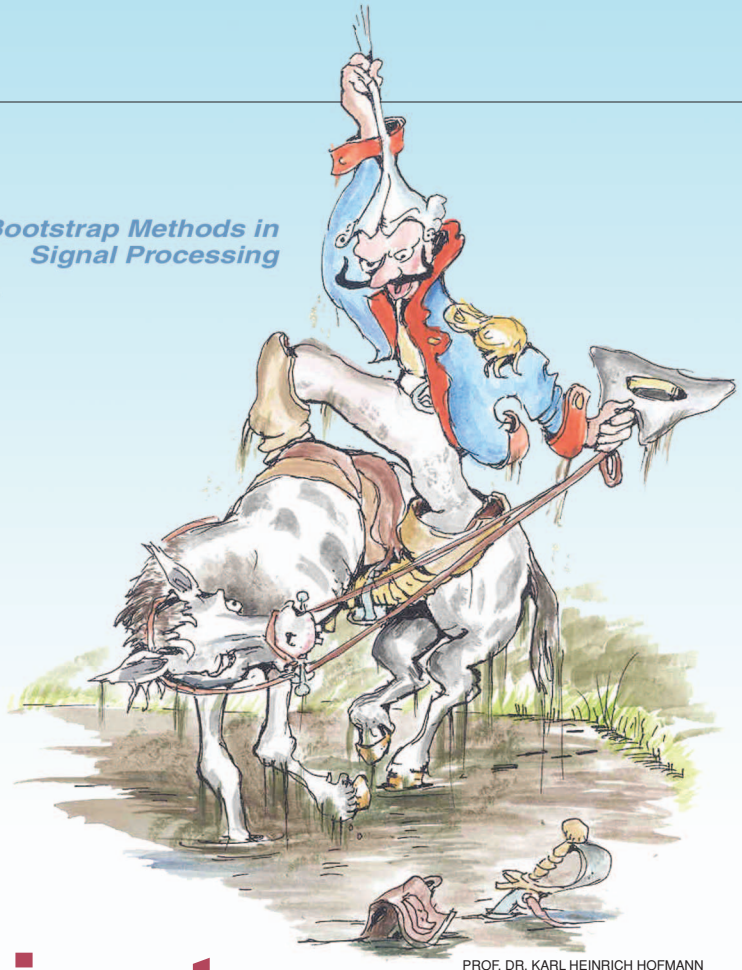
David J. Thomson

Bootstrap Methods in Signal Processing



PROF. DR. KARL HEINRICH HOFMANN

Jackknifing Multitaper Spectrum Estimates

Identifying variances of complicated estimation procedures

This article discusses some examples of jackknifing multitaper estimates of spectra, coherences, and frequency estimates. Examples include barometric pressure data, where spectrum with an extremely large range plus many narrow-band processes are seen. Analysis of dropped-call rates in cellular phone systems and their coherence with solar radio data illustrates further uses of the jackknife and some of the complexities encountered in processes with many spectral lines. The third example is a re-examination of the 663-year record of Nile River levels, a process claimed to be long-memory. There are persistent terms, but not of the form usually classed as long-memory. We begin by explaining the terms in the title.

The term jackknife was introduced into statistical usage by John W. Tukey in an unpublished 1962 memorandum with a Meaning Given in [1]:

> Tukey adopted the name 'jackknife' for this procedure, since a boy scout's jackknife is symbolic of a rough-and-ready instrument capable of being utilized in all contingencies and emergencies.

I recall John giving a very similar description himself. While an outline of the method will be given in the following, it is, briefly, a general method to estimate the variance of complicated estimation procedures without making overly restrictive assumptions. It was originally introduced [2] as a method of bias reduction, but as most of the bias in time-series problems is subtly

1053-5888/07/$25.00©2007IEEE

different from that in general statistics, bias reduction is less useful than the variance estimate.

Multitaper refers to methods for estimating power spectra, coherences, and related quantities using an orthogonal set of data tapers, specifically the discrete prolate spheroidal or Slepian sequences. This method was introduced in [3] and is outlined in a following section. Multitaper methods are ideally suited to jackknifing and have been used with the jackknife almost since their discovery. They were described outside Bell Labs at Scripps Institution of Oceanography in 1983 and at the 1984 Brighton ISIT, with details appearing in [4]. The combination has been responsible for both scientific and engineering progress.

In this article, *spectrum* refers to the power spectrum or power spectral density used to describe the distribution of energy as a function of frequency and not in the physicist's sense of density of charged particles as a function of energy or the mathematician's usage to characterize operators or eigenvalues.

Finally, *estimates* implies that we are working with data and specifically, attempting to estimate the power spectral density as a function of frequency from a time-series.

Much of this research was initiated as part of my study of problems in various communications systems at Bell Labs: geometric distortions in millimeter waveguides and optical fibers, abnormal rates of dropped calls in cellular phone systems, outages in communications satellites, induced voltages on ocean cables, and non-white noise backgrounds. These are mostly problems that occur after the system is designed and installed, and consequently, one is dealing with field measurement data rather than laboratory measurements. Typically, these are neither simple nor Gaussian. It was realized that finding engineering solutions required better understanding of the physics. This understanding, in turn, depended on advances in signal processing. Some of this work [5]–[7] has begun to change our view of space physics. A summary of the solar and space physics background, estimation procedures, and additional references is given in [8]. As a specific example, the most important outcome of work that began as an investigation of communications satellite outages was showing that what superficially appeared to be harmonics of solar rotation in energetic particle fluxes were not harmonically related [9]. This demonstration depended on jackknife confidence intervals for frequency estimates.

This section is followed by short introductions to the jackknife, multitaper estimates, and their combination. These are followed by examples: microbarometer data, dropped calls in a cellular phone system, and the Nile River data. The article ends with some general notes on the methods and a summary.

THE JACKKNIFE

Details on the jackknife are given in numerous texts and papers (see, e.g., [4] and [10]–[12]) so only a summary is given here.

Assume that we have a sample of K independent observations, $\{x_i\}$, $i = 1, \ldots K$, drawn from some distribution characterized by a parameter θ, which is to be estimated. Here, θ is usually a spectrum or coherence at a particular frequency or a simple parameter such as the frequency of a periodic component. Denote an estimate of θ made using all K observations by $\hat{\theta}_{\text{all}}$. Next, subdivide the data into K groups of size $K-1$ by deleting each entry in turn from the whole set, and let the estimate of θ with the ith observation deleted be

$$\hat{\theta}_{\setminus i} = \hat{\theta}\{x_1, \ldots x_{i-1}, x_{i+1}, \ldots x_K\} \tag{1}$$

for $i = 1, 2, \ldots, K$, where the subscript $\setminus$ is the set-theoretic sense of without. Using $\bullet$ in the statistical sense of averaged over, define the average of the K delete-one estimates as

$$\theta_{\setminus\bullet} = \frac{1}{K}\sum_{i=1}^{K}\hat{\theta}_{\setminus i} \tag{2}$$

and the jackknife variance of $\hat{\theta}_{\text{all}}$ as

$$\widehat{\text{Var}}\{\hat{\theta}_{\text{all}}\} = \frac{K-1}{K}\sum_{i=1}^{K}(\hat{\theta}_{\setminus i} - \theta_{\setminus\bullet})^2 . \tag{3}$$

The scale factor reflects the fact that the various delete-one estimates $\hat{\theta}_{\setminus i}$ are not independent. The jackknife variance estimate is conservative, and Reeds [13] showed that under a set of conditions that 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 to the population values under only slightly weaker conditions than are required for maximum-likelihood. Almost sure convergence of the jackknife variance estimate, however, requires a moment condition; for the estimators considered here, this is satisfied.

To conclude this section, I must emphasize that the jackknife variance estimate (3) should always be compared with its expected value under standard assumptions. If the estimate disagrees significantly with its expected value, we must look for the cause. In my experience, the most common causes in spectrum estimation problems have been unsuspected periodic components (see the section on Nile data), complicated structure in the continuum spectrum (detectable with quadratic-inverse methods [14]), and nonstationarity. In long series, nonstationarity is best shown by dynamic spectrum estimates and their singular value decomposition [15], but in short series, nonstationary quadratic-inverse methods [7] are preferred.

MULTITAPER ESTIMATES OF SPECTRA

While the theory of multitaper estimates is given in several papers ([3], [7], [14], [16]), the cook book recipe for computing a multitaper spectrum estimate consists of the following four steps.

First, choose a time-bandwidth product $C_o = NW = BT$ where N is the sample size, W is the bandwidth in standardized units, $T = N\delta t$ (the total time duration of the sample), δt is the sample spacing, typically in seconds, and B is the analysis bandwidth in physical units, typically in Hertz. If C_o is too small, the estimate will be unstable and may not have enough dynamic range, but if C_o is too large, the estimate may not

have adequate frequency resolution. For a given choice of C_o, there are $K = \lfloor 2C_o \rfloor$ data tapers, the dimension of the time-frequency region. Similarly, except for the origin and Nyquist frequency with real-valued data, each dimension contributes two degrees-of-freedom (DoF), so one has $\nu = 2K = 4C_o$ DoF. Because the energy concentration of the higher-order tapers is poorer than that of the low-order ones, it is common to choose K one or two less than $\lfloor 2C_o \rfloor$.

Second, compute the data tapers. These are a set of special functions known as discrete prolate spheroidal sequences, or, in honor of their inventor, David Slepian, Slepian sequences. Slepian sequences are defined as the real, unit-energy sequences on $[0, N-1]$ having the greatest energy in a bandwidth W. I use Slepian's notation $v_n^{(k)}(N, W)$ [17] for sample n of the kth sequence, or for short, $v_n^{(k)}$. These sequences are solutions of the symmetric Toeplitz matrix eigenvalue equation

$$\lambda_k v_n^{(k)} = \sum_{m=0}^{N-1} \frac{\sin 2\pi W(n-m)}{\pi(n-m)} v_m^{(k)} \tag{4}$$

for $0 \le n \le N-1$ and are defined by this equation for n outside this interval. Note that solving this equation directly is not the way to compute Slepian sequences. Use the tridiagonal form in Appendix B of [14] instead.

The eigenvalues are bounded between zero and one with the first

$$K \approx \lfloor 2NW \rfloor \tag{5}$$

of them large, i.e., nearly one, and the rest nearly zero, so the dimensionality of the subspace is approximately $2NW$.

Third, compute the eigencoefficients

$$y_k(f) = \sum_{t=0}^{N-1} x(t) v_t^{(k)} e^{-i2\pi ft}, \tag{6}$$

where $\{x(t)\}$ is the sequence of data samples and f is the frequency in standard units. These eigencoefficients are Fourier transforms of the data multiplied by a window, or taper, exactly the form suggested by Tukey [18]. Pragmatically, we begin with a time-bandwidth product, $C_o = NW$, between about three and six, with four to ten windows. This gives bandwidths of a few times the Rayleigh resolution.

Fourth, estimate the spectrum using the adaptive weighted forms introduced in [3]. Using a minimum mean-square-error estimate of the ideal eigencoefficients $x_k(f)$ of the form $\hat{x}_k(f) = d_k(f) y_k(f)$, where the $d_k(f)$'s are weights, this is essentially a Wiener filter on the Slepian basis. In terms of the spectrum $S(f)$ and bias $B_k(f)$, the weights are

$$d_k(f) = \frac{\sqrt{\lambda_k} S(f)}{\lambda_k S(f) + B_k(f)} \tag{7}$$

with a corresponding spectrum estimate

$$\hat{S}(f) = \frac{\sum_{k=0}^{K-1} |d_k(f) y_k(f)|^2}{\sum_{k=0}^{K-1} |d_k(f)|^2}. \tag{8}$$

Because $S(f)$ and $B_k(f)$ in (7) are unknown, we use estimates and iterate. See, e.g., [3] and [7] for details. The resulting spectrum estimate, $\hat{S}(f)$, is the solution of

$$\sum_{k=0}^{K-1} \frac{\lambda_k (\hat{S}(f) - |y_k(f)|^2)}{[\lambda_k \hat{S}(f) + \hat{B}_k(f)]^2} = 0, \tag{9}$$

where $\hat{B}_k(f)$ is an estimate of the bias of $|y_k(f)|^2$ at frequency f. A simple estimate of, and an upper bound, for $\hat{B}_k(f)$ is $(1-\lambda_k)\sigma^2$, where σ^2 is the process variance. Denote the eigenspectra corresponding to the individual tapers by $\widehat{S}_k(f) = |d_k(f) y_k(f)|^2$. When the range of the spectrum is small, the weights are typically all nearly one, and $\widehat{S}_k(f) \approx |y_k(f)|^2$.

When the range of the spectrum is not too large, that is, the bias terms satisfy $\hat{B}_k(f) \ll |y_k(f)|^2$, the multitaper estimate (9) reduces to

$$\widehat{S}_c(f) = \frac{1}{K} \sum_{k=0}^{K-1} \widehat{S}_k(f). \tag{10}$$

If the spectrum varies slowly in the narrow band $(f-W, f+W)$, the various y_ks will be approximately uncorrelated because the Slepian sequences are orthonormal and the estimate will be distributed as χ^2_{2K}. In this case, the multitaper estimate is approximately maximum likelihood [19].

Since the spectrum estimate is, at least approximately, a χ^2 random variate, the expected value of its logarithm is biased [7],

$$\mathbf{E}\{\ln \hat{S}\} = \ln S + B_\chi(K), \tag{11}$$

where the bias $B_\chi(K)$ is given by

$$B_\chi(K) = \psi(K) - \ln K \tag{12}$$

with ψ being the digamma function. Similarly, the variance of such an estimate is

$$\mathbf{Var}\{\ln \hat{S}\} = \psi'(K), \tag{13}$$

where ψ' is the trigamma function.

If the data being studied were Gaussian and stationary, these formulae would be adequate, but with real data, these conditions rarely hold, and we must obtain estimates from the observations themselves. The first problem is to find an appropriate form of the estimation equations. Minimum requirements for such estimators are first, that they work well in the stationary Gaussian case and second, that they are sensitive enough to detect departures from assumptions. The jackknife is an ideal tool for this.

JACKKNIFING MULTITAPER SPECTRA

To jackknife multitaper spectrum estimates, my approach has been to begin with (9), omit the jth eigencoefficient from the weight

calculation, and take $\theta_{\setminus j} = \ln \widehat{S}_{\setminus j}(f)$ at each frequency. This treats the eigencoefficients as exchangeable data and effectively deletes each taper in turn. This is called "jackknifing over tapers" or, as the terms taper and window are used synomonously in spectrum estimation, "jackknifing over windows." When the approximation (10) is valid, one computes the delete-one values as

$$\ln \widehat{S}_{\setminus j}(f) = \ln \left[\frac{1}{N-1} \sum_{\substack{k=0 \\ k \neq j}}^{K-1} \widehat{S}_k(f) \right] \tag{14}$$

or, in general, by including the adaptive weighting and solving

$$\sum_{\substack{k=0 \\ k \neq j}}^{K-1} \frac{\lambda_k (\widehat{S}_{\setminus j} - |y_k(f)|^2)}{\left[\lambda_k \hat{S}(f) + \hat{B}_k(f)\right]^2} = 0\,. \tag{15}$$

Then take the average

$$\ln \widehat{S}_{\setminus \bullet}(f) = \frac{1}{K} \sum_{j=1}^{K} \widehat{S}_{\setminus j}(f) \tag{16}$$

and compute the variance estimate

$$\widehat{V}_J(f) = \frac{K-1}{K} \sum_{j=1}^{K} [\ln \widehat{S}_{\setminus j}(f) - \ln \widehat{S}_{\setminus \bullet}(f)]^2, \tag{17}$$

as in the general jackknife prescription.

Experiments with these estimators show that the variances of $\ln \widehat{S}(f)$ and $\ln \widehat{S}_{\setminus \bullet}(f)$ are nearly identical. The jackknife variance is slightly conservative, and [20] showed that (17) is too large by a factor of approximately $(K - (1/2))/(K-1)$. Thus, I recommend replacing the preceding equation with

$$\widehat{V}_J(f) = \frac{(K-1)^2}{K\left(K - \frac{1}{2}\right)} \sum_{j=1}^{K} [\ln \widehat{S}_{\setminus j}(f) - \ln \widehat{S}_{\setminus \bullet}(f)]^2\,. \tag{18}$$

The variance of this estimate was given in [20] in terms of trigamma functions. Their asymptotic expansion gives

$$\mathrm{E}\{\widehat{V}_J(f)\} \asymp \frac{(K-1)^3\,(K-3)}{\left(K - \frac{1}{2}\right) K(K-2)^3}\,, \tag{19}$$

a formula that is reasonably accurate for $K \geq 4$.

JACKKNIFE VERSUS DIRECT VARIANCE

A common question is, "Why go through the complications of delete-one estimates and the jackknife procedure instead of simply finding the average and variance directly?" One could compute the log-average

$$\overline{\ln S(f)} = \frac{1}{K} \sum_{k=0}^{K-1} \ln \widehat{S}_k, \tag{20}$$

where $\widehat{S}_k(f) \approx |y_k(f)|^2$ and the corresponding variance

$$\widehat{\mathrm{Var}}_d\{\ln S(f)\} = \frac{1}{K-1} \sum_{k=0}^{K-1} \left[\ln \widehat{S}_k - \overline{\ln S(f)}\right]^2\,. \tag{21}$$

This is Lehmann's test for homogeneity of variances [4]. From the preceding section, $\mathrm{E}\{\ln S_k(f)\}$is biased by $\psi(1) \approx -.57721$, and, as this is removable, the bias is not a problem. In contrast, for the variance estimate (21), we have

$$\mathrm{Var}\{\ln \hat{S}_k(f)\} = \psi'(1) = \frac{\pi^2}{6} \tag{22}$$

so that

$$\widehat{\mathrm{E}\{\mathrm{Var}_d\{\ln S(f)\}\}} = \frac{1}{K-1}\psi'(1) = \frac{1}{K-1}\frac{\pi^2}{6}, \tag{23}$$

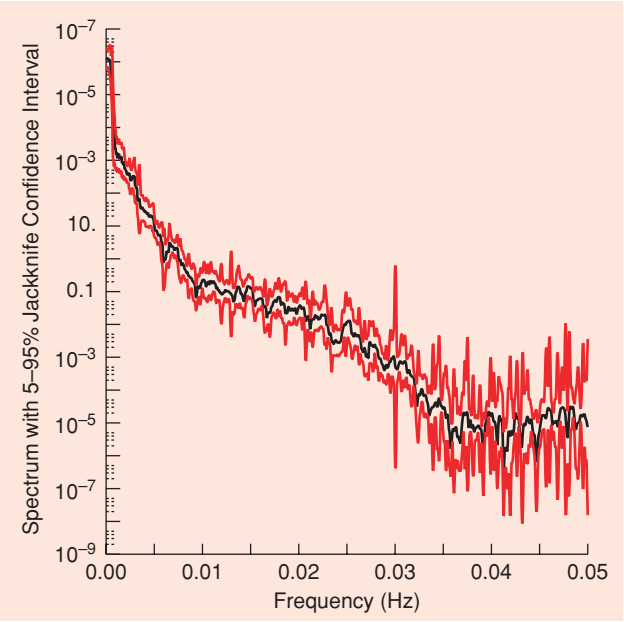
and this is unacceptably large. Taking $K = 4$ as an example, the exact form of (19) gives 0.2795, about one-half the variance of 0.5483 given by the obvious approach (23). The heuristic explanation is that for stationary Gaussian data, the individual eigenspectra $S_k(f)$ have a χ_2^2 (exponential) distribution so that the most probable value is zero and the distribution of $\ln S_k(f)$ has a very heavy lower tail. The delete-one estimates $\widehat{S}_{\setminus j}(f)$ on the other hand, have a χ^2_{2K-2} distribution, so the logarithms are much better behaved. Just going to a χ_4^2 from a χ_2^2 reduces the variance of $\ln S$ from 1.6449 to 0.6449, a factor of 2.55, so the jackknife approach gives a more efficient estimate of variance.

BAROMETER DATA

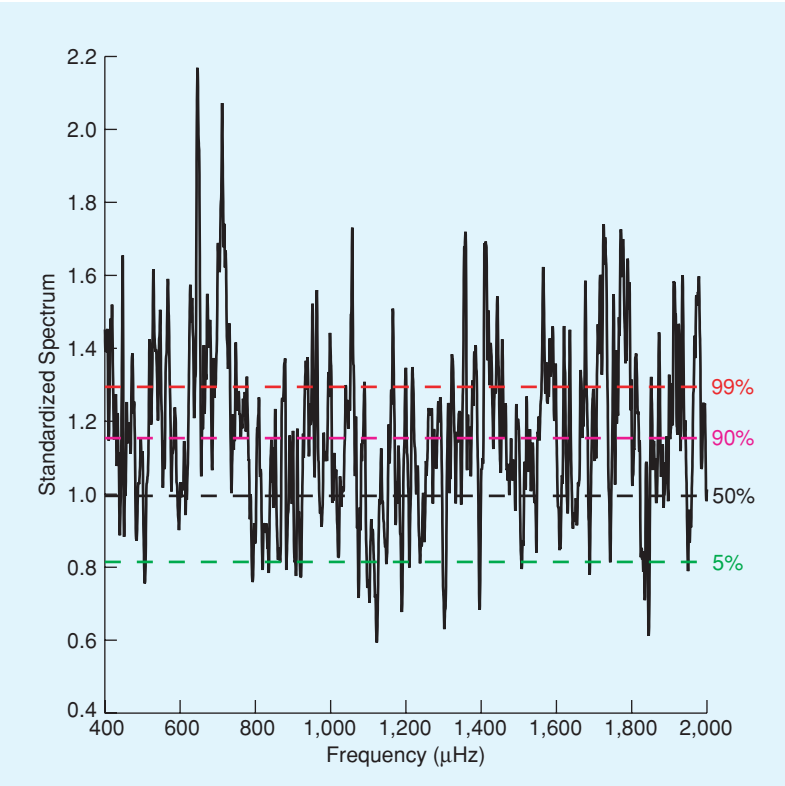
Barometric pressure data provides examples of two distinct phenomena; first, it contains many high-Q, high-significance peaks; second, it provides an excellent example of some of the dangers in spectrum estimation. Many seismic stations also have microbarometers, and the data shown here is from Black Forest Observatory (BFO). The data is sampled at a 10 s interval and has been carefully filtered before sampling. The least-significant-bit sensitivity of this instrument is $-1/303$ Pa.

The spectra shown in Figure 1 uses 800 samples (2.2 hours) typical of the block lengths used in many dynamic spectra. Using a time-bandwidth $C_o = 5$ and $K = 8$ tapers gave the spectrum shown. This spectrum has a range of $\sim 10^{12}$, somewhat less than the 10^{17} found using longer data spans. The raw data for this sample (not shown) appears to have an approximately linear trend, but as this is just part of the day-to-day fluctuation of atmospheric pressure associated with passage of storm fronts and the like, there is no justification for subtracting a trend from the data.

The jackknife 5–95% confidence region, shown by the upper and lower red curves, goes from relatively tight limits at some low frequencies, approximately what one would expect for a χ^2_{16} distribution, to a factor of $\sim 10^6$ at high frequencies. This range results mostly from loss of DoF associated with the large dynamic range, but a significant part comes from unresolved peaks in the spectrum.



[FIG1] Estimated spectrum with jackknife 5% and 95% confidence intervals for the barometer data. Note that, as opposed to a range of ∼4 expected under a Gaussian hypothesis, the range estimated by the jackknife is often closer to 10^6. To emphasize this point, the range predicted using stationary Gaussian statistics is too small by about a factor of 100,000.



[FIG2] A higher resolution prewhitened spectrum of the barometer data for the interval 21 March – 15 May 1999. The data was low-pass filtered and decimated to 1 min, prewhitened, and analyzed on ten blocks of 10,000 samples each and the spectra averaged. The resulting estimate has ∼144 DoF, so the largest peak is about 10σ above background. The horizontal dashed lines marked 5%, 50%, 90%, and 99% are quantiles for a χ^2_{144} distribution scaled to unit mean.

Prewhitening and using longer data sections (6.9 days) give the spectrum shown in Figure 2. Here, the largest peaks are about 9.9σ above background. Leaving aside the question of their origin (see [8]), there are 98 peaks above the 99% significance level in this frequency range that cover 26% of the total bandwidth. Estimating the noise level, or baseline, in such spectra must be done carefully because the peaks bias the average and even the median significantly. These spectra have a noncentral χ^2 distribution at modal frequencies and a central χ^2 distribution between modes. I have found (see [6, Appendix C]) that the lower 5% point of the prewhitened spectra gives a reasonable estimate the baseline. The variance of $\ln S$ is 2.92 times its expected value. Applying Bartlett's M-test across frequency gives a probability that this portion of the spectrum could be a random sampling fluctuation from a white spectrum of $\sim 10^{-42}$. Most of the rest of the spectrum shows similar structure, so one must conclude that we are not dealing with a trivial process that can be described by a low-order autoregressive (AR) or similar parametric model.

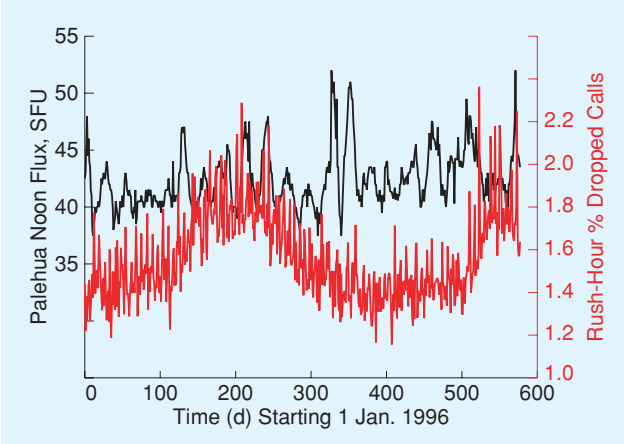
DROPPED CALLS IN CELLULAR PHONE SYSTEMS

A perplexing problem in communications systems is that of excessive dropped calls in cell phone systems. Because the systems operate in an interference-limited, fading environment, dropped calls were expected. Experience showed that, instead of the expected Poisson statistics, the rate fluctuated wildly from day to day and had a strong seasonal dependence. Exploratory analysis involving more than 30 time series of possible explanatory variables showed that one of the more promising was the U.S. Air Force's solar radio noon flux measured by their radio solar telescope network (RSTN) [21]. This was surprising because the peak noon flux was under 100 solar flux units (SFU) (1 SFU $= 10^{-22}$W/m^2-Hz) and the average ∼50 SFU, several decibels below the thermal noise level of an ideal receiver and approximately 40 dB below typical signal levels. More information and references are given in sections 10 and 19 of [8]. Figure 3 shows 590 days of the percentage of calls that fail during the afternoon rush hour in a typical midwestern U.S. AMPS (analog FM) system during the 1996–1997 solar minimum. The data is for the busiest hour, typically 6:00 p.m. on workdays, 2:00 or 3:00 p.m. on weekends, and this switching may reduce the coherence slightly. The data were truncated at 590 days to avoid confusion due to a change in system parameters. Solar radio data were not available at the 900 MHz band used by the cellular phone system so, as a compromise, the 606 and 1415 MHz noon fluxes were averaged. There are, as usual, a few missing data in both series, and these were interpolated before averaging. As shown in Appendix A of [6],

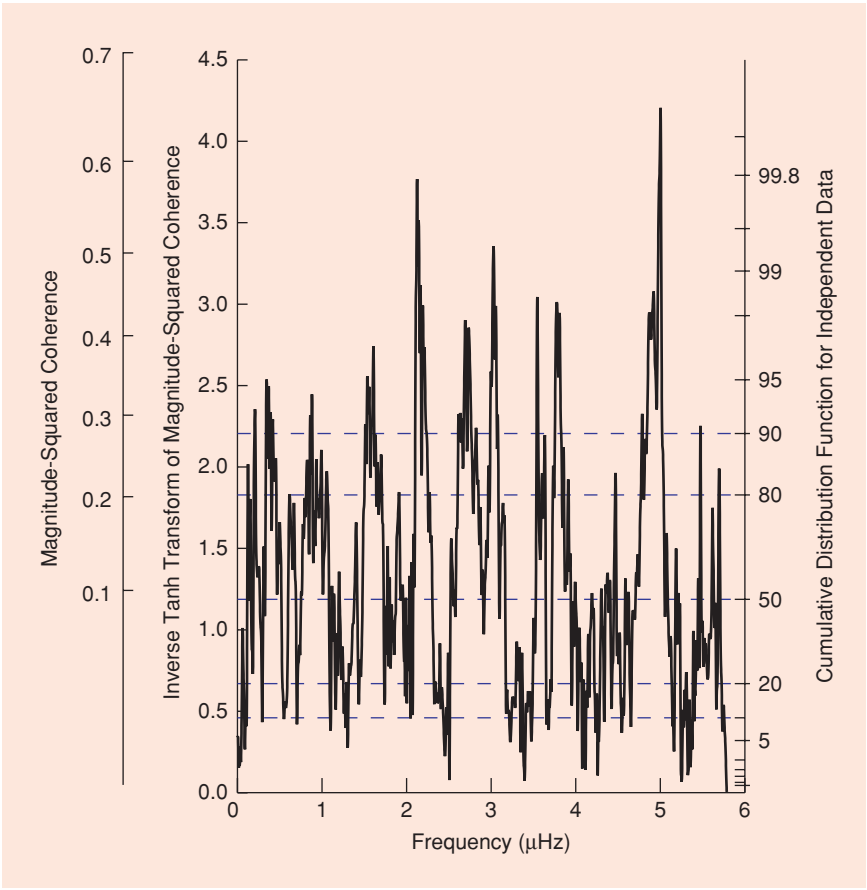
interpolating 5% of the sample does not change the spectrum very much but may reduce high coherences slightly. Figure 4 shows an estimate of the magnitude-squared-coherence (MSC) between the two series using $C_o = 5$ and $K = 8$. The very skewed distributions of sample MSC can be transformed to approximately Gaussian by use of a $\tanh^{-1}$ transform, see [4] for details and references and [22] for a second example. I use $\sqrt{2K-2}\tanh^{-1}(|c_{\setminus j}(f)|)$ transforms of the delete-one coherence estimates

$$c_{\setminus j}(f) = \frac{\sum_{\substack{k=0\\k\neq j}}^{K-1} \hat{x}_k(f)\hat{y}_k^*(f)}{\left[\sum_{\substack{k=0\\k\neq j}}^{K-1} |\hat{x}_k(f)|^2 \sum_{\substack{k=0\\k\neq j}}^{K-1} |\hat{y}_k(f)|^2\right]^{\frac{1}{2}}}, \qquad (24)$$

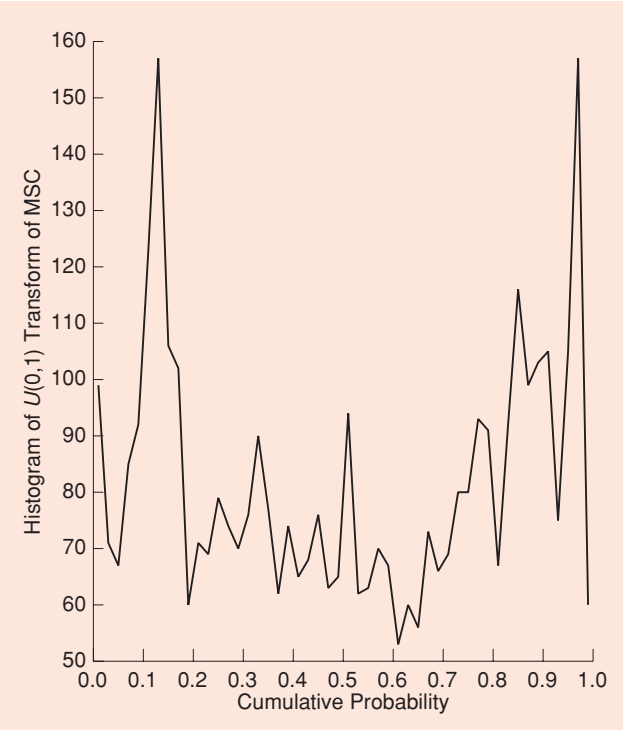
where $\hat{x}_k(f)$ and $\hat{y}_k(f)$ are the weighted eigencoefficients of the two series. The scaling is such that, if the data were ideal (stationary, Gaussian, with all spectral details resolved), the variance of the estimate should be 1.0. Here, the average jackknife variance is 1.0034, very close to expected. The minimum and maximum, however, are 0.058 and 7.2, respectively, somewhat unusual. It may also be seen that the plot appears to be bimodal, i.e., the MSC is either very low or very high, and we find that 91.1% of the estimates are above the nominal 10% level but 13.4% are above the 90% level. This is examined more closely in connection with Figure 5. To see the detailed structure, Figure 6 shows the normal ($\tanh^{-1}$) transform of MSC and the jackknife estimate of its variance between 4–5.6 μHz. The variance has a very spiky appearance, i.e., the variance is either excessively high or low. This seems to be very common with such data, and I attribute it to three causes: first, at frequencies where there is an isolated high-Q mode, the data is essentially deterministic and the variance is low; second, at frequencies where two or more modes are interfering, the variability is high; and third, we have a finite data segment and are attempting to achieve high frequency resolution (the bandwidth of the estimate is 100 nHz) resulting in only $\sim$16 DoF, so the estimate is naturally fairly variable. This figure shows the ratio of the normal MSC to the jackknife standard deviation and a comparison with frequencies observed on the Ulysses spacecraft [5]. When these frequencies were first encountered in a study of outages on communications satellites, they superficially appeared to be harmonics of solar rotation. Closer examination [9] showed that they were not harmonics. Indeed, had we not had jackknife error estimates on these frequency estimates, the presence of



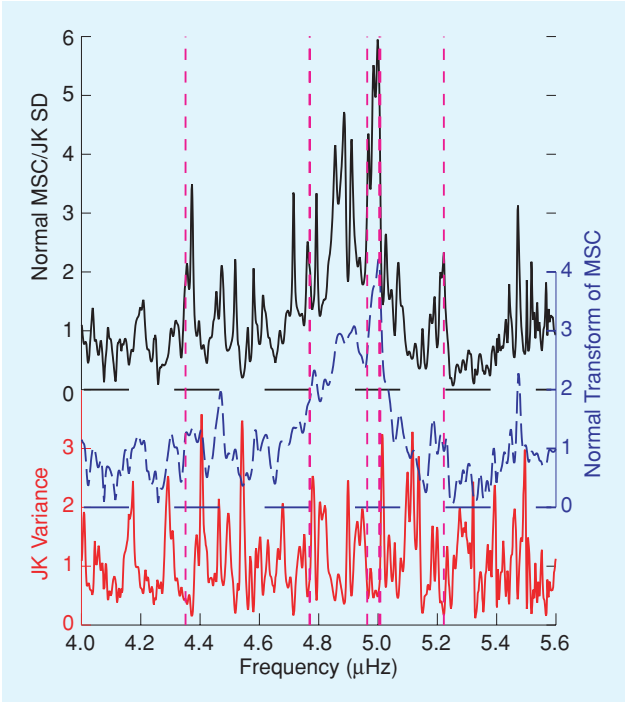
[FIG3] Data used in the dropped call study. The upper, thick, black curve shows the average of the 606 and 1415 MHz noon flux measured at Palehua, HI on the left-hand scale. Some missing data was interpolated in both series before averaging. The lower, red curve shows the rush-hour dropped call rate in percent for an advanced mobile phone service (AMPS) system operating in the U.S. Midwest on the right-hand scale.



[FIG4] Magnitude-squared-coherence (MSC) between the rush-hour dropped call rate and the Palehua noon flux. The three axes show MSC in conventional units; a $\tanh^{-1}$ transform of MSC in approximate $N(0, 1)$ units; and, on the right, the cumulative distribution for independent data.



[FIG5] **A histogram of the MSC estimates (shown in Figure 4) across frequency. For a null hypothesis of independence, the MSC estimates have been converted to $U(0, 1)$ by a probability integral transform. The bimodal character is obvious.**



[FIG6] **Transformed MSC, jackknife variance, and the ratio of transformed MSC to jackknife standard deviation between 4.0 and 5.6 μHz. The center blue dashed curve shows the normal transform of MSC from Figure 4 and the bottom red line the jackknife variance estimate. The ratio of the normal MSC to the jackknife standard deviation is shown by the top black line. The average is about 1.27σ with one peak at 6σ. The red dashed vertical lines are from [5, Table 2] or their aliases.**

discrete frequencies in the solar wind would likely have gone unnoticed. Similarly, had it not been for the accident of studying both the space physics and dropped call data nearly simultaneously, the agreement between the frequencies in the coherence and those seen on Ulysses and Voyager would probably have been missed. These Ulysses frequencies, or their aliases, are marked by the vertical dashed lines in Figure 6. Not all peaks are marked because Table 2 in [5] only listed 35 out of the many thousands of modes that are predicted to exist in this frequency range and because Ulysses is in an almost sidereal orbit, frequencies will disagree by $\approx(32\ m)$ nHz from those observed on Earth, where m is the longitudinal quantum number from the spherical harmonic expansion of the mode on the Sun.

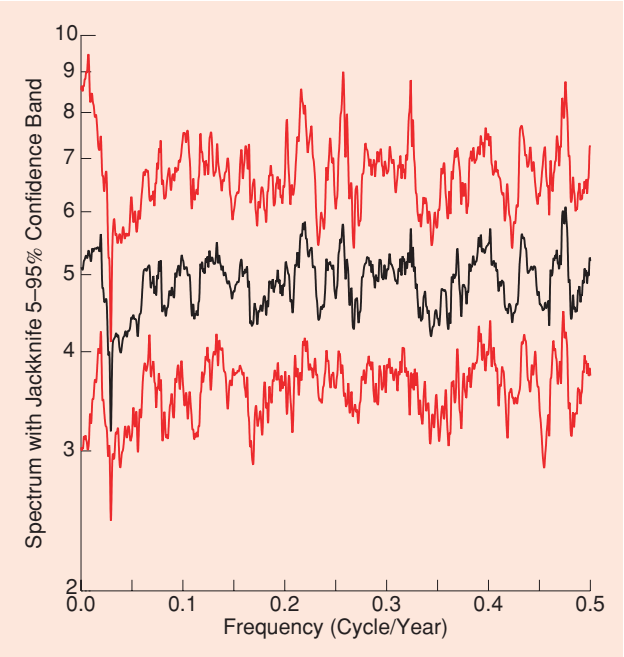
Returning to the bimodal distribution of coherences, Figure 5 is a histogram of the MSC in standardized units. The individual MSCs $|c|^2$ were transformed to a uniform $U(0, 1)$ distribution by a probability integral transform

$$u = 1 - (1 - |c|^2)^{K-1}, \tag{25}$$

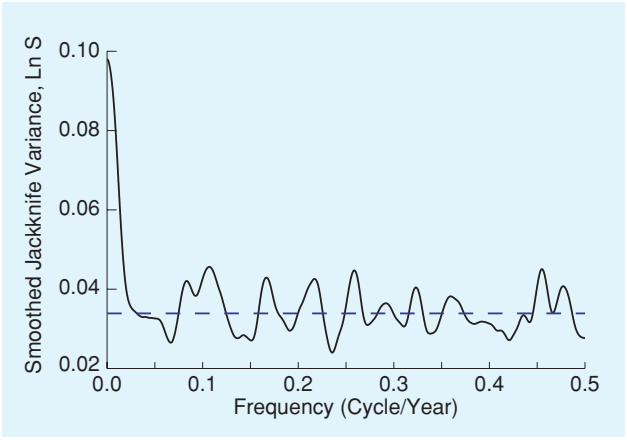
the cumulative distribution function of MSCs estimated between independent processes [4]. The u's were binned. If the dropped calls and solar data were independent, this histogram should approximate a uniform distribution, which it clearly does not. The conclusion is that the coherence is high if one is at a frequency corresponding to a solar mode and low at frequencies between modes. An explanation of the peak at low coherence levels is less certain, but one possibility is the presence of solar modes in the RSTN data that have little effect on cell phones, possibly by having the wrong polarization. A moderately strong component in one series of a coherence estimate and not the other suppresses estimates below that expected if the series were independent. The phase of the coherence is complicated, but remembering that the solar data is in fact noon flux (made at local noon to have a vertical path through the atmosphere) and, similarly, that the dropped call data is for the afternoon rush hour, one should expect the presence of aliases. I note that it is useful to plot the phase only at frequencies where the MSC has a significant local maximum because in problems where the coherence has a bimodal distribution, the phase between modes wanders randomly and causes more confusion than understanding, which is unsuitable for a tutorial paper.

THE NILE DATE

There have been numerous papers published in the last few years on so-called long-memory processes. These are processes supposedly having a fractional power-law spectrum near the origin, i.e., a spectrum of the form $S(f) \propto f^{-\alpha}$ with $0 < \alpha < 1$. The autocorrelations of such processes decay very slowly, and because their spectral representations are correlated on three lines, they are an intermediate form between stationary and



[FIG7] Prewhitened spectrum of the Nile River annual minima with jackknife 5% and 95% confidence limits. The raw data, minus the average, was prewhitened with an AR-20 prediction error filter and a time-bandwidth product $C_o = 16$ with $K = 30$ windows was used here.



[FIG8] Smoothed jackknife variance corresponding to the spectrum shown in Figure 7. The horizontal dashed blue line shows the expected value, and one notes significant departures from it.

nonstationary processes [23]. Moreover, such processes seem to be inevitably described for frequencies around the origin without frequency translation, which is very odd from an engineering perspective.

The evidence for such a physically unreasonable claim is often nothing more than an unwindowed periodogram, known for over a century to be an inconsistent estimate of the spectrum.

Given this, it is of interest to see what the jackknife and multitaper estimates can show. For this I chose a classic example, the 663 years of recorded minimum levels of the Nile River. Farm taxes were based partly on the level of the annual flood, and a continuous series from 622–1281, inclusive, is available [24].

Figure 7 shows a low resolution spectrum of the complete series after prewhitening with an AR-20 prediction error filter. The prewhitening reduces the range of the spectrum from almost 1,000 to $\lesssim 2$ and has three major effects: first, like most AR estimates, it shows the overall shape of the spectrum but almost nothing of interest; second, the adaptive weights for the prewhitened data are all close to unity; and third, one can distinguish the three lines in Figure 7 more easily than one can when the full range must be plotted. Here, $C_o = 16$ and $K = 30$ windows were used. The adaptive weighting gave a minimum of 59.84 DoF and an average jackknife variance on $\ln \hat{S}$ of 0.03587, close to the expected $\psi'(30) \approx 0.03398$. The maximum, however, is 0.1649 and correlation coefficient between $\widehat{V_{jk}}$ and $\hat{S}$ is 0.194, highly significant at this sample size. Figure 8 also shows a very high jackknife variance near the origin and also numerous smaller but suggestive features. Long-memory processes have a spectrum concentrated near the origin, thus, if I were not a natural skeptic, I would stop here and say, "case proven." As emphasized above, my opinion is that peculiarities in jackknife variances and other statistics show that the process needs to be examined more closely, and this is done in the next subsection.

TEST FOR LONG MEMORY

A possible test for long memory is to compute the canonical coherences (CCs) between parts of the process separated in time [25]. It is known that CCs define the mutual information between processes, and if a process is truly long memory, information must persist. Begin by defining the eigencoefficients beginning at time t, centered on frequency f, over a block of N samples

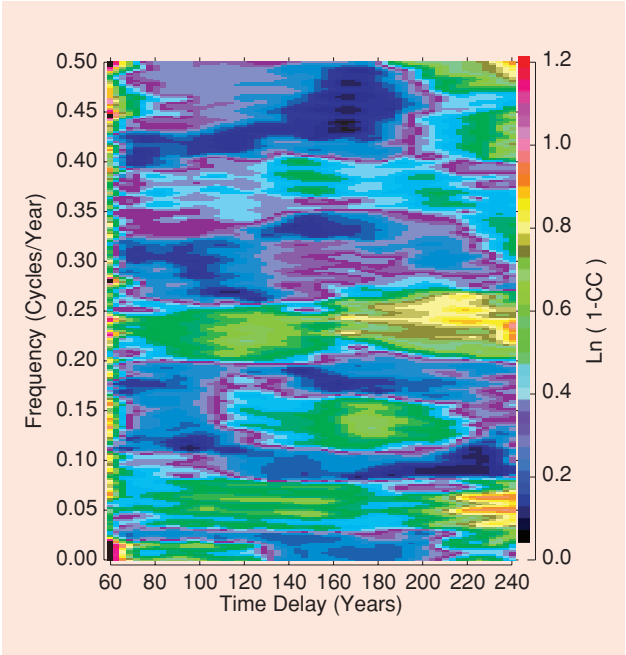
$$y_k(t, f) = \sum_{n=0}^{N-1} x(t+n) v_n^{(k)} e^{-i2\pi nf}. \tag{26}$$

These may be inverted to recover the data, so they contain all the information for the time-frequency region and the CCs between the eigencoefficients at time t and those at $t+\tau$, a measure of the mutual information. To estimate the CCs, define the matrices

$$\mathrm{G}(f, \tau) = \begin{bmatrix} y_0(1+\tau, f) & \cdots & y_{K-1}(1+\tau, f) \\ y_0(2+\tau, f) & \cdots & y_{K-1}(2+\tau, f) \\ \vdots & \vdots & \\ y_0(J+\tau, f) & \cdots & y_{K-1}(J+\tau, f) \end{bmatrix}, \tag{27}$$

where J denotes the number of time blocks and compute the CCs, denoted $C(f, \tau)$, between $\mathrm{G}(f, 0)$ and $\mathrm{G}(f, \tau)$ [25].

Figure 9 shows $\ln(1 - C(f, \tau))$ computed using $c_0 = 3.5$, $K = 4$, $N = 90$-year blocks with a block offset of three years. The first 60 years are omitted because, with 90-year blocks, the overlapping blocks are necessarily highly coherent. The striking



[FIG9] Canonical coherences between the eigencoefficients at time t and times $t+\tau$ for the Nile River data.

feature of this plot is that features at several frequencies persist and that frequencies near the origin are less persistent than some of the others. These frequencies and the delays where high coherence is seen are suggestive: 0.05 c/y of the Lunar 18.6-year nodal period and the Solar 22-year Hale magnetic cycle. Delays near 78 years suggest the Gleisberg cycle; and the 104, 208, and 230 years are the Suess frequencies seen in ^{14}C [15]. The pattern at low frequencies suggests the 78-year Gleisberg and 104-year Suess cycles with the 208- and 230-year Suess cycles predominating.

Doing a harmonic F-test over the full record gives the results seen in Figure 10. This test [3], [14], [15] minimizes the squared error between the eigencoefficients, the $y_k(f)$ from (6), and those expected from a line component. If we have an isolated periodic component with amplitude μ at some frequency f_o, the expected values of the eigencoefficients are

$$\mathrm{E}\{y_k(f_o)\} = \mu V_k(0), \tag{28}$$

where $V_k(f)$, the kth Slepian function, is the Fourier transform of the corresponding Slepian sequence. Usually, both the complex amplitude μ and the line frequency f_o are unknown. For a given frequency f, we estimate the amplitude by ordinary least-squares regression and tests for the significance of the regression. Specifically, w choose μ to minimize the residual sum-of-squares

$$r^2(f,\mu) = \sum_{k=0}^{K-1} |y_k(f) - \mu(f) V_k(0)|^2 , \tag{29}$$

so that

$$\widehat{\mu}(f) = \frac{\sum_{k=0}^{K-1} y_k(f) V_k(0)}{\sum_{k=0}^{K-1} V_k^2(0)} \tag{30}$$

and test the hypothesis with an F-test, the ratio of the energy explained by assuming a line component to the residual energy. Standardizing to have an F-distribution with 2 and $2K-2$ DoF, this gives

$$F(f) = \frac{\frac{1}{2}|\hat{\mu}(f)|^2 \sum_{k=0}^{K-1} |V_k(0)|^2}{\frac{1}{2K-2} r^2(f, \hat{\mu}(f))} . \tag{31}$$

Next, I choose the frequency where $F(f)$ is maximum as an estimate of f_o with $\widehat{\mu}(\hat{f}_o)$ from (30) giving the amplitude and phase. Under what John Tukey described as over-Utopian conditions (stationary Gaussian noise, a locally flat spectrum, and a constant amplitude and frequency sinusoid) some tedious algebra shows that this estimator approaches the Cramér-Rao bound for frequency estimates. The variance of the estimate depends on the signal-to-noise ratio that, in practice, is rarely known, but can be estimated by $F(\hat{f}_o)$. (The variance is larger than the bound by the efficiency factor Ξ_K^{-1} given in section VII of [3], typically by 5–20%.) Moreover, with the Nile data, we do not know what we have except for the single realization. This may well contain outliers and errors in dates and is almost certainly neither stationary nor Gaussian. Consequently, we attempt to estimate errors by jackknifing over windows. This is done by cycling through (30) and (31), withholding each eigencoefficient in turn, and doing the standard jackknife variance estimates. I have also found it useful to keep quantities such as the minimum and maximum of $\hat{f}_{o\setminus j}$. In doing this, it is necessary to zero-pad the data so the fast Fourier transform (FFT) mesh frequencies have a finer spacing than the expected standard deviation. I use moderate zero-padding followed by a slow Fourier transform to refine peak frequencies. Also, although it was assumed above that the periodic terms are isolated (i.e., separated from neighboring lines by a bandwidth of W or more), in practice, the F-test seems to work reasonably for separations down to one or two Rayleigh resolutions provided that the lines have roughly equal amplitudes.

Returning to the Nile data, I find six low frequency lines where the F-test is above the 95% significance level, five of them corresponding to named periods within the estimation accuracy. The jackknife errors on these frequency estimates are larger than expected and are approximately equal to the Rayleigh resolution, $1/T$. This, together with what appear to be systematic departures from known periods, suggests possible errors in the time scale. The estimated periods are 712 years, the first harmonic of the Bond cycle; the 238-year Suess; 75-year Gleisberg; 21-year Hale; and 18.3-year Lunar Node. The accepted periods are 735, 231, 78, ~22, and 18.6, years, respectively.

Thus, the evidence for long memory seems to be insignificant relative to that for Newtonian dynamics and Solar influences on climate.

NOTES ON THE JACKKNIFE

Despite the fact that Quenouille's discovery of the jackknife was in a time series application, most of the theoretical development

and applications of resampling techniques assume independent, identically distributed (i.i.d.) data. Of the methods that have been developed for time series problems (see e.g., chapter 9 of [26] or [27]), many rely on the assumption that either the residuals of low-order parametric models or the estimates computed on different data segments are approximately i.i.d. There may be interesting data for which these assumptions hold, but based on the data I have worked with, they are not common. These data come from problems in engineering [8], [28], climate [7], [14], [15], [22], space physics [5], [6], seismology [29], [30], astrophysics, medical imaging, and numerous other sources. Much of these data either include dozens to thousands of modes or have a large dynamic range and usually have both.

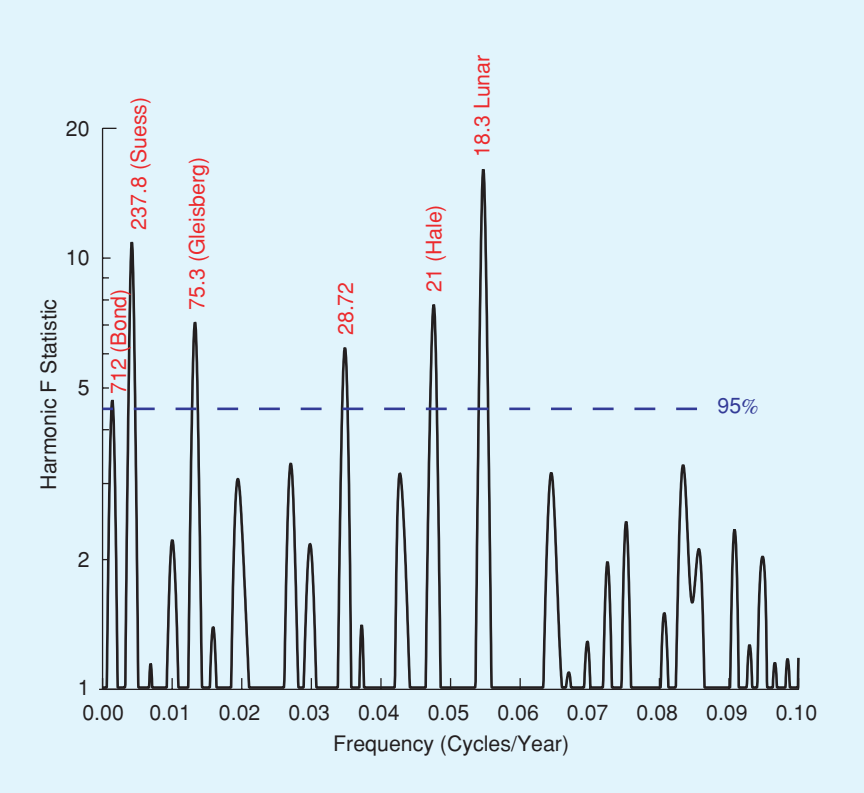
THE JACKKNIFE VARIANCE ESTIMATE IS CONSERVATIVE, AND UNDER A SET OF CONDITIONS THAT APPEAR TO BE REASONABLE FOR MOST TIME-SERIES DATA ENCOUNTERED IN THE PHYSICAL SCIENCES, JACKKNIFED MAXIMUM LIKELIHOOD ESTIMATES BEHAVE PROPERLY.

I have used the jackknife in preference to the bootstrap for four reasons.

- Some computational simplicity is important when large data sets are involved. The barometer data set used in the example in this article has 414,720 samples and is relatively small compared to the full data sets that have durations of many years. In a multitaper context, the number of data points in the time domain is irrelevant; the jackknife is applied over the K nominally uncorrelated eigencoefficients. In addition, many jackknife procedures may be done by a simple one-step downdating procedure.
- Bootstrap estimates rely heavily on random number generators, and experience has shown that even good random number generators cause unexpected artifacts in spectrum estimation problems.
- It is my hope that the statistical techniques I develop will be useful tools for scientists and engineers. The idea of random weighting used in bootstrapping often gets a skeptical reception from such people.
- Finally, the most important reason is Tukey's original motivation of having a rough-and-ready tool that can usually give a warning when the analysis is in trouble.

In many scientific and engineering problems, the ability to separate periodic or other defined narrow-band signals from the nondeterministic parts of the signal is crucial. It should be noted that the parametric models that appear to currently dominate the signal processing and statistical literature on time series are simply inadequate for these chores. As a specific example, solar and space physics data may contain several thousand periodic terms corresponding to normal modes of the sun. [Harvey (see [8, section 4]) estimates that there are 10 million solar modes.] Other problems aside, the frequencies could not be estimated from an AR model because of numerical problems alone. Note, however, that if a multitaper complex demodulation [7] is used to reduce the dimensionality of the problem to $2NW$ from N, then parametric methods are useful.

Data with many periodic or approximately periodic components are not restricted to problems involving the sun. Terrestrial seismology is nearly as complicated, and trying to count the number of gears, bearings, and rollers in manufacturing equipment that can leave artifacts in products such as optical fibers can be dismaying.



[FIG10] A Harmonic F-test for periodic components done over the full Nile series using $C_o = 4$ and $K = 5$ giving a test with two and eight DoF. Of the six lines where the significance exceeds the 95% level (shown by the horizontal dashed blue line), five are close to named periods.

SUMMARY

This article has summarized some theory on multitaper estimates, the jackknife, and their use together, and given some examples where the jackknife indicates problems with the basic estimate. These examples all use jackknifing over windows, as opposed to exchanging data blocks or residuals from parametric models. Indeed, in most of the problems I have encountered, spectra with both large ranges and many line components appear to be the rule rather than the exception, and it is not clear how any other method could work.

In spectrum estimation problems, the jackknife is rarely an end in itself, but when studying scientific and engineering data where the basic inferences depend on having accurate estimates of spectra or descriptions of the data under study, it is an invaluable diagnostic of possible problems. In such problems, one rarely cares if a particular statistic is significant at the 94% or 96% level, but a variance estimate that is ten or 100 times too large or too small requires serious attention. Thus, when the jackknife variance has a suspicious average or is either extremely low or extremely high at some frequencies, exploratory data analysis is mandatory.

ACKNOWLEDGMENTS

I thank Maja-Lisa Thomson for her helpful comments and Rudolf Widmer-Schnidrig of the Black Forest Observatory for the barometer data. Over the years my work in this area has benefited from discussions with Alan Chave and Frank Vernon. The Canada Research Chair program has provided funding for this work.

AUTHOR

David J. Thomson (djt@mast.queensu.ca) graduated from Acadia University and has a Ph.D. in electrical engineering from Polytechnic Institute of Brooklyn (now PUNY). He retired as a distinguished member of Technical Staff from the Communications Analysis Research Department of Bell Labs, Murray Hill, New Jersey in 2001, having previously worked on the WT4 Millimeter Waveguide System and the Advanced Mobile Phone Service projects. He is a Fellow of the IEEE and a member of the American Geophysical Union and the American Statistical Association. He is a professional statistician in the Statistical Society of Canada and a chartered statistician in the Royal Statistical Society. He has been chair of Commission C of USNC-URSI and associate editor for *Radio Science* and *IEEE Transactions on Information Theory*. He was a Green Scholar at Scripp's Institution of Oceanography and a Houghton lecturer at Massachussettes Institute of Technology. In 2002, he became a Canada research chair in statistics and signal processing in the Department of Mathematics and Statistics of Queen's University, Kingston, Ontario, and is a P.Eng. in Ontario.

REFERENCES

[1] R.G. Miller, Jr., "A trustworthy jackknife," *Ann. Math. Statist.*, vol. 35, no. 4, pp. 1594–1605, 1964.

[2] M. Quenouille, "Approximate tests of correlation in time series," *J. Roy. Stat. Soc., Ser. B*, vol. 11, no. 1, pp. 68–84, 1949.

[3] D.J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[4] D.J. Thomson and A.D. Chave, "Jackknifed error estimates for spectra, coherences, and transfer functions," in *Advances in Spectrum Analysis and Array Processing*, Simon Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1991, vol. 1, ch. 2, pp. 58–113.

[5] D.J. Thomson, C.G. Maclennan, and L.J. Lanzerotti, "Propagation of solar oscillations through the interplanetary medium," *Nature*, vol. 376, no. 6536, pp. 139-144, 1995.

[6] D.J. Thomson, L.J. Lanzerotti, and C.G. Maclennan, "The interplanetary magnetic field: Statistical properties and discrete modes," *J. Geophys. Res.*, vol. 106, no. A8, pp. 15,941–15,962, 2001.

[7] D.J. Thomson, "Multitaper analysis of nonstationary and nonlinear time series data," in *Nonlinear and Nonstationary Signal Processing*, W. Fitzgerald, R. Smith, A. Walden, and P. Young, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2001, pp. 317–394. .

[8] D.J. Thomson, L.J. Lanzerotti, F.L. Vernon, III, M.R. Lessard, and L.T.P. Smith, "Solar modal structure of the engineering environment," *Proc. IEEE*, vol. 95, no. 5, 2007.

[9] D.J. Thomson, C.G. Maclennan, and L.J. Lanzerotti, "Recurrences of interplanetary interaction regions at southern solar latitudes and approximate harmonics," *Adv. Space Res.*, vol. 20, no. 1, pp. 103–106, 1997.

[10] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: SIAM, 1982.

[11] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[12] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*. London, U.K.: Chapman & Hall, 1998.

[13] J.A. Reeds, "Jackknifing maximum likelihood estimates," *Ann. Statist.*, vol. 6, no. 4, pp. 727–739, 1978.

[14] D.J. Thomson, "Quadratic-inverse spectrum estimates: applications to paleoclimatology," *Philos. Trans. R. Soc. London A, Math. Phys. Sci.* vol. 332, no. 1627, pp. 539–597, 1990.

[15] D.J. Thomson, "Time series analysis of holocene climate data," *Philos. Trans. R. Soc. London A, Math. Phys. Sci.,* vol. 330, no. 1615, pp. 601–616, 1990.

[16] D.B. Percival and A.T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

[17] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1429, 1978.

[18] J.W. Tukey, "An introduction to the calculations of numerical spectrum analysis," in *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 25–46.

[19] P. Stoica and T. Sundlin, "On nonparametric spectral estimation," *Circuits Syst. Signal Process.*, vol. 18, no. 2, pp. 169–181, 1999.

[20] D.J. Thomson, "Jackknifing multiple-window spectra," in *Proc. ICASSP*, 1994, vol. VI, pp. 73–76.

[21] J.P. Castelli, J. Aarons, D.A. Guidice, and R.M. Straka, "The solar radio patrol network of the USAF and its application," *Proc. IEEE*, vol. 61, no. 9, pp. 1307–1312, 1973.

[22] C. Kuo, C. Lindberg, and D.J. Thomson, "Coherence established between atmospheric carbon dioxide and global temperature," *Nature*, vol. 343, no. 6260, pp. 709–714, 1990.

[23] T.O. Øigård, A. Hanssen, and L.L. Scharf, "Spectral correlations of fractional Brownian motion," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, pp. 031114-1–031114-6, 2006.

[24] J. Beran, *Statistics for Long-Memory Processes*. London, U.K.: Chapman & Hall, 1994.

[25] D.J. Thomson, "A test for 'long-memory' processes," in *Proc. 2003 IEEE Workshop Statistical Signal Processing*, St. Louis, MO, 2003, pp. 541–544.

[26] S.N. Lahiri, *Resampling Methods for Dependent Data*. New York: Springer-Verlag, 2003.

[27] H.R. Künsch, "The jackknife and bootstrap for general stationary observations," *Ann. Statist.*, vol. 17, no. 3, pp. 1217–1241, 1989.

[28] D.J. Thomson, "Spectrum estimation techniques for characterization and development of WT4 waveguide," *Bell Syst. Tech. J.*, vol. 56, pp. Part I, 1769–1815, no. 9, Part II, 1983–2005, no. 10, 1977.

[29] J. Park, C.R. Lindberg, and D.J. Thomson, "Multiple-taper spectral analysis of terrestial free oscillations: Part I," *Geophys. J. R. Astron. Soc.*, vol. 91, no. 3, pp. 755–794, 1987.

[30] F. Vernon, R.J. Mellors, and D.J. Thomson, "Broadband signal enhancement of seismic array data: Application to long-period surface waves and high-frequency wavefields," in *Proc. 17th Seismic Research Symp. Monitoring Comprehensive Test Ban Treaty*, 1995, pp. 807–814. SP