THE ASTROPHYSICAL JOURNAL, 263:835–853, 1982 December 15

STUDIES IN ASTRONOMICAL TIME SERIES ANALYSIS. II. STATISTICAL ASPECTS OF SPECTRAL ANALYSIS OF UNEVENLY SPACED DATA

JEFFREY D. SCARGLE
Theoretical and Planetary Studies Branch, Space Science Division, Ames Research Center, NASA
Received 1982 January 11; accepted 1982 April 26

ABSTRACT

Detection of a periodic signal hidden in noise is frequently a goal in astronomical data analysis. This paper does not introduce a new detection technique, but instead studies the reliability and efficiency of detection with the most commonly used technique, *the periodogram*, in the case where the observation times are unevenly spaced. This choice was made because, of the methods in current use, it appears to have the simplest statistical behavior. A modification of the classical definition of the periodogram is necessary in order to retain the simple statistical behavior of the evenly spaced case. With this modification, periodogram analysis and least-squares fitting of sine waves to the data are exactly equivalent. Certain difficulties with the use of the periodogram are less important than commonly believed in the case of detection of strictly periodic signals. In addition, the standard method for mitigating these difficulties (tapering) can be used just as well if the sampling is uneven. An analysis of the statistical significance of signal detections is presented, with examples.

Subject heading: numerical methods

I. THE SIGNAL DETECTION PROBLEM

Time series analysis has long been important in astronomy, but the application of automatic data acquisition systems has recently emphasized the need for this kind of mathematics. The first paper of this series (Scargle 1981, hereafter Paper I) reviewed astronomical time series in general and developed time domain analysis techniques for random phenomena. The opposite case, namely deterministic processes, leads naturally into the opposite (i.e., the frequency) domain. This paper is concerned with the detection of a special kind of deterministic signal, namely those which are strictly periodic. While time-domain techniques are well suited for the random case, the frequency domain has advantages for periodic processes (see, e.g., Blackman and Tukey 1958; Jenkins and Watts 1968; or, for an astronomical flavor, Brault and White 1971).

This paper discusses a specific frequency-domain approach to the detection problem, namely estimation of the power spectrum by means of the periodogram. The treatment includes the common situation where the observation times are not evenly spaced. In § II*a* it is shown that the periodogram's infamous statistical difficulty is not severe if the signal is rigidly periodic. In addition, the same smoothing techniques used to improve statistical and leakage properties in the evenly spaced case can be used if the sampling is uneven. If the sample times are at the observer's disposal, there is a new degree of freedom in the tailoring of the shape of the spectral response function. Further, uneven spacing can be desirable if aliasing (the appearance of high-frequency signals in the low-frequency part of the spectrum) is an important problem (see Appendix D). This paper does not propose a new technique for the detection of periodic signals, but rather presents a statistical analysis of the standard workhorse technique, namely periodogram analysis. When an error in the classical definition of the periodogram for unevenly sampled data is corrected (§ II*b*), three results follow: (1) the statistical behavior of the periodogram for uneven spacing is essentially identical to that for the case of even spacing (Appendix A); (2) periodogram analysis is exactly equivalent to least-squares fitting of sinusoids to the data (Appendix C); and (3) time-translation invariance is retained (Appendix B). Section III is an elaboration of Groth's (1975) statistical analysis, and includes expressions for the false alarm rate and the detection efficiency of a signal detection scheme using either periodogram or least-squares analysis. In § IV it is shown how to maximize the efficiency of a detection scheme. Specific examples of all of these concepts appear in § V. Some of the statistical results are applied to the planetary detection problem in an accompanying paper (Black and Scargle 1982, hereafter BS).

The basic problem considered throughout this paper is: A physical variable X is measured at a set of times t_i; the resulting time series data, $\{X(t_i), i = 1,2,\ldots,N_0\}$, are

assumed to be the sum of a signal and random observational errors:

$$X_i = X(t_i) = X_s(t_i) + R(t_i). \tag{1}$$

Throughout this paper the signal will be taken to be strictly periodic. Because of the assumed additive relationship between the signal and the errors in measuring it, the latter is often called noise. We assume throughout that the errors at different times are independent; that is, $R(t_i)$ is statistically independent of $R(t_j)$ for i not equal to j. We also assume that $R(t_i)$ is normally distributed with zero mean and constant variance, σ_0^2. The problem then is to establish the existence of a signal, despite the presence of the noise. This is called *signal detection*. Additional problems are the estimation of the harmonic content (i.e., the amplitudes of the fundamental and its multiples) as well as the period of the signal.

Good procedures for solving this detection problem will obviously make use of the periodic nature of the expected signal. One example is folding and averaging the data with respect to various periods, and then examining the appearance of the resulting average curves. This examination is often subjective. An objective procedure related to folding is least-squares fitting of sine waves of various periods to the data (e.g., Barning 1963; Vanicek 1969, 1971; Lomb 1976; Faulkner 1977). Another approach is periodogram analysis (Schuster 1898; Bartlett 1950; Wehlau and Leung 1964; Gray and Desikachary 1973; Deeming 1975; Groth 1975; Faulkner 1977). Some authors have pointed out the close relationship of these two methods (Lomb 1976; Faulkner 1977; Meisel 1978, 1979). The present paper establishes, apparently for the first time, that (with the proposed modifications) these two methods are exactly equivalent.

Several other techniques should be mentioned for the sake of completeness. Lafler and Kinman (1965) describe a procedure which involves trial-period folding followed by a minimization of the differences between observations of adjacent phase. The procedure proposed by Feraz-Mello (1981) directly attacks the nonorthogonality of the basis functions when the sampling is uneven, using the Gram-Schmidt orthogonalization procedure. Kuhn (1982) discusses two procedures for recovering an approximation to the discrete Fourier transform. The omission of such approaches from this paper is not to be taken as a negative comment, but has been made simply because it appears that the statistical properties of these methods would be very difficult to unfold.

II. THE PERIODOGRAM

A basic tool of spectral analysis is the discrete Fourier transform (DFT) which can be defined for an arbitrarily sampled[1] data set, $\{X(t_i), i = 1,2,\ldots,N_0\}$, as

$$\mathrm{FT}_X(\omega) = \sum_{j=1}^{N_0} X(t_j) \exp(-i\omega t_j). \tag{2}$$

The periodogram is then conventionally defined as

$$\begin{aligned} P_X(\omega) &= \frac{1}{N_0} |\mathrm{FT}_X(\omega)|^2 \\ &= \frac{1}{N_0} \left| \sum_{j=1}^{N_0} X(t_j) \exp(-i\omega t_j) \right|^2 \\ &= \frac{1}{N_0} \left[\left(\sum_j X_j \cos \omega t_j \right)^2 + \left(\sum_j X_j \sin \omega t_j \right)^2 \right] \end{aligned} \tag{3}$$

(Schuster 1898; Thompson 1971; Deeming 1975). This function will be called the *classical periodogram*. It can be evaluated for any value of the frequency. The reason for using the periodogram is that if X contains a sinusoidal component of frequency ω_0, then at and near $\omega = \omega_0$ the factors $X(t)$ and $\exp(-i\omega t)$ are in phase and make a large contribution to the sums in equation (3). At other values of ω the terms in the sum are randomly positive and negative, and the resulting cancellation yields a small sum. Hence the presence of a sinusoid is indicated by a large value of P near one value of ω—i.e., as a distinct narrow peak in the spectrum.[2] If the observation times are evenly spaced, at interval Δt, it is customary to take $\Delta t = 1$, $t_j = j$, and $X_j = X(t_j)$, so that

$$P_X(\omega) = \frac{1}{N_0} \left| \sum_{j=1}^{N_0} X_j \exp(-ij\omega) \right|^2. \tag{4}$$

While this expression can also be evaluated at any frequency, it is traditionally evaluated only at a special set of

$$N = N_0/2 \tag{5}$$

evenly spaced frequencies (see Appendix D). That equation (4) can be quickly evaluated at frequencies (D1)

[1] The set of observation times, $\{t_i\}$, in deference to probability theory, is called the *sampling*. We encounter *even sampling* ($\Delta t_i = t_{i+1} - t_i$ = constant), and *uneven sampling* (arbitrary t_i's).

[2] We use the terms (*power*) *spectrum* and *periodogram* interchangeably, although strictly speaking the power spectrum is a theoretical quantity defined as an integral over continuous time, and of which the periodogram is merely an estimate based on a finite amount of discrete data.

with the fast Fourier transform (FFT) explains, in part, its popularity.

a) Critique of the Periodogram

Modern work (e.g., Richards 1967; Tukey 1967; Kay and Marple 1981) has moved away from equations (3) and (4) because of two problems: statistical difficulties and spectral leakage. The main statistical problem is that the function $P(\omega)$ is very noisy, even when the data are only slightly noisy. Moreover, the noise does not diminish in amplitude with increasing sample size. For if X is a normally distributed noise process, the relative variance $\sigma_P/\langle P\rangle$ is of order unity (Bartlett 1950 and references therein; Richards 1967), no matter how many data points there are. The reason is that as more data are added, the number of available frequencies increases in proportion (see eq. [5]), so the noise is not averaged out.

This noise problem has been repeatedly emphasized in the literature. Richards (1967) says that the periodogram, our equation (4), "is almost useless for practical computations, except in cases so simple as to be of little interest. ... Experience has repeatedly shown that, with a noisy signal...[4] gives a very erratic spectrum, $P(\omega)$, which fails to converge no matter how large T [the total time interval] is made or how small Δt is chosen." Tukey (1967) makes similar remarks, noting that *spectral* can refer to either spectra or spectres. He adds: "If we dealt with problems involving the superposition of a few simple periodic phenomena, as do astronomers interested in binary stars and related problems [that's us!], we can learn much from the periodogram."

What saves the periodogram in such problems is that, as more data are acquired, even though the size of the noise remains large, the signal-to-noise ratio (which is the relevant quantity) increases. For if the observed process is

$$X(t_i) = X_0 \sin(\omega_0 t_i + \phi) + R(t_i), \quad i = 1,2,\ldots,N_0, \tag{6}$$

it can be shown that the expected value of the power due to the signal, at the signal frequency, is

$$P_X = N_0(X_0/2)^2, \tag{7}$$

and that due to the observational errors is (the variance)

$$P_R = \langle R^2\rangle = \sigma_0^2. \tag{8}$$

Hence the signal-to-noise ratio,

$$P = P_X/P_R = N_0(X_0/2\sigma_0)^2, \tag{9}$$

increases proportionally to the number of samples, N_0. The explanation is that the power per unit bandwidth of a monochromatic signal, in the passband containing the signal frequency, increases because the bandwidth is a decreasing function of N_0, whereas the noise power is constant per unit bandwidth. For signals with a continuous spectrum[3] the same argument leads to a constant signal-to-noise ratio, the result alluded to above.

The second problem, *spectral leakage*, is simply that for a sinusoidal signal at a given frequency, ω_0, the power in the periodogram not only appears at ω_0, but also leaks to other frequencies. This problem is inherent to frequency analysis with a finite amount of data. There are several forms of spectral leakage. Leakage to nearby frequencies (sidelobes) is due to the finite total interval over which the data is sampled. Leakage to distant frequencies is due to the finite size of the interval between samples.

In particular, the well-known phenomenon of *aliasing* is a leakage of power from high frequencies to much lower frequencies. The way in which it arises makes it sensitive to a very precisely maintained evenness to the sampling (see Fig. 8.4 of Kanasewich 1975, for a graph which makes this point particularly well). Hence anything from a slight to a major unevenness in the spacing substantially reduces aliasing. The theoretical framework for this statement has been well established (e.g., Beutler 1966, 1970; Masry and Lui 1975; Higgins 1976; Wiley 1978; Gaster and Roberts 1975, 1977; Kar, Hornkohl, and Farmer 1981; Ludeman 1981). Error-free recovery of a band-limited signal [i.e., reproduction of the entire function $X(t)$ from the samples $X(t_i)$] can be achieved with irregular sampling as long as the mean sampling rate exceeds the Nyquist rate (i.e., the average number of samples per unit time must exceed twice the highest frequency component in the signal). Indeed, only the (infinite) past need be sampled at such a rate to ensure error-free recovery. Surprisingly, such recovery is possible in some cases even if the mean sampling rate is less than the Nyquist rate (Beutler 1966). Beutler (1970) also exhibits an example in which one-sided sampling (i.e., of the past only, or of the future only) at a mean rate which is an arbitrarily small fraction of the Nyquist rate produces alias-free recovery of the spectrum of the process! Another surprising result is that the spectrum of a process can be estimated even if the sampling times are not recorded—only the order of the samples need be retained (Beutler 1970). Because these results depend on sampling over an infinite time interval, their practical significance is unclear.

Beutler (1970) also discusses the effects on aliasing of two different perturbations from even sampling: ran-

[3] In general, random processes have continuous spectra, while deterministic (e.g., periodic) signals have discrete, or line, spectra. This explains, in part, why frequency-domain techniques are good for periodic signals, while time-domain ones are good for random signals.

dom independent deletion of samples (see also Sturrock and Shoub 1981) and *jittered* sampling, in which the sampling times are randomly perturbed about the evenly spaced values. Such jitter is common when electromechanical devices control the sampling, as in some forms of Fourier transform spectroscopy. See also Deeming (1975), although his examples are with semiregular sampling for which the reduction of the aliasing is less dramatic than it is, say, for random sampling.

Astronomical sampling is typically irregular enough that aliasing is effectively eliminated. However, there is sometimes an enforced regularity that produces distant sidelobes that are very similar in effect to conventional aliasing (see Fig. 2 of Deeming 1975, and our Fig. 3 below). The most common example is the yearly periodicity imposed by the influence of the Sun on the observations. Typically, even when such specific periodicities are not present, the sampling tends to be semiregular—i.e., intermediate between randomly and evenly spaced. The result is significant leakage of power into sidelobes (see Fig. 3 below), which can cause problems.

The canonical attack on both the statistical and leakage problems is the use of procedures which are all equivalent to smoothing in the spectral domain. One example is the multiplication of the data by a function which goes smoothly to zero at the ends of the sampling interval. This is called data *windowing* or *tapering*. Another example is the analogous treatment of the autocorrelation function. By elementary means it can be shown that these are both equivalent to convolving the spectrum with a *spectral window function* (Harris 1978). Such convolution reduces the variance because it averages (smooths) the spectrum. At the same time spectral leakage can be controlled, because the window function can be chosen (*tailored* or, even more colorfully, *carpentered*) so that the amplitudes of the sidelobes are reduced. Many different windows have been proposed, tested, and used. Harris (1978) presents graphs of the sidelobe suppression for about 45 data windows. What at first appears to be a different averaging procedure, *segmented averaging* (analyzing subsegments of the data separately and then averaging), is really equivalent to windowing the autocorrelation function (Richards 1967).

It is important to realize that all of these spectral smoothing techniques, although developed for evenly sampled data, can be readily applied to the periodogram with arbitrary sampling (see Thompson 1971). In the case of direct time-domain windowing, this will be outlined below (§ V*d*) in the section on window carpentry. A disadvantage of any such smoothing, shared by both the evenly and the unevenly sampled cases, is that the spectrum values at different frequencies are no longer independent, so that the joint statistical properties are more complicated.

A different approach to the leakage problem is to try to remove it from the spectrum. A variety of such techniques has been suggested (e.g., Wehlau and Leung 1964; Fitch and Wehlau 1965; Barning 1963; Gray and Desikachary 1973; Meisel 1978; Swan 1981), in both the time domain (where it is called *prewhitening*) and the frequency domain. Unfortunately, most of these deconvolution techniques are somewhat ill conditioned in the sense that the inevitable observational noise is amplified in the process.

In summary, the statistical and leakage problems are problems with the use of the *unsmoothed* periodogram, not with the use of the periodogram itself, and not with the extension to uneven spacing. In addition, the statistical problem is less severe than commonly believed when the signal is strictly periodic. The periodogram is not claimed to be the best tool for even this restricted problem, as I have not made a systematic comparison with other techniques. But the simplicity of the statistical behavior of the periodogram does make it useful when evaluation of the reliability of a possible detection is important.

b) A New Definition of the Periodogram

The statistical distribution of the periodogram (see § III for comments on the meaning of this concept) is simple and well known for the even-sampling case (e.g., Groth 1975). The most important result is that if X is pure Gaussian noise, P_X is exponentially distributed (Groth's eq. [13] with $n=1$). The analogous result for equation (3) does not seem to have been previously derived, and is given in Appendix A. Indeed, the statistical behavior of the spectral estimator in equation (3) as it stands is considerably more complicated than that of the even sampling periodogram in equation (4). However, a slightly modified version of the periodogram has the same exponential distribution as in the even-sampling case. This redefinition is

$$P_X(\omega)=\frac{1}{2}\left\{\frac{\left[\sum_j X_j\cos\omega(t_j-\tau)\right]^2}{\sum_j\cos^2\omega(t_j-\tau)}+\frac{\left[\sum_j X_j\sin\omega(t_j-\tau)\right]^2}{\sum_j\sin^2\omega(t_j-\tau)}\right\}, \quad (10)$$

where τ is defined by

$$\tan(2\omega\tau)=\left(\sum_j\sin 2\omega t_j\right)\Big/\left(\sum_j\cos 2\omega t_j\right). \quad (11)$$

The term *slightly modified* was used because the actual values are typically not changed much (see Figs. 4 and 5

below), even though the form is significantly changed. I propose to replace the classical definition of the periodogram, equation (3), with equation (10), which is preferable for two reasons: it has a simple statistical behavior (Appendix A), and is equivalent to the reduction of the sum of squares in least-squares fitting of sine waves to the data (Appendix C). Like the classical periodogram, it reduces to equation (4) if the spacing is even, and has time-translation invariance (Appendix B). The computation of (10) is not substantially more difficult than that of (3), so that even though the numerical differences are small, the theoretically preferable form should probably be used in all but the most casual applications.

III. STATISTICS OF THE PERIODOGRAM

I want to clearly define what is meant by the statistical behavior of the periodogram because it is such an important concept. X is often a random variable—e.g., pure noise, or noise plus a signal. Hence P_X, through any of equations (3), (4), or (10), is also a random variable. The statistical distribution of this random variable is important... so important that we sculptured the very definition of the periodogram to ensure simple statistical behavior. The basic point is: since the periodogram of noisy data is noisy (§ II*a*), *surprisingly large spurious spectral peaks can occur and be erroneously taken to indicate the presence of a periodic signal*. Hence it is important to critically analyze the statistical significance of a suspected spectral feature, by answering the question: "What is the probability that this feature could have arisen from chance (noise) fluctuations?"

a) The Distribution of $P(\omega)$

The answer to this last question is completely contained in the probability distribution of the random variable $P_X(\omega)$ for the case where X is pure noise. We now summarize the well-known results (following Groth 1975) for the evenly spaced case, keeping in mind that we have arranged in Appendix A for the same distribution to apply to the case of arbitrary sampling.

We emphasize the cumulative distribution function (CDF) for three reasons: (1) Estimates of the CDF can be constructed without binning of the data (see paper I). In contrast, estimates of the differential probability distribution, $p(z)$, invariably depend on the arbitrary selection of the bin sizes and locations. There is always some information loss in binning. (2) The CDF of the maximum of a set of random variables is equal to the product of the CDFs of the variable (Papoulis 1965, § 7.1, application 3). This feature is especially useful since we are usually interested not in the power at a known preselected frequency, but in the maximum power over a set of frequencies. (3) Most of the subsidiary quantities of interest here (thresholds, false alarm rates, missed signal probabilities) can be read off a graph of the CDF directly, in terms of values of the ordinate or abscissa (see Fig. 1 below). A disadvantage of the CDF is that the differences between various distributions tend to be washed out by the integration that leads to the CDF (eq. [13] below). Hence all CDFs tend to look alike, whereas the differential distributions are more distinctive (but correspondingly noisier).

The starting point of the statistical analysis is the simple but very useful result that the power at a given frequency is exponentially distributed (Appendix A). Letting $Z = P_X(\omega)$, we have for the probability distribution

$$p_Z(z)\,dz = \Pr(z < Z < z + dz) = \exp(-z)\,dz. \quad (12)$$

Here and henceforth the noise variance will be taken to be unity (i.e., P will be measured in units of σ_0^2). Hence the cumulative distribution function is

$$F_Z(z) = \Pr\{Z < z\} = \int_0^z p_Z(z')\,dz'$$

$$= 1 - \exp(-z). \quad (13)$$

A more useful quantity is $\Pr\{Z > z\} = \exp(-z)$, which gives the statistical significance of a large observed power at a preselected frequency. That is, as the observed power (really, the signal-power to noise-power ratio) becomes larger, it becomes exponentially unlikely that such a power level (or greater) can be due to a chance noise fluctuation.

Now consider the maximum value (peak) in the spectrum. Let $Z = \max_n P(\omega_n)$, where the maximum is over some set of N frequencies such that the $P(\omega_n)$ are independent random variables (see Appendix D). Then the multiplicative property mentioned above yields for this case

$$\Pr\{Z > z\} = 1 - F_Z(z)$$

$$= 1 - [1 - \exp(-z)]^N [Z = \max_n P(\omega_n)]. \quad (14)$$

This formula contains the *statistical penalty* for inspecting a large number of frequencies and selecting the largest value of P. For if N independent experiments are carried out, even if each one individually has a very small probability of succeeding, the chance of one of them succeeding is very large if N is large enough (approaching certainty as N approaches infinity). Indeed, a simple calculation shows that the expected value of the maximum of a pure noise spectrum, over a set of N frequencies at which the power is independent, is

$$\langle Z(\max) \rangle = \sum_{k=1}^{N} 1/k,$$

a series well known to diverge logarithmically with N. The lesson: If many frequencies are inspected for a spectral peak, expect to find a large peak power even if no signal is present.

If a signal is present, the distributions are different. Let $P_s(\omega)$ be the power which the signal alone (i.e., with the noise magically turned off) would produce, and denote the signal-to-noise ratio as $P = P_S/P_R$. Then the CDF of the observed signal-to-noise ratio Z is [Groth 1975, eq. (16) with $n=1$]

$$F_Z(z) = 1 - \exp[-(z+P)]\phi(z,P), \qquad (15)$$

where ϕ is an integral of a Bessel function, with the series expansion

$$\phi(x,y) = \sum_{m=0}^{\infty} \sum_{k=0}^{m} x^k y^m/(k!m!). \qquad (16)$$

If the signal is present at a single unknown frequency among a set of N frequencies, then the CDF of $Z = \max_n P(\omega_n)$ is

$$F_N(z) = [1-\exp(-z)]^{N-1} F_Z(z), \qquad (17)$$

where again it is assumed that the $P(\omega_n)$ are independent. The factors of $[1-\exp(-z)]$ come from the $N-1$ frequencies at which the signal is not present, while the other factor is from the signal frequency (which is, of course, unknown).

b) *The Joint Distribution of $P(\omega)$ and $P(\omega')$*

At the important matter of the mutual dependence of the $P(\omega_n)$, the evenly and unevenly sampled cases finally part company. If the data are evenly spaced, the $P(\omega_n)$, with the ω_n given by equation (D1) below, are strictly independent random variables. But this independence is lost if the sampling is uneven. The underlying mathematical reason is that, while the sine and cosine functions are orthogonal with respect to summation over evenly spaced time, this orthogonality disappears if the times are unevenly spaced. There appears to be no way to restore orthogonality without transforming to basis functions which mix various frequencies together. This mixing defeats the purpose of spectral analysis, and cannot be considered a viable procedure in the present context.

But luckily if the frequency grid is well chosen, the degree of dependence between the powers at the different frequencies is usually small. To see this, we need the correlation coefficient between $P(\omega)$ and $P(\omega')$ for arbitrary ω and ω'. Lomb (1976) calculated this quantity, and showed that it is equal to the window function $G(\omega)$, evaluated at $\omega - \omega'$. The window function, defined in Appendix D, is a very useful quantity: It contains all relevant information about dependencies and correlations. If $G(\omega)$ has a set of evenly spaced nulls at the frequencies $\omega_n = n\omega_1$, $n = 1,2,3,\ldots$, then the $P(\omega_n)$ are uncorrelated. [Note that the nulls must be evenly spaced in order for all of the $P(\omega_n)$ to be uncorrelated with each other.] This does not mean that they are independent, as independence is a stronger condition than uncorrelation (Paper I). But it will be assumed that if G is small or zero at a set of evenly spaced frequencies, ω_n, then the $P(\omega_n)$ are nearly independent. This assumption can be rigorously justified for Gaussian processes, because for them lack of correlation does imply independence. With a wide variety of sampling schemes $G(\omega)$ does have nulls, or relatively small minima, that are approximately evenly spaced (see the examples in Figs. 3 and 4). Such nulls comprise a set of natural frequencies at which to evaluate the periodogram. At these frequencies the $P(\omega)$ form a set of approximately independent random variables—thus closely simulating the situation with evenly spaced data.

c) *The False Alarm Probability*

It is desired to find a power level, z_0, such that if we claim the detection of a signal only if the observed power exceeds this level, we will be wrong (fooled by fluctuations) only a small fraction, say p_0, of the time. From the distribution in equation (14) this detection threshold is:

$$z_0 = -\ln\left[1-(1-p_0)^{1/N}\right], \qquad (18)$$

where N is the number of frequencies searched for the maximum. The false alarm probability, p_0, is a fixed small number (in examples we take $p_0 = 0.01$). Note that, for small p_0,

$$z_0 \approx \ln(N/p_0), \qquad (19)$$

$$= 4.6 + \ln(N) \quad \{\text{for } p_0 = 0.01\}. \qquad (20)$$

Thus with $N = 30$, Z must be greater than 8 to permit reporting a signal with 99% confidence. This signal-to-noise ratio seems very high, but is not as striking when converted to the amplitude signal-to-noise ratio (see eq. [9]), viz., $X_0/\sigma_0 \approx 1$.

d) *The Detection Efficiency*

From the discussion in the previous section it can be seen that only rather large signals can be detected reliably. If the power falls below the threshold, z_0, the signal will not be detected. The probability of thus missing a signal of power P is given by the correspond-

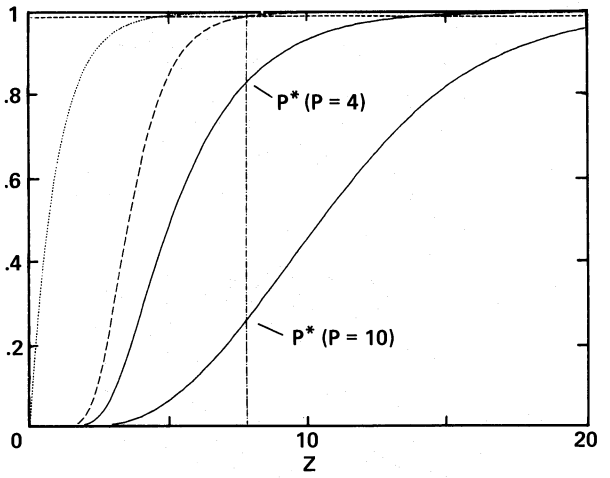

FIG. 1.—The dotted curve is the cumulative distribution function (CDF) for the power at a single, preselected frequency, in the no-signal case (eq. [13]). The point where this curve crosses $1-p_0$ (indicated by the horizontal dashed line) gives the value of z such that the probability of a noise fluctuation exceeding z is p_0. Similarly, the dashed curve is the CDF for the maximum over $N=25$ frequencies, again with no signal present (eq. [14]). The value of z where this curve reaches $1-p_0$ (p_0 is the desired small false alarm rate) is called z_0 (eq. [18]). The reason that this value of z, indicated in the figure by a vertical dot-dash line, is called the detection threshold is that signal powers above this threshold are spurious only a fraction p_0 of the time. The solid lines are CDFs for the maximum power when signals are present, the upper one with $P=4$ and the lower one with $P=10$. Since detection can be claimed only if P exceeds z_0, the probabilities of not detecting these signals are given by the corresponding ordinates of the CDF at z_0 (i.e. p^*; see eq. [22]).

ing CDF evaluated at the detection threshold:

$$p^*(N,P)=\Pr\{\text{miss}\}=F_N(z_0)$$
$$=(1-p_0)^{1-1/N}\{1-\exp[-(z_0+P)]\phi(z_0,P)\}, \quad (21)$$

where ϕ is given in equation (16), p_0 is a constant, and z_0 is given as a function of N in equation (18). The factor $(1-p_0)^{1-1/N}$ in this equation is a very slowly varying function, ranging monotonically from 1 to $1-p_0=0.99$ as N goes from 1 to ∞. Dropping this essentially constant factor yields

$$p^*(N,P)=1-\exp[-(z_0+P)]\phi(z_0,P). \quad (22)$$

This equation can be used to compare the detection efficiencies of two different observational schemes, in terms of their parameters N and P. The detection efficiency is defined as the probability of detecting a signal of power P, and is just

$$\text{DE}=1-p^*(N,P). \quad (23)$$

Note from equation (9) that P depends on not only the signal amplitude, X_0, but also the observational parameters σ_0 and N.

Figure 1 demonstrates the concepts of detection efficiency and false alarm rate, by showing how they relate to the CDF of the peak power with and without a signal present. One can see from the figure that the signal-to-noise ratio must be relatively large before the probability of missing the signal is low—especially if the chance of falsely claiming signal detection is desired to be very small.

IV. MAXIMIZATION OF THE DETECTION EFFICIENCY

One application of these results is the adjustment of the parameters to maximize the detection efficiency. In practice this is best done subject to a constraint that P have some functional relationship to N. For example, it may be true that the error variance σ_0^2 is proportional to N. Consider the case of a set of N_1 data points, with error variance σ_1^2, averaged[4] to form a set of $N_2<N_1$ points of greater accuracy ($\sigma_2<\sigma_1$). The assumption that the errors are independent and normally distributed yields

$$\sigma_2=(N_2/N_1)^{1/2}\sigma_1, \quad (24)$$

i.e., $\sigma\propto N^{1/2}$. With this relationship P is constant with N (see eq. [9]) and it is then easy to see that p^* is a monotonically increasing function of N, for $F_N(z_0)$ is monotonically increasing with z_0 (all CDFs are nondecreasing by virtue of the definition in eq. [13]). And from equation (18) one can see that z_0 is an increasing function of N. It follows, using equation (21), that p^* is an increasing function of N. Thus with $\sigma^2\propto N$ the missed signal rate is minimized by making N as small as possible. The main reason for this behavior is the statistical penalty discussed in § III*a*. Smaller N means that fewer trials are being performed, so that a spectral peak of a given size becomes more significant.

This result implies that one way to improve the detection efficiency is to heavily average the data, even to the point that $N=1$ (i.e., just two data points). This would be true but for one problem: averaging the data increases the effective sampling interval Δt, thus decreasing the Nyquist frequency. It is undesirable for the Nyquist frequency to fall below the signal frequency, as it can then be very difficult (but not impossible) to detect the signal. Lacking *a priori* knowledge about the signal frequency, one cannot average the data without running this risk. These considerations suggest trying a

[4]By this is meant that groups of data points are replaced by their mean values. It is important that the groups be nonoverlapping, so that the corresponding means are statistically independent. This makes the statistical properties very simple, in contrast to the situation when running (nonindependent) means are used.

range of degrees of averaging of the data. Unfortunately, a rigorous statistical analysis of such nonindependent trials would be difficult.

There is another important way that the detection efficiency can be improved, namely by decreasing the number of frequencies that are inspected. The relationship that the number of frequencies is half the number of data points, equation (5), only means that the maximum number of statistically independent $P(\omega)$'s is $N_0/2$. But one could choose on physical grounds to inspect a restricted number of frequencies, say many fewer than $N_0/2$. This would leave P unchanged, because it is determined by N_0 (eq. [9]), whereas z_0 would be reduced because it depends on N (eq. [18]). Hence, the detection efficiency derived using equation (22) is always improved if N is decreased below $N_0/2$. This should be done if and only if one is confident that no interesting signals could be present in the range of frequencies ignored—otherwise one would be discarding information and risking missing a real signal actually in the data. For a fixed N, it is easy to show that p is an increasing function of N_0/σ_0^2, and so is maximized (for fixed σ_0) by making N_0 as large as possible, and is constant if $N_0 \propto \sigma_0^2$. This confirms the intuitively compelling notion that the acquisition of more data should improve the detection efficiency.

It appears to be impossible to overemphasize that *choices of this kind must be made before the data are analyzed*—otherwise the statistical analysis of the results is completely changed.

Further, it is easy to see that if P is a nonincreasing function of N (and this includes the constant case discussed above), the p^* is an increasing function of N, and again N should be as small as possible. The remaining case, P an increasing function of N (or, if we are comparing two distinct schemes, $N_2 > N_1$ and $P_2 > P_1$), cannot be treated as easily, because the results depend on how strongly increasing the function is. Figure 2 shows contours of constant p^* in the P-N plane, based on approximations given in Appendix E. This figure verifies the above conclusions for the constant P case, and shows how strongly P must increase with N to reverse the situation and make p^* a decreasing function of N.

V. SAMPLE COMPUTATIONS

Sample computations are useful to demonstrate the theoretical concepts presented above.

a) The Classical Periodogram

The response of the classical periodogram (eq. [3]) to a sinusoidal signal is completely described by the window function (eqs. [D3] and [D4]) through the relation in equation (D2). Note that the window depends only on the observation times, $\{t_i\}$. Figure 3 shows window functions calculated for the epochs of photographic

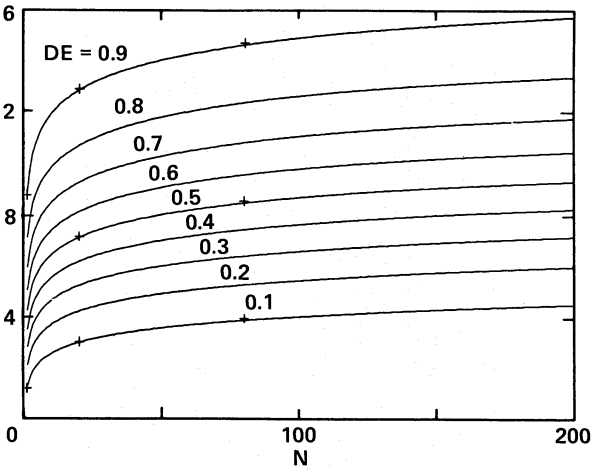

FIG. 2.—Contours of constant detection efficiency (DE = 1 − p^*), as functions of the signal strength, P, and the number of samples, N. The curves, labeled with the value of DE, are from the approximation given in eq. (E12), while the +'s are exact and were calculated from eqs. (16), (21), and (23).

plates obtained for a parallax program at the U.S. Naval Observatory. Dr. Harrington kindly provided the raw data for three stars in which a perturbation due to an unseen (stellar) companion was discovered in the course of this program (Behall and Harrington 1976; Harrington 1977). Even though all three stars were observed in more or less the same way, in the fashion typical of parallax programs, the vagaries of astronomical observations show up as differences in the three windows. Nevertheless there are similarities, and all three are generally similar to the standard "sinc^2" window for even spacing. In particular, all three have a narrow central peak, the *main lobe*, the width of which is of order $2\pi/T$. At the first null of the sinc^2 function, one window has a null, and two have minima. The sidelobe structure is different from star to star, but all can be described as having roughly evenly spaced lobes, with amplitudes considerably larger than those of the rapidly declining sinc^2 function. The real windows have one thing absent in the sinc^2 function, namely rather large peaks on either side of the main lobe and displaced from it by 1 cycle per year ($\omega = 2\pi$ radians per year). This is due to the 1 yr rough periodicity characteristic of any parallax program, as discussed above (§ II*a*). The sinc^2 window has one thing absent in the real windows, namely a series of large peaks on either side of the main lobe and displaced from it by integer multiples of 1 cycle per sampling interval (i.e., the Nyquist frequency). This aliasing is due to the regularity of the sampling, as described in § II*a*.

b) The Modified Periodogram

We now wish to compare the classical periodogram with the modified form in equations (10) and (11). As

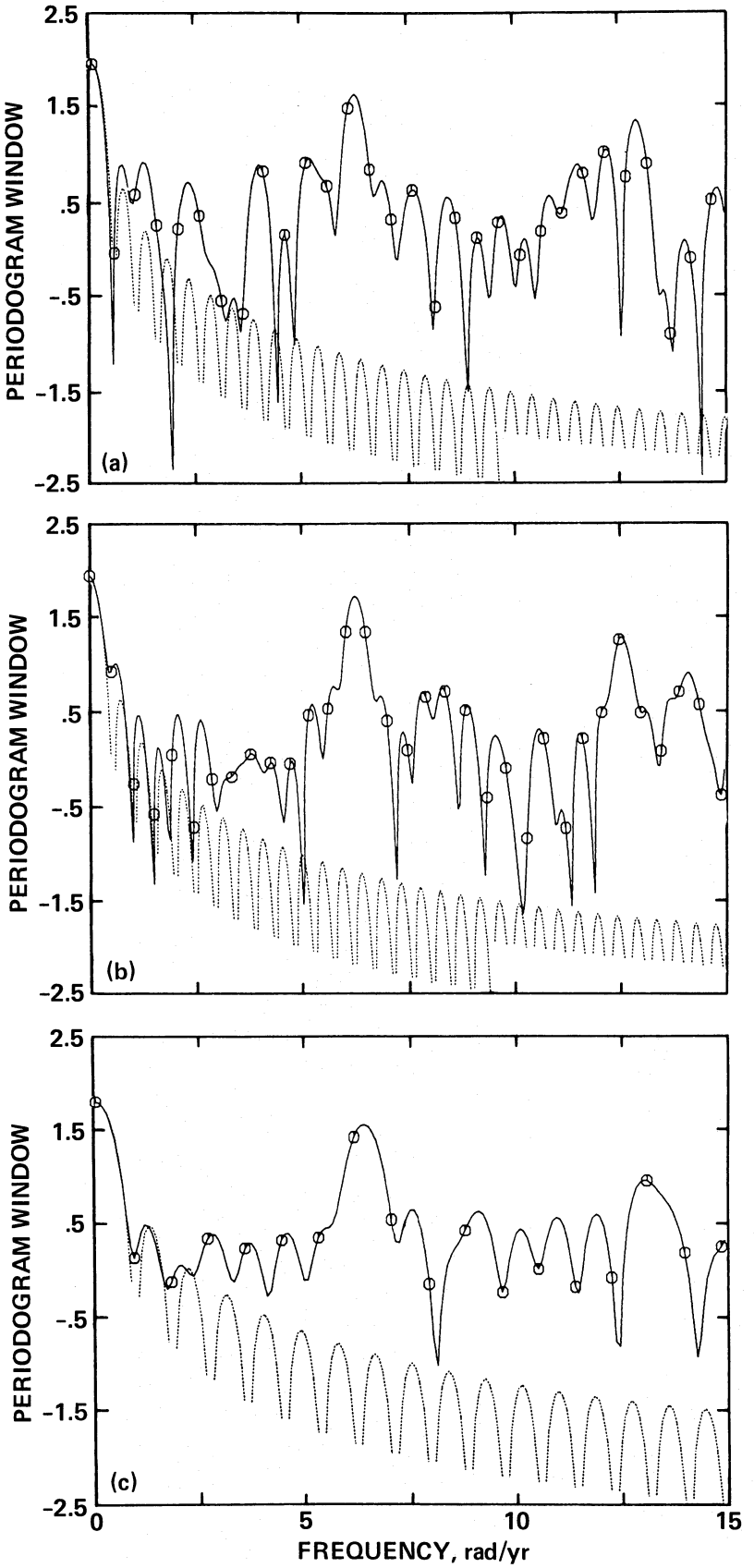

FIG. 3.—Logarithmic plots of the classical periodogram windows obtained from astrometric data for the stars: (*a*) G96–45, (*b*) G146–72, and (*c*) Wolf 1062. In each case, the solid line is the window calculated from the classical formulas (D3) and (D4), while the dotted line shows for comparison the window function for even spacing with the same total time interval and number of data points. The open octagonal symbols are the window function evaluated at the grid of frequencies defined in eq. (D1), while the solid curve is oversampled by a factor of 10 relative to this grid.

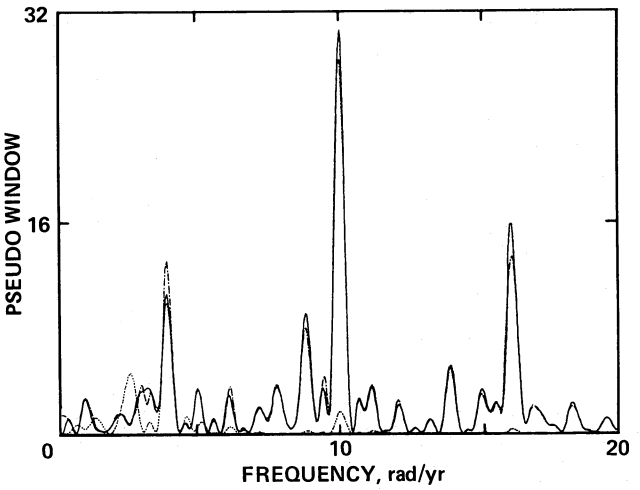

FIG. 4.—Comparison of the classical periodogram (*solid line*) from eq. (3) with the modified form (*dot-dash line*) from eq. (10). Since, as explained in the text, the spectral response of the modified periodogram cannot be written in the form in eq. (D2), strictly speaking the window function is not defined. Hence what is shown here is a pseudowindow, namely the response to a high-frequency sinusoid ($\omega_s = 10$), with sampling as in Fig. 3*a*. The dotted line is the classical periodogram computed from the same data interpolated to even spacing.

detailed in Appendix D, the response of the modified periodogram to a sinusoidal input cannot, strictly speaking, be described in terms of a window function because it cannot be written in the form in equation (D2). But we can directly calculate the response to a sinusoid of a specific frequency, sampled at the particular set of t_i's under consideration. The synthetic signal will be of the form

$$X_s(t_i) = \sin \omega_s t_i + aR_i, \tag{25}$$

where the term aR_i represents the noise. The process R was generated by adding five random variables uniformly distributed on the interval $(-\frac{1}{2}, \frac{1}{2})$, thereby producing a pseudo-Gaussian random variable of variance 0.64.

Figure 4 shows the modified periodogram for the case $\omega_s = 10$ and $a = 0$, with sampling as in Figure 3*a*. It is presumed that this frequency is large enough that the overlap terms and negative frequency spillover described in Appendix D are not important, so that the periodograms shown are much like effective, or approximate, windows shifted to the frequency origin ω_s. The asymmetry of the curves in this figure is probably due mostly to the overlap and spillover effects (similar computations with higher signal frequencies give windows that are more nearly symmetric). It can be seen in Figure 4 that the difference between the modified periodogram and the classical periodogram is not large. The biggest effect is a small change in the relative amplitudes of the main lobe and the quasi-alias sidelobes.

Figure 4 also shows the periodogram (*dotted line*) of the same data linearly interpolated to even spacing. Such interpolation is very bad for a high frequency signal, as can be readily seen from the figure, since much of the oscillation is lost in the interpolation, by being replaced with linear segments. Hence it should not be surprising that the spectrum of the interpolated data almost completely misses the main peak. Interpolation is not nearly so bad for low frequency sinusoids, but probably should never be used if an alternative is available. One of the main points of this paper is that an alternative is available—the periodogram. For more information on how interpolation affects the power spectrum, consult Horowitz (1974), who discusses the effects of different orders of spline interpolation on the power spectrum.

Figure 5 depicts another view of the effect of the modification of the periodogram. It shows the denominator terms which appear in the modified periodogram, as a function of frequency (see eq. [10]). The 10–20% variations seen here are typical of moderately irregular sampling.

c) *Noisy Data*

Figure 6 shows how the periodogram degenerates as increasing amounts of noise are added to the process analyzed in Figure 4. The power signal-to-noise levels, calculated from equation (9) with $X_0 = 1$, $N_0 = 107$, and $\sigma_0 = 0.64$a, are: (*a*) 65, (*b*) 2.6, and (*c*) 0.65. The signal is clearly detected in (*a*), while in (*b*) the noise is such that detection could be claimed only for a rather large threshold (and thus running considerable false alarm risk). Quantitatively, the peak power divided by the

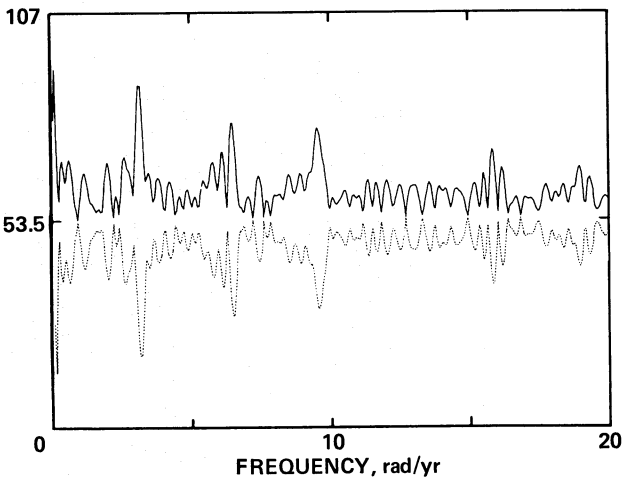

FIG. 5.—The expressions $\Sigma_j \cos^2 \omega t_j$ (*top*) and $\Sigma_j \sin^2 \omega t_j$ (*bottom*). These are the denominators in the corrected periodogram, eq. (A14). The classical periodogram, eq. (3), is obtained if these denominators are approximated with the constant $N_0/2$, indicated by the tick mark in the middle of the ordinate scale ($N_0 = 107$ in this example). The times $\{t_i\}$ are as in Fig. 3*a*.

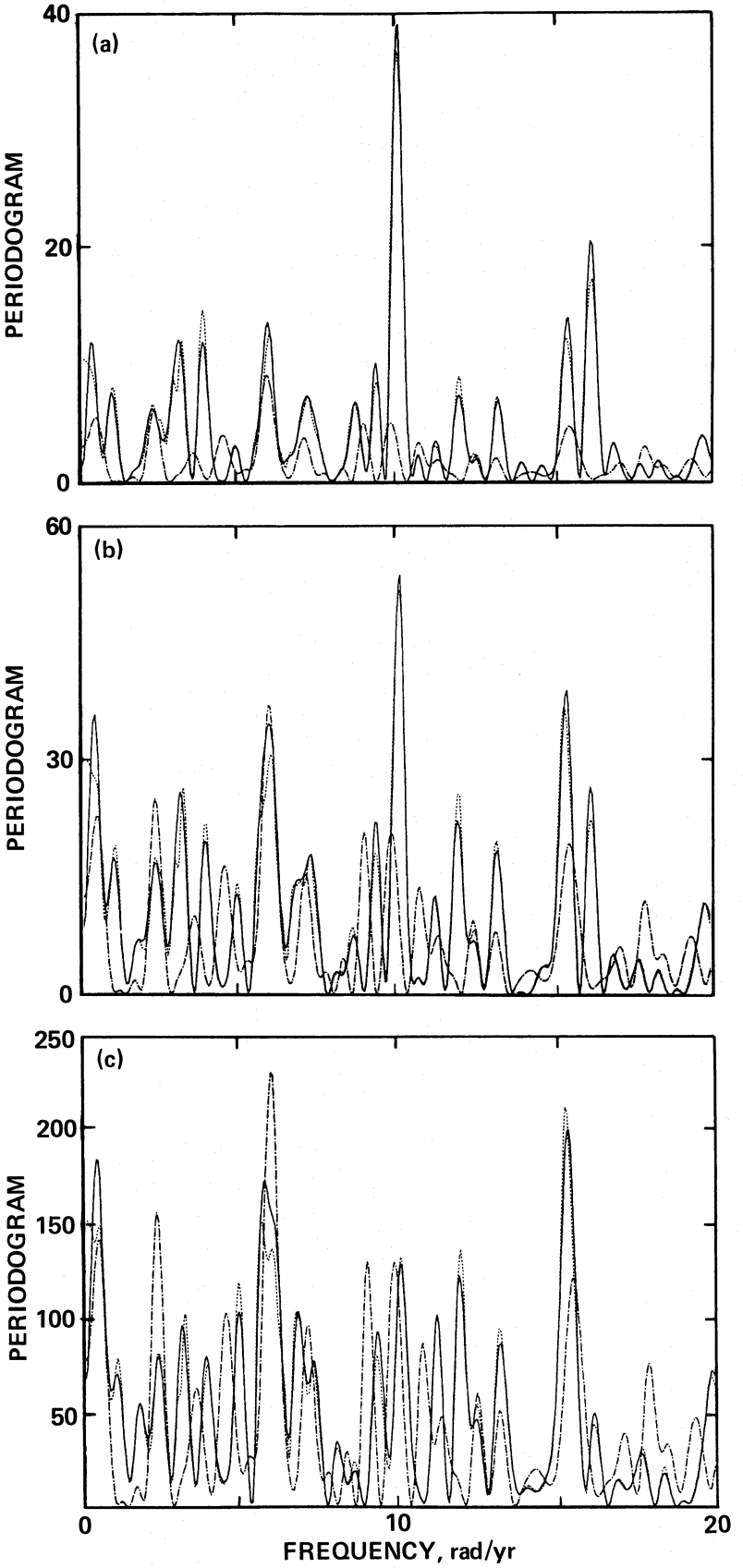

FIG. 6.—Pseudowindows determined by analyzing artificial data, as in Fig. 4, but noise has been added according to eq. (25), with: (*a*) $a=2$, (*b*) $a=4$, and (*c*) $a=10$ (the corresponding signal-to-noise ratios are 16, 4, and 0.65). The solid line is the classical periodogram, and the dotted line is the modified one; the dot-dash line is the (classical) periodogram of the noise alone.

noise power in Figure 6*b* is about 8.2; and with $N_0 = 107$, equation (14) shows that this signal-to-noise ratio corresponds to detection with 97.0% confidence—moderately good, but not overwhelming. If this is an acceptable detection, then a sine wave of amplitude 1 has been detected in the presence of noise of variance more than 2.5 times as great. This result demonstrates the claims made in § II*a* that the periodogram is well suited to the problem of detecting a periodic signal in the presence of noise. In Figure 6*c* the main peak is no longer even at the signal frequency; detection could not be claimed, as the peak signal-to-noise is about 5, corresponding to roughly 50% chance of being a random accident.

d) *Window Carpentry*

Spectral leakage is another issue in the use of the periodogram. Indeed, the sidelobes of the periodogram windows shown in Figures 3 and 4 are substantially larger than the lobes of the sinc^2 window, indicating that there is significant leakage away from the input frequency. In addition the quasi-alias peaks, due to the annual regularity in the sampling, are bothersome and would be a major source of confusion in dealing with all but the simplest signals.

The classical approach to plugging up spectral leakage is through windowing of the data or mathematically equivalent techniques. The basic idea is to multiply the data samples by a function of the time index that goes smoothly to zero at the ends of the data, but is unity over the central region of the sampling interval. Rather subtle differences in the shape of this function (which is sometimes called the *data window*) can make large differences in the way in which the sidelobe amplitudes fall off with frequency. The paper by Harris (1978) gives a catalog of many such data windows and the corresponding sidelobe structure.

The periodogram for unevenly spaced data allows two different forms of spectral window adjustment: (1) application of time-domain (data) windows, and (2) adjustment of the locations of the sampling times (which can be done, of course, only if these are at the experimenter's disposal). The first of these can be effected simply by replacing $X(t_i)$, everywhere it appears in the periodogram formula (eqs. [3] and [10]), with $w(t_i)X(t_i)$ (Thompson 1971). The function $w(t)$ can be represented either as a continuous function of time or as a set of weights, $w_i = w(t_i)$, one for each of the sampling times.

Several numerical experiments have been performed, the details of which will be given in a subsequent paper in this series. Here only a brief summary will be given. Experiments were carried out in which the weighting function $w(t)$ was varied, or the sampling times t_i were varied, or both. One must choose some aspect of the window function which is to be minimized; examples are (1) the amplitude of a particular sidelobe, (2) the sum of the amplitudes of the sidelobes from the first up to some specific frequency, such as the Nyquist frequency, (3) the amplitude of the peak at the Nyquist frequency (this is one way to measure the amount of aliasing), (4) the width of the main lobe (this is a measure of resolution). For an ideal system, the delta-function spectral response corresponds to zero for each of the quantities listed above. Indeed, this is the reason for minimizing the quantities in the first place.

The results of these numerical studies can be summarized as follows: (1) the time points (t_i) control the power in the window function which leaks to the Nyquist frequency and beyond (i.e., the aliasing), while (2) the weights $\{w_i\}$ control the sidelobes, and (3) there is very little cross-talk between (1) and (2). This is not unexpected from elementary considerations, such as those given in § II above. Window functions closely approximating the Hanning weights, and other standard weights, are the solutions to optimization problems in which the weights are free parameters and the integrated sidelobe power is the quantity to be minimized.

In addition, more specialized problems have been considered. For example, suppression of the window response over a specific range of frequencies can be easily accomplished. Such "passband" windows might be of use when a weak signal is suspected at a frequency offset by a known amount from a strong signal. Some additional experiments were preformed, in which the weights $\{w_i\}$ were allowed to be complex numbers. This is equivalent to introducing a variable phase shift between the components, not unlike "beam steering" in radio astronomy, where interferometer antennas are connected with variable delay lines. The general result seems to be what would be expected from this analogy, namely that the phase shifts cause the power to move away from the main lobe—and this is generally undesirable. On the other hand, there seem to be some cases, such as the passband experiments mentioned above, where the complex weights appear to be of some influence on the final window shape.

VI. SUMMARY AND FURTHER WORK

Some general aspects of the use of the periodogram for the detection of periodic signals hidden in noise have been given, with emphasis on the analysis of unevenly spaced data. The statistical distribution of this estimator of the power spectrum is simple and easy to use, especially if the classical definition is slightly modified. In particular, simple formulae can be used to (1) define a threshold power level above which a peak in the periodogram indicates that a signal is almost certainly (with probability $[1 - p_0]$) present (eq. [18]), (2) calculate the probability of a chance noise fluctuation exceeding a given power level (eq. [14]), and (3) calculate the probability that a signal of a given amplitude (relative to the noise) will be detected (eqs. [21] and [23]). The discussion in § II should clarify some of the practical issues in

the use of the periodogram, and aid one in deciding when it is a good tool to use.

Further work needs to be done in window carpentry. In particular, the relationship between the standard window functions and the minimization problems sketched in § V*d* should be clarified. Such problems should probably be formulated in terms of a weighting function (in the frequency domain) which specifies the relative importance of different aspects of the spectral response function. It is presumed that the minimization problem corresponding to each such function defines a unique data window, and that the standard data windows can be defined in this way. In addition, a systematic study of the nature of the spectral window, as a function of the sampling, needs to be carried out. The possibility that complex weights can be of use also needs study. Preliminary remarks on these matters have been made above, and more detailed results will be published subsequently.

It is a pleasure to thank Dave Black for suggesting the problem (see BS) which led to this work, as well as for continued encouragement. Paul Swan and Gary Villere made useful suggestions regarding the presentation. Robert Harrington provided the data used in some of the numerical examples. The referee made several useful suggestions.

APPENDIX A

STATISTICAL DISTRIBUTION OF THE PERIODOGRAM

Consider the following generalization of the discrete Fourier transform (DFT):

$$FT_X(\omega) = (N_0/2)^{1/2} \sum_{j=1}^{N_0} X(t_j)\left[A \cos \omega t_j + iB \sin \omega t_j\right]; \tag{A1}$$

A and B are as yet unspecified functions of ω, and may depend on the sampling, $\{t_j\}$, but not on the data, $\{X(t_j)\}$, nor on the summation index j. The corresponding periodogram is

$$\begin{aligned} P_X(\omega) &= (1/N_0)|FT_X(\omega)|^2 \\ &= (A^2/2)\left[\sum_j X(t_j) \cos \omega t_j\right]^2 + (B^2/2)\left[\sum_j X(t_j) \sin \omega t_j\right]^2. \end{aligned} \tag{A2}$$

If $A = B = (2/N_0)^{1/2}$, equations (A1) and (A2) reduce to the classical definitions. The basic rationale behind these definitions is that for even sampling FT_X reduces to the DFT (and in the limit $\Delta t \to 0, N \to \infty$, is proportional to the Fourier transform); similarly for P_X and the power spectrum. But this reduction is not unique: there are other choices for A and B which reduce to the same expressions for even sampling, as we shall soon see. Hence, additional conditions must be imposed to determine A and B. In particular, the statistical distribution of the generalized periodogram will be made as closely as possible the same as it is in the evenly spaced case.

This can be achieved with simple choices for A and B. Consider the important case in which X is pure noise—independently and normally distributed noise, with zero mean and constant variance σ_0^2. Then the quantity

$$C(\omega) = A \sum_j X(t_j) \cos \omega t_j \tag{A3}$$

is a linear combination of independent normal random variables, since the $A \cos \omega t_j$ are simply constant coefficients in this context. But a linear combination of normally distributed random variables is also normal (Parzen 1962, § 3.4, Theorem 4A, p. 90). The mean and variance of C are: $\langle C \rangle = 0$ and

$$\sigma_c^2 = \langle C^2(\omega) \rangle = A^2 \sum_j \sum_k \left\langle X(t_j) X(t_k) \right\rangle \cos \omega t_j \cos \omega t_k \tag{A4}$$

$$= A^2 N_0 \sigma_0^2 \sum_j \cos^2 \omega t_j, \tag{A5}$$

since the cross terms ($j \neq k$) vanish due to the assumed independence. Similarly,

$$S(\omega) = B \sum_j X(t_j) \sin \omega t_j \tag{A6}$$

is normal with zero mean and variance

$$\sigma_s^2 = B^2 N_0 \sigma_0^2 \sum_j \sin^2 \omega t_j . \tag{A7}$$

Now equation (A2) can be written

$$P_X(\omega) = \tfrac{1}{2}\left[C^2(\omega) + S^2(\omega)\right], \tag{A8}$$

so that P is the sum of the squares of two normally distributed, zero-mean random variables. It is well known (e.g., Papoulis 1965, § 7.1, example 7-7, pp. 194–195) that such a sum has an exponential probability distribution, but *only if the variances of the two normal variables are the same*. Let X and Y be two zero-mean random variables with variances σ_1^2 and σ_2^2, respectively. Then the methods in Papoulis (1965) yield that $Z = X^2 + Y^2$ is distributed according to

$$P_Z(z) = \frac{\exp\left(-z/2\sigma_2^2\right)}{2\sigma_1\sigma_2} G\left[(z/4)\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\right], \tag{A9}$$

where

$$G(x) = \exp(-x) I_0(x) \tag{A10}$$

and I_0 is the modified Bessel function of the first kind. For $\sigma_1 = \sigma_2 = \sigma$, this reduces to the usual result for the sum of squares of two normal variables of equal variance (and mean), namely the exponential distribution

$$P_Z(z) = \left(1/2\sigma^2\right) \exp\left(-z/2\sigma^2\right). \tag{A11}$$

But if $\sigma_1 \neq \sigma_2$, distribution (A9) is quite different from (A11). From equations (A5) and (A7) it can be seen that the choices

$$A(\omega) = Q(\omega)\left(\sum_j \cos^2 \omega t_j\right)^{-1/2} \tag{A12}$$

and

$$B(\omega) = Q(\omega)\left(\sum_j \sin^2 \omega t_j\right)^{-1/2} \tag{A13}$$

give the necessary equality of the variances, namely $\sigma_c = \sigma_s (= \sigma_0)$. In these equations $Q(\omega)$ is an arbitrary function of ω. Its proper value, $Q(\omega) = 1$, is determined by the condition that P have the same mean value as in the evenly spaced case. The resulting periodogram, namely

$$P_X(\omega) = \frac{1}{2}\left[\frac{\left(\sum_j X_j \cos \omega t_j\right)^2}{\sum_j \cos^2 \omega t_j} + \frac{\left(\sum_j X_j \sin \omega t_j\right)^2}{\sum_j \sin^2 \omega t_j}\right], \tag{A14}$$

has exactly the same (exponential) probability distribution as for even spacing. However, the joint distributions are different, as discussed in Appendix D. For evenly spaced t_j it can be shown with elementary trigonometry that $A(\omega) = B(\omega) = (2/N_0)^{1/2}$ whenever $\omega = \omega_n$ (see eq. [D1]). But A and B can be very far from these values for other

values of ω. For uneven spacing, A and B are not equal, even at the ω_n. Nevertheless, unless the sampling is pathological, $A \approx B \approx (2/N_0)^{1/2}$ for the relevant values of ω. Thus $A = B = (2/N_0)^{1/2}$, which reduces (A1) and (A2) to the classical definitions, is not a bad approximation in many cases.

The form of the power spectrum in (A14) was also arrived at, using a least-squares regression not unlike that of Lomb (1976), by Kar, Hornkohl, and Farmer (1981). They propose this form as an approximation to the power spectrum, for randomly sampled data, which is computationally faster and requires less computer storage than conventional methods, and in one example also provides somewhat superior stop-band rejection.

APPENDIX B

TIME-TRANSLATION INVARIANCE OF THE PERIODOGRAM

Invariance to time translation is a useful property possessed by the classical periodogram. That is, if there is a shift of the time origin, say $t_j \to t_j + T_0$ for every j, then the periodogram in equation (4) (or eq. [3]) is unchanged: there is simply a phase factor $\exp(i\omega T_0)$ of modulus unity inside the absolute value. Our periodogram, equation (A14), does not have this property. Even though the actual changes in the values brought about by time translation are usually small, it is nevertheless satisfying to find a way of restoring invariance. After all, the power spectrum is meant to measure harmonic content of signals without regard to phase. There are many ways to restore invariance. The following procedure—while it appears capricious at the moment—is chosen for reasons that will become transparent in Appendix C. Insert a delay τ into all of the time arguments in equation (A14) as follows:

$$P_X(\omega) = \frac{1}{2}\left\{\frac{\left[\sum_j X_j \cos\omega(t_j - \tau)\right]^2}{\sum_j \cos^2\omega(t_j - \tau)} + \frac{\left[\sum_j X_j \sin\omega(t_j - \tau)\right]^2}{\sum_j \sin^2\omega(t_j - \tau)}\right\}, \tag{B1}$$

and define

$$\tau = (1/2\omega)\tan^{-1}\left[\frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}\right]. \tag{B2}$$

The formula for the tangent of a sum can be used to show that if t_j becomes $t_j + T_0$, then τ becomes $\tau + T_0$; hence T_0 cancels out in the arguments $\omega(t_j - \tau)$ in equation (B1). Further, this change does not alter the statistical results in Appendix A. Equation (B1), with (B2), is the definition we adopt: it has time-translation invariance, a simple statistical behavior virtually identical to that for even sampling, and it reduces to the ordinary periodogram, equation (4), if the sampling is even (for in this case $\tau \equiv 0$). The computation of (B1) is somewhat more complicated than that of the classical periodogram, equation (3). The only tricky part is dealing with the 2π ambiguity in the arctangent function in (B2). The secret is to impose continuity on τ as a function of ω, and to use sufficiently high frequency resolution so that no phase jumps are missed. It is also important to note that

$$\lim_{\omega \to 0} \tau(\omega) = (1/N_0)\sum_{j=1}^{N_0} t_j = \langle t \rangle, \tag{B3}$$

and to circumvent the indeterminacy in the second term of (B1) at $\omega = 0$ [both the numerator and denominator are $O(\omega^2)$]. The computation time depends on N_0 in the same way for (B1) as for the classical periodogram. Note that τ, A, and B can be calculated once and for all for a given sampling.

APPENDIX C

EQUIVALENCE OF PERIODOGRAM AND HARMONIC LEAST-SQUARES ANALYSIS

An alternate approach to the detection problem outlined in § I is the fitting, in the least-squares sense, of sine waves directly to the data. That is, we define

$$X_f(t) = A\cos\omega t + B\sin\omega t, \tag{C1}$$

and seek to minimize the mean square difference between this model and the data, viz.,

$$E(\omega) = \sum_{j=1}^{N_0} \left[X(t_j) - X_f(t_j) \right]^2. \tag{C2}$$

Since A and B enter the residual in a linear way, in contrast to ω, they may be determined by standard linear least-squares techniques. The resulting minimized value of E, $E_{\rm MIN}(\omega)$, can then be minimized as a function of ω numerically or graphically. Equivalently, one may define the reduction in the sum of squares as

$$\Delta E(\omega) = \sum_{j=1}^{N_0} \left[X(t_j) \right]^2 - E_{\rm MIN}(\omega), \tag{C3}$$

and seek a maximum in $\Delta E(\omega)$. The details of this procedure can be found in Lomb (1976), who generalizes (C1) to

$$X_f(t) = A \cos \omega(t-\tau) + B \sin \omega(t-\tau). \tag{C4}$$

Lomb introduces the redundant parameter τ because, if it can be chosen so that

$$\sum_{j=1}^{N_0} \cos \omega(t_j - \tau) \sin \omega(t_j - \tau) = 0, \tag{C5}$$

then his complicated explicit formula for $\Delta E(\omega)$ simplifies to

$$\Delta E(\omega) = \frac{\left[\sum_j X(t_j) \cos \omega(t_j - \tau) \right]^2}{\sum_j \cos^2 \omega(t_j - \tau)} + \frac{\left[\sum_j X(t_j) \sin \omega(t_j - \tau) \right]^2}{\sum_j \sin^2 \omega(t_j - \tau)}. \tag{C6}$$

The solution to equation (C5) is precisely the expression for τ which gives time translation invariance (eq. [B2]). Note that, despite their very different meanings and derivations, $\Delta E(\omega)$ and $P_X(\omega)$ are exactly the same (eqs. [10] and [C6])! A maximum in the periodogram occurs at the same frequency which minimizes the sum of squares of the residuals of the fit of a sine wave to the data. And clearly any theoretical results for the periodogram—e.g., the statistical discussion above—apply equally to least-squares analysis, and vice versa.

APPENDIX D

NATURAL FREQUENCIES AND THE SPECTRAL WINDOW

In practice one is faced with the problem of choosing a finite set of frequencies at which to evaluate the periodogram. For the case of even spacing there is a well-known natural set of frequencies, defined by

$$\omega_n = 2\pi n / T \qquad \{n = -N_0/2, \ldots, +N_0/2\} \tag{D1}$$

(T is the total time interval). The significance of this set of frequencies is that the DFT (eq. [2]), evaluated at these frequencies, contains just enough information to recover the original data. Indeed, an explicit formula can be written for X_n in terms of the FT(ω_n). Since the periodogram of real data is symmetric [$P(-\omega) = P(\omega)$], all of its information is contained in the positive frequencies, $n = 0, 1, 2, \ldots, N (= N_0/2)$. Roughly half of the information in the data has been thrown away by going from the DFT to the periodogram (the absolute value discards the phase, but retains the amplitude), so half as many frequencies are necessary. Evaluation of the periodogram at intermediate frequencies gives plots that look smoother. For example, all of the periodograms shown in this paper are plotted at 10 times $N_0/2$ frequencies. This oversampling is really a kind of interpolation that adds no information. Furthermore, the random variables $P(\omega_n)$ are independent of each other—whereas the $P(\omega)$ at intermediate frequencies are dependent variables.

An intuitive way of looking at the meaning of the set of frequencies defined above is that the fundamental frequency, $\omega_1 = 2\pi/T$, corresponds to a sine wave of period equal to the whole interval T. This is roughly the lowest

frequency about which there is information in the data. The so-called Nyquist frequency, $\omega_N = \frac{1}{2}(2\pi/\Delta t) = \pi N_0/T$ ($\Delta t = T/N_0$ is the sampling interval) is roughly the highest frequency about which there is information, because Δt is the shortest time interval spanned.

If the sampling is uneven, the fundamental frequency is basically unchanged in both meaning and value, since the interval $T = \max(t_i) - \min(t_i)$ is still well defined. However, the meaning of the Nyquist frequency is more complicated. The highest frequency about which there is information is π divided by $\Delta t_{\min} = \min(t_{i+1} - t_i)$, but the average value of Δt might better be used in defining a generalized Nyquist frequency—but which mean is appropriate: algebraic, geometrical, harmonic,...? We will now see that the best way of choosing the natural frequencies is through consideration of the spectral response function, sometimes called the *spectral window*.

Roughly speaking, the spectral window is used to describe the response of the entire data analysis system to a monochromatic (i.e., single frequency) sine wave. This is not a unique definition, to the extent that the periodogram is a function of the phase of the sine wave. However, this dependence is usually weak, and one can meaningfully average over phase. Deeming (1975, eq. [8]) has shown that the expectation value of the classical periodogram is equal to the convolution of the true power spectrum with the spectral window function for the particular sampling used (see eq. [D3] below). Thus the window function is a kind of Green's function. One reason for its importance lies in the fact that all spectral leakage effects (aliasing, sidelobes, interference phenomena, ghosts, etc.) are manifested directly in the window and can be easily evaluated quantitatively (Deeming 1975). Lomb (1976) has derived expressions for the periodogram due to a sinusoidal signal. The exact expression is rather messy, but when averaged over phase and simplified with the approximations that reduce the full modified periodogram in equation (10) to the classical periodogram in equation (3), it reduces to

$$P_s(\omega) = |W(\omega - \omega_s) + W(\omega + \omega_s)|^2; \tag{D2}$$

P_s is the periodogram due to a sine wave of frequency of ω_s, and W is the DFT of the time-domain observing window:

$$W(\omega) = (1/N_0) \sum_{j=1}^{N_0} \exp(i\omega t_j). \tag{D3}$$

It can be seen from (D2) that $P_s(\omega)$ consists of three contributions: the function

$$G(\omega) = |W(\omega)|^2, \tag{D4}$$

with its origin shifted to the signal frequency [i.e., $|W(\omega - \omega_s)|^2$], plus a similar term shifted to $-\omega_s$ [i.e., $|W(\omega + \omega_s)|^2$], plus the overlap term $2W(\omega - \omega_s)W(\omega + \omega_s)$. If $W(\omega)$ is narrowly peaked about $\omega = 0$ and ω_s is not too small, then for $\omega > 0$ the second and third terms are negligible, and the basic response is $G(\omega - \omega_s)$. Hence G is called the *periodogram window*, or *spectral window*. For the modified periodogram we can derive, from expressions given by Lomb (1976) for least-squares analysis, the following formula for the phase-averaged sine wave response:

$$2P_s(\omega) = \frac{(\Sigma \cos \omega_s t_j \cos \omega t_j)}{(\Sigma \cos^2 \omega_s t_j)(\Sigma \cos^2 \omega t_j)} + \frac{(\Sigma \cos \omega_s t_j \sin \omega t_j)}{(\Sigma \cos^2 \omega_s t_j)(\Sigma \sin^2 \omega t_j)}$$
$$+ \frac{(\Sigma \sin \omega_s t_j \cos \omega t_j)}{(\Sigma \sin^2 \omega_s t_j)(\Sigma \cos^2 \omega t_j)} + \frac{(\Sigma \sin \omega_s t_j \sin \omega t_j)}{(\Sigma \sin^2 \omega_s t_j)(\Sigma \sin^2 \omega t_j)}, \tag{D5}$$

where the $-\tau$ terms have been suppressed for simplicity, but are assumed present with every t_j. It does not appear possible to rewrite this expression in the form (D2), so a window function in the sense of the expression in equation (D4) cannot be defined. In particular, the following seemingly straightforward generalization of the classical window function (cf. eqs. [3], [D4], and [10]),

$$P(\omega) = 1/2 \left\{ \frac{\left[\sum_j \cos \omega(t_j - \tau)\right]^2}{\sum_j \cos^2 \omega(t_j - \tau)} + \frac{\left[\sum_j \sin \omega(t_j - \tau)\right]^2}{\sum_j \sin^2 \omega(t_j - \tau)} \right\}, \tag{D6}$$

is not a correct window function. In lieu of having a formula for the window, one can calculate the periodogram for synthetic data consisting of a sine wave of high frequency, so that the negative frequency and overlap contributions would be expected to be small. Figure 4 shows a pseudowindow determined in this way. Similar computations verify that the shape of this pseudowindow does not depend much on the frequency of the sine wave (as long as it is large). There is a mild phase sensitivity, in that the relative amplitudes of the sidelobes change somewhat as the phase of the sine wave is altered. As described in the text, the pseudowindows have the general character of the classic window: main lobe, nulls near $\omega = 2\pi/T$, sidelobes, pseudo-alias peaks, etc.

APPENDIX E

DETECTION EFFICIENCY CONTOURS

Equation (21) is an explicit formula for the missed signal probability, which is 1 minus the detection efficiency. Since it involves Bessel function integrals, it is not a convenient expression. A simple but accurate formula can be derived for the contours of equal p^* in the N-P plane. Along such contours

$$dp^* = (\partial p^*/\partial N)\, dN + (\partial p^*/\partial P)\, dP, \tag{E1}$$

so that the contour slope is

$$dP/dN = -(\partial p^*/\partial N)/(\partial p^*/\partial P), \tag{E2}$$

and from the form in equation (22) it is readily shown that

$$dP/dN = P(dz/dN)\Lambda(z_0 P), \tag{E3}$$

where

$$\Lambda(x) = \psi_1(x)/\psi_2(x), \tag{E4}$$

$$\psi_1(x) = \sum_{m=0}^{\infty} x^m/(m!)^2, \tag{E5}$$

$$\psi_2(x) = \sum_{m=0}^{\infty} x^{m+1}/\left[(m+1)(m!)^2\right] \qquad (\psi_1 = \psi_2'), \tag{E6}$$

and from equation (18)

$$dz_0/dN = \frac{N^{-2}(1-p_0)^{1/N}\ln(1-p_0)}{(1-p_0)^{1/N}-1}. \tag{E7}$$

The following asymptotic forms are easily derived:

$$\Lambda(x) \sim \begin{cases} x^{-1} + \frac{1}{2} + O(x) & \text{as } x \to 0 \quad \text{(E8)} \\ x^{-1/2} & \text{as } x \to \infty. \quad \text{(E9)} \end{cases}$$

The first comes from the series expansions for ψ_1 and ψ_2, while the second is based on numerical evaluations. The approximation

$$\Lambda(x) \approx (1+x)^{-1}\left(x^{1/2} + \frac{1}{2} + x^{-1}\right) \tag{E10}$$

is good over the entire range of x, but somewhat awkward to deal with. But z_0 is ~ 8 (cf. eq. [20]) and P must be greater than 1 in order for the detection efficiency to be reasonably high, so that typical values of $x = z_0 P$ will be large

compared to 1. Accordingly we use (E9), together with the close approximation $dz_0/dN \approx N^{-1}$, to obtain:

$$dP/dN = (P/N)\left[P \ln (N/p_0)\right]^{-1/2}. \tag{E11}$$

This equation has the exact integral

$$P = \left\{\left[\ln (N/p_0)\right]^{1/2} - C\right\}^2, \tag{E12}$$

where C is a constant of integration. Exact evaluation of equation (22) shows that the error in P calculated from this approximation is less than 0.1 over the entire range of P and N covered in Figure 1, and the typical relative error is on the order of 1%. The error is smaller for the larger values of Pz_0, as expected. In these computations it was verified that dropping of the factor $(1-p_0)^{1-1/N}$ in going from equation (21) to (22) results in an entirely negligible error.

REFERENCES

Barning, F. J. M. 1963, *Bull. Astr. Inst. Netherlands*, **17**, 22.
Bartlett, M. S. 1950, *Biometrika*, **37**, 1.
Behall, A. L., and Harrington, R. S. 1976, *Pub. A.S.P.*, **88**, 204.
Beutler, F. J. 1966, *S.I.A.M. Rev.*, **8**, 328.
_____. 1970, *IEEE Trans.*, **IT-16**, 147.
Black, D. C., and Scargle, J. D. 1982, *Ap. J.*, **261**, 854 (BS).
Blackman, R. B., and Tukey, J. W. 1958, *The Measurement of Power Spectra from the Point of Communications Engineering* (New York: Dover).
Brault, J. W., and White, O. R. 1971, *Astr. Ap.*, **13**, 169.
Deeming, T. J. 1975, *Ap. Space Sci.*, **36**, 137; erratum *Ap. Space Sci.*, **42**, 257.
Faulkner, D. J. 1977, *Ap. J.*, **216**, 49.
Ferraz-Mello, S. 1981, *A.J.*, **86**, 619.
Fitch, W. S., and Wehlau, W. 1965, *Ap. J.*, **142**, 1616.
Gaster, M., and Roberts, J. B. 1975, *J. Inst. Math. Appl.*, **15**, 195.
_____. 1977, *Proc. R. Soc. London*, *A*, **354**, 27.
Gray, D. F., and Desikachary, K. 1973, *Ap. J.*, **181**, 523.
Groth, E. J. 1975, *Ap. J. Suppl.*, **29**, 285.
Harrington, R. S. 1977, *Pub. A.S.P.*, **89**, 214.
Harris, F. J. 1978, *Proc. IEEE*, **66**, 51.
Higgins, J. R. 1976, *IEEE Trans.*, **IT-22**, 621.
Horowitz, L. L. 1974, *IEEE Trans.*, **ASSP-22**, 22.
Jenkins, G. M., and Watts, D. G. 1968, *Spectral Analysis and its Applications* (San Francisco: Holden-Day).
Kanasewich, E. R. 1975, *Time Sequence Analysis in Geophysics* (Edmonton: University of Alberta Press).
Kar, M. L., Hornkohl, J. O., and Farmer, W. M. 1981, in *Proc. International Conference on Acoustics, Speech, and Signal Processing* (Piscataway, N.J.: IEEE Service Center), pp. 89–93.
Kay, S., and Marple, S. 1981, *Proc. IEEE*, **69**, 1380.
Kuhn, J. R. 1982, *Ap.J.*, **87**, 196–202.
Lafler, J., and Kinman, T. D. 1965, *Ap. J. Suppl.*, **11**, 216.
Lomb, N. R. 1976, *Ap. Space Sci.*, **39**, 447.
Ludeman, L. C. 1981, *IEEE Trans.*, in press.
Masry, E., and Lui, M.-C. C. 1975, *S.I.A.M. J. Appl. Math.*, **28**, 793.
Meisel, D. D. 1978, *A.J.*, **83**, 538.
_____. 1979, *A.J.*, **84**, 116.
Papoulis, A. 1965, *Probability, Random Variables and Stochastic Processes* (New York: McGraw-Hill).
Parzen, E. 1962, *Stochastic Processes* (San Francisco: Holden-Day).
Richards, P. I. 1967, *IEEE Spectrum*, **4**, 83.
Scargle, J. D. 1981, *Ap. J. Suppl.*, **45**, 1 (Paper I).
Schuster, A. 1898, *Terrestrial Magnetism* (now *J. G. R.*), **3**, 13.
Sturrock, P. A., and Shoub, E. C. 1981, Stanford University, Institute for Plasma Research, Report No. 824R.
Swan, P. 1981, preprint.
Thompson, R. O. R. Y. 1971, *IEEE Trans.*, **GE-9**, 107.
Tukey, J. W. 1967, in *Spectral Analysis of Time Series*, ed. B. Harris (New York: Wiley), pp. 25–46.
Vanicek, P. 1969, *Ap. Space Sci.*, **4**, 387.
_____. 1971, *Ap. Space Sci.*, **12**, 10.
Wehlau, W., and Leung, K. C. 1964, *Ap. J.*, **139**, 843.
Wiley, R. G. 1978, *IEEE Trans.*, **COM-26**, 135.

JEFFREY D. SCARGLE: Mail Stop 245-3, Theoretical and Planetary Studies Branch, Space Science Division, Ames Research Center, NASA, Moffett Field, CA 94035