

# Quadratic-inverse spectrum estimates: applications to palaeoclimatology

BY DAVID J. THOMSON

*AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill,  
New Jersey 07974, U.S.A.*

This paper describes some new methods for the analysis of time series and their application to find new results in palaeoclimate. The new statistical theory includes a quadratic inverse theory for unbiased estimation of power spectra, an associated test for spectral resolution, maximum-likelihood spectrum estimates, and detailed explanations of some topics in the detection and estimation of periodic components. A new technique for estimating transfer functions is described. This methodology is used to analyse series describing global ice volume over the past 700 000 years as recorded by proxy oxygen isotope ratios from deep-sea cores. We find many of the periodic components predicted by the Milankovitch theory. However, systematic departures are found from the predicted frequencies. These are accompanied by phase modulation that can be attributed to changes in the precession constant of the Earth caused by glaciation-induced changes in the Earth's principal moments. An estimate of the transfer function from ice volume to precession implies that the Earth's crust requires more than 160 000 years to compensate for mass redistribution, and overcompensates with a delay of about 24 000 years.

## Introduction

The analysis of time-series describing climate and the interrelations between such series provides difficult tests of statistical methods. In this paper I outline a quadratic inverse theory extension of the multiple-window method and illustrate some newer multiple-window techniques using palaeoclimate as an example. The major objective in the development of multiple-window statistical techniques has been to find a set of frequency-domain estimation procedures where all the critical assumptions are testable in a consistent way. Such tests should also be compatible with current scientific practice and thus, for example, not require replication but make efficient use of it when available. Currently the most powerful tools for understanding such problems are quadratic inverse theory, to be described in §4, and the method of jackknifing over windows described in Thomson & Chave (1990).

As an example, I apply these techniques to  $\delta^{18}\text{O}$  isotope records from deep-sea cores and to discrepancies between the results of this analysis and 'standard' theory. First, many of the astronomical line components predicted by the Milankovitch theory are detected embedded in a continuous background spectrum. The harmonic  $F$  test, described in §§5 and 6, is used for detection, with the significance level of the test exceeding 99.9% on the 41 000 year obliquity line. Similarly, a multiple-line test (§6) shows that a minimum of four of the predicted precession components are required for an adequate fit to the data in the precession band. However, the estimated

frequencies of these lines show both systematic biases from the frequencies predicted by celestial mechanics and phase modulation, depending on the source of the line. In particular, spectral lines originating from obliquity have a different frequency shift from those caused by precession. We propose here that the precessional frequency shifts are a consequence of a time-varying precessional constant resulting from changes in the Earth's principal moments of inertia caused by glacial mass redistribution. The shifts in the obliquity lines, on the other hand, are a consequence of systematic errors in the timescale resulting from 'tuning' in an attempt to match the strongest obliquity and precession lines simultaneously. The estimated transfer function between the differential phases on obliquity and precession and the SPECMAP  $\delta^{18}\text{O}$  stack (Imbrie *et al.* 1984) has a zero at a period of *ca.* 160000 years, with compensation at longer periods and an impulse response with maximum overshoot at a delay of *ca.* 24000 years. One implication is that the average difference in moments of inertia of the Earth is closer to hydrostatic than current values.

Because these results were obtained using a new method of estimating power spectra and related quantities we begin this paper with a brief description of conventional spectrum estimation procedures. This is followed with a 'cookbook' description of the multiple-window method of spectrum estimation for continuous spectra (details are contained in Thomson (1982)). Section 3 describes quadratic inverse theory with the maximum-likelihood estimates following. Details of the line estimation process occupy §5, problems with multiple lines §6, and transfer function estimation is outlined in §7. Section 8 describes the data in general. In §9 a single time series, the Southern Atlantic core V22-174 is used to demonstrate the single line *F*-test, and analysis of the residuals. Section 10 illustrates the multiple-line test and §11 the differential precession calculations, and estimation of the transfer function using the SPECMAP stack.

## 1. Historical spectrum estimates

In a new article on controversies in climate studies, R. A. Kerr (1984) remarked, 'A large part of the problem is that everybody uses different mathematical techniques to extract the subtle periodic signals from the random climatic noise that nearly overwhelms them in the record.' Compounding this, many of these statistical methods are known by several different names as in Schonwiese (1987). Currently the terms 'chaos' and 'fractals' are too often invoked as fashionable explanations for complicated, but stationary, behaviour. Because the statistical method used in this paper is still not well known I begin with a short comparison of the common methods of spectrum estimation with emphasis on their differences and shortcomings. Extensive reviews are given in Lindberg (1986) and Cox (1981). This is followed by a brief description of the multiple-window method used in this paper.

Initially I assume that the available data consists of a single series of  $N$  contiguous observations spaced at unit time intervals,  $x_0, x_1, \dots, x_{N-1}$  and that it is a representative random sample (under the usual probabilistic assumptions) of a wide sense stationary process. It is initially assumed that the observed process is zero-mean and also without periodic or other deterministic components, but methods described in later sections allow the presence of these. Theoretical variance calculations assume that the process in the frequency domain is gaussian. Similarly, in theoretical developments, I also assume that the data is 'clean', i.e. free of outliers or bad data points. However, as error-free data are extremely unlikely in practice

this assumption is not made in the examples, and robust methods are used. Denote the spectrum of the process by  $S(f)$ , where  $f$  is the frequency in cycles per time unit. Frequency is assumed to be a continuous variable on the principal domain  $-\frac{1}{2} \leq f < \frac{1}{2}$  but in convolutions and similar operations is extended periodically. In this work the total, or Nyquist, frequency band  $[-\frac{1}{2}, \frac{1}{2}]$  is split into two domains: a band of width  $2W$ ,  $0 < W \leq \frac{1}{4}$ , centred on a particular frequency  $f$ ,  $(f-W, f+W)$  referred to as *inner* or *local* or *interior* domain; and the rest of the frequency domain, referred to as *remote*, *broadband*, or *exterior* domain. Letting  $f$  denote the centre frequency of a frequency band  $(f-W, f+W)$  and  $\xi$  denote an incremental frequency restricted to  $|\xi| < W$ , the symbols ‘ $\oplus$ ’ and ‘ $\ominus$ ’ will be used to denote addition and subtraction with the second operand restricted to the inner domain; for example  $f \oplus \xi$  stands for  $f + \xi$  with the understanding that  $|\xi| < W$ . Complex conjugates are denoted by either an overbar or by a superscript  $*$ , ‘real’ data have no imaginary part, and ‘complex’ is not used as a synonym for ‘complicated’. Bold-face letters denote column vectors or matrices with matrix transpose denoted by superscript ‘T’ and conjugate transpose by a superscript ‘ $\dagger$ ’. Open-face letters are used for; the expected value operator  $\mathbb{E}$ , the Dirichlet kernel  $\mathbb{D}_N(f)$ , and the inner band projection operator  $\mathbb{P}_W(f, g)$ , with the last two defined in Appendix A. In this paper I use the standard definition of the discrete Fourier transform of the sequence  $x_n$ ,

$$y(f) = \sum_{n=0}^{N-1} x_n e^{-i2\pi n f} \quad (1)$$

with the corresponding inverse transform

$$x_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} y(f) e^{+i2\pi n f} df. \quad (2)$$

In the examples, where the time interval, or time between successive data samples, is  $\delta t$ , the Nyquist, or folding, frequency is  $1/(2\delta t)$ , the Rayleigh resolution frequency is  $1/(N\delta t)$ . Fast Fourier transform (FFT) algorithms are used to do the arithmetic; as these work most efficiently when the number of points transformed,  $M$ , factors into small primes one takes  $M > N$ , and preferably  $M \geq 2N$ , and ‘zero pads’. This is equivalent to defining  $x_N, \dots, x_{M-1}$  to be zero and letting the sum in (1) run from  $n = 0$  to  $M-1$ . Standard FFT algorithms compute the transform (1) on a frequency mesh spaced  $1/(M\delta t)$ .

Stationary stochastic processes have a spectral, or Cramér (1940) representation

$$x_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi n \nu} dX(\nu) \quad (3)$$

for all  $n$  where  $dX(\cdot)$  is an orthogonal increment process (Koopmans 1974). This representation is valid for any wide-sense stationary process and, of great importance here, its second moment

$$\mathbb{E}\{|dX(f)|^2\} = S(f) df \quad (4)$$

is, by definition, the *spectral density function*  $S(f)$  of the process. In addition  $dX$  is uncorrelated at different frequencies and

$$E\{dX(f) \overline{dX(g)}\} = S(f) \delta(f-g) df dg, \quad (5)$$

where  $\delta(f)$  is the Dirac delta function. This is a useful expression for simplifying

integrals if the usual cautions on the use of the delta function are observed; note that this expression does not apply at the singular points of the spectrum. Information on spectral representations is also given by Doob (1952), Brillinger (1975), and Priestley (1981), among others.

When it comes to the analysis of real data, many of the above assumptions cannot be verified; one must assume that whatever is true for the data is not so radically different from the probabilistic assumptions as to completely invalidate the analysis. A standard problem is that data are not sampled equally in time, so the sampling process must be assumed to represent a continuous time process in some sense. This problem is much more difficult than the equally spaced one as, for example, there are no simple equivalents to the Nyquist frequency or Parseval's theorem, and useful equivalents of the Slepian sequences require additional physical constraints. For an introduction to these problems see Bronez (1988) or Parzen (1983).

The oldest spectrum estimate in common use is the *direct* estimate (Schuster 1898; Tukey 1967),

$$\hat{S}_D(f) = \left| \sum_{n=0}^{N-1} x_n D(n) e^{-i2\pi n f} \right|^2. \quad (6)$$

Here the sequence  $D(n)$ , which directly multiplies the data, is known as a *data window* or *taper* and is normalized so that

$$\sum_{n=0}^{N-1} D^2(n) = 1.$$

The expected value of a direct estimate is

$$\mathbb{E}\{\hat{S}_D(f)\} = S(f) * \{|\tilde{D}(f)|^2\}, \quad (7)$$

where the on-line  $*$  denotes convolution and  $\tilde{D}$  is the discrete Fourier transform of  $\{D(n)\}$ . Parseval's theorem and the normalization of  $D$  show that  $\hat{S}_D$  is unbiased when the spectral density  $S$  is constant. In this application  $|\tilde{D}(f)|^2$  is referred to as a *spectral window* but, as this term is also used in other areas of spectrum estimation with varying implications, it must be qualified. The frequency resolution of the estimate is determined by the width of this window. Because the direct estimate at each frequency is a sum of squares it is distributed as a  $\chi^2$  random variable with two degrees of freedom, i.e. it is exponentially distributed, so its mode is at zero and its variance does not decrease with sample size. Consequently the width of the spectral window does not control the variance of the estimate and the trade-off in window selection is between frequency resolution and bias. The data window  $D$  must be carefully chosen (see Brillinger 1981; Harris 1978). If a prolate spheroidal wave function (Slepian & Pollak 1961; Landau & Pollak 1961; Eberhard 1973; Kaiser 1974; Slepian 1978) is used for a data taper, the bias from all frequencies remote from the frequency of interest decreases exponentially as a function of (window width)  $\times N$ . This bound is robust in that it does not depend on stationarity or normality assumptions, and follows from the Cauchy inequality and the properties of the prolate spheroidal wave functions. Near the opposite extreme, taking  $D$  constant gives the classical, or unwindowed, periodogram and should never be used (see MacDonald 1989).

Because of their exponential distribution, direct estimates are generally useful only if smoothed (averaged over frequency); that is  $\bar{S}_{D,G}(f) = \hat{S}_D(f) * G(f)$ , where  $G$  is a 'smoothing window' with  $G(f) \geq 0$  and  $\int G(f) df = 1$ . Such an operation

obviously decreases the variance of the estimate at the expense of frequency resolution. However, because of the correlations implicitly generated by the convolution (7) between values of  $\hat{S}_D(f)$  at neighbouring frequencies, the variance of  $\bar{S}_{D,G}$  depends on both  $D$  and  $G$ . Corresponding to a given data taper,  $D$  there is an optimum (in the sense of minimum mean-square bandwidth) smoothing window  $G$  which may be found by extensions of the method of Papoulis (1973) given in Thomson (1977). However, as calculations in §IV of Thomson (1982) show, the multi-taper method significantly outperforms such smoothed estimates. (Efficiency calculations for spectrum estimates have the usual statistical connotations (see Jones 1962; Zhurbenko 1978, 1980, 1983).)

An alternative approach is to use Welch's (1961, 1967) method and compute several direct estimates on overlapped subsets of the data, then average the spectrum estimates across subsets. As before, this decreases the frequency resolution, here by a factor of  $N$  (length of subsets), but is efficient if the data are stationary (Thomson 1977). It is most useful for detecting non-stationarity as in Thomson (1977), where Bartlett's  $M$ -test (Pearson & Hartley 1969) for homogeneity of variance was computed as a function of frequency. A related use is in robust estimation from large data-sets, again in Thomson (1977), Martin & Thomson (1982), or Chave *et al.* (1987). Work in progress suggests that for mildly non-stationary data one does better using multiple-window estimates on each subsection, particularly with multivariate problems. One reason is that multivariate non-stationarity appears in a multitude of forms and may not imply that the constituent series are individually non-stationary. Here the advantage of multiple-window methods is that different forms of robust estimation are possible, for example a transfer function and its variance can be estimated on each subsection, then these estimates combined robustly. Similarly, the usual multivariate tests for equality of covariance matrices can be applied to the spectral matrices estimated on each data segment.

To augment my earlier work (Thomson 1990, section titled 'Estimates to be avoided'), spectral estimates based on the simple periodogram or unwindowed FFT (or, equivalently, on sample autocorrelations) including indirect estimates (also known as Blackman-Tukey (1958) estimates) or, worse, adding misfit to bias, the autoregressive or 'maximum-entropy' estimates, must be considered both obsolete and extremely dangerous. Some of these estimators are asymptotically unbiased (see Grenander & Rosenblatt 1958; Berk 1974), however, when considering these or similar estimates with ordinary data it is well to remember a 'folk theorem': If your spectrum estimate explicitly requires sample autocorrelations you are almost certainly doing something wrong. One must also be careful with 'maximum-entropy' and related estimates as one name is often used to refer to several distinct estimates (Kay & Marple 1981); others, such as the Chao & Gilbert (1980) AR form work on prefiltered data (see Thomson 1986). Some of these, e.g. the so-called 'covariance form' (see Markel & Gray 1976), have been specifically tailored for speech applications and effectiveness in specialized applications does not imply that they are good general data analysis tools. Others (see Marple 1987) are optimized for array problems where few elements but many 'snapshots', or replications, are available. In addition, numerous specific time series methods have been developed; see the papers in Franke *et al.* (1984), Rao & Gabr (1980), Brillinger & Krishnaiah (1983), Gani & Priestley (1986), or the *Proceedings of the Nth Spectrum Estimation Workshop*, sponsored by the Acoustics, Speech, and Signal Processing Society of the IEEE. There are other 'new' methods that should be avoided, for example, Brethorst's

(1988) method is just an orthogonalized least-squares fit assuming global white noise; the method of ‘linear spectrum estimation’ fails in all but trivial cases and is biased (see Stigler & Wagner 1987); and the ‘Sompi’ method has too many faults to summarize concisely (see Lindberg & Thomson 1990). Although more detailed consideration of these lies beyond the scope of this paper, those contemplating use of such methods should first read Tukey (1984) or Gutowski *et al.* (1978).

Unlike conventional statistics where one may attempt to estimate a few parameters from data, in spectrum estimation problems one is attempting to estimate a *function*. It is often claimed that because of this indeterminacy some criteria must be invoked (implicitly or explicitly) to enforce a degree of ‘reasonableness’ on the solution, and this is used as a justification for maximum entropy estimates. Similarly, it is stated that direct estimates can only be computed at the ‘Fourier frequencies’,  $k/N$ , with  $k$  an integer and so must be interpolated in some way. Both assertions are patently incorrect. Not only are direct estimates continuous functions of frequency, they are also entire functions, i.e. analytic everywhere. As such, one could hardly ask for a more reasonable solution, but the implicit requirement that the solution be an entire function is too strong for many uses.

The spectrum estimation problem is intrinsically ill posed and is more akin to Backus–Gilbert theory than to classical statistics (see Backus & Gilbert 1967, 1968, 1970; Parker 1977a, b). They differ, however, in that Backus–Gilbert theory is typically concerned with finding approximate solutions of a linear integral equation whereas, in spectrum estimation problems, one typically is more interested in obtaining accurate estimates of the second moments of such a solution. For these estimates I introduce *quadratic inverse theory* that gives solutions using a second basis set of special functions. The reader is cautioned that some of the methods described here are new and still somewhat empirical; several years from now the recommended procedures will probably have evolved. However, despite their present evolutionary state, multiple-window methods probably constitute the best available method for the analysis of physical time series, and particularly so when available data are limited.

## 2. Multiple-window estimates

The single-window estimates described in the preceding section are discouraging in several respects. (i) What theory was used to introduce the data window  $D$ ? Clearly it is easy to show that some choices of  $D$  are better than others, but this is not the same as starting from first principles and finding that the estimator should be of this form. (ii) Note that although all the data are equally valid, a ‘good’  $D$  will weight data in the centre of the series more heavily than data near the ends of the series. (Recall that the periodogram, which provides ‘uniform weighting’, has very poor properties.) Worse, if one regards the weighting from the Shannon viewpoint, wherein information is measured by the reduction in uncertainty, the ends of a series should be weighted heavier than the centre as the variance for predicting data at the ends of a series is typically much higher than that for interpolating a point near the centre, i.e. points near the ends are more informative than those near the centre. (iii) The smoothing operation was again introduced in an *ad hoc* manner. The assumption here is that the spectrum being estimated is smooth. Although one may be willing to make such an assumption about the continuous part of the spectrum for the width of the basic spectral window, it becomes much less satisfactory when there are line

components or when smoothing increases this bandwidth considerably. (iv) The above procedures ignore distinctions between continuous background spectra (for which they work well) and line components (which are badly smeared). (v) The direct estimates are entire functions of frequency, which is often too strong a constraint while the parametric alternatives require acceptance of unacceptable constraints. (vi) The statistical efficiency of the admissible estimators is poor. In this regard one observes that the bandwidth,  $W$ , of the better windows is typically more than  $1/N$  ( $4/N$  for the Parzen or Papoulis windows) whereas fundamental theorems of Landau & Pollak (1961, 1962), Slepian & Pollak (1961) require at least  $2NW$  basis functions to represent the information in a bandwidth  $W$  so the effective basis set corresponding to common single-window estimates is deficient.

Consideration of these and related problems led to the multiple-window method. The basis of this method is a Fredholm integral equation of the first kind expressing the projection from the random orthogonal measure (or orthogonal increment process) in the Cramér spectral representation onto the discrete Fourier transform of the sample

$$y(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbb{D}_N(f-f') dX(f') \quad (8)$$

produced by substituting (3) in (1). Because the Fourier transform  $y(f)$  in (1) may be inverted to give the original data, it is trivially a sufficient statistic and so may be treated as the data. The multiple-window class of spectrum estimates is the result of an approximate least-squares solution of equation (8) using a local eigenexpansion. Note, however, that because one is attempting to estimate a function from a finite sample there is not a unique solution. The best one can hope for is a ‘reasonable’ one. The smoothed multiple-window solution is obtained by the following simple steps:

1. Given the set of  $N$  observations, one chooses an analysis bandwidth  $W$ . Typically  $W$  is between  $2/N$  and  $20/N$  with  $4/N$  a good initial choice. If  $W$  is too small the estimate is statistically unstable, but if  $W$  is too large, resolution is poor. The choice of  $W$  dictates the number  $K$  of windows; there are  $\lfloor 2NW \rfloor$  windows ( $\lfloor \cdot \rfloor$  denoting the least integer or ‘floor’ function) having most of their energy in the frequency band  $(-W, W)$  but, as the higher-order windows become less concentrated,  $K$  is usually taken 1 to 3 smaller than this.

2. The data series is multiplied by each of a set of  $K$  data windows and a discrete Fourier transform done on each to produce the  $K$  functions:

$$y_k(f) = \sum_{t=0}^{N-1} x_t v_t^{(k)}(N, W) e^{-i2\pi ft}. \quad (9)$$

The data windows,  $v_t^{(k)}(N, W)$ ,  $k = 0, 1, \dots, K-1$ , are known as *discrete prolate spheroidal sequences* or *Slepian sequences*. (The suggestion to refer to the discrete prolate spheroidal wave functions and sequences as *Slepian* functions and sequences was made by Professor Robert Parker of the Institute of Geophysics and Planetary Physics of Scripp’s Institution of Oceanography. Since the term ‘discrete prolate spheroidal wave function’ is unwieldy and properties of the spheroidal coordinate system rarely (if ever) occur in signal processing and statistical applications it seems appropriate to name these functions after David Slepian, who first (1953) described their properties for these purposes.) They are orthonormal and have the property that of any set of  $K$  orthonormal sequences of duration  $N$ , their Fourier transforms, or *Slepian functions*  $V_k(f)$  given in (A 5) have the maximum energy concentration in

the bandwidth  $(-W, W)$ . The data windows depend on  $N$ , the number of observations, and the bandwidth parameter  $W$ . These parameters are important in that, asymptotically the sidelobe height or amplitude of the spectral windows outside the inner band decreases exponentially in  $NW$  for the  $K$  lowest-order windows. Properties of these sequences and their discrete Fourier transforms, known as Slepian functions, or discrete prolate spheroidal wave functions, used in the following, are contained in Slepian (1978) and in Appendix A. Efficient methods for computing the Slepian sequences are described in Appendix B (see also Slepian 1964, 1983).

Considerable insight into the multiple-window method may be obtained by substituting the Cramér representation of  $x_t$  (3) in (9) to give

$$y_k(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} V_k(\xi) dX(f - \xi). \quad (10)$$

The  $y_k(f)$ s should be considered as crude estimates of the ideal bandlimited eigencoefficients

$$x_k(f) = \int_{-W}^W \mathcal{V}_k(\xi) dX(f - \xi). \quad (11)$$

Here, in addition to the change in domain of the integration, the  $\mathcal{V}_k(f)$ , defined in (A 21), are orthonormal on the inner interval and so differ from the  $V_k(f)$ s by a factor of  $1/\sqrt{\lambda_k}$ . The  $y_k(f)$  are the raw coefficients of the eigenexpansion in the interval  $(f - W, f + W)$  and are biased, in the sense of including information from the outer interval.

The absolute square of the  $k$ th eigencoefficient  $\hat{S}_k(f) = |y_k(f)|^2$  called the  $k$ th eigenspectrum is a direct spectrum estimate, and hence is distributed as  $\chi_2^2$ .

3. Leaving to the next section the problems associated with line components, one forms a local least-squares estimate of the eigencoefficients

$$\begin{aligned} \hat{x}_k(f) &= d_k(f) y_k(f) \\ &= [\lambda_k^{\frac{1}{2}} \hat{S}(f) / (\lambda_k \hat{S}(f) + (1 - \lambda_k) \sigma^2)] y_k(f) \end{aligned} \quad (12)$$

and spectrum by iteratively solving this equation with

$$\hat{S}(f) = \sum_{k=0}^{K-1} |\hat{x}_k(f)|^2 / \sum_{k=0}^{K-1} d_k^2(f) \quad (13)$$

for  $\hat{S}(f)$ . Here

$$\sigma^2 = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2$$

is the process variance. The  $\lambda$ s are the Slepian eigenvalues and give the fraction of the energy of  $v_n^{(k)}(N, W)$  in the frequency band  $(-W, W)$  relative to its total energy in the Nyquist band  $[-\frac{1}{2}, \frac{1}{2}]$ . If one starts this procedure with  $\hat{S}(f) = \frac{1}{2}(S_0(f) + S_1(f))$  it typically converges to a few percent accuracy in two or three iterations.

The standardization given in the denominator of (13) implies that  $\hat{S}(f)$  is unbiased for globally white noise, that is  $\hat{S}$  is unbiased when the spectrum is constant over the whole frequency domain. Note that  $\hat{S}(f)$  is not a simple windowed transform of the data but the root of a rational equation in such forms. As such the estimate of the autocovariance obtained by taking the Fourier transform of  $\hat{S}(f)$  is not necessarily zero for lags greater than  $N$ .

For small  $k$  the eigenvalues,  $\lambda_k$ , are exponentially close to 1 in  $NW$  and so in many problems the above estimate  $\hat{S}(f)$  is approximately

$$\bar{S}(f) = \frac{1}{K} \sum_{k=0}^{K-1} S_k(f), \quad (14)$$

just the average of the  $K$  lowest-order eigenspectra. The first term in this series,  $\hat{S}_0(f)$ , is simply a direct estimate of the type discussed in §1 and, if  $NW = 4$ , will be similar to the estimate obtained with a Parzen (1967), Papoulis (1973), or Thomson *et al.* (1976) data window. The higher terms in the series are also direct estimates that, if the true spectrum does not vary too much in the interval  $(f-W, f+W)$ , will be almost uncorrelated with one another because the data tapers are orthogonal. The average spectrum will thus have resolution  $\pm W$  and stability characterized by  $2K$  degrees of freedom. In addition to the normalization given in (13) other choices are possible. Replacing the denominator of (13) with  $K$  gives an unbiased estimate for locally white noise. The local least-squares weighting in (12), essentially a Wiener filter in disguise, has the advantage of simplicity because the estimates at different frequencies are done separately; it can be improved by including a ‘sidelobe canceller’, where one iteratively estimates the part of  $y_k(f)$  from the exterior domain and subtracts it to estimate  $x_k(f)$ .

These estimates do not conserve power in a rigorous sense so that Parseval’s theorem is only satisfied approximately in expected value. Because most of the power in many physically interesting processes is concentrated in a small part of the frequency band the sample variance typically has relatively few degrees of freedom so ‘correction’ of (13) by a factor of

$$\frac{1}{N} \sum_{n=0}^{N-1} x_n^2 / \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{S}(f) df$$

usually results in an estimate with both a higher variance and a more complicated distribution than  $\hat{S}(f)$ .

The use of several orthogonal data windows initially strikes one as being strange, particularly after hearing years of argument on the subject of whether one should or should not use any data window or if one should base the analysis on sample autocovariances. However, the additional complexity entailed by multiple data windows is more than offset by the advantages of the method. (i) The estimation process is obtained by straightforward application of well-established procedures. It begins with a Fredholm equation relating the observations to the relevant theoretical functions. An approximate solution of this equation is then found using an eigenexpansion. The different data tapers fall out of the method rather than being introduced heuristically and then justified. The method is, in a sense, equivalent to Backus–Gilbert inverse theory. (ii) The statistical efficiency of the method is much better than that of conventional estimates, typically 80–95 % as opposed to 20–40 %. (iii) The estimate is consistent. For fixed  $W$ , the variance of the estimate tends to zero as  $1/N$  as  $N \rightarrow \infty$ . Note that the estimates have nearly  $2K$ , or  $4NW$ , degrees of freedom, typically 10 in the spectra presented in the following sections, but have the same basic frequency resolution as the common windows characterized by 2 degrees of freedom before smoothing. There are almost  $2K$  degrees of freedom because the solution of equations (12) and (13) effectively varies the truncation point,  $K$ , as a

function of frequency. Because the intrinsic dimensionality of the process restricted to  $(f-W, f+W)$  is *ca.*  $2NW$  one must have  $K \leq [2NW]$ . Thus, when the spectrum is highly coloured, regions where the spectrum is large will have more stable estimates (higher effective  $K$ ) than where the spectrum is small. (iv) As will be seen in the next section, the use of multiple windows allows one to separate line components from the continuous background spectra and to give a test for their significance. (v) The multiple-window method may be jackknifed to estimate both parameters and estimates of their variance. Bias, resolution, and variance are all controlled by the parameter  $W$  in the multiple-window method. The difficult part, choosing  $W$ , is addressed in §3. Examples of multiple-window estimates are given in Forster & Vezzosi (1987), Lindberg (1986), Hinnov & Park (1988), Park *et al.* (1987*a, b*), Vernon (1989), Walden (1990*a, b*), Mullis & Scharf (1990), and elsewhere.

### 2.1. Jackknifed multi-window estimates

That multi-taper estimates can be jackknifed is exceptionally important in applications. The usual approach in spectrum estimation has been to estimate variances or confidence intervals using  $\chi^2$  distributions parametrized by the degrees of freedom of the estimate and independently of the data; see Izenman & Sarkar (1987) for a recent refinement of this method. The justification for making a gaussian assumption is that the output of a narrow-band filter, here the  $y_k(f)$ s considered as functions of the starting time of the record, is close to gaussian for reasonable conditions on the input (see Rozanov 1961; Rosenblatt 1961; Mallows 1967; Brillinger 1975). Other relevant properties of filtered processes are available in Rice (1948, 1982), Papoulis (1967), Kailath (1974), and Khugin & Yakovlev (1977). Many series encountered in practice, however, are either subtly non-stationary or have fine structure in their spectrum that is not resolved in the estimate. The uncertainties resulting from these departures from a stationary process with a well-resolved spectrum should be reflected in uncertainty estimates. With multiple-window methods such effects usually receive different weighting by the different data tapers and so increase the jackknife estimate of variance. Thus I strongly recommend that one compute both the standard gaussian theory and jackknife variance estimates, then plot their ratio. If the two estimates are similar one has greater confidence in the estimate; however, when they differ significantly the discrepancy must be explained; see Thomson (1990) for an example.

Jackknife variance estimates smaller than gaussian variances are unusual; the only cause I know for this is a ‘superstationary’ process consisting of many unresolved line components so the distribution of the estimated spectrum is non-central  $\chi^2$ . An interesting example of such a series is given by spacings between successive zeros of the Riemann zeta function (Odlyzko 1987). Letting  $\alpha$  denote the ratio of the non-central to central power, the apparent degrees of freedom will be increased by a factor  $(1+\alpha)^2/(1+2\alpha)$ . Excess jackknife variance is much more common and usually suggests unresolved structure, non-stationarity, or simply outliers.

For independent, identically distributed observations the jackknife is a standard method (see Mosteller & Tukey 1977; R. G. Miller 1974; Efron 1982). If the spectrum is resolved, as defined in the following section, and if the local band does not contain a periodic component, the eigencoefficients at a given frequency are close to iid and the adaptation of the jackknife to spectrum estimation given in Thomson & Chave (1990) is as follows: denote by  $\hat{S}_{jj}(f)$  a solution of (12) and (13) with the  $j$ th

window deleted, that is, the sums in (13) run from  $k = 0$  to  $k = K - 1$  with  $k \neq j$ . Compute the average

$$\ln \hat{S}_{\text{all}}(f) = \frac{1}{K} \sum_{j=0}^{K-1} \ln \hat{S}_{jj}(f) \quad (15)$$

and sample variance

$$\hat{V}(f) = \frac{K-1}{K} \sum_{j=0}^{K-1} [\ln \hat{S}_{jj}(f) - \ln \hat{S}_{\text{all}}(f)]^2. \quad (16)$$

Given a spectrum estimate with  $\nu$  degrees of freedom made from a random sample from a stationary gaussian process the variance of  $\ln \hat{S}$  is given by  $\psi'(\frac{1}{2}\nu)$  where  $\psi$  is the digamma function (Abramowitz & Stegun 1965). Taking  $\nu \approx 2K$  one has

$$\psi'(K) = \frac{1}{6}\pi^2 - \sum_{j=1}^{K-1} j^{-2}, \quad (17)$$

where the sum is zero for  $K = 1$ . The jackknife bias estimate on  $\ln \hat{S}(f)$  given by  $(K-1)(\ln \hat{S}_{\text{all}}(f) - \ln \hat{S}_{\text{all}}(f))$  when applied directly to the estimated spectrum becomes a factor

$$\hat{B}(f) = [\hat{S}_{\text{all}}(f)/\hat{S}_{\text{all}}(f)]^{K-1}, \quad (18)$$

where  $\hat{S}_{\text{all}}(f)$  denotes the original solution of (13) using all the windows. For other functions such as coherence, different transformations are appropriate; Thomson & Chave (1990) give several detailed examples.

### 3. Quadratic inverse theory

I have named the methods to be described in this section quadratic inverse theory to emphasize both the philosophical connection between inverse theory and spectrum estimation, and simultaneously that the problem is not that of simply obtaining an approximate solution of a Fredholm equation as in linear inverse theory but instead, that of determining the second moments of such a solution. In retrospect, the multiple-window method as originally introduced was more a philosophy of spectrum estimation than a distinct method, and several new estimates of the spectral density function and related entities were suggested. The simplest and least understood of these was the ‘high-resolution’ estimate

$$\hat{S}_h(f \ominus \xi) = \frac{1}{N} \left| \sum_{k=0}^{K-1} \hat{x}_k(f) \overline{\mathcal{V}_k(\xi)} \right|^2, \quad (19)$$

where  $f \oplus \xi$  and  $f \ominus \xi$  denote  $f + \xi$  and  $f - \xi$  respectively, with the understanding that  $\xi$  is restricted to  $(-W, W)$  and  $\mathcal{V}_k(\xi)$ , defined in (A 21) is orthonormal on the inner band  $(-W, W)$ . The study of this high-resolution estimate has been responsible for raising numerous questions and although the quadratic inverse theory described here answers some of these, it perhaps leads to more mysteries than it answers. For example, it is shown below that although (19) is unbiased for slowly varying spectra, it underestimates fine spectral structure. I emphasize that this theory is still being actively developed; consequently, in this section, I am again concerned exclusively with the continuous part of the spectrum. Details of the full mixed problem with line components, while an obvious but tedious extension, must be deferred to another paper.

In part, quadratic inverse theory arose from an attempt to answer the question, ‘How does one choose the bandwidth  $W$ ?’ With continuous spectra this harks back to the papers of Bartlett (1950), Blackman & Tukey (1958), Parzen (1957, 1967), and is at the heart of the ‘resolution-variance’ tradeoff commonly considered in papers on indirect spectrum estimates. Recently Hurvich (1985, 1988) has made some progress applying cross-validation and inverse correlation ideas to single-window estimates. With multiple-window estimates this question may be posed as follows: denote by  $\hat{\mathbf{X}}(f)$  the complex vector of the  $K$  estimated eigencoefficients (12)

$$\hat{\mathbf{X}}(f) = [\hat{x}_0(f), \hat{x}_1(f), \dots, \hat{x}_{K-1}(f)]^T$$

and an estimate of its covariance matrix (as a function of frequency) by  $\hat{\mathbf{C}}$ ,

$$\hat{\mathbf{C}}(f) = \hat{\mathbf{X}}(f) \hat{\mathbf{X}}(f)^T. \quad (20)$$

Note that  $\hat{\mathbf{C}}$  is not the sample autocovariance matrix. Also, as mentioned earlier, the estimated eigencoefficients  $\hat{x}_k(f)$  should be thought of as approximations to the unobservable ideal eigencoefficients given by (11). Recall that we are interested in estimation in the inner band; this led to the weights of the  $y_k(f)$  given in (12) and in the reshaping operation described in §5 below. A standard eigensolution of the fundamental integral equation (8) gives the approximate narrow-band expansion

$$d\hat{\mathbf{X}}(f \ominus \xi) \approx \sum_{k=0}^{K-1} x_k(f) \bar{\mathcal{V}}_k(\xi) d\xi \quad (21)$$

suitable for representing inner-band processes about a given frequency. Incidentally the Fourier transform of (21) over the inner band

$$x(f; t) = \sum_{k=0}^{K-1} x_k(f) \lambda_k^{\frac{1}{2}} v_t^{(k)}(N, W)$$

is a useful complex demodulate of the process. The covariance matrix  $\mathbf{C}$  has components

$$C_{jk}(f) = \mathbb{E}\{x_j(f) \bar{x}_k(f)\}. \quad (22)$$

Using the orthogonal increment property (5) of the spectral representation for a stationary process and substituting (11) for each eigencoefficient in (22) gives

$$C_{jk}(f) = \int_{-W}^W \mathcal{V}_j(\xi) \bar{\mathcal{V}}_k(\xi) S(f - \xi) d\xi. \quad (23)$$

hence, if  $S(f)$  is constant over the interval  $(f - W, f + W)$ , the diagonal elements of  $\mathbf{C}$  are the different eigenspectra,  $\mathbb{E}\{|x_k(f)|^2\}$ , and

$$\mathbf{C}(f) = S(f) \mathbf{I},$$

where  $\mathbf{I}$  is the  $K \times K$  identity matrix. Similarly, if  $W$  is chosen such that the spectrum varies significantly across the band  $(f - W, f + W)$ ,  $\mathbf{C}(f)$  will not be diagonal.

**Definition.** Conditioned on the observed data span  $N$ , I take  $\mathbf{C}(f)$  being diagonal as the definition of the spectrum being resolved.

This definition is pragmatic and statistical in nature as opposed to probabilistic; in essence the question being asked is, ‘Given the available data, has the estimation procedure obscured fine structure that is theoretically observable in a sample of this size?’ Whether the true spectrum contains fine structure observable with a larger sample is not answered; as usual, such questions can only be answered with a larger sample or better physical understanding. Clearly one rarely expects to have perfectly

resolved spectra or  $\mathbf{C}(f)$  to be exactly diagonal. In practice, one attempts to have ‘reasonably’ resolved spectra with corresponding ‘reasonably’ diagonal estimates of  $\mathbf{C}$ .

For gaussian data the standard test for diagonal form is one of Wilks’s classic likelihood ratios known as the ‘sphericity test’, (Anderson 1984). This test, however, requires that the estimated matrix  $\hat{\mathbf{C}}(f)$  be non-singular and thus requires replication of the data-set or subdivision of a longer set. Because subdivision of a longer data-set of length  $T$  into shorter segments of length  $L$  gives an immediate loss of resolution by a factor of  $T/L$  it seems silly to subdivide only to be able to tell that spectra estimated on the shorter segments are poorly resolved! One way of circumventing this difficulty is to assume a less general alternative family of covariance matrices than is implicit in the sphericity test, and my initial attempt was to use a Taylor (or polynomial) expansion of the spectrum within the local bandwidth. Algebraic and numerical complexity, however, made it obvious that this was unsatisfactory and the following expansion in eigenfunctions of the asymptotic Fejér kernel is both more satisfactory, and by the trace-orthogonality property of the induced basis matrices, algebraically much simpler than the Taylor series.

The expected value of the high-resolution estimate  $\hat{S}_h(f \oplus \xi)$  given by (19) is, using (23) and rearranging terms

$$\mathbb{E}\{\hat{S}_h(f \oplus \xi)\} = \frac{1}{N} \int_{-W}^W |\mathbb{P}_W(\xi, \eta)|^2 S(f + \eta) d\eta \quad (24)$$

where the inner domain projection operator  $\mathbb{P}_W$  is defined by (A 23). Because the first  $2NW$  eigenvalues are close to 1, comparing (A 19) and (A 23) shows that  $\mathbb{P}_W(\xi, \eta)$  is almost a convolution operator. Denote by  $B_l(N, W, K; \xi)$ , hereinafter written  $B_l(\xi)$ , the real eigenfunctions of the degenerate symmetric kernel  $|\mathbb{P}_W(\xi, \eta)|^2$

$$g_l B_l(\xi) = \int_{-W}^W |\mathbb{P}_W(\xi, \eta)|^2 B_l(\eta) d\eta \quad (25)$$

ordered by the eigenvalues  $g_0 \geq g_1 \geq \dots \geq g_{2K-1} \geq 0$  and normalized so that

$$\frac{1}{2W} \int_{-W}^W B_l(f) B_m(f) df = \delta_{lm}. \quad (26)$$

Substituting the following expansion of the spectrum on the inner band

$$S(f \ominus \xi) \approx \sum_{l=0}^{L-1} b_l(f) B_l(\xi) \quad (27)$$

where  $L \leq 2K - 1$ , in (23) produces a components-of-variance-like expansion of  $\mathbf{C}(f)$ ,

$$\mathbf{C}(f) \approx \sum_{l=0}^{L-1} b_l(f) \mathbf{B}^{(l)}, \quad (28)$$

where the elements of the  $K \times K$  complex matrix  $\mathbf{B}^{(l)}$  are given by

$$B_{jk}^{(l)} = \int_{-W}^W \mathcal{V}_j(\xi) \mathcal{V}_k(\xi) B_l(\xi) d\xi \quad (29)$$

or, using (A 16) and (A 21)

$$B_{jk}^{(l)} = \frac{1}{\epsilon_j \bar{\epsilon}_k \sqrt{(\lambda_j \lambda_k)}} \int_{-W}^W U_j(-\xi) U_k(\xi) B_l(\xi) d\xi$$

so the  $\mathbf{B}^{(l)}$ 's are real symmetric for  $l$  even and imaginary skew-symmetric for  $l$  odd. These matrices have the remarkable property

$$\text{tr}\{\mathbf{B}^{(l)}\mathbf{B}^{(m)}\} = g_l \delta_{lm}, \quad (30)$$

where tr denotes trace. This trace-orthogonality is a general feature of product kernels and may be verified by using (29) to expand both matrices in (30), identify both sums as the projection kernel (A 23), then substitute (25). For a single series  $\hat{\mathbf{C}}(f) = \hat{\mathbf{X}}\hat{\mathbf{X}}^\dagger$  is obviously of rank 1. An estimate of  $\mathbf{C}(f)$  of the form

$$\tilde{\mathbf{C}}(f) \approx \sum_{l=0}^{L-1} \hat{b}_l(f) \mathbf{B}^{(l)} \quad (31)$$

is typically of full rank. Minimizing the squared difference (formally the matrix Frobenius norm) between  $\hat{\mathbf{C}}$  and  $\tilde{\mathbf{C}}$  and using (30) gives the estimated expansion coefficient

$$\hat{b}_l(f) = (1/g_l) \text{tr}\{\hat{\mathbf{C}}(f)\mathbf{B}^{(l)}\}$$

or, for a single series, as the standard quadratic form

$$\hat{b}_l(f) = (1/g_l) \hat{\mathbf{X}}^\dagger \mathbf{B}^{(l)} \hat{\mathbf{X}}. \quad (32)$$

We show below that the  $\hat{b}_l(f)$  are unbiased and, when the local spectrum is white, uncorrelated. Note that  $B_l(f)$  denotes a local basis function for the spectrum but  $\mathbf{B}^{(l)}$  denotes the corresponding  $K \times K$  basis matrix for the eigencoefficient covariance matrix, and that both expansions have the same coefficients. Consequently, estimation may be done in either domain and questions about one (for example the sphericity test) posed in the other. The standardization (26) of the functions  $B_l(f)$  defined on  $|f| < W$  was chosen so the resulting expansion coefficients

$$b_l(f) = \frac{1}{2W} \int_{-W}^W S(f-\nu) B_l(\nu) d\nu$$

are directly comparable in magnitude with  $S$ .

For the usual multiple-window estimates the eigenvalues  $\lambda_k$  are all close to 1 so that  $\mathbb{P}_W(f, g)$  is by (A 19), approximately the Dirichlet kernel. The kernel of (25) is thus similar to the Fejér kernel, and the eigenfunctions  $B_l$  like those of

$$\tilde{g}_l \tilde{B}_l(\xi) = \int_{-W}^W \left[ \frac{\sin\{N\pi(\xi-\eta)\}}{\sin\{\pi(\xi-\eta)\}} \right]^2 \tilde{B}_l(\eta) d\eta. \quad (33)$$

Because its kernel is degenerate there are  $2K \approx 4NW$  non-zero eigenvalues of (25). For the asymptotic form (33) there are still about  $4NW$  non-zero eigenvalues and these tend to

$$g_l \rightarrow 2NW - \frac{1}{2}l, \quad l = 0, 1, \dots, 4NW. \quad (34)$$

Properties of these eigenvalues are similar to those of  $\text{sinc}^2$  available in the optics literature where the continuous-time counterpart of equation (33) characterizes incoherent imaging. For the  $\text{sinc}^2$  kernel it is known (Gori 1974; Gori & Palma 1975) that the eigenvalues are non-degenerate and the eigenfunctions are complete. Because only about  $4NW$  eigenvalues are non-zero structure corresponding to eigenfunctions with  $l$  greater than  $4NW$  is not observable. Superficially one might expect that because

$$\text{tr}\{\mathbf{B}^{(l)}\mathbf{B}^{(m)}\} = (\text{vec } \mathbf{B}^{(l)}) (\text{vec } \mathbf{B}^{(m)}),$$

where ‘vec’ is the column stacking operation, one could choose  $K^2$  such matrices by a Gram–Schmidt operation in contrast to the near  $2K$  obtained here. Recalling,

however, that the kernel  $\mathbb{P}_W$  consists of sums of terms  $\mathcal{V}_j(f) \bar{\mathcal{V}}_k(\xi)$  corresponding, in the time domain, to  $v_n^{(j)}(N, W) * v_n^{(k)}(N, W)$  (and so is time-limited to  $2N-1$  and bandlimited to  $W$ ) the dimensionality is clearly about  $4NW$ , and consequently resolution is limited to twice the Rayleigh resolution. Equations (25) and (33) are handily solved using the Slepian functions  $V_k(2N-1, W; f)$  as a basis. Thus for the continuum part of the spectrum, superresolution is limited to 2. A plot of Fisher information (reciprocal variance) against superresolution (Rayleigh resolution divided by achieved resolution) is a triangle with intercepts of  $1/2NW$  and 2. The differences between this result and corresponding superresolution for line components is striking: with line components superresolution is proportional to signal-to-noise power ratio (see equation (47) below) and much better than Rayleigh resolution may be obtained with reasonable signals; in contradistinction resolution of the continuum spectrum is restricted to twice the Rayleigh resolution.

### Moments

Empirically, the estimated  $b_l$ s are stable and their variance and distributions are of interest. Expanding (32) one has

$$\hat{b}_l(f) = \frac{1}{g_l} \sum_{j=0}^{K-1} \bar{B}_{jk}^{(l)} x_j(f) \bar{x}_k(f). \quad (35)$$

Taking expected values via (23) and assuming that the spectrum in  $(f-W, f+W)$  may be expressed by the expansion (27) the trace-orthogonality (30) gives

$$\mathbb{E}\{\hat{b}_l(f)\} = b_l(f)$$

for  $g_l > 0$  so that the estimate (32) is unbiased. Thus we emphasize: If the spectrum has a local expansion expressible in the lowest order  $4NW$  eigenfunctions of  $\text{sinc}^2$ , unbiased estimates of the expansion coefficients are given by the estimator (32).

To compute the variances we begin with (32) and expand to obtain

$$\mathbb{E}\{\hat{b}_l \bar{\hat{b}}_m\} = \frac{1}{g_l g_m} \times \sum_{h,i,j,k} \bar{B}_{hi}^{(l)} B_{jk}^{(m)} \mathbb{E}\{x_h \bar{x}_i \bar{x}_j x_k\}.$$

We assume complex gaussian statistics, expand the fourth moment (see K. S. Miller (1974), p. 82), and discard the product of mean-value terms to obtain

$$\text{Cov}\{\hat{b}_l, \bar{\hat{b}}_m\} = \frac{1}{g_l g_m} \times \sum_{h,i,j,k} \bar{B}_{hi}^{(l)} B_{jk}^{(m)} \mathbb{E}\{x_h \bar{x}_j\} \mathbb{E}\{x_k \bar{x}_i\}. \quad (36)$$

(As usual, there is an additional term for real series of importance only at zero and the Nyquist frequencies.) There are two cases: the general problem and the simple one where the spectrum is assumed locally white with a constant spectral density  $S$ . We begin with the latter as it is, by definition, a fully resolved spectrum; recall that for this case  $\mathbb{E}\{x_h \bar{x}_j\} = S \delta_{hj}$ . The remaining double sum is the trace defined in (30) and gives

$$\text{Cov}\{\hat{b}_l, \bar{\hat{b}}_m\} = (S^2/g_l) \delta_{lm}.$$

Thus, when the spectrum is resolved, the estimated spectrum components obtained from (32) are uncorrelated. Moreover, since by (34) there are typically several large  $g_l$ s the variances of the lower order  $b_l$ s are not significantly larger than that of  $\bar{S}$ . For the resolved estimate one has

$$\mathbb{E}\{\hat{b}_l\} = (S/g_l) \text{tr}\{\mathbf{B}^{(l)}\}. \quad (37)$$

Not surprisingly, formulae for variances and covariances of the  $\hat{b}$ s with an unresolved spectrum are considerably more complicated than given above. Begin with (36) and expand the two expectations using (28) written as

$$\mathbb{E}\{x_j(f)\bar{x}_k(f)\} = \sum_l b_l(f) B_{jk}^{(l)}.$$

Some algebra results in

$$\text{Cov}\{\hat{b}_l(f), \overline{\hat{b}_m}(f)\} = \frac{1}{g_l g_m} \times \sum_{p,q} b_p b_q \text{tr}\{\mathbf{B}^{(p)} \mathbf{B}^{(q)} \mathbf{B}^{(l)} \mathbf{B}^{(m)}\}. \quad (38)$$

### *Connection with the high-resolution solution*

Because (19) is the absolute square of a complex gaussian random variable, it is distributed as a multiple of  $\chi^2_2$  and hence inconsistent unless smoothed in some way. The simplest smoother is an unweighted average over the interior frequency domain ( $f-W, f+W$ ), leading to the estimate

$$\bar{S}(f) = \frac{1}{2W} \int_{-W}^W \hat{S}_k(f \ominus \xi) d\xi. \quad (39)$$

By the orthogonality of the Slepian functions (39) reduces to (14), a consistent estimate. If the spectrum is locally white the different terms in (14) are uncorrelated and it will be distributed as  $\chi^2$  with  $2K$  degrees of freedom. Now replace the simple smoothing by averaging over the inner band by the projection of (19) on the subspace spanned by the basis functions  $B_l(f)$  and generalize (39) by writing

$$\bar{S}_{(l)}(f) = \frac{1}{2W} \int_{-W}^W \hat{S}_h(f \ominus \xi) B_l(\xi) d\xi.$$

Expanding the summations in (19), integrating, and identifying terms in (29) and (32), given expanded in (35), shows that  $\bar{S}_{(l)}(f) = g_l b_l(f)/2NW$  so while  $\bar{S}_{(0)} \approx b_0$  the higher coefficients are biased down.

Returning to the problem of attempting to tell if the spectrum is resolved, note that Wilks's sphericity test (see, for example, Wilks 1962; Anderson 1984) is Bartlett's  $M$  test applied to the eigenvalues of an estimate of  $\mathbf{C}(f)$ . This clearly requires the tested matrix to be full rank and so cannot be applied to  $\mathbf{X}\mathbf{X}^\dagger$ . Thus I apply it, in modified form, to the projection  $\tilde{\mathbf{C}}(f)$ . Now, assume that the spectrum is almost resolved so  $\tilde{\mathbf{C}}$  is nearly a sum of  $S(f)\mathbf{I}$  and terms small enough for perturbation expansions to be valid. Denoting the eigenvalues of  $\tilde{\mathbf{C}}$  by  $c_j$ ,  $M$  is approximately  $\kappa \text{var}\{c\}/\text{ave}^2\{c\}$  where  $\kappa$  is a constant. Because the matrix Frobenius norm is invariant under orthogonal transformations,

$$\text{var}\{c\} = (1/K) \|\tilde{\mathbf{C}}(f) - S_{\text{av}}(f) \mathbf{I}\|_F^2,$$

where  $S_{\text{av}}(f)$  is the average spectrum over  $(f-W, f+W)$ ,

$$S_{\text{av}}(f) = \frac{1}{K} \text{tr } \tilde{\mathbf{C}}(f) = \frac{1}{K} \sum_{l=0}^{L-1} b_l(f) \text{tr } \mathbf{B}^{(l)}$$

and some algebra gives

$$\text{var}\{c\} = \sum_{l=0}^{L-1} g_l [b_l(f)]^2 - K[S_{\text{av}}(f)]^2 \quad (40)$$

and I use

$$r_a(f) = \text{var}\{c\}/[S_{\text{av}}(f)]^2 \quad (41)$$

as an initial resolution test. For ‘equal-energy’ ripples in the spectrum, test sensitivity decreases linearly with ripple frequency and disappears for details in the spectrum finer than half the Rayleigh resolution, that is, for  $\delta f < 1/2N$ . Because features in the spectrum with odd symmetry about a given frequency do not bias a univariate spectrum estimate at that frequency, the sum in (40) could be restricted to  $l$  even. Recalling that  $S_{\text{av}}^2(f)/g_l$  is the variance of  $\hat{b}_l(f)$  when the spectrum is resolved, the test is of the form

$$\sum_{l=0}^{L-1} \frac{[b_l(f) - \hat{\mathbb{E}}_0\{b_l(f)\}]^2}{\text{var}(b_l(f))}$$

and so distributed approximately as  $\chi_{L-1}^2$ . If one tests the hypothesis that the spectrum is well approximated by a constant plus linear term, the analysis is similar except that now a ‘trend’ is removed from the odd-order terms and covariances (38) must be computed. In either case a large remaining misfit in (41) would imply that the model is inadequate.

#### 4. Maximum-likelihood estimates

The problem addressed in this section is that of obtaining estimates of the spectral density that are, in some sense, approximately maximum likelihood. As will be seen in the following, an equivalent problem is finding a maximum-likelihood estimate of the matrix  $\mathbf{C}(f)$  described in the preceding section. If the series were replicated one could use  $\hat{\mathbf{C}}(f) = \text{ave } \hat{\mathbf{X}}(f) \hat{\mathbf{X}}(f)^\dagger$  but, because this estimate is singular when less than  $K$  series are available, additional constraints are necessary. In this section I describe a partial maximum-likelihood estimate of  $\mathbf{C}$  where the estimate is constrained to lie in the subspace spanned by the  $\mathbf{B}^{(l)}$ s.

If the quadratic inverse coefficients  $b_l(f)$  were restricted to be non-negative, equations (20) and (31) would be a standard form for components of variance (see Anderson 1973; Malley 1986; Rao & Kleffe 1988; Hocking *et al.* 1989). In this theory, also known as Eisenhart’s model II, or structured or patterned covariances, the observed covariance matrix is assumed to be a sum of known non-negative definite matrices with positive scalar coefficients. Thus, although this theory may be used as a guide, the quadratic inverse basis matrices are complex, are individually not definite (the odd-order  $\mathbf{B}$ s have symmetric positive and negative eigenvalues), and may have negative or, in cross-spectrum calculations, complex coefficients, so components-of-variance methods are not directly applicable. Notwithstanding, the least-squares solution developed in the preceding section is, in essence, the ‘fundamental linearization’ described in Malley (1986, ch. 3). This approach has the advantages of relative simplicity, of providing explicit quadratic estimators, of allowing interpretations as equivalent windows, and may be used to derive sampling properties not available for more complicated estimators. A major disadvantage of the expansion

$$\hat{S}(f \ominus \eta) = \sum_{l=0} \hat{b}_l(f) B_l(\eta)$$

is that it is not necessarily positive even though it is an approximate expansion of  $\hat{S}(f \ominus \eta)$ . The simplest explanation is that  $\hat{S}(f \ominus \eta)$  corresponds to an unconstrained projection of the rank one matrix  $\hat{\mathbf{X}}\hat{\mathbf{X}}^\dagger$  onto the full rank matrix  $\tilde{\mathbf{C}}$  so minor

perturbations from sampling to Gibb's phenomena or round-off occasionally give negative spectrum estimates. The problem of constraining spectrum estimates to be positive is old and has been approached in many ways. For a stationary process the autocovariance matrix is Toeplitz (Grenander & Szego 1984), and this is invoked implicitly in indirect methods. More recently, methods such as embedding an  $N \times N$  Toeplitz matrix in a larger circulant have been developed by Burg *et al.* (1982), Miller & Snyder (1987), Dembo *et al.* (1989), and Van Veen & Scharf (1990).

Here I propose a different approach and assume that the observed process may be decomposed into a set of band-limited processes each having bandwidth  $W$ . This assumption is prompted in part by the numerical solutions of the Karhunen–Loëve equations described in Thomson (1982), where it was observed that the eigenvectors appear much more like frequency shifted Slepian sequences than like the sines and cosines of standard asymptotic theory (Davenport & Root 1958). Additional stimulus for this approach is given by the orthogonal increment properties of the spectral representation and it is philosophically similar to Whittle's (1952, 1953) likelihood arguments (see Dzhaparidze (1986) for recent work). Another major motivation is, as in Cox (1975), the reduction in dimensionality with the goal of having 'all or nearly all the relevant information contained in the partial likelihood'. For time series, the approximate dimensionality of a time-bandwidth region was the subject of the famous series of papers by Slepian *et al.* (1961) with the results defined by the optimal set of prolate spheroidal wave functions. In practice a series may be decomposed into sub-bands if the dimensionality,  $K = 2NW$  of each sub-band is not too small, so that most of the  $\lambda$ s are close to unity (see Slepian 1976; Hannan & Thomson 1971). My experience with a range of spectra (not including their line components) is that the matrix  $Q$ , eq. (12.7) of Thomson (1982) is concentrated along the main diagonal with, as the Gershgorin bounds described there showed, small interaction terms.

For motivation recall that the Karhunen–Loëve expansion is an orthogonal decomposition of a finite sample of a random process with the property that the expansion coefficients are uncorrelated. As is well known (Davenport & Root 1958) given a sample of size  $N$  from a stationary series with a known autocovariance function  $R(\tau) = \mathbb{E}\{x_t x_{t+\tau}\}$  the discrete Karhunen–Loëve expansion is obtained by finding the orthonormal eigenvectors  $\hat{\psi}_n(t)$  of the Toeplitz matrix  $\hat{R}(i-j)$  and expanding the data on them. These expansion coefficients,  $\hat{c}_n$ ,

$$\hat{c}_n = \sum_{t=0}^{N-1} \hat{\psi}_n(t) x_t$$

have expected value  $\mathbb{E}\{\hat{c}_n\} = 0$  and covariances  $\mathbb{E}\{\hat{c}_n \hat{c}_j\} = \theta_n \delta_{nj}$  where  $\hat{\theta}_n$  is the corresponding eigenvalue. Thus for complex gaussian data the likelihood of the observation may be written

$$L(\{x\}) = \prod_{n=0}^{N-1} \frac{1}{\pi \hat{\theta}_n} \exp\{-|\hat{c}_n|^2/\hat{\theta}_n\}. \quad (42)$$

In Thomson (1982, §XII) it was shown that writing the Karhunen–Loëve eigenequation in the frequency domain gave

$$\hat{\theta}_n \hat{\Phi}_n(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{S}(\xi) \mathbb{D}_N(f - \xi) \hat{\Phi}_n(\xi) d\xi, \quad (43)$$

where the eigenfunctions,  $\hat{\Phi}_n(f)$  are the Fourier transforms of the eigenvectors  $\hat{\psi}_n(t)$ . These functions share the property of the Slepian functions of being both orthonormal on  $[-\frac{1}{2}, \frac{1}{2}]$  and also orthogonal with the spectrum as weight over the same interval.

Now consider the likelihood given by (42) as a functional of the spectrum; the usual calculus of variations methods applied to the likelihood equation (42) under perturbations on the spectrum gives  $\hat{\theta}_n = |\hat{c}_n|^2$  for  $n = 0, 1, \dots, N-1$  as conditions for the spectrum estimate  $\hat{S}(f)$  to be maximum-likelihood where the ‘hat’ denotes primary dependence on  $\hat{S}$  and not  $\{x\}$ . As in the papers cited above, such a maximum can be found by a variant of the EM algorithm (see Laird *et al.* 1987; Meilijson 1989) but such solutions are highly non-unique. A more desirable estimate is defined by maximizing the likelihood subject to the covariance matrix lying in the space spanned by the  $\mathbf{B}^{(l)}$ s. To do this first make the narrow-band approximation above, then truncate the maximum-likelihood conditions  $\hat{\theta}_n = |\hat{c}_n|^2$  to  $n = 0, \dots, K-1$ . Projecting  $\hat{C}(f)$  onto the space spanned by the  $\mathbf{B}^{(l)}$ s usually results in  $\tilde{C}(f)$  being of full rank but not necessarily positive. One should note that the space spanned by the  $\mathbf{B}^{(l)}$ s is restrictive; for example the coefficients  $b_l(f)$  are real for real spectra. In addition when  $l$  is even  $B_{jk}^{(l)}$  is real and symmetric with zero elements if  $j+k$  is odd; for odd  $l$  the  $B_{jk}^{(l)}$  are skew symmetric with elements that are pure imaginary when  $j+k$  is odd and zero when  $j+k$  is even. Thus the even parity elements of  $\tilde{C}$  are real, the odd parity elements imaginary, and the matrix hermitian. Expanding  $\hat{\Phi}_n(f)$  in a series of Slepian functions

$$\hat{\Phi}_n(f \oplus \xi) = \sum_{k=0}^{K-1} a_{nk}(f) \mathcal{V}_k(\xi)$$

that are a complete basis for this problem (43) becomes

$$\hat{\theta}_n a_{nk}(f) = \lambda_k \sum_{j=0}^{L-1} \int_{-W}^W \hat{S}(f-\zeta) \mathcal{V}_j(\zeta) \overline{\mathcal{V}_k(\zeta)} d\zeta a_{nj}(f).$$

The integral is (23) and the equation may be symmetrized by the usual multiplication and division by  $\sqrt{\lambda_k}$ . Replacing  $\hat{S}$  with the expansion (27) and letting  $\mathbf{A}_n(f)$  denote the vector  $[a_{n0}(f), \dots, a_{nK-1}(f)]^T$  gives

$$\hat{\theta}_n \mathbf{A}^{-1} \mathbf{A}_n = [\sum_{l=0} b_l(f) \mathbf{B}^{(l)}] \mathbf{A}_n$$

that we solve by the following variant of the EM algorithm:

1. Start with  $\hat{C}^{(0)} = \hat{X}(f) \hat{X}^\dagger(f)$  plus possibly a small diagonal term.
2. At the  $j$ th iteration,  $j = 1, 2, \dots$ , project  $\hat{C}$  onto the span of the  $\mathbf{B}^{(l)}$ s

$$\tilde{C}^{(j)} = \sum_{l=0} g_l^{-1} \text{tr}\{\hat{C}^{(j-1)} \mathbf{B}^{(l)}\} \mathbf{B}^{(l)},$$

3. Find the eigendecomposition of  $\tilde{C}^{(j)} = \mathbf{A}^{(j)} \boldsymbol{\Theta}^{(j)} \mathbf{A}^{(j)\dagger}$ , where  $\boldsymbol{\Theta}^{(j)}$  is the diagonal matrix of eigenvalues  $\theta_k^{(j)}$ .

4. Find the expansion coefficients  $c_k^{(j)} = \mathbf{A}_k^{(j)\dagger} \hat{X}(f)$ , and replace the eigenvalues with

$$\tilde{\theta}_k^{(j)} = \frac{1}{2}[\max(0, \theta_k^{(j)}) + |c_k^{(j)}|^2].$$

5. Generate an updated estimate of the covariance matrix,  $\hat{C}^{(j)}$  by the pseudo-spectral representation

$$\hat{C}^{(j)} = \sum_{k=0}^{K-1} \mathbf{A}_k^{(j)} \tilde{\theta}_k^{(j)} \mathbf{A}_k^{(j)\dagger}.$$

6. If the sample likelihood ratio (42) found using the  $cs$  and  $\theta$ s in step 5 have not stabilized return to step 2.

The basic method usually converges in a few iterations. The expansion of the spectrum implicit in the coefficients of the  $B^{(l)}$  matrices result in positive expansions of the spectrum using (27) and the time-domain autocovariance matrix is Toeplitz by construction.

## 5. ‘Lines’ and line frequency estimation

Much of the interest in climate records centres on detection of periodic components and estimation of their parameters. This problem has also been receiving renewed attention in several other areas, in part, from the recent emphasis on high-resolution spectrum estimation techniques, and also because the need for such methods exists. This is, however, not without danger for two reasons. First, the situation is much like that existing after the discovery of the periodogram. At that time the sampling properties of the periodogram were not well understood and, partly as a result, innumerable ‘periodic’ phenomena were ‘discovered’. Later investigation of these ‘discoveries’ gave both the periodogram and the search for periodicities a bad reputation. Second, because purportedly ‘high-resolution’ techniques are being used where they are poorly understood, it is likely that history will repeat itself. This danger is increased because these methods are frequently applied ‘because they are computationally efficient’ and without proper data analysis. A major objection to ‘conventional’ spectrum estimates is that they do not separate first and second moments so ‘line’ components are likely to be identified as any improbable looking bump in the spectrum. (This is a major argument against AR and Burg methods; such spectra often contain a mix of real and spurious lines without any objective criteria for telling them apart.)

In contrast, the multiple-window method provides a simple likelihood ratio test for periodic components. Like other multiple-window methods, this test is localized in frequency so inferences on lines spaced  $W$  or further apart are essentially independent. (Similarly, if line pairs or multiplets are separated from other groups by at least  $W$ , each group may be treated independently of the others.) In addition, this is a dual purpose statistic: first, for a known frequency it tests for the presence of a line component; second, the frequency where the test is maximized is an accurate estimate of line frequency. Because this test is based on familiar least-squares concepts it provides both the statistical significance tests and diagnostics common to such methods. In this section I describe the simple test, its peculiarities and some cautions on its use, the ‘integrated test’, and mention jackknifing such tests. This is followed by a description of the multiple-line test, comments on degree-of-freedom estimation for it, and some useful variants.

We now assume the process to have the harmonically centred stationary representation of Munk & Hasselmann (1964), that is the extended Cramér representation

$$x(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi\nu n} dX(\nu),$$

where now the orthogonal increment process has a non-zero first moment

$$\mathbb{E}\{dX(f)\} = \sum \mu_m \delta(f - f_m) df,$$

$\delta$  being the Dirac delta function and, replacing (5), the spectrum of the wide-sense stationary component of the process is the second central moment

$$S_c(f) df = \mathbb{E}\{|dX(f) - \mathbb{E}\{dX(f)\}|^2\}.$$

Examining the moments of the eigencoefficients one has

$$\mathbb{E}\{y_k(f)\} = \sum \mu_m V_k(f - f_m)$$

so, from the section on quadratic inverse theory, if the continuous portion of the spectrum  $S_c(f)$  does not vary too rapidly over  $(f - W, f + W)$  one finds for  $j, k < 2NW$

$$\text{Cov}\{y_j(f), \bar{y}_k(f)\} \approx S_c(f) \delta_{jk},$$

i.e. the different coefficients are approximately uncorrelated. Thus if the line frequencies were known, one could find the amplitudes by minimizing

$$\sum_{f, k} |y_k(f) - \sum \mu_m V_k(f - f_m)|^2.$$

Proceeding in the usual way, this minimization results in a matrix that, if the different line frequencies are separated by at least  $W$ , is strongly diagonal. Thus, as a result of the energy concentration properties of the Slepian functions, the different coefficients may be estimated independently of each other. To illustrate this estimation procedure consider the simplest procedure for a single isolated line. Since  $\mathbb{E}\{y_k(f_0)\} = \mu V_k(0)$  one may estimate  $\mu$ , given  $f_0$ , by ordinary least-squares. The basis of the procedure is to test for the presence of a line component at frequency  $f$  using a generalized likelihood ratio test,

$$\max L(\{X\} | \mu) / L(\{X\} | \mu = 0),$$

that, under a gaussian assumption, becomes the variance ratio test.

Given  $\mathbf{Y}$ , defined to be the vector of eigencoefficients, and  $\mathbf{V}$ , the vector of Slepian functions I minimize the mean square error between the observed coefficients and their expected values

$$\begin{aligned} e(f) &= \|\mathbf{Y}(f) - \hat{\mu} \mathbf{V}(f - f_0)\|^2 \\ &= \sum_{k=0}^{K-1} |y_k(f) - \hat{\mu} U_k(f - f_0)|^2 \end{aligned} \quad (44)$$

at the single frequency  $f_0$ , resulting in a simple linear regression estimate

$$\hat{\mu}(f) = \mathbf{V}^\dagger(0) \mathbf{Y}(f) / \|\mathbf{V}(0)\|^2. \quad (45)$$

Like other least-squares regression procedures one may test this one for significance by use of a standard  $F$ -test. Comparing the sum-of-squares explained by the regression to the residual sum-of-squares, both normalized by their respective degrees-of-function, one obtains

$$F(f) = (2K - \nu) |\hat{\mu}(f)|^2 \|\mathbf{V}(0)\|^2 / (\nu \|\mathbf{Y}(f) - \hat{\mu}(f) \mathbf{V}(0)\|^2), \quad (46)$$

where the constants provide proper scaling for  $F$  with  $\nu$  and  $2K - \nu$  degrees of freedom. For testing at a known frequency  $\nu$  is 2, but if frequency is estimated as well  $\nu = 3$ .

### 5.1. Comments on the harmonic $F$ test

Clearly the line amplitude estimation procedure and associated  $F$  test described in equations (46) and (47) apply only to a single isolated line. If several lines are present and separated in frequency by at least  $W$  then the estimation problems are essentially independent by virtue of the energy concentration properties of the Slepian functions. Empirical evidence suggests further that the  $F$ -test works well down to the Rayleigh resolution,  $1/N\delta t$ , but if lines are spaced closer in frequency than this limit the test rapidly breaks down. With real data one must remember that lines come in pairs at positive and negative frequencies with conjugate coefficients and also that a non-zero average acts like a line at zero frequency. Note that despite the attention given to ‘zero-mean processes’ the ordinary average is no more distinguished than any other frequency except, with real processes, the spectrum is symmetric about the frequency origin; with complex processes the average is completely undistinguished.

The frequency for which  $F$  is a maximum is a good estimate of  $f_0$  for isolated lines. Under the usual assumption of moderate local signal-to-noise power ratio it may be shown that the variance of this estimate is within a few percent of the Cramér–Rao bound, specifically for (46)

$$\text{var}\{\hat{f}\} = 6\Xi_K^{-1}(2\pi T)^{-2}S_c(f)/S_A(f), \quad (47)$$

where  $\Xi_K$  is the efficiency factor defined in §VII of Thomson (1982),  $f$  is frequency in units of cycles per time unit,  $T$  is the total time span of the data,  $S_c(f)$  is the continuous background (or noise) power spectral density and  $S_A(f)$  is the apparent (periodogram) spectral density of the line component. In particular, for a line of the form  $a \cos(2\pi ft + \phi)$

$$S_A(f) = \frac{1}{4}a^2T,$$

$S_A(f)/S_c(f)$  is the local signal-to-noise ratio of the line in question and the given observation time; except in trivial cases it bears no relation to the total power in the signal. Formula (47) is, except for the factor  $\Xi$ , the Cramér–Rao bound (see Rife & Boorstyn 1976; Walker 1971; Rice & Rosenblatt 1988). For the time-bandwidth products used here the estimator efficiency,  $\Xi_K$  is between 80 and 95% so with  $T = 780000$  years and a signal-to-noise ratio of 30 one obtains standard deviations of frequency estimates of ca.  $\pm 10^{-7}$  c a $^{-1}$ <sup>†</sup>, that is about one tenth of the Rayleigh resolution. The following points should be noted.

1. When computing  $F$  using an FFT mesh one must ‘pad’ the tapered series with lots of zeros. The reason is that the  $F$  line shape is so narrow on strong periodic signals that if  $\Delta f \approx 1/N$  there is a high probability of the line frequency being far enough from the closest mesh frequency to be missed. If you are interested in line spectra, pad the series to  $4N$  to  $10N$  with zeros, use an FFT as an initial screen, then use a single frequency Fourier transform to refine the frequency estimate. It is usually best to jackknife here so that both a frequency estimate and a tolerance are obtained. See Gentleman (1969) for a good single-frequency Fourier transform algorithm.

2. Remember that typical time series have many observations. With 10000 data points one expects to get one ‘false alarm’ at the 99.99% point! Such a result will

<sup>†</sup> c a $^{-1}$  is here defined as the unit for cycles per year; c Ma $^{-1}$  is the unit for cycles per million years.

not, usually, justify a paper. (On the other hand the difference between the 90 and 99.9% points is not that large: with 2 and 8 degrees of freedom the 90% point is at 3.01, 99% at 8.65, and 99.9% is at 18.5. Thus if you see an  $F$  value of 100 you probably have something.)

3. For most series I plot  $F(f)$  as a separate figure (see the examples) but to avoid the inconvenience of both an  $F$  plot and a plot of the residual spectrum (where the mean values of the significant coefficients have been subtracted) I reshape the spectrum. To do this: first, subtract the effect of the significant line from the expansion coefficients, that is replace  $x_k(f)$  with  $x_k(f) - \hat{\mu}(f_0) V_k(f-f_0)$  near lines. Secondly, compute the residual spectrum; thirdly, include the estimated line components in the spectrum. In the last operation I add a portion with shape given by the  $F$  line shape (out to the first local minimum each side of the peak) scaled in amplitude to preserve power. The width of the line is thus proportional to the frequency uncertainty of the estimate.

4. The single-line  $F$  test can fail on multiple lines when they are spaced less than  $W$  apart. A roughly rectangular ‘bump’ in the residual spectrum is a good clue to unresolved lines. In the examples, the spectrum shown in figure 2 was computed without removing lines; the region around  $24 \text{ c Ma}^{-1}$  is typical.

5. With complicated phenomena (such as meteorological series) the  $F$  plot is often bewildering. Here a  $\chi^2$  test on peak significance levels seems helpful. Also mechanized searches for harmonics, sidebands about multiples of the annual cycle, etc., appear to be necessary adjuncts.

### 5.2. The integrated $F$ -test

Among the variants proposed in Thomson (1982) was the ‘integrated  $F$ -test’ where, instead of using just the eigencoefficients at frequency  $f$  and the  $V_k(0)$ s used in equation (45) one minimizes the integral of (44) over the band  $(f_0-W, f_0+W)$ . Both estimates and a range of intermediate forms may be covered by including a weight function  $w$  in the integral

$$\int_{f_0-W}^{f_0+W} \sum_{k=0}^{K-1} |y_k(f) - \hat{\mu} V_k(f-f_0)|^2 w(f-f_0) df,$$

so the point estimate corresponds to  $w(f) = \delta(f)$  and the integrated form to  $w(f) = 1$ . An obvious effect of the integration is to increase the bandwidth of the estimate and when  $w(f) = 1$  the effective frequency band extends from  $f_0-2W$  to  $f_0+2W$  instead of the nominal  $f_0 \pm W$  of the point estimate. The effective sidelobes of the integrated estimate, however, are low. Second, regarding the problem purely in regression terms, one would be remiss if considerations of leverage, outliers, and the equivalent ‘hat’ or projection matrix were ignored. For the point weighting the projection matrix has diagonal elements  $U_j^2(0)/\sum_{k=0}^{K-1} U_k^2(0)$  and so is zero for odd  $j$ . Similarly, for frequency estimation, the projection matrix vanishes for  $j$  even. The integrated form does not have such an obvious matrix interpretation but, with uniform weighting, the different coefficients all have approximately equal influence.

### 5.3. Jackknifed estimates of line parameters

In §5.1 a Cramér–Rao bound for the variance of a frequency estimate was given that applies when the signal-to-noise power ratio is above threshold. However, at low values of signal-to-noise ratio frequency estimates are susceptible to large errors

(Rife & Vincent 1970). The problem is that one usually does not know the signal-to-noise ratio *a priori* and a safer procedure than assuming formula (47) to be valid, is to jackknife the frequency estimation over windows. Using a ‘differential update’ on the regression problem, i.e. after solving the problem on  $[0, T]$  finding the change in solution associated with changing the sampling times so the data block becomes  $[dT, T+dT]$ , one may show that Rice’s (1948) results on frequency modulation apply here. Considering the estimators as functions of the starting time of the block, one finds that it is common for a pseudo-frequency to have a ‘click’ (i.e. a  $2\pi$  phase discontinuity with an associated unbounded frequency error), but the case where such behaviour occurs simultaneously (or even nearly so) in two or more pseudo-frequencies appears to be extremely rare. As usual, the pseudo-values can be used to estimate the variance of the frequency estimate thus warning when the estimate is unreliable. Tolerances computed in this way agree reasonably with the usual high signal-to-noise ratio formulae. An alternative is to estimate the signal-to-noise power ratio directly from the data by conventional methods and then apply the formula, but this approach is not as reliable both because estimates of signal-to-noise ratios are excessively variable and formula (47) is invalid near threshold. Since the peak  $F$  value is an estimate of the signal-to-noise power ratio (SNR), the same jackknifed approach as above may be applied to it and results in a much more stable estimate of SNR. Because the signal-to-noise ratio has a non-central  $F$  distribution the jackknife is much more effective in this application if the different  $F$  estimates are log-transformed before recombining. In the examples, see table 1, standard deviations of estimated frequencies are made by both methods and agree reasonably.

## 6. The multiple line $F$ -test

A standard data analysis problem is the detection of multiple lines at known frequencies; as an example, the 41000 year obliquity ‘line’ discussed in the examples is a group of four closely spaced lines. Also, both the 18 ka and 23 ka precession ‘lines’ are similarly split, but, in light of §9, must be treated specially. A controversial area of climatology, that of detecting the 8.85 and 19.6 year lunar cycles, is similar but difficult because with many series both the positive and negative parts must be treated together with the process mean and also because of other known or suspected cycles in the data. Known periods also frequently show up as ‘sidebands’ or modulation products on harmonics of the annual cycle and so the problem becomes one of testing for the presence of suspected line components in the presence of known lines. Our basic approach is to use a generalized likelihood-ratio test. Currently, because the statistics of the process are unknown, I use a narrow-band expansion in Slepian sequences instead of a Karhunen–Loëve expansion but in view of §3, anticipate replacing this with a mixed components of variance model. Also, because the spectrum of geophysical data is often complicated near strong lines, we use ordinary least-squares instead of generalized (or gaussian multivariate theory) least squares but decrease the effective number of degrees of freedom to include the window overlap effects.

The test we develop is a ‘partial  $F$ ’ test. Such tests are described, for example, in Draper & Smith (1981, p. 101) and, in the case considered here, are closely related to  $C_p$  (Mallows 1973). Formally we wish to test the hypothesis that the signal consists of only the  $J$  periodic components  $f_1, f_2, \dots, f_J$ , against the hypothesis that the signal

Table 1. Estimated periodic components, core V22-174  
 (Time step is 4930 years, standardized to O1 line.)

no.	ID <sup>a</sup>	$F_{\min}$ <sup>b</sup>		$F_{\text{all}}$ <sup>c</sup>		$F_{\max}$ <sup>b</sup>		FTSD <sup>e</sup>		$P_{\min}$ <sup>b</sup>		$P_{\text{all}}$		$P_{\max}$		$P_{\text{ta}}$		phase		$F$ -test <sup>f</sup>
		e Ma <sup>-1</sup>	e Ma <sup>-1</sup>	e Ma <sup>-1</sup>	e Ma <sup>-1</sup>	e Ma <sup>-1</sup>	e Ma <sup>-1</sup>	years	years	years	years	years	years	years	years	years	deg	deg		
1	E1	1.2503	1.3585	1.4968	0.165	0.269	668057.0	736088.0	799748.0	412885	4.80 ± 1.3	28 ± 28	7.0							
2	E2	10.3999	10.5135	10.5851	0.140	0.172	94471.5	95115.2	96154.4	94945	16.72 ± 2.7	137 ± 28	11.6							
3	O3	18.7176	18.8349	18.9102	0.139	0.229	52881.2	53092.6	53425.4	53615	3.51 ± 2.4	87 ± 29	9.9							
4	O1	24.2692	24.3871	24.5025	0.155	0.084	40812.1	41005.1	41204.4	41000	7.14 ± 0.8	-113 ± 26	115.4							
5	E	29.0706	29.1800	29.2925	0.152	0.211	34138.4	34269.9	34398.9	34061	3.06 ± 1.8	-44 ± 21	8.7							
6	$P^2$	31.1553	31.3774	31.7357	0.449	0.283	31510.2	31870.0	32097.2	31817	1.92 ± 1.6	156 ± 87	4.0							
7	$P^4$	37.5416	37.5984	37.6678	0.089	0.194	26547.8	26596.8	26637.0	26675	2.53 ± 0.8	31 ± 23	10.0							
8	$P1$	42.4797	42.7107	42.8959	0.330	0.204	23312.2	23413.3	23540.6	23716	4.01 ± 1.5	168 ± 47	10.1							
9	$P2$	44.7336	44.9388	45.3546	0.462	0.256	22048.4	22252.4	22354.5	22428	3.94 ± 3.2	34 ± 81	7.2							
10	$P3$	52.7415	53.1865	53.2671	0.430	0.213	18773.2	18801.7	18960.3	18976	2.89 ± 1.7	93 ± 109	7.5							
11	$P15$	67.3649	67.5079	67.7063	0.233	0.214	14769.6	14813.0	14844.5	15025	1.72 ± 0.8	158 ± 30	8.1							
12	$P20$	78.5905	78.7149	78.9898	0.286	0.213	12659.8	12704.0	12724.1	12972	1.12 ± 0.6	-130 ± 53	14.5							
13	$P4$	86.7567	86.8279	86.9440	0.146	0.228	11501.6	11517.0	11526.4	11507	1.12 ± 0.3	-113 ± 38	12.2							

<sup>a</sup> Tentative identification with tables in Berger (1979).  $P^2$  and  $P^4$  are squared and fourth powers of precession.

<sup>b</sup> Maximum and minimum delete-one jackknife estimates.

<sup>c</sup> Estimate using all windows.

<sup>d</sup> Jackknife estimate of standard frequency deviation from trimmed  $F$ .

<sup>e</sup> Estimated frequency standard deviation from trimmed  $F$ .

<sup>f</sup> Maximum  $F$  with two and eight degrees of freedom on delete-one and full estimates.

consists of  $J+1$  components  $f_1, f_2, \dots, f_J, f_{J+1}$ . We begin with an expansion of the data,  $x_t$ , on a fine grid of frequencies with  $M$  mesh points  $g_m$ ,  $m = 1, 2, \dots, M$  where, typically, the  $f_s$  are a subset of the  $g_s$

$$y_k(g_m) = \sum_{t=0}^{N-1} x_t v_t^{(k)} \exp(-i2\pi g_m t).$$

Collect the eigencoefficients at the different frequencies into a column vector,  $\mathbf{Y}$  of length  $MK$

$$\mathbf{Y} = [y_0(g_1), y_1(g_1), \dots, y_{K-1}(g_1), y_0(g_2), y_1(g_2), \dots, y_{K-1}(g_M)]^T. \quad (48)$$

Under  $H_0$  the matrix of explanatory variables has  $J$  columns, one for each of the assumed line frequencies

$$\mathbf{X}_J = \begin{bmatrix} V_0(g_1 - f_1) & V_0(g_1 - f_2) & V_0(\cdot) & V_0(g_1 - f_J) \\ V_1(g_1 - f_1) & V_1(g_1 - f_2) & V_1(\cdot) & V_1(g_1 - f_J) \\ \vdots & \vdots & \vdots & \vdots \\ V_{K-1}(g_1 - f_1) & V_{K-1}(g_1 - f_2) & V_{K-1}(\cdot) & V_{K-1}(g_1 - f_J) \\ V_0(g_2 - f_1) & V_0(g_2 - f_2) & V_0(\cdot) & V_0(g_2 - f_J) \\ V_1(g_2 - f_1) & V_1(g_2 - f_2) & V_1(\cdot) & V_1(g_2 - f_J) \\ \vdots & \vdots & \vdots & \vdots \\ V_{K-1}(g_M - f_1) & V_{K-1}(g_M - f_2) & V_{K-1}(\cdot) & V_{K-1}(g_M - f_J) \end{bmatrix}$$

so that

$$\mathbf{Y} = \mathbf{X}_J \boldsymbol{\mu}_J + \mathbf{R}_J.$$

For purposes of hypothesis testing, the quantity of primary interest is the residual sum of squares  $R_J^{ss} = \mathbf{R}_J^\dagger \mathbf{R}_J$ . Similarly, under  $H_1$ ,  $\mathbf{X}$  has  $J+1$  columns and one again computes the residual sum of squares  $R_{J+1}^{ss}$ . The energy explained by the introduction of the extra line is thus  $R_J^{ss} - R_{J+1}^{ss}$ , which depends on only two degrees of freedom. Assuming  $v$  degrees of freedom for the composite mesh there are thus  $v - 2(J+1)$  for  $R_{J+1}^{ss}$  so the variance ratio statistic is

$$F = \frac{1}{2}[R_J^{ss} - R_{J+1}^{ss}] / [R_{J+1}^{ss}/(v - 2(J+1))].$$

A standard estimate of the degrees-of-freedom for the composite mesh is given in the next section but one should have about  $2(K + 2N(f_M - f_1))$ , or just  $2BT$ , where the bandwidth is that of both the mesh and the windows. I emphasize that here  $f_{J+1}$  has been treated as a known parameter; in the common case where one also estimates  $f_{J+1}$  an additional degree of freedom must be allowed for it.

#### Notes on Multiple-Line Detection.

1. In my experience the most common application of the multiple-line test is for two suspected lines with one of them at a known frequency. For this problem one should not expect results similar to those obtained with a single isolated line; the Cramér–Rao bounds for the different parameters tend to degrade badly when the line spacing becomes less than twice the Rayleigh resolution (Rife & Boorstin 1976; Messer & Adar 1989). Note that while one frequency may be known exactly, the corresponding amplitude and phase are usually unknown. Moreover, simple numerical conditioning also degrades substantially with frequencies spaced less than Rayleigh.

2. Multiple, i.e. more than two, line tests degrade further and should be treated with extreme caution. I do not advise using these if the line frequencies are unknown.

Note, however, that when one has a set of known frequencies the problem of selecting the subset of ‘significant’ lines is the same as selecting a significant subset of explanatory variables from a set of possibilities in ordinary regression problems. A standard method for this is Mallows’s (1973)  $C_p$  statistic (see also Thompson 1978), remembering that with lines (except for the zero frequency term) each one counts for two degrees of freedom.

### 6.1. Degrees of freedom for non-orthogonal basis

A recurring problem in spectrum estimation is that of estimating the degrees of freedom (DF) of a particular estimate. In simple situations, when the different estimates are from independent samples, when the data tapers are orthogonal, etc., one simply counts the estimates. In general, however, one works with overlapped segments, and with estimates closely spaced in frequency and the situation is less simple. Corresponding to (48) above I consider the products of the  $K$  different tapers times the  $M$  corresponding complex exponentials

$$B_{nl} = \{v_n^{(k)} \exp(i2\pi ng_j)\},$$

where  $l$  indexes  $k$  and  $j$  as in (48) as a set of  $P = KM$  non-orthogonal complex basis vectors ‘concentrated’ near the same frequency that I write as a  $N \times P$  matrix  $\mathbf{B}$ . Taking  $x$  to be a typical realization from a stationary, complex white gaussian process of variance  $S$ , unnormalized spectrum estimates  $\hat{S}_j = |\mathbf{x}^\dagger \mathbf{B}_j|^2$  will be individually distributed as  $\chi^2_2$  with  $E\{\hat{S}_j\} = S \mathbf{B}_j^\dagger \mathbf{B}_j$ . Denoting by  $\mathbf{R} = \mathbf{B}^\dagger \mathbf{B}$  the grammian (or ‘covariance’) matrix of the basis, the average of the  $S_j$ s has expected value  $E\{\bar{S}\} = S \text{tr } \mathbf{R}$ . If the different  $S_j$ s were uncorrelated and the diagonal elements of  $\mathbf{R}$  equal,  $\bar{S}$  would have a  $\chi^2$  distribution with  $2P$  degrees of freedom. Because of the non-orthogonality of the basis vectors, the variance will be larger than expected for uncorrelated data so the effective degrees of freedom less than  $2P$ . Using gaussian theory for the fourth moments, the variance of  $\bar{S}$  is

$$\sigma_s^2 = \frac{S^2}{\text{tr}^2 \mathbf{R}} \sum_{j=1}^P \sum_{k=1}^P |\mathbf{B}_j^\dagger \mathbf{B}_k|^2.$$

Recalling that a  $\chi^2_\nu$  variate has mean  $\nu$  and variance  $2\nu$ , one obtains the effective degrees of freedom

$$\hat{\nu} = 2 \text{tr}^2 \mathbf{R} / \text{tr} \mathbf{R}^2 \quad (49)$$

by matching moments. Note that this formula depends only on the basis properties of the windows and is valid when the noise spectrum is constant within the bandwidth of interest.

### 6.2. Linear FM model

Because perfectly periodic signals rarely exist and linearly modulated or ‘chirped’ signals are common, one must include departure from periodicity in line tests. Given that a signal is almost periodic, a useful variant is to assume a linear frequency drift of the form  $x(t) = \mu \exp((i\omega_0 + \alpha t)t)$ . Because we are interested in almost periodic phenomena we assume  $\alpha$  small and expand in a Taylor’s series,

$$(\partial x(t)/\partial \alpha)|_{\alpha=0} = i\mu t^2 e^{i\omega_0 t}.$$

Examining the expansion coefficients, the  $t^2$  term becomes a second derivative

$$x_k(\omega) \approx \mu U_k(\omega - \omega) - i\mu\alpha U_k''(\omega - \omega).$$

Recall the differential equation for the discrete prolate spheroidal wave functions (Slepian 1978, eq. (17))

$$\frac{d}{d\omega}[\cos \omega - \cos 2\pi W] \frac{dU_k}{d\omega} + [\frac{1}{4}(N^2 - 1) \cos \omega - \theta_k] U_k = 0,$$

where  $\omega = 2\pi f$ , so that for  $k$  even

$$U''_k(0) = (\theta_k - \frac{1}{4}(N^2 - 1)) U_k(0) / (1 - \cos 2\pi W)$$

and  $U''_k(0) = 0$  for  $k$  odd by symmetry. Thus one has

$$x_k(\omega) \approx \mu U_k(0) - i\mu\alpha U''_k(0).$$

One now proceeds as usual but now estimates both  $\mu$  and  $\mu\alpha$  noting that, as with the usual single-frequency  $F$  test, the parameters are determined solely by the even coefficients. This simple technique works reasonably for frequency drifts over the window length up to about half the Rayleigh resolution.

A similar method works both for refining frequency estimates from a mesh point and for estimating attenuation again assuming the changes from simple periodicity are small. Inference from a short sample of apparently attenuating sinusoidal data must, in any case, be treated carefully (see Park *et al.* 1987b; Bolt & Brillinger 1979; Dahlen 1982; Walden 1990a; Zhu *et al.* 1989). An obvious alternative explanation to an attenuating sinusoid is that two closely spaced frequencies are present. Orthogonalizing  $e^{i2\pi ft}$  and  $e^{i2\pi(f+\delta)t}$  for small  $\delta$  by a Gram–Schmidt procedure gives  $e^{i2\pi ft}$  and  $t e^{i2\pi ft}$  as unnormalized basis functions. Using the same argument as above, the  $t$  in the second basis corresponds to a first derivative term and so appears on odd order coefficients instead of on the even order coefficients. Thus attenuation and fine frequency splitting are detectable but difficult to distinguish in short samples.

## 7. Estimation of coherences and transfer functions

Many of the interesting and controversial questions in time series analysis occur when examining the relations between two or more time series. There are legitimate reasons for controversial results being obtained; usually the questions being asked are difficult, often the available data limited or a proxy, and the physics of the problem may be poorly understood so that little theoretical guidance on the choice of allowable models is available. When analysing such time series it is well to remember that nature can be amazingly complex and that many of the theoretical constructs used in stochastic process theory, for example linearity, ergodicity, normality, and particularly stationarity, are mathematical fairy tales. For example there are, by virtue of the ‘big bang’ if nothing else, no stationary time series in the strict mathematical sense. Thus, although it is necessary to know the theory of stationary processes, one should not adhere to it dogmatically when analysing data from physical sources, particularly when the observations span an extended period (see, for example, Larsen 1980; Chave *et al.* 1987; Jones *et al.* 1989; Egbert & Booker 1989; Egbert 1989; Chave & Thomson 1989; Lanzerotti *et al.* 1986).

In addition to the above legitimate reasons for obtaining controversial results, many appear to result from ignoring the guidance of the ‘stationary gaussian process’ fairy tale and the use of methods that violate Einstein’s maxim, ‘as simple as possible but not simpler’. Perhaps the worst, and most common, example is to claim association between two series by simply plotting them together as a function of time; see the example cited by Eddy (1983). Such plots are a useful tool but cannot

be used in place of the usual statistical tests and diagnostics. When analysing time series problems one must remember that almost all the methods, and problems, of multivariate analysis are transferable; the major difference is that in time series almost everything is a complex function of frequency and that forms of estimates beyond those suggested by ordinary multivariate theory are sometimes required (see Preisendorfer & Mobley 1988; Mooers 1973).

### *Estimates of coherence*

The simplest such analogy with ordinary multivariate problems is the coherence, the frequency-domain counterpart of the ordinary correlation coefficient (see Carter 1987). Intuitively, if one filters two series through narrow band filters, samples their outputs at the appropriate Nyquist rate, and then computes a complex correlation coefficient between the two sets of samples one has an estimate of coherence. This estimate is obviously a function of the centre frequency of the filters and, as the filters I consider are ‘one sided’, that is the centre frequency is  $f$ , the filter passes the band from  $f-W$  to  $f+W$  but not the band centred on  $-f$ , the output of the filter is complex. Note that coherence is primarily an exploratory tool useful, for example, in deciding if two series are related or effectively independent. Because it is complex it has both magnitude and phase with the phase conveying information on the relative delays between the two series; specifically the phase slope is an estimate of envelope (or group) delay. Also, like the ordinary correlation coefficient, coherency estimates are subject to outliers, high leverage points, and to the local mean value function.

Assume now that we have two series  $x_t$  and  $y_t$ ,  $t = 0, 1, \dots, N-1$  and wish to compute an estimate of coherence. Temporarily assuming that there are no average or line components, we begin by isolating the energy in a band  $(f-W, f+W)$  and express it as an orthogonal expansion of Slepian functions. As before, we assume that the two series have a bivariate Cramér representation and that the observable component of this representation is

$$d\hat{X}(f \oplus \nu) = \sum_{k=0}^{K-1} \hat{x}_k(f) \mathcal{V}_k(\nu) d\nu \quad \text{and} \quad d\hat{Y}(f \oplus \nu) = \sum_{k=0}^{K-1} \hat{y}_k(f) \mathcal{V}_k(\nu) d\nu$$

for  $|\nu| < W$ . In this section, as in §3, I assume that the eigencoefficients,  $\hat{x}_k(f)$ , etc., are the best available estimates of these coefficients; that is they have been adaptively weighted or have had sidelobe effects subtracted and are not simply the raw estimates given by (9).

Given these estimates one can form estimates of the spectra and cross-spectra as

$$\hat{S}_{xx}(f \oplus \nu) = \frac{1}{N} \left| \sum_{k=0}^{K-1} \hat{x}_k(f) \mathcal{V}_k(\nu) \right|^2.$$

Simply averaging over the inner bandwidth gives, for example, a cross-spectrum estimate of the form

$$\hat{S}_{xy}(f) = \frac{1}{N} \sum_{k=0}^{K-1} \hat{x}_k(f) \bar{\hat{y}}_k(f) \quad (50)$$

that has been used as a basis for the usual multiple-window estimate of coherence (Thomson 1982; Park *et al.* 1987a; Kuo *et al.* 1990; Vernon 1989; Schreiner & Dorman 1990). Walden (1990b) has shown that this estimate is superior to

conventional coherence estimates. None the less, if the relative phase between the two processes varies significantly across the inner band, (50) can be badly biased down.

Now assume that  $y$  represents an ‘output’ and  $x$  an ‘input’ process, and that the two are related by a transfer function  $H(f)$ , as in Hannan (1970), Brillinger (1975), Shumway (1988) (but distinct from the meaning in Sachs *et al.* (1977)). Then one has

$$d\hat{Y}(f+\nu) = H(f+\nu) d\hat{X}(f+\nu). \quad (51)$$

Multiplying by  $\overline{\mathcal{V}_k}(\nu)$  and integrating over the inner frequency interval gives

$$\hat{y}_k(f) = \sum_{j=0}^{K-1} \hat{x}_j(f) \int_{-W}^W \mathcal{V}_j(\nu) \overline{\mathcal{V}_k}(\nu) H(f+\nu) d\nu \quad (52)$$

or

$$Y(f) = H(f) X(f).$$

Now expand the transfer function  $H$  in a Taylor series assumed to converge in the inner region  $(f-W, f+W)$

$$H(f+\nu) \approx H(f) + H'(f) \nu + \dots \quad (53)$$

and define the  $K \times K$  matrix  $D$  as a special case of the integral in (52)

$$(D)_{jk} = \int_{-W}^W \mathcal{V}_j(\nu) \overline{\mathcal{V}_k}(\nu) \nu d\nu$$

so equation (51) becomes

$$Y(f) \approx H(f) X(f) + H'(f) D X(f),$$

just an ordinary linear regression equation for  $H$  and  $H'$ . In the examples second estimates of  $H$  are computed by simply integrating the estimate of  $H'$ . In doing such calculations one must be aware of the usual regression problems and familiarity with Scheffé (1959), Muirhead (1982), or Seber (1984) is advisable. Regression diagnostics, particularly those checking for leverage, as in Belsley *et al.* (1980), Cook & Weisberg (1982), or Chatterjee & Hadi (1988) are especially important near the origin and periodic components.

### 6.3. An improved estimate of coherence

An obvious extension of this method is to the problem of estimating coherence. With the standard estimate (50) problems arise when the effective transfer function across the inner band varies rapidly, Cleveland & Parzen (1975), so that the matrix  $H$  is significantly non-diagonal. Consider a matrix  $F$  composed of the two  $K \times 2$  sub-matrices

$$F_1 = [X, DX] \quad \text{and} \quad F_2 = [Y, DY]$$

with

$$F = [F_1, F_2] \quad (54)$$

and the complex canonical correlations between  $F_1$  and  $F_2$ . As is well known if one partitions the covariance matrix of  $F$

$$F^* F = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

then the canonical correlations are given by the eigenvalues of

$$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}. \quad (55)$$

Information on the relative scale of the derivative terms is obtained from the eigenvectors implicit in the construction of the canonical correlations, and significance can be tested by the usual methods as discussed in Lawley (1956, 1959), Glynn & Muirhead (1978), or Muirhead (1982). Miyata (1970) gives a detailed description of complex canonical correlations. The canonical correlation form is also obtained for transfer function problems if in (53) the Taylor's series is replaced by the rational form  $(a + bv)/(1 + cv)$ . Variants such as using the  $B_t(v)$  in place of the Taylor series, local compound smoothers in place of the simple integration mentioned above, and remote reference methods are being studied. Finally, because these forms use the basis properties of the Slepian functions explicitly and assume structure within the inner band, jackknifing estimates such as (55) over windows is presently not well understood and probably is inadvisable.

## 8. The data

As examples I use two sets of records of the relative concentrations of the stable oxygen isotopes,  $^{18}\text{O}$  and  $^{16}\text{O}$ , measured in foraminifera skeletons from deep sea cores. The measured oxygen concentrations reflect two correlated variables: firstly, the ambient isotopic ratio, controlled largely by global ice volume; secondly, the temperature when the organism was growing. The latter effect is about 20% of the former and tends to be positively correlated so that data from benthic foraminifera is used as a sensitive proxy for global ice volume, see the papers in the two volume set edited by Berger *et al.* (1984). Because biological uptake is sensitive to both isotopic concentration and temperature, data from planktonic forams may be used to recover local sea surface temperature. To study sensibly such data it is necessary to be familiar with the Milankovitch theory; changes on timescales from 10000 to 500000 years in the Earth's orbit largely control long-term behaviour of the global climate. A good introduction is provided by Hays *et al.* (1976); many current results are in the journal *Paleoceanography* (see Arthur & Garrison 1976). The celestial mechanical calculations for this problem have been refined by Berger (1977, 1978a, b) and he has also written an extensive review (1988) of Milankovitch theory. Note that the orbital parameters split into three major groups: obliquity, or tilt of the Earth's axis, dominated by an almost periodic component with a 41000 year period; precession, or orientation of the axis, split into two bands concentrated around 19000 and 23000 year periods; and finally, eccentricity of the orbit that changes on 100000 and 400000 year timescales. None of these motions are simple periodic functions but do have non-harmonic trigonometric series representations; see Berger (1978a) for expansions and tables of the dominant frequencies. The reader should also note that although Milankovitch theory is attractive, it is not universally accepted (see Evans & Freeland 1977) and considerable scepticism exists that the relatively tiny changes in orbital parameters (for example, the tilt of the axis changes by less than one degree) are capable of causing glaciation at the latitude of New York City. Consequently, numerous other mechanisms have been proposed and it does appear that terrestrial responses modify the apparent amplitudes of the different lines considerably. Much work has been done on modelling this dependence (see Imbrie *et al.* 1979; Imbrie & Imbrie 1980; Le Treut & Ghil 1983; Hyde & Peltier 1987; Oglesby & Park 1989).

The cores are sampled as a function of depth, and interpolated to approximately equal spacing in time using available geological information to establish a time base.

Since the presence of several of the strongest lines have been established many times over, and beyond any reasonable doubt, the ‘known’ frequencies of these components are further used to fine tune the timescale. Details of this process are available in Imbrie *et al.* (1984), Kominz *et al.* (1979), Morley & Hays (1981), Shure & Chave (1984), Press *et al.* (1986), and Schiffelbein & Dorman (1986). However, because of the shifts in the precessional frequencies to be discussed below (and possibly others), the present timescale must be considered interim. Similarly, the quality of the data has now reached a level where interpolation procedures previously considered adequate are becoming marginal. Finally, the analysis is iterative in that the spectrum estimates obtained from the best current data can be used to improve understanding of the process and so to refine timescales and interpolation procedures. This, in turn, leads to a revision of the data and so on. It is, of course, recognized that such a procedure is somewhat circular and the possibility of generating ‘lines’ at the presumed frequencies by ‘tuning’ the timescale has been studied by Pisias & Moore (1981). Given the constraints on the process (for example a legitimate timescale cannot change the ordering of the data and must agree with known geological and radioactive isotope time markers within acceptable bounds (Prell *et al.* 1986; Shackleton *et al.* 1987, 1988) it appears to be difficult to generate even one ‘strong’ line, let alone a process of the complexity seen in real cores. Moreover, one is not tuning to a single line but to a narrow-band process and one of the most convincing features of the process is the agreement in both amplitude and phase obvious in fig. 9 of Imbrie *et al.* (1984). Given these cautions, the data will be used with the existing published timescales and interpolations and consequences of the analysis on the data and timescales left to future work. Thus the reader is warned that some of the estimates and conclusions are conditional and may be changed in future as the timescales and interpolations are improved. With some indicated exceptions, estimated frequencies are given without consideration of possible deficiencies in the present timescales.

## 9. Example 1: V22-174

The first data-set is from the core V22-174, kindly supplied by Professor N. G. Pisias. The depth-to-time transformation used the ‘tune-up’ timescale of Hays *et al.* (1976) that resulted from taking obliquity as a clock. This set consists of 156 data points at a nominal spacing of 5000 years and extends from the present through the Brunhes–Matuyama magnetic reversal (approximately 730 000 years ago). This data is shown in figure 1. In this, and similar plots following, the time axis runs backwards to the conventions normally used in statistics and engineering (geologists typically plots functions of age or depth); i.e. the present is the origin and moving to the right moves backwards in time. Similarly, the vertical axis is a proxy for global ice volume. Preliminary analysis of V22-174 showed that a uniform shrinkage of the sampling interval from the nominal 5000 year spacing to 4930 years brought most frequencies into somewhat better agreement with the Berger frequencies and this value is used hereinafter. Apart from this uniform shrinkage the maximum timing error appears to be less than 2600 years.

Initial estimates of the spectrum and line components were made using the multiple-window method with  $NW = 3.5$  and  $K = 5$ . These values were selected as a reasonable compromise after trying several, and reflect the complexity of the spectrum and the large number of line components probably existing. Because the ratio of maximum to minimum spectrum estimates is small, larger values of  $NW$  are

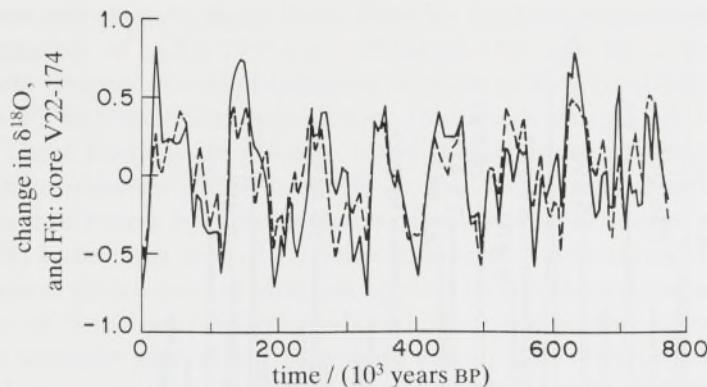


Figure 1. The raw  $\delta^{18}\text{O}$  data for core V22-174 (solid line) and the estimated fit (dashed line) from the periodic components detected by the harmonic  $F$  test shown in figure 3 and listed in table 2. The data consist of 156 samples at an estimated sample spacing of 4930 years.

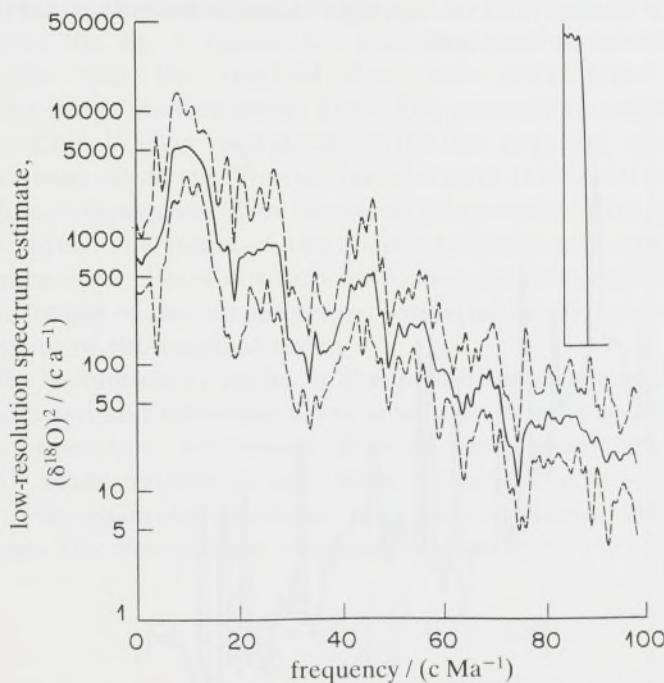


Figure 2. Low resolution spectrum estimate for the V22-174  $\delta^{18}\text{O}$  data made assuming no line components. In this example a time-bandwidth product of 3.5 and five windows were used. The roughly rectangular ‘bumps’ in the spectrum are line components; see figure 3. The dashed lines are the 5 and 95% confidence limits determined by jackknifing over windows as in (16); note that the limits are wide near the edges of the line ‘bumps’. The insert shows that the effective spectral window has no sidelobes on this scale.

not needed for bias protection. The raw spectrum, computed using (12) and (13) initially ignoring the presence of line components, is shown in figure 2. The major ‘lumps’ of power near  $10$ ,  $25$ , and  $40$  to  $50\text{ c Ma}^{-1}$  are largely eccentricity, obliquity, and precession respectively. Figure 3 is a plot of the simple harmonic  $F$  statistic (46) with lines for which  $F$  exceeds the 90% point labelled with their estimated periods. The threshold was set lower than I recommend with known sampling times to allow for some departure from sinusoidal behaviour and also

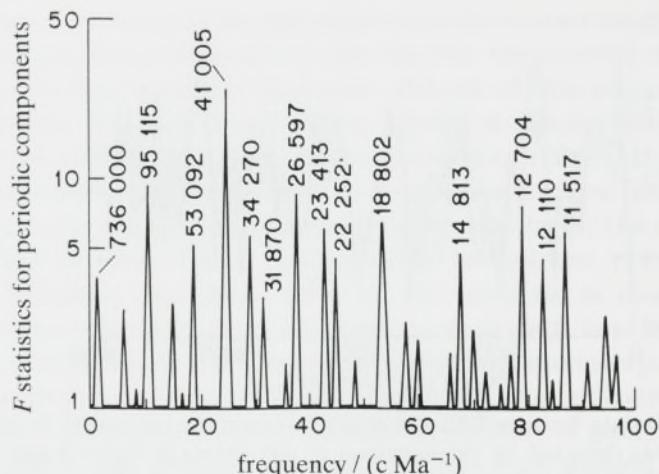


Figure 3. The  $F$  statistic (46) for periodic components using the same estimation parameters as in figure 2. Peaks where the  $F$  statistic (with two and eight degrees of freedom) exceed the 90% point are labelled with the estimated period in years.

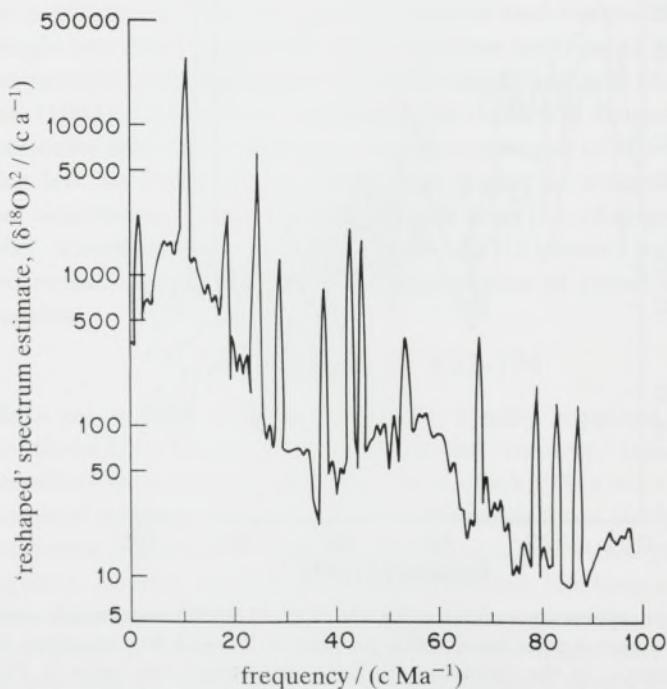


Figure 4. The reshaped spectrum estimate, see §5.1, combining the information about periodic components from figure 3 with the residual spectrum as in figure 2. Energy associated with the periodic components marked in figure 3 is coherently subtracted from the eigencoefficients used to compute the spectrum of figure 2. The 'baseline' in this figure represents the residuals, mostly continuum spectrum. Energy in the line components has been included by taking the shape of the  $F$  test (so the width is proportional to frequency uncertainty) and scaling the height to preserve power.

because here one expects many lines. For this analysis frequencies were computed at a mesh spacing of  $4.9 \times 10^{-8} \text{ c a}^{-1}$  obtained by use of a 4096-point FFT. The approximate frequencies so obtained were then refined to obtain the frequencies for which the *F* test was maximized and tolerances estimated by jackknifing over windows. These frequencies are also listed, together with their estimated tolerances and tentative identifications in table 1. For these identifications the theoretical frequencies computed by the methods of celestial mechanics and given in Berger (1977, 1978) have been used. In addition, several ‘cross-terms’ have been identified by comparison with a similar analysis applied to the theoretical series. Judged by the occurrence of lines near their predicted frequencies the agreement is excellent; agreement with the predicted frequencies less so. The reshaped spectrum is shown in figure 4. the reader may wish to compare this estimate with those of Moore *et al.* (1982), Siddiqui & Wang (1984), or Wigley (1976).

Next a ‘trend’ generated by taking a trigonometric series formed using the estimated frequencies, amplitudes and phases of the 13 strongest lines, those marked in figure 3, was subtracted from the data. Figure 1 shows the data with the trend superimposed; the fit is reasonable and explains about 57% of the variance. Following this step the residual data was prewhitened using a fifth-order autoregressive (or prediction error) filter. The prediction coefficients are obtained by solving the Yule–Walker equations with the required estimates of the autocovariances obtained as the Fourier transform of the spectrum left after removing the above line components. The choice of a low-order filter here is deliberate: the desire is to reduce the range of the spectrum while not attempting to follow fine details. Higher-order filters not only introduce considerable distortion but also, by reducing the length of the output series, correspondingly reduce available resolution in the spectrum of the residual series.

Besides the reduction in range in the prewhitened spectrum with its attendant gains in resolution and efficiency there is an even more important reason to use this prewhitening operation: robustness. Because one is usually interested in spectra on a logarithmic scale, outliers in time series do not have to be large to cause serious distortions in the estimated spectrum, only large relative to the innovations variance of the process. The innovations variance is given by

$$\sigma_I^2 = \exp \left\{ \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln S(f) df \right\}$$

and is the one step linear-prediction variance of the process and is the basis for using log spectra instead of simple power. (The prediction variance is also known as Wiener entropy and the confusion with Shannon entropy is responsible for much mischief; for a discussion of both Wiener and Shannon entropies see Berger (1971).) Processes encountered in the physical sciences are often predictable within a small fraction of the process variance,  $\sigma^2$ , usually  $\sigma_I^2/\sigma^2 \approx 10^{-1}$  to  $10^{-6}$  with 0.01 to 0.001 being common. Thus a time series ‘outlier’ may be small compared with the overall scale of the process. If a series contains only a few isolated outliers their effect may be greatly reduced by use of a device known as a ‘robust filter’. In this procedure, data at a given time is compared with its prediction from preceding points of the filtered series. If the prediction error is small the data are retained; if large they are replaced by the prediction; at intermediate errors a weighted combination of data and prediction is used. Details are given in Thomson (1977), Kleiner *et al.* (1979), and

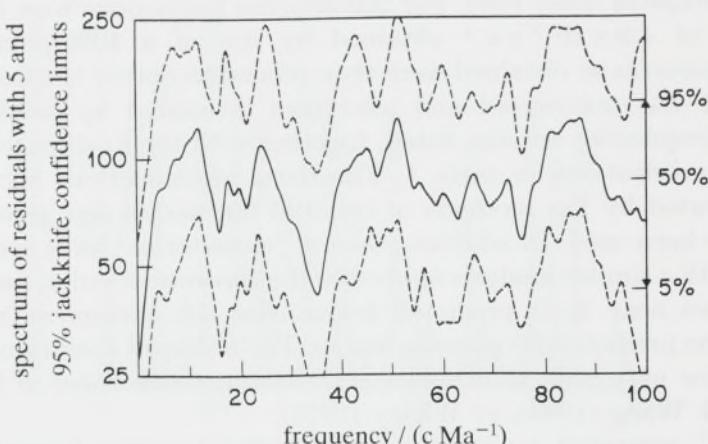


Figure 5. Estimated spectrum of the residuals with 5 and 95% jackknife confidence limits. These residuals are found by: first, subtracting the fit shown in figure 1 from the original data; second, prewhitening these with a fifth-order autoregressive prediction error filter. The length of the arrow on the right margin shows the 5 and 95% points expected with a  $\chi^2_{16}$  distribution.

Martin & Thomson (1982). Note that the name ‘robust filter’ is also used for the closely related problem of finding Wiener filters that are robust to uncertainties in the assumed signal and noise spectral densities; see Kassam & Poor (1985) for a review. (Here the uncertainty is because the filter is being derived from an estimate of the spectrum.) However, in the V22-174 record no data were changed, partly because predictability of the residual process is unusually low,  $\sigma_I^2/\sigma_r^2 \approx 0.30$ , where  $\sigma_r^2$  is the variance of the residuals left after subtracting the periodic terms, but primarily because of the care taken in preparing such data. Incidentally, if one attempts to directly apply a robust filter to the original series without removing the line components several points are erroneously rejected. Finally, the 151 data points resulting from these operations we reanalysed using three procedures as follows.

1. The simple spectrum of these residuals was computed with a wider band,  $NW = 5$  and  $K = 8$  multiwindow approach with the jackknife variance estimate. The average jackknife variance of  $\ln \bar{S}_r(f)$  is 0.144, close to the expected  $\psi'(8) = 0.133$  and the maximum is 0.69, again with the range expected. This is shown in figure 5.

2. The first six (lowest-order) quadratic inverse coefficients were computed, again with  $NW = 5$  and  $K = 8$ ; some are shown as ratios  $b_l(f)/\bar{S}_r(f)$  in figures 6 and 7. Recall from §3 that  $\hat{b}_l(f)$  is unbiased if the local spectrum has an expansion in the  $B_l(f)$ s whereas  $\bar{S}(f)$  is unbiased only when the even part of the local spectrum is constant. The fact that  $b_0(f)$  differs significantly from  $\bar{S}(f)$  implies that the simple estimate is biased by almost a factor of 2 at low frequencies. One may also estimate the bias of  $\bar{S}$  using the jackknife; figure 6 shows both this estimate and the ratio  $b_0(f)/\bar{S}(f)$ , and although the general picture is similar, it is clear that  $\bar{S}$  misses much fine structure. Figure 8 shows the unresolved structure test (41). At low frequencies, about  $3 \text{ c Ma}^{-1}$ , the spectrum is apparently poorly resolved. This is not surprising as, for example, the theoretical eccentricity expansion in table 3 of Berger (1978a) lists 7 components with periods between  $2 \times 10^5$  and  $10^6$  years and there are similar periods in obliquity. The peak near  $32 \text{ c Ma}^{-1}$  is more of a puzzle; visually it occurs on the lower side of the ‘hole’ in the residual spectrum at  $36 \text{ c Ma}^{-1}$  and falls roughly between the obliquity and lower precession bands. One explanation is that the line with apparent period 31 870 years removed in the fit is a composite; the closest

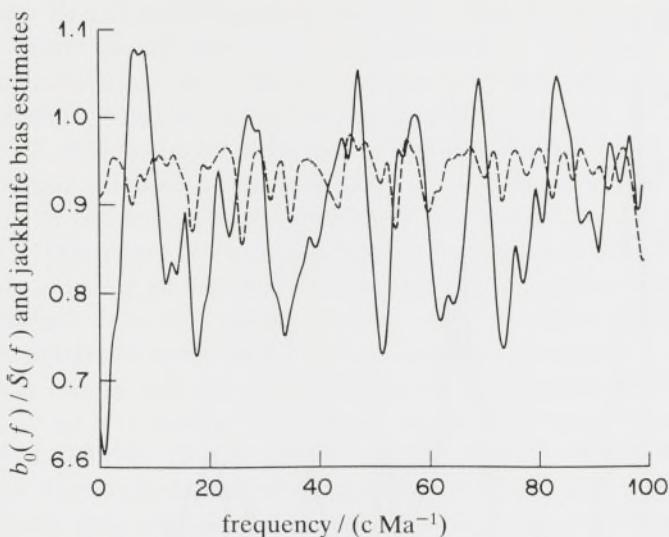


Figure 6. Two estimates of the bias of  $\bar{S}(f)$ : the solid line is a plot of the ratio of the quadratic inverse coefficient  $b_0(f)$  to  $\bar{S}(f)$ ; the dashed line is the jackknife estimate of bias given by (18). Both show that  $\bar{S}(f)$  should be biased down by a factor of approximately 0.9 and, while some of the fine structure shows coincidental agreement, the quadratic inverse coefficient generally requires larger changes than does the jackknife. The curves have been smoothed over a Rayleigh resolution.

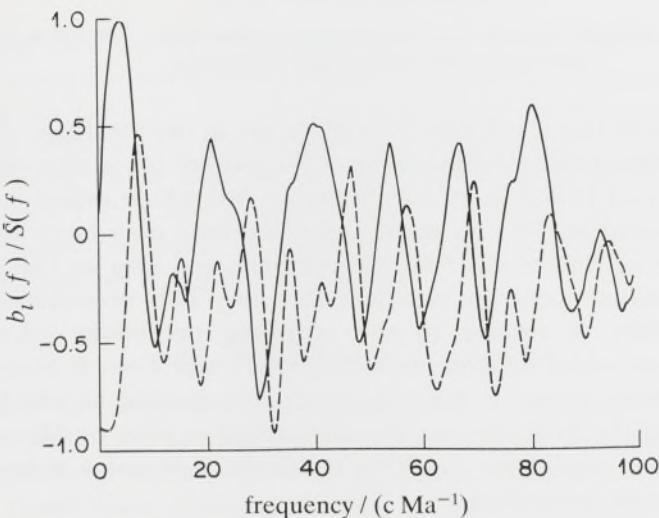


Figure 7. Quadratic inverse coefficients  $b_1(f)$  (solid) and  $b_2(f)$  (dashed) a fraction of  $\bar{S}(f)$ . The curves have been smoothed over a Rayleigh resolution.

simple theoretical period is 30365 years obliquity, but the frequency error is more than the Rayleigh resolution. However, a similar analysis of squared precession shows a period of 31800 years and there are several other cross terms in the vicinity. Inhibiting the removal of the 31870 year line reduces the peak in the unresolved structure test to near normal levels so that it may be spurious. It is also noteworthy that the ‘hole’ in the spectrum near  $36 \text{ c Ma}^{-1}$  is matched by one near  $73 \text{ c Ma}^{-1}$ , suggesting a ‘moving average’ like phenomenon with a recurrence time of about 27400 years. It is tempting to speculate that the ‘overshoot’ time of about 24000 years in the transfer function described below is more than coincidentally related.

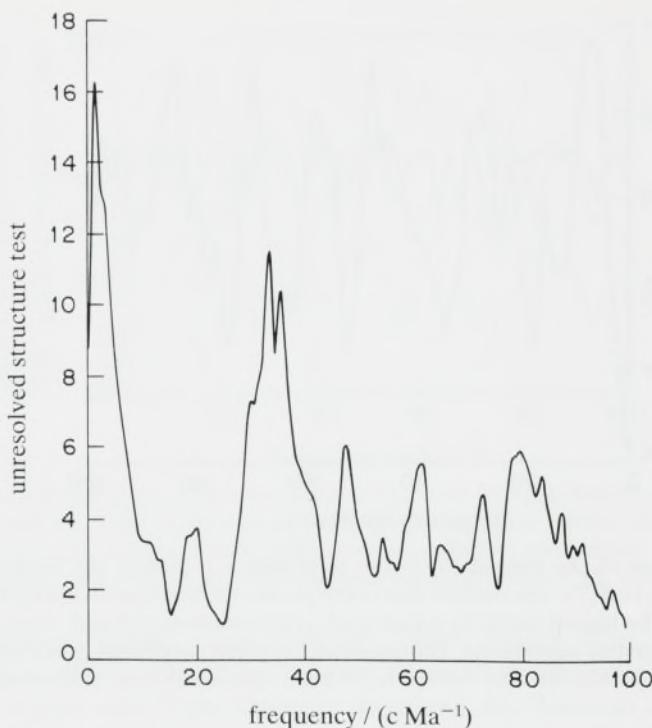


Figure 8. The quadratic inverse test for unresolved structure. The curve has been smoothed over a Rayleigh resolution.

3. The analysis, including the  $F$ -test, was repeated on the residuals. Again, several low-level lines were detected, including some of the weaker precession signals and one at an apparent period of 12084 years with a high  $F$  but whose power is only slightly less than the lines found in the first pass. Detection here may have been enhanced both by the removal of the stronger interfering lines and also by the prewhitening filter making the residual spectrum flat near this line. This procedure, also used in Thomson *et al.* (1986), is similar to the stripping procedure ‘clean’ used for deconvolution and described in Roberts *et al.* (1987) and Tsao & Steinberg (1988). However, as Tan (1986) shows, it must be used with caution as effective sidelobes from slight misfits on the first pass can be interpreted as lines on the second. Thus, although the process is useful for checking residuals, it is not a substitute for the simultaneous line fitting procedures.

## 10. Example 2: the SPECMAP stack

The second data-set analysed is a composite formed by ‘stacking’, or ensemble averaging, records from five cores all aligned using the SPECMAP timescale and described in detail by Imbrie *et al.* (1984). Briefly the series consists of 783 samples at a nominal sampling rate of 1000 years beginning with the present and ending about 782000 years before present (BP). The data consist of three cores that cross the Brunhes–Matuyama magnetic boundary; V28-238, V22-174 (again), DSDP 502b and two shorter cores, V30-40, and RC11-120. This timescale was produced by making a compromise between obliquity and precession, with a bias in favour of precession. Though the consistency of the timescale is the best currently available

back to near the Brunhes–Matuyama boundary (the lower parts, those older than 617 000 years, may be off somewhat (Ruddiman *et al.* 1989; N. J. Shackleton, personal communication), the absolute accuracy is not as good as with obliquity tuning. The present analysis gives an explanation for this discrepancy. Looking ahead to figure 13, which shows estimates of the differential phase between the stack data and the Berger calculation for obliquity, it may be seen that (ignoring the error near the bottom of the core) the differential is less than 1.3 rad or  $\pm 0.11$  of a cycle. For a period of 41 000 years the implied peak discrepancy is thus less than 4500 years and part of this is linear phase drift from a simple frequency offset.

With the stacked data the signal-to-noise ratio is large enough to examine the multiple-line structure in more detail. As an example I consider the set of frequencies in the precession bands and use  $C_p$  (Mallows 1973) to select those that explain the observed data most succinctly. For this I begin with nine lines with theoretical periods of 18873, 18976, 19155, 19261, 22428, 22818, 23293, 23716, and 27027 years. These are augmented with mesh periods of 17241 and 20727 years to form a grid of eleven mesh frequencies. A time-bandwidth product of 3.5 was used and only the four lowest-order windows were retained; for this configuration equation (49) estimates  $\nu = 25.4$  degrees of freedom. Also, to whiten the background spectrum in the precession band, the raw data,  $r_t$ , were filtered with the first-order prediction error filter  $x_t = r_t - 0.93660357r_{t-1}$  before computing the eigencoefficients. Here I define

$$C_p(j) = R^{\text{ss}}(j)/\sigma^2 - \nu + 4j,$$

where the background power is estimated by  $\sigma^2 = R^{\text{ss}}(9)/(25.4 - 2 \times 9)$  where  $R^{\text{ss}}(j)$  is the residual sum-of-squares with  $j$  complex amplitudes estimated. Beginning with  $j = 9$  and then for  $j = 8, 7, \dots, 1$ , the line having the lowest partial- $F$  statistic was deleted and the resulting  $C_p$  statistics are shown in figure 9. A reasonable model is that with four lines, those with periods of 23716, 22428, 19155, and 18873 years. These are numbers 1, 2, 4, and 7 respectively in table 2 of Berger (1978a). Note that two lines are chosen from each precession band (near the edges of both) and that the fits with single lines are significantly poorer.

## 11. Changes in precession

Comparison of the estimated and astronomical frequencies of the different line components in table 1, or in a similar analysis on the SPECMAP stack, suggests that there are systematic differences between them that depend on the origin of the line. Particularly, those lines associated with precessional effects hint at different frequency shifts from those associated with changes in the obliquity of the axis and the ellipticity of the orbit. A possible reason for these apparent differences is that the astronomical calculations assume a static Earth and take no account of internal dynamics. This certainly appears reasonable insofar as ellipticity and, to a lesser extent, obliquity is concerned; there does not appear to be any simple way for internal changes to affect the ellipticity of the orbit. Similarly, the predicted changes in obliquity occur because the ecliptic of date tilts relative to the ecliptic of reference while the tilt of the axis remains fixed with respect to the ecliptic of reference; again there does not appear to be any mechanism whereby internal changes on the Earth can cause major changes in the orientation of the ecliptic within the geologically short time periods under consideration. Although internal changes can change the tilt of the axis about the reference ecliptic, the possibility of causing ‘large’ shifts in

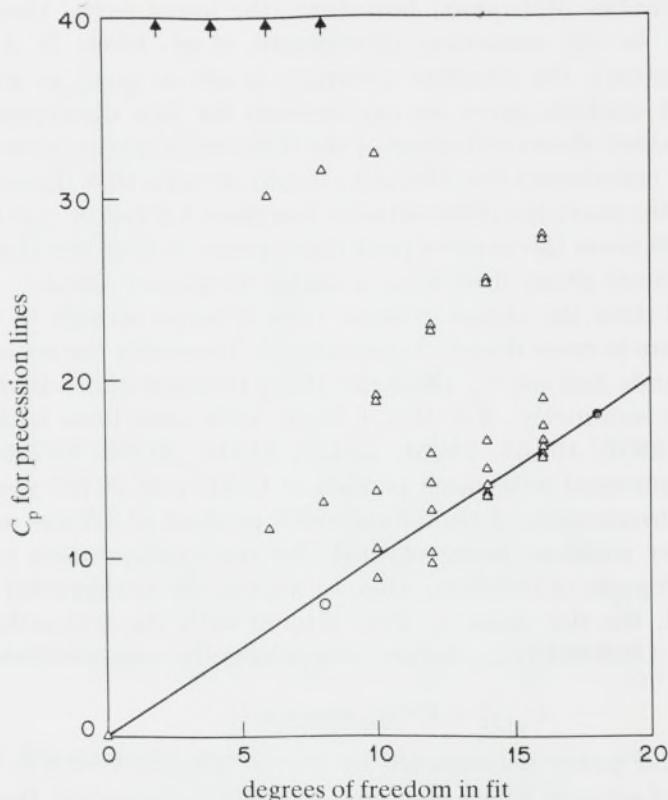


Figure 9. Mallows's  $C_p$  statistic for significant periodic components in the two precession bands. The bottom set with eight degrees of freedom (circled) is adequate and requires two lines in each band. The ordinate has been truncated at 40; individual values go to 240 with a single line. An initial fit with 18 degrees of freedom (solid dot) was used for calibration.

obliquity, i.e. more than a few seconds of arc, appears to be small. Work on this question goes back to Darwin (1877) and has been extensively studied since (see Munk & MacDonald (1960) for an interesting overview, Lambeck (1980), and Peltier (1982) for more recent work) the general conclusion appears to be that the effects of geological changes on obliquity during the Quaternary must be small. Moreover, such changes are largely non-cumulative; restoring the mass distribution largely restores the obliquity (Wu & Peltier 1984).

Precession, however, is a different matter. There are two components of general precession: *planetary* precession, that results from changes in the ecliptic by perturbations from the other planets, and *luni-solar* precession, that comes from the couple induced by the solar and lunar potentials with the non-spherical figure of the Earth. For my purposes an adequate formula for the luni-solar precession constant (Brouwer & Clemence 1961; Danby 1962) is

$$\omega = \kappa(C - A)/nC$$

with  $\kappa$  a constant and the important quantities for this discussion are  $n$ , the angular velocity of the Earth about its axis,  $C$ , the moment of inertia about the polar (or rotational) axis, and  $A$ , the moment of inertia about the equatorial axis, all three of which may be changed by terrestrial changes in mass distribution. Note that precessional effects are purely a result of the non-spherical figure of the Earth; if  $C$

and  $A$  were equal there would be none. Indeed, because the Earth is reasonably spherical we have implicitly assumed rotational symmetry, i.e. that the moments  $A$  and  $B$  about both equatorial axis are equal,  $(C-A)/C$  is small (*ca.* 0.003 276) so small changes in  $C$  and  $A$  can result in significant changes in their difference. Because glaciation involves the transfer of about  $60 \times 10^{18}$  kg of water from the ocean to the polar ice caps out of a total mass of the Earth of  $5.974 \times 10^{24}$  kg, changes in the moments of inertia should be expected. Calculations are complicated, however, because the Earth is not rigid and responds to surface loading by redistributing mantle material, but the mechanics of this redistribution have been enigmatic. Calculations by Peltier (1982, 1985) and Wu & Peltier (1982, 1983, 1984) assuming a particular model for the mantle give a mixture of exponential decay modes with time constants for the longest major component of about 12 000 years. Such a short adjustment time, however, would imply that surface rebound from the Wisconsinian glaciation would be nearly complete so that the figure of the Earth should be close to that expected at hydrostatic equilibrium. Jeffreys (1976) gives  $H_{\text{he}} = 0.003\,237\,9$  for  $(C-A)/C$  at equilibrium, about 20 standard errors from the measured value of 0.003 276. Note, however, that Goldreich & Toomre (1969) have argued that this discrepancy is an artefact of the way the principal axes are defined; see the discussion in Lambeck (1980, ch. 11) and Tanimoto (1989).

The analysis given in the following section uses the methods developed here to estimate the transfer function between glacial loading and changes in precession. This transfer function is negative at long periods as required by eventual mantle mass redistribution, near zero at a period of about 180 000 years, and positive at shorter periods. Thus glacial loading, with its characteristic 100 000 year cyclicity, occurs primarily in the positive part of the transfer function, so resulting changes in precession appear larger than if the crust were rigid. This, in turn, implies that the average figure of the Earth over recent geological time is much closer to that predicted by hydrostatic theory, and that the current value of  $H$  is not only atypical because of the contemporary low glaciation but also because of filter ‘overshoot’. As will be seen below, this result is obtained by two essentially independent ways: first, by the mean slope between obliquity and precession; second, by the detailed transfer function response.

Proceeding with details of the calculation, Rochester & Smylie (1974) have shown that the trace of the inertia tensor is conserved so changes in the two moments are related by  $\delta(C+2A) = 0$  and therefore  $\delta(C-A)$  is just  $\frac{3}{2}\delta C$ . As an approximation to these changes I vastly oversimplify the problem and assume that the Earth is rigid, that glaciation occurs in a uniform spherical cap from northern latitude  $\theta$  that I take to be  $55^\circ$  N (implicitly filling the Arctic ocean with Antarctica), and that the rest of the Earth is ocean. Although such a cap has a considerable moment of inertia, it is formed by a reduction in global sea level originally with an even larger moment of inertia, and consequently  $\delta C$  is negative. I approximate the moment from sea level change as that of a thin spheroidal shell and write the total change in the moment of inertia about the polar axis as  $\delta C = \delta C_c - \delta C_s$ . Crowley (1983) and Fairbanks (1989) estimate that the peak glaciation consisted of from  $50$  to  $60 \times 10^6$  km $^3$  of water with an accompanying reduction in global sea-level of about 120 m so I choose  $M_i = 60 \times 10^{18}$  kg. The polar moment of the spheroidal shell is

$$\delta C_s = \frac{2}{3}M_i a^2 = 1.627 \times 10^{33} \text{ kg m}^2.$$

To compute the moment of the cap we take the radius at  $55^\circ$  latitude,  $a_\theta = 6363.8$  km

(Allen 1976) and approximate the surface as locally spherical. With this approximation the area of the cap is  $2\pi a_\theta^2 \sin \theta$ , about 9% of the total surface area of the globe, and its moment about the polar axis

$$\delta C_c = \frac{1}{3} a_\theta^2 M_i [2 - \sin \theta (\cos^2 \theta + 2)] / (1 - \sin \theta)$$

so the total change in  $C$  is  $\delta C \approx -1.204 \times 10^{33}$  kg m<sup>2</sup>. The current value of  $C$  is  $0.3306 M_\oplus a^2 \approx 8.0372 \times 10^{37}$  kg m<sup>2</sup> so that  $\delta C/C \approx -15.105 \times 10^{-6}$  and recalling that  $nC$  is the angular momentum about the polar axis and hence constant the change in the luni-solar precession constant may be written

$$\delta\omega = \frac{3}{2}\omega(\delta C/C)/[(C-A)/C].$$

The relative change in  $C-A$  is  $-22.657 \times 10^{-6}/0.003276$  or  $-6.916 \times 10^{-3}$ . The luni-solar precession for 1950.0 (Berger 1978a) being 50.439273 seconds per year, the change at peak glaciation will be about -0.349 seconds per year assuming a rigid crust.

Although such a change is astronomically gross, is an annual change in precession of 0.3 seconds detectable in the geological record? Assume, for the moment, that the timescale is accurate and consider the error in estimating the frequency of a line component. For the strongest lines one might obtain (optimistically) a signal-to-noise power ratio,  $S_A/S_c$  of 200 (the apparent signal-to-noise of the four precession lines discussed above is about 54 but much of the 'noise' appears to be the unresolved lines) so that for  $T = 783000$ ,  $\text{var}\{\delta f\} = 1.24 \times 10^{-15}$  corresponding to a standard deviation of  $3.52 \times 10^{-8}$  c a<sup>-1</sup>. This is equivalent to 0.046 seconds per year so that precessional frequency changes about 0.35 seconds per year would be easily detectable under idealized conditions. Unfortunately actual conditions are far from ideal. In the first place we are not dealing with a constant frequency offset but one changing with glaciation. Thus the precession lines will appear to be frequency modulated. Second, the timescale is not known perfectly but must be partly derived from the data. These difficulties are reduced somewhat because there are many lines in the spectrum.

Having decided that there is possibly enough signal in the data to be useful, one must now make more realistic assumptions. Changing the geographical distribution of ice loading to one more compatible with geological evidence is tedious but straightforward; changing the assumption of a rigid Earth requires a realistic model of the Earth's visco-elastic properties. The latter is a problem because, during periods of extended glaciation, considerable subsidence and crustal deformation occur, presumably accompanied by flow in the mantle. For example it is estimated that the Laurentide ice sheet reached thicknesses of 2500 m, while the crust was depressed about 750 m. These changes, however, occur slowly with a time constant of about 12000 years estimated in Le Treut & Ghil (1983). Thus the changes in precession will not be proportional directly to the global ice mass but instead to a high-pass filtered version of it. Calculations of the changes both in the moments of inertia and their visco-elastic filtered versions have been computed by Wu & Peltier (1984). Their calculations are in rough agreement with those inferred from the stack data, but agreement of the differential phase is poorer than with the filter found here. In particular I find larger changes in precession than those predicted by Peltier's model. Among possible explanations for this disagreement are: (1) the signal-to-noise ratio is simply not high enough; (2) glaciation was heavier than believed; (3) the visco-elastic properties of the mantle are substantially more rigid than assumed by Peltier;

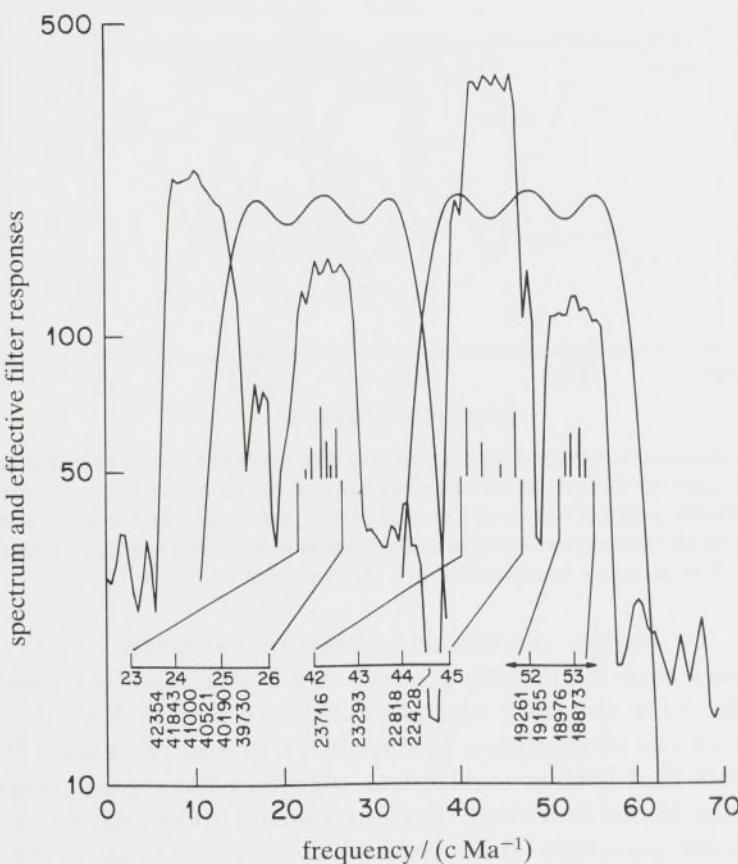


Figure 10. A low-resolution spectrum of the Imbrie stack in the precession and obliquity bands with the effective ‘transfer functions’ of the coherence filters superimposed. The inserts show the theoretical (Berger) detailed structure of both obliquity and the 19k precession band; the length of the arrow is one Rayleigh resolution.

(4) the astronomical calculations of the planetary precession or obliquity are inadequate; (5) the age scales contain subtle, correlated, artefacts. This problem is further complicated because the present ice volume is exceptionally low by historical standards so that the precessional constants derived from recent observations are probably anomalous.

To test these hypotheses, time-varying spectra and coherences between the data and theoretical forcing functions were computed on ‘short’ blocks that were slid along the stack. To track the phase to the desired accuracy requires a signal reasonably free of interfering effects and a good signal-to-noise ratio. (Similar problems have been extensively studied in the electrical engineering literature on frequency modulation (FM) and the ‘capture’ effect makes perfect rejection of interfering signals or exceptionally high signal-to-noise ratios unnecessary; frequency demodulation works well provided one has a signal-to-interference of several decibels and is above threshold (see Middleton 1960; Schwartz *et al.* 1966; Van Trees 1968). None the less, these requirements eliminate all but a few lines. Referring to figure 10, which is a schematic spectrum, the obliquity band at  $24.39 \text{ c Ma}^{-1}$  is restricted below by an ellipticity peak at  $18.8 \text{ c Ma}^{-1}$  and above by one at  $29 \text{ c Ma}^{-1}$ . Thus the available bandwidth is  $29.0 - 18.8 = 10.2 \text{ c Ma}^{-1}$ . With a sampling rate of 1000 years the Nyquist frequency is  $500 \text{ c Ma}^{-1}$  so that the normalized bandwidth is  $W =$

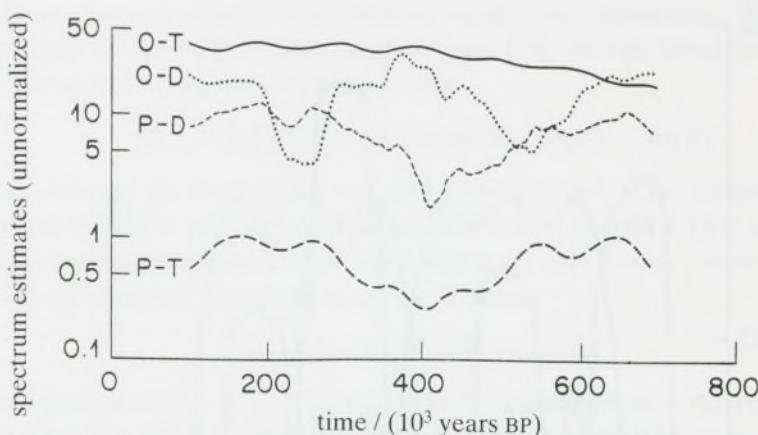


Figure 11. Spectrum estimates plotted as a function of the central time in the analysis block. The curves labelled ‘O-T’ and ‘O-D’ are for the theoretical and Stack obliquity respectively, with the analysis centred on 41000 years. Curves ‘P-T’ and ‘P-D’ are similarly for precession centred on 21083 years. The dip in the precession spectra near 400000 years before present coincides with the ellipticity minimum. The analysis bandwidths are those shown in figure 10.

$5.1/500 = 0.0102$ . To obtain reasonable coherence estimates (eight degrees of freedom) we choose a time-bandwidth product of 2.9 with resulting passbands from  $16.89$  to  $31.89 \text{ c Ma}^{-1}$  for obliquity and from  $39.93$  to  $54.93 \text{ c Ma}^{-1}$  for precession. Figure 10 also shows the composition of the obliquity and precession ‘lines’ on an expanded scale. Note that both are multiplets, obliquity having six components, the two major precession bands four each. The line spacing and signal-to-noise ratios in the weaker sub-lines prohibits their resolution even without problems in the timescale. This is why coherences are used. Note, incidentally, that when discussing significance levels we are computing coherences against a deterministic driving function, not between two random signals, so the interpretation should be as in Munk & Cartwright (1966) or Shumway (1988). Now we compute the spectra, coherency, and phase of the coherency between the data and obliquity at a centre period of 41000 years and similarly between data and precession at a centre period of 23000 years. These quantities are plotted, as functions of the centre time of the block, in figures 11 to 13. Note that the spectra of data and theoretical quantities have similar time dependence. Also the different spectra, as one would expect with eight degrees of freedom, do not become small anywhere. Similarly the magnitude-squared coherences are all acceptable, showing that the calculation is dominated by signal, not noise. More important, as mentioned earlier, the phase of each coherency is small at all times, indicating that the timescale is good.

The difference in these two phases is shown in figure 14. In this figure the dominant characteristic is the near linear slope terminating in the rapid break at  $\text{ca. } 6.6 \times 10^5$  years BP. Near this time the stack is down to two cores, and sampling of these is becoming sparse so the break is most probably attributed to low signal strength and the dating error mentioned earlier. A linear least-squares fit to the differential phase up to this point has a slope of 0.21 seconds per year. This average phase-slope is, of course, just the frequency offsets between phase and obliquity that has been noticed by Imbrie, Pisias, and others. The slope is not well determined and narrowing the filter bandwidths gives a more stable slope estimate of about 0.31 seconds per year at the expense of details in the transfer function.

Suppose one tunes to a theoretical frequency,  $\omega_T$ , by phase matching when the

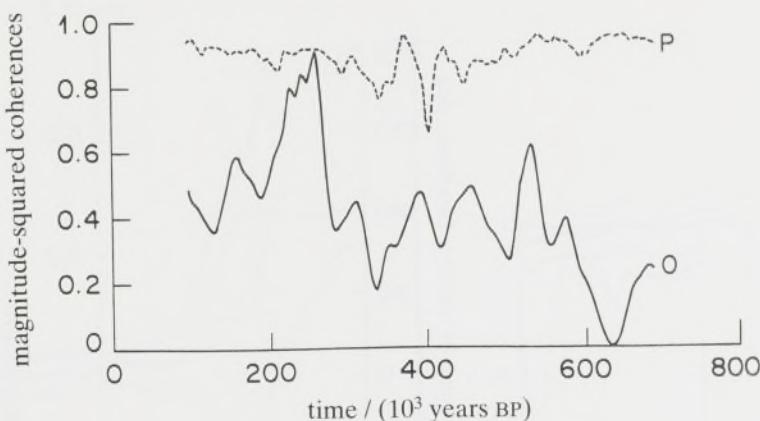


Figure 12. Estimates of magnitude-squared coherence between the theoretical and stack precession (dashed line) and obliquity (solid line) plotted as functions of the central time in the analysis block. These correspond to the spectrum estimates shown in figure 11. The 50 % point for a single estimate is 0.206 (Lee 1981) and the precession coherence is higher because the tuning emphasized it.

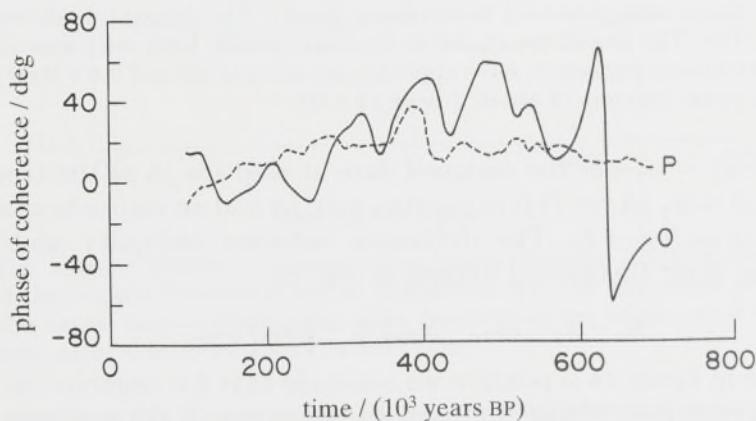


Figure 13. Phase of the coherences between the theoretical and stack precession (dashed line) and obliquity (solid line) plotted, as in the last two figures, as functions of the central time in the analysis block.

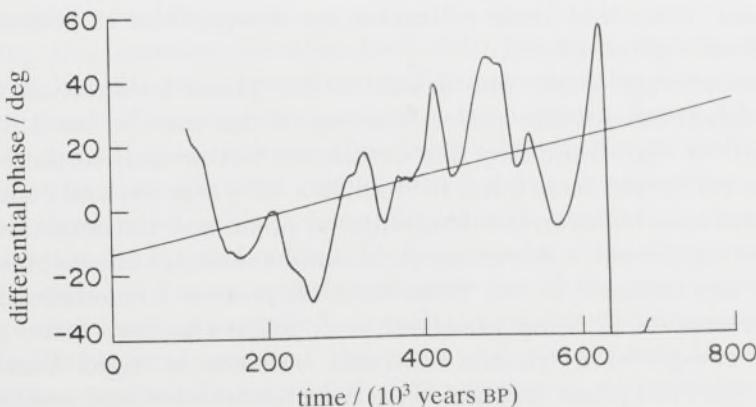


Figure 14. Differential phase, obliquity minus precession, for the coherence phases of figure 13. The solid line is a robust slope estimate and corresponds to increasing the average precession by 0.21 seconds per year.

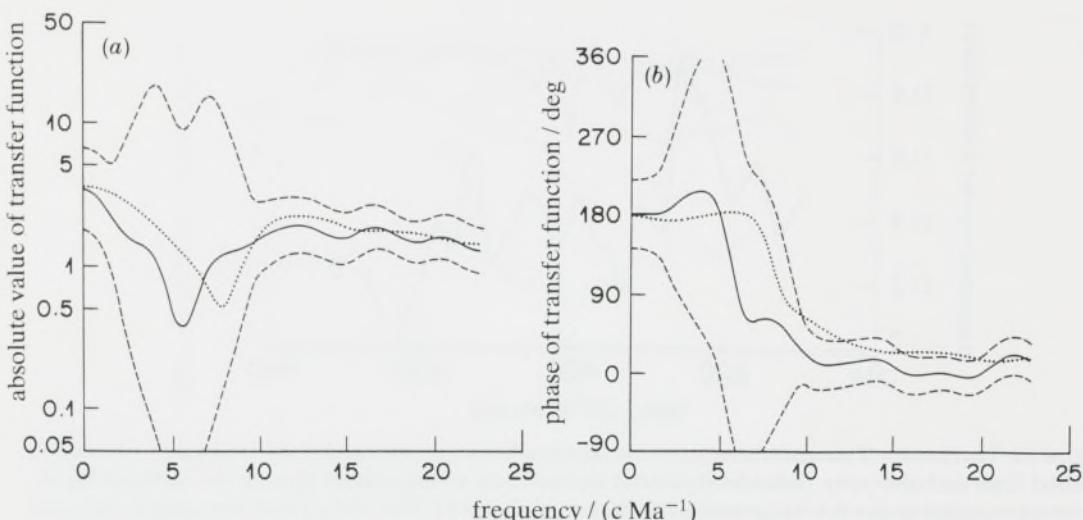


Figure 15. Two estimates of the transfer function from global ice volume to differential phase. The first (solid line) is the direct estimate of  $H(f)$  from (53); the top and bottom dashed lines are  $\pm 1$  standard deviation limits using standard multivariate theory. The dotted line shows the integral of  $\hat{H}'(f)$  started at  $\hat{H}(0)$ . The phases are shown in the lower panel; both estimates of the transfer function are negative at zero frequency, have approximate zeros at  $5.6$  and  $7.8 c \text{ Ma}^{-1}$  respectively, and have near zero-phase maxima of about  $2$  near  $12 c \text{ Ma}^{-1}$ .

actual frequency  $\omega_A = \omega_T + \delta$ ; the assigned date at step  $j$  is  $j\Delta$  so the true time  $T(j)$  is given by  $\omega_A T(j) = \omega_T j\Delta$ , or  $T(j) = [\omega_T / (\omega_T + \delta)] j\Delta$  and an estimate of a frequency  $\omega_0$  will be  $\hat{\omega} = \omega_0 \omega_T / (\omega_T + \delta)$ . The difference between obliquity and precession frequencies minus their theoretical difference will be

$$\hat{\omega}_0 - \hat{\omega}_p - (\omega_{To} - \omega_{Tp}) \approx -\delta \omega_{To} / \omega_{Tp}.$$

Because the slope in figure 14 is positive we conclude that  $\delta$  is negative, in agreement with the above prediction. Using the centre frequencies of the analysis filters, the factor  $\omega_{To} / \omega_{Tp}$  is approximately  $24.39 / 47.43$  or  $0.514$ , implying that  $\delta$  is about  $0.411$  seconds per year while the more reliable slope estimate of  $0.31$  scales to  $0.60$  seconds per year. Subtraction of these from the current luni-solar precession constant of  $50.439$  gives  $50.03$  and  $49.84$  seconds per year respectively as estimates of the luni-solar precession constant. Note that these estimates are independent of both stack units and the simple theory given above.

Subtracting the average slope, one is left with a phase modulation curve that resembles the original ice-volume plots. Because of the narrow-band filters used, detailed agreement of high-frequency fine details are not present in the phase. This process gives a useful output from  $9.5 \times 10^4$  to  $6.32 \times 10^5$  years BP, and computing the low-frequency coherence between the demodulated phase and the original stack data shows modest, but significant, coherences. As the stack includes effects of eccentricity and several lines not included in the demodulation process I conclude that glacial modulation of precession is being observed and, using the derivative procedures described in §7, computed a transfer function between a 'rigid Earth' forcing function and the observed phase changes. This transfer function is shown in figure 15. Note that, at low frequencies, less than about  $5.7 c \text{ Ma}^{-1}$  (corresponding to a period of 175000 years) the transfer function is larger than 1 and negative; forcing with periods significantly longer than 175000 years will be compensated and not change

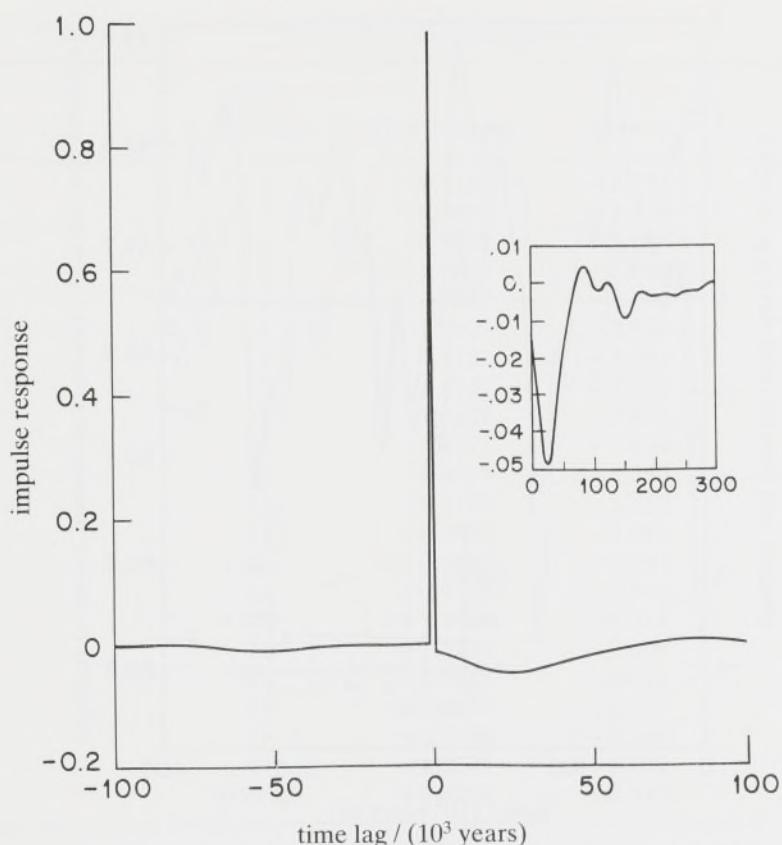


Figure 16. The impulse response corresponding to the transfer function plotted in figure 15, assuming that the transfer function at higher frequencies is 1. The response is primarily causal. The inset shows the causal part excluding the delta function at the origin; note that the maximum overshoot occurs after a delay of about 24 000 years.

precession significantly. Because we do not have an absolute zero reference for the data and the series is short compared with the periods involved, the estimated transfer function of about  $-3$  may not be significantly different from  $-1$ . Above the zero, the transfer function is again larger than one but nearly positive real implying that forcing with periods of 100 000 years or less have a greater effect on precession than my simple model suggests.

Converting this transfer function back into the time domain gives an impulse response that, despite the processing used to estimate the changes in precession, is essentially causal! This estimated impulse is shown in figure 16; the anti-causal portion (corresponding to negative time lags) is small. Heuristically, I take this as a strong indication that physical reality and not just a statistical artefact is being observed as there is no explicit time direction or constraint to estimate a causal transfer function built in. Looking at the impulse response more carefully, note that it has a relatively large negative overshoot with the maximum response about 24 000 years after loading is applied; overall the response resembles a massively slow, damped oscillation implying that the crust and mantle may be significantly more spring-like than previously suspected. For precession calculations in particular the response from forcing 24 000 years ago implies that the present values will be greatly influenced by the last glaciation. Running the stack data through this filter produces the results shown in figure 17. This filtered stack is in the same units as the original,

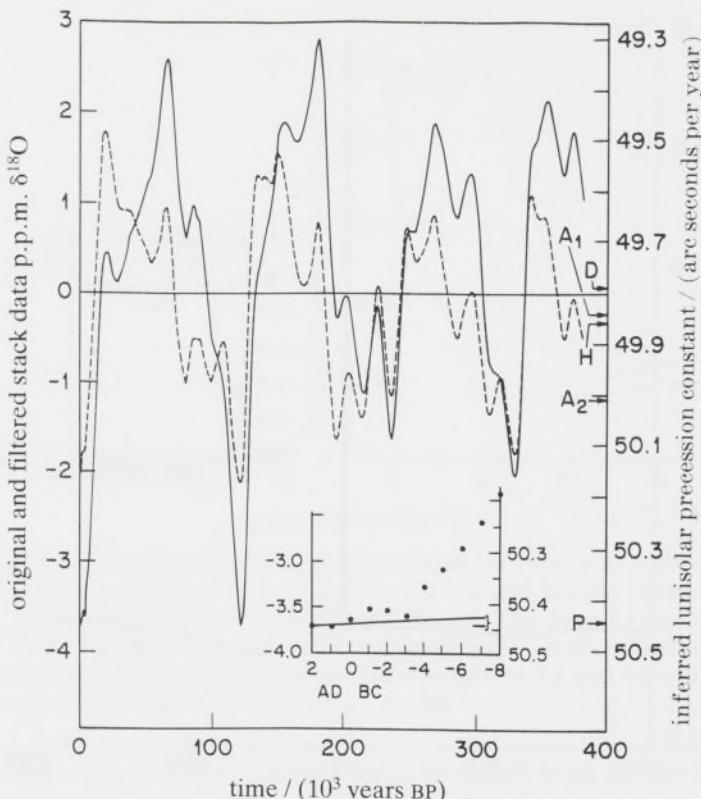


Figure 17. The raw  $\delta^{18}\text{O}$  data for the stack (dashed line) and the results obtained by running the stack data through the filter described in the previous two figures (solid line) with the output in the same units as the stack. Note that the transient response of the filter nearly doubles the effect of the last glaciation. The insert shows the filtered stack for the past 10 000 years, the solid line in the insert shows the slope expected from current LAGEOS data; the overall agreement is good.

that is  $\delta^{18}\text{O}$  parts per million by volume (p.p.m.), but the present amplitude is nearly doubled. The character of the data is also changed; during the slow increases in glaciation there is partial equalization, but the rapid collapses of the ice sheets are too quick and thus substantial overshoots occur (recall that about 100 000 years passed from the beginning of the Wisconsinan to maximum glaciation, whereas collapse of the ice sheets occurred in a mere 12 000 years). Referring to the abbreviated listing of the stack given in the second column of table 2, one sees that the difference between maximum glaciation at  $1.9 \times 10^4$  years BP and current conditions is  $(1.7783 - 2.1157) = 3.894$  corresponding to 0.349 seconds per year obtained with the simple theoretical model. Including the correction for  $\omega_{\text{To}}/\omega_{\text{Tp}}$  described above, one obtains an overall calibration factor of  $0.349/(3.894 \times 0.514) = 0.1763$  seconds per p.p.m.  $\delta^{18}\text{O}$ . The third column of table 2 gives the filter output; scaling the present value of  $-3.704$  gives  $-0.653$  seconds per year or an ‘average’ lunisolar precession of 49.789 seconds per year. Although this estimate is much closer to the 49.853 seconds per year expected at hydrostatic equilibrium than the present value of 50.439 seconds per year it should not be taken too seriously.

Yoder *et al.* (1983) suggested that the non-tidal secular acceleration of the Earth’s rotation was due to deglaciation and found reasonable agreement with Peltier’s (1983) estimate. Taking a crude derivative of the stack output over the past 5000 years one has from table 2  $(3.7043 - 3.6203)/(5000 \times 0.514 \times 3.894) \approx 8.39 \times 10^{-6}$  and

Table 2. Filtered stack data

time (ka BP)	stack-average	filtered
0	-2.1157	-3.704
1	-2.0637	-3.718
2	-1.9367	-3.646
3	-1.7917	-3.545
4	-1.7657	-3.552
5	-1.8097	-3.620
6	-1.4727	-3.301
7	-1.2747	-3.109
8	-1.0437	-2.875
9	-0.7747	-2.594
10	-0.4617	-2.262
11	-0.1107	-1.883
12	+0.2603	-1.477
13	+0.6283	-1.067
14	+0.9753	-0.672
15	+1.2833	-0.311
16	+1.5293	-0.008
17	+1.6913	+0.214
18	+1.7673	+0.351
19	+1.7783	+0.422
20	+1.7503	+0.454
average	+0.0274	+0.347

scaling by the peak glaciation factor  $\delta C/C = 15.105 \times 10^{-6}$  gives a net product of  $\dot{C}/C \approx 12.7 \times 10^{-11} \text{ a}^{-1}$ . This is about twice the non-tidal acceleration given by Mignard (1986) of  $6.6 \times 10^{-11} \text{ a}^{-1}$  as an average over the same 3000 year interval and with current LAGEOS estimates of  $(5.2 \pm 1.2) \times 10^{-11} \text{ a}^{-1}$ . Because we are near minimum glaciation, the first derivative of the filtered stack is near zero and estimates of it consequently susceptible to minor sampling noise (using just the past 1000 years gives the opposite sign!). Despite this, the insert in figure 17 agrees qualitatively with fig. 2 of Mignard (1986) including the minimum between 1000 and 2000 b.c.

Considering the uncertainties introduced by tuning, the possibility of aliases from 'high-frequency' climate variations as documented by Dansgaard *et al.* (1984) or Yiou *et al.* (1990), the arbitrary choice of  $55^\circ \text{N}$  in the forcing calibration, compromises in choosing the centre frequencies, bandwidths, and time-bandwidth products for the moving coherences, and the signs from the different subtractions, complex conjugations, and time reversals, one should remember Jeffreys's (1976) footnote (p. 210) on small corrections and not be too surprised if even the signs of the various estimated corrections made here are wrong! However, they do agree with simple intuition and I believe that this model explains many of the problems in developing Quaternary timescales. I recommend that, even if the full transfer function model is not used, future timescales be developed by tuning to the equilibrium precession constant rather than the present value.

## 12. Conclusions

To summarize, this paper has included both new methods for the analysis of time series and some examples of their use. In palaeoclimate studies, the occurrence and agreement of the periodic components found in the data with Milankovitch theory should eliminate any lingering doubts about the fundamental validity of the theory. Details of the theory require refinement, however, as amplitudes of the components do not seem to be in good agreement with predictions of the simple insolation model. I have proposed glacial modulation of precession via changes in the Earth's principal moments of inertia as an explanation for the observed discrepancy between obliquity and precession tuning. Details of the transfer function estimated between global ice volume and changes in precession should not, perhaps, be taken too seriously as they depend on our current age scales: the existence of such a transfer function at the estimated levels requires the age scales to be modified. Despite these uncertainties, I believe that the luni-solar precession constant must be modified to be valid on Pleistocene timescales.

In statistical analysis of time series, the multiple-window method has been extended to include structure in the inner band. The quadratic inverse theory shows that structure in the spectrum can be resolved down to near half the Rayleigh resolution and gives a test for the presence of resolvable, but unresolved, structure. Also, the method gives a new basis set for constructing approximate maximum-likelihood estimates of the spectrum.

Finally, where does this paper leave these subjects? For the analysis of palaeoclimates an obvious next step is to produce an equivalent of the SPECMAP stack using obliquity tuning; this can then be used to derive better transfer functions to precession than the one given here. Hopefully, the improvement in timescales will be adequate to allow identification of the nonlinear terms needed to bring insolation formulae into agreement with observations.

In the area of spectrum estimation, the quadratic inverse theory probably raises more problems than it has solved; for univariate stationary series the full quadratic inverse problem in mixed spectra is of interest, as are extensions to bispectra and higher moments. A new method of estimating bivariate transfer functions was sketched but considerable refinement is desirable, and the same is true of the approximate maximum-likelihood procedure. Methods to include physically derived smoothness constraints in transfer functions are needed as are ways to include the multiple-window continuity equations. A vast area needing work is that of mildly non-stationary time series; the quadratic inverse method appears to carry over to this problem, but there are many missing details. For the analysis of non-stationary processes the wavelet formalism developed by Daubechies (1988, 1990) appears natural and Walden (1990b) has used multiple-window methods to estimate wavelets.

It is my pleasure to thank Professor N. Pisias for his V22-174 data, Professor J. Imbrie for the stack data and much advice and encouragement, and C. R. Lindberg, L. J. Lanzerotti, C. L. Mallows and J. C. Lagarias for comments on the typescript.

## Appendix A. Slepian functions and sequences

The notation introduced by Slepian (1978) for the discrete prolate spheroidal wave functions, here called Slepian functions, is convenient for analysis but statistical

work, and in particular programming, is somewhat awkward. There are two reasons for this: first, the relations between the  $U$ s and  $v$ s, Slepian's (1978) eq. (26)

$$U_k(N, W; f) = \epsilon_k \sum_{n=0}^{N-1} v_n^{(k)}(N, W) \exp[i2\pi f(n - \frac{1}{2}(N-1))], \quad (\text{A } 1)$$

where  $\epsilon_k$  is 1 for  $k$  even and  $i$  for  $k$  odd, has the advantage of both  $U$  and  $v$  being real, but uses a positive exponential transform from time to frequency whereas conventional usage in engineering and time-series analysis dictates use of a negative Fourier transform for analysis. The footnote on p. 283 of Stratton (1941), however, claims advantages for the opposite choice in electromagnetic theory problems where both space and time variables are involved. Expressions containing mixes of the conventional (negative exponent) Fourier transforms with the  $U$ s also led to conceptional problems distinguishing between products and convolutions. Finally, even though both the sequences and functions are real, many people are not used to seeing Fourier transforms written in centred form. This appendix formalizes and extends the notation introduced in Thomson (1990).

I assume  $W$  is the analysis bandwidth width with  $0 < W < \frac{1}{2}$ , and that  $N$  is the number of data samples. The Slepian sequences are solutions of the symmetric Toeplitz matrix eigenvalue problem

$$\lambda_k v_n^{(k)} = \sum_{m=0}^{N-1} \frac{\sin\{2\pi W(n-m)\}}{\pi(n-m)} v_m^{(k)} \quad (\text{A } 2)$$

standardized to be real, positive at the centre of the range for even sequences, and to have positive central slope for odd sequences with, as usual

$$\sum_{n=0}^{N-1} v_n^{(j)}(N, W) v_n^{(k)}(N, W) = \delta_{j,k} \quad (\text{A } 3)$$

and are ordered by their eigenvalues

$$1 > \lambda_0 > \lambda_1 > \dots > \lambda_{N-1} > 0. \quad (\text{A } 4)$$

Taken as an extrapolation rule, equation (A 2) also defines the Slepian sequences for  $n$  outside  $[0, N-1]$ . Slepian sequences, functions, and eigenvalues are assumed to depend implicitly on both  $N$  and  $W$ ; when the explicit meaning is needed they are denoted by  $v_n^{(j)}(N, W)$ ,  $\lambda_j(N, W)$ , etc. Be cautioned, however, that although (A 2) is an algebraic eigenvalue equation, solving it is emphatically not the preferred method for computing these sequences and the method sketched in Appendix B is numerically much better. The reason is that about  $2NW$  of the eigenvalues are exponentially close to 1 (asymptotically in  $NW$ ), and the remainder are similarly close to 0, and with sensitivity of such computations including terms of  $(\lambda_j - \lambda_k)^{-1}$ , numerical conditioning is vile.

The Slepian functions (discrete prolate spheroidal wave functions) are here defined by

$$V_k(f) = \sum_{n=0}^{N-1} v_n^{(k)}(N, W) e^{-i2\pi n f}. \quad (\text{A } 5)$$

The  $V$ s ( $V_k(N, W; f)$  in explicit notation) are thus complex functions of frequency so, by elementary properties of the Fourier transform,  $\overline{V}_k(f) = V_k(-f)$ .

Comparison with Slepian's equation (26) gives

$$V_k(f) = \epsilon_k^{-1} e^{-i2\pi f(\frac{1}{2}(N-1))} U_k(-f) \quad (\text{A } 6)$$

and

$$U_k(f) = \epsilon_k e^{-2\pi f(\frac{1}{2}(N-1))} V_k(-f). \quad (\text{A } 7)$$

The inverse transform, Slepian's eq. (28), becomes

$$v_n^{(k)}(N, W) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi n f} V_k(f) df \quad 0 \leq n \leq N-1 \quad (\text{A } 8)$$

$$= 0 \quad \text{otherwise.} \quad (\text{A } 9)$$

Similarly, the extension of the definition of the  $v_s$  outside  $[0, N-1]$  is given by the analogy of Slepian's eq. (29)

$$v_n^{(k)}(N, W) = \frac{1}{\lambda_k} \int_{-W}^W e^{+i2\pi n f} V_k(f) df \quad \forall n. \quad (\text{A } 10)$$

For the integral equation the Dirichlet kernel is given by

$$\mathbb{D}_N(f-f') = \sum_{n=0}^{N-1} e^{-i2\pi n(f-f')} = \frac{\sin\{N\pi(f-f')\}}{\sin\{\pi(f-f')\}} e^{-i2\pi(f-f')(\frac{1}{2}(N-1))} \quad (\text{A } 11)$$

so Slepian's integral equation

$$\lambda_k U_k(f) = \int_{-W}^W \frac{\sin\{N\pi(f-f')\}}{\sin\{\pi(f-f')\}} U_k(f') df' \quad (\text{A } 12)$$

becomes

$$\lambda_k V_k(f) = \int_{-W}^W \mathbb{D}_N(f-f') V_k(f') df' \quad (\text{A } 13)$$

and the orthogonality conditions are

$$\int_{-W}^W V_j(f) \overline{V_k(f)} df = \lambda_j \delta_{jk} \quad (\text{A } 14)$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} V_j(f) \overline{V_k(f)} df = \delta_{jk}. \quad (\text{A } 15)$$

In addition note the useful identity

$$V_k(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbb{D}_N(f-f') V_k(f') df'. \quad (\text{A } 16)$$

The spectral representation of a matrix gives

$$\frac{\sin\{2\pi W(n-m)\}}{\pi(n-m)} = \sum_{j=0}^{N-1} \lambda_j v_n^{(j)} v_m^{(j)} \quad (\text{A } 17)$$

and, via the standardization implicit in (A 14), Mercer's theorem is

$$\frac{\sin\{N\pi(f-f')\}}{\sin\{\pi(f-f')\}} = \sum_{j=0}^{N-1} U_j(f) U_j(f') \quad (\text{A } 18)$$

for  $|f|, |f'| \leq W$ , and

$$\mathbb{D}_N(f-f') = \sum_{j=0}^{N-1} V_j(f) \overline{V}_j(f'). \quad (\text{A } 19)$$

Taking the Fourier transform of (A 19) with respect to  $f'$  gives the identity similar to that noted by Frieden (1969)

$$e^{-i2\pi n f} = \sum_{j=0}^{N-1} v_n^{(j)} V_j(f) \quad (\text{A } 20)$$

valid for  $|f| \leq W$  and  $0 \leq n \leq N-1$ . It is convenient to have a symbol for an orthonormal version of the  $V_k$ s on the inner domain and I define

$$\mathcal{V}_k(f) = V_k(f)/\sqrt{\lambda_k} \quad (\text{A } 21)$$

so the equivalent of (A 14) becomes

$$\int_{-W}^W \mathcal{V}_j(f) \overline{\mathcal{V}}_k(f) df = \delta_{jk} \quad (\text{A } 22)$$

with the inner domain projection operator

$$\mathbb{P}_W(f, g) = \sum_{k=0}^{K-1} \mathcal{V}_k(f) \overline{\mathcal{V}}_k(g) = \sum_{k=0}^{K-1} \frac{1}{\lambda_k} V_k(f) \overline{V}_k(g). \quad (\text{A } 23)$$

## Appendix B. Computation of Slepian sequences

The appendix of Thomson (1982) gave an approximate method for computing discrete prolate spheroidal wave functions where gaussian quadrature was used to convert their integral equation to an algebraic eigenvalue equation. This method works well for reasonably small time-bandwidth products but, because the eigenvalues of the integral equation are exponentially close to either 0 or 1, produces mixtures of the eigenfunctions for larger time-bandwidth products.

Here I describe an improved method for computing the Slepian, or discrete prolate spheroidal, sequences directly. This method, which I have been using since late 1982, finds the eigenvalues of the differential equation; these are well separated and the procedure is much more stable than its predecessor. In his 1978 paper Slepian gave a solution to the differential equation for the discrete wave functions as a trigonometric series. The coefficient matrix of this series, Slepian's eq. (14), had eigenvalues  $\theta$  and successive terms in the series obeyed a second-order difference equation, Slepian's eq. (25). The matrix corresponding to this difference equation is doubly symmetric and tri-diagonal. Its diagonal elements are

$$\sigma_{ii} = (\frac{1}{2}(N-1)-i)^2 \cos(2\pi W) \quad (\text{B } 1a)$$

and on the sub-diagonal

$$\sigma_{i,i-1} = \frac{1}{2}i(N-i). \quad (\text{B } 1b)$$

Since the differential eigenvalues,  $\{\theta\}$ , are well separated, the form provides an excellent computational basis. This method has also been used by Durrani & Chapman (1984).

The original procedure took advantage of a trick suggested by Slepian (1977) by using the known symmetries of the sequences to divide the problem into odd and

even sub-problems. Usually an odd–even reduction of a symmetric matrix results in a non-symmetric matrix, but Toeplitz matrices and those corresponding to symmetrized polar kernels are exceptions. Here, although (B 1) is neither of these, it is both tri-diagonal and doubly symmetric (and so ‘locally Toeplitz’ about the centre of symmetry) and retains the symmetric tri-diagonal structure when split into odd and even sub-problems. This results in two problems of size  $\frac{1}{2}N$  (or  $1 + \frac{1}{2}N$ ) for the even sequences when  $N$  is odd) whose eigenvalues are split twice as far as those of the full problem. The EISPACK routines (Smith *et al.* 1976) BISECT and TINVIT have been used to find Slepian sequences for  $N$  up to about 20000 points with no observed problems on a Cray X-MP using single precision. Problems have been observed, however, with sequences as short as a few hundred points with 32 bit floating point representations.

Slepian functions are computed by Fourier transforming the corresponding sequences, and the eigenvalues from

$$\lambda_k = \int_{-W}^W |V_k(f)|^2 df \gtrsim \int_{-\frac{1}{2}}^{\frac{1}{2}} |V_k(f)|^2 df. \quad (\text{B } 2)$$

Round-off considerations make it advisable to separately compute both  $\lambda_k$  and  $1 - \lambda_k$ .

## References

- Abramowitz, M. & Stegun, I. A. 1965 *Handbook of mathematical functions*. Applied Mathematics Series 55. U.S. Dept of Commerce.
- Allen, C. W. 1976 *Astrophysical quantities*, 3rd edn. London: Athlone Press.
- Anderson, T. W. 1973 *Ann. Statist.* **1**, 135–141.
- Anderson, T. W. 1984 *An introduction to multivariate statistical analysis*, 2nd edn. Wiley.
- Arthur, M. A. & Garrison, R. E. (eds) 1986 *Paleoceanography* **1**, 369–586.
- Backus, G. E. & Gilbert, J. F. 1967 *Geophys. Jl R. astr. Soc.* **13**, 247–276.
- Backus, G. E. & Gilbert, J. F. 1968 *Geophys. Jl R. astr. Soc.* **16**, 169–205.
- Backus, G. E. & Gilbert, J. F. 1970 *Phil. Trans. R. Soc. Lond. A* **266**, 123–192.
- Bartlett, M. S. 1950 *Biometrika* **37**, 1–16.
- Belsley, D. A., Kuh, E. & Welsch, R. E. 1980 *Regression diagnostics*. New York: Wiley.
- Berger, A. L. 1977 *Celestial Mech.* **15**, 53–74.
- Berger, A. L. 1978a *J. atmos. Sci.* **35**, 2362–2367.
- Berger, A. L. 1978b Contribution no. 18, Institut D'Astronomie et de Geophysique Georges LeMaitre, Universite Catholique de Louvain.
- Berger, A. L. 1988 *Rev. Geophys.* **26**, 623–658.
- Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B. (eds) 1984 *Milankovitch and climate*, vol. I, II. Dordrecht: Reidel.
- Berger, T. 1971 *Rate distortion theory*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Berk, K. N. 1974 *Ann. Statist.* **2**, 489–502.
- Blackman, R. B. & Tukey, J. W. 1958 *Bell System Tech. J.* **37**. (Reprinted by Dover.)
- Bolt, B. A. & Brillinger, D. R. 1979 *Geophys. Jl R. astr. Soc.* **59**, 593–603.
- Bretthorst, G. L. 1987 Ph.D. thesis, Washington University, St Louis, U.S.A.
- Brillinger, D. R. 1975 *Time series, data analysis and theory*. Holt, Rinehart & Winston.
- Brillinger, D. R. 1981 *IEEE Trans. ASSP-29*, 1075–1076.
- Brillinger, D. R. & Krishnaiah, P. R. (eds) 1983 Time Series in the frequency domain, *Handbook of statistics*, vol. 3. Amsterdam: North-Holland.
- Bronez, T. P. 1988 *IEEE Trans. Acoust. Speech Signal Processing ASSP-36*, 1862–1873.
- Phil. Trans. R. Soc. Lond. A* (1990)

- Brouwer, D. & Clemence, G. M. 1961 *Methods of celestial mechanics*. New York: Academic Press.
- Burg, J. P., Luenberger, D. G. & Wenger, D. L. 1982 *Proc. IEEE* **70**, 963–974.
- Carter, G. C. 1987 *Proc. IEEE* **75**, 236–255.
- Chao, B. F. & Gilbert, F. 1980 *Geophys. J. R. astr. Soc.* **63**, 641–657.
- Chatterjee, S. & Hadi, A. S. 1988 *Sensitivity analysis in linear regression*. New York: Wiley.
- Chave, A. D. & Thomson, D. J. 1989 *J. geophys. Res.* **94**, 14215–14225.
- Chave, A. D., Thomson, D. J. & Ander, M. E. 1987 *J. geophys. Res.* **92**, 638–648.
- Cleveland, W. S. & Parzen, E. 1975 *Technometrics* **17**, 167–172.
- Cook, R. D. & Weisberg, S. 1982 *Residuals and influence in regression*. New York: Chapman and Hall.
- Cox, D. R. 1975 *Biometrika* **62**, 267–276.
- Cox, D. R. 1981 *Scand. J. Statist.* **8**, 93–115.
- Cramér, H. 1940 *Ann. Math.* **41**, 215–220.
- Crowley, T. J. 1983 *Rev. Geophysics Space Phys.* **21**, 828–877.
- Dahlen, F. A. 1982 *Geophys. Jl R. astr. Soc.* **69**, 537–549.
- Danby, J. M. A. 1962 *Fundamentals and celestial mechanics*. New York: Macmillan.
- Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N., Hammer, C. U. & Oeschger, H. 1984 *Climate processes and climate sensitivity* (ed. J. E. Hansen & T. Takahashi), pp. 288–298. Geophysical Monograph 29. Washington, D.C.: American Geophysical Union.
- Darwin, G. H. 1887 *Phil. Trans. R. Soc. Lond. A* **167**, 271–312.
- Daubechies, I. 1988 *IEEE Trans. Inf. Theory* **IT-34**, 605–612.
- Daubechies, I. 1990 *IEEE Trans. Inf. Theory*. (In the press.)
- Davenport, W. B. Jr & Root, W. L. 1958 *Random signals and noise*. New York: McGraw-Hill.
- Dembo, A., Mallows, C. L. & Shepp, L. A. 1989 *IEEE Trans. Inf. Theory* **IT-35**, 1206–1212.
- Doob, J. L. 1953 *Stochastic processes*. Wiley.
- Draper, N. R. & Smith, H. 1981 *Applied regression analysis*. New York: Wiley.
- Durrani, T. S. & Chapman, R. 1984 *IEEE Trans. Acoust. Speech Signal processing* **ASSP-32**, 716–721.
- Dzhaparidze, K. 1986 *Parameter estimation and hypothesis testing in spectral analysis of stationary time series*. New York: Springer-Verlag.
- Eberhard, A. 1973 *IEEE Trans. AU-21*, 37–43.
- Eddy, J. A. 1983 *Weather and climate responses to solar variations* (ed. B. M. McCormac), pp. 1–15. Boulder: Colorado University Press.
- Efron, B. 1982 *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: SIAM.
- Egbert, G. D. & Booker, J. R. 1989 *J. geophys. Res.* **94**, 14227–14247.
- Egbert, G. D. 1989 *J. geophys. Res.* **94**, 14249–14265.
- Evans, D. L. & Freeland, H. J. 1977 *Science, Wash.* **198**, 528–530.
- Fairbanks, R. G. 1989 *Nature, Lond.* **342**, 637–642.
- Forster, F. & Vezzosi, G. 1987 *Proc. ICASSP*, 53.3.1–53.3.4.
- Franke, J., Hardle, W. & Martin, R. D. (eds) 1984 *Robust and nonlinear time series analysis*. New York: Springer-Verlag.
- Frieden, B. R. 1969 *Prog. Optics* **X**, 312–407.
- Gani, J. & Priestley, M. B. (eds) 1986 *J. appl. Prob.* **A23**.
- Gentleman, W. M. 1969 *Comp. J.* **12**, 160–165.
- Glynn, W. J. & Muirhead, R. J. 1978 *J. Multivariate Analysis* **8**, 468–478.
- Goldreich, P. & Toomre, A. 1969 *J. geophys. Res.* **74**, 2555–2567.
- Gori, F. 1974 *J. opt. Soc. Am.* **64**, 1237–1243.
- Gori, F. & Palma, C. 1975 *J. Phys. A* **8**, 1709–1719.
- Grenander, U. & Rosenblatt, M. 1957 *Statistical analysis of stationary time series*. Wiley.
- Phil. Trans. R. Soc. Lond. A* (1990)

- Grenander, U. & Szego, G. 1984 *Toeplitz forms and their applications*, 2nd edn. New York: Chelsea.
- Gutowski, P. R., Robinson, E. A. & Trietel, S. 1978 *IEEE Trans. Geosci. Electronics* **GE-16**, 80–84.
- Hannan, E. J. 1970 *Multiple time series*. New York: Wiley.
- Hannan, E. J. & Thomson, P. J. 1971 *J. appl. Prob.* **8**, 157–169.
- Harris, F. J. 1978 *Proc. IEEE* **66**, 51–83.
- Hays, J. D., Imbrie, J. & Shackleton, N. J. 1976 *Science, Wash.* **194**, 1121–1132.
- Hinnov, L. A. & Park, J. 1988 *The Earth's rotation and reference frames for geodesy and geodynamics* (ed. A. K. Babcock & G. A. Wilkins), pp. 221–226. Dordrecht: Kluwer.
- Hocking, R. R., Green, J. W. & Bremer, R. H. 1989 *Technometrics* **31**, 227–239.
- Hurvich, C. M. 1985 *J. Am. statist. Ass.* **80**, 933–940.
- Hurvich, C. M. 1988 *Biometrika* **75**, 485–490.
- Hyde, W. T. & Peltier, W. R. 1987 *J. atmos. Sci.* **44**, 1351–1374.
- Imbrie, J. & Imbrie, J. Z. 1980 *Science, Wash.* **207**, 943–953.
- Imbrie, J., McIntyre, A. & Mix, A. 1979 *Climate and geo-sciences* (ed. A. Berger *et al.*), pp. 121–164. Dordrecht: Kluwer.
- Imbrie, J., Hays, J. D., Martinson, D. G., McIntyre, A., Mix, A. C., Morley, J. J., Pisias, N. G., Prell, W. L. & Shackleton, N. J. 1984 *Milankovitch and climate* (ed. A. Berger *et al.*), pp. 269–305. Dordrecht: Reidel.
- Izenman, A. J. & Sarkar, S. K. 1987 *J. Am. statist. Ass.* **82**, 271–275.
- Jeffreys, H. 1976 *The Earth*. Cambridge University Press.
- Jones, A. G., Chave, A. D., Egbert, G., Auld, D. & Bahr, K. 1989 *J. geophys. Res.* **94**, 14201–14213.
- Jones, R. H. 1962 *Skandinavisk Aktuarietidskrift* **45**, 39–69; 135–153.
- Kailath, T. 1974 *IEEE Trans. Inf. Theory* **IT-20**, 146–180.
- Kaiser, J. F. 1974 *IEEE Int. Symp. Circuits Sysgems Proc.*, pp. 20–23.
- Kassam, S. A. & Poor, H. V. 1985 *Proc. IEEE* **73**, 433–481.
- Kay, S. M. & Marple, S. L. Jr 1981 *Proc. IEEE* **69**, 1380–1419.
- Kerr, R. A. 1984 *Science, Wash.* **224**, 587.
- Khurgin, Y. I. & Yakovlev, V. P. 1977 *Proc. IEEE* **65**, 1005–1029.
- Kleiner, B., Martin, R. D. & Thomson, D. J. 1979 *Jl R. statist. Soc. B* **41**, 313–351.
- Kominz, M. A., Heath, G. R., Ku, T.-L. & Pisias, N. G. 1979 *Earth planet. Sci. Lett.* **45**, 394–410.
- Koopmans, L. H. 1974 *The spectral analysis of time series*. New York: Academic.
- Kuo, C., Lindberg, C. & Thomson, D. J. 1990 *Nature, Lond.* **343**, 709–714.
- Laird, N., Lange, N. & Stram, D. 1987 *J. Am. statist. Ass.* **82**, 97–105.
- Lambeck, K. 1980 *The Earth's variable rotation: geophysical causes and consequences*. Cambridge University Press.
- Landau, H. J. & Pollak, H. O. 1961 *Bell Sys. Tech. J.* **40**, 65–84.
- Landau, H. J. & Pollak, H. O. 1962 *Bell Sys. Tech. J.* **40**, 1295–1336.
- Lanzerotti, L. J., Thomson, D. J., Meloni, A., Medford, L. V. & MacLennan, C. G. 1986 *J. geophys. Res.* **B 91**, 7417–7427.
- Larsen, J. C. 1980 *J. Geomagn. Geoelect.*, Kyoto **32** (Suppl. I), SI 89–SI 103.
- Lawley, D. N. 1956 *Biometrika* **43**, 128–136.
- Lawley, D. N. 1959 *Biometrika* **46**, 59–66.
- Lee, P. F. 1981 *IEEE Trans. Acoust. Speech Signal Processing* **ASSP-29**, 117–119.
- Le Treut, H. & Ghil, M. 1983 *J. geophys. Res.* **88**, 5167–5190.
- Lindberg, C. R. 1986 Ph.D. thesis, University of California, San Diego, Scripps Institution of Oceanography.
- Lindberg, C. R. & Thomson, D. J. 1990 *J. geophys. Res.* (In the press.)
- MacDonald, G. J. 1989 *Rev. Geophys.* **27**, 449–469.
- Malley, J. D. 1986 *Optimal unbiased estimation of variance components*. Berlin: Springer-Verlag.
- Mallows, C. L. 1967 *J. appl. Prob.* **4**, 313–329.

- Mallows, C. L. 1973 *Technometrics* **15**, 661–675.
- Markel, J. D. & Gray, A. H. Jr 1976 *Linear prediction of speech*. Berlin: Springer-Verlag.
- Marple, S. L. Jr 1987 *Digital spectral analysis*. Inglewood Cliffs, New Jersey: Prentice-Hall.
- Martin, R. D. & Thomson, D. J. 1982 *Proc. IEEE* **70**, 1097–1115.
- Meilijson, I. 1989 *Jl R. statist. Soc. B* **51**, 127–138.
- Messer, H. & Adar, Y. 1989 *Signal Processing* **18**, 413–424.
- Middleton, D. 1960 *Statistical communication theory*. New York: McGraw-Hill.
- Mignard, F. 1986 *Earth rotation: solved and unsolved problems* (ed. A. Cazenave), pp. 93–110. Dordrecht: Reidel.
- Miller, K. S. 1974 *Complex stochastic processes*. Reading, Massachusetts: Addison-Wesley.
- Miller, M. I. & Snyder, D. L. 1987 *Proc. IEEE* **75**, 892–907.
- Miller, R. G. 1974 *Biometrika* **61**, 1–15.
- Miyata, M. 1970 *J. mar. Res.* **28**, 202–214.
- Mooers, C. N. K. 1973 *Deep-Sea Res.* **20**, 1129–1141. (Also see **29**, 1267–1269 (1982) for comments and corrections by J. H. Middleton.)
- Moore, T. C. Jr., Pisias, N. G. & Dunn, D. A. 1982 *Mar. Geol.* **46**, 217–233.
- Morley, J. J. & Hays, J. D. 1981 *Earth planet. Sci. Lett.* **53**, 279–295.
- Mosteller, F. & Tukey, J. W. 1977 *Data analysis and regression*. Reading, Massachusetts: Addison-Wesley.
- Muirhead, R. J. 1982 *Aspects of multivariate statistical theory*. New York: Wiley.
- Mullis, C. T. & Scharf, L. L. 1990 *Advances in spectrum estimation* (ed. S. Haykin). Prentice-Hall.
- Munk, W. H. & Cartwright, D. E. 1966 *Phil. Trans. R. Soc. Lond. A* **259**, 533–581.
- Munk, W. H. & Hasselmann, KI. 1964 In *Studies on oceanography* (ed. K. Yoshida), pp. 339–344. Tokyo. (Reprinted by Universities of Washington Press, 1965.)
- Munk, W. H. & MacDonald, G. F. J. 1975 *The rotation of the Earth*. Cambridge University Press.
- Odlyzko, A. M. 1987 *Math. Comp.* **48**, 273–308.
- Oglesby, R. & Park, J. 1989 *J. geophys. Res.* **94**, 14793–14816.
- Papoulis, A. 1967 *Proc. IEEE* **55**, 1677–1686.
- Papoulis, A. 1973 *IEEE Trans. Inf. Theory* **IT-19**, 9–12.
- Park, J., Lindberg, C. R. & Thomson, D. J. 1987 *Geophys. Jl R. astr. Soc.* **91**, 755–794.
- Park, J., Lindberg, C. R. & Vernon, F. L. III 1987 *J. geophys. Res.* **92**, 12675–12684.
- Parker, R. L. 1977a *Rev. Geophys. Space Phys.* **15**, 446–456.
- Parker, R. L. 1977b *A. Rev. Earth planet. Sci.* **5**, 35–64.
- Parzen, E. 1957 *Ann. Math. Statist.* **28**, 329–348.
- Parzen, E. 1967 *Time series analysis papers*. San Francisco: Holden-Day.
- Parzen, E. (ed.) 1983 *Time series analysis of irregularly observed data*. New York: Springer-Verlag.
- Pearson, E. S. & Hartley, H. O. 1969 *Biometrika tables for statisticians*, 3rd edn. Cambridge University Press.
- Peltier, W. R. 1982 *Adv. Geophys.* **24**, 1–146.
- Peltier, W. R. 1983 *Nature, Lond.* **304**, 434–436.
- Peltier, W. R. 1985 *A. Rev. Fluid Mech.* **17**, 561–608.
- Pisias, N. G. & Moore, T. C. Jr 1981 *Earth planet. Sci. Lett.* **52**, 450–458.
- Preisendorfer, R. W. & Mobley, C. D. 1988 *Principal component analysis in meteorology and oceanography*. Amsterdam: Elsevier.
- Prell, W. L., Imbrie, J., Martinson, D. G., Morley, J. J., Pisias, N. G., Shackleton, N. J. & Streeter, H. F. 1986 *Paleoceanography* **1**, 137–162.
- Priestley, M. B. 1981 *Spectral analysis and time series I & II*. Academic Press.
- Rao, C. R. & Kleffe, J. 1988 *Estimation of variance components and applications*. Amsterdam: North-Holland.
- Rao, T. S. & Gabr, M. M. 1984 *An introduction to bispectral analysis and bilinear time series models*. New York: Springer-Verlag.
- Phil. Trans. R. Soc. Lond. A* (1990)

- Rice, J. A. & Rosenblatt, M. 1988 *Biometrika* **75**, 477–484.
- Rice, S. O. 1948 *Bell Sys. Tech. J.* **27**, 109–157.
- Rice, S. O. 1982 *Proc. IEEE* **70**, 692–699.
- Rife, D. C. & Boorstyn, R. R. 1976 *Bell Sys. Tech. J.* **55**, 1389–1410.
- Rife, D. C. & Vincent, G. A. 1970 *Bell Sys. Tech. J.* **49**, 197–228.
- Roberts, D. H., Lehar, J. & Dreher, J. W. 1987 *Astr. J.* **93**, 968–989.
- Rochester, M. G. & Smylie, D. E. 1974 *J. geophys. Res.* **79**, 4948–4951.
- Rosenblatt, M. 1961 *Q. appl. Math.* **18**, 387–393.
- Rozanov, Y. A. 1961 *Theor. Prob. Appl.* **6**, 321.
- Ruddiman, W. S., Raymo, M. E., Martinson, D. G., Clement, D. M. & Backman, J. 1989 *Paleoceanography* **4**, 353–412.
- Sachs, H. M., Webb, T. III & Clark, D. R. 1977 *A. Rev. Earth planet. Sci.* **5**, 159–178.
- Scheffé, H. 1959 *The analysis of variance*. Wiley.
- Schiffelbein, P. & Dorman, L. 1986 *J. geophys. Res.* **91**, 3821–3835.
- Schonwiese, C.-D. 1987 *J. Climate appl. Meteorology* **26**, 1723–1730.
- Schreiner, A. E. & Dorman, L. M. 1990 *J. acoust. Soc. Am.* (In the press.)
- Schuster, A. 1898 *Terr. Magn.* **3**, 13–41.
- Schwartz, M., Bennett, W. R. & Stein, S. 1966 *Communications systems and techniques*. New York: McGraw-Hill.
- Seber, G. A. F. 1984 *Multivariate observations*. New York: Wiley.
- Shackleton, N. J. 1987 *Quaternary Sci. Rev.* **6**, 183–190.
- Shackleton, N. J. & Imbrie, J. 1990 *Climatic Change* **16**. (In the press.)
- Shackleton, N. J., Imbrie, J. & Pisias, N. G. 1988 *Phil. Trans. R. Soc. Lond. B* **318**, 679–688.
- Shackleton, N. J., Duplessy, J.-C., Arnold, M., Maurice, P., Hall, M. A. & Cartlidge, J. 1988 *Nature, Lond.* **335**, 708–711.
- Shumway, R. H. 1988 *Applied statistical time series analysis*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Shure, L. & Chave, A. D. 1984 *J. geophys. Res. B* **89**, 2497–2499.
- Siddiqui, M. M. & Wang, C.-C. 1984 *J. geophys. Res.* **89**, 7195–7201.
- Slepian, D. 1953 *IRE Trans. IT-3*, 68–89.
- Slepian, D. 1964 *Bell Sys. Tech. J.* **43**, 3009–3057.
- Slepian, D. 1976 *Proc. IEEE* **64**, 292–300.
- Slepian, D. 1977 *Bell Laboratories Tech. Memo.* 77-1218-8.
- Slepian, D. 1978 *Bell Sys. Tech. J.* **57**, 1371–1429.
- Slepian, D. 1983 *SIAM Rev.* **25**, 379–393.
- Slepian, D. & Pollak, H. O. 1961 *Bell Sys. Tech. J.* **40**, 43–64.
- Sjith, B. T. et al. 1976 *Matrix eigensystem routines – EISPACK guide*. Lecture Notes in Computer Science no. 6. Springer-Verlag.
- Stigler, S. M. & Wagner, M. J. 1987 *Science, Wash.* **238**, 940–944.
- Stratton, J. A. 1941 *Electromagnetic theory*. New York: McGraw-Hill.
- Tan, Sze M. 1986 *Mon. Not. R. astr. Soc.* **220**, 971–1001.
- Tanimoto, T. 1989 *Geophys. Res. Lett.* **16**, 389–392.
- Thompson, M. L. 1978 *Int. statist. Rev.* **46**, 1–19; 129–146.
- Thomson, D. J. 1977 *Bell Sys. Tech. J.* **56**, 1769–1815; 1983–2005.
- Thomson, D. J. 1982 *Proc. IEEE* **70**, 1055–1096.
- Thomson, D. J. 1986 *Proc. Third IEEE ASSP workshop on spectrum estimation and modelling*, pp. 103–105.
- Thomson, D. J. 1990 *Phil. Trans. R. Soc. Lond. A* **330**, 601–616.
- Thomson, D. J. & Chave, A. D. 1990 *Advances in spectrum analysis* (ed. S. Haykin), ch. 2. Prentice-Hall.

- Thomson, D. J., Lanzerotti, L. J., Medford, L. V., MacLennan, C. G., Meloni, A. & Gregori, G. P. 1986 *Geophys. Res. Lett.* **13**, 525–528.
- Thomson, D. J., Robbins, M. F., MacLennan, C. G. & Lanzerotti, L. J. 1976 *Phys. Earth planet. Interiors* **12**, 217–231.
- Toraldo di Francia, G. 1969 *J. opt. Soc. Am.* **59**, 799–804.
- Tsao, J. & Steinberg, B. D. 1988 *IEEE Trans. Ant. Prop.* **36**, 543–556.
- Tukey, J. W. 1967 In *Spectral analysis of time series* (ed. B. Harris). Wiley.
- Tukey, J. W. 1984 *A celebration in geophysics and oceanography – 1982 in honor of Walter Munk*, pp. 100–103. Scripps Institution of Oceanography Reference Series 84–5, March 1984. (Reprinted in *The collected works of J. W. Tukey*, vol. II (ed. D. R. Brillinger), pp. 1143–1153. Monterey, California: Wadsworth.)
- Van Trees, H. L. 1968 *Detection, estimation, and modulation theory*. New York: Wiley.
- Van Veen, B. D. & Scharf, L. L. 1990 *IEEE Trans. acoust. Speech Signal Processing*. (In the press.)
- Vernon, F. L. III 1989 Ph.D. thesis. University of California, San Diego, Scripps Institution of Oceanography.
- Walden, A. T. 1990a *Geophys. Prospect.* **38**. (In the press.)
- Walden, A. T. 1990b *Geophys. J. Int.* (Submitted.)
- Walker, A. M. 1971 *Biometrika* **58**, 21–36.
- Welch, P. D. 1961 *IBM J. Res. Devel.* **5**, 141–156.
- Welch, P. D. 1967 *IEEE Trans. AU-15*, 70–74.
- Whittle, P. 1952 *Trabajos Estadistica* **13**, 43–57.
- Whittle, P. 1953 *Arkiv För Matematik* **2**, 423–434.
- Wigley, T. M. L. 1976 *Nature, Lond.* **264**, 629–631.
- Wilks, S. S. 1962 *Mathematical statistics*. New York: Wiley.
- Wu, P. & Peltier, W. R. 1982 *Geophys. Jl R. astr. Soc.* **70**, 435–486.
- Wu, P. & Peltier, W. R. 1983 *Geophys. Jl R. astr. Soc.* **74**, 377–450.
- Wu, P. & Peltier, W. R. 1984 *Geophys. Jl R. astr. Soc.* **76**, 753–792.
- Yiou, P., Genthon, G., Ghil, M., Jouzel, J., Le Trout, H., Barnola, J. M., Lorius, C. & Korotkevitch, Y. N. 1990 Preprint.
- Yoder, C. F., Williams, J. G., Dickey, J. O., Schultz, B. E., Eanes, R. J. & Tapley, B. D. 1983 *Nature, Lond.* **303**, 757–762.
- Zhu, T., Chun, K.-Y. & West, G. F. 1989 *Bull. Seismo. Soc. Am.* **79**, 1054–1069.
- Zhurbenko, I. G. 1979 *Scand. J. Statist.* **6**, 49–56.
- Zhurbenko, I. G. 1980 *Theor. Prob. Appl.* **25**, 466–480.
- Zhurbenko, I. G. 1983 *Theor. Prob. Appl.* **28**, 409–419.