8. Can you see that because the main field points upward in the southern hemisphere, the rocking platform causes a signal that is 180° out of phase between $Z$ and $X$, just as we see in Fig 9; high coherence is also predicted by this mechanism.

One lesson from the power and cross spectra is that 90% of the bandwidth of the record from this Project Magnet data is devoted to noise! Only signal with wavelength less than 30 km are geophysical, and wiggles on a smaller scale must not be interpreted as geology. For interpretational purposes we should filter the traces with a low-pass filter. Indeed

**Figure 9:** Coherence and phase spectra between $X$ and $Z$.



the very longest wavelength signals are also noise: on a flight of such long duration, the time varying magnetic field appear as long wavelength spatial signals, and these have not been properly corrected. Fig 8 shows the $Y$ component has more power than the $Z$ near $k_x = 0$, which the Power Sum rule says can never happen; the improper phase spectrum there also suggests a problem.

## 7. Stationary Processes on a Sphere

The following is one way of approaching the problem. It must be noted that, as usual in the literature involving spherical harmonics, everyone feels free to define different normalizations, so factors of $\pi$, $l$, $l+1$ and $2l+1$ appear in various places in the papers of various authors.

The proposition is that the statistics of a stationary random process on the unit sphere, called $S^2(1)$, cannot depend on the position $\hat{\mathbf{r}}$. We will assume that $V(\hat{\mathbf{r}})$, the process has zero mean, so that $\mathcal{E}[V]=0$ is obviously position independent. The second-order statistics are once again captured by the autocovariance:

$$\tilde{R_V}(\Delta) = \mathrm{cov}[V(\hat{\mathbf{r}}),\, V(\hat{\mathbf{s}})] \tag{7.1}$$

$$= \mathcal{E}[V(\hat{\mathbf{r}})\, V(\hat{\mathbf{s}})] \tag{7.2}$$

where $\Delta$ is the angle at the center of the sphere between the two unit vectors. On the sphere being independent of location means that every unit vector is the same as every other, that any reorientation of the sphere must leave $\tilde{R_V}$ unaltered. It is convenient to use the cosine of $\Delta$ rather than $\Delta$ itself in the definition:

$$R_V(\cos\Delta) = R_V(\hat{\mathbf{r}}\cdot\hat{\mathbf{s}}) = \mathcal{E}[V(\hat{\mathbf{r}})\, V(\hat{\mathbf{s}})] \tag{7.3}$$

If this is the autocovariance, what is the corresponding PSD? How do we decompose a function on a sphere into different wavelength components? Answer: spherical harmonics! The analog of the relation that the autocovariance is the FT of the PSD is the expansion of $R_V$ :

$$R_V(\hat{\mathbf{r}}\cdot\hat{\mathbf{s}}) = \sum_{l=0}^{\infty} S_l\, P_l(\hat{\mathbf{r}}\cdot\hat{\mathbf{s}}) \tag{7.4}$$

Here $S_l$ is the variance of $V$ in the part of the function with spherical harmonic degree $l$. Since $P_l(1)=1$ we see that

$$R_V(1) = \mathrm{var}[V] = \sum_{l=0}^{\infty} S_l \tag{7.5}$$

This result is the equivalent of the fact that the variance of random process on a line or in the plane is the area under the PSD. The sequence $S_l$ is my candidate for the **Spherical Power Spectrum**.

The inverse of (7.4), since Legendre polynomials are orthogonal is

$$S_l = (l+\tfrac{1}{2}) \int_0^{\pi} R_V(\cos\theta)\, P_l(\cos\theta)\, \sin\theta\, d\theta \tag{7.6}$$

$$= \sqrt{(2l+1)\pi} \int_{S^2(1)} R_V(\hat{\mathbf{z}}\cdot\hat{\mathbf{s}})\, Y_l^0(\hat{\mathbf{s}})^*\, d^2\hat{\mathbf{s}} \tag{7.7}$$

where the spherical harmonic is normalized to $\int |Y_l^m|^2 = \|Y_l^m\|^2 = 1$. Equations (7.6) and (7.7) are the analog to the fact that the PSD is the FT of the autocovariance.

An alternative definition proceeds as follows: imagine that the stationary process is written out as a spherical harmonic expansion on $S^2(1)$:

$$V(\hat{\mathbf{r}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_{lm} Y_l^m(\hat{\mathbf{r}}) \tag{7.8}$$

where $C_{lm}$ must be random complex coefficients. Plug this into (7.3). It turns out the only way that you can get $R_V$ to be independent of the vectors $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$ in (7.3) is to say that

$$\mathrm{cov}\,[C_{lm}, C_{nk}^*] = 0, \text{ unless } l = n, \text{ and } m = k \tag{7.9}$$

In other words the coefficients in the expansion must be uncorrelated random variables. Furthermore, we find that

$$\mathcal{E}\,[\,|\,C_{lm}\,|^2] = \sigma_l^2 \tag{7.10}$$

independent of the order $m$. Then from (7.3) and (7.8) we find

$$R_V(\hat{\mathbf{r}} \cdot \hat{\mathbf{s}}) = \mathcal{E}\,[V(\hat{\mathbf{r}})V(\hat{\mathbf{s}})] = \mathcal{E}\,[V(\hat{\mathbf{r}})V(\hat{\mathbf{s}})^*] \tag{7.11}$$

$$= \mathcal{E}\left[\sum_l \sum_m C_{lm} Y_l^m(\hat{\mathbf{r}}) \sum_n \sum_k C_{nk}^* Y_n^k(\hat{\mathbf{s}})^*\right] \tag{7.12}$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \mathcal{E}\,[\,|\,C_{lm}\,|^2] Y_l^m(\hat{\mathbf{r}}) Y_l^m(\hat{\mathbf{s}})^* \tag{7.13}$$

We used (7.9) to go from (7.12) to (7.13). Now we use the Spherical Harmonic Addition Formula together with (7.10)

$$R_V(\hat{\mathbf{r}} \cdot \hat{\mathbf{s}}) = \sum_{l=0}^{\infty} \frac{(2l+1)\sigma_l^2}{4\pi} P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{s}}) \tag{7.14}$$

Compare this to (7.4) and we see that

$$\sigma_l^2 = \frac{4\pi S_l}{2l+1} \tag{7.15}$$

The seemingly unnecessary factor of $4\pi$ here could be eliminated if we were willing to use spherical harmonics normalized as $\|\,Y_l^m\,\|^2 = 4\pi$.

The surprising fact is that on a sphere there are only isotropic stationary processes analogous to those described by equations (5.16-21), and none corresponding to the more general (5.1).

Spherical power spectra are used to describe the geomagnetic and gravitational potential fields, and also in seismology and geodynamics to characterize velocity and temperature structure at a particular radius within the Earth. One of the most successful uses of the stochastic model on the sphere has been geomagnetic field over geological time: if the dipole and a few of the other harmonics are excluded, the remaining field appears to be spatially stationary and to have a "white" spectrum (Constable and Parker, 1988).

## References

Berger, J., D. C. Agnew, Parker, R. L., and W. E. Farrell. Seismic system calibration, 2. Cross-spectral calibration using random binary signals. Bull. Seism. Soc. Am., 69, 271-288, 1979.

Bracewell, R. N., *The Fourier Transform and Its Applications*, McGraw-Hill, New York, 1978.

Constable, C. G. and Parker, R. L., Statistics of the geomagnetic secular variation for the past 5 m. y., J. Geophys. Res., 93, B10. 11569-81, 1988.

Egbert, G. D., and Booker, J. R., Robust estimation of geomagnetic transfer functions. GJRAS, 87, 173-94, 1986.

Gamble, T. D., Clarke, J., and Goubau, W. M., Magnetotellurics with a remote magnetic reference. *Geophysics,* 44, 53-68, 1979.

Kendall, M. G., and Stuart, A., *Advanced Theory of Statistics* Vol 2, Griffin, London, 1966.

Parker,R. L., and O'Brien, M. S., Spectral analysis of vector magnetic field profiles, *JGR*, 102, B11, pp 24815-24, 1997.

Percival, D. B., and Walden, A. T., *Spectral Analysis for Physical Applications - Multitaper and Conventional Univariate Techniques,* Cambridge, 1993.

Priestley, M. B., *Spectral Analysis and Time Series,* Academic Press, New York, 1981.

Rice, J. A., *Mathematical Statistics and Data Analysis,* Brooks-Cole Pub. Co., Monterey, CA, 1988.

Riedel, K. S., and Sidorenko, A., Minimum bias multiple taper spectral estimation, IEEE Trans. Sig. Proc., 43, 188-195, 1995.

Stevenson, J. M., Hildebrand, J. A., Zumberge, M. A. and Fox, C. G., An ocean bottom gravity study of the southern Juan de Fuca Ridge: J. Geophys. Res., 99, 4875-4888. 1994.

Watts, A. B., *Isostasy and Flexure of the Lithosphere*, Cambridge Univ. Press, Cambridge, 2001.