3

Spectral analysis

"If we dealt with problems involving the superposition of a few simple periodic phenomena, as do astronomers interested in binary stars and related problems, we can learn much from the periodogram. Sadly, however, almost no one else has this kind of data. As a result, the periodogram has been one of the most misleading devices I know."

John Tukey (1967)

3.1 INTRODUCTION

As mentioned in the Introduction, spectral analysis has been absolutely key in establishing the importance of astronomical forcing for climate. Yet spectral analysis is considered by many to be an arcane field to be approached only by experts. There is good reason for this: there is a long history of incorrect results published by good scientists who fell into spectral traps. Many people prefer to ignore the field, or at least not get involved. Yet that really isn't possible for scientists who want to form their own judgements about the relative contributions of proposed astronomical mechanisms.

In this chapter we will give an introduction to spectral analysis. We will attempt to describe the fundamental techniques, as well as alert the reader to traps and pitfalls. We will describe some commonly-used "tricks of the trade" such as zero padding, that make interpretation of spectra simpler and more intuitive.

We feel obliged, however, to offer a few words of caution. Although there are many ways to do an incorrect spectral analysis, there is no "correct way". The method chosen will depend on the taste and intuition of the analyst. In the literature on climate you will find proponents of the Blackman–Tukey method, of the Multitaper method, of the Maximum Entropy method, and even proponents of the unwindowed periodogram. You will find advocates of log plots, linear plots and square root plots. Be aware that all approaches to spectral analysis entail assumptions, and that the differences in these assumptions often account for the different results of the

analyses. If you know what those assumptions are, you can make a reasoned judgement of what conclusions you should draw.

In many ways, the spectral analysis of climate data is a specialised application. Many of the methods that are optimum for other kinds of data are not optimum for paleoclimate data. The traditional focus of statisticians are issues such as "leakage" (how does power at one frequency affect another), "bias" (is the estimate, on average, centred at the correct value), and "consistency" (does the precision of the estimate improve when more data are analysed). Yet these issues have been remarkably unimportant in the history of climate analysis! Far more important have been issues traditionally ignored (or at least, not emphasised) in the traditional methods. These include systematic errors, the uncertainty in the time scale, the coherence of the background across many different sites, and the need for high resolution—rather than accurate estimation of either the amplitude or the frequencies.

With a few exceptions, we will emphasise results, and not proofs. There are many textbooks that prove the theorems that we state; see, for example, those by Priestley (1981) and Percival and Walden (1993). It is more important for the paleoclimatologist to understand the limitations of the various methods rather than the derivations.

3.2 THE FOURIER TRANSFORM

Any function can be expanded in terms of polynomials. The resulting expression is known as a Taylor Series. But even if you understand the Taylor expansion, it still is not obvious (although it is true) that, instead of polynomials, you can use harmonic functions: sines and cosines. Basically, *any* function can be expressed as a sum of sine functions and cosine functions. The result is known as a Fourier series, named after the French mathematical physicist Joseph Fourier, who invented it to help solve heat flow problems.

When should you use polynomials, and when should you use harmonic functions? If the underlying forces are harmonic, then use of a harmonic series often results in the need for a much simpler and shorter expression. If upon doing the analysis you discover that a few terms dominate, then you can try to identify them with known forces that have the same periods.

Mathematicians write sums in two ways. When the sum consists of a countable number of pieces, they use the capital Greek letter Σ (which corresponds to the Roman S). This letter is now often called "the summation symbol". When the sum consists of a continuous number of terms, they use a script S, which has evolved into an integral sign $\int$. We will begin with the integral version but, since the data consist of discrete points, this will soon turn into a discrete sum, known as the digital Fourier transform.

3.2.1 Basic properties

Let the symbol $h(t)$ represent a continuous function of time, such as the insolation Then the ordinary Fourier transform, also called the Fourier amplitude, is a functio

of frequency f:

$$H_T(f) \equiv \int_{-\infty}^{\infty} h(t)e^{2\pi i f t}dt$$

The subscript on the function H_T is the letter "T" standing for "True". (We'll get rid of this in a moment, as soon as we show that we can never calculate H_T from real data, but only estimates of it.)

Everything that we are about to do can be done without the use of complex numbers. If we were to do so, we would get two Fourier transforms: a cosine transform and a sine transform. But these correspond the real and imaginary parts of the complex transform H. We will continue using complex numbers primarily because, typically, they make the equations half as long and more transparent. In the end, the spectral functions that we derive will all use real numbers.

Part of the charm of the Fourier transform comes from the fact that we can solve this equation explicitly for $h(t)$, yielding an equation that looks very similar to the previous one:

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

Thus, given data $h(t)$, we can find the Fourier transform $H(f)$, and given $H(f)$, we can find $h(t)$. The two contain exactly the same information, expressed in different ways.

The spectral power P is defined as the square of the Fourier amplitude:

$$P_T(f) = |H_T(f)|^2$$

Note that the spectral power does not tell us the phases of the sine waves. In an information theory sense, it contains exactly half of the information. Of course, the phases can be obtained directly from the Fourier amplitude.

But once we start talking about real data, rather than theoretical functions, we run into a problem. Real data do not span infinite time. Even worse, if h is a climate proxy, we might have it sampled only at a few (say 300) discrete points over a time range of a few hundred thousand years. Because of this limitation, we can never calculate the true Fourier transform H_T, and never get the true spectral power P_T. The best that we can do is to get an *estimate* of P_T. The entire theory of spectral analysis is devoted to methods to obtain such an estimate.

One important estimate of H_T can be made from discrete samples of $h(t)$. Suppose that all we have available to use is the value of $h(t)$ at times t_j. We define these values as the set h_j:

$$h_j \equiv h(t_j)$$

Then an estimate of the Fourier transform can be made by the equation:

$$H(f) \equiv \sum_{j=1}^{N} h_j e^{2\pi i f t_j}$$

Notice that we have turned the integral into a sum, and that H no longer has a subscript T. This H is not the true transform; it is only an "estimate" of the true H_T.

There are important theorems that can be proven for the special case that the data are equally spaced in time. If the interval between points is Δt, then define the "Nyquist frequency" as

$$f_N = \frac{1}{2\Delta t}$$

The Nyquist frequency is the highest frequency that will appear in the Fourier transform. To appreciate how important the Nyquist frequency is to your everyday life, consider the following fact. The engineers who designed music compact disks (CDs) chose the sampling period Δt to be 1/44000 sec, so that the Nyquist frequency would be 22,000 cycles/sec—above the hearing range of most people. (Some audiophiles still argue that this was not high enough, since although the human ear cannot hear individual tones at frequencies this high, inaudible signals at higher frequencies are *aliased* into the audible range. We discuss aliasing in Section 3.2.8.)

By convention, we let $f_0 = 0$ represent the average ("DC") level of the data. We will usually subtract the average value from the data before we do a spectral analysis, and this will automatically give us $H_0 = 0$. (If you ever forget to subtract the mean from the data, the thing you will notice is a very high value in the first bin of the spectral power, usually overwhelming the rest of the transform!) The lowest frequency of real interest is the one that gives one full cycle in the time interval T:

$$f_L = \frac{1}{T}$$

Now we have the highest frequency f_N and the lowest f_L. We still have to decide what other frequencies to evaluate. It turns out that a very convenient choice is to pick multiples of the low frequency f_L:

$$f_k = kf_L$$

To simplify our discussion, let's assume that N is even. If we allow k to run between 1 and $N/2$, this gives us $N/2$ frequencies going from f_L to f_N.

Why do we get only $N/2$ frequencies from N data points? The reason is that for each frequency, we obtain a complex number. The real part is the cosine coefficient, and the imaginary part is the sine coefficient. So we are obtaining $N/2 + N/2 = N$ amplitudes from N data points. Why are there at most only $N/2$ data points in the spectral power plot? The reason is that we are not plotting the phases, which account for another $N/2$ data points.

It turns out that there is some value in letting k run over a larger range, even though the frequencies that we calculate in doing so will not give us any useful information. We start with $k = 0$; that includes $f = 0$ in the calculation. If we let k run from 0 to $N - 1$, we get the following nicely symmetric pair of equations:

$$H_k \equiv \sum_{j=0}^{N-1} h_j e^{2\pi ijk f_L t_j}$$

$$h_j = \sum_{k=0}^{N-1} H_k e^{-2\pi ijk f_L t_j}$$

This set of equations for discrete data is usually called the "discrete Fourier transform". H is the complex Fourier transform of h.

These equations have another important feature. If we knew, somehow, that $h(t)$ were "band-limited" (it contains no frequencies below f_L or above f_N), then we can prove that the discrete transform above will yield the exact true function H_T! This means that there is a relationship between the continuous function $h(t)$ and the discrete values H_k. It is given by:

$$h(t) = \sum_{k=0}^{N-1} H_k e^{-2\pi k f_L t}$$

The only difference between this and the previous equations is that the function $h(t)$ here is continuous, and the subscript k has been dropped from the t in the sum. This result means that the function $h(t)$ is completely determined by its samples h_k. The band-limited condition is often met for many functions of paleoclimate interest, such as eccentricity and insolation. The condition is often not met for geologic data. Nevertheless, this equation means that the discrete Fourier transform can often serve as a excellent method for estimating the true spectrum.

3.2.2 Glacial Fourier transform

We will now discuss the numerical calculation of the Fourier transform. We begin with a method we call the "Glacial Fourier Transform". This method is extremely easy, but almost never used, and yet it is probably the best method to start with for a beginner. Most experts will tell you to use the "Fast Fourier Transform" or FFT. Don't listen to them! We think the FFT, with the burden of jargon that you must learn before you can even start to use it, serves as a barrier to keep beginners away from playing with and understanding the Fourier transform.

The equation for the discrete Fourier transform (from the previous section) can be broken into its real and imaginary parts:

$$\begin{aligned} H(f) &= \sum_{j=1}^{N} h_j(\cos(2\pi f t_j) + i\sin(2\pi f t_j)) \\ &= \sum_{j=1}^{N} h_j \cos(2\pi f t_j) + i\sum_{j=1}^{N} h_j \sin(2\pi f t_j) \\ &\equiv H_R(f) + iH_I(f) \end{aligned}$$

(Note that although the quantity $iH_I(f)$ is imaginary, the function $H_I(f)$ is real.) For paleoclimate work, these real and imaginary components of H can be evaluated from this equation on a small computer. Once the data h(k), the ages t(k), the desired frequency f, and the number of data points N have been read into the computer, the code in Basic calculates the real and imaginary components (Hr and Hi) of the

Fourier amplitude and then sums and squares them:

```
Hr = 0
Hi = 0
for j = 1 to N
        Hr = Hr + h(j) * cos(6.2831853 * f * t(j))
        Hi = Hi + h(j) * sin(6.2831853 * f * t(j))
next j
P = Hr^2 + Hr^2
```

As we mentioned earlier, you will almost never see code like this, since computer programmers consider it inefficient. But what is inefficient for the computer may be simple and clear for the scientist. The method will not be mentioned in texts on numerical methods. But that is silly. On a modern computer, the computation is so fast, that it is more important to write easily understood code than to write fast code. As an example, we applied the Glacial Fourier transform to calculate 250 frequencies for 500 data points, using the computer language Matlab™ on a laptop computer. The entire computation took 0.12 seconds.

This is a good time to look at the equation, and get a sense of what the Fourier transform does. It takes your data, multiplies it by a sine wave, and then sums the results. If the data oscillate in phase with the sine wave, so that they are positive together and negative together, then all the terms in the sum are positive and the Fourier amplitude is large. If they drift into phase and then out of phase, then half of the values in the product will be positive and half will be negative, and the sum will be close to zero. You can also see that the sum is not particularly sensitive to sharp changes in the data (e.g. sudden terminations); it is more sensitive to the bulk behaviour of the data, e.g. are most of the data points positive when the sine wave is positive?

You can calculate the phase from

$$\phi = \tan^{-1}\left(\frac{H_I}{H_R}\right)$$

Unfortunately, the inverse tangent function in many computer languages, called *atan*, returns a value limited to the range $-\pi/2 < \phi < \pi/2$, which only covers half of the angular range. This problem can be avoided in Fortran, Matlab, and in some Basic interpreters, by using the alternate function atan2:

$$\phi = \text{atan2}(H_I, H_R)$$

We like the program Matlab™ for our computations of the Fourier transform.

Matlab represents the list of values $h_1, h_2, h_3, \ldots, h_n$ by a single symbol h. The Matlab code for the Glacial Fourier transform becomes very simple:

```
Hr = sum(h. * cos(2 * pi * f * t));
Hi = sum(h. * sin(2 * pi * f * t));
```

If you think of the word "sum" in these equations as integral signs, then the code looks just like the equations defining the Fourier transform.

In paleoclimate studies, the only time you need to use the faster FFT is for Monte Carlo studies, when it is necessary to calculate thousands of spectra. Also, if you are using someone else's spectral analysis program, it probably has an FFT built in. To learn to use the FFT, of course, you must learn a few more things:

- how to interpolate your data to uniformly spaced time intervals;
- how to pad the data with zeros, so the number of data points will be an exact power of 2;
- how to interpret the mysterious order of the output vector, which will have every value appearing twice;
- how to calculate the frequencies that go along with the FFT, by using the Nyquist formula;
- how to interpolate between the FFT frequencies to get the value at the frequencies that you really want (e.g. at the known astronomical frequencies).

We give Matlab routines in Appendix 2 that do all these steps.

3.2.3 The uncertainty principle

The Fourier transform has an elegant theorem associated with it called the "uncertainty principle". We are going to say more about it than anyone in paleoclimate needs to know, just because we find its connection to quantum mechanics very interesting. If you don't, then skip this section. All we will need is the final equation.

The uncertainty theorem of Fourier analysis is directly related to the uncertainty principle of quantum mechanics. The uncertainty principle is most simply stated in the following way. Suppose the function $h(t)$ is significantly different from zero in only a limited region near $t = t_0$. Let the width of this region be Δt. Then the spectral power will also have a significant value only over a limited range of frequencies Δf. Moreover, there is a simple relationship between these widths:

$$\Delta t \, \Delta f \geq \frac{1}{4\pi}$$

This inequality can be proven mathematically if we define Δt to be the root-mean-square (RMS) width of $h(t)$, and Δf to the RMS width of $H(f)$. The equality is true only if $h(t)$ happens to have a particular shape known as a Gaussian. The relation is the basis of the famous Heisenberg Uncertainty Principle of Physics. If we multiply both sides of the equation by Planck's constant h, and then use the

quantum-mechanics relationship that the energy $E = hf$, we get

$$\Delta t \, \Delta E \geq \frac{h}{4\pi}$$

We can derive a similar physics equation for momentum. If our quantum mechanical wave is a function of position x, the wavefunction can be expanded in terms of spatial frequencies f. We define the wavenumber k by $k = 2\pi f$. The uncertainty equation now reads

$$\Delta x \, \Delta k \geq \frac{1}{2}$$

Multiplying by $h/2\pi$, and substituting the quantum mechanics definition of momentum $p = hk/2\pi$, we get the equation that Heisenberg is most famous for:

$$\Delta x \, \Delta p \geq \frac{h}{4\pi}$$

But let us now return to the mathematical version of the equation. Suppose that we have a record that lasts for a time T. What is Δt? It is the root-mean-square deviation from the average time. What is that? We can make a quick guess from the fact that all the points are less than $T/2$ from the average time, so Δt must be less than $T/2$. If you calculate the RMS for a flat distribution, it is not hard to show that the correct value is $\Delta t = T/\sqrt{12} \approx 0.29T$. So according to the Uncertainty Principle,

$$\Delta t \, \Delta f = \frac{T \Delta f}{\sqrt{12}} \geq \frac{1}{4\pi}$$

$$\Delta f \geq \frac{\sqrt{3}}{2\pi} \frac{1}{T} \approx \frac{0.28}{T}$$

This, of course, is still an inequality. To get an equality, we can perform the calculation explicitly by performing a Fourier transform of a pulse of duration T. Moreover, when we do this we can conveniently pick Δf to be the full-width at half-maximum (*FWHM*) of the spectral power, rather than the RMS deviation from the average of the Fourier amplitude. From this calculation, we get the result:

$$\Delta f = \frac{0.89}{T}$$

(The numerical constant 0.89 is actually $1.3916 \times 2/\pi$, and 1.3916 is the solution to the transcendental problem $(\sin x/x)^2 = 1/2$.) This equation is often simplified even further to read:

$$\Delta f \approx \frac{1}{T}$$

and this is the way you will frequently see it stated in the literature, sometimes with the "$\approx$" symbol replaced with a "$=$" symbol. In this explanation, T was the duration of a pulse. The same equation results if instead T is the duration of data containing a

sine wave. (To show this, we can represent the data as a product of an infinite sine wave and a pulse, and then use a result known as the "convolution theorem".)

Applied to paleoclimate data, this equation gives a relationship between the width of the peak in the spectral power, and the duration of the signal. So, for example, if you do an analysis of data that has a duration of 600 kyr, and the data contain a perfectly periodic wave that stays at constant frequency for that period, then the *FWHM* of the spectral peak will be $0.89/600 = 0.0015$ cycle/kyr. If the record is longer, then the width will be correspondingly narrower.

But you can do much worse. If windowing is applied to the data (see the sections to follow) then the resolution is degraded. For example (see Priestley, 1981, Sections 7.3–7.4), the Parzen window degrades the resolution by a factor of $6/\pi = 1.91$. And you can also fool yourself into thinking that you can do better, e.g. with Maximum Entropy Analysis (see Section 3.8).

3.2.4 Periodogram

The "periodogram" was invented by Arthur Schuster in 1898 for the purposes of studying climate. It was the first form of numerical spectral analysis and, although experts frequently prefer other methods, it is still the most widely used. As conceived by Schuster, the periodogram is the absolute value squared of the Fourier transform, applied to evenly-spaced data. Even though the data exist only at discrete times, the periodogram could be evaluated and plotted at any frequency—provided that you had the patience and time to do the computation. Thus the periodogram is the technical term for the method of estimating spectral power that we have been describing.

Many competitors to the periodogram have been invented subsequently. As we said in the Introduction, there is no one "correct" method for spectral estimation. One important advantage of the periodogram is that, unlike many other popular methods (e.g. maximum entropy analysis and multitaper analysis), its statistical properties are well understood. For example, if we represent our model for the data as a sum of sine waves plus error, then Whittle (1952) showed that the periodogram gives the best maximum likelihood estimate of those frequencies. If you are really interested in these formal aspects of spectral analysis, you will find them described in a readable review by MacDonald (1989).

It has become popular in statistical circles to object to the use of a simple periodogram. For example, M.B. Priestley (1981) states that

> *"the periodogram is an extremely poor (if not a useless) estimate of the spectral density function." (p. 420)*

In the otherwise insightful book, *Numerical Recipes*, W. Press et al. state (1993):

> *"... the variance of the periodogram estimate at a frequency f_k is always equal to the square of its expectation value at that frequency. In other words, the standard deviation is always 100 percent of the value, independent of N!"* (p. 422)

But these statements are highly misleading for the student of paleoclimate, since they are meant to apply to continuous spectra or resolved spectral peaks. (A resolved peak is one which is broader than the Δf resolution limit of the Uncertainty Principle.) If we have an unresolved spectral line, as is often the case in astronomy or astronomically-driven phenomena, then the periodogram is often an excellent method. John Tukey, one of the founders of modern spectral analysis, summarised this conclusion (Tukey, 1967):

> *"If we dealt with problems involving the superposition of a few simple periodic phenomena, as do astronomers interested in binary stars and related problems, we can learn much from the periodogram. Sadly, however, almost no one else has this kind of data. As a result, the periodogram has been one of the most misleading devices I know."*

Paleoclimate is precisely one of those "related problems" that is dominated by a few simple periodic phenomena. In this chapter we will show the application of many spectral methods to paleoclimate data, and it will often turn out that the periodogram, sometimes with a simple modification called a taper, is the most useful method to apply. Particularly for the study of the 100-kyr cycle, the signal to noise is huge (often 10:1 or greater). There is negligible leakage from adjacent peaks, and bias from a sloped background is often unimportant. For the study of such phenomena, there is little value to be obtained in reducing the resolution by the use of other methods, and there is much to be gained by the high resolution of the periodogram.

The periodogram has been maligned because of a curious feature: signal to noise of a continuous spectrum cannot be improved by taking a longer set of data. The reason for this is simple: in the periodogram, the bin size is proportional to $1/T$, and thus for long records, the frequency resolution improves proportionately—so that the power in one bin will not increase. If the signal is continuous or the spectral feature resolved, then the signal to noise level does not change. Other techniques have been developed for this purpose, including the "Blackman–Tukey method", the windowed periodogram, and the multi-taper method. These methods also have other advantages as well; they are devised to reduce systematic effects known as leakage and bias.

But you cannot gain something without losing something, and each of these methods gains its signal-to-noise advantage by throwing away frequency resolution. The loss of resolution can be extremely misleading if the key difference between models lies in the shape of the spectral peak. In the periodogram, the information that you gain about the spectrum does increase with the length of the spectrum since you measure the frequency with greater accuracy. But, historically, statisticians have been more concerned with power estimation than with frequency estimation, hence their dislike of the method. In paleoclimate, the situation is exactly the opposite. We are more concerned with the detection of narrow spectral peaks than in a precise estimate of their power levels. This makes the periodogram, in many cases, ideal.

For the unresolved spectral feature, the periodogram can be superior to many of the other standard techniques, because it puts the power into the fewest number of

bins. It can be shown that, if the power at a frequency f is normalised with respect to the local noise (i.e. the local noise has value 1), then the standard deviation σ_P of the power P is given by:

$$\sigma_P = \sqrt{2P - 1}$$

This implies that the fractional uncertainty of the power is

$$\frac{\sigma_P}{P} = \frac{\sqrt{2P - 1}}{P} \approx \sqrt{\frac{2}{P}}$$

Since the spectral power of an unresolved peak increases with time, the uncertainty in the estimate of the power decreases rapidly, as the square-root of time (in contradiction to the overgeneralised statement of Press et al. given above). For a pure sine wave, the power increases proportionally to time, whereas the noise (local background) remains constant, so the signal-to-background estimate improves as the square-root of time.

In paleoclimate work, the favoured approach has been the Blackman–Tukey method which we will describe in Section 3.3. In fact, the periodogram can be considered a special case of the Blackman–Tukey method, applied with the Blackman–Tukey lag parameter set to unity.

3.2.5 Sharp feature in the time domain

In paleoclimate data, there are often features in the time domain that are abrupt, such as the "sudden terminations" of the glacials. It is important to note what effect such a feature has on the spectrum. Mathematically we call a brief spike a "delta function"; an abrupt change in level (e.g. an abrupt termination) is a "step function".

The Fourier transform of a delta function is particularly easy to calculate, since it is zero except at one time:

$$H_T(f) \equiv \int_{-\infty}^{\infty} \delta(t - t_0) e^{2\pi i f t} dt = e^{2\pi i f t_0}$$

$$P_T(f) = |H_T|^2 = 1$$

This shows that the spectral power of a delta function is constant, i.e. it contributes equally at all frequencies. For a digital Fourier transform, the result is approximately the same. We recommend that the reader, particularly the reader new to digital transforms, try this digitally. Represent a delta-function by setting the data equal to zero except at one point. Then do a Fourier transform. The result is shown in Fig. 3.1 below for the age range 0–600 kyr, with one point at $t = 300$ kyr set equal to 1. The function is not quite constant, although most of the points have value near 1. The function goes to zero at $f = 0$ because we subtracted the average from the data points before performing the transform.

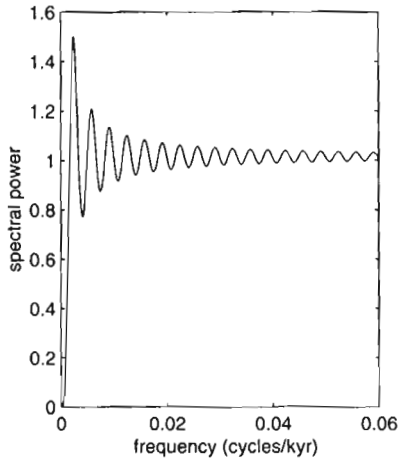


Fig. 3.1. Spectrum of delta function.

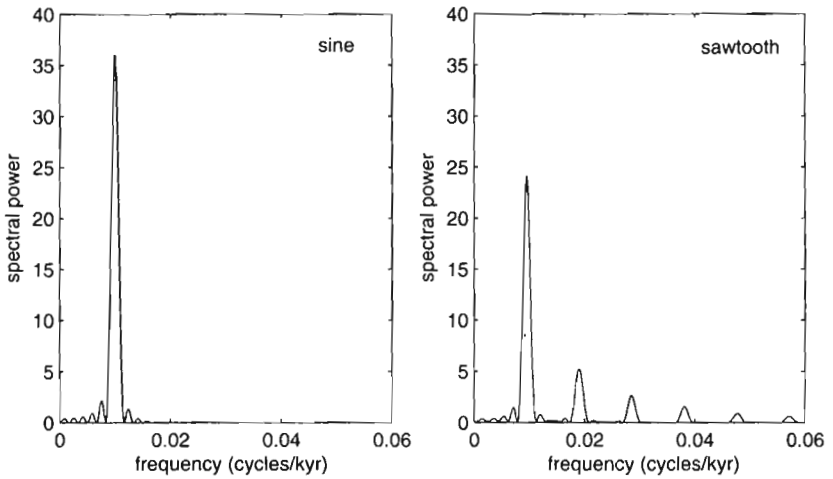


Fig. 3.2. Spectra of sine and sawtooth.

It is very interesting to see the effect that sudden terminations have on data. In Fig. 3.2 we show the spectral power of a pure sine wave and that of a 100 kyr sawtooth (i.e. triangular) wave:

You'll note that the primary difference between the two are the reduced power in the fundamental peak (at 0.01) and the presence of the harmonics at multiples of the fundamental frequency. The area under both curves is the same, by our normalisation convention (mean power = 1). The harmonic peaks are so small that they can be easily lost in, or cancelled by, background (if it were present, as it would be in real data). The net result is that the properties, or even the existence of the abrupt terminations, can be hidden or lost in the frequency domain. They are best studied in the time domain.

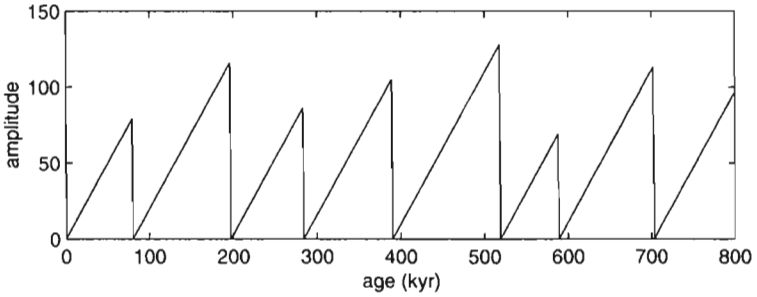


Fig. 3.3. Randomised sawtooth.

Another possibility, however, is that the glacial cycles are astronomically driven, but that the sudden terminations have a random component. The first day of snow rarely occurs on the first day of winter. The ice does not usually crack on the first day of spring. We have simulated this by generating a sawtooth in which the abrupt transition occurs every $100\,\text{kyr} \pm \sigma$ where we set σ equal to a random number between $\pm 20\,\text{kyr}$. The data generated in this way are shown in Fig. 3.3.

The spectrum for these data is shown below (Fig. 3.4). The 100 kyr peak is still strong, but the secondary peaks no longer show a simple regular behaviour. The harmonics, multiples of the fundamental 0.01 frequency (and have periods of 50 kyr and less) are gone—destroyed by the $\pm 20\,\text{kyr}$ random terminations! The spectral power of these harmonics has been spread to nearby frequencies. This again illustrates that the analysis of the terminations is not readily performed in the frequency domain.

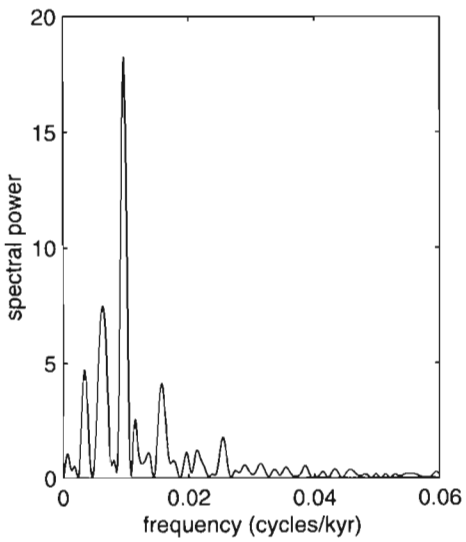


Fig. 3.4. Spectrum of randomised sawtooth.

3.2.6 Fast Fourier Transform (FFT)

The FFT is simply an extremely efficient way to calculate the Fourier transform for the case of equally-spaced data. It is such a popular technique that many analysts will extrapolate their data to evenly-spaced points just so they can use this method. An ordinary digital Fourier transform takes N data points, and calculates the real and imaginary amplitudes at $N/2$ frequencies. To do this requires N^2 transcendental calculations (i.e. computation of sines or cosines). Based on a suggestion of Richard Garwin, Cooley and Tukey developed a method of doing the identical calculation in a way that was proportional to $N \log_2 N$ rather than to N^2. For very large computations this provides an enormous saving in computer time.[1] For example, in astronomy, it is frequently necessary to compute the Fourier transforms of $N = 10^7$ data points. The ratio of computation times for the FFT vs. the Glacial Fourier Transform (GFT) is then $\log_2 N/\mathrm{N} = 2.3 \times 10^{-6}$. A computation that takes one second using the FFT would take five days using the GFT.

Many people who use the FFT find it confusing initially. The FFT requires data at equally-spaced time points, and is most efficient when the number of points is an exact power of two. The process of interpolation to evenly-spaced data introduces additional biases and systematic error. The format of the output can also be confusing. For real data consisting of N data points y_j, each taken at time t_j, the power spectrum output from the FFT appears as a set of $N + 1$ data points. The first and last data points are the same, and they represent the power at frequency zero. The second through to the $N/2 + 1$ data points represent the power at evenly-spaced frequencies up to the Nyquist frequency ($= 1/2\Delta t$, where Δt is the time between equally-spaced data points). The spectral power for a given frequency is spread out over several bins; an optimum determination of the power requires combining this information. In the standard FFT, the original data are overwritten with the Fourier transform, erasing it. This procedure dates back to the days when computer memory was a limiting expense. It is hardly ever worthwhile today.

Despite these complications (which are routinely handled by a computer program, once it is written and debugged, and forgotten by the user), the FFT is worth using whenever speed is important. Especially when doing a thousand or more Monte Carlo simulations (a procedure we highly recommend whenever you encounter a new kind of data), the speed of the FFT proves valuable.

3.2.7 Padding with zeros

The FFT calculates the amplitude for a particular set of pre-chosen frequencies. If there are N data points, then $N/2$ complex amplitudes are calculated at $N/2$

[1] Although the practical use of the FFT began with the Cooley/Tukey implementation, there is a long history of mathematicians recognising that faster computations could be done. Gauss observed that he could break a Fourier series into two parts, and reduce the computation significantly, but he never published his work. Danielson and Lanczos in 1942 derived the entire algorithm. Could a faster algorithm be invented some day? For standard computing (Turing machine), the answer is no.

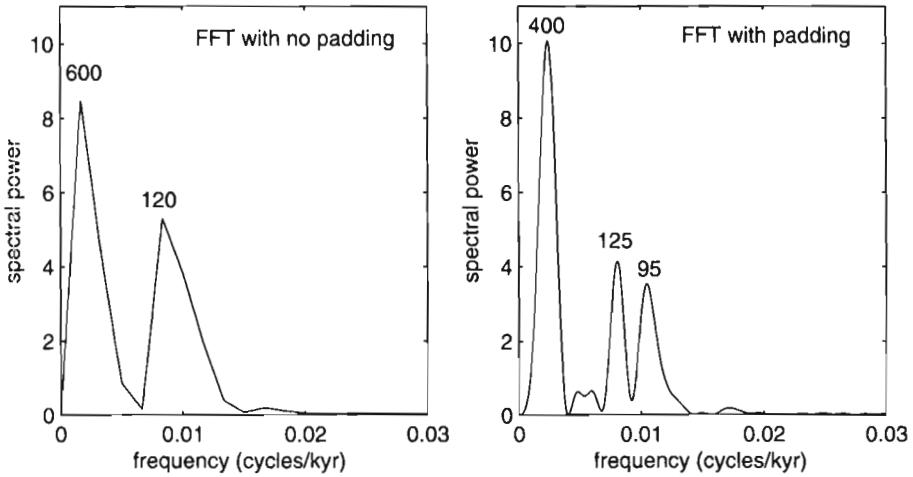


Fig. 3.5. Effect of padding on spectrum of eccentricity.

different frequencies. The lowest (non-zero) frequency is $1/T$, where T is the duration of the record. This frequency corresponds to one cycle over the entire data set. All other frequencies are multiples of this. Unfortunately, none of these may correspond to the true frequencies present in the record. For example, suppose we have 200 data points, covering a period of 0 to 500 kyr. Then the lowest frequency is $1/500 = 0.002$. If the data are dominated by a 41 kyr signal, then the frequency of that signal is $1/41 = 0.0244$. This is not an even multiple of 0.002, so the FFT will not calculate the amplitude at this frequency.

A simple and elegant way around this is to subtract the mean from the data, and then pad it with zeros, i.e. extend the data beyond T (to, say, $10 \times T$) by adding zeros to it. The FFT takes about 10 times longer, but computers are fast. The ideal way is to pad to an exact power of 2. With this padding, the minimum frequency is now 0.0002, and the FFT *will* calculate the amplitude for the frequency 0.0244. (Of course, 1/41 is not exactly 0.0244...)

We illustrate the value of padding by calculating the spectral power of eccentricity for the period 0 to 600 kyr, with and without zero padding. The results are plotted in Fig. 3.5. Note that without padding, the structure of the spectral peaks is invisible. That is because much of the information is hidden in the phase, which is not included in the plot of spectral power. With the padded data, intermediate frequencies are calculated, and the clear contributions of the peaks with periods of 400, 125 and 95 kyr are evident.

3.2.8 Aliasing

A problem arises if there are fluctuations in your data that are more rapid than your sampling. (This is the same as saying that there are frequencies in the data that are higher than the Nyquist frequency.) Suppose, for example, that you have samples of Greenland Ice that represent the temperature every 100 years. But your sampling is

not precisely spaced by a year, so sometimes you measure winter ice, and other samples are summer ice. Even if there is no long-term variation in the temperature, you will see jumps up and down in your data. Unless you know the time scale exceptionally well, you may not be able to tell that the variations you see are actually seasonal. Yet, the Fourier transform will "explain" all these fluctuations as being due to cycles with periods greater than 100 years. It will do this because the transform is a mathematical identity; the data can *always* be expressed in terms of sines and cosines, even though the Fourier transform uses only a period longer than twice the sample period. The maths finds a fit, even though the real frequency is not included in the expansion series. The Fourier transform is lying to you!

As an example, suppose you measure data at 50 kyr intervals, i.e. at $t = (0, 50, 100, 150, 200, \ldots)$, and get the values $y = (1, -1, 1, -1, 1, \ldots)$. Let us assume that these data points resulted from an oscillating signal with a period of 20 kyr, i.e. they are explained by the function $y = \cos(2\pi ft)$ with $f = 1/20 = 0.05$. But if we didn't know that, we might think that they were due to frequency $f = 1/60$. We plot the data points along with both sine waves in Fig. 3.6, and you will notice that both frequencies agree with the data points. There is no way for the Fourier transform to know which of the frequencies is the correct one, since it has only the sampled data.

Which will the Fourier transform pick to represent the data? It will give a large amplitude at frequency $f = 1/60$, despite the fact that we created the signal from $f = 1/20$. This misidentification is given the name "aliasing". It occurs whenever the sampling is less frequent than the period of the true oscillation. For a further discussion of the mathematical aspects of aliasing, see Press et al. (1993).

Aliasing can be a problem in ice data, as the example illustrates. It is usually not a problem in sea floor data, since the periods of interest (from 19 kyr to 400 kyr) are usually much longer than the sample interval. Any phenomena that tend to smooth the data (such as bioturbation) tend to eliminate aliasing.

We mentioned earlier (Section 3.2.1) that some audiophiles complain that music CDs suffer from aliasing. The sample period for these CDs was chosen to be 1/44000

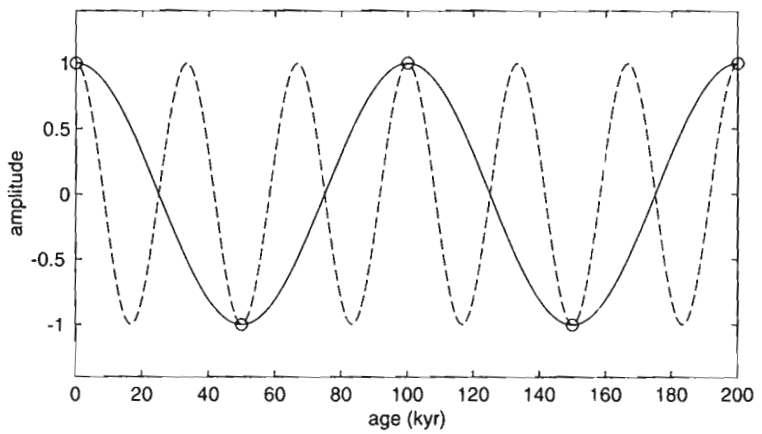


Fig. 3.6. Aliasing. The circles are the data points.

sec, so that the Nyquist frequency is 22000 cycles/sec, above the audible range. But musical instruments produce sounds at higher frequencies. When music from these instruments is recorded digitally, the sound is aliased to lower frequencies. When this happens, the CD contains sounds at audible frequencies which would not be present in the live performance. Recording engineers can eliminate such aliased sounds by putting in a filter that removes the higher frequencies prior to the music being digitised. If they are absent in the original, then they are not aliased. However, any filter applied to the signal has *some* effect at lower frequencies, particularly since the ear does not respond in a linear way to sound, and this too can introduce an objectionable distortion. Audiophiles will say that they prefer listening to music to which no filter has been applied. Of course, any recording medium (such as records) has a frequency response, and therefore acts as a filter. The question isn't whether or not we should apply a filter, but which filter has the most objectionable effect on the music.

3.3 BLACKMAN–TUKEY METHOD

The Blackman–Tukey method was the most commonly applied spectral analysis tool for paleoclimate studies in the last three decades of the twentieth century and, as such, it deserves special attention. It was introduced in a classic book (Blackman and Tukey, 1958) that introduced many scientists to fundamental techniques of spectral analysis. It was published before the rediscovery of the Fast Fourier Transform by Tukey. Blackman and Tukey were primarily concerned with the difficult case of analysing a continuous spectrum that was biased by the presence of sidelobes from strong peaks. Part of the attraction of the book is that it offered a prescription for scientists to follow. The book also offered an abundance of warnings—that the algorithm and suggested parameters should not be followed slavishly, but that it should be used only as a starting point. As an added bonus, the method made efficient use of tricks to reduce the required calculation of sines and cosines. This was important at that time since the book predated the FFT by seven years.

The Blackman–Tukey method is based on a fundamental theorem of Fourier analysis that states that the Fourier transform of a correlation is equal to the product of the Fourier transforms. The correlation of two functions A and B is defined as

$$C(\tau) \equiv g \otimes h \equiv \int_{-\infty}^{\infty} g(t+\tau)h^*(t)dt$$

The star * indicates complex conjugation. The correlation theorem states:

$$FT(g \otimes h) = FT(g)FT^*(h)$$

A special case of this, when $g = h$, is known as the Wiener–Khintchine theorem:

$$FT(g \otimes g) = |FT(g)|^2 = P$$

where P is the spectral power. In words, the Fourier transform of the autocorrelation function is identical to the spectral power. In the Blackman–Tukey method, an estimate of the spectral power is obtained by taking the Fourier transform of only part of the autocorrelation function.

The algorithm involves several steps. The first, optional, step is called prewhitening; it is designed to reduce the bias in frequency determination that happens when there is a strongly varying background under a weak peak. The next step is also optional: the data are tapered by application of a window; for further discussion of this, see Section 3.4. But the heart of the method is the calculation of the partial autocorrelation function, defined by

$$A_{1/N}(\tau) = \int_0^{T/N} f(t+\tau)f(t)dt$$

Since we are using data with real values, we have assumed that $f = f^*$. The key difference between this and the full autocorrelation function is that the integral is evaluated only to T/N, where T is the length of the data record. If $N = 1$, then this is the full autocorrelation function, and the method gives the same answer as the ordinary periodogram. A more typical value is $N = 3$; when this is chosen (and it is the value that Blackman and Tukey recommended as a starting value), then we say that "a lag of 1/3" was used. Alternatively, we can specify the number of data points that appear in the overlap integral. Thus, for example, if there are 99 data points, and $N = 3$ was used, then we might say that there were $99/N = 33$ lags used.

Note that this "partial" autocorrelation function is symmetric, i.e. $A(-\tau) = A(\tau)$. If we take the Fourier transform of this we get P_{BT}, the Blackman–Tukey estimate of the power for 1/3 lag:

$$P_{BT}(f) = \int_{-\infty}^{+\infty} e^{2\pi i f t} A_{BT}(\tau)d\tau \equiv FT(A_{BT})$$

Since the autocorrelation function is symmetric, only the real part is non-zero. Therefore, in the pre-FFT days, we could save half of the computation time by computing only the symmetric part:

$$P_{BT}(f) = 2\int_0^{+\infty} \cos(2\pi f) A_{BT}(\tau)$$

If the lag were one, then this procedure gives the same estimate of the power as we get by squaring the Fourier transform:

$$P(f) = \int_{-\infty}^{+\infty} e^{2\pi i f \omega \tau} A(\tau) \equiv FT(A)$$

However, if only part of the full autocorrelation function is used, the result is different. To see what the effect is, consider the pedestal function

$$B(\tau) = 1 \quad \text{for} \quad -T/3 < \tau < T/3$$

$$B(\tau) = 0 \quad \text{otherwise}$$

Then we can write

$$A_{1/3}(\tau) = A(\tau)B(\tau)$$

and the Blackman–Tukey estimate of the power is

$$P_{BT} = FT(AB) = FT(A) \otimes FT(B)$$
$$\equiv P \otimes b$$

Where $b \equiv FT(B)$. This shows that the effect of the Blackman–Tukey method is to blur (or smooth) the spectral power function. The amount of smoothing is the same that you get by convolving the spectral power with the blurring function b. For the simple pedestal function $B(\tau)$ given above, b is

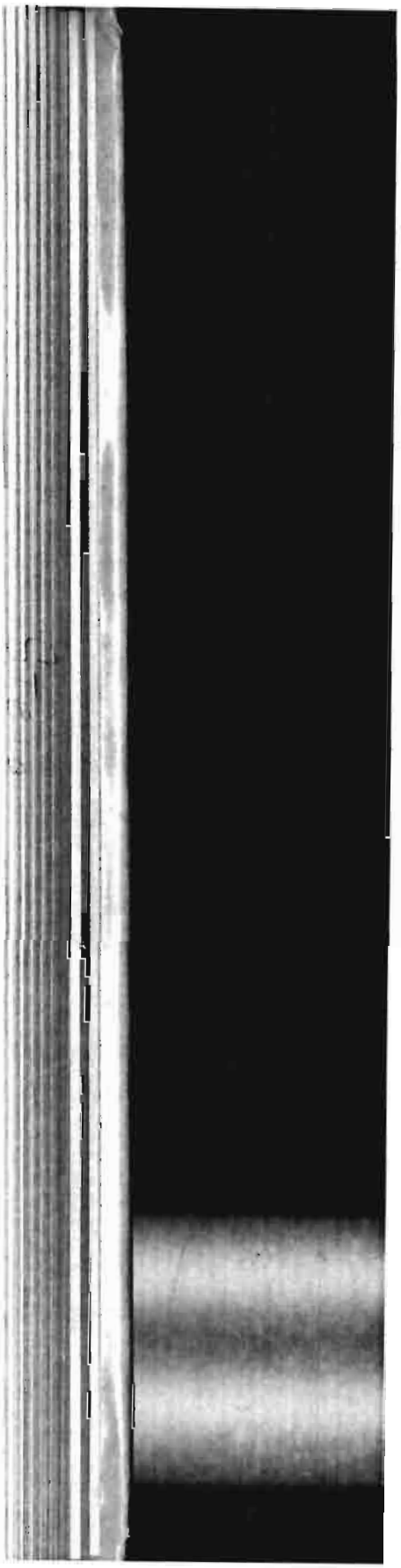
$$b(f) = \int_{-\infty}^{+\infty} e^{2\pi i f t} B(t)dt = \int_{-T/3}^{T/3} e^{2\pi i f t} dt$$
$$= \int_{-T/3}^{T/3} [\cos(2\pi f t) + i \sin(2\pi f t)]dt$$
$$= \frac{\sin\left(\frac{2\pi f T}{3}\right)}{\pi f}$$

The approximate width of this blurring function is given by the condition $\Delta f = 3/(2T)$. Recall that for the full Fourier transform, the resolution is approximately $1/T$. In using 1/3 lags, we introduced another blurring, and we have lost a factor of about 3/2 in resolution.

The kind of blurring that we get with the Blackman–Tukey method is fundamentally different than the blurring that we get when we use a taper (Section 3.4). With a taper, it is the Fourier transform that is blurred; with Blackman–Tukey, it is the spectral power. This difference is important. A spectral amplitude that is rapidly varying will be averaged to zero if the blurring is done with a taper. However, with the Blackman–Tukey method, a rapidly varying amplitude does not necessarily average to zero, since the process of squaring can make the function positive over the region of blurring.

If resolution is lost, why use the Blackman–Tukey method? One reason is because blurring is sometimes desired. Drift in the time scale behaves like phase modulation, and can introduce spurious frequencies. We will discuss this at length in Section 5.1. Blackman–Tukey blurring can average these sidelobes into the main peak, and thereby give a better estimate of the true power. Note that windowing of the data with a taper function will not have the same effect, since the sidelobes will not have the same phase, and if averaged in amplitude, they can reduce the strength of the spectral peak. We'll demonstrate this property with examples.

But this advantage only works if the variations in sedimentation that are being compensated are small enough that the width is only broadened a little. Suppose the sedimentation rate were different in the two periods 0–300 kyr. and 300–600 kyr.



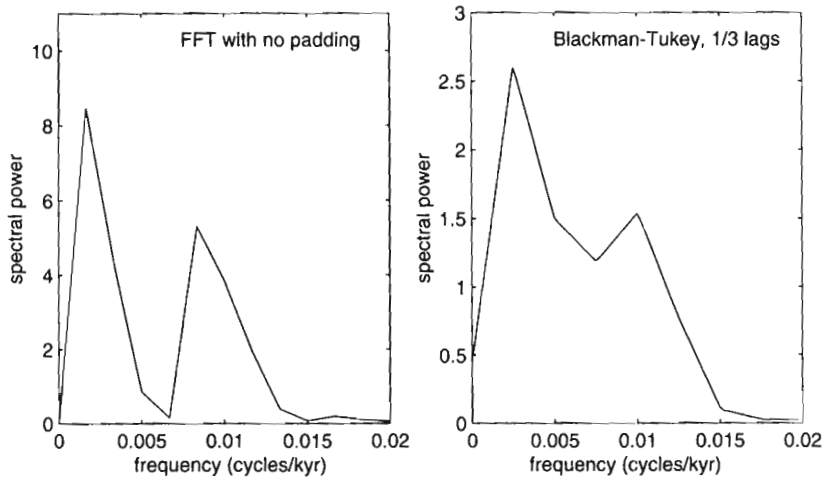


Fig. 3.7. Eccentricity spectrum, using FFT and 1/3 lags Blackman–Tukey.

Then the apparent frequencies of the 100 yr cycle would be different in those two records. The Blackman–Tukey method would average the two peaks together. Note that using large lags (e.g. 1/3) makes the least sense when tuning is used, since tuning (if done correctly) removes the low frequency drifts.

3.3.1 Example: eccentricity

As a vivid example of the loss of resolution that can result by the unthinking application of the Blackman–Tukey method with 1/3 lags, we show the spectrum of eccentricity for the period 0–600 kyr, calculated in three different ways. In Fig. 3.7 we show the spectrum from an unpadded FFT, and from the Blackman–Tukey method. In the FFT plot, the separation between the 400 kyr peak ($f = 0.0025$) and the 100 kyr peak ($f = 0.01$) is clear. In the Blackman–Tukey plot, the peaks are not even resolved (by the Rayleigh criterion, the dip in the centre should drop to half the maximum of the nearest peak). Of course, it makes no sense to use 1/3 lags for eccentricity, since the time scale is perfectly known. In Fig. 3.8 we show the spectrum calculated using the FFT with zero padding. With this method, the full triplet of eccentricity peaks is easily resolved.

3.3.2 Example: Site V12-122

We'll illustrate the value of the Blackman–Tukey method by showing several spectra from an historically interesting data set published by Broecker and Donk, for Site V12-122, for the relatively short time period of 0 to 370 kyr. First, we show the data, in Fig. 3.9.

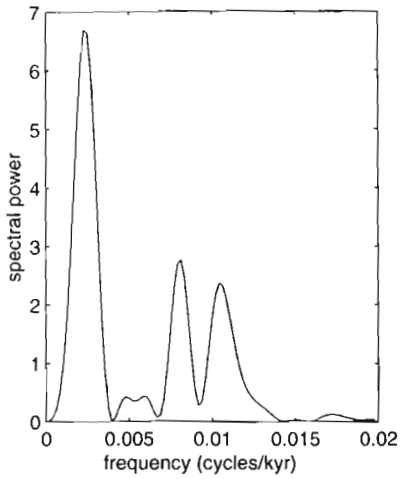


Fig. 3.8. Eccentricity spectrum using FFT with zero-padding.

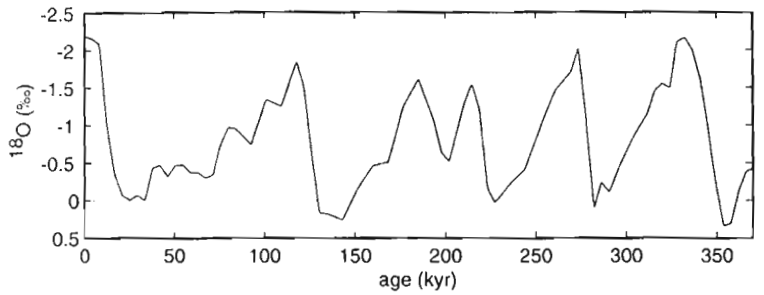


Fig. 3.9. $\delta^{18}O$ for Site V12-122.

If we plot the unwindowed periodogram for these data, we get the spectrum shown in Fig. 3.10.

The Blackman–Tukey spectrum, with lag of 2/3, is shown in Fig. 3.11. The spectrum appears smoother and more readily interpretable. There appear to be two major peaks, one near frequency 0.01, and one near 0.03 cycles/kyr.

In Fig. 3.12 we show the result of using a lag of 1/2. The pattern is becoming evident. The shorter the lag used, the greater is the blurring of the spectrum. In Fig. 3.13, we show the results of using a lag of 1/3. And finally, in Fig. 3.14, we show the Blackman–Tukey spectral estimate for a lag of 1/4.

From looking at these, we would probably pick the lag of 1/2 as the most useful for determining the frequencies and relative powers of the peaks. What would you pick? (Remember, there is no "right" answer.)

As we mentioned earlier, there is another reason for using the Blackman–Tukey method. The Blackman–Tukey book was published in 1958, several years prior to the introduction of the FFT. Prior to the FFT, the calculation of Fourier transforms

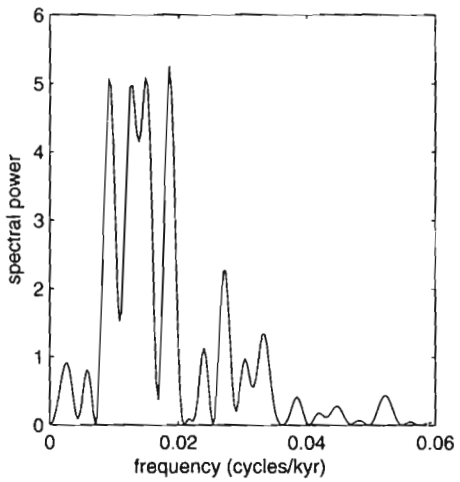


Fig. 3.10. Spectrum of V12-122, no window.

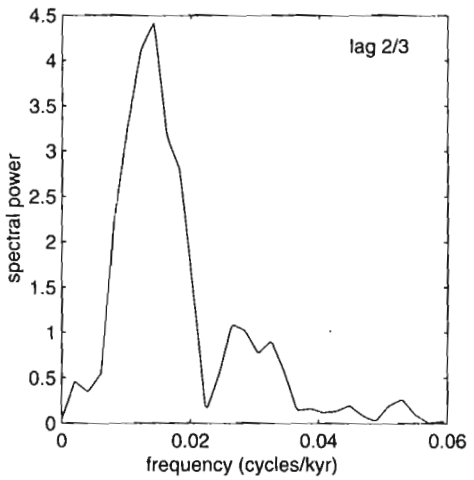


Fig. 3.11. Blackman–Tukey Spectrum V12-122, lag 2/3.

was a slow process that used a lot of computer time. A Fourier transform of N points would require the calculation of N^2 sines and cosines. For the computers of the day, even relatively small values for N (such as 1000) were a considerable expense. The Blackman–Tukey method saved substantial time by reducing the size of the data set by a factor of the lag. If the data had chatter, then there was no loss; in fact, if the sidelobes were brought back into the main peak, the estimate of spectral power was improved.

But even more time was saved by the fact that the autocorrelation function is symmetric. This meant that in taking its Fourier transform, only the cosine components need be calculated. This saved an additional factor of two. The corre-

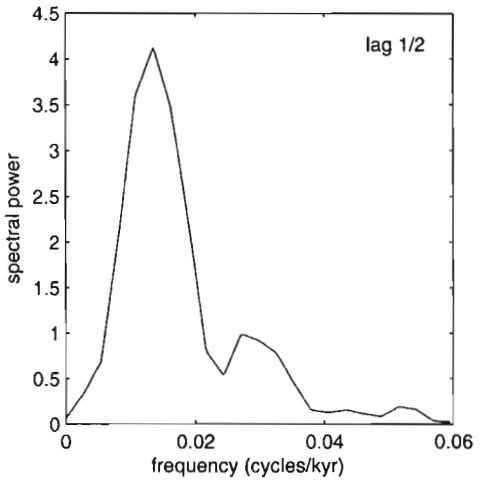


Fig. 3.12. Blackman–Tukey of V12-122, lag 1/2.

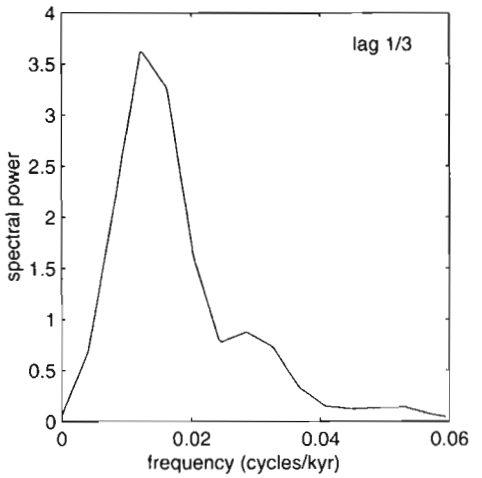


Fig. 3.13. Blackman–Tukey of V12-122, lag 1/3.

sponding disadvantage is that the method does not give any of the phases of the Fourier components. But if all you need is the power estimate, then this could be considered a worthwhile saving of time, at least in the pre-FFT days of slow computers.

A second disadvantage of the Blackman–Tukey method is that you have to estimate the proper lag to do before you do the calculation. If you then want to apply a different lag, you have to redo the entire calculation. In their book, Blackman and Tukey recommended starting with the value 1/3 for the lag, but

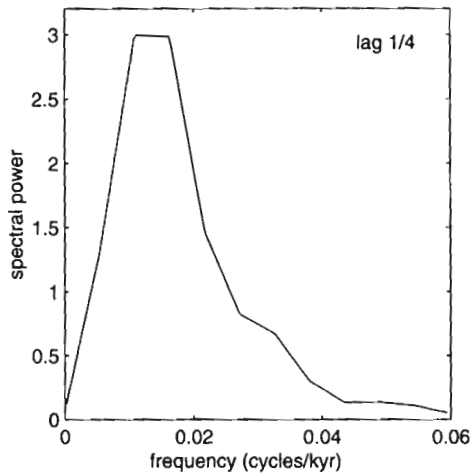


Fig. 3.14. Blackman–Tukey of V12-122 with lag 1/4.

then varying it until it gave the best performance. However, the paleoclimate community has rarely done this in practice. In the period 1976 to 1995, the vast bulk of spectral analysis carried out by the paleoclimate community used the Blackman–Tukey method with 1/3 lags. Few papers ever mention other lags. (Actually, we are not aware of *any* paper that used the Blackman–Tukey method with lag other than 1/3, although we haven't studied every paper written during that period.) Lags of 1/3 became the *de facto* standard, even in software programs distributed within the community. The result was a loss of resolution, and the missed discovery that the 100 kyr peak was single and narrow, in disagreement with the predicted structure from the Milankovitch models.

3.4 TAPERS

Tapering is a very simple, very clever mathematical technique that is extremely popular in spectral analysis. Instead of doing spectral analysis on the data $f(t)$, the analysis is carried out on the data multiplied by a taper function $g(t)$, also called the window function. The function g is usually taken to be a simple, slowly varying function, often going to zero (or close to it) near the edges. Several important tapers, for the interval $t = 0$ to T, are:

Sine taper	$g(t) = \sin(\pi t/T)$
Hanning (offset cosine) taper	$g(t) = (1/2)(1 - \cos(2\pi t/T))$
Hamming taper	$g(t) = .54 - .46\cos(2\pi t/T)$
Parzen or Bartlett (triangle) window	$g(t) = 1 - (t - T/2)/(T/2)$
Welch (parabolic) window	$g(t) = 1 - (t - T/2)^2/(T/2)^2$
Daniell (untapered or rectangular) window	$g(t) = 1$

The last of these is included only because windowing is so common that even an untapered (unwindowed) spectrum is sometimes referred to as if it were.

The effect of tapering is sometimes described as reducing the artificial presence of the high frequencies generated by fact that the data end suddenly at $t = 0$ and $t = T$. In fact, if discontinuities are present in the original signal, then the Fourier transform introduces distortions at those locations, a property known as the Gibbs phenomenon. Of course, tapering also introduces distortions. You just have to decide which kinds of distortions are more acceptable.

The ability of tapering to suppress high frequencies generated by sharp changes finds a very practical use in optics (especially radar) where the radiation pattern is the Fourier transform of the aperture function. A tapered aperture had significantly reduced sidelobes. This effect is shown in the following example. Let

$$h = \sin(2\pi t/95) + \sin(2\pi t/125)$$

represent data with two oscillations, one with period 95 kyr and the other with period 125 kyr, analogous to eccentricity. We will apply a Hanning taper to the data. (In Appendix 2, we give explicit code for calculating the spectral power with and without the taper.) The results are shown in the two plots of Fig. 3.15. The solid line is the spectrum of the untapered data; the dotted line is that of the tapered data. Both plots show the spectral power; the left one is a linear plot, and the right one is a semilog plot.

It is clear from the linear plot on the left, that application of the taper has severely degraded the resolution. It is equally clear from the right plot that the sidelobes of the tapered spectrum are much reduced. If your primary interest is in the resolution of the two peaks, then the untapered periodogram is superior. If your primary interest is to reduce the bias that sidelobes from these peaks could apply to other nearby peaks, then the tapered spectrum is superior.

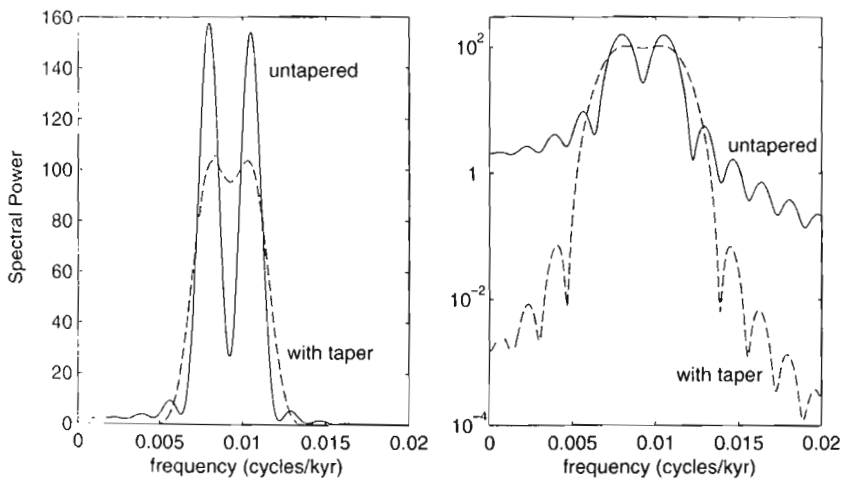


Fig. 3.15. Effect of tapering (dotted line) on a doublet peak.

The effect of the windowing can most easily be understood when the spectral analysis is done by a Fourier transform. The Fourier transform of a product is equal to the convolution of the Fourier transforms of the individual functions:

$$FT(fg) = FT(f) \circ FT(g)$$

Because the taper functions are broad and slowly varying, their Fourier transforms $FT(g)$ are narrow. The effect of tapering the data then is to convolve the Fourier transform with the narrow Fourier transform of the taper function. This amounts to smoothing. The difference between this and the Blackman–Tukey smoothing is that the smoothing is done in the complex Fourier space before the spectrum is squared.

The effect of tapering on the sidelobes can also be understood in terms of this convolution. An untapered sine wave of length T is equivalent to an infinite sine wave multiplied by a boxcar function. The Fourier transform of this is the transform of an infinite sine wave, i.e. a δ-function, convolved with the Fourier transform of the boxcar. For a tapered sine wave, the Fourier transform is the δ-function convolved with the Fourier transform of the taper function. If the taper function has lower sidelobes than the boxcar, then the spectral power of the tapered sine wave will also have lower sidelobes.

It is also possible to invent tapers that increase the resolution of the spectrum, by adding additional power at the ends, and decreasing the amplitude in the centre of the data. For example, we could multiply our data by the taper

$$g(t) = (1 + \cos(2\pi t/T))/2$$

The spectrum that we get with this taper is shown Fig. 3.16. It has better resolution than the untapered plot, and a more dramatic separation of the two peaks, but the sidelobes are getting uncomfortably large, and the fraction of the power in the main peaks is decreasing, making the spectrum more susceptible to background.

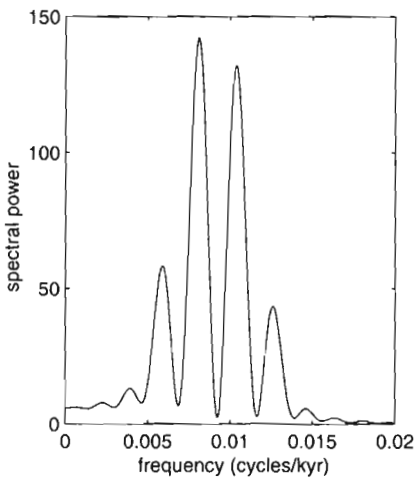


Fig. 3.16. Spectrum of the doublet using a high-resolution taper.

Some authors are adamantly opposed to tapering. Yuen has argued that tapering is "statistically unsound" (Yuen 1979, quoted in Percival and Walden, 1993, p. 215). Fougere calls tapering "tampering" (quoted in Percival and Walden, 1993, p. 215).

3.5 LOMB–SCARGLE PERIODOGRAM

Lomb (1976) and Scargle (1982) improved on the simple periodogram by a slight alteration. What they showed is that if the cosine and sine coefficients are normalised separately (we'll show explicitly what we mean by this in a moment) then the classic periodogram can be used with unevenly spaced data, and yet the statistical behaviour of the power is identical to the behaviour you would expect if you had evenly-spaced points.

To calculate the Lomb–Scargle periodogram of a data set (t_k, y_k), first calculate the usual mean and the variance:

$$\bar{y} = \frac{1}{N}\sum_{k=1}^{N} y_k$$

$$\sigma^2 = \frac{1}{N-1}\sum_{k=1}^{N}[y_k - \bar{y}]^2$$

For every frequency f, define the time constant τ by

$$\tan(4\pi f\tau) = \frac{\sum \sin(4\pi f t_k)}{\sum \cos(4\pi f t_k)}$$

Then the Lomb–Scargle periodogram estimate of the spectral power $P(f)$ at frequency f is given by

$$P(f) = \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_k (y_k - \bar{y})\cos 2\pi f(t_k - \tau)\right]^2}{\sum_k \cos^2 2\pi f(t_k - \tau)} + \frac{\left[\sum_k (y_k - \bar{y})\sin 2\pi f(t_k - \tau)\right]^2}{\sum_k \sin^2 2\pi f(t_k - \tau)}\right\}$$

This equation is less imposing than it looks. It has two terms, one for the cosine transform, the other for the sine transform. Each term is normalised separately. The only complication is that each frequency uses a different time offset τ. Other than these changes, the equation looks just like an ordinary digital Fourier transform.

The Lomb–Scargle method has several advantages over the classic periodogram. One, obviously, is that much paleoclimate data are not evenly spaced. Although this can be handled by interpolation, the statistical effects of such interpolation can be complicated. Secondly, there is a limit to the ordinary periodogram that comes about from a process called aliasing. What this means is that two signals of different frequencies can have identical sets of values if the samples are taken at exactly even spacing. With unevenly-spaced data, this effect can be substantially reduced. The net result is that the Lomb–Scargle periodogram can measure frequencies that would be aliased in evenly-spaced data.

One note of caution: although superior to many others, it takes a degree of courage (recklessness?) to use the Lomb–Scargle method in the paleoclimate community. We have had referees, unfamiliar with the method, urging rejection of our papers for using a method so "obscure"! (Obscure, in this context, presumably means that the referee had not heard of it.)

3.6 MULTITAPER ANALYSIS

We discussed in Section 3.4 how applying a taper (or window) to the data can reduce the sidelobes of spectral lines, and thus reduce the "leakage", i.e. the effect that strong lines can have in putting excess power at adjacent frequencies. D. Thomson recognised that the following problem had not been solved in a general way: how can you achieve the best sensitivity for finding a small signal of known frequency in the presence of another signal which had a different frequency and much greater power? The solution turned out to be a procedure in which several different tapers were applied to the data, and the resulting powers then averaged. This method is called "multitaper analysis".

The method is widely misunderstood. It has been incorrectly referred to as having "high resolution", although its basic idea is to trade resolution for sidelobe reduction. To apply the method, you must choose the number of tapers that you will apply. If you use only a few tapers, the resolution will not be degraded as much, but you will not get as great a sidelobe reduction. For the method to be optimum, the tapers should be Slepian functions. These are not easily computed, and numerical integration is usually employed. There are built-in Slepian functions available in the optional Matlab Spectral toolbox, and they are quite fast. If you don't have this, you should consider an alternative suggested by Walden et al. (1995): you can do almost as well using sine wave tapers instead of Slepian functions.

The advantages and disadvantages of the multitaper method are illustrated in Fig. 3.17. This plot shows two analyses of the signal $y = \sin(2\pi f_1 t) + \sin(2\pi f_2 t)$ where $f_1 = 1/95$ and $f_2 = 1/125$ cycles/kyr. (These, of course, are two important frequencies present in the eccentricity variation.) Shown is a periodogram of the untapered function, and multitaper analysis using four tapers. The periodogram is superior at resolving the two close peaks, but the multitaper spectrum is superior at suppressing sidelobes outside the main peak.

3.7 MAXIMUM LIKELIHOOD ANALYSIS

The maximum likelihood method is a generic approach to problems in statistical estimation. If you have a model with adjustable parameters (e.g. the amplitudes of eccentricity, obliquity and background), then you can calculate as a function of those parameters the relative probability that your model would yield data that match the real data. The parameters are then adjusted to maximise this relative probability. The relative probability that the model and a set of parameters account for the data

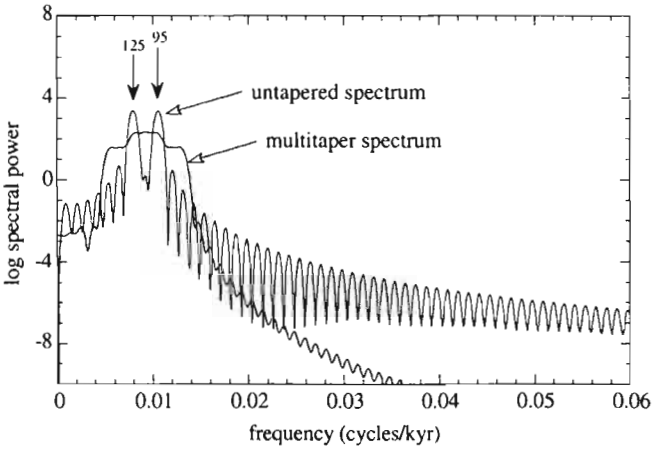


Fig. 3.17. Multitaper method compared to periodogram.

is called the likelihood. Many methods apply this approach, and differ only in the details, e.g. what specific parameters they adjust.

In this section we will discuss an approach to spectral analysis developed by MacDonald (1989). If the record is a noise process with a structured signal, this method provides the best means of signal detection in the maximum likelihood sense. We will describe the method, and then apply it to one of the historically most interesting data sets, the stack used in the 1976 article, "Pacemaker of the Ice Ages" (Hays et al., 1976). We will show that the data clearly verify the discovery of the important 41 and 23 kyr peaks, but that the claim in that paper that they saw a 19 kyr peak is questionable.

The method consists of the following steps:

1. First select an acceptable false alarm probability. The selection is a matter of judgement. If the probability is set too low, physically significant peaks in the spectrum may be missed; if it is set too high, peaks that are due to noise will be mislabelled as true peaks. A typical value might be 0.05. Then the chance of identifying a noise peak as a true peak is 5%.
2. Calculate the normalised periodogram. (Note: if the record is a noise process without any signal, this method is inappropriate.) If the data are not equally spaced, use the Lomb periodogram instead of the FFT.
3. Identify the maximum peak and test it against the false alarm probability.
4. If the maximum peak meets the false alarm test, determine by least squares the amplitude and phase of the sinusoid representing the peak.
5. Using the amplitude and phase information from the least-squares fit, subtract the sinusoidal curve from the data. (Because the record length is left unchanged, this method is superior to band-filtering out the peak.) Removal of the peak also removes the disturbing sidelobes as the analysis continues, and is the appropriate form of prewhitening. After peak removal, the variance in the

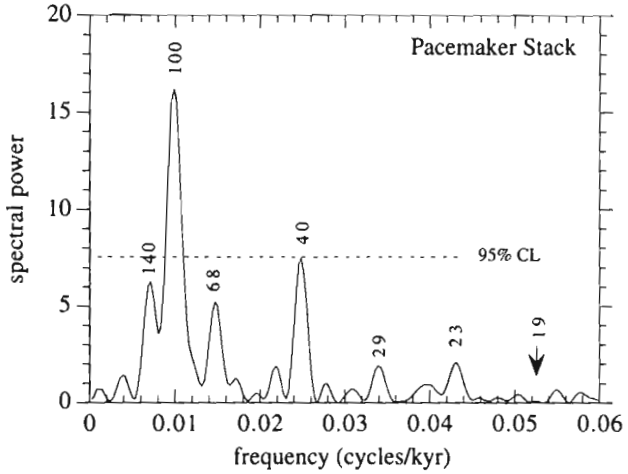


Fig. 3.18. Spectrum of Pacemaker stack.

total record is reduced. The change in variance provides guidance as to the rate of convergence of the peak removal process.

6. Recalculate the periodogram, and repeat the process of sine wave removal, until no further peaks meet the false alarm criterion. The remaining series is an estimate of the continuum part of the spectrum of the original series. The continuum residual can then be analysed to determine the nature of the noise or of the deterministic continuum.

We now apply this procedure to the "Pacemaker stack" of Hays et al. (1976). The first step is to do a spectral analysis. We did this using the Lomb–Scargle method on the $\delta^{18}O$ data of the stack. The resulting spectrum is shown in Fig. 3.18.

The maximum peak in the spectrum is at f = 0.01 cycles/kyr. Next, we do a least-squares fit of the Pacemaker data to a sine wave with this frequency. The results are shown in Fig. 3.19, plotted on top of the original $\delta^{18}O$ data. Next subtract this sine wave from the data. The data, with the 100 kyr cycle removed, is shown in Fig. 3.20.

A new spectrum is calculated, and it is found that the strongest remaining peak is at 41 kyr period. We obtain the best fit sine wave of this period, and subtract it from the data. The process then continues. The four strongest periods removed in this procedure turn out to be 102, 41, 140 and 69 cycles per kyr. When these have all been removed, the spectrum appears as shown in Fig. 3.21.

The next peaks that are removed are at 23 and then at 46. When these have been subtracted, the resulting plot looks as follows (Fig. 3.22).

The strongest remaining peak is at 30 kyr, followed by a pair at 20 and 19. How many of these peaks were statistically significant? That is trickier to answer than you might think. One conclusion is clear: the 19 kyr peak that was reported as significant in the original Pacemaker paper, has extremely low statistical significance. There are eight other peaks of greater statistical significance.

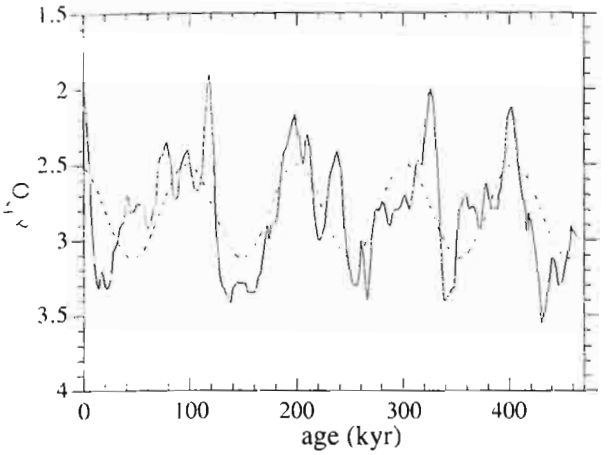


Fig. 3.19. Pacemaker stack, fit to 100 kyr cycle.

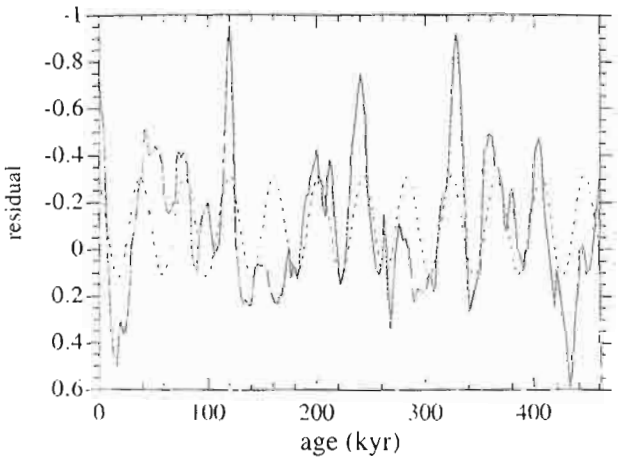


Fig. 3.20. Pacemaker residuals, 100 kyr removed, fit to 41.

The evidence presented in favour of the 19 kyr peak, in the original Pacemaker paper, was a Blackman–Tukey analysis using 1/3 lags. A plot using this method is shown in Fig. 3.23, on a log plot. (The 19 is too small to see on a linear plot.)

The small peak at 19 is what led to the reported presence of this peak. But was it real, or just a bump on top of a continuous red noise spectrum? Let us analyse the data yet another way. In Fig. 3.24 we plot the Fourier spectrum of the Pacemaker stack using a Hanning taper. A line marks the expected location of the 19 kyr peak. The location of the 19 kyr peak occurs less than half way up the side of a peak near the 20 kyr period. We conclude that the 19 kyr peak is not present in this analysis. It was not present in the Maximum Likelihood analysis. It was doubtful in the original

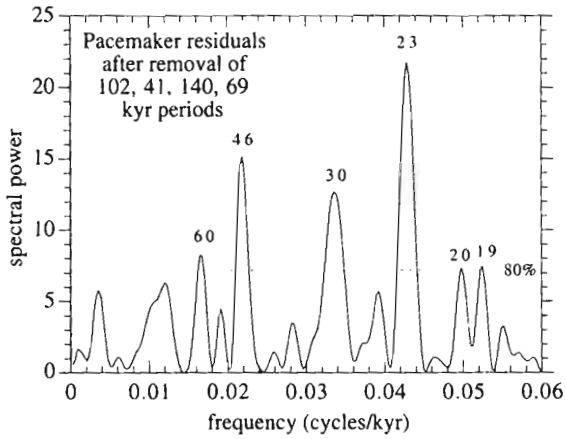


Fig. 3.21. Pacemaker stack spectrum after removal of 4 cycles.

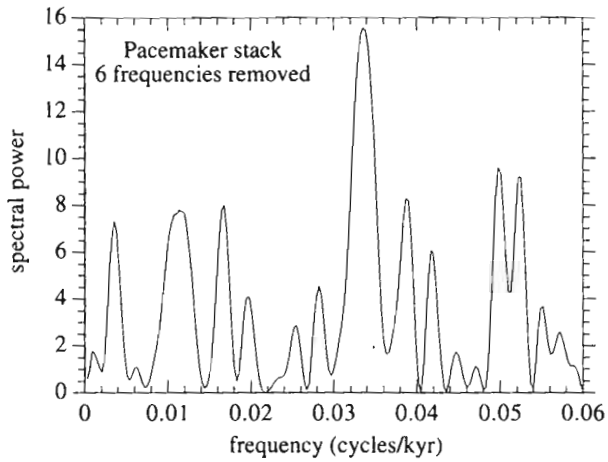


Fig. 3.22. Pacemaker stack spectrum, 6 frequencies removed.

Blackman–Tukey analysis. We conclude that its presence is not statistically significant.

But the 19 kyr peak appears to show up in other data, e.g. the Vostok atmospheric $\delta^{18}O$ (Fig. 4.19) and the terrigenous component from Site 721 (Fig. 4.23). Is it worthwhile to reanalyse old data? Yes, it is necessary, if we are truly to look at all the data as if for the first time. To untangle the mechanisms of the ice ages, it is very important to know where signals appear, and where they don't. Many people who study paleoclimate believe that the 19 kyr signal showed strongly in the Pacemaker stack—after all, that is where it was "discovered" (after having been predicted theoretically). To solve a jigsaw puzzle, it is just as important to remove pieces that were improperly placed as it is to put new pieces in their correct locations.

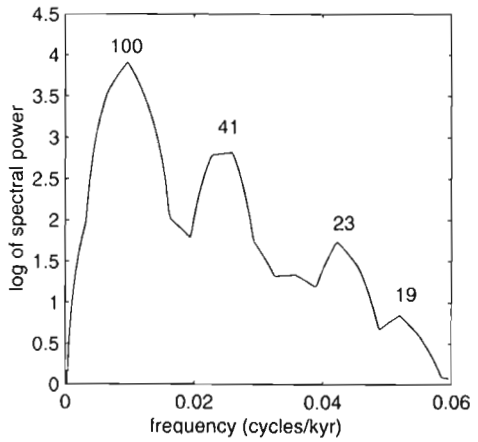


Fig. 3.23. Pacemaker spectrum, log of Blackman–Tukey with 1/3 lags.

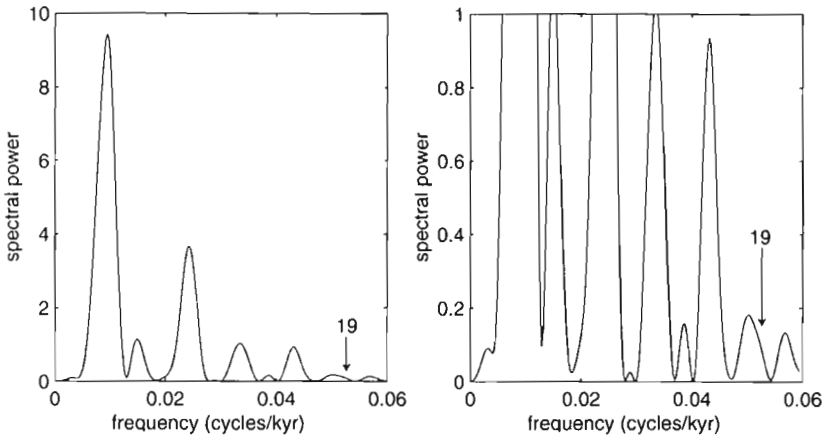


Fig. 3.24. Pacemaker spectrum, Hanning taper.

3.8 MAXIMUM ENTROPY METHOD

The maximum entropy method (MEM) applies a mathematical approach that is strikingly different from most of the other approaches described in this book. It assumes that the true power spectrum can be approximated by an equation that has a power series—but instead of making the spectrum proportional to that power series, it puts the power series in the denominator! For a readable description of why this makes some sense, see Press et al. (1993). It is full of advantages and disadvantages, just the ingredients to inspire passion among spectral analysts. The maximum entropy method finds the spectrum which is closest to white noise (i.e. it has a maximum randomness or "entropy") while still having an autocorrelation function that agrees with the measured values—in the range for which there are measured values. It has a curious feature: because it extrapolates the data to times outside the

measured range, it yields spectral lines that are often much narrower than you would get by any other method. But don't be fooled by this into thinking that the method can resolve spectral lines better than an ordinary periodogram. The narrowness of these lines is one of the mathematical artifacts that can be misleading. It is worthwhile quoting Press et al. (1993) about MEM:

> *"The maximum entropy property has caused MEM to acquire a certain "cult" popularity; one sometimes hears that it gives an intrinsically "better" estimate than is given by other methods. Don't believe it. MEM has the very cute property of being able to fit sharp spectral features, but there is nothing else magical about its power spectrum estimates .... With noisy input functions, if you choose too high an order, you will find spurious peaks galore! Some experts recommend the use of this algorithm in conjunction with more conservative methods, like periodograms, to help choose the correct model order, and to avoid getting too fooled by spurious spectral features."*

Tukey is quoted (by Priestley 1981, p. 606) as saying that the maximum entropy method is suitable only for relatively "smooth" spectra. This makes it highly unsuitable for data with a few narrow lines, as we have in paleoclimate. Recall that Tukey supported the use of the ordinary periodogram for such cases.

Nevertheless, you will see the maximum entropy method used in paleoclimate data, and there will be referees of your papers who will recommend—maybe even demand—that you use it! A program to use it, based on an optional package available in Matlab, is given in Appendix 2. Programs in C, Fortran and Basic are available in Press et al. (1993). We show the spectral power calculated for the SPECMAP $\delta^{18}O$ data, with both a linear and a log plot, in Fig. 3.25. The analysis was done for the time interval 0 to 600 kyr. The maximum entropy parameters were set as follows: the number of coefficients was 250, and the number of frequencies was 800.

For comparison we show in Fig. 3.26 the maximum entropy spectral estimate for eccentricity and orbital inclination. Note how the resolution obtained appears even better than for the periodogram. However, notice that the ratio of spectral peaks is quite different than in the periodogram.

3.9 CROSS SPECTRUM AND COHERENCY

Coherency is a statistical measure that is used to measure the similarity of two records $a(t)$ and $b(t)$ in frequency space. Suppose, for example, $a(t)$ is a climate proxy, and we would like to determine whether it is dominated by a theoretical driving force $b(t)$. We can examine the spectra of the two signals, and see if their amplitudes are similar. This procedure can be made more formal by looking at the following product function of the Fourier transforms:

$$C(f) = A(f)B^*(f)$$

where A is the Fourier transform of a, and B^* is the complex conjugate of the Fourier transform of b. This simple product is given the fancy name "cross spectrum". The cross spectrum of a signal with itself is the spectral power. The cross

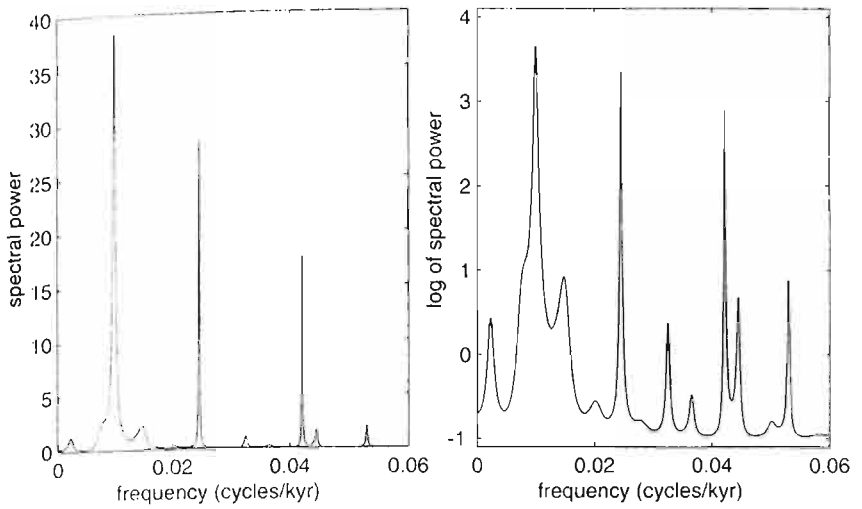


Fig. 3.25. Maximum entropy spectra of SPECMAP.

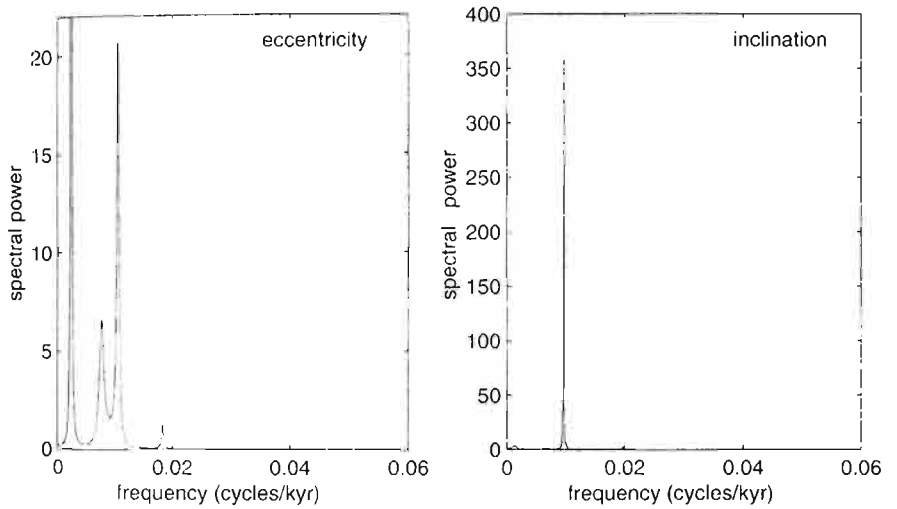


Fig. 3.26. Maximum entropy spectrum of eccentricity and inclination.

spectrum is large only at those frequencies in which there is significant spectral power in both a and b. Some books like to define the cross spectrum as the Fourier transform of the cross correlation function. But, by the correlation theorem, that definition is the same as ours.

If you are particularly interested in whether the two signals are in phase with each other, regardless of amplitude, then we can take the cross spectrum, square it, and

divide by the spectral powers of the individual signals. This gives:

$$c(f) = \frac{|C(f)|^2}{P_a(f)P_b(f)}$$

There is a problem with this definition. It is identically equal to 1 at all frequencies. This is mathematically trivial, but conceptually it happens owing to the fact that a Fourier transform is a sum of perfect sine waves, and a perfect sine wave in signal $a(t)$ that has the same frequency as a sine wave in $b(t)$ will always have the same relative phase. In other words, they differ by a phase factor δ, but this is independent of time. When we take the absolute square, we get $|\exp(i\delta)|^2 = 1$. So measuring the coherency of perfect sine waves makes no sense. Coherency is valuable only if we have two signals that are varying with time, and that implies that they are not single frequencies, but (at the minimum) a band of frequencies. Then it really is interesting to see if they stay in phase, since if one is driving the other, they will.

We build this concept of a band into the definition of coherence by averaging over a frequency band. So the truly useful definition of coherency becomes:

$$\text{coherency}(f) \equiv \gamma^2(f) \equiv \frac{|\langle C(f)\rangle|^2}{\langle P_a(f)\rangle\langle P_b(f)\rangle}$$

By the brackets we mean that adjacent frequencies have been averaged. A typical way to do this is to take two spectra that have N frequencies and average them three at a time, resulting in a coherency that has $N/3$ independent values. If the coherency is calculated by the Blackman–Tukey method, and 1/3 lags are used, then an averaging of adjacent points is automatic. This averaging is essential to make the coherency into something useful.

We should mention that some authors prefer to show γ, the square-root of γ^2. Unfortunately, the function γ is also often called the coherency, so if the authors don't tell you, you may not know which convention they used (unless they use the symbol γ).

Although the coherency is extremely useful in many fields, its value is limited by the need to do the averaging. Resolution is lost, just as spectral resolution is lost in the Blackman–Tukey method. As an illustration of this, we have calculated the coherency of the Earth's eccentricity and orbital inclination for the last 400 kyr in Fig. 3.27. Note that the coherency is extremely high (>0.9) in the vicinity of 0.01, i.e. the 100 kyr peak. Because of the degraded resolution, coherency cannot distinguish the 100 kyr peak of eccentricity from the 100 kyr peak of orbital inclination, any better than the Blackman–Tukey method can.

Coherency measures only the phase relationship; it is insensitive to amplitude, since in the definition we divided by the spectral powers. The process of tuning, described in Chapter 5, adjusts a parameterised sedimentation rate to try to bring the fluctuations of the data into agreement with those of the target. Thus a tuned data set is automatically made coherent with the target model. Since coherency is insensitive to amplitude variations, it does not detect disagreements in amplitude, such as the "Stage-11" problem (see Section 8.2 on page 233). A small number of

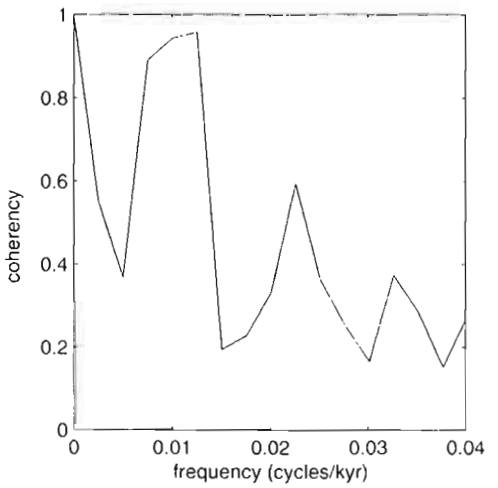


Fig. 3.27. Coherency of eccentricity and inclination.

tuning points (less than one every 100 kyr) can bring the phase of any $\delta^{18}O$ signal into agreement with that of any target (e.g. eccentricity); since amplitudes are ignored by coherency, the two signals will be in apparent complete coherency. The same procedure will yield agreement between $\delta^{18}O$ and orbital inclination. However, orbital inclination and eccentricity cannot be so adjusted, since their time scales are well known. But with one tuning adjustment, we could make an adjusted eccentricity coherent with inclination over 800 kyr! This illustrates the limitation of the use of coherency: it becomes virtually useless—maybe worse than useless (since it represents circular reasoning)—when working with tuned data.

3.10 BISPECTRA

The bispectrum of a data set is the simplest of the statistical quantities known as "higher order" spectra. Bispectra measure an aspect of data which is ignored in the usual spectral estimation (whether it be periodogram, Blackman–Tukey, or multitaper), which is the coherency relationship between several frequencies present in the data. Bispectra have proven invaluable in the analysis of many complex phenomena, both in mechanics and in geophysics.

Briefly, a bispectrum shows a peak whenever two conditions are fulfilled: (1) a triplet of frequencies f_1, f_2, and f_3, are present in the data that obey the relationship $f_1 + f_2 = f_3$, and (2) the phase relationship between the three frequencies is coherent for at least a short averaging time for a band near these frequencies. Such frequency relationships occur frequently in astronomy. They are generated in the nonlinear processes that create the orbital changes for the Earth. The usefulness of the bispectrum is simple: if a driving force (e.g. eccentricity or inclination) has frequency

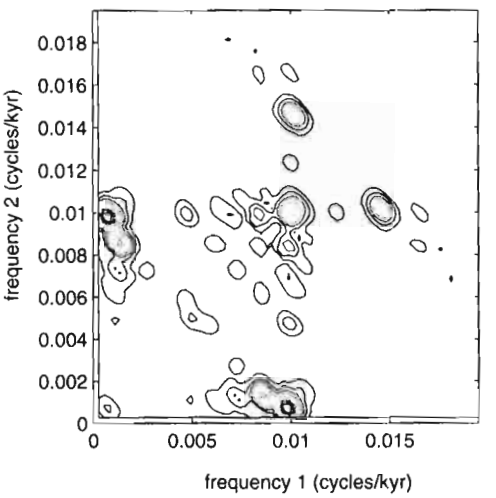


Fig. 3.28. Bispectrum of SPECMAP stack. Note that the figure is symmetric about the diagonal.

triplets which are coherent, then we expect the response (i.e. climate) to contain the same frequency triplets. So, for example, if $\delta^{18}O$ is driven by eccentricity, it should have the eccentricity triplets in it.

We have included an appendix that describes the properties of bispectra in some detail. In this chapter, we will confine ourselves to giving several examples of its use. A plot of the bispectrum has two axes, representing the frequencies f_1 and f_2. We then calculate the amplitude of a coherent signal f_3 in the data, and plot that amplitude at the location (f_1, f_2). Thus, any amplitude on the bispectral plot indicates the presence of a triplet. We make full use of these plots later, e.g. in Section 8.6.

For an example, we plot in Fig. 3.28 the bispectrum for the SPECMAP $\delta^{18}O$ data, using the corrected time scale discussed in Section 5.4. In Fig. 3.29 we show the bispectra for the same time interval of eccentricity and orbital inclination.

The way to interpret these is as follows. Nonlinear processes in a driving force, either eccentricity or inclination, result in coherent relationships between triplets of frequencies. These same relationships are likely to appear in the bispectra of the climate proxy. In addition, nonlinear processes in climate could add additional frequencies to the bispectra of the climate proxy that do not appear in the driving force. In Fig. 3.28, for example, we see several small peaks that are not present in either eccentricity or inclination, for example, at $(f_1, f_2) = (0.01, 0.01)$. This small peak implies the existence in the data of a harmonic at 0.02. This harmonic is not in either driving force, and it should not necessarily be expected to be.

It is clear from the figures that the bispectrum for inclination is a much better match to that of SPECMAP than is the bispectrum of eccentricity. We will return to this result when we discuss the bispectrum of Site 806 in Section 8.6.

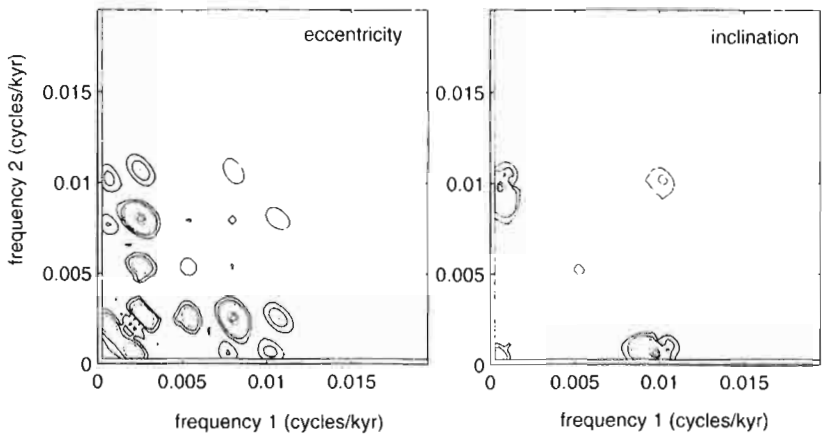


Fig. 3.29. Bispectra of eccentricity and orbital inclination.

3.11 STATISTICS OF THE FOURIER POWER SPECTRUM

The most common errors made in spectral analysis come from poor estimates of the effects of background. By background, we mean all the fluctuations in the data that come from something other than the feature (e.g. the spectral peak) under investigation. Background is not necessarily *noise*. It may be true climate variations, driven by some mechanism other than astronomy. In fact, the background may be similar in structure at different locations around the world.

An extremely important feature of the spectral power is that it does not obey the statistics of a normal distribution. We will show that the statistics, under very general conditions, are exponential rather than Gaussian. What this means is that fluctuations are much more common than would be expected from experience with other data. If the mean of the noise level is M, then the relative probability Prob of a peak fluctuating up to a power p is given by

$$\text{Prob}(p) = e^{-\left(\frac{p}{M}\right)}$$

rather than by the normal (Gaussian) equation

$$\text{Prob}(p) \neq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{p^2}{2\sigma^2}}$$

The reason for this behaviour is straightforward. The Fourier amplitudes (but not the Fourier power) *do* follow normal statistics. This is because the Fourier transform is a linear transform of the data points, which themselves are assumed to obey Gaussian statistics. Both the real and imaginary parts are linear combinations of the data points h_k. By the Central Limit Theorem of statistics we expect the individual Fourier amplitudes to have Gaussian statistics for both their real and imaginary parts. So far, no surprises. If the h_i average to zero, then the expectation value

for the real and imaginary parts of the Fourier amplitudes R and I is zero, and we have the normal equations

$$\text{Prob}(R) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{R^2}{2\sigma^2}}$$

$$\text{Prob}(I) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{I^2}{2\sigma^2}}$$

But the results will be very different for the Fourier power. The probability of obtaining a power p is the same as the probability of obtaining an amplitude A such that $p = A^2$. The probability of obtaining A is the joint probability of obtaining any combination of R and I such that $R^2 + I^2 = A^2$. This consists of all the points that lie in a ring in the complex plane of radius A and thickness dA. Therefore we write:

$$\begin{aligned}\text{Prob}(p)dp &= \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{R^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{I^2}{2\sigma^2}}\right) 2\pi A dA \\ &= \frac{\sqrt{2\pi}}{\sigma} e^{-\frac{A^2}{2\sigma^2}} A dA \\ &= \sqrt{\frac{\pi}{2\sigma^2}} e^{-\frac{p}{2\sigma^2}} dp\end{aligned}$$

This is the desired result. It states that the probability of obtaining a power P is proportional to $\exp(-p/M)$, where M is the mean power ($= 2\sigma^2$), i.e. it shows that the distribution is exponential in character rather than Gaussian.

The exponential rule is extremely important but counter intuitive. It has several implications that can surprise the unwary. These include the fact that large fluctuations are very probable, particularly if you are accustomed to working with data that have a normal distribution. The probability of getting a power equal to three times the mean is $\exp(-3) = 5\%$. Thus, in random data, we expect to see peaks at levels three times the average in 5% of the frequency bins. If there are 20 independent frequencies, then we expect, on average, one of them to show a power greater than three times the mean, just from statistical fluctuations. Thus "large" fluctuations are common in the Fourier power spectrum. This contrasts with normally-distributed data, in which the probability of exceeding three standard deviations is 0.27%.

The rule also implies that the most likely power in any frequency is zero! (That is when $\exp(-P/M)$ has its maximum value.) The mean power, of course, is M. But it is about twice as likely to get bins below the mean as it is to get peaks above the mean. A similar paradox occurs in physics of any process with an exponential distribution, e.g. radioactivity. If we have one atom which has a mean life of τ, then when is the most likely time for it to decay? The answer is, at $t = 0$. That is when $\exp(-t/\tau)$ has its maximum value. This is easier to understand if you take a large number of such atoms. Indeed the maximum decay rate will be at $t = 0$, and it will decay exponentially after that. This result demonstrates that the most likely time to decay is at $t = 0$.

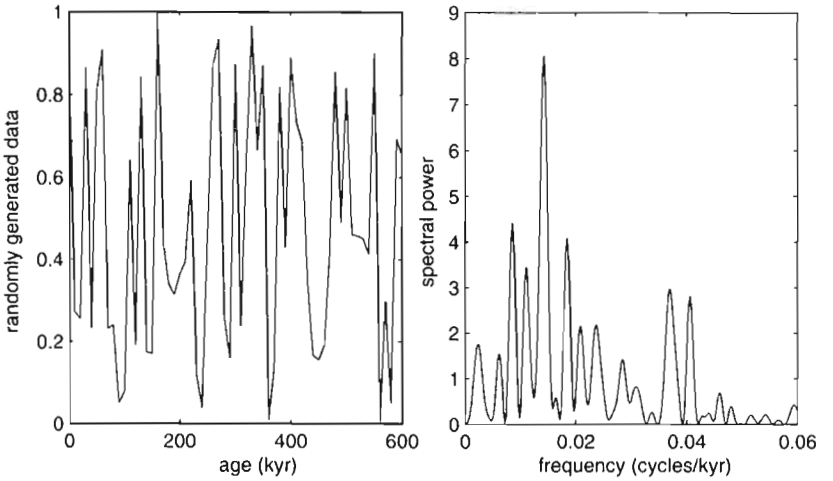


Fig. 3.30. Random signal and its spectrum.

The net effect of these results is that even random data appear, to scientists familiar with normal distributions, to have statistically significant peaks in it. This is best demonstrated with Monte Carlo simulations. In Fig. 3.30, we generated semi-random data by picking random numbers between 0 and 1 to represent the data every 10 kyr. Data for intermediate points (every kyr) was generated by interpolation. The data and spectral power for one such plot is illustrated above. This plot was the fourth one generated.

The apparent structure in the spectrum plot is misleading. It was selected over the first three because it appeared to show a structure that a paleoclimatologist might find interesting. There is a peak (maybe a split peak) near 0.01 and another set near 0.04. Could these be the eccentricity peak and the precession peak? With a little bit of time-scale adjustment, perhaps the data could be made to fit this model. And yet, the data were generated by a random-number generator!

The highest peak in the data has value 8. The spectrum has a mean of 1, so we calculate (incorrectly, it will turn out) that the probability of getting such a peak is $\exp(-8/1) = 3 \times 10^{-4}$. However, there are approximately 60 independent data points, and so 30 independent frequencies in the plot. This was the fourth one generated, so we looked at a total of 120 different frequencies to find this one. So the probability is $120 \exp(-8) = 4\%$. This still seems statistically significant! In fact, one might say "the peak is significant at the 96% confidence level". We publish. But it is nonsense! These are randomly generated data. Where did we go wrong?

The mistake we made is that we underestimated the background. To show this, we generated 1000 different spectra, using the same random procedure that we used to create Fig. 3.30, and then averaged the spectra. The result is shown in Fig. 3.31.

It is clear that the background is not the same at all frequencies. This is an example of a "red" background that we will discuss it at more length in Section

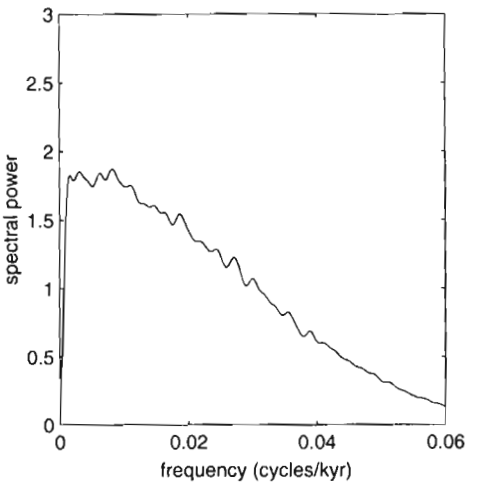


Fig. 3.31. Red spectrum from sum of many randomly-generated backgrounds.

3.13.1. Note that the background is high in the very region that we found the peak we thought was significant. The average spectral power goes to zero at $f = 0$ because we subtracted the mean before doing the Fourier transform.

It is not an accident that what appeared to be the most significant peak occurred in this region. From this plot, you can see that the background level in the vicinity of the peak is approximately 1.8, rather than 1. To estimate the statistical significance of the peak, we need to calculate the ratio of its power to that of the background at the same frequency. So the correct estimate of the probability is $100 \times \exp(-8/1.8) = 1.2 = 120\%$. So we expect to find (on average) 1.2 peaks of similar statistical significance in every four random data sets.

The above example is meant to illustrate how easy it is to overestimate the statistical significance of peaks, unless the background is well understood. It is rarely the case that we have a complete model for the background, so what is the next best approach? If we examine the spectrum again, we see that there is evidence that the background in the vicinity of the peak is unusually high. In fact, if we draw a background line through the data *in the vicinity of the peak*, it is quite plausible that the background level is approximately 2 units. (Remember, according to the exponential rule, about $1/e$ of the data should be above the line, on average.) Had the background level been estimated in this manner, then the mistaken "high confidence level" prematurely estimated for the data would not have been derived.

The only way to be safe is to be conservative. For a peak truly to exist, it must be well above estimates of local background. Of course, if you are conservative, then you might miss some discovery. The result is that the rule has been frequently broken in the literature by scientists whose papers got accepted only because they were reporting a discovery. Another result is that the literature is full of reported frequencies that have turned out not to exist, once better data were available.

3.12 PLOTTING POWER, LOG(POWER) AND AMPLITUDE

When plotting spectral power, the choice of axes becomes important, and controversial. The plot that shows spectral power on a linear scale is very popular, but many statisticians object strongly. They insist that only log plots are acceptable, and anyone who makes a linear plot is simply showing their ignorance of statistics. Many paleoclimate papers contain all their spectra using an ordinate proportional to the square-root of the spectral power. They label the axis "amplitude".

The choice of axis is not just a matter of taste. Each of the plots has its advantages and disadvantages. Each of the plots shows something to its best advantage, and shows something else poorly. To illustrate this, we will use a particular data set known as the SPECMAP stack; we will describe these data more fully in Section 4.2. We show the same spectrum (of SPECMAP data for the period 0–600 kyr) in three plots. Fig. 3.32 shows the spectral power; Fig. 3.33 shows the logarithm (to base e) of the spectral power, and Fig. 3.34 shows the square-root of the spectral power.

The value of the first plot, linear in spectral power, is that it gives an immediate sense of how much of the variance comes from each cycle and how much comes from noise. This is useful since variance has the nice property that amplitude doesn't have: it is additive. By that we mean that the variance of the data is equal to the sum of the variances of the individual components. So, for example, you will often read in the literature statements such as "the 41 kyr peak accounts for 50% of the variance". To the extent that such statements are useful in evaluating signal strength, using an ordinate that is proportional to the spectral power is best. If the data are normalised to unit variance and zero mean, by subtracting the mean and dividing every point by the sample standard deviation, then the area under the peak shows the fraction of the variance contained in any frequency interval. Subtraction of the mean is also useful

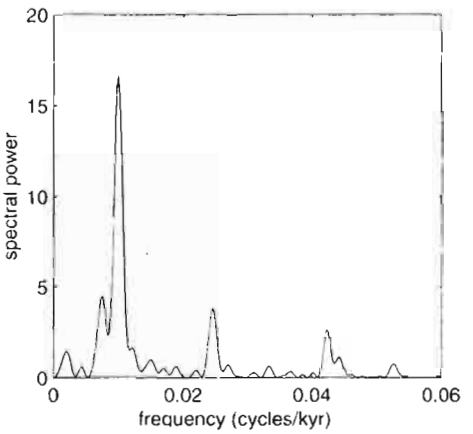


Fig. 3.32. Spectral power for SPECMAP.

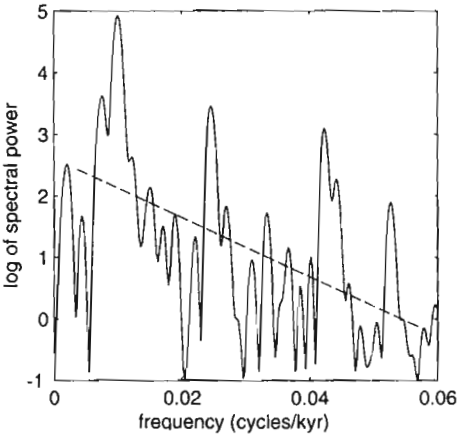


Fig. 3.33. $\log_e$ of spectral power, SPECMAP. The dashed line is an estimate of the background. Roughly one-third of the data points should rise above this line.

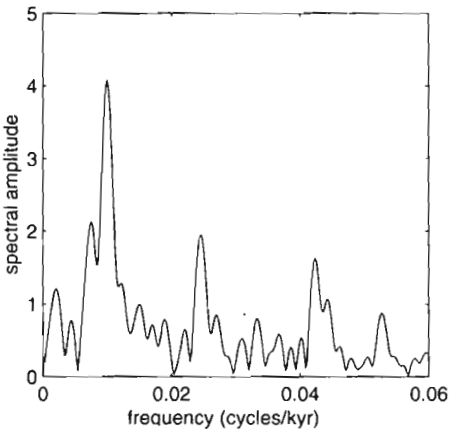


Fig. 3.34. Square-root of spectral power, SPECMAP.

since any line at zero frequency contaminates the estimate of low frequency power. Normalising also aids in comparison of spectra from different sites.

The linear plot also gives the most direct sense of the width of the peaks, since the full-width at half maximum can be seen without even reading the y-axis. The linear plot is the best plot if you are concerned about detecting evidence of an incorrect time scale (due to chatter or drift, which we discuss in Section 5.1). This sensitivity to time scale error is sufficient to justify its use in many cases.

The log plot puts the logarithm of the power on the y-axis. This plot is sometimes called a "semilog" plot, since only the ordinate is logarithmic, and not the abscissa. This kind of plot is the best one for making the statistical significance of peaks evident, particularly for small peaks. As we showed in Section 3.11, the statistical

significance of a peak of power P over a background B is $\exp(-P/B)$. In paleoclimate data the background B is often different at different frequencies. In the log plot, peaks of equal statistical significance will be the same linear distance above the local background level. That's why the log plot is favoured by statisticians whose primary interest is evaluating the statistical significance of small peaks. However, a log plot makes it very difficult to see time-scale error, since the multiple peaks characteristic of that kind of error are hidden among a welter of background whose presence is emphasised in the log plot.

In addition, the log plot can easily give a misimpression of the *climate* significance of each peak. A peak can be statistically significant simply because the background level in its region is small. But that doesn't mean that the peak plays an important role in climate. (This distinction is often confused in the public eye. A release of radiation that is statistically significant can be reported to the public as a "significant release", and they will misinterpret it.)

Some people believe that the best ordinate to give a sense of climate significance is the square-root plot. The square-root of the spectral power is the magnitude of the spectral amplitude, and the square-root plot is often labelled "amplitude". The amplitude tells you the peak-to-peak variation in the climate proxy. Unlike the spectral power, however, the sum of the spectral amplitudes does not equal the total amplitude variation in the signal.

In the amplitude plot, Fig. 3.34, it is relatively easy to see that the amplitude of the 41 kyr peak is two to three times less than the amplitude of the 100 kyr peak. (The uncertainty of 2 to 3 arises because of the uncertainty in the placement of the background level.) One disadvantage of the amplitude plot is that the statistical significance of the peaks is hidden. To calculate it you must find R the ratio of the peak height to the background level, square it, and then calculate $\exp(-R^2)$. This is not hard to do—but it is not as evident on the plot itself. All three of these plots have been popular in the paleoclimate literature; often two kinds are given in the same paper.

3.13 MONTE CARLO SIMULATION OF BACKGROUNDS

Monte Carlo simulations have proved to be indispensable in physics research. Typically, a simulation consists of the computer generation of signal buried in computer-generated noise. The combination is then studied as if it were real, to see whether the numerical methods applied to the data will pull true signal out of noise. Such tests are frequently done before an experiment is even built, to see if it can achieve the sensitivity that is needed.

A typical application is as follows. A student has performed a spectral analysis of $\delta^{18}O$ data, and notices what appears to be structure near the base of the peak, a "shoulder". Does this indicate that the data are not properly tuned, i.e. that the time scale is not correct, so some spectral power is being leaked to adjacent frequencies? Or does it indicate that the peak has real structure? Or is it just a statistical fluctuation? One way to address this question is to run 100 or more Monte Carlo simula-

tions. For each simulation, the computer generates a real signal that is a pure sine wave, with no shoulder, and a "random background signal" that is added to the sine wave. The spectral power is calculated for each Monte Carlo simulation, and the student looks for a shoulder. If they see one in, say, 3 out of the 100 simulations, then they know that random events can simulate the effect, but not very often. They might properly conclude that the shoulder is not a random event but real, with a "97% confidence level" since it did not occur in 97% of the simulations. But could it be due to a poor time scale? To do this test, another 100 simulations are run, but for each one the time axis is now changed in a small way that simulates what might happen with real $\delta^{18}O$ data. Again, the student studies the simulated events to see how often a shoulder appears.

The key technique in Monte Carlo simulation is the creation of background that behaves in a way similar to the background you have in the real data. If the background in the simulation is not an accurate mirror of the background in the data, then the conclusions drawn may be wrong. In this section we will discuss several ways of generating background, and we give Matlab programs for these methods in Appendix 2.

We have already presented one example of Monte Carlo simulation for Fig. 3.31. We will repeat the procedure here. We create our background by generating random numbers every 10 kyr, and doing a linear interpolation in between. This yields data and spectra shown in Fig. 3.35. The spectrum plot shows the spectrum generated from the data shown. Superimposed on the spectrum plot is the average spectral power found by repeating the entire process 100 times, and averaging the results.

As you can see in the spectral power plot, there appears to be a strong peak near frequency 0.017 cycle/kyr. Yet we know that these data were generated from random numbers, so we know this peak is not real. It is an result of the fact that spectral power shows statistical fluctuations which are exponential in their nature, and so large excursions are more common than in other types of analysis.

We now estimate the statistical significance of this peak. The mean value of the background in the vicinity of the peak is at spectral power = 2. The peak rises to a value of 10. Therefore the probability of having it occur by chance, at this frequency, is approximately Prob = $\exp(-10/2) = 0.0067$. There are about 45 frequency bins in

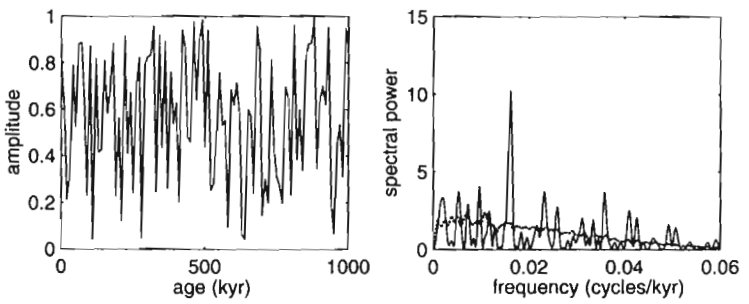


Fig. 3.35. Monte Carlo background and spectrum.

the plot, and we would have been equally interested in any other peak that was similarly high above the background. Thus, the probability that we would find something of similar height is $0.0067 \times 45 = 0.30 = 30\%$. So in 30% of our Monte Carlo simulations, we expect to see peaks with comparable significance—even in data that we *know* contain no real peaks.

Fluctuations such as seen in Fig. 3.35 are common, yet even experienced scientists can be fooled. We strongly urge every student of spectral analysis to spend a considerable amount of time generating Monte Carlo plots, before publishing anything from real data. There is no technique that is equally effective in showing how easily we can fool ourselves. And the primary difference between a scientist and a layperson is that the scientist is aware of how easily he can be fooled.

3.13.1 Red and white noise

In both of the examples given so far, Figs 3.31 and 3.35, the average spectrum tended to peak near the low frequencies. This is an example of "red noise". The name derives from optical spectra, where the red signal is the low frequency component. Many phenomena in physics show red noise. In many cases, the power spectrum of the background shows a strong enhancement at the low frequency end proportional to $1/f$. Such noise is called (without showing much originality) "one over f noise", or sometimes "flicker noise". Noise in paleoclimate work is frequently red. This may be because processes in the climate and in the ocean records tend to suppress rapid variations, rather than because low frequencies are enhanced.

In Fig. 3.36 we have generated another simulation of background, using a slightly different technique. It was created by generating a new random number for every thousand years, and then taking a 15 kyr running average of the results. This is the kind of background we would have, for example, if climate were random but the sea floor sediment was continually mixed to a depth of a few centimetres.

The plot on the left is an example of the data generated in this method. The plot on the right is the spectral power for this data set, and superimposed on it (as before) is a plot of the average spectral power from 100 such simulations. The choice of smoothing period allows us to adjust the redness of the spectrum. For example, if we smooth the same random numbers we used above over 10 kyr, rather than 15, we get

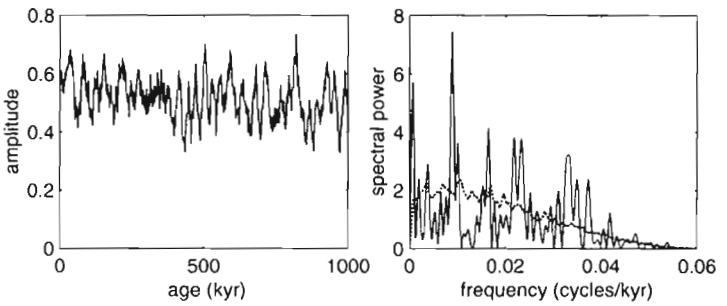


Fig. 3.36. Background from 15 kyr smoothing, and spectrum.

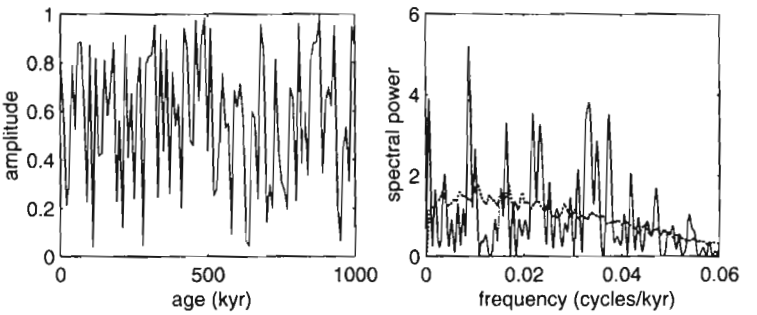


Fig. 3.37. Background from 10 kyr smoothing, and spectrum.

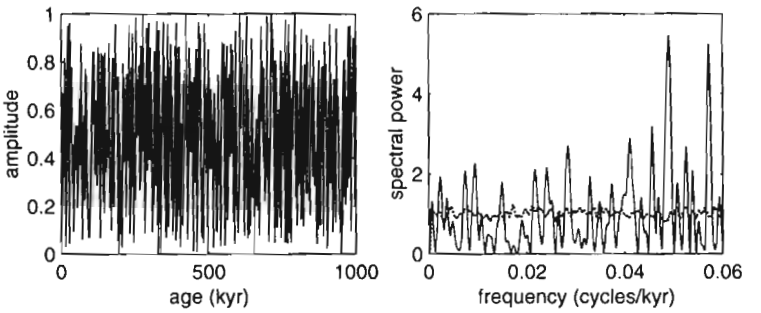


Fig. 3.38. White background, and spectrum.

the data and spectrum shown in Fig. 3.37. Note that the spectrum is less red, i.e. its average value does not fall off as quickly at high frequencies.

For contrast, we generate data with no smoothing. Let every data point, every 1 kyr, be random. The data and spectrum are shown in Fig. 3.38. Note the flatness of the average spectrum. This is the kind of signal known as white noise, since all frequencies are equally present. (This is the terminology, even though for the colour "white" in optics, all that is demanded is equal contributions of red, blue and green, the three colours of the eye sensitivity.)

A small spectral peak, appearing on top of a red background, will have the frequency at which it achieves maximum power shifted slightly towards lower frequencies. This is an example of "bias". It is particularly troublesome if the goal of your analysis is the precise determination of frequencies. To minimise the bias, you can use a method called "prewhitening" that tends to flatten the background. This method is discussed at length in the standard books on spectral analysis. Once the spectrum is computed, you must then compensate for the prewhitening by adjusting the measured spectral powers.

3.13.2 Prewhitening

Prewhitening can be used to compensate for red noise, or for the presence of an

adjacent large spectral peak. The techniques are quite different. For red noise, prewhitening is sometimes done by doing the spectral analysis not on the data, but with a new set of points equal to the differences of adjacent data points. This is equivalent to the process of taking a derivative of the data. We will show below that taking the derivative is equivalent to multiplying the Fourier transform by a factor of frequency f. If the Fourier transform is defined as

$$H(f) \equiv \int_{-\infty}^{\infty} h(t) e^{2\pi i f t} dt$$

the Fourier transform of the derivative of the data is

$$\begin{aligned} FT\left(\frac{\partial h(t)}{\partial t}\right) &= \int_{-\infty}^{\infty} \left(\frac{\partial h(t)}{\partial t}\right) e^{2\pi i f t} dt \\ &= -\int_{-\infty}^{\infty} h(t) \left(\frac{\partial}{\partial t} e^{2\pi i f t}\right) dt \\ &= -2\pi i f \int_{-\infty}^{\infty} h(t) e^{2\pi i f t} dt \\ &= -2\pi i f H(f) \end{aligned}$$

where we integrated by parts to shift the derivative to the exponential term. Thus taking differences multiplies the Fourier transform by a term proportional to the frequency f. The spectral power is multiplied by f^2. Taking differences often overcompensates for the observed red noise, which is typically proportional to $1/f$. Blackman and Tukey (1958) recommend that you try milder prewhitening, a partial differencing: replace the data h not by Δh (equivalent to taking the derivative) but instead by $(h - 0.6\Delta h)$. But they emphasise that you must look at the effect that such prewhitening has on your data, and then change it to do a better job at flattening the spectrum. There is no single prescription that works for all data.

Given the high signal to background levels seen in paleoclimate data, and the uncertainty in the time scale, there is little to be gained by the prewhitening methods discussed above, and most scientists no longer use them for such data. It is usually sufficient to subtract the mean from the data (thus assuring that the zero frequency bin, which represents the constant offset, will be zero), and sometimes to remove the "trend", i.e. the best fit to a constant slope. Removing the mean is common, but many experts prefer not to remove the trend since doing so may bias both the power and the frequency estimates.

Another form of prewhitening that is valuable is the systematic removal of strong peaks, as we discussed in Section 3.7. It consists of systematically finding the peaks, and removing them one at a time, before proceeding with further analysis. When the peaks have been removed, the bias that they might have caused is eliminated. It is an important technique, not so much for reducing bias, as it is for discovery of weak peaks on the tails of stronger peaks, and for the estimate of the statistical significance of such weak peaks.

3.14 FREQUENCY ESTIMATION: THE HWHM RULE

The best estimate for the frequency of a spectral peak can be obtained by using the maximum likelihood method: make a preliminary estimate of the frequency of the strongest peak. Using the parameters of this peak as a starting point, perform a maximum-likelihood fit (or a minimum chi-square fit) to the data. In this fit, the frequency as the starting point should be a variable. Unfortunately, this method is cumbersome, and rarely worth the effort. It is much simpler, and usually accurate enough, to pad the data with lots of zeros, and then find the frequency from the maximum of the periodogram.

But how accurate is this estimate? The primary error often comes from the inherent bias from noise and, indeed, the elimination of such bias forms the bulk of many texts on spectral analysis. Many methods have been developed to reduce such bias, but most of these methods depend on the presence of well-characterised noise. And lack of bias does ensure accuracy. The noise itself is a true climate variation often reflecting local or even global conditions. So, for example, the benthic $\delta^{18}O$ signal may reflect global ice, but it also includes a component reflecting local climate variations such as systematic temperature and salinity variations (as you would get, for example, if the deep thermohaline circulation changed).

The net result of this complexity is that you are often reduced to estimating the effects of such bias, rather than attempting to compensate for them. The uncertainty should be reflected in the error estimate for the frequency. For much paleoclimate data the result is relatively simple: the uncertainty in the frequency estimate will be approximately equal to $\pm w$, where w is the half-width at half-maximum of the peak in the spectrum. Thus the frequency is determined only over a range of the full-width at half-max of the peak. We call this the "HWHM rule". Obviously, this is not a precise result that works in all cases. But we will show below why it is a reasonable approximation for much paleoclimate data. More importantly, it is rare that one can determine the frequency to an accuracy better than this. You will see papers in which the period of the peak is stated to be 100 kyr, and the authors will claim this is incompatible with a theoretical prediction that the period should be 95 kyr. But although the peak of the spectrum may be at 100 kyr, that does not mean that it is an accurate determination of the true frequency, since background variations can shift the peak.

The uncertainty in frequency arises from the effect of background on the peak. Let us assume that the amplitude at maximum, in absence of background, is A_M. There is, of course, a phase associated with the amplitude. Typically the phase will vary as rapidly as the amplitude, so that as we go away from the maximum value to the half power point the phase will change by 45 degrees. The background also varies in amplitude and phase over a similar change in frequencies.

To get an idea of the relative values of the signal and background, we show in Fig. 3.39 the $\delta^{18}O$ spectrum for Site 552 (Shackleton and Hall, 1984). The peak near $f = 0.01$ cycles/kyr stands well above the background. Let us try to estimate the background. It appears to be higher at low frequencies than at high ones, and the high frequencies may include some non-background signals (e.g. the peak at 41 kyr

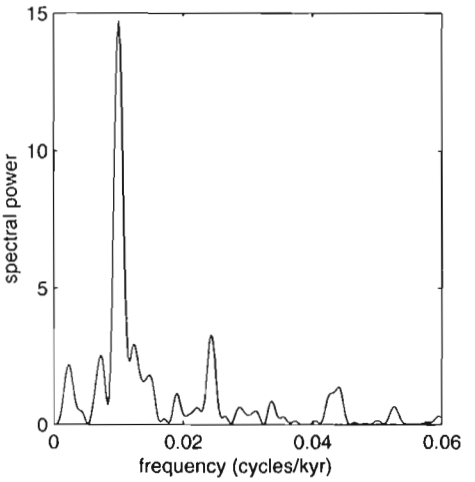


Fig. 3.39. Site 552 spectrum.

period). There does appear to be structure on either side of the 100 kyr peak, at a level of about 2 units. (The plot is normalised so that the average power is one unit.) For the sake of discussion, let us accept this level as the background estimate. The power of the peak is at 14 units. So the ratio of amplitudes of S/B is approximately $\sqrt{14/2} = 2.6$.

The question at hand is: how much does the presence of B affect our estimate of S, in particular, our estimate of the frequency at which S reaches its maximum? We will estimate this by looking at the effect of changing the background, while keeping the signal unchanged. (This can easily be done with Monte Carlo simulations, and the reader is encouraged to try it.) Suppose, for example, that the amplitude S_M of $\sqrt{14} = 3.7$ were partly due to constructive interference. To get its true value, subtract the amplitude of the background $B = \sqrt{2} = 1.4$, to get $S_{\mathrm{TRUE}} = 2.3$. If, instead of constructive interference, the background B happened to give destructive interference, then the amplitude at this point would have been $2.3 - 1.4 = 0.9$. The peak would not have been visible at all at this frequency. Back at the present half-power points, the relationship between the background and the signal could differ by 90 degrees, so the amplitude at this frequency would be $\sqrt{(2.32^2 + 1.42^2)} = 2.7$, and the power at this frequency would be $2.7^2 = 7.3$, well above the noise. The peak would be seen, but far removed from the correct frequency.

Anyone who wants to put uncertainty limits on a frequency estimate will learn a great deal by doing Monte Carlo simulations. However it is clear that a peak cannot be moved much further than its half maximum. In addition, it is clear that unless the signal to background level is unusually high, that a movement of that much in frequency has a reasonably high probability of happening. A more sophisticated analysis requires a better characterisation of the background.

To illustrate the usefulness, and limitations, of the HWHM rule, we generated 40.000 Monte Carlo data sets, each containing a true 100 kyr peak, but with varying

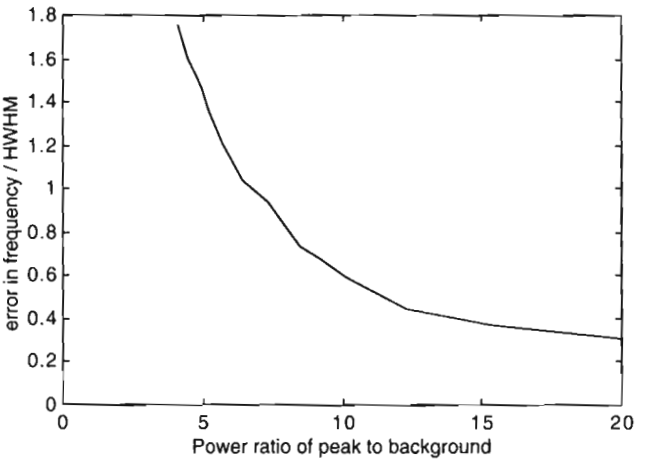


Fig. 3.40. Frequency determination vs. signal-to-background.

levels of random background. (The background was generated by choosing random numbers every 10 kyr, and interpolating in between.) For each data set, the periodogram was calculated, and the frequency of the maximum power (with period between 80 and 120 kyr) was recorded. The average error was the frequency difference between this value and the true value (0.01). In Fig. 3.40 we show the error, normalised to the half width at half maximum, plotted as a function of background level.

The plot demonstrates that at very high signal/background levels (20), the frequency can be determined to within about 30% of the half width of the peak. For more typical data (signal/background = 7), the accuracy is equal to the half width. When the signal/background drops to 5 or below, the frequency determination becomes very inaccurate.

3.15 SPECTRAL WIDTH AND SHAPE DETERMINATION

Suppose we have a climate proxy $h(t)$ which consists of a pure sine wave present for a time T. Moreover, assume that $h(t)$ is zero outside of this time. Then the untapered periodogram will show a peak proportional to the square of a sinc function:

$$\left(\frac{\sin(2\pi(f-f_0)T)}{2\pi(f-f_0)T}\right)^2$$

If the observed spectral peak has this shape, then nothing intrinsic to the shape is observable, and we say the peak is *unresolved*. However, if the signal is the sum of two or more sine waves, or has an amplitude that varies with time, then the shape can become complicated. The case of two sine waves is particularly interesting, since such *doublets* are common in orbital parameters. The other important case for

paleoclimate work is the effect of background on our ability to detect and measure such a doublet.

If two spectral lines are sufficiently close in frequency, then they will appear to be a single peak with a somewhat broader shape. How close can they be? This issue has received a great deal of study and analysis; see, for example, a lengthy section in Priestley (1981). The limit is set, in part, by the shape of the spectral peak, and that, in turn, depends on the method of analysis. The most widely used equation is the "Rayleigh Criterion", originally applied to images in a telescope. Rayleigh decided that two stars of equal brightness could be "resolved" if the maximum intensity of one star fell on the first zero (minimum intensity) of the other. However, in optics the telescope image of a star is equal to the square of the Fourier transform of the aperture—i.e. it is the spectral power! So the Rayleigh criterion can easily be used in spectral analysis. For a simple sine wave, the first zero in the spectral power occurs at

$$\Delta f = \frac{1}{T}$$

This looks very similar to the uncertainty principle criterion—with good reason, since the width of the spectral lines determines the resolution. We will illustrate this equation with an example. The eccentricity of the Earth's orbit contains oscillations with period 95 and 125 kyr. Let us consider a simple function consisting of two sine waves with these periods:

$$h(t) = \sin(2\pi t/95) + \sin(2\pi t/125)$$

They should be "Rayleigh resolved" when

$$T = \frac{1}{\Delta f} = \frac{1}{\frac{1}{95} - \frac{1}{125}} = 400$$

To illustrate this, we show in Fig. 3.41 the spectral power for $h(t)$ for the time interval $t = 0$ to 400 kyr.

Several features of this plot are worth noting. The first is that the two peaks appear to be "super resolved"—the spectral power in between the two peaks goes to zero. This is much better than Rayleigh required. The reason is that the two signals, at frequencies 0.008 and 0.0105 cycles/kyr, interfere with each other to

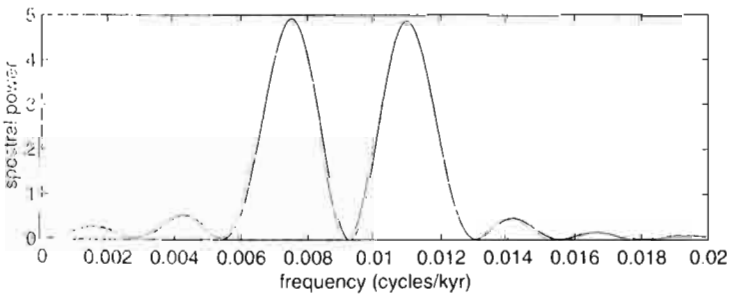


Fig. 3.41. Rayleigh Criterion example: 95/125 kyr peaks.

give the cancellation. A similar feature does not occur with stars, because stars are *incoherent*, and do not interfere with each other. But the super resolution is not always achieved. Since it requires a cancellation, it depends on the relative phases of the two signals. If we add a 90 degree phase shift to one of the signals, the dip between the two peaks is not nearly as deep.

A second important feature is that the maxima of the individual peaks do not occur at the expected frequencies. The peak on the left actually reaches its maximum value at $f = 0.0075 = 1/133$, i.e. it corresponds to period 133 kyr, not at 125 kyr. This disagreement is a consequence of the fact that this peak is sitting on the side-lobes of the other peak. This tilts the peak and, in this case, moves the maximum to a lower frequency (higher period). The degradation for close peaks is known as the Cramèr-Rao bound (Rife and Boorstyn, 1974, 1976; van Trees, 1968). In general, the estimate of the frequency degrades when the frequencies become closer than $2/T$, and that is the effect we are seeing here. However, if precise frequency estimation is important (and only if low levels of background suggest it is possible; see Section 3.14) it is easy to avoid the Cramèr-Rao restriction by doing the frequency estimation using the pre-whitening approach advocated by MacDonald (see Section 3.7). In paleoclimate data, the Cramèr-Rao limit is rarely important, since the uncertainty (the half-width at half-maximum) will be much larger than the bias.

In Fig. 3.42–Fig. 3.44, we plot the separation of the two peaks for various time intervals T. For $T = 250$ kyr (Fig. 3.42), the peaks are unresolved; however it is clear

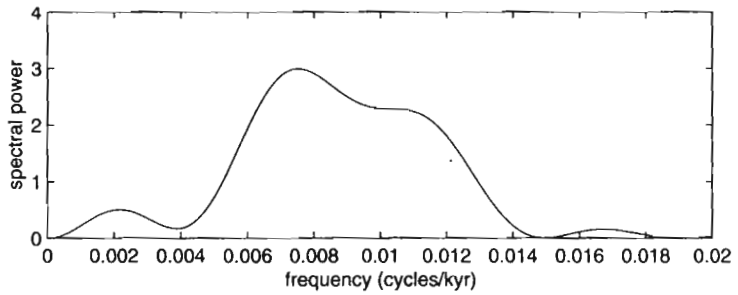


Fig. 3.42. 95/125 kyr peaks with interval $T = 250$ kyr.

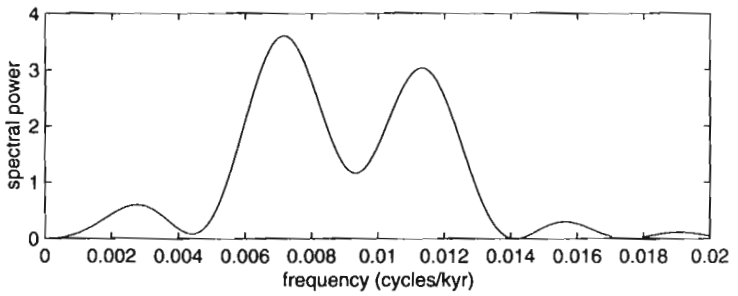


Fig. 3.43. 95/125 kyr peaks with interval $T = 300$ kyr.

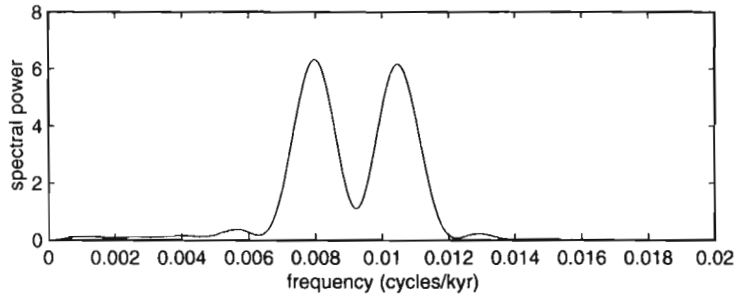


Fig. 3.44. 95/125 kyr peaks with interval $T = 600$ kyr.

that it is broad and is more than a single frequency. For $T = 300$ kyr (Fig. 3.43), the peaks are not separated, yet are resolved.

The spectral power for $T = 600$ kyr is shown in Fig. 3.44. Note that the peaks are not as well separated as they were for $T = 400$ kyr. With this time interval, the bias in the frequency maxima is much smaller, but still present. The low frequency peak reaches its maximum at $0.00796 = 1/125.5$, rather than at $0.08 = 1/125$. This Cramèr-Rao bias could be serious in the communications industry, but it is far too small to be important in paleoclimate work.

Next we consider the effects of background on shape determination. Such effects can be substantial, since background fluctuations can be large. If the background level is B, then the probability that it reaches a level C for any given data point is $\exp(-C/B)$. (This is an immediate consequence of the fact that the background has an exponential distribution, as discussed in Section 3.11.) This effect of background is best illustrated by a numerical example, similar to the one we gave when discussing frequency estimation. Let the spectral power of a peak, in the absence of background, be P. Let the average spectral power of the background, in the vicinity of the peak, be A. Let us assume that $P = 3B$. That seems like a reasonable ratio; it means that the probability that a random background peak could reach as high as P is $\exp(-3) = 0.05$, corresponding to a 95% confidence level. But although it is not terribly likely that the background will fluctuate that high, it is much more likely that the background will cancel the peak into insignificance. How can that be? Let us assume that the actual background in the vicinity of the peak has value B, but that it also happens to have the opposite phase from P. (This will happen about 25% of the time.) The Fourier amplitude of the peak is $\sqrt{P}$. The Fourier amplitude of the background is $\sqrt{B}$. The net amplitude is $\sqrt{P} - \sqrt{B} = (\sqrt{3} - 1)\sqrt{P} = 0.732\sqrt{P}$, and the observed power will be $(0.732\sqrt{P})^2 = 0.54P$. Recall that by itself, P was three times background. This partially cancelled peak will only be 1.6 times background. The experimenter will determine that its statistical significance is only $\exp(-1.6) = 0.2$. That places it at the 80% confidence level.

The lesson here is that background can be more bothersome in its tendency to *cancel* real peaks, than in its ability to fluctuate up and cause false ones. This is particularly important in the case of theories that attribute the 100 kyr peak to eccentricity. The eccentricity model attributes the 100 kyr peak to the 95 kyr cycle

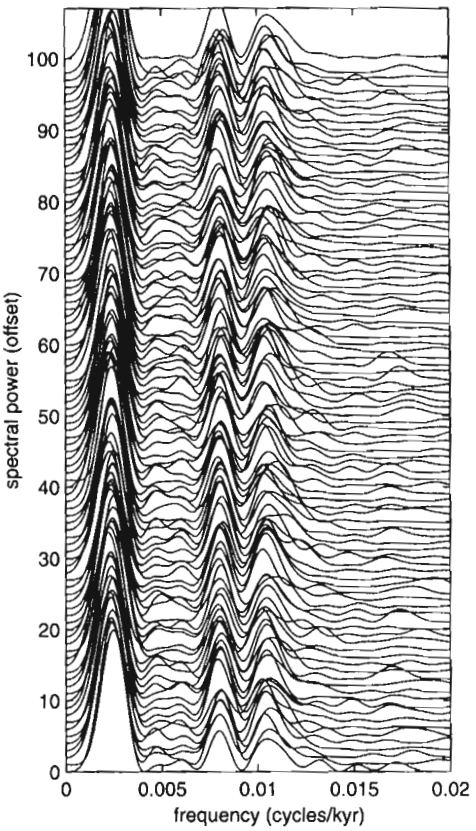


Fig. 3.45. Monte Carlo simulations: eccentricity + background.

of eccentricity, and the statistics are usually compatible with this identification (since you can rarely measure the frequency of a peak in paleoclimate data to better than a full-width at half-maximum). But the 125 kyr peak is missing. Could it have been accidentally cancelled by background? The answer is yes—if the background level is as high as one-third of the peak level.

However, if the background is significantly lower than one-third of the peak level, such cancellation becomes increasingly improbable. In Fig. 3.45 we show 100 Monte Carlo simulations of eccentricity added to background, in which the *average* level of background was ten times lower than the height of the expected 125 kyr peak. In the 100 spectra shown (each one with a different random background), the doublet nature of the peak is not significantly cancelled. This is not surprising, since to cancel would require both that the phase of the background be opposite to that of the peak (a 25% chance) and that the background spectral power fluctuate upwards to a significant fraction of the peak power.