PSE Melvyl Request

From: Judith Ng
Sent: Thursday, September 24, 2015 5:00 PM
To: PSE Melvyl Request
Subject: FREE CDDS: AE

Saito 2015092458

FREE CDDS: AE
PSE TK5102.5 .A334 1991 v.1

Our Number
7090626
Status Idle Authorisation Status Un-Authorised Printed Date 24 Sep 2015

Request From UCD Shields Library

Bibliographic Details Title Advances in Spectrum Analysis and Array Processing ISBN/ISSN Control Number Publisher Prentice-Hall Verification source MELVYL-UCLinks-sfx:citation Article Details Volume/Part 1 Article Title Jackknifed error estimates for spectra, coherence, and transfer functions Article Author Thomson, David D J Article Date 1991 Pagination 58-113

Action	Date	Item Notes
REQUEST		Please send me an electronic copy (i.e., pdf file) of this article instead of hardcopies.

Patron Details Name *SAITO, NAOKI (Faculty 04 [dds])* Email saito@math.ucdavis.edu Patron Barcode 21175002752314

Rota

- OCLC - Direct Request

Pickup Location UCD Shields Library

2

Jackknifed Error Estimates for Spectra, Coherences, and Transfer Functions

David J. Thomson and Alan D. Chave

2.1 INTRODUCTION

Most frequency-domain methods for analyzing time series are copies of classical statistical procedures parameterized by frequency. Once Fourier transforms of a data sequence are taken, estimating a power spectrum involves similar steps to computing a variance, coherence is analogous to correlation, and transfer functions correspond to linear regression. An essential feature of accepted statistical methods is that they provide *both* an estimate *and* a measure of its accuracy. Traditionally these accuracy estimates, or confidence intervals, were determined analytically. However, because "analytically" commonly involved distributions derived from the Gaussian, the resulting error estimates were frequently unsatisfactory in practice. Following Tukey's [1] suggestion that Quenouille's [2] method of bias reduction could be extended to estimate variances, the *jackknife method* was developed and began to depose distribution-based methods in many statistical problems. Among the reasons for this displacement were the simplicity of the jackknife, its applicability in complicated situations, and its lack of distributional assumptions, resulting in greater reli-

ability in practice. Since then, other nonparametric and distribution-free methods for describing the statistical properties of estimated parameters, notably the bootstrap and cross-validation, have also come into common statistical use.

In the simplest form of the jackknife, we take a set of N observations $\{x_1, x_2, \ldots, x_N\}$ and form estimates of a parameter by deleting each observation in turn. Thus, in addition to the usual estimate based on all N observations, we have N estimates each based on $N - 1$ observations $\{x_2, x_3, \ldots, x_N\}$, $\{x_1, x_3, \ldots, x_N\}$ $\ldots$, $\{x_1, x_2, \ldots, x_{N-1}\}$. Combinations of these give estimates of both bias and variance valid under a wide range of parent distributions. For an introduction to the jackknife, see R. G. Miller [3], Mosteller and Tukey [4], or Efron [5]. An elementary review can be found in Efron and Gong [6].

Like any other statistical procedure, it is obvious that no time-series analysis is complete without placing reliable confidence limits on the result, and that an estimation procedure for which these cannot be found should be regarded as of limited use. Among other problems, spectrum estimates have the same susceptibility to deviant distributions as other statistics and so should be natural candidates for jackknifing. However, the usual requirement that the data used in the jackknife be independent, and the obvious fact that we cannot take a time series of N points, delete an interior point, and treat it as a series of $N - 1$ points, has had an unfortunate result summarized in Miller's [3] review paper:

> An area in which the jackknife has had little or no success is time series analysis. This is an ironic twist because it was for a time series problem that Quenouille originally proposed the idea. ... For example, no one has successfully found a way to make it provide valid estimates of the variability of smoothed estimates of the spectral density.

It is the purpose of this chapter to supply such methods.

Why use the jackknife for time-series problems in preference to more established* methods? It is widely claimed that Fourier transforms of data are nearly Gaussian, so standard multivariate theory should provide accurate tolerances. Furthermore, distribution theory provides "optimum" estimates for a given problem, whereas the jackknife, being a general-purpose tool, is suboptimal in many cases. While these arguments are undoubtedly true for some time series, we claim that they can be grossly misleading and will give examples of this with real data. Trouble often begins when distribution-based confidence limits are used for spectral estimates computed from data. The following list is a short compendium of reasons why such calculations can go astray.

- Because of the serial correlations present in time series, it is difficult to assess the data distribution using common statistical tests devised for independent and identically distributed (IID) data in the time domain. These correlations

*That is, Pre, 1960s.

typically cause such tests to indicate short-tailed distributions even when moderate-sized outliers are present. Independence of the Fourier transforms of windowed sections of data at a given frequency is a better approximation for reasonable subsection offsets, so that most distributional tests could be employed in the frequency domain *if* the series were stationary. Unfortunately, this also implies that an N-point distribution test requires N *segments* as opposed to N samples, so much more data is required and sensitivity to nonstationarity is increased. Thus, in either domain, the results of distribution tests are likely to be unreliable.

- Even when it is granted that the original, time-domain, data are not Gaussian, it is almost invariably assumed that Fourier transforms of them are multivariate normal. It is commonly argued that the discrete Fourier transform is simply a sum of many terms, and hence should be nearly normal by a loose central limit argument. While this argument appears to be reasonably valid in many series, it fails in the presence of moderate nonstationarity and when outliers are present, particularly when the process has a complicated (i.e., nonwhite) spectrum.
- Even for stationary Gaussian processes, most of the probability distributions associated with spectra are complicated, especially for multivariate problems (e.g., the complex Wishart distribution and the derived distributions of coherence and phase, the complex multivariate-t distribution for transfer functions in Brillinger [7] or Goodman [8].) Consequently, asymptotic forms or limited simulations of the true distribution are generally used. The range of validity of these approximate forms is often difficult to establish, and their undetected breakdown can lead to significant errors.
- Many of the approximate distributions are χ^2 or t characterized by a "degrees-of-freedom" parameter, and accurate computation of the degrees of freedom is essential to obtain useful confidence limits, especially when few degrees of freedom are available. While this is straightforward in the absence of estimator-induced correlations and heteroscedasticity (heterogeneity of variance), computation of spectra from real data can encounter pitfalls. The use of a data taper or window causes significant correlation of nearby frequencies within a given raw estimate, and, when the Welch method with overlapped subset estimates is used, the estimates from different subsets are serially correlated. In addition, any signal whose bandwidth is comparable to the reciprocal series, or even subset, length will be highly correlated between subsets. Quantitative correction for these effects is usually difficult, and the use of simplifications for the degrees of freedom can be misleading.
- The best developed theory is for stationary processes, often with ergodic or mixing conditions attached, so that in modeling it is almost invariably assumed that the data are at least locally stationary. Real data, in contradistinction, are typically moderately nonstationary. One consequence of a stationarity assumption is that information at different frequencies is uncorrelated; in a nonstationary process such correlations can be significant.

- The presence of a deterministic component or spectral line at a given frequency introduces noncentrality into the distribution model. While the standard distributions for spectral quantities are complicated, their noncentral counterparts are almost hopelessly so, and parametric confidence limits in the presence of line components are not usually computed. Note that deterministic elements in data are common: trends, diurnal, seasonal, and calendar effects (with aliases) are common; and true periodic phenomena such as tides in geophysical data or intentional carriers in engineering data are often encountered.
- Less commonly recognized, but even more serious, leakage from high energy portions of the spectrum caused by poorly chosen or ad hoc data windows can significantly change the distribution in low-energy portions of the spectrum. In the simplest case of a spectrum estimate, such leakage can convert nominally independent χ^2 distributions into noncentral distributions with highly correlated (but random) noncentrality.
- Finally, and obviously, the presence of outliers can dramatically alter the effective degrees of freedom in an estimate.

These complications generally reduce the effective degrees of freedom, and unless properly accounted for, standard distribution-based confidence limits will be too small. Jackknifing or similar methods are not a guarantee against all possible errors: there can be, for example, no method, statistical or otherwise, to predict the behavior of a nonstationary series when only a short, apparently stationary, realization is observed. The principal disadvantages cited against the jackknife procedure are an increase in computational burden, possibly poorer performance than the more general bootstrap, and the imagined loss of efficiency compared to utopian statistical estimators. We do not find the arguments in favor of the bootstrap compelling for spectrum estimation work: first, it is not established that the bootstrap's performance is uniformly better than the jackknife; primarily, however, it is the intent of this paper to describe methods to, at a minimum, get error estimation for spectrum estimates into the correct decade, and the error estimates themselves perhaps within a factor of 2. Changes of a few percent on such tolerances are largely irrelevant and, at present, usually not worth the large increase in computation overhead entailed by the bootstrap. Moreover, once the proper entities for jackknifing have been identified, the extension to the bootstrap is conceptually trivial.

In this chapter, the jackknife is adapted to the estimation of confidence limits on spectra, coherences, transfer functions, and some related quantities. The jackknife procedure is first outlined and computational procedures are described, especially for complex entities. An emphasis is placed on the use of jackknifing transforms because of the marked nonnormality of many of the distributions associated with spectral quantities. The jackknife is then applied to the computation of confidence limits on power spectra, coherences, and transfer functions using the Welch method, and to Thomson's [9] multiple-window method of spectral analysis. We have used two kinds of data. The first consists of simulations and are not presented;

here we have checked that the ensemble average of the jackknife variance estimates agrees with ensemble variances, and, when possible, that estimated variances agree with Gaussian theory. The examples given illustrate the advantages of jackknifed confidence limits when the data distribution is complicated (and unknown) and the degrees of freedom are not readily apparent; these all use real data. Because of the relative simplicity of the method and its greater reliability with real data, it is recommended that the jackknife be used routinely as a replacement for, or preferably, in parallel with parametric estimates. In many of the examples we use both methods; when they differ careful study is required, but we believe the jackknife.

Finally, before turning to details we note that it has been proposed, Swanepoel and van Wyk [10], to bootstrap spectrum estimates by fitting a parametric model and treating the prediction residuals as permutable. Why not do this in preference to the frequency-domain techniques advocated here? Our answer is simply that each of the present authors has analyzed many different time series from scientific and engineering problems and neither of us has yet seen an example which could reasonably be described by an autoregressive model. This is not to say that such examples do not exist, but rather that the probability of encountering such an example appears to be so low that starting with the assumption that a given time series can be described by a low-order AR or ARMA model is a prescription for trouble.

.2 THE JACKKNIFE

Let $\{x_i\}$, $i = 1, \ldots, N$ be a sample of N independent observations drawn from some distribution characterized by a parameter θ which is to be estimated, and let $\hat{\theta}$ be an estimate of θ. Denote an estimate of θ made using all N observations by $\hat{\theta}_{all}$. Further, suppose that the data are subdivided into N groups of size $N - 1$ by deleting each entry in turn from the whole set, and let the estimate of θ based on the ith subset (the subset with the ith observation deleted) be

$$\hat{\theta}_{(i)} = \hat{\theta}\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N\}. \tag{2.1}$$

Much of the literature on the jackknife uses *pseudovalues*

$$p_i = N\hat{\theta}_{all} - (N-1)\,\hat{\theta}_{(i)}, \tag{2.2}$$

with the idea that they serve as substitute data in standard statistical procedures. The *jackknife estimate* of θ is the mean of the N pseudovalues:

$$\tilde{\theta} = \frac{1}{N}\sum_{i=1}^{N} p_i$$

$$= N\,\hat{\theta}_{all} - \frac{N-1}{N}\sum_{i=1}^{N}\hat{\theta}_{(i)}. \tag{2.3}$$

This quantity was originally introduced as a lower bias replacement for the standard estimate $\hat{\theta}_{all}$. Specifically, if the expected value of $\hat{\theta}_{all}$ can be written as

$$\mathbf{E}\{\hat{\theta}_{all}\} = \theta + \frac{a}{N} + O(\frac{1}{N^2}), \tag{2.4}$$

where **E** is the expectation operator, then (2.3) removes the a/N term, so that the bias of $\tilde{\theta}$ is reduced to $O(1/N^2)$. Unfortunately, the variability of the jackknife estimate (2.3) can be large for some statistics, Hinkley [11], and it should be used as a substitute for the standard value $\hat{\theta}_{all}$ only with caution. In addition, if $\hat{\theta}_{all}$ is a linear unbiased efficient estimate, in the sense of the Gauss-Markov theorem, $\tilde{\theta}$ being different requires $\text{var}\{\tilde{\theta}\} \geq \text{var}\{\hat{\theta}_{all}\}$.

A more important application for the jackknife is in the nonparametric estimation of the variance of an arbitrary statistic. The *jackknife variance* of $\hat{\theta}_{all}$ is obtained simply by using the pseudovalues in the standard formula

$$\text{var}\{\hat{\theta}_{all}\} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (p_i - \tilde{\theta})^2$$

or, more conveniently, directly from the delete-one estimates

$$\text{var}\{\hat{\theta}_{all}\} = \frac{N-1}{N} \sum_{i=1}^{N} \left[\hat{\theta}_{(i)} - \theta_{(\cdot)}\right]^2, \tag{2.5}$$

where

$$\theta_{(\cdot)} = \frac{1}{N} \sum_{i=1}^{N} \hat{\theta}_{(i)}. \tag{2.6}$$

While (2.5) was originally introduced as a variance estimate for $\tilde{\theta}$ by Tukey [1], simulations with small samples suggest that it more accurately predicts the variance of $\hat{\theta}_{all}$, Hinkley [11]. An important property of the jackknife estimate of variance is its conservatism; Efron and Stein [12] have shown that the expected value of the jackknife variance is always larger than the true variance even when the data are not identically distributed; see also Shao and Wu [13]. Moreover, Cover and Thomas [14] have shown that for generally correlated data, the differential entropy

$$H = \boldsymbol{E}\{-\ln p(x)\} = -\int p(x) \ln p(x)\, dx,$$

where $p(x)$ is a *multivariate probability density function* that satisfies the inequality

$$H_{all} \leq \frac{N}{N-1} H_{(\cdot)}.$$

the quantity H_{all} is the *differential entropy* of the N-dimensional distribution, and $H_{(\cdot)}$ is the average of the N-differential entropies $H_{(i)}$ of the $N-1$ dimensional distributions obtained by deleting the jth variate. Recall that among distributions with equal *covariance matrices* C, entropy is maximized by the Gaussian with the maximum value being $\frac{1}{2} \ln(2\pi e)^n$ det C. This again demonstrates the conservative nature of the jackknife.

Another advantage of the jackknife approach is computational simplicity with complicated statistics. For example, compare the parametric form for the variance of the magnitude-squared coherence implicit in (2.54) to (2.5). In addition, many jackknife procedures may be done by a simple one step downdating procedure. In these cases we compute the full model $\hat{\theta}_{all}$ first, then compute each of the $\hat{\theta}_{(j)}$'s by downdating the original calculation.

While the special case of one-at-a-time data omission has been described, the more general instance of *k*-at-a-time omission is well defined. It is not usually applied because the delete-one jackknife is more efficient and does not depend on an arbitrarily chosen subdivision for the groups. The computational burden of the delete-one jackknife is greater, but is not excessive unless large data sets are under consideration. Moreover, in a time-series context, the number of data points in the time domain is irrelevant; the jackknife is applied over independent Fourier transforms occupying a small part of the frequency domain, typically a greatly reduced number. The delete-one jackknife will be used exclusively in the remainder of this chapter.

Under general conditions, it has been shown that both $(\hat{\theta} - \theta)/\tilde{\sigma}$ and $(\tilde{\theta} - \theta)/\tilde{\sigma}$ are asymptotically normally distributed, Miller [3]. This allows approximate confidence limits to be placed on $\hat{\theta}$ using normal-theory statistics in the usual way. However, asymptotic behavior does not necessarily guarantee an accurate estimate for small samples. While Tukey [1] suggested that $(\tilde{\theta} - \theta)/\tilde{\sigma}$ would be distributed as Student's t with $N - 1$ degrees of freedom for small N, a general proof of this has proven elusive. Hinkley [15] has investigated the small sample behavior of the jackknife in some detail, and suggests that substantial errors can accrue if the Student t approximation is used blindly on statistics having markedly nonnormal distributions. However, if the statistic is transformed to one with a more normal form, as is commonly done in applied statistics, then the Student t model is reasonably accurate. There is also some evidence that confidence limits obtained with the delete-k jackknife are more reliable with small samples and when the statistic is markedly nonnormal, Hinkley [15], and it may be useful in cases where extreme accuracy is important.

The use of transformations with the jackknife are essential when the statistic being jackknifed is bounded or its distribution is strongly non-Gaussian. A heuristic justification of this statement follows from examining the pseudovalues (2.2) when $\hat{\theta}$ is a variance estimate. There is nothing to prevent negative pseudovalues from occurring, and the jackknife does not know a priori that the variance must be positive. Large or frequent negative pseudovalues will result in badly biased confidence limits. However, jackknifing the logarithm of the variance eliminates this problem, and negative pseudovalues simply appear as small estimates of the variance. A more rigorous justification of transformations with the jackknife was given by Cressie [16], and a general procedure to find the correct transformation is now available, justifying statistical lore on variance stabilizing transformations that make the distribution more Gaussian. Examples include the logarithmic transformation for variance and the inverse hyperbolic transformation for the correlation coefficient. These forms carry over directly when jackknifing power spectra or coherences.

Some additional complications appear when the jackknife is applied to the *regression problem*

$$\mathbf{Y} = \mathbf{A}\beta + \mathbf{e} \tag{2.7}$$

as in the computation of transfer functions. Regression problems are unbalanced, since the sample sizes and possibly their variances are different for the vector **Y** and the matrix **A**. Miller [17] used the ordinary balanced jackknife (2.1)–(2.5) on (2.7), deleting rows of the matrix to form the pseudovalues $\tilde{\beta}_i$, and proved the asymptotic normality of the jackknifed solution vector and its variance under general conditions. Hinkley [18] examined the small sample properties of the unbalanced jackknife, and suggested that the ensuing estimate of the variance was biased. This occurs because the pseudovalues (2.2) are defined symmetrically (in the sense that ordering is irrelevant), while the model (2.7) is asymmetric because the different rows have different leverages. (This is simply because elements of **A** far from their column means have much more effect, or leaverage, on the estimated β's than elements close to the means.) He proposed the use of a *weighted pseudovalue*

$$\tilde{\beta}_i = \hat{\beta} - N(1 - h_i)(\hat{\beta} - \hat{\beta}_{(i)}), \tag{2.8}$$

where the $\{h_i\}$ are the diagonal elements of the *hat matrix* $\mathbf{H} = \mathbf{A}(\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger$. **H** is the projection matrix from **Y** onto the column space of **A**. Its diagonal elements

$$h_i = \mathbf{a}_i^\dagger (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{a}_i \tag{2.9}$$

(with $\mathbf{a}_i$ as the ith *row* of **A** and † denoting *conjugate transpose*) measure the distance of single model points from the center of the model, and the lack of balance is reflected in their size. When $h_i = 1/N$, the ordinary pseudovalue (2.2) is obtained. The weighted jackknife estimate is just the mean of the weighted pseudovalues (2.8), and is identical to the solution of (2.7). The variance estimate is

$$\text{var}\{\hat{\beta}\} = \frac{1}{N(N-p)} \sum_{i=1}^{N} (\tilde{\beta}_i - \hat{\beta})(\tilde{\beta}_i - \hat{\beta})^\dagger, \tag{2.10}$$

where p is the number of columns in **A**. Hinkley [18] also proved that (2.10) was robust in the presence of inhomogeneity of error variance in (2.7), in contrast to the standard estimate of variance. This property is useful when computing confidence limits on transfer functions in the frequency domain.

For several parameters, a long-tailed (possibly infinite variance) distribution of the estimate is expected even when the original data are Gaussian. In such cases, the *trimmed jackknife* of Hinkley and Wang [19] applied directly to the pseudo values $\{p_i\}$ appears to offer a distinct improvement. It is easy to show that for simple statistics using ordinary averages, the pseudo-values and original data are identical and, consequently, a single outlier affects only the corresponding pseudo-value. Note, however, that this implies that not only is the usual estimate $\hat{\theta}_{all}$ contaminated, but also all but one of the delete-one estimates have compensating errors. Forming the trimmed mean on p_i reduces the severity of this problem. The formulae for trimmed mean and variance estimates from Hinkley and Wang [19] for real p_i are as follows:

Denoting the sorted values of p_i by $p_{(i)}$, the α-trimmed mean is

$$\hat{\theta}_a = \frac{1}{N - 2r} \sum_{i=r+1}^{N-r} p_{(i)} \tag{2.11}$$

where $\alpha = r/N$ and the corresponding *Winsorized variance* is given by

$$var\{\hat{\theta}_\alpha\} = \frac{N}{(N-2r)^2}\left[\frac{1}{N_i}\sum_{i=r+1}^{N-r}(p_{(i)} - \hat{\theta}_\alpha)^2 + a(p_{(r+1)} - \hat{\theta}_\alpha)^2 + a(p_{(N-r)} - \hat{\theta}_\alpha)^2\right] \tag{2.12}$$

Note that r values from each end of the distribution are "pulled in," analogous to using a Huber influence function. For complex parameters sorting is undefined so simple estimators of this type are unavailable and it is perhaps easiest to embed explicit robust estimators of the type discussed in Chave *et al* [20] in a jackknife procedure. Also, with complex data it appears possible to contaminate all the pseudo values by an outlier in $\hat{\theta}_{all}$ without corresponding problems in all the delete-one estimates so one should heed Tukey's advice "The proper place to put the robustness is in the definition of $\hat{\theta}$."

Finally, an important paper by Reeds [21] shows that under a set of conditions which appear to be reasonable for most time-series data encountered in the physical sciences, jackknifed maximum likelihood estimates behave properly. Specifically, the jackknife mean and variance estimates converge (in several senses) to the population values under only slightly weaker conditions than are required for maximum likelihood estimates. Almost sure convergence of the jackknife variance estimate, however, requires a moment condition; for the estimators considered here this is satisfied. Parr [22] shows that stringent conditions on the Fréchet differential of the functional form of the statistic being jackknifed will ensure consistency of the jackknife variance and bias estimators. Shao and Wu [23] show that for sufficiently smooth functionals the deleted jackknife is consistent and asymptotically unbiased.

2.3 SPECTRUM ESTIMATES

A major problem in time-series analysis is the choice of an algorithm that yields a spectrum estimate from a finite sequence of data such that the result is not badly biased, yet remains statistically consistent and reasonably efficient. In this chapter we use three types of spectrum estimates: conventional section averaging estimates, the newer multiple-window methods, and combinations of these. At first glance, the first two appear to be distinct approaches, but, as will be seen, they are mathematically similar. We do *not* use periodograms, smoothed periodograms, or any of the equivalent methods based on autocovariances: these are all hopelessly obsolete. In this paper the use of band averaging is restricted to secondary quantities, such as the

phase of a coherence estimate. Also, we do *not* use AR, MA, ARMA, or other parametric methods: in our experience, their performance is typically even poorer than that of the preceding group. Indeed, we attribute the failure of the jackknife in spectrum estimation problems noted in the introduction largely to the use of defective spectrum estimation methods. With such methods jackknifing will probably fail: this is hardly an indictment of the jackknife as the basic estimates are so badly biased that estimation of their variance is irrelevant.

We assume that a run of T contiguous samples of a stationary time series $x(t)$ is available. The sampling times are taken to be $0, 1, \ldots, T-1$, frequency is denoted by f, and radian frequency is denoted by $\omega = 2\pi f$. The spectrum of the sampled process will be denoted by $S(f)$ on the semiclosed interval $[-\frac{1}{2}, \frac{1}{2})$ and is extended periodically. We are primarily interested in processes where the spectrum is highly colored, (i.e., not in point processes) and we do not treat estimation of periodic or mixed processes explicitly in this chapter. Frequently, the T data samples are the result of a prewhitening operation applied to a slightly longer series: a final step of correcting for the prewhitening filter must be taken. Finally, in formulas for probability distributions of various parameters given for comparison purposes, we will invariably assume that the original process is Gaussian.

In the *section-averaging method* introduced by Welch [24], the available data are divided into N subsections each containing L samples and offset by b samples from the previous subsection. A general average (and possibly trend or known periodic components) is subtracted, the data in each section are multiplied by a data window or taper D, and Fourier transformed. The superior performance of prolate spheroidal, or Slepian, sequences as low-bias data windows is well established, Thomson [25]; they are used exclusively here. These windows are characterized by a *time-bandwidth product* c and have the property that the energy concentration of their Fourier transform in a bandwidth $W = c/L$ is the maximum possible. Note that W is defined by the subset length L, not the total data length T. As $T = L + (N - 1)b$, we obtain

$$W = \frac{c[1 + (N - 1)(b/L)]}{T} \tag{2.13}$$

for the bandwidth in units of the Rayleigh resolution $1/T$ of the complete data set. As we will see below, the stability or effective degrees of freedom of such an estimate depends on both the number of subsets N and their overlap ratio b/L.

It is useful to think of spectrum estimation in terms of the information contained in a given frequency band, and we recall that the sampling theorem requires at least $2\,W{\cdot}T$ samples to represent it. Thus, except near frequencies 0 and $\frac{1}{2}$ where the coefficients are real, we may ideally obtain an estimate with $4WT$ degrees of freedom. Consequently, we attempt to represent the information in the band $(f - W, f + w)$ by the set of N complex coefficients

$$x_k(f) = \sum_{t=0}^{L-1} D(t)\, x[t + (k - 1)b] e^{-i2\pi ft}, \tag{2.14}$$

where $k = 0, 1, \ldots, N-1$; these coefficients are used for inferences. As a simple example, consider the problem of estimating the spectrum of a univariate series, for which we form:

$$\hat{S}_{sa}(f) = \frac{1}{N}\sum_{k=0}^{N-1} |x_k(f)|^2. \tag{2.15}$$

An alternative is to define a set of N *data windows,* $D_n(t)$ as $D_k[t + (k-1)b] = D(t)$ for $0 \le t \le L - 1$ and 0 otherwise, as shown by

$$x_k(f) = \sum_{t=0}^{T-1} x(t) D_k(t) e^{-i2\pi ft}. \tag{2.16}$$

We may then consider the D_k's to be a set of basis functions with the $x_k(f)$ as the expansion coefficients of the process in the given band. If the windows are nonoverlapping, that is, if $b \ge L$, the windows are orthonormal:

$$\sum_{t=0}^{T-1} D_j(t) D_k(t) = \delta_{jk}.$$

Moreover, under reasonable conditions, the x_k's are uncorrelated so that inferences are simpler. This condition also implies that the basis is incomplete (i.e., $N << 2W{\cdot}T$) and statistically inefficient, so overlapped sections are generally used. The consequence is that the x_k's are serially correlated, so that inferences are more difficult. This will be discussed in Section 2.4, where we will also show that despite the violation of the usual independence assumption, the jackknife is still applicable.

In *multiple-window analysis* of time series, the use of an orthogonal basis is formalized. Specifically, we expand the part of the signal in a fixed bandwidth $f - W$ to $f + W$ in a series of *discrete prolate spheroidal wave functions,* hereafter referred to as *Slepian functions.* The coefficients of this expansion are given by the Fourier transform of the data windowed by the corresponding Slepian sequence:

$$x_k(f) = \sum_{t=0}^{T-1} x(t)\, v_t^{(k)}(T, W) e^{-i2\pi ft}. \tag{2.17}$$

These are then used to make inferences about the process in the band. In this approach the effects of energy at frequencies outside the band may be regarded as noise, and a least-squares solution in the selected band is obtained.

The Slepian sequences are orthonormal and have the property that, of any set of N sequences of duration T, their Fourier transforms have the maximal energy concentration in the bandwidth (−W, W). Indeed, it was precisely the study of such problems that resulted in the famous Slepian and Pollak [26] and Landau and Pollak [27] papers. In spectrum estimation terms, they provide a set of data windows with low sidelobes and leakage. These data windows depend on T, and the chosen bandwidth W, with the sidelobe behavior of the corresponding spectral windows outside the band $(-W, W)$ asymptotically exponentially decreasing with $W{\cdot}T$. There are

$\lfloor 2\,W{\cdot}T \rfloor$ windows having most of their energy in this band but, as the higher order windows become less concentrated, N is usually taken to be slightly smaller than this.

The Slepian functions are defined as solutions of the integral equation

$$\lambda_k U_k(T, W; f) = \int_{-W}^{+W} \frac{\sin T\pi(f - f')}{\sin \pi(f - f')} U_k(T, W; f')df', \tag{2.18}$$

where the eigenvalues λ_k give the fraction of the energy of $U_k(T, W; F)$ in the bandwidth $(-W, W)$. Their Fourier transforms are the *Slepian sequences*

$$v_t^{(k)}(T, W) = \frac{1}{\varepsilon_k \lambda_k} \int_{-W}^{W} U_k(T, W; f) e^{-i2\pi f\,(t - \frac{T-1}{2})} df, \tag{2.19}$$

with $\varepsilon_k = 1$ for k even and i for k odd. They are orthonormal on $[0, T-1]$. For details on the many fascinating properties of these functions, see Slepian [28].

If there are no line components in the band*, we compute a simple *multiple-window spectrum estimate* by solving the nonlinear equation

$$\sum_{k=0}^{N-1} \frac{\lambda_k[\hat{S}(f) - \hat{S}_k(f)]}{[\lambda_k \hat{S}(f) + (1 - \lambda_k)\sigma^2]^2} = 0 \tag{2.20}$$

for $\hat{S}(f)$ with σ^2, the process variance, and $\hat{S}_k(f) = |x_k(f)|^2$. Typically, this equation is solved iteratively starting with $\hat{S}(f) = [\hat{S}_0(f) + \hat{S}_1(f)]/2$. The solution $\hat{S}(f)$ is not a simple windowed transform of the data, but the root of a rational equation in such forms. Thus the autocovariance function obtained by Fourier transforming $\hat{S}(f)$ is not necessarily zero for lags greater than T. Note that this is an approximation in that the $(1 - \lambda_k)\,\sigma^2$ terms are a bound on bias from out-of-band energy, used instead of a more complicated convolution. For details, see Thomson [9] and Chapter 1.

To understand the behavior of this estimate, recall that the eigenvalues λ_k are exceptionally close to 1 (e.g., $NW = 4$ gives $1 - \lambda_0 = 3 \times 10^{-10}$), so for many problems the above estimate $\hat{S}(f)$ becomes

$$\overline{S_{mw}}(f) \approx \frac{1}{N} \sum_{k=0}^{N-1} S_k(f). \tag{2.21}$$

Each term in this series is a direct spectrum estimate, and so distributed as χ^2_2. Furthermore, if we assume that the true spectrum is flat over the interval $(f - W, f + W)$, the $\hat{S}_k$'s are uncorrelated by the orthogonality of the Slepian functions, so the estimate (2.21) will have a χ^2 distribution with $2N$ degrees of freedom.

* Simultaneous estimation of line and continuous spectrum components is covered in Thomson [9, Section XIII].

In the comparison of section-averaging and multiple-window methods, two practical problems have been ignored, namely, robustness and nonstationarity. For both problems overlapped section processing often permits use of robust estimates at a slight cost in efficiency, Chave et al. [20]. A method appearing to be both efficient and robust is to combine the two methods: use multiple windows on each section, then combine the results of different sections robustly. We take B sections of length L, but now with K windows on each section, $K \leq [2W \cdot L]$, and $N = B \cdot K$. Clearly much less section overlap is needed for efficiency than when a single low-order prolate window per section is used. Also, different robust estimators are possible; for example, we may estimate a transfer function on each section, then use a robust location estimate on the transfer function estimates directly. The smoothing may be performed using either simple or robust procedures; the latter are preferred owing to their inherent immunity from outlier contamination, Chave and others [20]. A further refinement is possible: in addition to simply obtaining an estimate on each section, we may also obtain its variance (obviously by jackknifing over the K windows within the section!); then combine sections by minimizing

$$\sum_{j=1}^{B} \rho\left(\frac{\hat{\theta}_j - \hat{\theta}}{\hat{s}_j}\right),$$

where $\hat{\theta}_j$ and $\hat{s}_j$ are the parameter and its error scale estimates for the jth section and ρ is a *loss function*. While the treatment of robust estimators is outside the scope of this chapter, we note that jackknifing such robust estimates is valid provided the influence function corresponding to ρ is sufficiently smooth.

A technical point concerns the exchangeability of the $x_k(f)$'s necessary for the validity of the jackknife. In section averaging, the covariance between coefficients from the intervals $[0, L-1]$ and $[b, b+L-1]$ using a data taper D is given by

$$\mathbf{E}\{x_1(f), \bar{x}_2(f)\} = \int_{-1/2}^{1/2} S(\eta) e^{-i2\pi\eta b} |\bar{D}(f-\eta)|^2 d\eta, \tag{2.22}$$

where S is the true spectrum of the process and $\bar{D}$ is the Fourier transform of the data window. This formula, which depends on the spectral representation of a stationary process and not on distributional assumptions, shows two sources of covariance in such estimates: first, a peak in S significantly sharper than D will generate covariance for $b > L$; second, the overlap of the windows generates a convolution term $D * D$, where $*$ denotes convolution. In the first instance, about all that can be done is to attempt to detect its presence or get a larger sample: theoretically, we count on ergodic, or better, mixing assumptions as in Rosenblatt [29]; practically, we test the coefficients from successive sections for abnormal autocorrelation or examine a Q-Q plot of the individual section spectrum estimates. The second topic is the subject of the next section.

In multiple-window methods the exchangeability requirement is that the bandwidth W be small enough that the spectrum is approximately resolved in the sense below. The equivalent of (2.22) becomes

$$\mathbf{E}\{x_j(f)\bar{x}_k(f)\} = \frac{1}{\varepsilon_j \bar{\varepsilon}_k} \int_{-W}^{W} U_j(T,W,\zeta)U_k(T,W,\zeta)S(f+\zeta)d\zeta + \eta, \qquad (2.23)$$

where ε_j is 1 for j even and i for j odd and η is bounded by the eigenvalue properties of the Slepian functions and asymptotically decreases exponentially with $W \cdot T$. Again, only the spectral representation is used, so that if the spectrum is resolved (i.e., S is constant over $(f - W, f + W)$) the coefficients are uncorrelated and have equal variance except for $O(1 - \lambda_k)$ terms. Mallows [30] has shown that the distribution of narrow-band processes such as this one (think of the $x_j(f)$'s as time series with "time" defined as the start of the observation interval) must be nearly Gaussian in a certain sense; since the first two moments agree, the distributions of the coefficients must be close in his sense. In practice, we may apply a distribution test to the different coefficients at a given frequency.

In the *combined multiple-window section averaging method* the assumptions and checks are the same as sketched above. In addition, the resolution assumption may be checked. The basis for this test is (2.23): if S is constant over $(f - W, f + W)$ the different coefficients will be uncorrelated, and we simply form their sample covariance matrix and test it for diagonal form. Let $x_l(f)$ be the vector of coefficients from the K windows on section l and

$$\hat{C}(f) = \frac{1}{B}\sum_{j=1}^{B} \mathbf{x}_j(f)\mathbf{x}_j^{\dagger}(f) \qquad (2.24)$$

be their sample covariance matrix at frequency f. For this problem, the formal Gaussian likelihood ratio test for diagonal form is the *complex sphericity test.* (The sphericity test, as it's name implies, tests the hypothesis that the multivariate distribution is the same along all coordinates against the alternative of a general elliptical shape. Thus the null hypothesis is that the covariance matrix is diagonal with equal variances, or proportional to the identity matrix, and consequently rotationally invariant.) Distributional theory for this test and a short table is contained in Nagarsenkar and Das [31]. In practice, the moderate nonstationarity found in much data make this test unrealistic and better results are obtained if the individual terms in (2.24) are normalized by power, that is, we compute

$$\hat{C}_n(f) = \frac{N}{B}\sum_{j=1}^{B} \frac{\mathbf{x}_j(f)\mathbf{x}_j^{\dagger}(f)}{\mathbf{x}^{\dagger}(f)\mathbf{x}(f)} \qquad (2.25)$$

in place of (2.24). The test of C_n for diagonal structure is known as one of uniformity (of directions on the K dimensional complex hypersphere) and is similar to the real case treated in Watson [32], Chapter 2. In applying these tests, remember that likelihood ratio tests (including the sphericity test) are notoriously sensitive to the assumption of normality so that theoretical values should not be taken too seriously. Also the complexity of the distribution of the test statistics makes determining the test significance by jackknifing over sections an attractive alternative to computing the distribution! For this application we recommend a log log transformation; the

first because log C is asymptotically χ^2, the second to make the approximately χ^2 variates more nearly normal. Again the underlying distributions are perverse enough that the usual t approximations for tail probabilities may be unreliable and it is probably safer to simply check situations where the test statistic is several jackknife standard deviations from the origin carefully.

Finally, one must remember that the spectrum is rarely, if ever, exactly resolved so, as usual, one must balance some lack of resolution against statistical stability of the estimated spectrum. This balance is also reflected in the sphericity test where the trade is between detecting serious lack of resolution and the sampling distribution of the test. The sampling distribution of the sphericity test when the null hypothesis is false is very complicated and beyond the scope of this work, however, some insight may be gained from following argument: First, assume that the spectrum is almost resolved so $\mathbf{C}$ is nearly proportional to $\mathbf{I}$ and low order Taylor series approximations to various statistics are valid. Second, note that the sphericity test is essentially Bartlett's M test applied to the eigenvalues of $\hat{\mathbf{C}}(f)$. Third, denoting the eigenvalues of $\hat{\mathbf{C}}$ by c_j, and assuming that their distribution is approximately χ^2 with μ degrees-of-freedom M is approximately $\frac{1}{2} K \mu \, var\{c\}/ave^2\{c\}$. Fourth, because the matrix Frobenius norm is invariant under orthogonal transformations, $K\,var\{c\} = \| \mathbf{C}/S_{av} - \mathbf{I} \|_F^2$ where $S_{av}(f)$ is defined to be the "average" value of $S(f)$ over $(f - W, f + W)$. Finally, using quadratic inverse theory, it may be shown that

$$\frac{1}{2WL}\|\mathbf{C}(f) - S_{av}(f)\mathbf{I}\|_F^2 \leq \frac{1}{2W}\int_{-W}^{W} [S(f+\xi) - S_{av}(f)]^2 \, d\xi$$

with near equality for $K \approx 2\,W\,L$ and the spectrum varying "slowly" over the local bandwidth. For "equal-energy" ripples in the spectrum the term on the right decreases approximately linearly with ripple frequency and essentially disappears for details in the spectrum finer than half the Rayleigh resolution, that is for $\delta f < 1/2\,L$.

.4 CORRELATION AMONG ESTIMATES

Among the problems mentioned in earlier sections are those of obtaining the correct equivalent degrees of freedom and those of correlation between estimates. Not surprisingly, the two are related. The concern is that covariance terms of the type $\mathbf{E}\{x_j(f)x_k(f)\}$ violate the usual jackknife assumption of independent errors. Note that this is distinct from the covariance arising in unbalanced problems where the error variables in (2.6) are assumed to be independent and identically distributed, but, because the different inputs have different leverages, the rows of $\mathbf{A}$ are not strictly exchangeable (in that diagonals of the hat matrix differ significantly) and the residuals will have different variances; here we assume that elements of $\mathbf{e}$ may be correlated. We give a brief outline of a method for computing the effective stability

of a spectrum estimate and of determining the effect of overlap covariances on variance estimates.

We assume B subsections each of length L samples, offset by b samples and K windows on each section, so $N = B \cdot K$. Define a $N \times T$ matrix $\mathbf{D}$ of basis functions (shown schematically in Fig. 2.1) whose jth row is given by

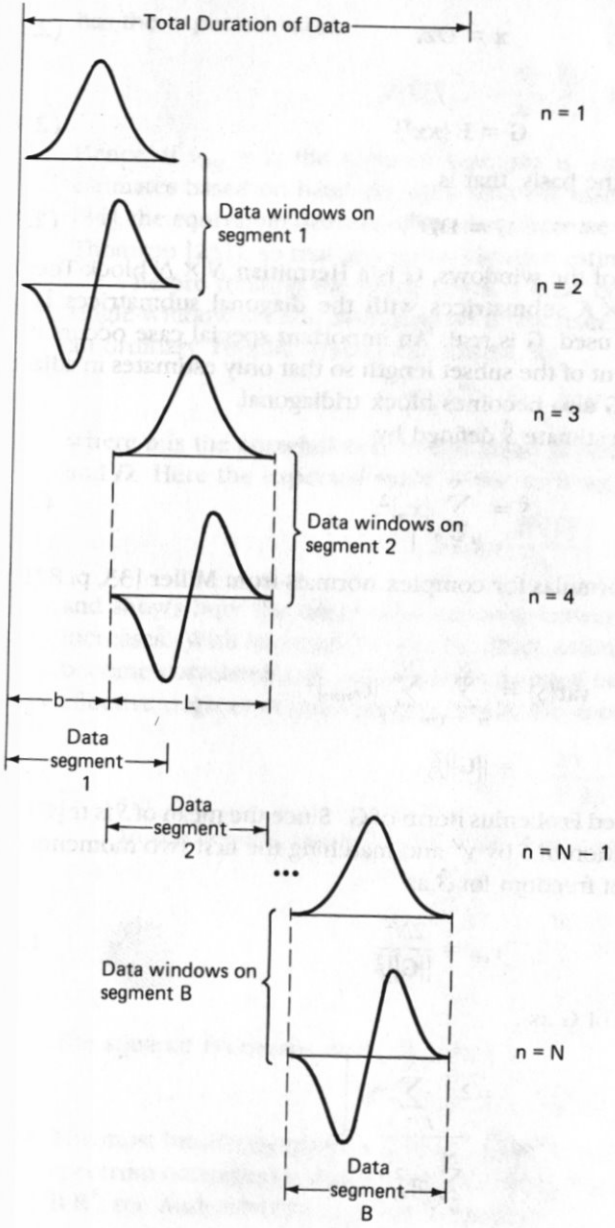


Figure 2.1 Schematic representation of a multiple-segment, multiple-window estimation process. Here two windows are used on each of $N/2$ overlapped segments.

$$D[j, t + 1 + (l - 1)b] = v_t^{(k)}(L, W)e^{-i2\pi ft}, \qquad 0 \le t \le L - 1 \tag{2.26}$$

for $j = K(l - 1) + k + 1$ with $l = 1, \ldots, B$; $k = 0, \ldots, K - 1$ and 0 otherwise. (For band-averaged estimates, the rows are simply $D(t)e^{-i2\pi fjt}$ for $j = 1, \ldots, N$.) Let $\mathbf{z}$ be a T vector of zero-mean complex Gaussian random variates with $\mathbf{E}\{\mathbf{z}\mathbf{z}^\dagger\} = \mathbf{I}_{T \times T}$, the $T \times T$ identity matrix, and define the N vector $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{D}\mathbf{z}. \tag{2.27}$$

The covariance matrix of $\mathbf{x}$

$$\mathbf{G} = \mathbf{E}\{\mathbf{x}\mathbf{x}^\dagger\} \tag{2.28}$$

is the Grammian matrix of the basis; that is,

$$\mathbf{G} = \mathbf{D}\mathbf{D}^\dagger. \tag{2.29}$$

Clearly, by the symmetries of the windows, $\mathbf{G}$ is a Hermitian $N \times N$ block-Toeplitz matrix consisting of B^2 $K \times K$ submatrices with the diagonal submatrices $\mathbf{I}_{K \times K}$. When a single frequency is used, $\mathbf{G}$ is real. An important special case occurs if the offset is more than 50 percent of the subset length so that only estimates in adjacent blocks are correlated, and $\mathbf{G}$ also becomes block tridiagonal.

Consider a spectrum estimate $\hat{S}$ defined by

$$\hat{S} = \sum_{n=1}^{N} |x_n|^2. \tag{2.30}$$

Using the fourth-moment formulas for complex normals from Miller [33, p. 82], we obtain the variance of $\hat{S}$ as

$$\begin{aligned} \operatorname{var}\{\hat{S}\} &= \sum_{n=1}^{N} \sum_{m=1}^{N} |G_{nm}|^2 \\ &= \|\mathbf{G}\|_F^2, \end{aligned} \tag{2.31}$$

which represents the squared Frobenius norm of $\mathbf{G}$. Since the mean of $\hat{S}$ is $\operatorname{tr}[\mathbf{G}] = N$, approximating the distribution of $\hat{S}$ by χ^2 and matching the first two moments gives the "equivalent" degrees of freedom for $\hat{S}$ as

$$\nu_{eq} = \frac{2N^2}{\|\mathbf{G}\|_F^2} \tag{2.32}$$

or, in the eigenvalues $\{\eta_j\}$ of $\mathbf{G}$ as

$$\nu_{eq} = \frac{2\left[\sum_{j=1}^{N} \eta_j\right]^2}{\sum_{j=1}^{N} \eta_j^2}. \tag{2.33}$$

Turning to the reliability of variance estimates, it can be shown that the estimated variance of $\bar{S} = \hat{S}/N$, given by

$$\hat{\mathbf{V}}\{\bar{S}\} = \frac{1}{N(N-1)} \sum_{j=1}^{N} (|x_j|^2 - \bar{S})^2, \tag{2.34}$$

has the expected value:

$$\mathbf{E}\{\hat{\mathbf{V}}\{\bar{S}\}\} = \frac{\mathbf{E}^2\{\bar{S}\}}{N-1}\left(1 - \frac{2}{\nu_{eq}}\right). \tag{2.35}$$

Hence, if $\nu_{eq} \approx 2$, the variance estimate is biased toward zero. In particular, for estimates based on band-averages with the essential data taper included, Brillinger [34], the equivalent degrees of freedom increase slowly with bandwidth (see Fig. 7 of Thomson [25]), so that any naive variance estimate is likely to be misleading.

Before treating the block-Toeplitz case, consider the special case of ordinary single-window section averaging with less than 50 percent overlap. Here **G** will be an ordinary Toeplitz tridiagonal matrix, so

$$\|\mathbf{G}_1\|_F^2 = N + 2(N-1)\rho, \tag{2.36}$$

where ρ is the correlation between adjacent spectrum estimates and depends on b and D. Here the expected value of the variance of $\bar{S}$ becomes

$$\mathbf{E}\{\hat{\mathbf{V}}\{\bar{S}\}\} = \frac{\mathbf{E}^2\{\bar{S}\}}{N}\left(1 - \frac{2\rho}{N}\right) \tag{2.37}$$

and shows how the effect of covariance between adjacent subsets decreases as N increases. With less than 50 percent offset, estimates two (and possibly more) apart become correlated and additional terms must be added. As an incidental note, the effective degrees of freedom here (with the approximation $N - 1 \approx N$), is

$$\nu_{eq} \approx \frac{2N}{1 + 2\rho}. \tag{2.38}$$

In the general case with less than 50% overlap, write **G** in submatrices as

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \mathbf{R} & \mathbf{0} & \cdot & \mathbf{0} \\ \mathbf{R}^\dagger & \mathbf{I} & \mathbf{R} & \mathbf{0} & \cdot \\ \mathbf{0} & \mathbf{R}^\dagger & \mathbf{I} & \mathbf{R} & \mathbf{0} \\ \cdot & \mathbf{0} & \mathbf{R}^\dagger & \mathbf{I} & \mathbf{R} \\ \mathbf{0} & \cdot & \mathbf{0} & \mathbf{R}^\dagger & \mathbf{I} \end{bmatrix}. \tag{2.39}$$

The squared Frobenius norm of matrix **G** is

$$\|\mathbf{G}\|_F^2 = B \cdot K + 2(B-1)\|\mathbf{R}\|_F^2. \tag{2.40}$$

The most intuitively pleasing form for $\|\mathbf{R}\|_F^2$ is in the canonical correlations between spectrum estimates on adjacent blocks and is simply the sum of the singular values of $\mathbf{R}\mathbf{R}^\dagger$; see Anderson [35, p. 492]. This gives

$$\mathbf{E}\{\hat{\mathbf{V}}\{\bar{S}\}\} = \frac{\mathbf{E}^2\{\bar{S}\}}{N}\left[1 - \frac{2(B - 1/B)\|R\|_F^2}{(N-1)K}\right], \tag{2.41}$$

which is asymptotically unbiased with N.

In both cases described by (2.37) and (2.41), the banded Toeplitz structure implies that the equivalent degrees of freedom increase with the block count, so that the estimate of the variance of $\bar{S}$ converges to $\mathbf{E}^2\{\bar{S}\}/N$. Because the jackknife estimate of variance is identical to (2.34), it also converges to $\mathbf{E}^2\{\bar{S}\}/N$. Also, as shown in the next section, jackknifed variance estimates are usually done with a logarithmic transformation, so, because the correlations between $\log \chi_2^2$ variates are essentially identical to those between the original χ_2 variables, the jackknife variance estimate remains asymptotically unbiased.

2.5 JACKKNIFING UNIVARIATE SPECTRUM ESTIMATES

At this stage the jackknife can be applied to obtain approximate confidence intervals on a power spectrum estimate for a univariate series. Taking the complex coefficients of (2.14) or (2.17) or a hybrid, denote a raw spectrum estimate by

$$\bar{S}_k(f) = |x_k(f)|^2$$

for either a section-averaging or multiple-window estimate. For the standard model of a stationary Gaussian process, estimates of power spectra are the sums of squares of normally distributed variates, and hence distributed as χ^2 or gamma. (For a combined multiple-window, multisection estimate, we take an average over the windows in each section, yielding $2K$ degrees of freedom per section.) For reference, the probability density function of such a spectrum estimate s with $\nu = 2m$ degrees of freedom may be written as

$$p(s) = \frac{(m/a)^m}{\Gamma(m)} s^{m-1} e^{-s\frac{m}{a}}, \tag{2.42}$$

where $\Gamma(m)$ is the gamma function. The mean and variance of s are a and a^2/m. For comparison with the corresponding jackknife results, we have for the mean

$$\mathbf{E}\{\ln s\} = \ln a + B_\chi(m),$$

where the bias $B_\chi(m)$ is given by Bartlett and Kendall [36] and Wishart [37] as

$$B_\chi(m) = \psi(m) - \ln m, \tag{2.43}$$

with ψ being the digamma function. Similarly, for the variance we have

$$\text{var}\{\ln s\} = \psi'(m), \tag{2.44}$$

where ψ' is the trigamma function. Formulas for computing both ψ and ψ' are given in Abramowitz and Stegun [38]. Because this variance is independent of the mean,

gross departures from the nominal model may be readily seen. The maximum-likelihood estimate of a given N independent samples from distribution (2.42) is, regardless of m, their arithmetic average

$$\bar{S} = \frac{1}{N}\sum_{j=1}^{N} \hat{S}_j. \tag{2.45}$$

Turning to the jackknife with more typical and less utopian data, statistical tradition implies that a logarithmic transformation of the variance (or power spectrum) stabilizes the estimation procedure by giving a more symmetric distribution than χ^2. We take our statistic to be the logarithm of the power spectrum $\theta = \ln S$ at a single frequency and denote an estimate of it from a finite data sample by $\hat{\theta} = \ln \hat{S}$. The delete-one values $\ln \hat{S}_{(j)}$ from (2.1) are

$$\ln \hat{S}_{(j)} = \ln\left(\frac{1}{N-1}\sum_{\substack{k=1 \\ k\neq j}}^{N} \hat{S}_k\right) \tag{2.46}$$

with their average

$$\ln \hat{S}_{(\cdot)} = \frac{1}{N}\sum_{j=1}^{N} \ln \hat{S}_{(j)} \tag{2.47}$$

defining a spectrum estimate $\ln \hat{S}_{(\cdot)}$. We will examine the differences between the jackknife estimate of the log power spectrum and $\ln \bar{S}$ from (2.45) below. This difference is related to the jackknife estimate of the variance of the log power spectrum given by

$$\begin{aligned}\tilde{\sigma}^2 &= \operatorname{var}\{\ln S\} \\ &= \frac{N-1}{N}\sum_{i=1}^{N} (\ln \hat{S}_{(i)} - \ln \hat{S}_{(\cdot)})^2.\end{aligned} \tag{2.48}$$

Because of the logarithmic transformation, $(\ln \hat{S}_{(j)} - \ln \hat{S}_{(\cdot)})/\tilde{\sigma}$ is nearly distributed as t_{N-1}, and approximate confidence intervals on either $\ln S$ or S can be constructed. The double-sided $1-\alpha$ confidence interval for the power spectrum is

$$\hat{S}e^{-t_{N-1}(1-\alpha/2)\tilde{\sigma}} < S \leq \hat{S}e^{t_{N-1}(1-\alpha/2)\tilde{\sigma}}. \tag{2.49}$$

For setting confidence intervals, see Hall [39]; Izenman and Sarkar [40] give improved classical expressions.

The foregoing formulas for jackknifed spectrum (or variance) estimates have interesting interpretations as the classical tests for homogeneity of variance (and simultaneously Gaussianity of distribution) of Bartlett and Lehmann (see Sugiura and Nagao [41], Ghosh [42]). Rewrite (2.47) as

$$\ln \hat{S}_{(\cdot)} = \ln \bar{S} + \ln \frac{N}{N-1} + \frac{1}{N}\sum_{j=1}^{N} \ln \beta_j', \tag{2.50}$$

with $\beta_j' = 1 - \beta_j$ and

$$\beta_j = \frac{S_j}{N\bar{S}}. \tag{2.51}$$

The distribution of the β's is individually beta and jointly Dirichlet when the original data are stationary and Gaussian. Denote the last term in (2.50) by M_0':

$$M_0' = \sum_{j=1}^{N} \ln \beta_j',$$

and recall that Bartlett's statistic for homogeneity of variances with equal degrees of freedom ν may be written as $M = -\nu M_0$, with

$$M_0 = \sum_{j=1}^{N} \ln \beta_j.$$

Except for a scale factor, M_0' is a Bartlett test applied to the delete-one averages $\hat{S}_{(j)}$. The Bartlett test assumes that the $\hat{S}_j$ are independent, while the $\hat{S}_{(j)}$ are obviously highly correlated; nonetheless, M' appears to convey similar information and empirically the correlation between M and M' is high.* Thus the difference between $\ln\bar{S}$ and the average of the delete-one values $\ln\hat{S}_{(\cdot)}$ may be interpreted as a downward "correction" term based on the homogeneity, or lack thereof, of the estimates. A second interpretation of this statistic is that when both a and m are estimated from (2.42) by maximum-likelihood, we obtain

$$B_\chi(\hat{m}) = \frac{1}{N}\sum_{j=1}^{N} \ln \beta_j$$

for the degrees of freedom, where B_χ is the χ^2 bias function (2.43). Thus the M' term can also be thought of as an "effective degrees-of-freedom" correction.

Turning to the jackknife estimate of variance, we write (2.48) as

$$L_0' = \sum_{j=1}^{N} (\ln \beta_j' - M_0')^2,$$

which is identical in form to Lehmann's test for homogeneity of variance:

$$L = \nu \sum_{j=1}^{N} (\ln \beta_j - M_0)^2$$

again with delete-one averages in place of independent variates. Hence, the jackknife estimate of variance provides a test of homogeneity of variance and Gaussianity of distribution. The presence of both the Bartlett and Lehmann forms is not surprising:

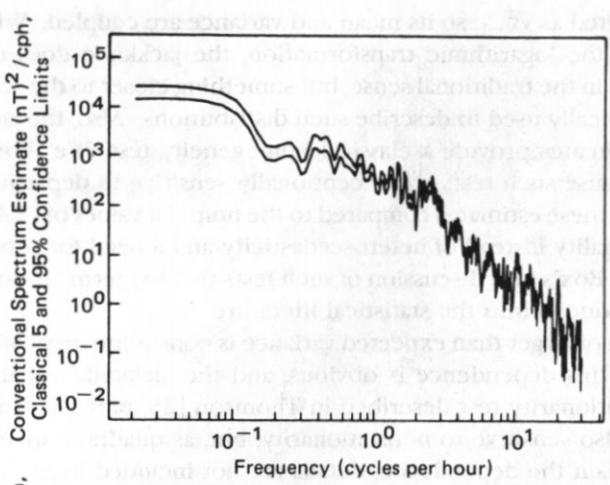
* See Han [43] for a formal discussion of the Bartlett test with correlated variances.

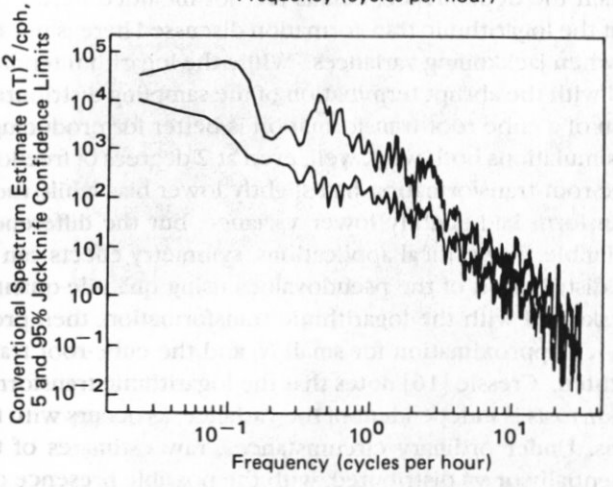
asymptotically M is distributed as χ^2_{N-1}, so its mean and variance are coupled. What is interesting is that, with the logarithmic transformation, the jackknife does not estimate mean and variance in the traditional sense, but something closer to the scale and degrees of freedom typically used to describe such distributions. Also, the jackknife bias and variance estimates provide a classical homogeneity test. We should consider this carefully because such tests are exceptionally sensitive to departures from normality: excesses in these estimates compared to the nominal values of (2.44) may well indicate nonnormality instead of heteroscedasticity and a need for robust procedures. Indeed, it is in Box's [44] discussion of such tests that the term "robust" appears to have been introduced into the statistical literature.

The usual explanation of larger than expected variance is nonstationarity. With multiple-section methods, this dependence is obvious, and the jackknife variance becomes essentially the stationarity test described in Thomson [25, part II]. Multiple-window methods are also sensitive to nonstationarity, but, as quadratic inverse theory is necessary to explain the dependence, details are not included here.

It should be noted that the logarithmic transformation discussed here is not the only type that can be used when jackknifing variances. While the logarithm tends to relieve problems associated with the abrupt termination of the sampling distribution at the origin, the application of a cube root transformation is better for producing a symmetric one. In limited simulations both work well, even at 2 degrees of freedom. Estimates based on the cube-root transformation had slightly lower bias while those based on a logarithmic transform had slightly lower variance, but the differences between the two were negligible. For critical applications, symmetry effects can be checked by examining the distribution of the pseudovalues using quantile-quantile plots; if they are markedly skewed with the logarithmic transformation, then problems can occur with the t_{N-1} approximation for small N, and the cube-root transformation should be substituted. Cressie [16] notes that the logarithmic transformation is only valid when the kurtosis is independent of the variance, as occurs with the normal and χ^2 distributions. Under ordinary circumstances, raw estimates of the power spectrum are exponentially or χ^2_2 distributed, with the possible presence of a small fraction of outliers, and this shouldn't pose problems. However, if strange results are obtained with the jackknife, then the distributions of the data, of the raw power estimates, and of the pseudovalues should be examined to detect anomalies.

Figure 2.2 illustrates application of the jackknife to power spectra. The data consist of one month of the north-south component of the Earth's magnetic field variations recorded at the Victoria, B.C., Observatory in July 1982. The geomagnetic field was particularly active during this time, and frequent short-duration magnetic storms and other events were observed, so that the data are locally nonstationary. Raw spectra were computed by Fourier transforming 71 percent overlapped two-day-long pieces of data after tapering with a prolate data window of time-bandwidth product 4, yielding 42 raw spectra. Correcting for the correlations caused by this overlap gives about 30 degrees of freedom (instead of the 84 obtained by counting sections), so the classical 5 to 95 percent points are 0.64 to 1.56 times the estimate.





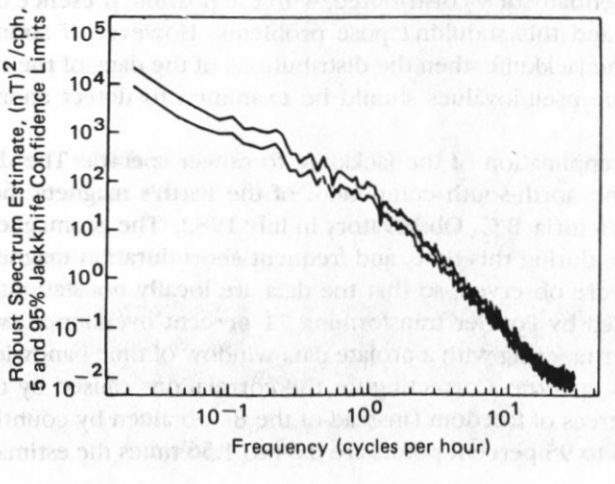


Figure 2.2 Three estimates of the spectrum of the H component from Victoria Observatory with 90 percent (5 to 95 percent) confidence interval. The top frame shows a conventional nonrobust estimate with confidence intervals from normal theory. The estimate shown in the center frame is also nonrobust but jackknife confidence intervals are used. The bottom frame shows a robust estimate with jackknife confidence intervals. Note that down weighting the few extreme segments gives a great difference in absolute level of the estimated spectrum, but confidence intervals of similar length to those shown in the top frame obtained from normal theory.

These estimates were section averaged using both an ordinary arithmetic average, Figs. 2.2(a) and 2(b), and a robust mean described by Chave et al. [20], Fig. 2.2(c). The two curves shown in each frame correspond to 5 and 95 percent confidence limits, Fig. 2.2(a) representing the classical 0.64 to 1.56 factors mentioned, and those in Figs. 2.2(b) and 2.2(c) obtained by jackknifing. Note, first, that there is a decade between the conventional and robust spectrum estimates, reflecting the dominance of the arithmetic mean by brief, energetic storm events that are eliminated by adaptive weighting during robust processing. Second, the width of the jackknife confidence bands in Fig. 2.2(b) is again about a decade larger than the classical ones of Fig. 2.2(a). Third, the width of the jackknife confidence intervals for the robust estimate, Fig. 2.2(c), are close to the distribution-based value. In this example, the problem is that the conventional average is dominated by a few outliers, the population heterogeneous, and a standard estimate of the equivalent degrees of freedom meaningless. Analysis of the data using the section-by-section multiple-window technique shows that a single event of roughly one-day duration dominates the data, so that the arithmetic average is controlled by 1 to 3 subsections and has only a few degrees of freedom. The jackknifed confidence limits account for this, as well as providing a correction for the overlap-induced correlation which is significant when the degrees of freedom are small. Note, however, that the jackknife confidence limits of Fig. 2.2(b) *do not* include the robust estimate, of Fig. 2.2(c), particularly at high frequencies. This is because the two estimates are of fundamentally different physical quantities: the first is simply the *average* power at a given frequency, storms included; the robust estimate gives the power that one would *usually* observe.

A particular feature of multiple-window estimates is that the result (2.20) is the solution of a rational equation. The behavior of this estimator is to downweight badly biased eigenspectra so that, effectively, the resulting spectrum has variable stability. As mentioned earlier, the estimate is χ^2_{2K} when all bias terms are negligible but, in general, is only known to be bounded between χ^2_2 and χ^2_{2K}. The statistic is thus an ideal candidate for jackknifing. That the solution of (2.20) is found by an iterative process rather than by an explicit estimator illustrates another strength of the jackknife; it may be applied to complicated estimation procedures! Formal justification of the jackknife's applicability in such iterative procedures is given in Jorgensen [45].

Figure 2.3 shows a spectrum with jackknife estimates of the 90 percent confidence intervals (based on 5 and 95 percentage points) computed using a multiple-window estimate. The data consists of 360 samples of one component of the Earth's magnetic field recorded on a magnetometer at Frobisher Bay, Northwest Territories, with one sample taken every 10 seconds. Because Baffin Island is both far from most sources of urban noise and also in a region of intense magnetic activity, we expect the spectrum to have a large range: for this hour's data the extremes are 10^{-4} and $10^{+5.5}$ in units of $(\text{nT})^2/\text{Hz}$ (nT is nano-Tesla). In this example, 8 windows were used with a time-bandwidth product, $N \cdot W$ of 4.5. Thus details spaced closer than 4.5/3600. or 1.25 mHz are not resolved and, in particular, the spectrum near zero frequency typically increases as c / f^v with $v \approx 1$–2 instead of flattening as it does here. In this

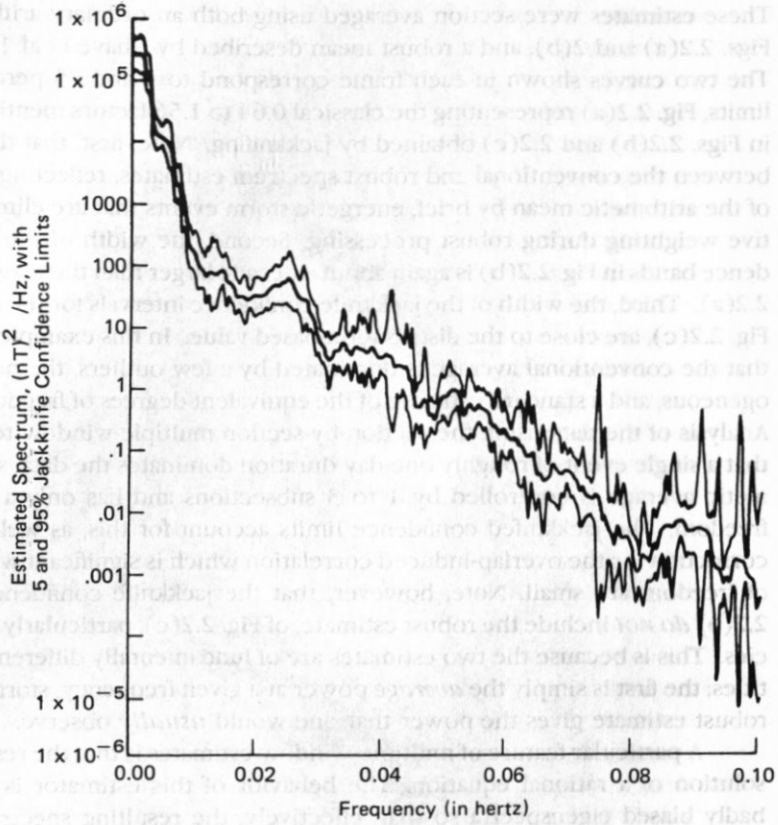


Figure 2.3 Multiple-window estimate, of the spectrum of an hour's data, of the vertical magnetic field in Frobisher Bay. The 90 percent confidence intervals were computed by jackknifing (2.20) and include effects of the decreasing degrees-of-freedom shown in Fig. 2.4. Note that a straight-line fit to the spectrum between about .03 and 0.10 Hz is outside the confidence region only over about 13% of this interval, and so is a reasonable approximation except near the "absorption band" around .08 Hz.; excluding this region only 6% of the region is missed.

example (2.20) was jackknifed directly with dramatically different results from those expected under Gaussian theory. In particular the maximum jackknife variance estimate for ln S of 15.4 exceeds the nominal value $\Psi'(8) = 0.1331$ from (2.44) by a factor of 116. We emphasize that this factor is applied to the *logarithm* of the spectrum! (For plotting purposes the variance estimates have been smoothed over a span of ± 3 points.) This may be surprising, but recall that the spectral representation of a time series only constrains the variance of the orthogonal increment process: it *does not* constrain either its distribution or its fourth moments. Neither is it required that the distribution be frequency independent. In this particular example there is, excepting the usual Fourier transform argument, no good reason to expect Gaussian statistics: on the contrary, the "dips" near 0.04 and 0.08 Hz. are peculiar and, in later

data segments, evolve into peaks, implying that complicated dynamics are involved. In addition to the large jackknife variances, the jackknife bias correction (not shown) also becomes large at higher frequencies having a range of $(-3.2, +0.4)$; given their other interpretations as homogeneity statistics given earlier, this is not surprising.

A second aspect of this example worthy of note is that *both* variance *and* degrees of freedom (the latter obtained from the coefficient weights) vary with frequency. An estimate of the degrees of freedom is shown in Fig. 2.4 from which it may be seen that they vary between 16 and 4.5, so when setting confidence intervals using (2.49) both frequency dependent variance and degrees of freedom must be used. At the high-frequency end of the spectrum, we have both a large jackknife variance and low degrees-of-freedom so that the extreme 5 to 95 percent confidence intervals cover *five orders of magnitude.*

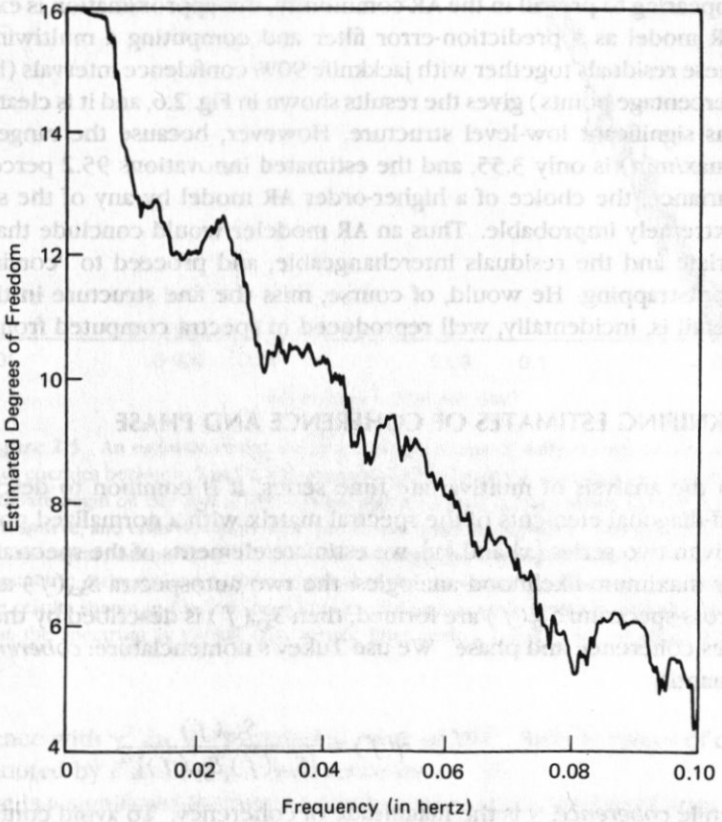


Figure 2.4 The estimated degrees-of-freedom used to compute the confidence intervals for the magnetometer data shown in Fig. 2.3. In this example, prewhitening was deliberately *not* used so that the uncertainty in the estimate used to generate the filter can be seen.

Our last univariate spectrum estimation is related to the problem of autoregressive modeling. As mentioned earlier, Swanepoel and van Wyk [10] have proposed bootstrapping ARMA spectrum estimates by treating the estimated innovations as interchangeable, and the idea has been elaborated on by Bose [46]. Our objection to this approach is simply that present ARMA modeling techniques usually result in oversimplified descriptions for most spectra encountered in practice. As a consequence, the estimated innovations contain significant information on details of the spectrum and so are not interchangeable. As an example of this effect we take data consisting of logarithms of daily counts of 0.5 to 1.5 Mev protons measured by the *Voyager 1* spacecraft starting on day 250 of 1977 and ending 800 days later, just before entering Jupiter's magnetosphere. For details, see Maclennan et al. [47]. Superficially, these data are well approximated by a fourth-order autoregressive model, specifically a product autoregression of the type discussed in McKenzie [48]. Figure 2.5 is a plot of a spectrum estimate and an AR-4 approximation. By standards appearing to prevail in the AR community, this approximation is excellent. Using the AR model as a prediction-error filter and computing a multiwindow spectrum of these residuals together with jackknife 90% confidence intervals (based upon 5 to 95 percentage points) gives the results shown in Fig. 2.6, and it is clear that the spectrum has significant low-level structure. However, because the range of this spectrum (max/min) is only 3.55, and the estimated innovations 95.2 percent of the residual variance, the choice of a higher-order AR model by any of the standard criteria is extremely improbable. Thus an AR modeler would conclude that AR-4 was appropriate and the residuals interchangeable, and proceed to "confirm" the model by bootstrapping. He would, of course, miss the fine structure in the spectrum. This detail is, incidentally, well reproduced in spectra computed from *Voyager 2* data.

.6 JACKKNIFING ESTIMATES OF COHERENCE AND PHASE

In the analysis of multivariate time series, it is common to describe the complex, off-diagonal elements of the spectral matrix with a normalized polar representation. Given two series $\{x\}$ and $\{y\}$, we estimate elements of the spectral matrix motivated by maximum-likelihood analogies: the two autospectra $S_{xx}(f)$ and $S_{yy}(f)$ and the cross-spectrum $S_{xy}(f)$ are formed; then $S_{xy}(f)$ is described by the subordinate entities coherence and phase. We use Tukey's nomenclature: *coherency* is the complex quantity

$$\tilde{\gamma}(f) = \frac{S_{xy}(f)}{[S_{xx}(f)\, S_{yy}(f)]^{1/2}} \tag{2.52}$$

while *coherence*, γ is the magnitude of coherency. To avoid confusion we normally use "complex coherency" and "magnitude-squared coherence" or MSC. For simplicity we usually drop the explicit frequency dependence and write

$$\gamma = |\tilde{\gamma}|$$

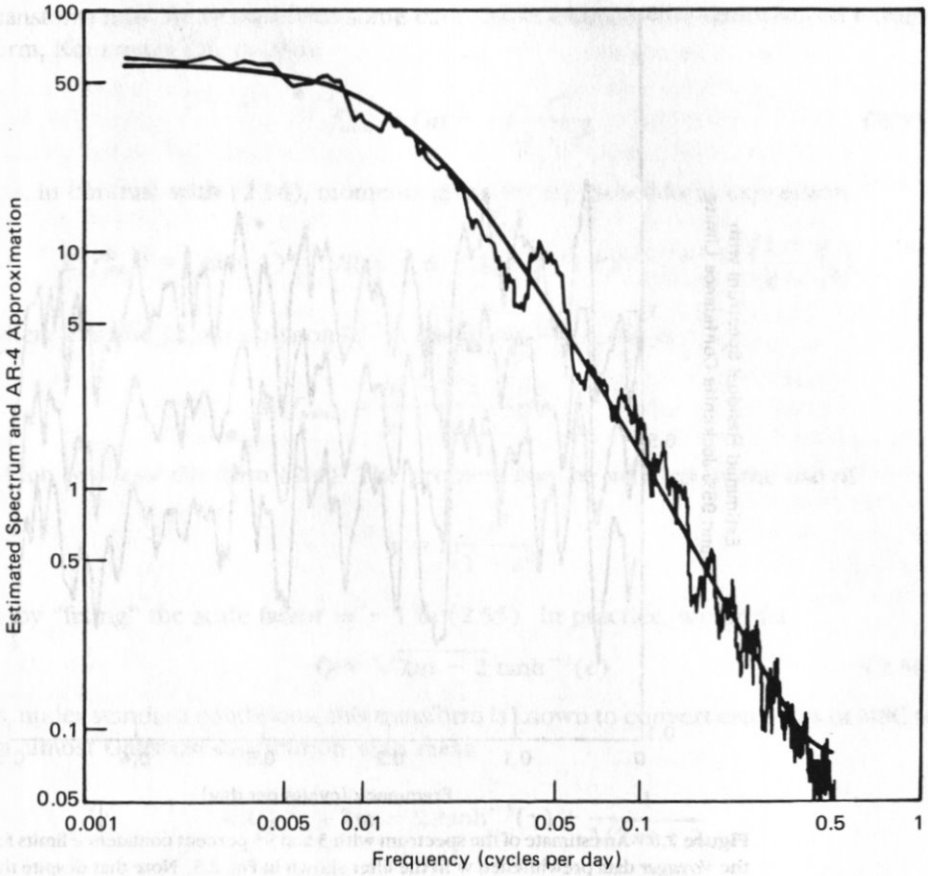


Figure 2.5 An estimate of the spectrum of logarithms of daily counts of protons with energies between .5 and 1.5 *Mev* measured by *Voyager I.* The data used for this estimate begin on day 250 of 1977, when the spacecraft is just outside Earth's magnetosphere, and ends 804 days later, just before entering Jupiter's. The solid line is the spectrum obtained by a fourth-order autoregressive approximation to the nonparametric estimate, and is the reciprocal of the prewhitening filter used to obtain the results shown in Fig. 2.6. Note that the frequency scale is also logarithmic and that the spectrum is typical of a simple filter with a break point of about 100 days.

for coherence with γ^2 for the population value of MSC. Sample values of coherency will be denoted by $\tilde{c}$ and sample coherence by $c = |\tilde{c}|$.

There is a significant literature on coherence estimation (see Carter [49] for a recent review), including one of the few time-series papers making use of the jackknife. In that work, Lee [50] jackknifes MSC estimates directly without making use of variance stabilizing transformations and with an emphasis on bias reduction.

For an estimate of MSC based on m independent samples (or approximately for

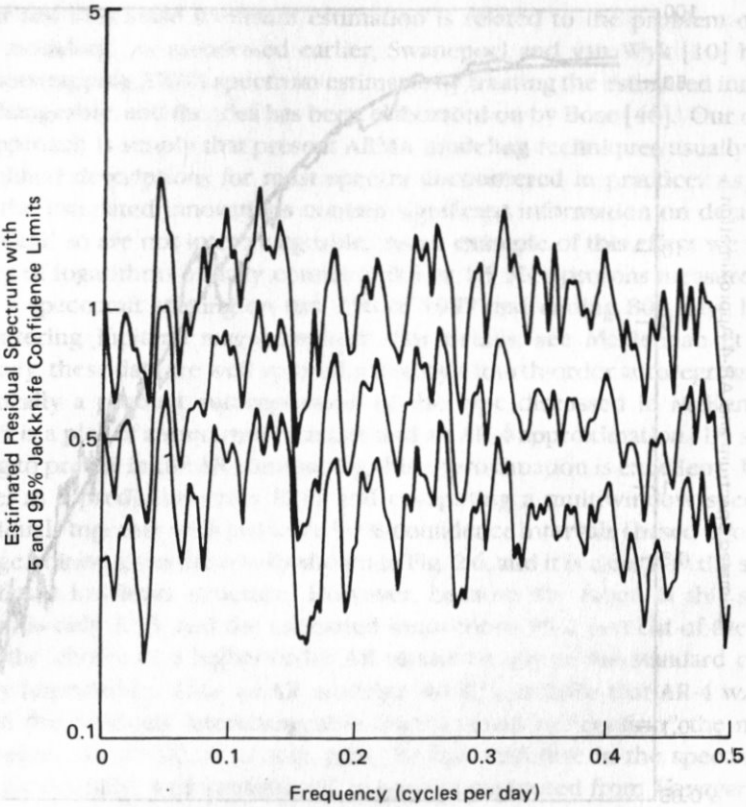


Figure 2.6 An estimate of the spectrum with 5 and 95 percent confidence limits for the *Voyager* data prewhitened with the filter shown in Fig. 2.5. Note that despite the apparently excellent fit in Fig. 2.5, the 5 and 95 percent levels overlap, so the prewhitened filter output cannot be considered white. Note also that the low frequency spectrum does not increase rapidly as would be the case for a power-law process, so that fractal interpretations of the data can also be excluded.

$2m$ equivalent degrees-of-freedom) from a stationary Gaussian process, the probability density function of c^2 is given by Hannan [51, p. 259] as

$$p(c^2) = \frac{(1-\gamma^2)^m}{m-1}(1-c^2)^{m-2}\,{}_2F_1(m, m; 1; \gamma^2 c^2), \tag{2.53}$$

where ${}_2F_1$ is Gauss's hypergeometric function. It is well known that the ordinary estimate of MSC is biased, with μth central moment given by Carter and others [52] as

$$\mathbf{E}\{c^{2\mu}\} = \frac{\Gamma(m)\Gamma(\mu+1)}{\Gamma(\mu+m)}(1-\gamma^2)^m\,{}_3F_2(m, m, \mu+1; 1, \mu+m; \gamma^2). \tag{2.54}$$

Like numerous other statistics, transformed estimates of coherence are more suitable for jackknifing (not to mention plotting) than are the raw estimates. The

transform must be chosen with some care. As an example, the standardized F-transform, Koopmans [53, p. 284]:

$$f_{\text{msc}} = (m-1)\frac{c^2}{1-c^2} \tag{2.55}$$

has, in contrast with (2.54), moments given by the closed-form expression

$$\mathbf{E}\{f_{msc}^b\} = (m+1)^{b-1}B(m-b-1, b+1)\,P_b^{(0,m-b-1)}\left(\frac{1+\gamma^2}{1-\gamma^2}\right),$$

where P is the Jacobi polynomial. In particular, the mean is

$$\mathbf{E}\{f_{msc}\} = \frac{m-1}{m-2}\left[\frac{1+(m-1)\gamma^2}{1-\gamma^2}\right],$$

which is *not* of the form (2.4). The problem may be avoided by the use of

$$y = \ln\frac{c^2}{1-c^2}$$

or by "fixing" the scale factor $m-1$ in (2.55). In practice, we prefer

$$\mathbf{Q} = \sqrt{2m-2}\,\tanh^{-1}(c) \tag{2.56}$$

as, under standard conditions, this transform is known to convert estimates of MSC to an almost Gaussian distribution with mean

$$\mathbf{E}\{\mathbf{Q}\} = \sqrt{2m-2}\,\tanh^{-1}(\gamma) + \frac{1}{\sqrt{2m-2}}$$

and unit variance, Koopmans [53], Amos and Koopmans [54]. If a plot of the transformed coherence and the ± 1 jackknife deviation limits is made, it is immediately obvious if the jackknife tolerance is not close to 1.

To jackknife coherence and phase estimates, we assume that N complex transform pairs, $x_k(f)$ and $y_k(f)$, $k = 1, 2, \ldots, N$ are available. These may be the result of a multiple-window analysis on a single section with N windows, from N distinct data segments, or from a combination of K windows on B segments, so $N = KB$. We then form the basic deleted estimates of coherency from the corresponding delete-one estimates of the spectral matrix

$$\tilde{c}_{\mathcal{J}} = \frac{\sum_{\substack{k=1\\k\neq j}}^{N} x_k(f)\overline{y_k(f)}}{\left[\sum_{\substack{k=1\\k\neq j}}^{N} |x_k(f)|^2 \sum_{\substack{k=1\\k\neq j}}^{N} |y_k(f)|^2\right]^{1/2}} \tag{2.57}$$

for $j = 1, 2, \ldots, N$ plus the standard estimate with nothing omitted. From these, we transform to the almost normal variates

$$Q_{\mathcal{J}} = \sqrt{2m - 2}\ \tanh^{-1}(|\tilde{c}_{\mathcal{J}}|) \tag{2.58}$$

and proceed to find estimates and tolerances as in Section 2.2. Because the mean and variance of Q do not depend on γ, and hence not on frequency, simple averages over frequency of the jackknife bias and variance estimates are sensitive problem indicators.

Figure 2.7 is a coherence plot for the gas furnace data, series J in Box and Jenkins [55]. As this set is short (296 samples) we used a multiple-window estimate with 9 windows and a time-bandwidth product of 5. In this figure the inner axis on the left is the normal transformed scale, plotted linearly, and represents values from (2.56). In these units, coherence estimates have unit standard deviation. The outer

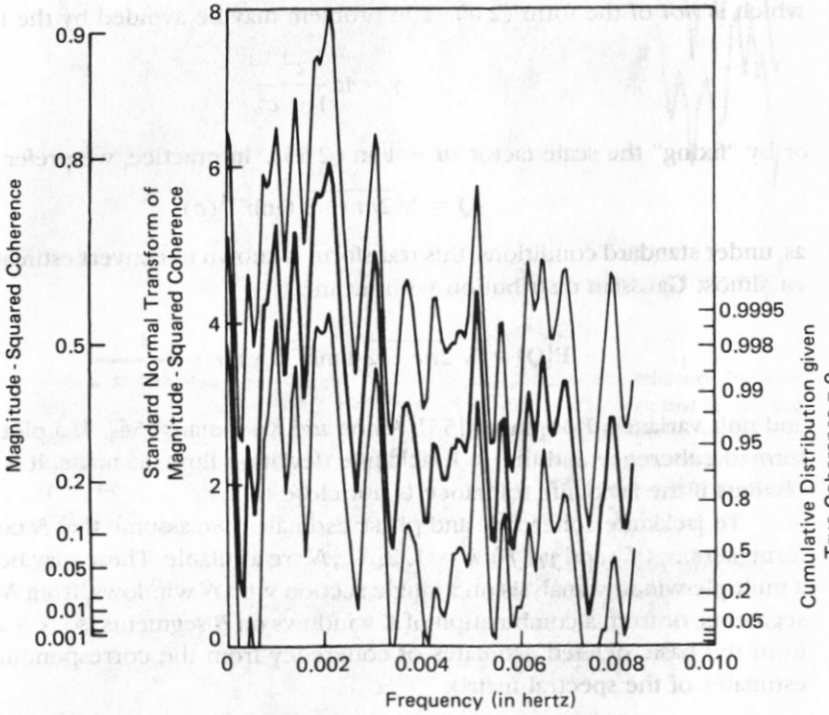


Figure 2.7 Estimated magnitude-squared coherence with ±1 jackknife deviations between the input methane feed and output CO_2 concentration for the Box and Jenkins "gas furnace" data. Jackknifing has been done on the approximately normal variates (equation 2.58), with results shown on the inner left ordinate. The outer left ordinate shows conventional values of *MSC*, while the right ordinate shows the cumulative distribution function assuming the true coherence to be zero. Here the jackknife estimates of standard deviation are about one, the expected level. Note the abrupt change in character at .0028 Hz.

left axis gives the corresponding values of magnitude-squared coherence obtained by inverting (2.58). The vertical axis on the right shows corresponding values of the exact cumulative distribution function of the magnitude-squared coherence assuming zero true coherence. These are computed using transformation (2.55) and correspond to integrals of (2.53) for $\gamma = 0$. The three curves plotted are the jackknifed estimate of MSC from the normal transform (2.58) and the jackknife estimates of ± 1 standard deviation again estimated after using transformation (2.58). Because the data used in this example are reasonably well behaved, they are, as expected, about ± 1 unit from the estimate on the transformed scale. At frequencies below 0.02 Hz the coherence is reasonable but, above that frequency, not significantly nonzero so that estimates of transfer functions and phases would be unreliable in this band. Applying the same method to the prewhitened series using the filter given by Box and Jenkins [55, p. 381] yields the same results.

For a second example, we compute the coherence between the north-south components of the Earth's magnetic field between a magnetometer at the Victoria, B.C., observatory and an ocean-bottom instrument located in Queen Charlotte Sound off the west coast of Canada during a geomagnetically active period of 1982. Here we used a combined method with two windows on each of five sections. The coherence is shown in Fig. 2.8 using the same transformations as before: note that the jackknife standard deviation is about 2, so that the coherence is considerably less significant than it would appear. Also, despite the fact that 1000 sample data segments were used, we must conclude that the Fourier transforms are far from Gaussian. Again, the cause of this is similar to that discussed with Fig. 2.2, where nonstationarity results in dominance of the estimate by a subset of the time series.

2.6.1 Phase Estimates

Coherency is characterized by both magnitude and phase with phase frequently being a more sensitive parameter than MSC. For example, measurement noise attenuates MSC but only increases the variance of a phase estimate, so averaged phases from different runs will converge where averaged MSC's will be biased downward. As a basis for comparison, Goodman [8, eq. 4.58] gives the probability distribution for the phase ϕ of a coherency estimate made using m independent samples from a population with true coherency $\tilde{\gamma}$. Then with

$$\phi = \text{phase}(\tilde{\gamma}) - \mathbf{E}\{\text{phase}(\tilde{\gamma})\},$$

where phase $(z) = \text{atan}(\text{Im}\{z\}/\text{Re}\{z\})$ is the four quadrant complex arctangent (the Fortran function atan2)

$$p(\phi) = \frac{\delta^m}{2\pi\Gamma(m)} \sum_{k=0}^{\infty} (2\gamma)^k \frac{\Gamma[m + (k/2)]\Gamma[1 + (k/2)]}{\Gamma(k+1)} \cos^k\phi, \qquad (2.59)$$

where $\delta = 1 - \gamma^2$. However, because it is circular, phase is a more complicated quantity to estimate and we use the *circular variance* for a unimodal phase distri-

for $j = 1, 2, \ldots, N$ plus the standard estimate with nothing omitted. From these, we transform to the almost normal variates

$$Q_{\mathcal{J}} = \sqrt{2m - 2}\ \tanh^{-1}(|\tilde{c}_{\mathcal{J}}|) \tag{2.58}$$

and proceed to find estimates and tolerances as in Section 2.2. Because the mean and variance of Q do not depend on γ, and hence not on frequency, simple averages over frequency of the jackknife bias and variance estimates are sensitive problem indicators.

Figure 2.7 is a coherence plot for the gas furnace data, series J in Box and Jenkins [55]. As this set is short (296 samples) we used a multiple-window estimate with 9 windows and a time-bandwidth product of 5. In this figure the inner axis on the left is the normal transformed scale, plotted linearly, and represents values from (2.56). In these units, coherence estimates have unit standard deviation. The outer

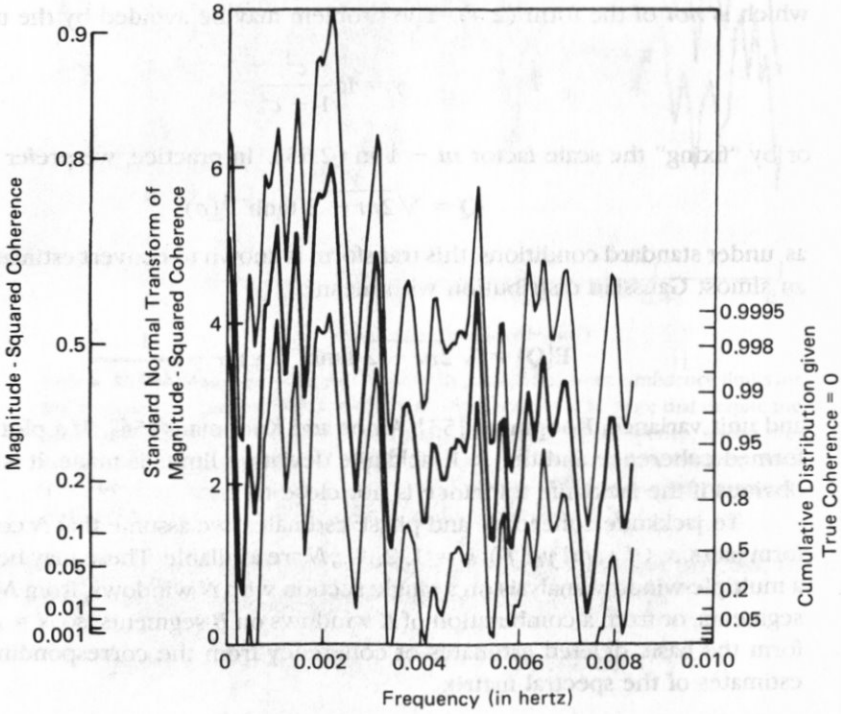


Figure 2.7 Estimated magnitude-squared coherence with ±1 jackknife deviations between the input methane feed and output CO_2 concentration for the Box and Jenkins "gas furnace" data. Jackknifing has been done on the approximately normal variates (equation 2.58), with results shown on the inner left ordinate. The outer left ordinate shows conventional values of *MSC*, while the right ordinate shows the cumulative distribution function assuming the true coherence to be zero. Here the jackknife estimates of standard deviation are about one, the expected level. Note the abrupt change in character at .0028 Hz.

left axis gives the corresponding values of magnitude-squared coherence obtained by inverting (2.58). The vertical axis on the right shows corresponding values of the exact cumulative distribution function of the magnitude-squared coherence assuming zero true coherence. These are computed using transformation (2.55) and correspond to integrals of (2.53) for $\gamma = 0$. The three curves plotted are the jackknifed estimate of MSC from the normal transform (2.58) and the jackknife estimates of ± 1 standard deviation again estimated after using transformation (2.58). Because the data used in this example are reasonably well behaved, they are, as expected, about ± 1 unit from the estimate on the transformed scale. At frequencies below 0.02 Hz the coherence is reasonable but, above that frequency, not significantly nonzero so that estimates of transfer functions and phases would be unreliable in this band. Applying the same method to the prewhitened series using the filter given by Box and Jenkins [55, p. 381] yields the same results.

For a second example, we compute the coherence between the north-south components of the Earth's magnetic field between a magnetometer at the Victoria, B.C., observatory and an ocean-bottom instrument located in Queen Charlotte Sound off the west coast of Canada during a geomagnetically active period of 1982. Here we used a combined method with two windows on each of five sections. The coherence is shown in Fig. 2.8 using the same transformations as before: note that the jackknife standard deviation is about 2, so that the coherence is considerably less significant than it would appear. Also, despite the fact that 1000 sample data segments were used, we must conclude that the Fourier transforms are far from Gaussian. Again, the cause of this is similar to that discussed with Fig. 2.2, where nonstationarity results in dominance of the estimate by a subset of the time series.

2.6.1 Phase Estimates

Coherency is characterized by both magnitude and phase with phase frequently being a more sensitive parameter than MSC. For example, measurement noise attenuates MSC but only increases the variance of a phase estimate, so averaged phases from different runs will converge where averaged MSC's will be biased downward. As a basis for comparison, Goodman [8, eq. 4.58] gives the probability distribution for the phase ϕ of a coherency estimate made using m independent samples from a population with true coherency $\tilde{\gamma}$. Then with

$$\phi = \text{phase}(\tilde{\gamma}) - \mathbf{E}\{\text{phase}(\tilde{\gamma})\},$$

where phase $(z) = \text{atan}(\text{Im}\{z\}/\text{Re}\{z\})$ is the four quadrant complex arctangent (the Fortran function atan2)

$$p(\phi) = \frac{\delta^m}{2\pi\Gamma(m)} \sum_{k=0}^{\infty} (2\gamma)^k \frac{\Gamma[m + (k/2)]\Gamma[1 + (k/2)]}{\Gamma(k+1)} \cos^k\phi, \qquad (2.59)$$

where $\delta = 1 - \gamma^2$. However, because it is circular, phase is a more complicated quantity to estimate and we use the *circular variance* for a unimodal phase distri-

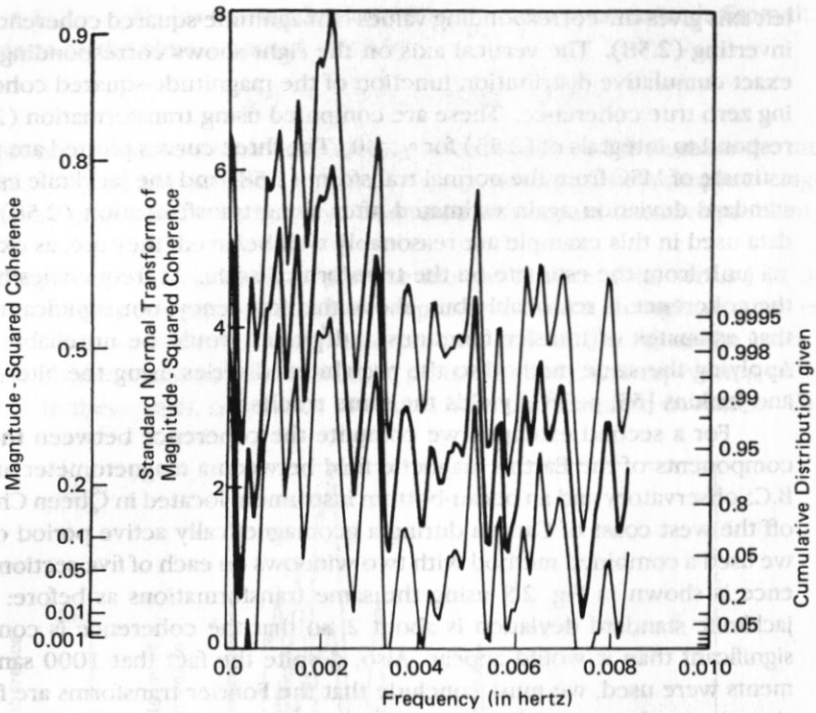


Figure 2.8 Estimates of magnitude-squared coherence and ±1 jackknife deviations between the north-south, *H*, magnetic field recorded at Victoria Observatory, and by a seafloor magnetometer in Queen Charlotte Sound on day 188 of 1982. Note that here the jackknife standard deviation estimates are about twice as large as expected. The variance being four times larger than expected implies that one cannot trust simple inferences from this data.

bution $p(\phi)$ having expected value 0, Mardia [56]:

$$\sigma_0^2 = \int_{-\pi}^{\pi} 2(1 - \cos\phi)p(\phi)d\phi. \tag{2.60}$$

Because $2(1 - \cos\phi) \approx \phi^2$ for small ϕ, the circular variance is intuitively satisfying for small errors and is more reasonable for large ones. The integration to obtain the first cosine moment

$$C_1 = \int_{-\pi}^{\pi} p(\phi) \cos\phi \, d\phi$$

is done by Wallis's formula. The result is

$$C_1 = \frac{\pi^{1/2}}{2} \frac{\Gamma(m + 1/2)}{\Gamma(m)} \delta^m \, {}_2F_1(m + 1/2, 3/2; 2; \gamma^2).$$

Here, the circular variance is given by

$$\sigma_0^2 = 2(1 - C_1).$$

As an example $\gamma^2 = 0.64$ and $m = 5$ give $\sigma_0^2 = 0.076962$ or, changing to degrees, $\sigma_0 = 15.895°$.

The corresponding phase estimates are obtained by minimizing

$$r = \sum_{j=1}^{N} 2[1 - \cos(\phi_j - \hat{\phi})]$$

with respect to $\hat{\phi}$. Manipulating the resulting condition

$$0 = \sum_{j=1}^{N} \sin(\phi_j - \hat{\phi})$$

by the usual trigonometric identities results in several equivalent formulas for $\hat{\phi}$, r, and σ_0.

The specific approach used here is to compute the phase factors

$$e_{\curvearrowright} = \frac{\tilde{c}_{\curvearrowright}}{|\tilde{c}_{\curvearrowright}|} \tag{2.61}$$

from (2.57) and their average

$$e_{\frown} = \frac{1}{m} \sum_{j=1}^{m} e_{\curvearrowright} \tag{2.62}$$

Define $\phi_{\frown} = \text{phase}\{e_{\frown}\}$, and the jackknife variance estimate comes from the mean resultant

$$\hat{V}\{\phi_{\frown}\} = 2(m - 1)(1 - |e_{\frown}|) \tag{2.63}$$

For small phase errors, this is equivalent to

$$\frac{m-1}{m} \sum_{j=1}^{m} \text{phase}^2\{e_{\curvearrowright} \overline{e_{\frown}}\},$$

but is more meaningful for large errors. (Note that this form is just a simple way to compute $\Sigma\,(\phi_{\curvearrowright} - \phi_{\frown})^2$ while allowing the difference between phases of 1° and 359° to be 2° apart.)

Figure 2.9 shows the phase estimate for the gas-furnace data whose coherence was shown in Fig. 2.7. The general slope corresponds to a delay of about 5.3 samples between input and output, in general agreement with the 5 sample delay given for the peak of the cross-correlation of the prewhitened series by Box and Jenkins [55]. Below 0.02 Hz, where the coherence is high, the phase is well determined but, particularly in the 0.02 to 0.05 Hz band the phase is essentially unknown. Curiously, the phase near the band edge has returned to nearly the same slope as the low-

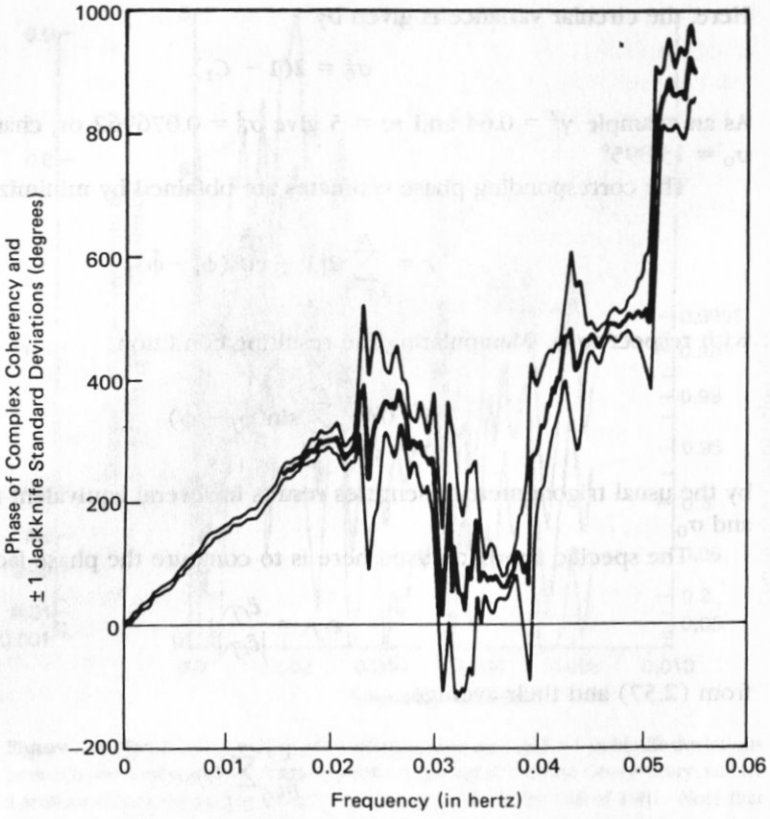


Figure 2.9 Phase of the complex coherency for the "Gas Furnace" data corresponding to the coherence estimate shown in Fig. 2.7. The approximately linear phase dependence on frequency below .02 Hz corresponds to a simple delay and is well determined. At many higher frequencies, however, the phase is essentially indeterminate as the jackknife ±1 deviation limits exceed 360 degrees. Between .037 and .054 *Hz*, the estimate should likely be 360° higher.

frequency portion so that, without the jackknife tolerances, we might be tempted to believe the apparent structure at intermediate frequencies.

Figure 2.10(a) shows the phase estimated between global temperature (as estimated by changes in ^{18}O from the deep sea sediment core V22-174, see Imbrie et al. [57] and the seasonal difference in solar insolation due to changes in the eccentricity of the Earth's orbit over the last 730,000 years. The data consists of 156 samples spaced about 4977 years apart. Eccentricity of the orbit was computed using the trigonometric series expansions in Berger [58]. Because of the limited data and presence of much structure in the spectrum, a multiple-window approach with six windows and the relatively low time-bandwidth product of 4 were used. As before, the figure shows the estimated phase and the jackknife ±1 standard deviation curves. These, however, are somewhat tighter than might be expected from the coherences, Fig. 2.10(b). The difference is that here the eccentricity is deterministic so that the

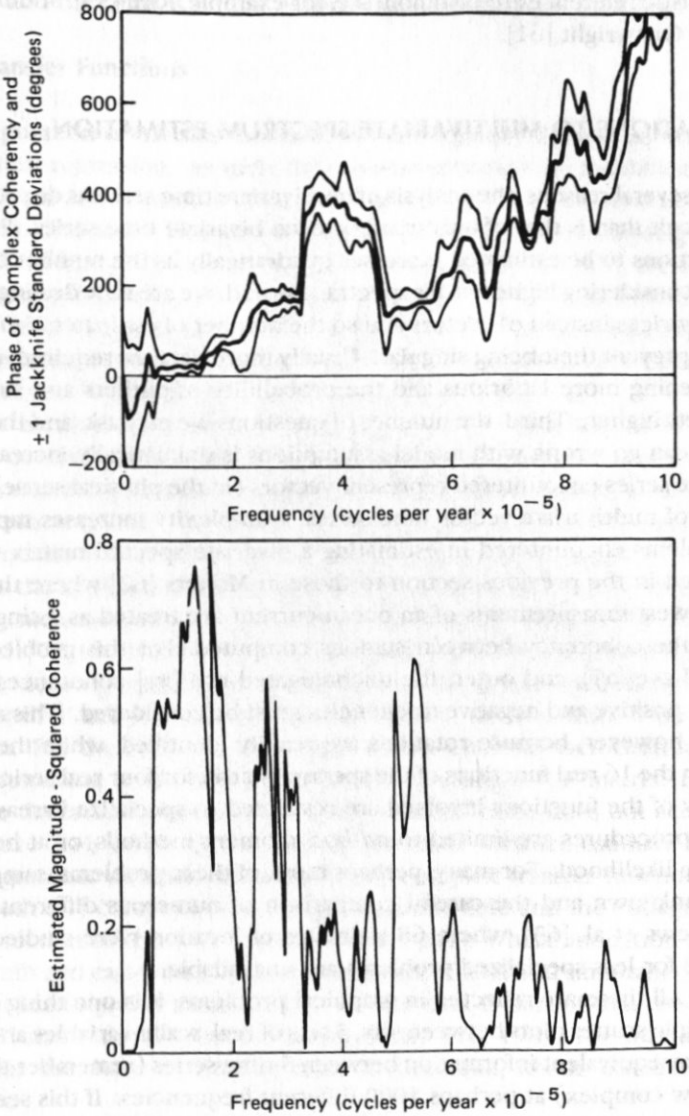


Figure 2.10 The upper panel shows the estimated phase and ± jackknife standard deviations between the δ^{18}O proxy global temperature record, from core V22-174, and the eccentricity of the Earth's orbit for the last 730,000 years. At very low frequencies, periods of 400,000 years and more, the record is probably nonstationary, details in the spectrum are unresolved, and there is confounding with the unknown mean value, so the coherences (shown in the lower panel) are low and the phase errors are large. The increasing phase errors at high frequencies are a result of poor signal-to-noise ratio, again reflected in poor coherences.

formulas given above do not apply. Formulas for phase distributions with a deterministic reference are common: see, for example, Rice [59], Middleton [60], or Munk and Cartwright [61].

2.7 APPLICATIONS TO MULTIVARIATE SPECTRUM ESTIMATION

For several reasons, the analysis of multivariate time series is disproportionately more difficult than is that of univariate or even bivariate time series. First, the number of functions to be estimated increases quadratically as the number of series even without considering higher-order spectra. Second, we are now dealing with p-dimensional matrices instead of vectors and so the number of estimates N must be greater than p to prevent their being singular. Usually more data are required, in turn making data screening more laborious and the probability of outliers and subtle nonstationary effects higher. Third, the number of questions we may ask, and the number of things that can go wrong with model assumptions is dramatically increased. Fourth, many of the series encountered represent vectors (in the physical sense) so the problem is one of multivariate-vector time series. Complexity increases rapidly: compare the problems encountered in estimating a bivariate spectral matrix with real data discussed in the previous section to those in Mooers [62] where the north-south and east-west measurements of an ocean current are treated as a single complex series and the coherency between stations computed. For this problem both inner, the usual ave $\{x\bar{y}\}$, and outer, the unconjugated ave $\{xy\}$ coherences are required and both positive and negative frequencies must be considered. This approach is important, however, because rotations are readily identified, while they are not as clear given the 16 real functions of the spectral matrix for four real series. Finally, because many of the functions involved are restricted to specialized areas, available estimation procedures are limited to *ad-hoc,* moment methods, or at best, Gaussian maximum likelihood. For many, perhaps most, of these problems, sampling distributions are unknown and the careful comparison of numerous different estimators (as in Andrews et al. [63], where 68 estimates of location were studied) that have been made for less specialized problems are unavailable.

All these are reflected in graphical problems; it is one thing to show a page of multiple scatter plots between, say, 5 sets of real, scalar variables and another to try to display equivalent information between 5 time series (remember the equivalent data is now complex) at perhaps 1000 different frequencies. If this seems extreme, consider the problems in the analysis of data from the EMSLAB experiment, see EMSLAB Group [64], where 29 such five-component vector time series were recorded simultaneously with an additional 118 three-component series. As a result, each problem tends to have its own programs tailored specifically to the physics of the problem and to the estimation of parameters of physical interest. Because the distributions of many of these parameters are unknown, even with Gaussian data, they are ideal candidates for jackknifing. The remainder of this section gives three examples of such problems: the first is a brief discussion of transfer function estimation by least

squares; the second covers transfer function estimation by a singular value decomposition; and the third is on the estimation of polarization parameters.

2.7.1 Transfer Functions

The computation of transfer functions is the frequency-domain equivalent of multivariate linear regression. As such, the problem inherits both the difficulties of regression (with the added complication of being complex) and the techniques devised to treat them. The latter includes numerous jackknife and resampling methods: see Miller [3], Hinkley [18], and the recent survey paper by Wu [65]. Details of complex least-squares problems are described by K. S. Miller [66]. Also, before attempting to compute transfer functions, we should be familiar with the usual regression problems, in particular those caused by collinearity and leverage: see Belsley and others [67], Cook and Weisberg [68], and Mosteller and Tukey [4].

The method we prefer for estimating transfer functions is to use the basic complex coefficients from the various series directly as elements of $\mathbf{Y}$ and $\mathbf{A}$ in (2.7), then solve it using a stable method such as QR decomposition, or singular value decomposition (SVD).* In addition to avoiding the numerical ill-conditioning of the normal equations encountered when the spectral matrix is explicitly computed, this facilitates the application of the jackknife by allowing the use of downdating procedures. In such methods, see, for example, Lawson and Hanson [71, Ch. 24], or Dongarra et al. [72, Ch. 10], we typically compute a QR decomposition for the full problem. This decomposition is saved and each of the delete-one estimates, where each row of both Y and A is deleted simultaneously, derived from it by a downdating operation. For the SVD, one may begin with a Cholesky factorization of the $N \times P$ matrix, downdate it, and use a SVD on the resulting $P \times P$ matrix. Because only a one-stage downdating operation is used, roundoff error does not accumulate.

Figure 2.11 shows the ratio of the jackknife variance estimate to the conventional regression variance estimate for the complex transfer function between the north-south (H) component of the geomagnetic field and the voltage induced on a 225-km section of the TAT-6 trans-Atlantic cable where induction from both the north-south and east-west (D) components are considered simultaneously and a conventional least-squares method is used. This data is part of that used in Lanzerotti et al. [73] and is for the first 2.5 hours of day 116, 1985. The magnetometer is located near the shore end of the cable in Greenhill, Rhode Island. Specifically, let $\mathbf{M}_2(f)$ be

$$\mathbf{M}_2(f) = \begin{bmatrix} H_1(f) & D_1(f) \\ H_2(f) & D_2(f) \\ \vdots & \vdots \\ H_N(f) & D_N(f) \end{bmatrix}, \tag{2.64}$$

*For a detailed discussion of singular value decomposition, QR decomposition, and Cholesky factorization, see Golub and Van Loan [69] and Haykin [70].

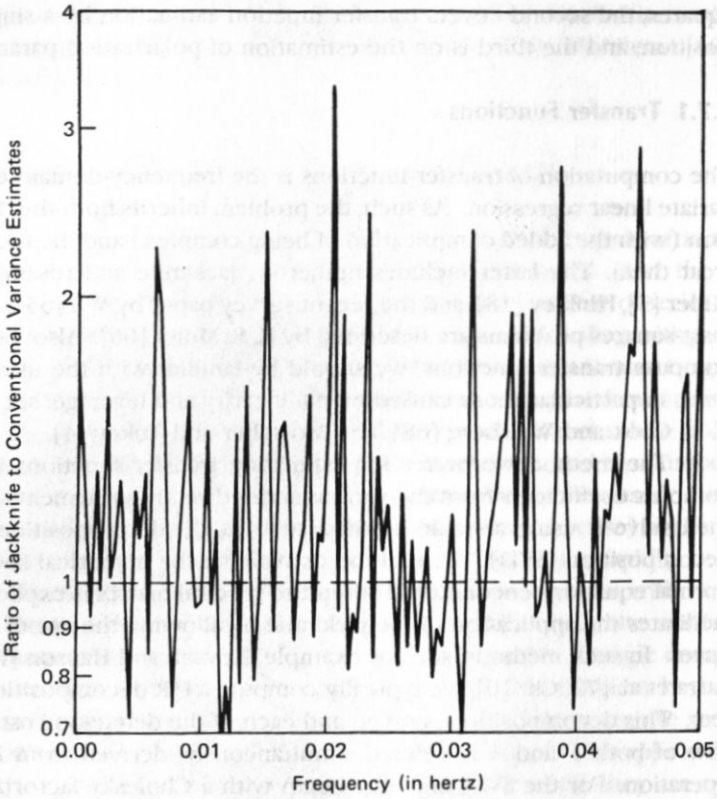


Figure 2.11 The ratio of variance estimated by the jackknife to the conventional (normal-theory) variance for the part of the complex transfer function between the north-south, H, component of the geomagnetic field, and voltage induced on the part of the TAT-6 cable on the continental shelf. Note that the average, across frequency, of this ratio is greater than one, and also that the ratio shows signs of systematic behavior.

and taking the corresponding voltage coefficients to be

$$\mathbf{V}(f) = [V_1(f), V_2(f), \ldots, V_N(f)]^T \tag{2.65}$$

we solve

$$\mathbf{V}(f) = \mathbf{M}_2(f)\,\mathbf{T}_2(f) + \mathbf{R}_2(f), \tag{2.66}$$

where $\mathbf{T}_2(f)$ is the transfer function vector and $\mathbf{R}_2$ the residuals. The conventional variance estimate for this problem is

$$\text{var}\{\hat{\mathbf{T}}_2\} = (\mathbf{M}_2^{\dagger}\,\mathbf{M}_2)^{-1}\frac{\|\mathbf{R}_2\|^2}{N-2} \tag{2.67}$$

and was computed using all N coefficients. From this figure it may be seen that the jackknife variance estimates usually exceed the conventional ones for this problem

by 20 to 30 percent but factors of 2 are common. Considering both components and several hours shows that factors of 4 and 5 are not unusual. A possible explanation is simply the nonstationarity and nonnormality of the geomagnetic field: this is unsatisfactory as this data period appears reasonable. A more likely explanation is that systematic oscillations in the spectrum and transfer functions suggest the presence of long time delays: this would imply that the spectrum is not resolved and the variance is being increased by the presence of correlations of the kind given by (2.23). See the appendix of the paper by Lanzerotti et al. [73].

A more interesting use of these methods is to determine if the model is two or three dimensional. For this problem it is known that usually the incident geomagnetic field has two independent components, H and D, and that the vertical component, Z, is linearly dependent on H and D. It is, however, possible for there to be an independent Z component and its presence and effect on the cable voltage is of interest. Thus we may consider the three-dimensional problem

$$\mathbf{M}_3(f) = \begin{bmatrix} H_1(f) & D_1(f) & Z_1(f) \\ H_2(f) & D_2(f) & Z_2(f) \\ \vdots & \vdots & \vdots \\ H_N(f) & D_N(f) & Z_N(f) \end{bmatrix} \tag{2.68}$$

and solve

$$\mathbf{V}(f) = \mathbf{M}_3(f)\,\mathbf{T}_3(f) + \mathbf{R}_3(f). \tag{2.69}$$

Clearly the additional energy explained by the assumption of an independent Z component is $\|\mathbf{R}_2\|^2 - \|\mathbf{R}_3\|^2$ with two degrees of freedom for the complex coefficients, so the partial F test (see, e.g., Draper and Smith [74]) is

$$F_p(f) = \frac{1/2\,\|\mathbf{R}_2\|^2 - \|\mathbf{R}_3\|^2}{[1/2(N-3)]\,\|\mathbf{R}_3\|^2}. \tag{2.70}$$

Such a test is shown in Fig. 2.12, and it may be seen that the levels are indecisive. (Here $N = 15$, so the 99 percent point is 5.61, larger than any of the values plotted. The original data segments are 900 samples long and a multiple-window estimate with a time-bandwidth of 10 was used, so the levels obtained could happen easily by chance.) It is well known, see, for example, Box [44], that an F statistic is sensitive to nonnormality. Figure 2.13 is a plot of the jackknife variance estimates of $\ln F_p$, and clearly they are both extremely unstable and often much larger than the expected value of 1.732. For this problem, the measurements available are insufficient for an unambiguous decision: in addition to the problems discussed, the magnetic field almost certainly contains gradient terms and, worse, the magnetic measurements are not noise free. Before concluding this example, we note that if additional measurements were available, the problem of deciding which regression terms to include would be even more difficult. In standard regression problems, the method of choice for making such decisions appears to be Mallows's [75] C_p statistic, see Thompson [76]. Like many other statistics considered here C_p depends on ratios of quadratic forms and could be similarly jackknifed or bootstrapped with the points on the usual

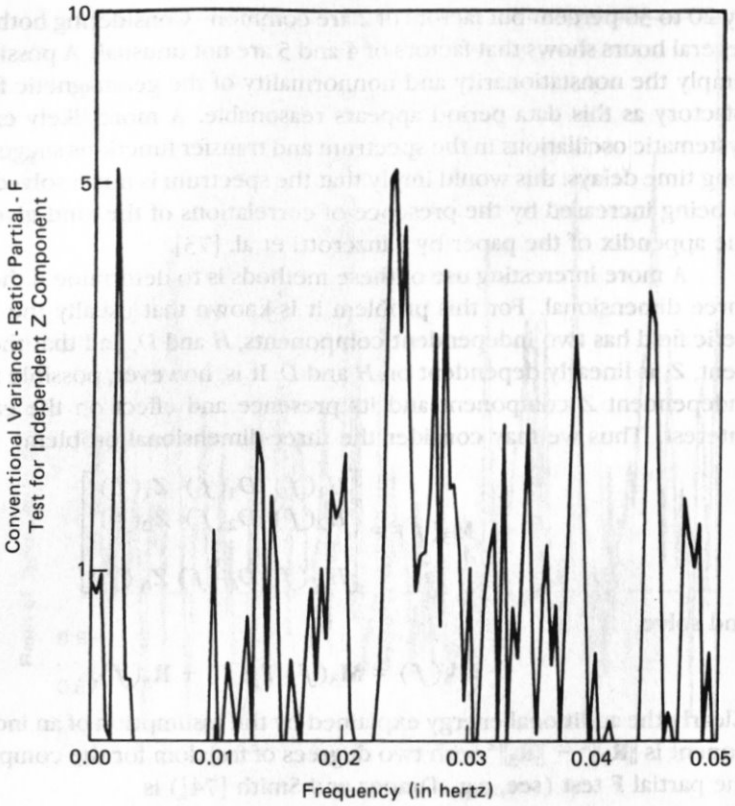


Figure 2.12 A partial F test for the presence of induced voltage of the TAT-6 cable, due to an independent vertical, Z, component of the geomagnetic field. The statistic is based on normal theory and is indecisive (see Fig. 2.13).

C_p plot replaced by error bars or boxplots. We have not solved the problem of including frequency dependence in such a plot!

2.7.2 Errors in Variables Regression

In many multivariate problems it is unreasonable to make the usual least-squares assumptions and assume that all the noise is in the "response" channel and that the "explanatory" series is noise free. A typical example occurs in magneto-variation studies where three orthogonal components of the Earth's magnetic field are recorded simultaneously with the two horizontal components, H and D, regarded as "input" and the vertical component, Z, as "output." The usual least-squares assumption that all the instrumental noise occurs additively on the output and the inputs are noise free is clearly ridiculous: the most reasonable assumption is that the instrument noise power is identical on all three channels. (But note that the *relative* noise can be significantly different.) For real data this problem has been treated by numerous

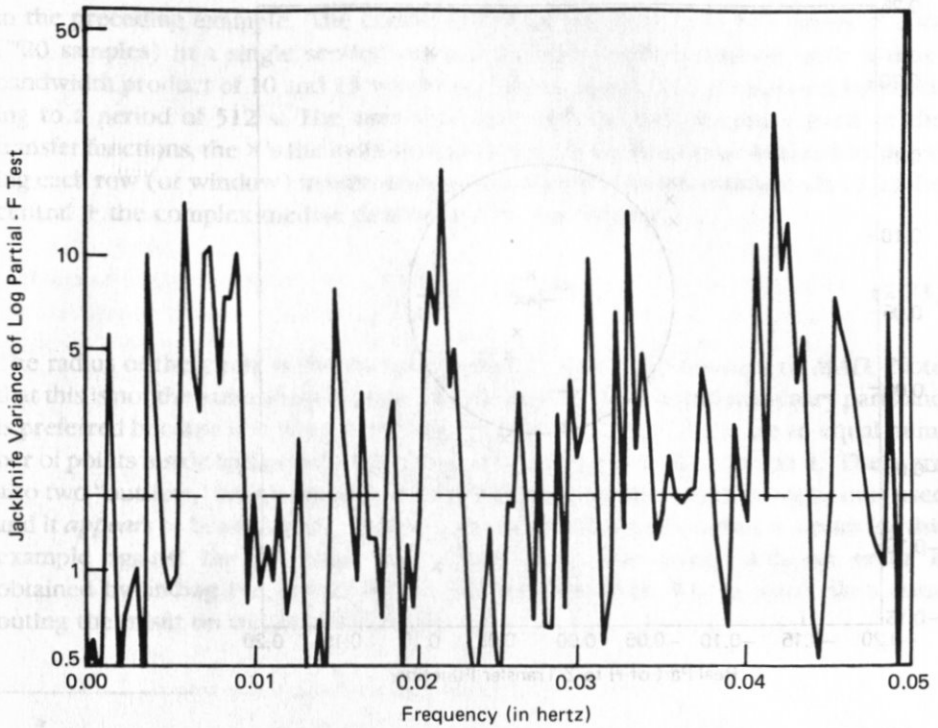


Fig. 2.13 A jackknife estimate of the variance of the logarithms of the partial-F statistic, shown in Fig. 2.12. The expected variance here is 1.73, shown by the dashed line. Note that the jackknife variance of the log F tests is plotted on a logarithmic scale (an additional logarithm) so the average is much higher than normal-theory would indicate.

authors, see Anderson [77] for a recent review and Fuller [78], but of particular relevance to the problems considered here are Park and Chave [79] and Golub and Van Loan [80]. The idea of applying the jackknife to such problems appears in Brillinger [81].

For this problem we consider the $N \times 3$ matrix $\mathbf{M}_3(f)$ noted earlier and compute its singular value decomposition:

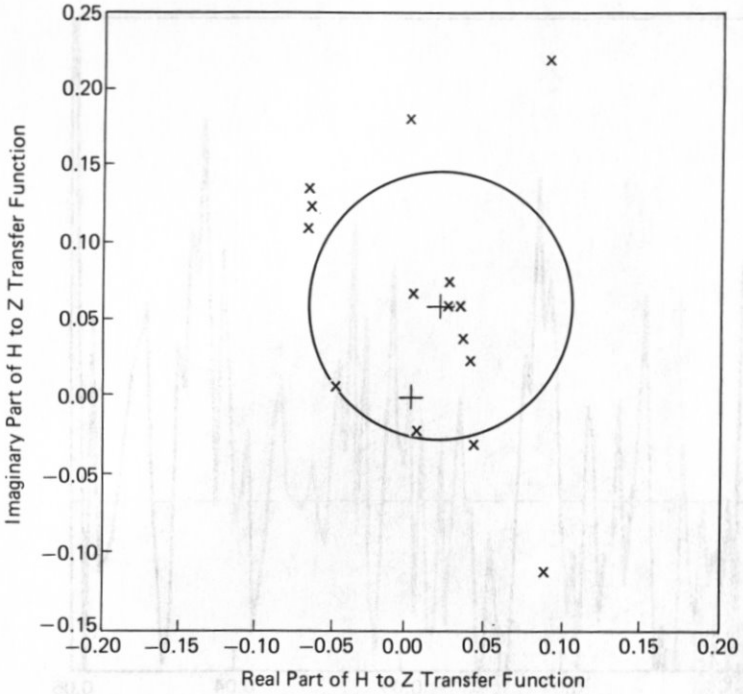
$$\mathbf{M}_3 = \mathbf{U}\Sigma\mathbf{V}^{\dagger}. \tag{2.71}$$

The transfer function is given by the right singular vector, V_3 corresponding to the smallest singular value σ_3. Explicitly, we assume σ_3 is small, so that

$$V_{31}\, H + V_{32}\, D + V_{33}\, Z \approx 0, \tag{2.72}$$

where $\mathbf{U}$ is the matrix of left singular vectors, $\mathbf{V}$ is the matrix of right singular vectors, and Σ is the diagonal matrix of singular values. Thus, for example, the H to Z transfer function is given by $-V_{31}/V_{33}$.

Figure 2.14 shows the H to Z and D to Z pseudo-values of single-frequency components of this transfer function for the Greenhill, R.I. magnetometer mentioned



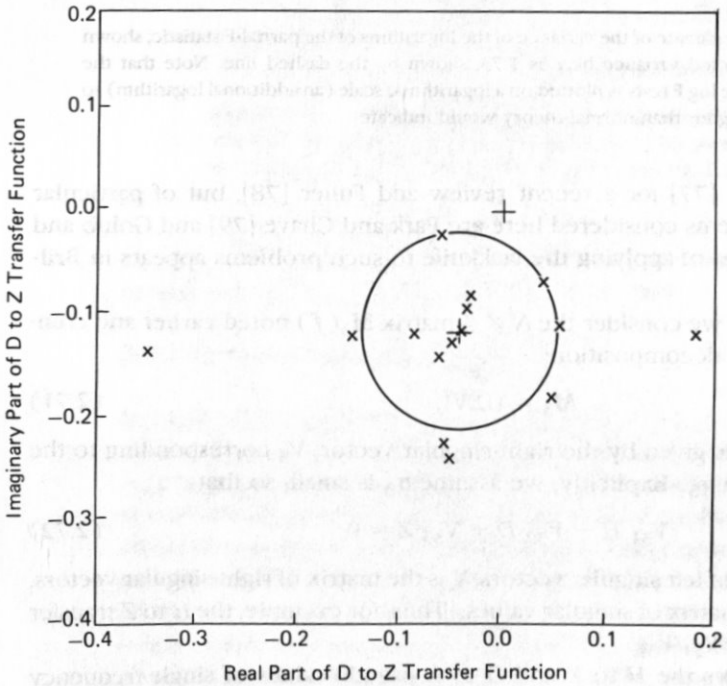


Figure 2.14 Upper left panel shows the pseudo-value estimates of the complex H to Z transfer functions for the Greenhill, Rhode Island, magnetometer at 1.95 mHz. The corresponding D to Z transfer function is shown in the lower panel. A multiple-window method with a time-bandwidth of 10 and 15 windows was used on a single two-hour data segment with the transfer functions computed by the SVD method. In these figures, both the pseudo-values and standard estimates are shown by ×'s, the central + is the complex median, and the circle the radial MAD.

in the preceding example. The coefficients were obtained from two hours of data (720 samples) in a single section using a multiple-window method with a time-bandwidth product of 10 and 15 windows. The frequency is 1.95 mHz corresponding to a period of 512 s. The axes represent the real and imaginary parts of the transfer functions, the ×'s the individual pseudo-transfer functions obtained by deleting each row (or window) in turn and combining with the full estimate via (2.2) the central + the complex median determined by minimizing

$$\sum_{j=1}^{N} \left| h_{(j)} - h_{med} \right|. \tag{2.73}$$

The radius of the circle is the complex *median absolute deviation* or *MAD*. Note that this is not the same as taking separate medians of the real and imaginary parts and is preferred because it is phase-invariant. However, as usual there are an equal number of points inside and outside the circle and, because N is odd, one on it. There are also two "outliers," which are somewhat mysterious as only one data segment is used and it *appears* to be stationary. Figure 2.15 shows the cross-validation errors for this example against the minimum eigenvalue, $\sigma_{3(j)}$. The cross-validation error is obtained by finding the delete-one transfer function from $\mathbf{V}_{(j)}$ as usual, then computing the misfit on the deleted jth row:

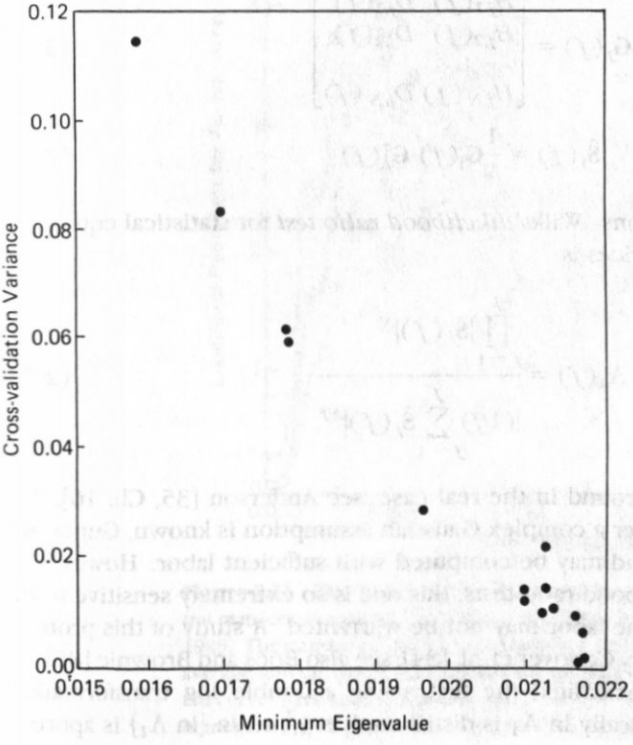


Figure 2.15 The cross-validation variance for predicting Z, from the SVD delete-one transfer function estimates shown in Fig. 2.14, against the corresponding minimum eigenvalue.

$$\rho_j^2 = |V_{31(j)}H_j + V_{32(j)}D_j + V_{33(j)}Z_j|^2.$$

Observe that omitting an outlying point from the SVD calculation both decreases the minimum eigenvalue and increases the cross-validation error. The nearly straight-line behavior seems to be coincidental.

2.7.3 Homogeneity and Polarization

A third application in this area is the computation of *complex multivariate homogeneity statistics* and *polarization parameters* used in electromagnetic and seismic problems to describe transverse waves. For this example we use recordings of the two horizontal components, H and D, of the geomagnetic field recorded simultaneously at three different, but closely spaced, stations. A major goal in making such recordings is to infer properties of the earth, in particular the electrical conductivity, underlying the stations. Making such inferences correctly, however, depends on numerous assumptions, one of which is that the incident field is a homogeneous plane wave. One test of this assumption is that the spectral matrices at the different stations should be "similar." Denoting the estimated spectral matrix at station j by $\hat{\mathbf{S}}_j(f)$, recall that it is simply the covariance matrix of the $N = 2$ matrix of coefficients

$$\mathbf{G}_j(f) = \begin{bmatrix} H_{j,1}(f) & D_{j,1}(f) \\ H_{j,2}(f) & D_{j,2}(f) \\ \vdots & \vdots \\ H_{j,N}(f) & D_{j,N}(f) \end{bmatrix} \tag{2.74}$$

$$\hat{\mathbf{S}}_j(f) = \frac{1}{N}\mathbf{G}_j(f)\,\mathbf{G}_j^{\dagger}(f) \tag{2.75}$$

computed at each of J stations. Wilks' *likelihood ratio test* for statistical equality of the estimated spectral matrices is

$$\Lambda_1(f) = \frac{\prod_{j=1}^{J} |\hat{\mathbf{S}}_j(f)|^N}{|(1/J)\sum_{j=1}^{J} \hat{\mathbf{S}}_j(f)|^{NJ}}. \tag{2.76}$$

For a derivation and background in the real case, see Anderson [35, Ch. 10]. The distribution of this test under a complex Gaussian assumption is known, Gupta [82] or Krishnaiah et. al. [83], and may be computed with sufficient labor. However, in common with similar likelihood ratio tests, this one is so extremely sensitive to the Gaussian assumption that the labor may not be warranted. A study of this problem for univariate data is given by Conover et. al. [84]; see also Boos and Brownie [85]. As a compromise, we have jackknifed the test, using a double log transformation. (Double because asymptotically $\ln \Lambda_1$ is distributed as χ^2_ν, so $\ln\{\ln \Lambda_1\}$ is approxi-

mately symmetric.) We then take the pragmatic approach that when the test is more than two or three jackknife deviations from nominal, the data should be checked more carefully.

As an example, we use a small fraction of the EMSLAB data mentioned earlier, from stations 1, 2, and 4. These stations are located on a line running nearly east-west near Lincoln City, Oregon, and are 10.6, 19.3, and 40.0 km, respectively, from the coast. Figure 2.16 shows a homogeneity calculation for an 18,000 second data block beginning at hour 5 of July 25, 1985 with coefficients calculated using a time-bandwidth product of 10 and 15 windows. Clearly, the test fails over most of the frequency band.

In this application, an informative diagnostic is given by the polarization parameters of the field at the different stations. A general description of polarization is

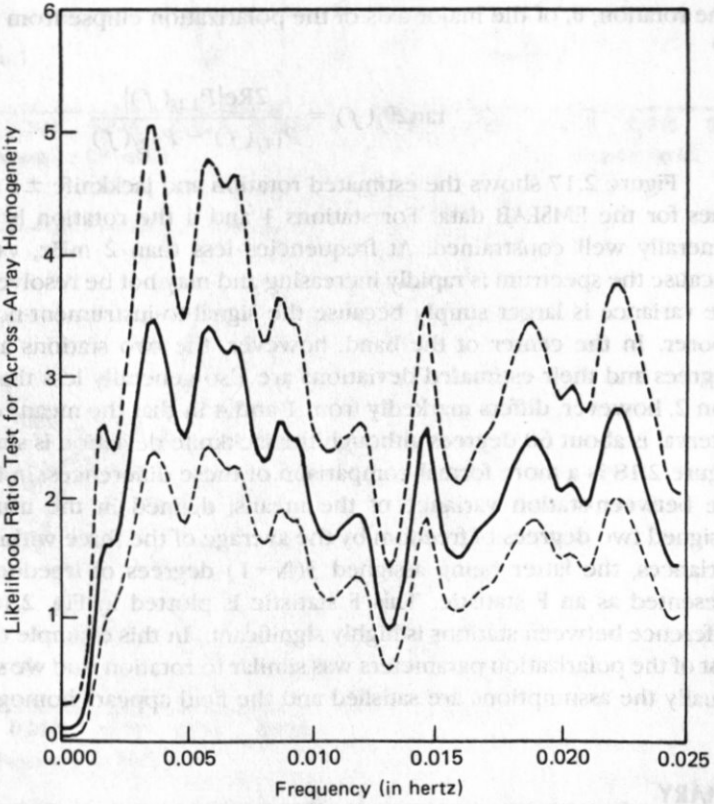


Figure 2.16 The Wilks' likelihood ratio statistic for homogeneity of the *H-D* spectral matrices at stations 1, 2, and 4 of the EMSLAB array for hours 5 to 10 of July 25, 1985. The statistic $1n\{1n\,\Lambda_1\}$ was jackknifed, with the heavy center line showing the average, and the lighter dashed lines the ± 1 jackknife standard deviations. The plotted curves have been nonlinearly smoothed, and it is obvious that in this time period homogeneity is a poor assumption for most of the frequency band.

given in Born and Wolf [86]. Means [87] describes the analysis of three component transverse wave data, and Park et. al [88] discuss three component seismic data. In this example, we are assuming that the direction of propagation is vertical downward and consequently use only the two horizontal components so that the simpler analysis of Fowler et. al. [89] can be used, and we apply the jackknife both to estimate variances of the parameters and to compare the results at different stations. Continuing the example, at each station we compute the polarization parameters and jackknife variances of them; then, using the polarization estimates for the three stations, we compute a grand mean and its variance; finally, we compare the variance between stations with the average jackknife variance at a single station. Specifically, take $D_j(f)$ to be the smaller eigenvalue of $\hat{S}_j(f)$ and let

$$\mathbf{P}_j(f) = \hat{\mathbf{S}}_j(f) - D_j(f)\,\mathbf{I}. \tag{2.77}$$

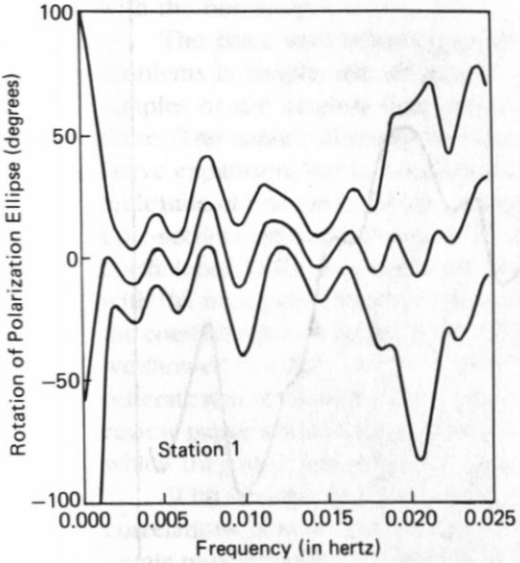
The rotation, θ, of the major axis of the polarization ellipse from the H-axis is given by

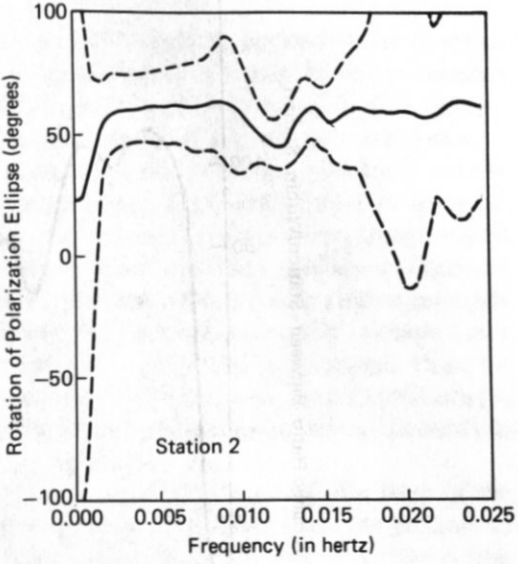
$$\tan 2\theta_j(f) = \frac{2\mathrm{Re}\{P_{12j}(f)\}}{P_{11j}(f) - P_{22j}(f)}. \tag{2.78}$$

Figure 2.17 shows the estimated rotation and jackknife ± 1 standard deviation lines for the EMSLAB data. For stations 1 and 4 the rotation is small, similar, and generally well constrained. At frequencies less than 2 mHz, variances are larger because the spectrum is rapidly increasing and may not be resolved. Above 15 mHz, the variance is larger simply because the signal-to-instrument-noise power ratio is poorer. In the center of the band, however, the two stations agree within a few degrees and their estimated deviations are also generally less than 20 degrees. Station 2, however, differs markedly from 1 and 4 in that the mean rotation during this interval is about 60 degrees although the jackknife deviation is similar to the others. Figure 2.18 is a more formal comparison of these differences: in it we have divided the between-station variance of the means, defined in the usual way, and here assigned two degrees of freedom by the average of the three within-station jackknife variances, the latter being assigned $3(N-1)$ degrees of freedom, with the result presented as an F statistic. This F statistic is plotted in Fig. 2.18, and clearly the difference between stations is highly significant. In this example the behavior of the rest of the polarization parameters was similar to rotation, and we should remark that usually the assumptions are satisfied and the field appears homogeneous.

2.8 SUMMARY

The purpose of this chapter has been a detailed description of methods for computing the variances of estimates of the various frequency-domain functions commonly used in time-series analysis without imposing unverifiable distributional or model assumptions on either the basic estimate or on the variance estimate. Motivation for





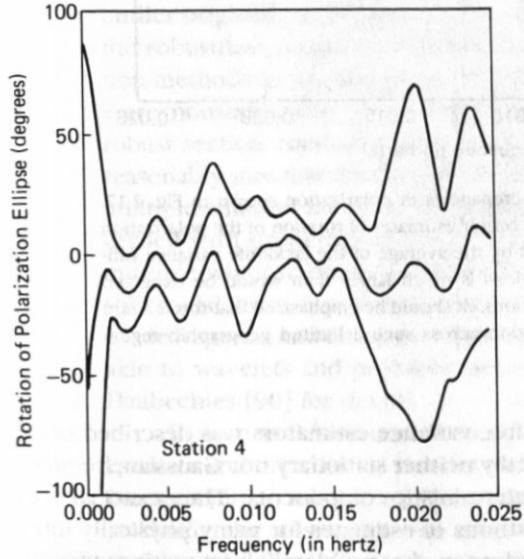


Figure 2.17 These three figures show estimates of the rotation of the polarization ellipse in the *H-D* plane for the EMSLAB data of Fig. 2.16. As usual, standard and jackknife ±1 standard deviation lines are plotted. Note that the orientation is well determined between 2 and 16 mHz, and that the estimates for stations 1 and 4 are essentially identical. The intermediate station, 2, however, is significantly different, with a mean polarization rotation about 60 degrees.

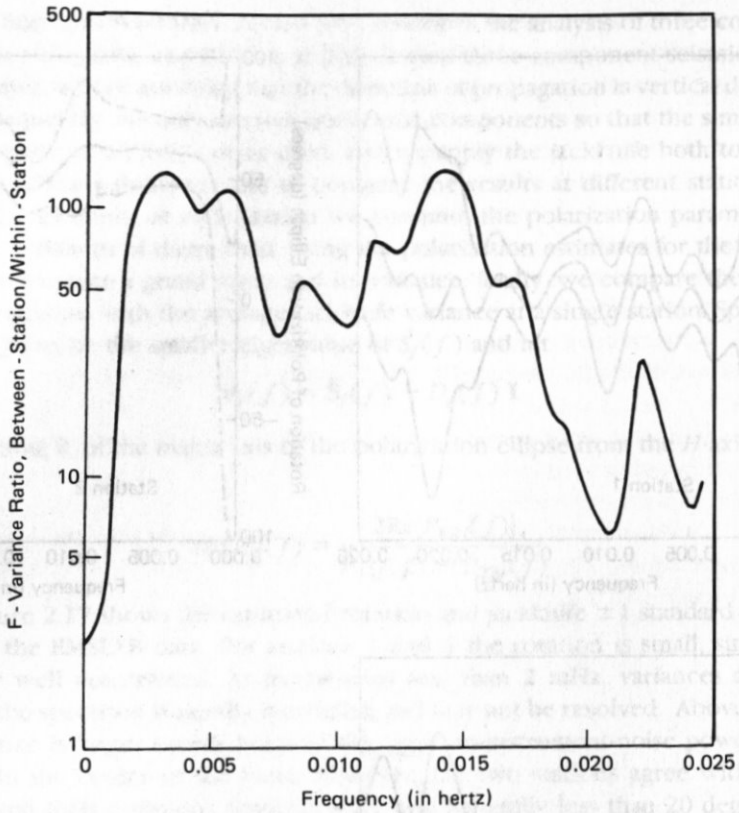


Figure 2.18 A summary of the discrepancies in polarization shown in Fig. 2.17. The variance computed between the center estimates of rotation of the polarization ellipse for the three stations, divided by the average of the jackknife variance estimates. The differences between stations is much larger than would be expected from estimates made at individual stations. It should be emphasized that this is a rare condition and that, usually, polarization across such a limited geographic region shows only small changes.

using nonparametric and distribution-free variance estimators was described in the introduction; time-series data are typically neither stationary nor Gaussian, frequently contain outliers, and often contain deterministic components. Thus, exact analytic expressions for the probability distributions of estimates for many physically interesting parameters are impractical, and even the problem of estimating effective degrees of freedom for approximate distributions in the unrealistic case of stationary Gaussian data can be difficult. The need for a robust and generally applicable alternative method of evaluating sampling properties of estimators is, of course, not restricted to time-series problems and several possibilities have been studied in the statistics literature. These methods, however, are restricted to independent samples and fail with correlated data. In our adaptation of these methods to time-series prob-

lems we have concentrated on the jackknife, but the same adaptation can be used with the bootstrap.

The basic idea behind our adaptation of resampling methods to time-series problems is simple; use an analysis method that transforms the highly correlated samples of the original time series into uncorrelated coefficients, then resample these. The canonical example of such a transformation is given by the Karhunen-Loève expansion, but it is unsuited for most spectrum estimation problems. Of the multitude of time-series methods available we found that multiple-window and multiple-section methods, plus their obvious combination, result in nearly uncorrelated coefficients. If the true spectrum is simple, the basic complex coefficients obtained with the multiple-window method are uncorrelated; with multiple-section methods the coefficients have an easily described Toeplitz correlation structure. In both cases we showed that the jackknife variance estimate is asymptotically unbiased. Thus, we reiterate that the failures of the jackknife in time-series problems noted in Miller's [3] review paper are, in our opinion, an indictment of the poor time-series methods to which they were being applied, not of the jackknife.

With multiple-section estimates we may argue that, because the form of the correlations is known from (2.22), we may filter the multisection coefficients to obtain uncorrelated estimates and improve on the results of Section 2.4. This is true in a restricted sense, but, because such filtering also has the effect of spreading an outlier originally restricted to one or two segments over many segments, it destroys the robustness properties that were the original motivation for using multiple-section methods in the first place. We remark, parenthetically, that if we expect moderate nonstationarity or clumpy outliers, then multiple-section methods with a robust section combiner are currently the estimators of choice; however, if we are reasonably sure that the data are stationary and statistical efficiency is important, then multiple-window methods are preferred. Much of Sections 2.3 and 2.4 were concerned with properties of the basis set used to compute the spectrum estimate. In multiple-window estimates the basis functions are the Slepian, or discrete prolate spheroidal wave, functions that are time-limited, optimally concentrated in frequency, orthogonal, and complete. The multiple-section basis, on the other hand, is more akin to wavelets and probably better described in terms of frames than bases; see Daubechies [90] for details.

In Section 2.5 we described a close connection between the jackknife variance estimate and the classical Bartlett and Lehmann tests for homogeneity of variance. It is reasonable to ask why such tests cannot be applied directly instead of through the jackknife; the answer is that adaptive methods, e.g., using (2.20) to determine weights, as well as robust procedures require moderate sample sizes and so preclude such direct application. Note that the use of delete-one estimates give the jackknife and bootstrap a much wider range of applicability than common variance formulas based directly on the uncorrelated coefficients. For instance most of the examples in Sections 2.6 and 2.7 *cannot* be computed in any other way; single-coefficient coherency estimates are all identically 1, and the eigenvalue distribution of rank-one matrices is exceptionally uninformative. In these applications we have used delete-one

formulas with complex data although these are usually treated as having two degrees of freedom, the difference being absorbed either in circular variances or, as with coherence, by estimating two real variances simultaneously. The problem of counting degrees of freedom deleted was addressed by Dempster [91] but needs further study with complex data.

Section 2.6 covered coherence and phase estimates in reasonable detail, with emphasis on the use of transformations. For magnitude-squared coherence the inverse hyperbolic tangent transform was used, and it was recommended that the frequency-averaged jackknife variance be used as a diagnostic. For phase the circular variance was computed.

In Section 2.7 a beginning was made on multivariate problems. Transfer function calculations were mentioned, but, as these parallel ordinary regression, our emphasis was on problems where much less is available. These included hypothesis tests, test of homogeneity across an array, polarization, and errors in variable regression. Similar examples of these methods appear in [92] and jackknife methods are rapidly being adopted by the electromagnetic geophysics community because it works well. Other recent examples in related fields are given in [93] and [94].

Finally, where does this paper leave the subject? First, several of the examples given here depart radically from standard theory. We emphasize that these examples were not specifically sought out to be particularly perverse, but were simply ones that the authors had come across in the course of our work during the last few years. Consequently, we believe that estimates of power spectra and related functions used for scientific inference should always be computed using robust methods together with jackknife or bootstrap confidence intervals and that this paper, taken together with those on robust estimation of spectra, permit this to be done for most such problems. Further, given such data, we must either seriously question the existing folklore of the field saying that Fourier transforms of nearly stationary series must be nearly Gaussian or find reasons for the discrepency. In these problems, a deficiency of variance usually implies the presence of deterministic components, but explaining excessive variance is largely an open problem. A start on this problem using the singular value decomposition of the time-frequency matrix of log spectral estimates made by a combined multiwindow, multisection method is in Thomson [95]. Other possible explanations include unresolved fine detail in the spectrum and nonlinear interactions.

Other outstanding problems are how to apply resampling methods to higher-order spectrum estimates such as the bispectrum. We have not treated mixed spectra in detail and only "scratched the surface" of multivariate problems. Progress on multivariate and the extensions to array and multidimensional processes is hindered by numerous problems: regression diagnostics for complex data; the degrees-of-freedom problem mentioned needs work, particularly given that with electromagnetic data we may, in effect, be deleting quarternions; not to mention the graphics problems of displaying multidimensional functions of frequency with uncertainties.

ACKNOWLEDGMENTS

It is our pleasure to thank Louis J. Lanzerotti, Carol Maclennan, and Les Medford for the Frobisher Bay and TAT-6 data; Rob Gold and Carol Maclennan for the *Voyager* records; Nick Pisias for the V22-174 $\delta^{18}O$ records and time-scale; and Alan Jones for the data used in the EMSLAB examples.

REFERENCES

1. J. W. Tukey, "Bias and Confidence in Not-Quite Large Samples," *Ann. Math. Stat.* **29**, (1958), 614.
2. M. Quenouille, "Approximate Tests of Correlation in Time Series," *J. Royal Stat. Soc.* **B 11**, pp 18-84, 1949.
3. R. G. Miller, "The Jackknife—A Review," *Biometrika* **61**, (1974), 1–15.
4. F. Mosteller and J. W. Tukey, *Data Analysis and Regression* (Reading, Mass.: Addison-Wesley, 1977.).
5. B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans* (Philadelphia: SIAM, 1982.).
6. B. Efron and G. Gong, "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-validation," *Am. Stat.* **37** (1983), 36–48.
7. D. R. Brillinger, *Time Series: Data Analysis and Theory* (San Francisco: Holden-Day, 1981).
8. N. R. Goodman, "On the Joint Estimation of the Spectra, Cospectrum, and Quadrature Spectrum of a Two-Dimensional Stationary Gaussian Process," Sci. Paper No. 10, Engineering Statist. Lab, New York Univ., 1957.
9. D. J. Thomson, "Spectrum Estimation and Harmonic Analysis," *Proc. IEEE* **70** (1982), 1055–1096.
10. J. W. H. Swanepoel and J. W. J. van Wyk, "The Bootstrap Applied to Power Spectral Density Function Estimation," *Biometrika* **73** (1986), 135–142.
11. D. V. Hinkley, "Improving the Jackknife with Special Reference to Correlation Estimation," *Biometrika*, **65** (1978), 13–21.
12. B. Efron and C. Stein, "The Jackknife Estimate of Variance," *Ann. Stat.* **9** (1981), 586–596.
13. J. Shao and C. F. J. Wu, "Heteroscedasticity—Robustness of Jackknife Variance Estimators in Linear Models," *Ann. Statist.*, **15** (1987), 1563–1579.
14. T. Cover and J. Thomas, "Determinant Inequalities via Information Theory," *SIAM J. Matrix Anal. and Appl.* **9** (1988), 384–392.
15. D. V. Hinkley, "Jackknife Confidence Limits Using Student *t* Approximations," *Biometrika* **64** (1977), 21–28.
16. N. Cressie, "Transformations and the Jackknife," *J. Roy. Stat. Soc.*, **B43** (1981), 177–182.
17. R. G. Miller, "An Unbalanced Jackknife," *Ann. Stat.* **2** (1974), 880–891.
18. D. V. Hinkley, "Jackknifing in Unbalanced Situations," *Technom.* **19** (1977), 285–292.

19. D. Hinkley and Hai-Li Wang, "A Trimmed Jackknife," *J. Royal Stat. Soc.* **B 42** (1980), 347–356.
20. A. D. Chave, D. J. Thomson, and M. E. Ander, "On the Robust Estimation of Power Spectra, Coherences, and Transfer Functions," *J. Geophys. Res.* **B92** (1987), 633–648.
21. J. A. Reeds, "Jackknifing Maximum Likelihood Estimates," *Ann. Stat.* **6** (1978), 727–739.
22. W. C. Parr, "Jackknifing Differentiable Statistical Functionals," *J. Royal Stat. Soc.* **B 47** (1985) 56–66.
23. J. Shao and C. F. J. Wu, "A General Theory for Jackknife Variance Estimation," *Ann. Stat.* **17** (1989), 1176–1197.
24. P. D. Welch, "A Direct Digital Method of Power Spectrum Estimation," *I.B.M. J. Res. Devel.* **5** (1961), 141–156.
25. D. J. Thomson, "Spectrum Estimation Techniques for Characterization and Development of the WT4 Waveguide," *Bell Syst. Tech. J.* **56** (1977), 1769–1815; (1977), 1983–2005.
26. D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—I," *Bell Syst. Tech. J.* **40** (1961), 43–64.
27. H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty—II," *Bell Syst. Tech. J.* **40** (1961), 65–84.
28. D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty—V: The Discrete Case," *Bell Syst. Tech. J.* **57** (1978), 1371–1429.
29. M. Rosenblatt, "Some Comments on Narrow Band-Pass Filters," *Quart. Appl. Math.* **18** (1961), 387–393.
30. C. L. Mallows, "Linear Processes Are Nearly Gaussian," *J. Appl. Prob.* **4** (1967), 313–329.
31. B. N. Nagarsenker and M. M. Das, "Exact Distribution of Sphericity Criterion in the Complex Case and Its Percentage Points," *Comm. in Stat.* **4** (1975), 363–374.
32. G. S. Watson, *Statistics on Spheres* (New York: Wiley-Interscience, 1983).
33. K. S. Miller, *Complex Stochastic Processes* (Reading, Mass.: Addison-Wesley, 1974).
34. D. R. Brillinger, "The Key Role of Tapering in Spectrum Estimation," *IEEE Trans. Acoustics, Speech, and Sig. Proc.* **ASSP-29** (1981), 1075–1076.
35. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis,* 2nd ed. (New York: John Wiley, 1984).
36. M. S. Bartlett and D. G. Kendall, "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation," *J. Royal Stat. Soc. Suppl.* **8** (1946), 128–138.
37. J. Wishart, "The Cumulants of the Z and of the Logarithmic χ^2 and t Distributions," *Biometrika* **34** (1947), 170–178.
38. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions,* (National Bureau of Standards, 1964).
39. P. Hall, "Theoretical Comparison of Bootstrap Confidence Intervals (with Discussion)," *Ann. Stat.* **16** (1988), 927–985.
40. A. J. Izenman and S. K. Sarkar, "Simultaneous Confidence Intervals for the Frequency Analysis of Multiple Time Series," *J. Am. Stat. Assoc.* **82** (1987), 271–275.

41. N. Sugiura and H. Nagao, "On Bartlett's Test and Lehmann's Test for Homogeneity of Variances," *Ann. Math. Stat.* **40** (1969), 2018–2032.

42. B. K. Ghosh, "On Lehmann's Test for Homogeneity of Variances," *J. Royal Stat. Soc.* **B-34** (1972), 221–234.

43. C. P. Han, "Testing the Homogeneity of a Set of Correlated Variances," *Biometrika* **55** (1968), 317–326.

44. G. E. P. Box, "Non-normality and Tests on Variances," *Biometrika* **40** (1953), 318–335.

45. M. A. Jorgensen, "Jackknifing Fixed Points of Iterations," *Biometrika* **74** (1987), 207–211.

46. A. Bose, "Edgeworth Correction by Bootstrap in Autoregressions," *Ann. Stat.* **16** (1988), 1709–1722.

47. C. G., Maclennan et. al., "Statistical Properties of Shock Accelerated Ions in the Outer Heliosphere," *Int. Assoc. Geomagn. Aeron.,* Exeter (1989).

48. E. McKenzie, "Product Autoregression: A Time-Series Characterization of the Gamma Distribution," *J. Appl. Prob.* **19** (1982), 463–468.

49. G. C. Carter, "Coherence and Time Delay Estimation," *Proc. IEEE* **75** (1987), 236–255.

50. J. C. Lee, "On Bias Reduction in Estimation of the Magnitude-Squared Coherence Function," *Bull. Inform. and Cyber.* **20** (1983), 107–114.

51. E. J. Hannan, *Multiple Time Series* (New York: John Wiley, 1970).

52. G. C. Carter, C. H. Knapp, and A. H. Nuttall, "Statistics of the Estimate of the Magnitude-Coherence Function," *IEEE Trans. on Audio and Electroacoust.* **AU-21** (1973), 388–389.

53. L. H. Koopmans, *The Spectral Analysis of Time Series* (New York: Academic Press, 1974).

54. D. E. Amos and L. H. Koopmans, "Tables of the Distribution of the Coefficient of Coherence for Stationary Bivariate Gaussian Processes," Sandia Corp. Monograph SCR-483 (1963).

55. G. E. P. Box and G. M. Jenkins, *Time Series Analysis* (San Francisco: Holden-Day, 1970).

56. K. V. Mardia, *Statistics of Directional Data* (New York: Academic Press, 1972).

57. J. Imbrie and others, "The Orbital Theory of Pleistocene Climate: Support from a Revised Chronology of the Marine δ^{18} O Record," In pp. 269–305 of *Milankovitch and Climate,* A. Berger and others, eds. (Dordrecht: D. Reidel, 1984).

58. A. L. Berger, "Long-Term Variations of Daily Insolation and Quaternary Climatic Changes," *J. Atmosph. Sci.* **35** (1978), 2362–2367.

59. S. O. Rice, "Statistical Properties of a Sine Wave Plus Random Noise," *Bell Syst. Tech. J.* **27** (1948), 109–157.

60. D. Middleton, *Statistical Communication Theory* (New York: McGraw-Hill, 1960).

61. W. H. Munk and D. E. Cartwright, "Tidal Spectroscopy and Prediction," *Philos. Trans. Roy. Soc. London* **Ser. A:** 1105, 259 (1966), 533–581.

62. C. N. K. Mooers, "A Technique for the Cross-Spectrum Analysis of Pairs of Complex-Valued Time Series, with Emphasis on Properties of Polarized Components and Rotational Invariants," *Deep-Sea Res.* **20** (1973), 1129–1141. Also, see Vol. 29 (1982), 1267–1269, for comments and corrections by J. H. Middleton.

63. D. F. Andrews and others, *Robust Estimates of Location* (Princeton, N. J.: Princeton University Press, 1972).

64. The EMSLAB Group (33 authors), "The EMSLAB Electromagnetic Sounding Experiment," *EOS, Trans. AGU* **69** (1988), 89, 98–99.

65. C. F. J. Wu, "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis (with Discussion)," *Ann. Stat.* **14** (1986), 1261–1350.

66. K. S. Miller, "Complex Linear Least Squares," SIAM Rev. **15** (1973), 706–726.

67. D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics* (New York: John Wiley, 1980).

68. R. D. Cook and S. Weisberg, *Residuals and Influence in Regression* (London: Chapman and Hall, 1982).

69. G. H. Golub and C. F. Van Loan, *Matrix Computations* (Baltimore, Md.: Johns Hopkins University Press, 1983).

70. S. Haykin, *Modern Filters* (New York: Macmillan, 1989).

71. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Englewood Cliffs, N. J.: Prentice-Hall, 1974).

72. J. J. Dongarra, C. B. Moler, J. R. Bunch, and G. W. Stewart, *LINPACK Users' Guide* (Philadelphia: SIAM, 1979).

73. L. J. Lanzerotti and others, "Electromagnetic Study of the Atlantic Continental Margin Using a Section of a Transatlantic Cable," *J. Geophys. Res.* **B91** (1986), 7417–7427.

74. N. R. Draper and H. Smith, *Applied Regression Analysis* (New York: John Wiley, 1981).

75. C. L. Mallows, "Some Comments on C_p," *Technometrics* **15** (1973), 661–675.

76. M. L. Thompson, "Selection of Variables in Multiple Regression," *International Stat. Rev.* **46** (1978), Part 1 1–9, Part II 129–146.

77. T. W. Anderson, "Linear Statistical Relationships," *Ann. Stat.* **12** (1984), 1–45.

78. W. A. Fuller, *Measurement Error Models* (New York: John Wiley, 1987).

79. J. Park and A. D. Chave, "On the Estimation of Magnetotelluric Response Functions Using the Singular Value Decomposition," *Geophys. J. Roy. Astr. Soc.* **77** (1984), 683–709.

80. G. H. Golub and C. F. Van Loan, "An Analysis of the Total Least Squares Problem," *SIAM J. Num. Anal.* **17** (1980), 883–893.

81. D. R. Brillinger, "Discussion on a Generalized Least-Squares Approach to Linear Functional Relationships," *J. Royal Stat. Soc.* **B-28** (1966), 297.

82. A. K. Gupta, "Distribution of Wilk's Likelihood-Ratio Criterion in the Complex Case," *Ann. Inst. Stat. Math.* **23** (1971), 77–87.

83. P. R. Krishnaiah, J. C. Lee, and T. C. Chang, "The Distributions of the Likelihood Ratio Statistics for Tests of Certain Covariance Structures of Complex Multivariate Normal Populations," *Biometrika* **63** (1976), 543–549.

84. W. J. Conover, M. E. Johnson, and M. M. Johnson, "A Comparative Study of Tests for Homogeneity of Variances, With Applications to the Outer Continental Shelf Bidding Data," *Technometrics.* **23** (1981), 351–361.

85. D. D. Boos and C. Brownie, "Bootstrap Methods for Testing Homogeneity of Variances," *Technometrics* **31** (1989), 69–82.

86. M. Born and E. Wolf, *Principles of Optics,* 6th ed. (Elmsford, N. Y.: Pergamon, 1980).

87. J. D. Means, "Use of the Three-Dimensional Covariance Matrix in Analyzing the Polarization Properties of Plane Waves," *J. Geophys. Res.* **77** (1972), 5551–5559.

88. J. Park, F. L. Vernon III, and C. R. Lindberg, "Frequency Dependent Polarization Analysis of High-Frequency Seismograms," *J. Geophys. Res.* **92** (1987), 12664–12674.

89. R. A. Fowler, B. J. Kotick, and R. D. Elliott, "Polarization Analysis of Natural and Artificially Induced Geomagnetic Micropulsations," *J. Geophys. Res.* **72** (1967), 2871–2873.

90. I. Daubechies, "The Wavelet Transform, Time-Frequency Localization, and Signal Analysis," *IEEE Trans. Information Theory,* (1990), in press.

91. A. P. Dempster, "Estimation in Multivariate Analysis," In pp. 315–334 of *Multivariate Analysis,* P. R. Krishnaiah, ed. (New York: Academic Press, 1966).

92. A. D. Chave and D. J. Thomson, "Some Comments on Magnetotelluric Response Function Estimation," *J. Geophys. Res.* **94** (1989), 14215–14225.

93. L. A. Hinnov, & J. Park, "Multi-Windowed Spectrum Estimates of the ILS Polar Motion," pp. 221–226 of *The Earth's Rotation and Reference Frames for Geodesy and Geophysics,* IAU symposium 124, A. K. Babcock and G. A. Wilkins, eds., Kluwer Academic, Dordrecht, 1988.

94. F. L. Vernon III, "Analysis of Data Recorded on The Anza Seismic Network," Ph.D. dissertation, University of California, San Diego, 1989.

95. D. J. Thomson, "Time Series Analysis of Holocene Climate Data," *Phil. Trans. Roy. Soc.,* 1990, in press.