GEOPHYSICAL DATA ANALYSIS

Class Notes by Bob Parker

CHAPTER 2: SPECTRAL ANALYSIS OF STOCHASTIC PROCESSES

1. Spectral Analysis

At this point in the class it should not be a surprise that we will introduce a decomposition based on frequencies, a **spectral analysis.** Many things are simpler when looked at through these glasses – any time-invariant system, and solutions of differential equations with constant coefficients, have already been beaten to death with filters and Fourier transforms. As I mentioned elsewhere, most physical processes have behavior that can be characterized by the frequency of the variation, because which set of physical laws provide the dominant approximation depends on frequency. Recall my example about the magnetic field at a point on the Earth's surface: at frequencies of one over 1 million years (3×10^{-14} Hz) the physics is that of the slow moving fluids of the core dynamo; at 1 over a year (3×10^{-8} Hz) the field is governed by the behavior of the solar wind; at 1×10^{8} Hz the magnetic field most likely originates from a local TV station. Given a natural process with a random appearance, one naturally would like to decompose it into components of various frequencies, in effect, to take its Fourier transform.

To create a viable theoretical basis we need to consider stationary stochastic processes, which are random functions extending throughout all time with time-invariant properties. For functions of a real variable (continuous-time signals) we have the classical Fourier transform:

$$\hat{x}(f) = \int_{-\infty}^{\infty} x(t)\, e^{-2\pi i f t}\, dt\,. \tag{1.1}$$

For sequences (discrete-time signals) we have the spectrum on the finite interval (–½, ½):

$$\hat{x}(f) = \sum_{n=-\infty}^{\infty} x_n\, e^{-2\pi i f n},\ -\tfrac{1}{2} \le f \le \tfrac{1}{2} \tag{1.2}$$

which allows reconstruction by the spectral representation:

$$x_n = \int_{-1/2}^{1/2} \hat{x}(f)\, e^{2\pi i f n}\, df\,. \tag{1.3}$$

Can we apply either of these to their corresponding stochastic processes? The answer is no, because to make sense, the integral in (1.1) or the sum in (1.2) must converge, which requires some kind of decrease in amplitude, or energy, as t or n gets large. For a stationary process, that does not happen.

Also, if we put a stochastic process in $f(t)$ in (1.1) we would obtain another random function. Our goal is to characterize $X(t)$ with an ordinary function describing its properties in frequency (as the autocorrelation function does in time) not generate another random process. We can see in an intuitive way what we require, as follows. Suppose we wanted to know how much variability the stationary process exhibits at a frequency f_0. We could build (or design) a narrow band-pass filter ϕ_{f_0} that only allowed signals through in the frequency band $(f_0 - \frac{1}{2}df, f_0 + \frac{1}{2}df)$ with unit gain. Now send the signal X through that filter, and out would come a stochastic process which would have a very limited frequency content. We measure its amplitude by the variance; we would expect the variance to proportional to the width df of the band-pass filter. Then define

$$S_X(f_0)\, df = \mathrm{var}\,[\phi_{f_0} * X] \tag{1.4}$$

that is the variance of the filtered process will be some positive value times df; it will vary as the center frequency f_0 is varied, and be proportional to the variance in X at that frequency. The variance of X in a frequency band is called the **power** in that band and so S_X is the **Power Spectrum** of X, or more grandly its **Power Spectral Density.** Equation (1.4) is our informal definition of $S_X(f_0)$. Notice this definition works equally well for continuous or discrete processes. In the days before computers, analog spectral analysers were built based on this principle: a large number of narrow band-pass filters followed by rectifiers to measure the variance in each band.

2. Two Definitions of the PSD

We begin with a strict definition of the **Power Spectral Density** (PSD) of a stationary process as a kind of Fourier Transform. Let us study the definition for a continuous stationary process. In what follows we will always assume the process $X(t)$ has **zero mean.** Two problems with the ordinary FT were noted in Section 1: (a) the FT of X would not be defined on the infinite interval, and (b) the answer would be a random process, not a statistic of X. We fix these two things as follows: first define $X_T(t)$ as the process $X(t)$ on the finite interval $(-T, T)$:

$$X_T(t) = \begin{cases} X(t), & -T \le t \le T \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Any particular realization of this process (which is not stationary) has bounded 2-norm and thus has an ordinary FT:

$$\hat{X}_T(f) = \mathcal{F}\,[X_T](f) = \int_{-\infty}^{\infty} X_T(t)\, \mathrm{e}^{-2\pi i f t}\, dt = \int_{-T}^{T} X_T(t)\, \mathrm{e}^{-2\pi i f t}\, dt\,. \tag{2.2}$$

Note that $\hat{X}_T$ is still a random function of f, however. So we find its magnitude, square it, and take the expected value: $\mathcal{E}\,[\,|\hat{X}_T(f)|^2]$. Our plan is to let T tend to infinity, but we can easily see that this number would grow to finity, and so we divide by the interval length $2T$ to tame the growth: we define the function of frequency

$$S_X(f) = \lim_{T\to\infty} \frac{1}{2T}\,\mathcal{E}\,[\,|\hat{X}_T(f)|^2] \tag{2.3}$$

$$= \lim_{T\to\infty} \mathcal{E}\left[\frac{1}{2T}\left|\int_{-T}^{T} X_T(t)\,\mathrm{e}^{-2\pi i f t}\,dt\right|^2\right]. \tag{2.4}$$

This is the definition of the PSD; it can be shown to exist for all stationary processes X with zero mean and a bounded variance. It is obviously real and non-negative.

Equation (2.4) defines what is often called the **two-sided** PSD, because we allow f to run from $-\infty$ to ∞. When X is real, the usual case, it is easily seen from the well-known properties of the FT that $S_X(f)$ is an even function of f and therefore only values for $f \geq 0$ need be specified. Quite commonly a **one-sided PSD** is used, given by $2S_X(f)$ for $f \geq 0$; we will see in a moment why it is convenient to put in the factor of two.

Looking at (2.4) we can observe that S_X at any particular f is obtained as some kind of second order moment of X – only products of X with itself are needed, no third order moments enter. We have already introduced another second-order moment of X, the autocovariance. Does S_X provide independent information about X, or is there a connection between R_X and S_X? Somewhat surprisingly, to me anyhow, is the following answer: the functions $R_X(t)$ and $S_X(f)$ *contain exactly the same information.* In fact, S_X is the FT of R_X

$$S_X(f) = \mathcal{F}\,[R_X] = \int_{-\infty}^{\infty} R_X(t)\,\mathrm{e}^{-2\pi i f t}\,dt \tag{2.5}$$

Equation (2.5) is sometimes used as an alternative the definition of the PSD.

Before we establish the truth of (2.5) we observe a few consequences. Since R_X is a real even function of t, (2.5) implies that S_X is a real and even in f. But the fact that S_X must be non-negative puts severe restrictions on what functions R_X are allowed to be autocovariances; clearly not every even R_X with an FT is going to have a positive FT. Now take the inverse transform of (2.5):

$$R_X(t) = \int_{-\infty}^{\infty} S_X(f)\,\mathrm{e}^{2\pi i f t}\,df\,. \tag{2.6}$$

Now recall from the definition of R_X that

$$R_X(0) = \mathcal{E}\,[X(t)\,X(t)] = \mathrm{var}\,[X] = \sigma_X^2 \tag{2.7}$$

remembering that X is a zero-mean process. Setting $t = 0$ in (2.5) gives the **important result:**

$$\sigma_X^2 = \int_{-\infty}^{\infty} S_X(f)\, df\,. \tag{2.8}$$

In words, *the area under the power spectrum is the process variance.* That is why we double S_X if we use the one-sided PSD, to preserve this property.

Now we verify (2.5). We start with the squared magnitude of the FT of X_T:

$$|\hat{X}_T|^2 = \hat{X}_T \hat{X}_T^* \tag{2.9}$$

Now recall that the FT of a convolution is the product of the FTs; further notice that, since X_T is real

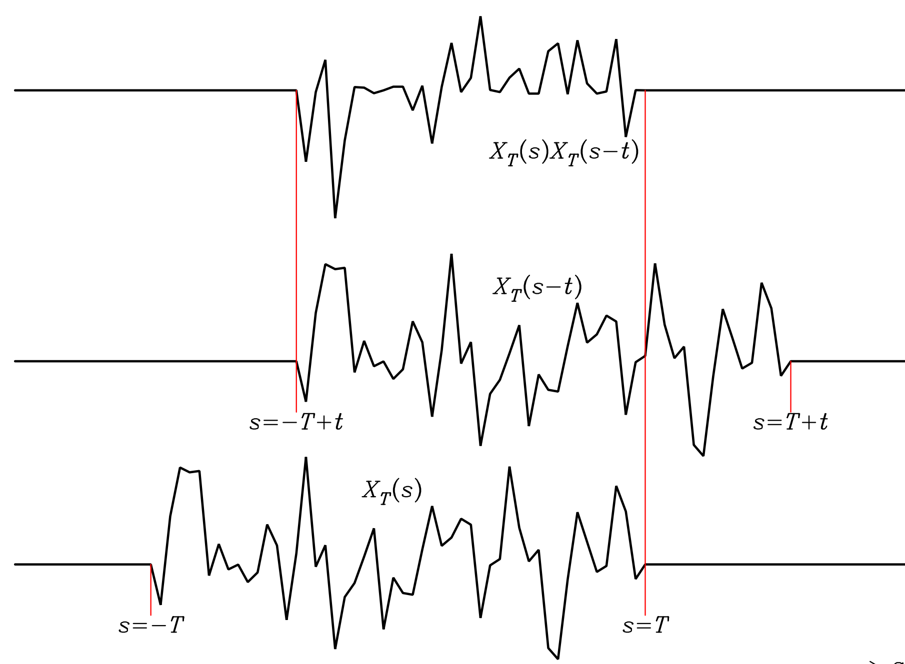
$$\hat{X}_T(f)^* = \int_{-\infty}^{\infty} X_T(t)\, e^{2\pi i f t}\, dt = \int_{-\infty}^{\infty} X_T(-t) e^{-2\pi i f t}\, dt = \mathcal{F}\,[X_T(-t)] \tag{2.10}$$

Combining the Convolution Theorem with (2.10) and (2.9), we have

$$|\hat{X}_T|^2 = \mathcal{F}\,[X_T(t) * X_T(-t)] = \mathcal{F}\,[\int_{-\infty}^{\infty} X_T(s)\, X_T(s-t)\, ds] \tag{2.11}$$

In (2.3) we have normalized by the interval $2T$. So let us put that into the

Figure 1: The integrand of (2.12).



definition of another function:

$$R_T(t) = \frac{1}{2T} X_T(t) * X_T(-t) = \frac{1}{2T} \int_{-\infty}^{\infty} X_T(s)\, X_T(s-t)\, ds \tag{2.12}$$

Then by (2.11)

$$\frac{|\hat{X}_T|^2}{2T} = \mathcal{F}\,[R_T] = \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi i f t}\, R_T(t)\, dt \tag{2.13}$$

Our definition of the PSD is (2.3); let us plug (2.13) into that

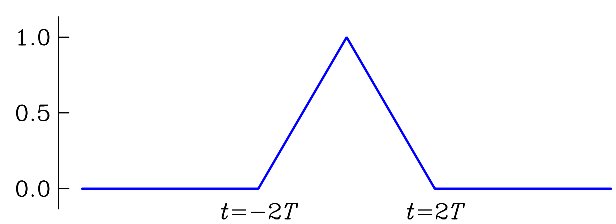
$$S_X(f) = \lim_{T\to\infty} \mathcal{E}\left[\frac{|\hat{X}_T|^2}{2T}\right] = \lim_{T\to\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-2\pi i f t} \mathcal{E}\,[R_T(t)]\, dt \tag{2.14}$$

From (2.12) we see that $R_T(t)$ is even in t and so we can always write $R_T(t) = R_T(|t|)$; in the following we will assume $t \geq 0$ and then replace t by $|t|$ at the end. We know $X_T(s)$ vanishes outside the interval $(-T, T)$ and therefore the integrand of (2.12) must vanish when $s > T$ or when $|s-t| > T$; see Figure 1. Therefore we can reduce the interval of integration in (2.12) to be on $(-T+t,\, T)$ instead of the whole real line. Also observe that once $t > 2T$ the nonzero sections cease to overlap, and the integrand is identically zero. These considerations lead to

$$R_T(t) = \begin{cases} \dfrac{1}{2T} \displaystyle\int_{-T+t}^{T} X_T(s)\, X_T(s-t)\, ds, & 0 \leq t < 2T \\ 0, & t \geq 2T\,. \end{cases} \tag{2.15}$$

Further simplifications ensue when we take the expected value, as dictated by (2.14); for the segment $0 \leq t < 2T$

Figure 2: The function $\Lambda_T(t)$.



$$\mathcal{E}\,[R_T(t)] = \frac{1}{2T}\int_{-T+t}^{T} \mathcal{E}\,[X_T(s)\,X_T(s-t)]\;ds = \frac{1}{2T}\int_{-T+t}^{T} R_X(-t)\;ds \quad (2.16)$$

where we have introduced R_X, the autocovariance of the process. Since $R_X(-t) = R_X(t)$, which is independent of s, we can evaluate s the integral explicitly:

$$\mathcal{E}\,[R_T(t)] = \frac{R_X(t)}{2T}\int_{-T+t}^{T} 1\cdot ds = R_X(t)\left[1 - \frac{t}{2T}\right], \quad 0 \le t \le 2T \qquad (2.17)$$

From (2.15) $\mathcal{E}\,[R_T(t)] = 0$ when $t \ge 2T$. Recalling that R_T is even, we can the negative t behavior from $R_T(t) = R_T(-t)$, and obtain the following complete description for the expected value of R_T:

$$\mathcal{E}\,[R_T(t)] = R_X(t)\,\Lambda_T(t) \qquad (2.18)$$

where

$$\Lambda_T(t) = \begin{cases} 1 - |t|/2T, & |t| \le 2T \\ 0, & |t| > 2T\,. \end{cases} \qquad (2.19)$$

See Figure 2 for a sketch of $\Lambda_T(t)$. Substituting (2.18) into (2.14) gives us this very plausible expression for the PSD:

$$S_X(f) = \lim_{T\to\infty} \mathcal{F}\,[R_X(t)\Lambda_T(t)] = \lim_{T\to\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-2\pi i f t} R_X(t)\,\Lambda_T(t)\;dt \qquad (2.20)$$

If it is permitted to put the limit in (2.20) inside the integral we have the result we predicted, equation (2.5), since $\Lambda_T(s) \to 1$ as $T \to \infty$. Priestley (*Spectral Analysis and Time Series*, p 213-4) uses the Lebesgue Dominated Convergence Theorem, and the further condition that

$$\int_{-\infty}^{\infty} |R_X(t)|\;dt < \infty \qquad (2.21)$$

to prove that it is permitted to reverse the order of the limit and the integral. For those interested, I give in the Appendix a proof of my own that makes a different set of assumptions about R_X.

3. Some Properties of the PSD

We continue to consider only continuous processes. By establishing the key fact that the autocovariance function and PSD are Fourier transforms of each other, we have shown that they contain the same information. But as we will see later, the PSD is much more useful for the interpretation of actual data because of the intuitive idea that a process divides up naturally as a sum of processes with different frequencies. So while the PSD is the property of the stationary process that is the most informative, the definition via the limit, equation (2.3) is usually very awkward to handle, and the alternative relation (2.5) through autocovariance is the most useful for doing theory, as we will illustrate.

First the simplest example, take white noise. Recall for a continuous process, a white noise has a delta function autocovariance, equation (3.2) in Chapter 1. Then The PSD is

$$S_W(f) = \int_{-\infty}^{\infty} R_W(t)\, e^{-2\pi i f t}\, dt = \int_{-\infty}^{\infty} s^2 \delta(t)\, e^{-2\pi i f t}\, dt = s^2 \tag{3.1}$$

which is a constant independent of frequency. White noise has the same power at every frequency, which is why the term is borrowed from physics, because white light is ideally composed of light with the same property, equal power at each frequency. Recall from (2.8) that the area under the PSD is the process variance. For white noise that is infinite, so continuous time white noise is not a physically realizable phenomenon. Notice that the different white noises of Chapter 1 all have exactly the same frequency content, a flat spectrum. The PSD is simply a second order property, and does not concern itself with other details of the distribution defining the stationary process.

Let us consider next what happens to the PSD of process if it has been filtered. In the continuous parameter case we will say

$$Y = g * X \tag{3.2}$$

where g is a filter function, possibly infinite in extent. When we know S_X we can calculate S_Y. As advertised the simplest approach is to compute the autocovariance of Y. Assume X has mean zero, then so does Y, and

$$R_Y(t) = \text{cov}\,[Y(s),\, Y(s+t)] = \mathcal{E}\,[Y(s)\,Y(s+t)] \tag{3.3}$$

$$= \mathcal{E}\left[\int_{-\infty}^{\infty} du\ g(u)\,X(s-u) \int_{-\infty}^{\infty} dv\, g(v)\,X(s+t-v)\right] \tag{3.4}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{E}\,[X(s-u)\,X(s+t-v)]\ g(u)\,g(v)\ du\ dv \tag{3.5}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_X(t+u-v)\; g(u)\, g(v)\; du\; dv\,. \tag{3.6}$$

Next we replace the autocovariance on the right with its representation in terms of S_X: we apply (2.6):

$$R_Y(t) = \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} df\;\; S_X(f)\, \mathrm{e}^{2\pi i f(t+u-v)}\; g(u)\, g(v) \tag{3.7}$$

$$= \int_{-\infty}^{\infty} df\;\; \mathrm{e}^{2\pi i ft}\; S_X(f) \int_{-\infty}^{\infty} du\;\; \mathrm{e}^{2\pi i fu}\; g(u) \int_{-\infty}^{\infty} dv\;\; \mathrm{e}^{-2\pi i fv}\; g(v) \tag{3.8}$$

$$= \int_{-\infty}^{\infty} df\;\; \mathrm{e}^{2\pi i ft}[S_X(f)\, \hat{g}(f)^*\; \hat{g}(f)] \tag{3.9}$$

$$= \mathcal{F}^{\,-1}[S_X(f)\, \hat{g}(f)^*\; \hat{g}(f)] \tag{3.10}$$

If we take the FT of (3.10) and again recognize the FT of R_Y is the PSD S_Y we see

$$S_Y(f) = \hat{g}(f)^*\, \hat{g}(f)\, S_X(f) \tag{3.11}$$

$$= |\, \hat{g}(f)\,|^2\, S_X(f)\,. \tag{3.12}$$

This is an **important result.** It is the stochastic process version of the Convolution Theorem applied to filtering: when one filters a deterministic signal, the FT of the resultant function is multiplied by the frequency response of the filter. Here the new spectrum is found by multiplying the original by the *squared magnitude* of the filter function. Equation (3.12) also puts on a firm foundation the idea described in Section 1 that we can obtain a power spectrum by applying a series of ideal, narrow band-pass filters the stochastic process.

4. PSD of Discrete Processes

Every result for a continuous-time stationary stochastic process has an analog in the discrete theory, and there are no surprises. I will state the results here without derivation. Here is the definition of the PSD as a limit of finite Fourier transforms:

$$S_X(f) = \lim_{N \to \infty} \frac{1}{2N} \mathcal{E}\,[\,|\sum_{n=-N}^{N} X_n e^{-2\pi i n f}\,|^2], \quad -½ \le f \le ½\,. \tag{4.1}$$

The alternative definition through the autocovariance is

$$S_X(f) = \sum_{n=-\infty}^{\infty} R_X(n)\, e^{-2\pi i n f}, \quad -½ \le f \le ½ \tag{4.2}$$

and of course one can obtain the autocovariance from the PSD with the coefficients of the Fourier series expansion in (4.2)

$$R_X(n) = \int_{-½}^{½} S_X(f)\, e^{2\pi i n f}\, df\,. \tag{4.3}$$

Setting $n = 0$ in (4.3) gives

$$\sigma_X^2 = \int_{-½}^{½} S_X(f)\, df \tag{4.4}$$

so that the variance is again the integral of the PSD over frequency. If we restrict ourselves to convolution filters, then filtering a discrete sequence gives the power spectrum

$$S_{g\,*\,X}(f) = |\,\hat{g}(f)\,|^2\, S_X(f) \tag{4.5}$$

where

$$\hat{g}(f) = \sum_{n=-\infty}^{\infty} g_n\, e^{-2\pi i n f}\,. \tag{4.6}$$

5. Aliasing in the PSD

Although we have seen how aliasing affects sampled data earlier, we should see how sampling modifies the PSD for stochastic processes. This is an important issue in the real world because sampled data may be the only record we have of an underlying continuous physical signal whose power spectrum we would like to know. We call a continuous-time signal $Y(t)$, and we derive from it the discrete process:

$$X_n = Y(n\Delta t), \quad n = 0, \pm 1, \pm 2, \cdots \tag{5.1}$$

where Δt is the sampling interval. As I asserted earlier it is almost always easier to derive properties of the spectrum from the autocovariance. So the autocovariance of the sampled series is

$$R_X(n) = \mathcal{E}\,[X_j\, X_{j+n}] \tag{5.2}$$

$$= \mathcal{E}\,[Y(j\Delta t)\, Y((j+n)\Delta t) = R_Y(j\Delta t)\,. \tag{5.3}$$

Thus the autocovariance of X_n is simply the sampled autocovariance of Y.

The PSD of the discrete process is given by (4.2), which we modify by including Δt in the exponent to scale the frequencies, replacing the Nyquist frequency of ½ in (4.2) by $1/2\Delta t$, and also scaling the expression by the same quantity:

$$S_X(f) = \Delta t \sum_{n=-\infty}^{\infty} R_X(n)\, e^{-2\pi i n f \Delta t}, \quad -1/2\Delta t \le f \le 1/2\Delta t\,. \tag{5.4}$$

Now substitute (5.3)

$$S_X(f) = \Delta t \sum_{n=-\infty}^{\infty} R_Y(n\Delta t)\, e^{-2\pi i n f \Delta t}\,. \tag{5.5}$$

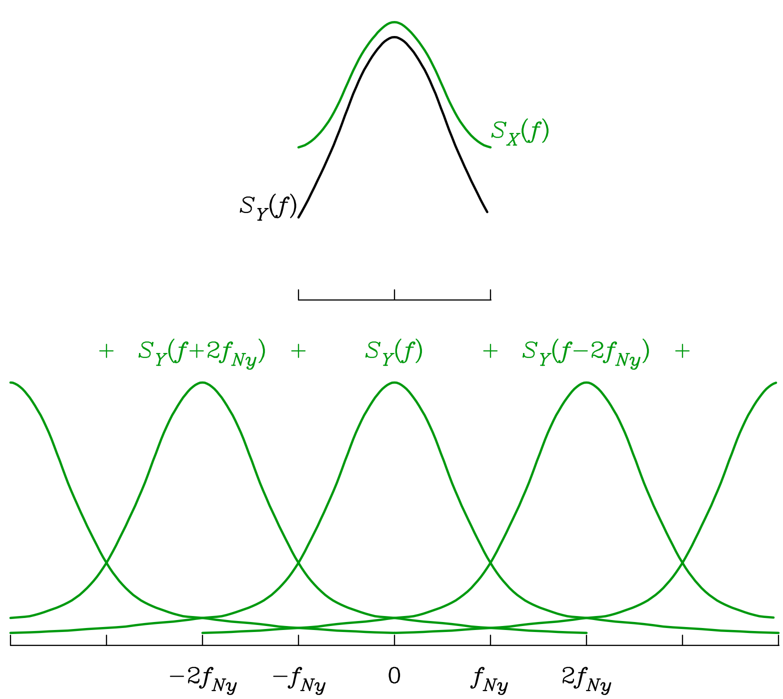
To sum this series we appeal to the **Poisson Sum Rule** given in our treatment of Fourier Theory

$$S_X(f) = \Delta t \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} R_Y(n\Delta t)\, e^{-2\pi i n f \Delta t}\, e^{-2\pi i n m}\, dn\,. \tag{5.6}$$

We change variables in the integral: set $t = n\Delta t$ and then recognize the definition of the continuous process PSD:

$$S_X(f) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} R_Y(t)\, e^{-2\pi i t(f + m/\Delta t)}\, dt \tag{5.7}$$

Figure 3: Aliasing in the PSD



$$= \sum_{m=-\infty}^{\infty} S_Y(f + m/\Delta t)\,. \tag{5.8}$$

Thus the discrete process PSD is a sum of spectra of the original continuous process, shifted by multiples of **twice the Nyquist frequency.** If the power in the continuous process has fallen off to low levels at the Nyquist frequency, the PSD of S_X will be a good approximation to S_Y, although in general S_X will be a factor of two or more above the “true” PSD at $f = 1/2\Delta t$. The moral is that to get a good PSD one must set the sampling rate high enough to avoid aliasing.

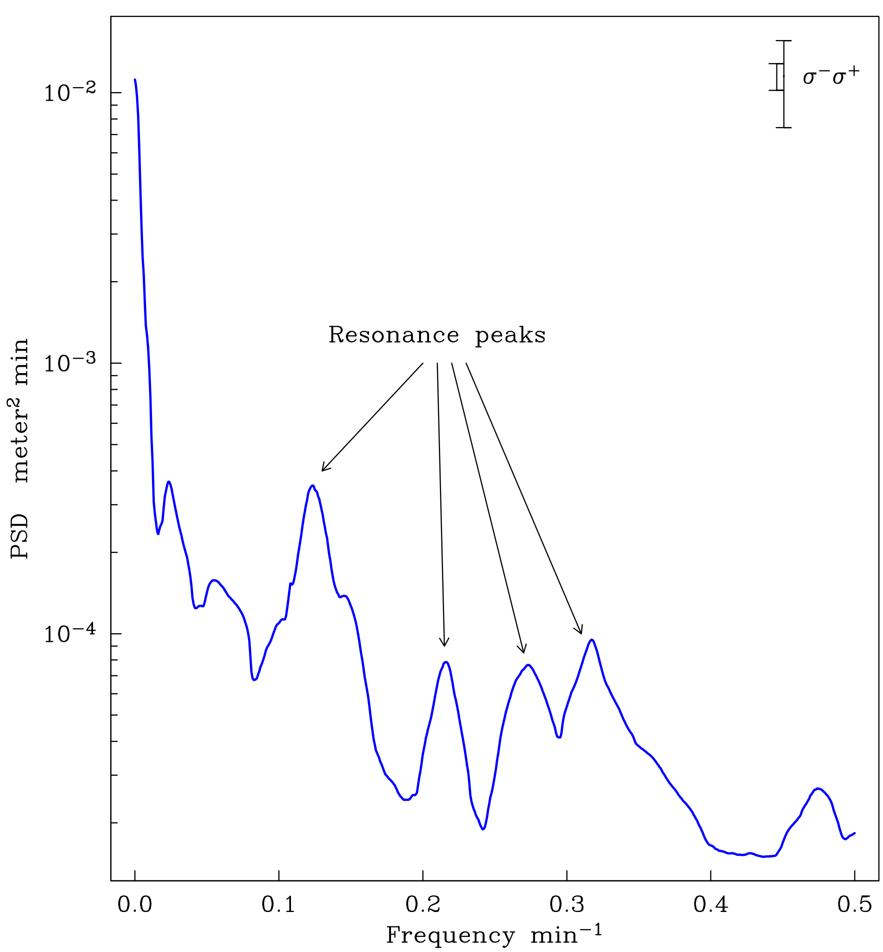
6. Illustrations

Let us look briefly at some PSDs estimated from real-world data. Exactly how the PSD is estimated is the subject of Chapter 3; it is a nontrivial topic which we must defer.

We begin with the Salton Sea seiche data. Below is a an estimate of the PSD of the 1301 data points. As I mentioned in Chapter 1 the data time series suggests a number of resonances excited by the wind. In the PSD we see a series of peaks, well resolved in frequency. The peaks are far from being delta-function lines, and this not a fault of the estimation process. The lake resonances are not perfect since we expect them to broad features due to frictional losses. The four peaks picked out are certainly significant, but other smaller ones are suggestive, though not delineated with confidence. I have plotted two bars, both representing the 1 standard deviation error estimates; the smaller one for places where the spectrum is smooth, the larger for regions of rapid variation.

You should notice several other important things about this graph. First it is plotted on linear-frequency, log-PSD axes. The spectrum covers a large dynamic range, which is almost universal in geophysical data. Plotted on a linear y axis nothing except the left side of the plot would be visible to the eye. **Always** plot your spectra with a log scale. A log

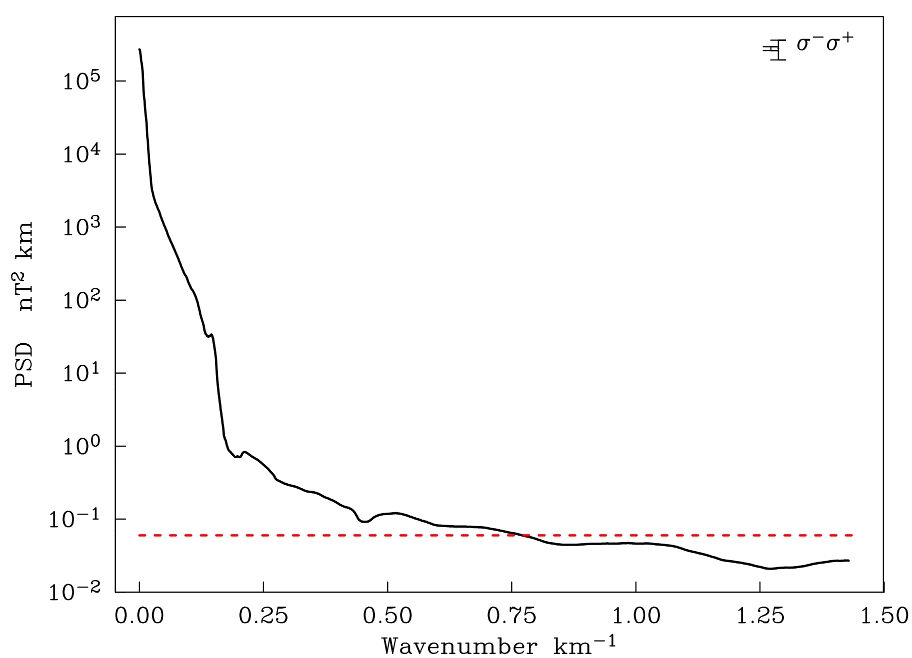
Figure 4: PSD of Salton Sea seiche data.



frequency axis is sometimes useful too, but not in this example. Next observe that rise in power at the low-frequency end. Such a rise is called a **red spectrum** because in optics red light has its power concentrated towards long wavelengths. Red spectra are the norm in geophysical work. This is because attenuation processes operate more efficiently at high frequencies or short wavelengths; most natural filters (like seismic losses, or upward continuation in potential fields) are low-pass filters. Finally, notice the units of the power spectrum: PSD is a measure of variance per unit frequency. Since the frequency here is one cycle per minute, and the series measures water depth, the units are those of squared-length multiplied by time. The Power Theorem for Fourier decomposition means that one could, for example, estimate how much kinetic energy is tied up in the largest resonance by finding the area under the PSD in an appropriate frequency range.

Although the autocorrelation function holds exactly the same information as the PSD, a plot of R_X is almost always useless. We plotted the autocovariance for one component of the magnetic data from the plane flying across the Pacific (Figure 7, Section 4): about the only thing one can deduce from it is that the data series is far from white noise. Let us look at the PSD. This will take two graphs. In Figure 5 I show the whole spectrum. Notice again the very large dynamic range in power. The field values are sampled at an interval of 350 m, so $f_{Ny} = 1/(2 \times 0.35) = 1.43$ km^{-1}. The magnetometer is reported as having an accuracy of ±1 nT. If we take this to mean there is uncorrelated noise with a uniform PDF, with $b = 1$ from Chapter 1 (3.5), we find the variance is $1/12 = 0.0833\,\mathrm{nT}^2$. The

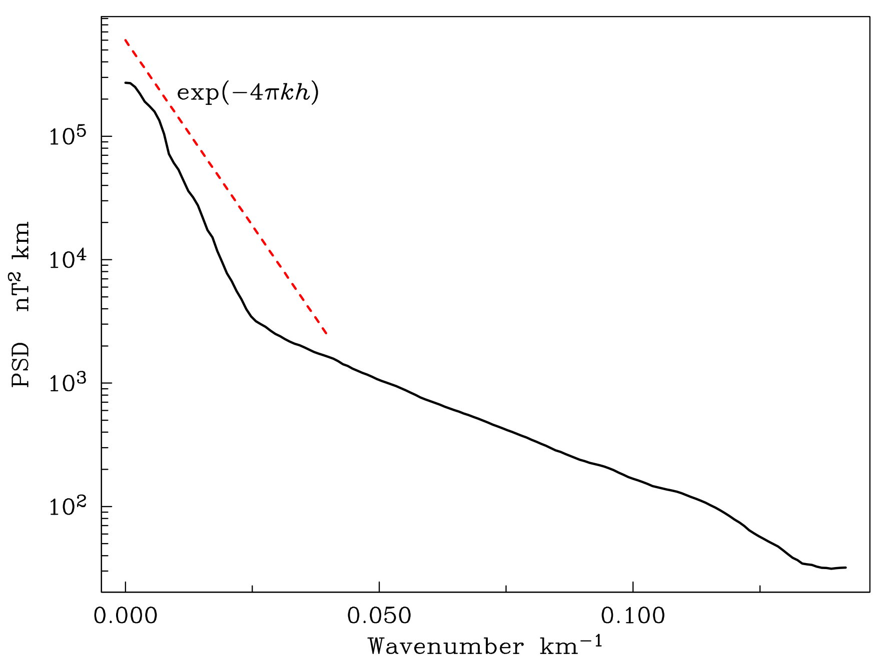
Figure 5: PSD of Z component in magnetic data.



dashed horizontal line corresponds to white noise with this variance: Figure 5 is a one-sided PSD plot, so the area under the dashed line is $\sigma_Z^2 = 1.43 \times 0.0582 = 0.0833\,\mathrm{nT}^2$. What this means is that two-thirds of the frequency content of the record is devoted to almost pure white noise, and is completely uninformative about geophysics — the true geophysical signal at wavenumbers greater than 0.5 km^{-1} (wavelengths shorter than 2 km) is evidently of such low amplitude, the noise of the magnetometer totally obscures it.

In Figure 6 I expand the frequency scale by a factor of ten to show the small wavenumber part of the spectrum more clearly. We see the PSD at this scale approximates two intersecting straight lines. When one looks into the theory of upward continuation of static magnetic fields the process is essentially that of low-pass filtering (as mentioned earlier) — roughly speaking, if $B_z(x)$ is the Z component on a profile at the level of the magnetic sources, then on a path of height h above the first line the field becomes $g * B_z$ where g is filter with response $\hat{g}(k) = \exp(-2\pi kh)$, an exponential fall with wavenumber k. Recall that the effect of the filter on the PSD is to square this response. I have plotted a line corresponding to $|\hat{g}|^2$ in Figure 6 for $h = 7 + 4 = 11\,\mathrm{km}$, the aircraft height plus the average ocean depth. It is plausible to assume that the PSD near the crustal sources of the field falls off fairly slowly, and after upward continuation the spectrum fits that prediction quite well. What then is the other straight-line segment? It can be shown that lack of stability in the gyro system orienting the coordinates for the measurements causes this effect

Figure 6: Lowest 10% in wavenumber of Figure 5.



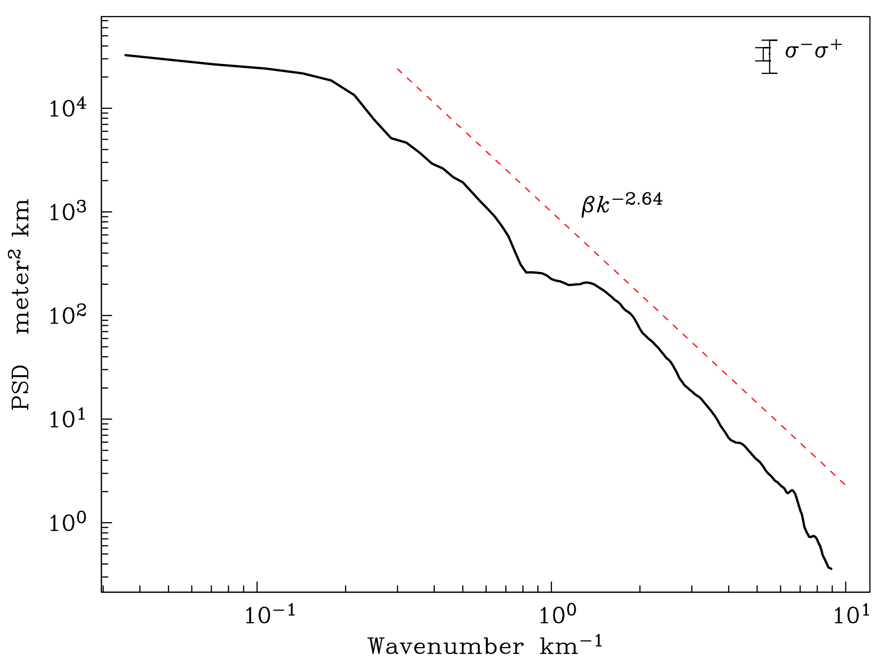
through nonlinear processes; for anyone interested, see Parker and O'Brien, *JGR,* v 102, pp 24815-24, 1997. So it turns out by looking at the power spectrum we have discovered that nothing above $k = 0.03\,\mathrm{km}^{-1}$ concerns geophysics in the observations; only the bottom one-thirtieth of the spectrum contains geophysical information, the rest is noise or one kind or another! It is worth remembering that a leveling off in the PSD at high frequency is often an indication that instrument noise has overwhelmed the signal under investigation at that point, and that the high frequency behavior in the record is probably not geophysical. Unfortunately, interpretation of high-frequency wiggles is a widespread occupation in data analysis.

We have seen in this example how the PSD has revealed a number of remarkable things about the original signal, properties that could not be deduced by visual inspection from the original sequence (shown in Chapter 1 Figure 6) nor from the autocovariance function in Figure 7. More still can be learned by combining the X, Y and Z components in their **cross spectra**, but we don't yet have the theory in hand for that.

Finally, we look at the PSD estimated from the stationary part of the marine bathymetry data set; this is shown in Figure 7. Notice here I have used a log wavenumber axis. The reason for this is to show that the power spectrum of the profile is quite well approximated by a straight line which on this graph means that the PSD has the form of a power law:

$$S_H(k) = \beta\, k^{-q} \tag{6.1}$$

Figure 7: PSD of ocean rise bathymetry.



A fit to the spectrum gives the estimate $q = 2.64$. Power-law behavior has the characteristic that there is no intrinsic length scale. This means that the stochastic process must look the same at any magnification — the hall-mark of a **fractal.** This idealization cannot be true for all scales, and clearly it breaks down in the graph at wavelengths longer than about 3 km. The geological process responsible for the terrain is repeated fracturing and faulting, and there are some theories (Fox and Hayes, *Reviews of Geophysics,* v 23, pp 1-48, 1985) predicting fractal behavior, although I do not believe it is possible to predict the exponent q very well.

Appendix: Proof of Equation (2.20)

In Section 2 I left the proof of the validity of the interchange of the integral and limit to a reference to Priestley. There is a reason why Priestley's proof is not entirely satisfactory, namely, the additional restriction (2.21) is too severe since it excludes autocovariance functions behaving like the sinc function, those associated with a discontinuity in S_X. Here is an alternative proof.

We examine the difference between the desired result and the function within the limit, and show that the difference vanishes under some mild restrictions, as T becomes large. We measure the discrepancy with the 2-norm:

$$\| f(t) \|^2 = \int_{-\infty}^{\infty} | f(t) |^2 \, dt \,. \tag{A1}$$

Define the number

$$\Delta_T = \| \mathcal{F}[\Lambda_T R_X - R_X] \| = \| \Lambda_T R_X - R_X \| \,. \tag{A2}$$

The second equality follows from the Power Theorem, the invariance of the 2-norm under the FT. Then, since Λ_T and R_X are both real and even

$$\Delta_T^2 = 2 \int_0^{\infty} (\Lambda_T(t) - 1)^2 R_X(t)^2 \, dt \tag{A3}$$

$$= 2 \int_0^{2T} (\Lambda_T(t) - 1)^2 R_X(t)^2 \, dt + 2 \int_{2T}^{\infty} R_X(t)^2 \, dt \tag{A4}$$

$$= 2 \int_0^{2T} \frac{t^2}{4T^2} R_X(t)^2 \, dt + 2 \int_{2T}^{\infty} R_X(t)^2 \, dt \,. \tag{A5}$$

To proceed we need to assume something more about the behavior of R_X. We know R_X is never greater than σ_X^2, so it is bounded; we will assume that it dies away for large t, but more rapidly than some power:

$$| R_X(t) | < \frac{c}{(1+t)^{\nu}} \tag{A6}$$

for some fixed values of c and ν. With this constraint we can see that

$$\Delta_T^2 < \frac{1}{2T^2} \int_0^{2T} \frac{c^2 t^2}{(1+t)^{2\nu}} \, dt + 2 \int_{2T}^{\infty} \frac{c^2}{(1+t)^{2\nu}} \, dt = \frac{A(T)}{T^2} + B(T) \,. \tag{A7}$$

Our interest lies in the behavior of Δ_T as T tends to infinity. By L'Hopital's rule of elementary calculus on the first term:

$$\lim_{T \to \infty} \frac{A(T)}{T^2} = \lim_{T \to \infty} \frac{A'(T)}{2T} \tag{A8}$$

$$= \lim_{T \to \infty} \frac{2Tc^2}{(1+2T)^{2\nu}} = 0, \quad \text{when } \nu > \tfrac{1}{2} \,. \tag{A9}$$

In the second term we find

$$B(T) < \int_{2T}^{\infty} \frac{2c^2}{t^{2\nu}}\, dt = \frac{2c^2}{2\nu - 1} \frac{1}{(2T)^{2\nu-1}} \tag{A10}$$

and, provided that $\nu > ½$, $B(T)$ tends to zero with large T. Thus if $\nu > ½$, both terms in (A7) tend to zero for large T and this means that Δ_T vanishes, and hence the discrepancy between the FT of R_X and the $\mathcal{F}[\Lambda_T R_X]$ also vanishes in the limit. In the context of bounded functions like $R_X(t)$ our class of functions in (A6) is much bigger than the one in Priestley's proof and includes the sinc function, for example.