

# Power (and Other) Spectra

## 1. Introduction

Looking at strainmeter or tiltmeter data “in the frequency domain” (that is, after taking some sort of Fourier transform) helps us to obtain more insight into the data. We look first at several kinds of Fourier transform that can be applied to the data, and discuss one (the power spectrum) in some detail, particularly how (and how not) to compute it.

### 1.1. Four Processes and Three Spectra

Suppose we have some strainmeter data, with drift, pressure-related fluctuations, tides, an earthquake, and electronics noise. The most important thing to realize is that the appropriate mathematical models for these are not the same, and that each model leads to a different kind of spectrum with (even) different units. The different classes are:

1. For the drift, or for that matter for the step associated with an earthquake, Fourier methods are strictly inapplicable: in the case of the drift because the signal is nonstationary, in the case of the step because the Fourier transform does not exist. We can fix both problems by taking the time-derivative of the data, which turns the drift (usually) into a stationary stochastic process, and the step into a transient. (If we choose to model the drift as a deterministic function, this will not be amenable to spectral analysis).
2. Periodic functions on  $(-\infty, \infty)$ , with period  $2T$ : these have a Fourier series expansion. The tides are a particular case of this (though the periodicity is not exact over an  $T$  for all tides). We can write

$$X(t) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} [a_n \cos(\pi n t / T) + b_n \sin(\pi n t / T)]$$

The total energy over the time  $(-T, T)$  is

$$\int_{-T}^T X^2(t) dt = 2T \left( \sum_{n=0}^{\infty} c_n^2 \right)$$

with  $c_n^2 = a_n^2 + b_n^2$ ; it thus grows without bound as  $T$  increases, but the amplitude of the cosine is constant. Such functions are thus best described using an **amplitude spectrum**, with  $c_n^2$  being the contribution to the total power from the term in the Fourier series of  $X(t)$  with frequency  $n/2T$ . The units for such a spectrum are the same as for the data.

3. Next we have non-periodic functions that are **transient**: they satisfy  $\int_{-\infty}^{\infty} X^2(t) dt < \infty$ . For these, the Fourier transforms  $\tilde{X}(f) = \int_{-\infty}^{\infty} X(t) e^{-2\pi i f t} dt$ , exists: the bound on the integral of  $X^2$  means that the total energy (as we call the integral of the square of the function) is finite, thus the average of energy divided by time (the power) is zero (for infinite time). We describe the such a

function in terms of an *energy spectrum* over frequencies. Parseval's theorem gives us that

$$\int_{-\infty}^{\infty} X(t)^2 dt = \int_{-\infty}^{\infty} |\tilde{X}(f)|^2 df$$

We can think of  $|\tilde{X}(f)|^2 df$  as the contribution to the energy from those components in  $X(t)$  with frequencies between  $f$  and  $f + df$ . This kind of spectrum is appropriate for any bounded transient signal, such as a seismic wave. The units for this would be, if the data is in m,  $\text{m}^2/\text{Hz}^2$  (assuming we take the second to be our unit of time).

4. Finally, we come to the class of processes (such as pressure-related changes, or electronic noise) that are best modelled as stationary stochastic processes. For these, we run into a complication if we attempt Fourier analysis. We cannot represent such a process as a Fourier series (because it is not periodic), nor as a Fourier integral (i.e., a Fourier transform), since there is no reason to expect the total energy in the process to be finite (in fact the stationarity requirement strongly suggests it will not be).

## 1.2. Defining the Power Spectrum

To describe such a process, statistical measures *must* be used: we can define the mean, variance, and higher moments of a random variable, but not its amplitude. The most intuitively appealing approach to the power spectrum is to imagine putting the signal into a narrowband filter that passes only frequencies between  $f_0 - \Delta f/2$  and  $f_0 + \Delta f/2$ . The output of such a filter will be another stochastic process with a variance  $\sigma^2$ . The power spectrum  $P(f_0)$  may then be calculated using

$$P(f_0) = \sigma^2/\Delta f$$

When  $P$  varies with frequency, the equivalent expression is

$$\sigma^2 = \int_{f_0 - \Delta f/2}^{f_0 + \Delta f/2} P(f) df$$

To proceed more formally, we first define a new function  $X_T(t)$ , which is a truncated version of the realization of a stochastic process

$$X_T(t) = \begin{cases} X(t) & -T \leq t \leq T \\ 0 & \text{otherwise} \end{cases}$$

Because this is transient, we may take its Fourier transform:

$$\tilde{X}_T(f) = \int_{-\infty}^{\infty} X_T(t) e^{-2\pi i f t} dt = \int_{-T}^T X_T(t) e^{-2\pi i f t} dt$$

Now we can find an energy spectrum for  $X_T$ , namely  $|\tilde{X}_T(f)|^2$ . This will be infinite if we let  $|T| \rightarrow \infty$ , so we use a power density distribution instead, normalizing by  $T^{-1}$  as we did for the Fourier series case. This gives

$$\lim_{T \rightarrow \infty} \frac{1}{2T} |\tilde{X}_T(f)|^2 df$$

which is the contribution to the total power in the frequency range  $(f, f + df)$ . The true power density function for the stochastic process must involve more than just one realization. Thus we average over the different realizations, and define the **two-sided power spectral density function** of  $X(t)$  by

$$h(f) = \lim_{T \rightarrow \infty} \left[ E \left[ \frac{|\tilde{X}_T(f)|^2}{2T} \right] \right] = \lim_{T \rightarrow \infty} \frac{1}{2T} E \left[ \left| \int_{-T}^T X(t) e^{-2\pi i f t} dt \right|^2 \right] \quad (1)$$

For this definition,  $h(f)df$  is the average of the contribution to the total power from components in  $X(t)$  with frequencies between  $f$  and  $f + df$ . This is defined for all  $f$ . If  $X$  is real, the Fourier transform (and hence  $h(f)$ ) is symmetric in  $f$ . It is then tempting, and common, to use positive frequency only, and add the symmetric parts together to get a **one-sided** power spectral density, which is just twice (1), for  $f \geq 0$ .

By equation (1), the units of this (for the same assumptions as before) would be  $\text{m}^2/\text{Hz}$  (in some engineering usage the square root is taken, to give units of (say)  $\text{m}/\text{Hz}^{\frac{1}{2}}$ ).

The averaging over realizations introduces a statistical aspect to the spectral representation. Take  $X(t)$  to be a **zero-mean** stationary stochastic process. There is then an alternative definition for the power spectral density function  $h(f)$ , in terms of the autocovariance function  $R(\tau)$

$$h(f) = \int_{-\infty}^{\infty} e^{-2\pi i f \tau} R(\tau) d\tau \quad (2)$$

i.e. ,  $h(f)$  is the Fourier transform of  $R(\tau)$ . We have from (2) the inverse transform

$$R(\tau) = \int_{-\infty}^{\infty} h(f) e^{2\pi i f \tau} df$$

from which we get

$$R(0) = E[X(t)X(t)] = \text{Var}[X] = \int_{-\infty}^{\infty} h(f) df$$

which means that the total power integrated over all frequencies is equal to the variance of  $X$  in the time domain. (This is also true for the one-sided spectrum, provided the domain of integration is  $f \geq 0$ ).

The simplest possible case is a white noise process, for which

$$R_X(\tau) = \sigma_X^2 \delta(\tau)$$

Thus  $h(f) = \sigma_X^2$ : the power spectral density is constant, with the variance equally distributed over all frequencies.

### 1.3. Filters and their Effect on $h(f)$

Suppose we are given a stationary stochastic process  $Y(t)$  with variance  $\sigma_Y^2$  and autocorrelation  $\rho_Y(\tau)$ . We pass the process through a linear time invariant system (a filter) to produce  $X(t)$ ; i.e., we have

$$X(t) = \int_{-\infty}^{\infty} g(u)Y(t-u)du$$

We will assume that  $\int_{-\infty}^{\infty} |g(u)| du < \infty$  i.e., the Fourier transform of the filter exists.

We now compute the effect  $g$  will have on  $h_Y(f)$ , to start with by finding the autocovariance of  $X$ :

$$\begin{aligned} R_X(\tau) &= \text{Cov}[X(t), X(t+\tau)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u)g(v)\text{Cov}[Y(t-u), Y(t+\tau-u)]dudv \\ &= \sigma_Y^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u)g(v)\rho_Y(\tau+u-v)dudv \end{aligned}$$

The spectral density function is the Fourier transform of  $R_X(\tau)$

$$\begin{aligned} h_X(f) &= \int_{-\infty}^{\infty} R_X(\tau)e^{-2\pi if\tau} d\tau = \sigma_Y^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u)g(v)\rho_Y(\tau+u-v)e^{-2\pi if\tau} d\tau dudv \\ &= \sigma_Y^2 \left\{ \int_{-\infty}^{\infty} g(u)e^{2\pi ifu} du \right\} \left\{ \int_{-\infty}^{\infty} g(v)e^{-2\pi ifv} dv \right\} \left\{ \int_{-\infty}^{\infty} \rho_Y(\tau+u-v)e^{-2\pi if(\tau+u-v)} d\tau \right\} \quad (4) \end{aligned}$$

Now we write  $\tau' = \tau + u - v$  and define the Fourier transforms

$$G(f) = \int_{-\infty}^{\infty} g(u)e^{-2\pi ifu} du \quad p_Y(f) = \int_{-\infty}^{\infty} \rho_Y(\tau')e^{-2\pi if\tau'} d\tau'$$

Substituting these into (4) we get

$$h_X(f) = \sigma_X^2 p_X(f) = \sigma_Y^2 p_Y(f) |G(f)|^2 = h_Y(f) |G(f)|^2$$

$G(f)$  is called the frequency response of the linear system. We see that the spectral density function of the output is thus just the spectral density function of the input times the squared modulus of the frequency response.

The important feature of this result is that for a time invariant linear system the value of the output spectral density function at frequency  $f$  depends purely on  $|G(f)|^2$  and the value of the input spectral density function at the same frequency  $f$ . The frequency domain representation has “disentangled” the convolution required in the time domain so that we can study the properties of the system separately at each frequency.

#### 1.4. A Note on Spectrum Units

As noted above, the correct units for a power spectrum are variance (amplitude squared) divided by frequency, so that the units for tilt spectra are  $\text{rad}^2 \text{Hz}^{-1}$ . The square root of this is often used in engineering; such spectra can be converted to conventional power spectra by squaring. Another engineering practice is to use decibels (dB) as a convenient way of handling quantities which vary by such large factors that logarithmic scales are appropriate (and rare is the spectrum for which this is not true). If two quantities have variances  $\sigma_1^2$  and  $\sigma_2^2$ , the ratio between them is  $10 \log_{10}(\sigma_1^2/\sigma_2^2)$  dB; for amplitudes  $a_1$  and  $a_2$  the expression is therefore  $20 \log_{10}(a_1/a_2)$  dB. To be useful, spectra must be plotted on an absolute scale, in which 0 dB ( $P = 1$ ) is a known physical value; for example, the tilt noise spectra in Figure 2 are in decibels relative to a power level of  $1 \text{ rad}^2 \text{Hz}^{-1}$ ; this reference level, or the strain equivalent  $1 \text{ } \epsilon^2 \text{Hz}^{-1}$ , should be used.

#### 1.5. Comparing Different Spectral Results

As an example of the use of power spectra, consider the statement made above that the earth tides dominate any low-passed strain record. For frequencies less than  $10^{-2}$  Hz, the noise spectrum is roughly  $5 \times 10^{-24} / f^2 \text{ } \epsilon^2 \text{Hz}^{-1}$  for  $f$  in hertz. When looking at a record for tides the eye classifies everything at much lower frequencies as ‘drift’; we may take  $2 \times 10^{-6}$  Hz (6 days) as our lower limit and  $10^{-2}$  Hz as our upper limit. Application of (8) then gives a standard deviation, or rms amplitude, of  $2 \times 10^{-9}$ , only 10% of the average tidal amplitude at most latitudes. (The peak-to-peak amplitude of a random process is about five times the rms value.) Tides are not stochastic processes; as a consequence they have well-defined amplitudes but not a power spectral level. We may, however, use an approximate rule for finding one: the rms amplitude of a sine wave of amplitude  $a$  is  $\frac{1}{2} a^2$ , and for a record length  $\tau$  the smallest resolvable frequency is  $\tau^{-1}$ , making the apparent spectral level  $\frac{1}{2} a^2 \tau$ . At tidal periods the strain noise level is about  $10^{-14} \text{ } \epsilon^2 \text{Hz}^{-1}$ , and so in a 1-year record a sine wave of amplitude  $1.5 \times 10^{-10}$  would have a 20 dB signal-to-noise ratio.

If  $y(t)$  is a transient, rather than a sinusoid, the signal-to-noise ratio is the square root of

$$\int_0^{\infty} \frac{|S(f)|^2}{P(f)} df$$

where  $S(f)$  is the amplitude spectrum of the signal. The value of this ratio determines whether or not a transient signal can be detected against background noise. If we use the same units for both signal and noise, and plot them in decibels (effectively squaring the amplitude), then the units are directly comparable.

## 2. Spectrum Estimation

We now turn to the problem of actually estimating the power spectral density of a stochastic process. Note that this is, in a fundamental sense, a difficult problem, since we are not trying to estimate, or set bounds on, a finite set of parameters; rather, we are trying to estimate a function  $h(f)$ , which cannot in general be

described by such a finite set of numbers. We will, given finite data, inevitably have to settle for incomplete resolution of the details of this function.

We assume that the process is discrete, with sample interval  $\Delta t = 1$  for simplicity. We also assume that (in principle at least) the process is infinite in length, though in practice we will have only  $N$  terms. We also assume that there are no purely periodic components (i.e., lines in the spectrum), that the spectral density  $h(f)$  is continuous, and that the process is ergodic. (Averaging in time is the same as averaging over realizations).

## 2.1. Parametric Spectral Estimation

We first touch on a very popular method of spectral estimation, with a very large literature. We describe it briefly because we think it is not a good way to estimate the spectrum for geophysical cases. Basically, it consists of modelling the spectrum by a relatively small number of parameters, determined in the time domain. This would be an excellent method of spectrum estimation if we knew that the process we were examining in fact had a spectrum of the assumed form—but we almost never do know this, and if it is not true, such spectral fitting will give answers that may be badly biased, and which at best must be regarded as determined as much by our assumptions as by the data. In particular, it becomes very difficult to make realistic estimates of the uncertainty of the estimated spectrum.

The basic approach derives from earlier results on the effects of filters on the spectrum: we showed that if we had a continuous process  $X$  that could be regarded as a filtered version of another process  $Y$

$$X = g * Y$$

then their spectral densities would be related by

$$h_X(f) = |G(f)|^2 h_Y(f)$$

Suppose now that  $h_Y(f) = \sigma^2$ , i.e.,  $Y$  is a continuous white noise process. Then  $|G(f)|^2$  would give us the spectrum for  $X$ . Thus one way to find the spectrum for  $X$  is to find a convolution  $\gamma$  that transforms  $X$  into white noise; i.e.,

$$\gamma * X = g^{-1} * X \approx \text{white noise}$$

Then  $\Gamma(f) = \mathbf{F}[\gamma]$  and the spectrum would be

$$h_X(f) = \left| \frac{1}{\Gamma(f)} \right|^2$$

This idea is usually implemented by assuming that  $\gamma$  can be represented by a finite-length autoregressive (AR) or moving average (MA) filter. In the AR case this is equivalent to modeling the process in the time domain by

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k} = \varepsilon_t \quad \text{for all } t$$

with  $\varepsilon_t$  independent and identically distributed (and usually Gaussian). The parameters  $a_i$ ,  $i = 1, \dots, k$  are estimated using maximum likelihood or least squares. Once we have found these, it is straightforward to compute the transfer function (using

the methods of Chapter 11) and thus find the spectrum. Note that if even we know the process, we still have the problem of deciding what order ( $k$ ) the filter must have been. A number of semi-automatic methods have been proposed for choosing  $k$ , based on repeated solution at various  $k$ 's. Maximum entropy, Yule-Walker estimation, Burg's method, and other methods, are all variations on this procedure.

There is a useful, if subsidiary, role for these techniques, which is to provide what is called **prewhitening**. As we will see, the methods of spectral estimation we prefer have the defect that they can be biased if the power spectral density covers a large range, with values at different frequencies differing by many orders of magnitude. This is not uncommon for actual data. The methods just mentioned can be used to devise a filter which, when used on the data, gives a series whose spectrum is much closer to being flat. We may then estimate the spectrum of this new series, with much less fear of bias, and find the spectrum of the actual data by dividing the estimated spectrum by the filter response.

## 2.2. The Raw Periodogram

Earlier we showed that our two definitions for the power spectral density function were equivalent

$$h_X(f) = \mathbf{E}[R_X(k)] = \lim_{N \rightarrow \infty} \frac{1}{N} E \left[ \left| \sum_{n=0}^{N-1} X_n e^{-2\pi i n f} \right|^2 \right]$$

The second definition, in terms of the expected value of the Fourier transform of the process, suggests an estimator

$$\hat{h}_X(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x_n e^{-2\pi i n f} \right|^2$$

This is known as the **periodogram** estimator. In this simple form it is not particularly good, but modifications of it form the basis of almost all efficient modern spectral methods. If we have white noise, the periodogram values  $\hat{h}_X(f)$  occur at  $N+1$  equally spaced frequencies  $f_m$ . The statistical results for the estimates are that they are independent, with mean and variance

$$E[\hat{h}_X(f_m)] = \frac{1}{N} N \frac{2N\sigma^2}{2} = \sigma^2 \quad \text{and} \quad \text{Var}[\hat{h}_X(f_m)] = \frac{1}{N^2} 4 \left( \frac{N\sigma^2}{2} \right)^2 = \sigma^4$$

The good news is that  $E[\hat{h}] = h$ : the periodogram estimate of the power spectral density function for a white process is unbiased. The bad news is twofold. First, this estimate has a very large variance; taking the standard deviation  $(\text{Var}[\hat{h}_X(m)])^{\frac{1}{2}}$  as measure of the uncertainty in our estimate for  $h_X(f)$ , we see that this is  $\sigma^2$ : the uncertainty of the estimate is the same as the estimate itself. And worse yet, this uncertainty is independent of  $N$ : as  $N \rightarrow \infty$ , the variance  $\text{Var}[\hat{h}]$  remains constant, so the periodogram is not a consistent estimator.

An intuitive way to see why this would be so is to note that, for  $N$  data, the periodogram gives us  $N/2$  spectral estimates. We might thus expect that each estimate has about the same amount of information, equivalent to two data, no matter how large (or small)  $N$  is. For many estimation problems, we have a fixed number of parameters and can expect them to be better estimated as we add more data (though this is not always true). Since here we are trying to estimate a function (in this case the power spectral density function) there is, in principle, no limit to the number of parameters needed—but the example of the periodogram shows us that unless we apply some limits, we cannot form a good estimate.

If we look at a nonwhite power spectral density, the expected value for the raw periodogram estimator is biased:  $\hat{h}_X(f_0)$  is the convolution of a function (called the Fejer kernel) with  $h_X(f)$  at surrounding frequencies. Unless the spectrum is completely flat there will be bias. If the true spectrum  $h_X(f)$  has a strong peak or a deep trough the results may be very deceptive: the low parts of the spectral estimate will be raised by energy actually present in the high parts. This phenomenon is known as **leakage**: power leaks into adjacent frequency bands obscuring details of the spectrum.

### 2.3. Improving the Periodogram

The raw periodogram is an extremely poor estimate of the power spectral density function. While it is asymptotically unbiased, it is not unbiased for nonwhite spectra and finite  $N$ , and may reach the asymptote only very slowly. It also has two other undesirable features:

- (1). It is not a consistent estimate of  $h(f)$  in the sense that  $\text{Var}[\hat{h}(f)]$  does not tend to zero as  $N \rightarrow \infty$ .
- (2). As a function of  $f$ ,  $\hat{h}(f)$  fluctuates wildly; to see that this is what to expect, remember that for any two fixed neighboring frequencies  $f_{1,f_2}$ ,  $\text{Cov}[\hat{h}_N(f_1), \hat{h}_N(f_2)]$  decreases as  $N$  increases.

We now discuss ways to improve it.

#### 2.3.1. Reducing Bias: Tapering

Since the bias is due to convolution of the true spectrum with the Fejer kernel, if we can find a way to replace the Fejer kernel with a function with smaller side-lobes then the leakage will be less and contamination from distant frequencies will not be such a problem.

One can guess (we will not prove) that multiplication of the original data series by an appropriate taper (also called a data window) might do the job. The Fejer kernel results from an implicit boxcar taper which truncates the data to a series  $N$  long. If we choose a taper which dies smoothly away to zero at the ends of the data series, then it will have smaller side lobes in the frequency domain. If the weights used in the taper are  $w_k$ , the expected value of the estimated spectrum becomes

$$E[\hat{h}_X(f)] = \int_{-\frac{1}{2}}^{\frac{1}{2}} W(f' - f) h_X(f') df'$$

where we have approximately

$$W(f) = |\mathbf{F}[w]|^2$$

The weights  $w_k$  are chosen so that  $W(f)$  falls off rapidly. As was the case in digital filter design, protection against bias from distant frequencies comes at the expense of broadening the central frequency lobe.

Another consequence of tapering is effectively to reduce the length of the data series. It should therefore not be surprising that the variance in the resulting estimate is always larger than for the untapered series (we discuss this further below). Our hope will be that the reduction in bias will result in a smaller overall mean square error. This will of course depend on the nature of the spectrum: if it is very nearly white the bias will be small to begin with, and perhaps not worth reducing—but in most cases this is not so.

### 2.3.2. Improving Consistency: Section Averaging

An obvious way to reduce the variance of the raw periodogram estimator is to obtain several independent periodogram estimates and then average them. There are a number of ways of doing this. One is to break the available time series into  $M$  parts of equal length, find the periodogram estimate on each one and then average the set together. This is consistent with the spirit of the definition of  $h_X(f)$  and the ergodic assumption

$$h_X(f) = \lim_{N \rightarrow \infty} \frac{1}{N} E\left[|\tilde{X}(f)|^2\right]$$

We use

$$\overline{\hat{h}_X(f)} = E[\hat{h}_X(f)] \approx \frac{1}{M} \sum_{j=1}^M \hat{h}_X^{(j)}(f)$$

For sufficiently large  $M$  we might guess that it is reasonable to invoke the Central Limit Theorem and say

$$\text{Var}[\overline{\hat{h}_X(f)}] \approx \frac{\text{Var}[\hat{h}_X(f)]}{M} \approx \frac{h_X(f)^2}{M}$$

Clearly by making  $M$  large enough we can reduce the variance in the spectral estimate to any level required.

We can also show that section averaging is the best way, in the maximum likelihood sense, of reducing variance in the spectral estimate.

However, this averaging has a price (to some extent unavoidable). The raw periodogram provided an estimate at each of  $N/2$  frequencies for  $h_X(f)$ . Section averaging reduces this by a factor of  $M$ . We thus reduce both the *resolution* of our spectral estimate, and also potentially increase the *bias* (because the Fejer kernel becomes broader).

### 2.3.3. Weighted Overlapped Section Averaging

Our suggested “simple” method for estimating the power spectral density combines the procedures of the previous two sections: we break the data into sections, taper each section, form the DFT of the result, and average over all sections. The tapering provides protection against bias; the averaging provides a lower-variance estimate than the raw periodogram. Because of the tapering applied to each section, it is appropriate to allow the sections to overlap. This method is known by the title given above, and also as Welch’s method, after its inventor.<sup>1</sup> We call this a “simple” method because it is not difficult to describe; while more information can be obtained using the multitaper method (to be described later), this method requires more background to explain and understand and more computation to estimate. While Welch’s method is now quite old, it remains very useful if there are large amounts of data relative to the level of detail in the spectrum that is needed.

For this method it is necessary to choose the data taper and the number of sections. There is no universally right answer for either of these; the choice depends largely on what the spectrum looks like, which makes the analysis an iterative process informed by personal judgement. If the dynamic range of the spectrum is large, then you need the good bias protection offered by, for example, the  $4\pi$  prolate data taper. If the dynamic range is lower, you might decide to retain better resolution near the central lobe in frequency of the data taper. The Hann taper is a common choice in this case. The choice of  $M$ , the number of data sections, depends on the kind of resolution/variance tradeoff you want to make. Too many sections will ultimately give poor frequency resolution, too few will give high variance in the spectral estimate.

We now discuss briefly the variance of the estimates from this method (following Welch). We take the estimator to be

$$\hat{h}(f) = \frac{1}{M} \sum_{j=1}^M \hat{h}_j(f)$$

Each estimate  $\hat{h}_j$  is a direct spectral estimate based on some subset of the data  $X_0, X_2, \dots, X_{N-1}$  and we assume that some appropriate data taper  $w_k$  has been applied to each section in order to reduce bias in the spectral estimate. Each section is displaced by  $D$  samples with respect to the origin of the previous one, and is  $L$  points long. If we do not have overlapping sections (i.e., if  $D = L$ ), we have  $M$  independent estimates of  $\hat{h}$ , each with 2 degrees of freedom: the variance in the spectral estimates is reduced by a factor of  $M$  by the section averaging process:

$$\text{Var } \hat{h}(f) \approx \frac{1}{M} \left[ E[\hat{h}(f)] \right]^2$$

However, usually we do overlap the tapered data sections. If each periodogram estimator is

<sup>1</sup> Welch, Peter D. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified, periodograms, *IEEE Trans. Audio Electroacoustics*, **AU-15**, 70-73.

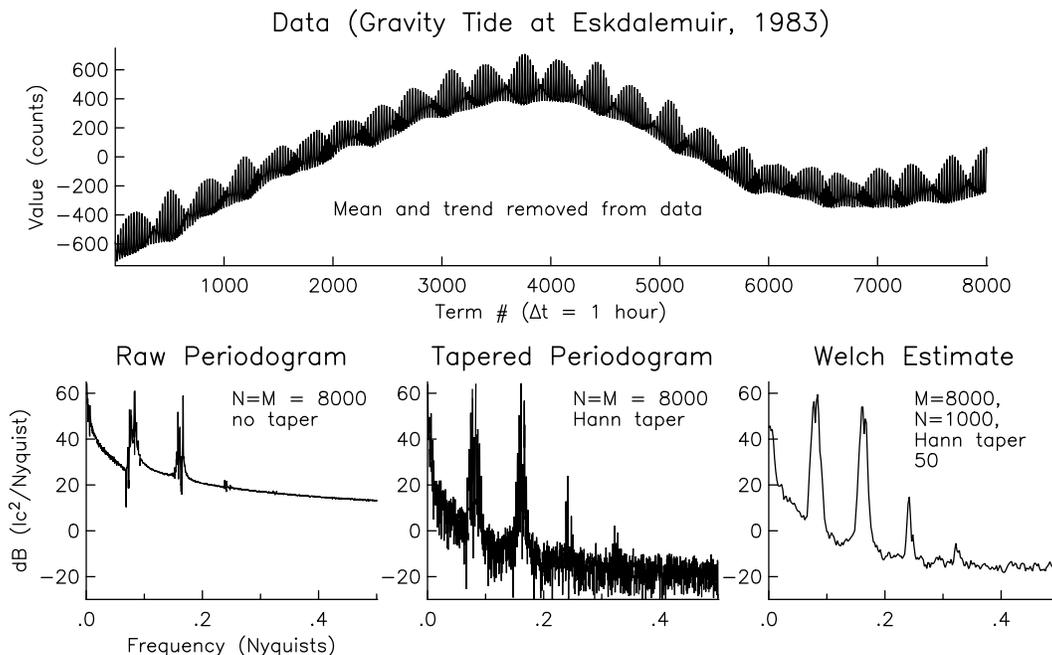


Figure 14-1. A data series (of gravity, from the Project IDA gravimeter at Eskdalemuir in Scotland), and three estimates of its power spectral density. The raw periodogram is so biased as to be useless. Tapering removes the bias but the estimates are so variable as to still not be useful. Welch's method gives a good estimate of the spectrum, at the price of losing resolution (shown by the loss of many separate peaks in the tidal bands).

$$\hat{h}_j = \frac{1}{L} \mathbf{E}[|w_n x_n|^2]$$

and the sections overlap, then even for a locally white spectrum there will be covariance in the spectral estimates: they can no longer be regarded as independent. Let

$$d_j = \text{Cov} \left[ \hat{h}_k(f_n) \hat{h}_{k+j}(f_n) \right]$$

and

$$\hat{h}(f_n) = \frac{1}{M} \sum_{i=1}^M \hat{h}_i(f_n) \quad \text{for } n = 0, 1, \dots, L/2$$

Then it can be shown that

$$\text{Var}[\hat{h}(f_n)] = \frac{1}{M} \left\{ d_0 + 2 \sum_{j=1}^{M-1} \frac{M-j}{M} d_j \right\}$$

Now if

$$\rho_j = \text{Correlation} \left\{ \hat{h}_k(f_n), h_{k+j}(f_n) \right\} = \frac{d_j}{d_0}$$

then this becomes

$$\begin{aligned} \text{Var}[\hat{h}(f_n)] &= \frac{d_0}{M} \left[ 1 + 2 \sum_{j=1}^{M-1} \frac{M-j}{M} \rho_j \right] \\ &= \frac{\text{Var}[h_k(f_n)]}{M} \left[ 1 + 2 \sum_{j=1}^{M-1} \frac{M-j}{M} \rho_j \right] \end{aligned}$$

If  $X(j)$  is a sample from a Gaussian process and  $h(f)$  is flat over the band sampled by our estimator, then we had

$$\text{Var}[\hat{h}_k(f)] = h^2(f_n)$$

It is also possible to show that

$$\rho_j = \left[ \sum_{k=0}^{L-1} w_k w_{k+jD} \right]^2 \left[ \sum_{k=0}^{L-1} w_k^2 \right]^{-2}$$

which is to say that the correlation depends only on the shape of the data taper for a locally white spectrum. Then,

$$\text{Var}[\hat{h}(f_n)] = \frac{\hat{h}^2(f_n)}{M} \left[ 1 + 2 \sum_{j=1}^{M-1} \rho_j \right]$$

and so when the spectrum is flat the variance is controlled by the shape and degree of overlap of the window. If the spectrum is locally complex, this estimate will be invalid. In the non-overlapping case  $\rho_j = 0$ . Therefore the *computationally* most efficient means of acquiring any desired variance is to have non-overlapping segments. It can be shown that for data windows shaped like  $1 - t^2$  or  $1 - |t|$  (data interval  $-1 \leq t \leq 1$ ) and 50% overlap the variance is

$$\text{Var}\hat{h}(f_n) \approx \frac{11}{9} \cdot \frac{h^2(f_n)}{M}$$

So the variance is inflated by the overlap if the number of segments remains the same, but an overall reduction in variance is achieved for fixed record length since overlapping segments increases the value of  $M$ . This expression is also approximately valid for a Hann taper, and it is this, with a 50% overlap, which would be our suggested first approach to estimating the power spectral density of any data series of reasonable length.

### 2.3.4. Prewhitening

We noted in Section 10.1 the advantage of “prewhitening” the data series; the reason for this may be clearer now that we have discussed bias in

the periodogram—the “whiter” the spectrum to be estimated, the less we need to be concerned about bias. In any case some kind of modification of the series is almost always needed, since most data series are non-stationary in some way, the simplest being a constant slope (non-stationary mean). The standard solution is to fit the “trend” (mean, best fitting line or other low-frequency component) using least squares fitting, remove this trend, and perform the spectrum estimation on the residual. Alternatively, the trend can be removed by first-differencing the series (or performing more elaborate prewhitening), computing the spectrum, and then correcting it for the frequency response. For first-differencing, this will always give an infinite value at zero frequency—but this cannot be estimated very well by these procedures anyway.

#### 2.4. Frequency Averaging

Section averaging of the periodogram estimate is not the only means of obtaining a consistent estimate of the power spectral density using the periodogram. Instead of averaging over time sections we could smooth the resulting periodogram in the frequency domain, choosing an interval around each frequency of interest and averaging the estimates for neighbouring frequencies together. This is also called **band averaging**. A more complicated approach, would be to weight these estimates in some way according to their distance from the nominal frequency of interest

$$\hat{h}_X(f_0) = \sum_{j=-M}^M w_j \hat{h}_X\left(f_0 - \frac{\Delta j}{M}\right)$$

The weight function  $w_j$  would be symmetric and chosen to provide a local average in frequency of the periodogram estimates. Not surprisingly, the smoothing results in a loss of resolution, and could introduce bias even if this has been kept small in the initial estimate (and we should be sure that it is).

#### 2.5. Autocovariance Windowing

The approach to spectral estimation through smoothing of the periodogram lead to a method of spectral estimation much used in the past: taking the Fourier transform of the estimated autocovariance function. This is a tempting course of action, given that we have defined the power spectral density in terms of the Fourier transform of the autocovariance: but as with the periodogram estimate, that some relationship holds does not always make it the basis for a good estimator. While this is still around, under the name of the Blackman-Tukey estimate, we know of no situation for which it is appropriate.