JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. A5, PAGES 10,355–10,368, MAY 1, 1999

Time series, periodograms, and significance

G. Hernandez

Graduate Program in Geophysics, University of Washington, Seattle

Abstract. The geophysical literature shows a wide and conflicting usage of methods employed to extract meaningful information on coherent oscillations from measurements. This makes it difficult, if not impossible, to relate the findings reported by different authors. Therefore, we have undertaken a critical investigation of the tests and methodology used for determining the presence of statistically significant coherent oscillations in periodograms derived from time series. Statistical significance tests are only valid when performed on the independent frequencies present in a measurement. Both the number of possible independent frequencies in a periodogram and the significance tests are determined by the number of degrees of freedom, which is the number of true independent measurements, present in the time series, rather than the number of sample points in the measurement. The number of degrees of freedom is an intrinsic property of the data, and it must be determined from the serial coherence of the time series. As part of this investigation, a detailed study has been performed which clearly illustrates the deleterious effects that the apparently innocent and commonly used processes of filtering, de-trending, and tapering of data have on periodogram analysis and the consequent difficulties in the interpretation of the statistical significance thus derived. For the sake of clarity, a specific example of actual field measurements containing unevenly-spaced measurements, gaps, etc., as well as synthetic examples, have been used to illustrate the periodogram approach, and pitfalls, leading to the (statistical) significance tests for the presence of coherent oscillations. Among the insights of this investigation are: (1) the concept of a time series being (statistically) band limited by its own serial coherence and thus having a critical sampling rate which defines one of the necessary requirements for the proper statistical design of an experiment; (2) the design of a critical test for the maximum number of significant frequencies which can be used to describe a time series, while retaining intact the variance of the test sample; (3) a demonstration of the unnecessary difficulties that manipulation of the data brings into the statistical significance interpretation of said data; and (4) the resolution and correction of the apparent discrepancy in significance results obtained by the use of the conventional Lomb-Scargle significance test, when compared with the long-standing Schuster-Walker and Fisher tests.

1. Introduction

Most geophysical measurements are typically recorded as a time series, and these time series are then analyzed for the presence of (statistically) significant occurrences, later to be interpreted in terms of the underlying mechanisms. The literature shows a wide and conflicting usage of methods employed to obtain meaningful information on coherent oscillations from measurements. This makes it difficult, if not impossible, to relate the findings reported by different authors. The tests employed in the statistical phase of analysis are central to an investigation, and their properties and behavior need to be clearly understood in order to be properly utilized. In this paper we critically investigate statistical analyses of time series in terms of the sampling character of the time series proper, measurement uncertainties, available degrees of freedom, and adverse effects that (apparently) innocent data manipulation can produce in the final result. Specifically, this investigation deals with periodogram analysis and testing.

Paper number 1999JA900026.
0148-0227/99/1999JA900026$09.00

Since the introduction of the periodogram in 1898 [*Schuster*, 1898] for the investigation of hidden periodicities, this has become the standard search for statistically significant coherent oscillations in time series. The study by *Walker* [1914] on the criteria for statistical significance, or reality, of these periodicities placed the use of periodogram techniques on a sound statistical basis. Further investigations on the tests of significance were carried out by *Fisher* [1929], who derived the exact probability distribution and showed that it is not necessary to use the large-sample asymptotic assumptions required in the earlier tests.

Periodogram methods are used in a wide range of observational geophysical disciplines, as well as astronomical and meteorological studies, of which the investigations of *Scargle* [1982] and *Hamilton and Garcia* [1986] are examples. The usage of periodogram techniques has received contemporary attention since their introduction in the *Numerical Recipes* collection [*Press, et al.*, 1986]. However, unqualified use of the statistical significance test provided by them is likely to provide overoptimistic and incorrect conclusions.

The results obtained from periodogram significance analysis simply provide information on the (statistical) probability that no, or one or more, coherent periodicities may exist in a given data sample. This information is then used in interpreting the source, properties, and behavior of these oscillations. The results from the present investigation highlight the conditions necessary, along with some pitfalls, in obtaining a realistic estimate of the statistical significance of periodicities existing in a periodogram derived from a time series.

The following sections give the basic assumptions, derivations, and notation necessary for understanding and using statistical tests on a periodogram derived from a time series. The techniques used for handling unequally spaced data, missing points, and unequally weighted data are also presented. Following these, the periodogram, its statistical properties, and significance tests are described. A concrete example is used to illustrate the deductions. The effects of data serial coherence (affecting the degrees of freedom available) in time series data, and of tapering, filtering, and de-trending (all in the context of the periodogram results) are also discussed.

2. Time Series

In general, it is possible to fit a time series $Y(t)$ with a function of the form:

$$Y(t_j) = Y_j = \sum_{k=1}^{k=K} \Big[a_k \cos(2\pi k T^{-1} t_j) + b_k \sin(2\pi k T^{-1} t_j) \Big] , \quad k = 0,\ 1,\ 2..... \tag{1}$$

By least squares methods, the general best fit (for any of the k coefficients) is given when $<\varepsilon^2>$ is a minimum, where

$$<\varepsilon^2> = \sum_{j=1}^{j=N} [Y_j - a_k \cos(2\pi k T^{-1} t_j) - b_k \sin(2\pi k T^{-1} t_j)]^2 , \tag{2}$$

and

$$\frac{\partial <\varepsilon^2>}{\partial a_k} = \frac{\partial <\varepsilon^2>}{\partial b_k} = 0 . \tag{3}$$

The solution is straightforward, the results being:

$$a_k = \frac{\sum_{j=1}^{j=N} Y_j \cos(2\pi k T^{-1} t_j)}{\sum_{j=1}^{j=N} \cos^2(2\pi k T^{-1} t_j)} - \frac{b_k \sum_{j=1}^{j=N} \sin(2\pi k T^{-1} t_j) \cos(2\pi k T^{-1} t_j)}{\sum_{j=1}^{j=N} \cos^2(2\pi k T^{-1} t_j)} , \tag{4}$$

$$b_k = \frac{\sum_{j=1}^{j=N} Y_j \sin(2\pi k T^{-1} t_j)}{\sum_{j=1}^{j=N} \sin^2(2\pi k T^{-1} t_j)} - \frac{a_k \sum_{j=1}^{j=N} \sin(2\pi k T^{-1} t_j) \cos(2\pi k T^{-1} t_j)}{\sum_{j=1}^{j=N} \sin^2(2\pi k T^{-1} t_j)} . \tag{5a}$$

Equations (4a) and (5a) show the lack of orthogonality between the coefficients. This is not clearly noticeable in the usual full derivation of the coefficients, such as the following for b_k:

$$b_k = \Bigg[\sum_{j=1}^{j=N} Y_j \sin(2\pi k T^{-1} t_j) \sum_{j=1}^{j=N} \cos^2(2\pi k T^{-1} t_j) - \sum_{j=1}^{j=N} Y_j \cos(2\pi k T^{-1} t_j) \sum_{j=1}^{j=N} \sin(2\pi k T^{-1} t_j) \cos(2\pi k T^{-1} t_j) \Bigg] \times \Bigg\{ \sum_{j=1}^{j=N} \sin^2(2\pi k T^{-1} t_j) \sum_{j=1}^{j=N} \cos^2(2\pi k T^{-1} t_j) - [\sum_{j=1}^{j=N} \sin(2\pi k T^{-1} t_j) \cos(2\pi k T^{-1} t_j)]^2 \Bigg\}^{-1} . \tag{5b}$$

When the length of the function $Y(t)$ is an integer multiple of the period T/k, for $k = 1, 2,..., $ and the t_j are equally spaced, i.e., when $t_j = \Delta t \times j$, where $\Delta t = TM^{-1}$, $M = 1, 2,..., N/2$, $j = 1, 2,...N$, and $T = N$, then the orthogonality rules apply:

$$\sum_{j=1}^{j=N} \cos(2\pi k T^{-1} t_j) \cos(2\pi n T^{-1} t_j) = \begin{cases} T/2 & n=k\neq 0 \\ T & n=k\equiv 0 \end{cases} \tag{6}$$

$$\sum_{j=1}^{j=N} \sin(2\pi k T^{-1} t_j) \sin(2\pi n T^{-1} t_j) = \begin{cases} T/2 & n=k\neq 0 \\ T & n=k\equiv 0 \end{cases} \tag{7}$$

$$\sum_{j=1}^{j=N} \cos(2\pi k T^{-1} t_j) \sin(2\pi n/T^{-1} t_j) = 0 . \tag{8}$$

Thus, Equation (4a), Equation (5a) and Equation (5b) become:

$$a_k = \frac{\sum_{j=1}^{j=N} Y_j \cos(2\pi k T^{-1} t_j)}{\sum_{j=1}^{j=N} \cos^2(2\pi k T^{-1} t_j)} = 2T^{-1} \sum_{j=1}^{j=N} Y_j \cos(2\pi k T^{-1} t_j) , \tag{9}$$

$$b_k = \frac{\sum_{j=1}^{j=N} Y_j \sin(2\pi k T^{-1} t_j)}{\sum_{j=1}^{j=N} \sin^2(2\pi k T^{-1} t_j)} = 2T^{-1} \sum_{j=1}^{j=N} Y_j \sin(2\pi k T^{-1} t_j) . \tag{10}$$

Note, however, that a_0 and b_0 are special cases, where

$$a_0 = T^{-1} \sum_{j=1}^{j=N} Y_j ; \quad b_0 = 0 .$$

Usually, the zero-frequency coefficient is of no interest; so the time series used in this study is that obtained by subtracting the mean value of the original series, i.e., redefining $Y_j = (Y_j - a_0)$. This operation is a simple redefinition of the ordinate axis and has the crucial property of preserving the variance. Implicit in this statement is that the time series is a realization of a stationary process of at least order 2 [*Priestley*, 1981]. That is to say, it has the same mean and variance at all time points, and the covariance between the values at any two time points depends on the interval between these time points and not the location of the points along the time axis.

Under these circumstances, the a_k and b_k are orthogonal up to the Nyquist limit, i.e., $K = k_{max} \le N/2$. Because of their independence and orthogonality, these frequencies are known as the natural frequencies. As was pointed out earlier, the original data points are required to be equally spaced, and implicitly each point has the same uncertainty, σ_k = constant. For the case where the time series is not evenly spaced, possibly because of missing data [*Little and Rubin*, 1987], or the data points have unequal weight, and the frequencies desired are not commensurate with the length of the data; then, harmonic fitting techniques must be employed. Under these circumstances, it is possible to approach the normal least squares Fourier series. Specifically, it must be mentioned that in the present context, not evenly spaced data are meant to be irregularly-spaced data. This fitting treatment is due to *Lomb* [1976]. The weight associated with the uncertainties is defined as $w_j = \sigma_j^{-2}$, where $\sigma_j^2 \ne \sigma_k^2$. In addition, an arbitrary phase shift term τ_k is added to each of the trigonometric terms in Equation (2a). This is a simple redefinition of the axis, which does not alter the function. Thus:

$$<\varepsilon^2> = \sum_{j=1}^{j=N} [Y_j - a_k \cos(2\pi k T^{-1}(t_j - \tau_k)) - b_k \sin(2\pi k T^{-1}(t_j-\tau_k))\,]^2\, w_j . \quad (2')$$

Treating Equation (2b) in the same manner as Equation (2), we obtain the following result:

$$a_k = \frac{\sum_{j=1}^{j=N} Y_j w_j \cos[2\pi k T^{-1}(t_j-\tau_k)]}{\sum_{j=1}^{j=N} w_j \cos^2[2\pi k T^{-1}(t_j-\tau_k)]} - \frac{b_k \sum_{j=1}^{j=N} w_j \sin[2\pi k T^{-1}(t_j-\tau_k)]\, \cos[2\pi k T^{-1}(t_j-\tau_k)]}{\sum_{j=1}^{j=N} w_j \cos^2[2\pi k T^{-1}(t_j-\tau_k)]} . \quad (4')$$

If we arbitrarily force the numerator of the second term of Equation (4b) to be zero, we have that:

$$b_k \sum_{j=1}^{j=N} w_j \sin[2\pi k T^{-1}(t_j-\tau_k)] \cos[2\pi k T^{-1}(t_j-\tau_k)] \equiv 0. \quad (11)$$

It is then possible to solve for τ_k:

$$\tau_k = T(4\pi k)^{-1} \tan^{-1}\left[\frac{\sum_{j=1}^{j=N} w_j \sin(4\pi k T^{-1} t_j)}{\sum_{j=1}^{j=N} w_j \cos(4\pi k T^{-1} t_j)}\right], \quad (12)$$

which, when replaced in Equation (4b), gives the desired answer:

$$a_k = \frac{\sum_{j=1}^{j=N} w_j Y_j \cos[2\pi k T^{-1}(t_j-\tau_k)]}{\sum_{j=1}^{j=N} w_j \cos^2[2\pi k T^{-1}(t_j-\tau_k)]} . \quad (9')$$

Although it is possible to obtain estimates of the variance for the determined coefficients, this does not directly give a measure of the importance of a particular frequency in the spectrum relative to the other possible independent frequencies present. Also, if a fitting method is used because of the irregular spacing, unequal variance, etc., in the data, then the orthogonality properties which assured independence of the resultant coefficients are no longer applicable. These properties must be investigated independently. Therefore, it becomes necessary to use other means to obtain a measure of the true number of degrees of freedom in the data series. This topic will be discussed later in Section 5. Implicit in the following discussion is the minimizing of the effects of discontinuities, or edge effects, of real data samples having limited length. 'Tapering' the data with an appropriate window is often employed to accomplish this purpose [*Blackman and Tukey*, 1959; *Priestley*, 1981; *Percival and Walden*, 1993]. This topic will also be discussed.

3. Periodogram

Schuster [1898] defined the periodogram as a measure of the relative power of a time series as a function of frequency. He was searching for 'hidden periodicities', or small periodic variations hidden behind irregular fluctuations. Here our notation will change to the more common usage of $\omega = 2\pi k T^{-1}$. The periodogram is defined [*Priestley*, 1981]:

$$I(\omega) = \left[\sum_{j=1}^{j=N} Y_j \cos(\omega t_j)\right]^2 \times \left[\sum_{j=1}^{j=N} \cos^2(\omega t_j)\right]^{-1} + \left[\sum_{j=1}^{j=N} Y_j \sin(\omega t_j)\right]^2 \times \left[\sum_{j=1}^{j=N} \sin^2(\omega t_j)\right]^{-1}, \quad (13)$$

which, in terms of the previous derivations, can be written as

$$I(\omega) = [A(\omega)]^2 + [B(\omega)]^2 = a^2(\omega) \sum_{j=1}^{j=N} w_j \cos^2(\omega t_j) + b^2(\omega) \sum_{j=1}^{j=N} w_j \sin^2(\omega t_j). \quad (14a)$$

For the case where the uncertainties of the data are the same (i.e., w_j = constant) for all values and the data points are evenly spaced, it is easy to show that:

$$I(\omega) = N2^{-1}[a^2(\omega) + b^2(\omega)] = N2^{-1} c^2(\omega). \quad (14b)$$

This equation shows the 'multiplicative' effect of the periodogram [*Priestley*, 1981). In this section it will be implicitly understood that the coefficients $a(\omega)$ and $b(\omega)$, and their associated quantities $A(\omega)$ and $B(\omega)$, are independent because of their orthogonal properties and/or their independently determined degrees of freedom of the time series.

For the null case, where the Y_j consists of a sequence of independent random variables of zero mean and variance σ_y^2, the set of $A(\omega)$ and $B(\omega)$ of Equation (13) is a linear combination of the Y_j set and has a multivariate normal distribution. The set of $I(\omega)$ has a distribution which is proportional to χ^2 in 2 degrees of freedom. Thus:

$$I_p(\omega) = \sigma_v^2\, \chi_2^2 , \quad (15)$$

where the χ_2^2 distribution is a simple exponential distribution having a mean ν and variance 2ν. However, note that this is applicable only to the (maximum) number $N/2$ of independent $I(\omega)$, i.e., natural frequencies. *Schuster* [1898] tested the largest periodogram ordinate with the statistic γ:

$$\gamma = (I_p)_{max}\, \sigma_y^{-2} ; \qquad 1 \le p \le N/2 . \quad (16a)$$

In practical applications the variance σ_v^2 must be estimated, preferably with an unbiased estimate such as the expectation value, rather than with the sample variance s_y^2. For completeness, when the sample variance is replaced by $2s_y^2$, the resulting expression of γ can be recognized as the conventional Lomb-Scargle (LS) periodogram statistic [*Press et al.*, 1986], e.g.,

$$\gamma_{LS} = (I_p)_{\max} (2 s_y^2)^{-1} ; \qquad 1 \le p \le N/2 . \tag{16b}$$

Therefore the statistical test consists of checking whether or not the value of γ differs from zero at some significance level for the χ_2^2 distribution, i.e., whether or not all $I_p = 0$. The probability distribution of a χ_2^2 is a simple exponential function [*Hoel*, 1954]:

$$f(z) = 2^{-1} \exp(-z/2) . \tag{17}$$

Hence, for any value of $z \ge 0$, the probability that I_p/σ_y^2 does not exceed z is given by:

$$p[(I_p/\sigma_y^2) \le z] = \int_0^z f(x)dx = 1 - \exp(-z/2) . \tag{18a}$$

Under the null hypothesis that γ represents one of the $N/2$ independently distributed exponential variables, then for any z:

$$\begin{aligned} p(\gamma > z) &= 1 - p[(I_p/\sigma_y^2) \le z, \text{ for all } p] \\ &= 1 - [1 - \exp(-z/2)]^{N/2} . \end{aligned} \tag{18b}$$

This makes it possible to test whether or not the largest value in a periodogram is statistically different from a zero mean distribution with variance σ_y^2. If such a nonzero peak exists, the distribution is unlikely to be random, and this is the end of the test. To be significant, γ must exceed the critical test value of:

$$\gamma_c = -2 \ln(1 - P_o^{1/n}) . \tag{19}$$

where $n = N/2$, or the possible number of natural frequencies, and P_o is defined as the confidence level probability, 1 - p. However, this implies that the sample variance (s_y^2) is an unbiased estimate of the variance (σ_y^2). A better unbiased estimate of the variance can be obtained with the Parseval-Rayleigh's theorem, that is, the power is equal in the time and frequency domains:

$$\sum_{p=1}^{p=N/2} I_p = \sum_{i=1}^{i=N} Y_i^2 . \tag{20}$$

Accordingly, the expectation value of the left side is (for a zero mean value data series):

$$< \sum_{p=1}^{p=N/2} I_p > = N\sigma_y^2 . \tag{21}$$

Therefore, placing the above unbiased estimate of σ_y^2 into Equation (15a), we define g^* and g:

$$g^* = (I_p)_{\max}/\sigma_y^2 = \frac{(I_p)_{\max}}{N^{-1} \sum_{p=1}^{p=N/2} I_p} = gN . \tag{15$'$}$$

For large N, when the sampling fluctuations of the denominator are negligible, then g^* will asymptotically have the same distribution as γ of Equation (16a). This asymptotic distribution is *Walker's* [1914] large sample test for $(I_p)_{\max}$, with the same result as given in Equation (19). One should note that as the average power of a time series increases, as provided by the denominator of Equation (15b), the value of the statistic g^* decreases. This result is indifferent as to whether the increase in power is due to the presence of noise or of signal. In fact, a spectrum too rich in signals asymptotically reaches the statistical properties of a noise spectrum and limits the amount of significant information that can be obtained from it.

Fisher [1929] derived an exact test for g of Equation (15a) by geometrical arguments, for N odd, later analytically proven by *Grenander and Rosenblatt* [1957]. Fisher showed that the probability of the power at one frequency over the total power of the set can be expressed (relative to the arbitrary level z) by:

$$p[g > z] = \sum_{i=1}^{i=a} \frac{(-1)^{i-1} n!(1-iz)^{n-1}}{i!\,(n-i)!} , \tag{22}$$

where a is the next largest integer greater than z^{-1}. The first term of the expansion of Equation (22) can be recognized as the expansion of a simple exponential, such that for large values of n ($n >> 10{,}000$):

$$p[g^* > z] \approx 1 - [1 - \exp(-z/2)]^n , \tag{23}$$

which is the same result obtained in Equation (18b).

The critical value z is the highest value in the periodogram to be tested. For the approximation of Equation (23), at some critical value (P_c) of the confidence level probability the asymptotic value of z to be exceeded at that level is (see Equation (19)):

$$z_c = -2 \ln[1 - (P_c)^{1/n}] . \tag{24}$$

However, *Whittle* [1952] has proposed that the quantity g_2^* could be used to continue the significance test to lesser amplitude peaks than the highest:

$$g_2^* = \frac{I_{p=2}}{N^{-1}[(\sum_{p=1}^{p=n} I_p) - I_{p\max}]} , \tag{25}$$

where $I_{p=2}$ is the next smaller peak after $(I_p)_{\max}$. This approach can be continued to the next peak, etc., as long as the probability is both statistically significant and less than the probability of the earlier peaks. The appropriate changes to the value of n in Equations (22) and (23) are also required. This procedure will give an estimate of K, the maximum number of periodic components possible in Equation (1). An objection can be raised to Equation (25), in that the denominator does not preserve the variance of the sample. This can be remedied by our defining a new quantity g_m^*:

$$g_m^* = \frac{I_{p=m}}{N^{-1} (\sum_{p=1}^{p=N/2} I_p)} . \tag{26}$$

In this modified expression, m represents the lesser amplitude peaks after the first one. In effect, this modification states that all those peaks above the critical levels of Equations (19) and (24) are to be considered to be statistically significant. It also gives a measure of K, the maximum number of periodic terms in Equation (1).

4. Real Data Example

The use of periodogram techniques to assess the statistical significance of the periodicities that may be present in a time series is better shown using a concrete example. Specifically, an actual, noisy, irregularly spaced, real world experimental sample with missing points, has been used for this illustration. This sample consists of measurements of wind at 92 km height by a medium-frequency (MF) radar at Scott Base, Antarctica, during 11 days in August 1996. For the sake of convenience and generalization, the time scale of these measurements has been changed from hours to seconds. The route of using typical real data has been taken because the results were not known, thus avoiding their being prejudiced as they would be in a synthetic example, as well as representing the usual case an investigator must face. This real data sample is shown in Figure 1. It goes almost without saying that many a synthetic example has been used to develop, test, and refine the numerical procedures employed here. Detailed results of the methodology using synthetic examples is given in the Appendix. Synthetic examples have been used sparingly in the main text and only when deemed necessary for clarity of presentation (see Figure 13).

Figure 2 illustrates the periodogram for the data. The original data have been given the unequal spacing *Lomb* [1976] treatment described earlier and, for the time being, it is presumed that the time series measurements are independent (in the statistical sense). As noted earlier in the time series section, in the process of calculating the periodograms the time series has been rendered stationary by subtracting its mean value. Figure 2 (bottom) illustrates the Schuster χ_2^2 distribution significance test results and the conventional Lomb-Scargle test [*Press et al.*, 1986], utilizing the sample variance as an estimate of the true variance, while Figure 2 (top) gives the Fisher-g significance test using the power as an unbiased estimate of the variance. As the reader can readily see, the measures of statistical significance obtained by the Fisher and the Schuster-Walker methods are in excellent agreement, while the conventional Lomb-Scargle method of testing significance [*Press et al.*, 1986] shows strong disagreement with the other two. This discrepancy will be addressed at the end of this section. The significance tests shown have been made only for the largest periodogram ordinate. In Figure 2 the natural frequencies have been calculated, presuming that the longest period represents the fundamental frequency and the shortest period is twice the average separation between points. When the frequency of a 'hidden' periodicity falls between a pair of the natural frequencies, its amplitude is substantially reduced. For the extreme case, when the sought after periodicity falls midway between two neighboring ordinates, the height of such a hidden peak is reduced by 4 π^{-2} [*Whittle*, 1952].

The obvious solution to the apparent graininess in the results is to have measured more points in the original data, i.e., more independent frequencies. However, this is not always possible. Another approach, which is numerically correct, is to calculate the periodogram amplitudes at closer spacing than the natural frequencies. This approach, shown in Figure 3, illustrates the possible frequencies available; but the independence necessary for the statistical tests is lost, and the statistical tests can no longer be made. However, based on this artifice, it is possible to redefine the fundamental frequency so as many as possible of the desired frequencies are included in the correct natural harmonic frequencies. After all, a priori, we do not know the real frequencies which are present in the data. This redefinition of the fundamental frequency can be achieved by cropping the data down to a suitable length. This is the approach which has been taken in the determination of the critical statistical level of Figure 3 (top). The significance tests in Figure 3 (bottom) are the Schuster-Walker and the conventional Lomb-Scargle [*Press et al.*, 1986] levels, respectively. For consistency of presentation, the previous result is shown again.

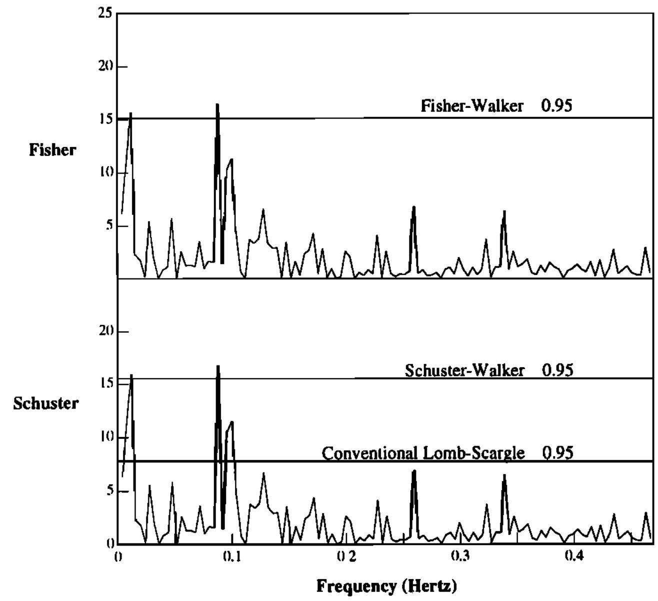

Figure 2. Periodogram of the data of Figure 1, where the statistical significance levels are shown for the (top) Fisher-g [*Fisher*, 1929] test and (bottom) both the *Schuster* [1898] and the conventional Lomb-Scargle [*Press et al.*, 1986] tests. The natural frequencies are shown.

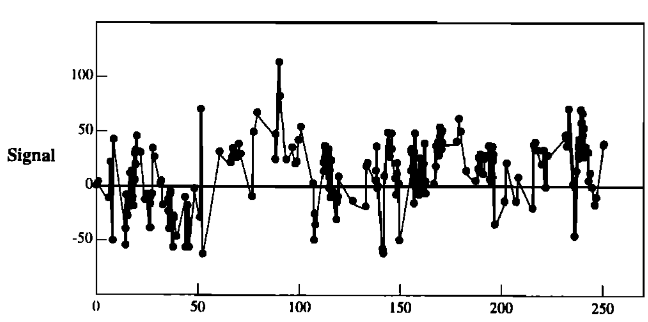

Figure 1. Noisy and unevenly spaced experimental data chosen for examination of hidden periodicities.

Since the conventional Lomb-Scargle statistical significance test [*Press et al.*, 1986] of the periodogram gives substantially different results than the Schuster-Walker and Fisher tests, the reason for this discrepancy requires further investigation. Because the three tests use the same basic quantities, one must look into the derivations of these statistical significance tests. The derivations for the Schuster-Walker and Fisher methods have been given previously; therefore those used in the conventional Lomb-Scargle will be examined now. Following *Kendall* [1948], the variance of a_k of Equation (9a), for equally weighted and equally spaced points, can be shown to be equal to:

$$s_{a_k}^2 = 4\sigma_y^2 T^{-2} \sum_{j=1}^{j=N} \cos^2(2\pi k T^{-1} t_j) = 2\sigma_y^2 T^{-1} , \quad (27)$$

and the variance for b_k is similarly obtained. Note that in the present derivations, $T = N$. From the definition of the normal distribution one finds the frequency function for two independent variables to be:

$$\begin{aligned} dF &= [\sigma(2\pi)^{1/2}]^{-1} \exp[-x^2(2\sigma^2)^{-1}]\,dx \\ &= (N^{-1}4\pi\sigma_y^2)^{-1/2} \exp\{-[a_k^2+b_k^2](4\sigma_y^2/N)^{-1}\} \\ &\quad \times da_k db_k . \end{aligned} \quad (28a)$$

From the definitions given in Equations (14b) and (21) one can recognize, using the different estimates of σ_y^2, that the argument of the exponential is:

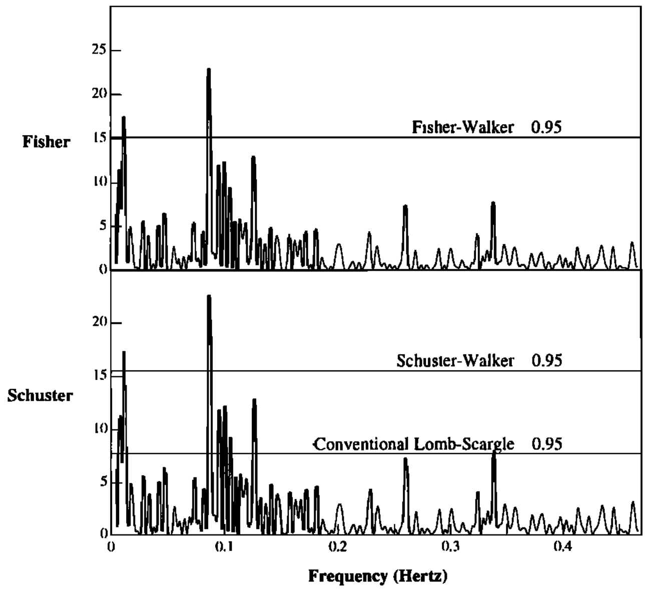

Figure 3. Same as Figure 2, except that many more frequencies than the natural frequencies have been calculated.

$$[a_k^2+b_k^2](4\sigma_v^2/N)^{-1} = I_p/(2\sigma_v^2)$$
$$= I_p(2N^{-1}\sum_{p=1}^{p=N/2} I_p)^{-1} , \quad (28b)$$

where the last term is half the value given in Equation (15b). A reader can recognize the center expression as the conventional Lomb-Scargle statistic, when σ_v^2 is replaced by s_v^2 [see Equation (16b)]. Using the best estimate of the variance, the appropriate Lomb-Scargle statistic to test is, therefore:

$$I_p(2N^{-1}\sum_{p=1}^{p=N/2} I_p)^{-1} \leq \eta , \quad (28c)$$

where η is an arbitrary number. Following *Scargle* [1982], when the a_k and the b_k of Equation (28a) are not orthogonal and independent, they do not have the same variance, and the probability is a simple exponential. That is, the probability that the statistic will not exceed this value is $P=(1-e^{-\eta})^{N/2}$ for all the possible frequencies. From this, it is simple to show that:

$$\eta_c = -\ln(1-P^{2/N}) , \quad (29)$$

which is to be compared with γ_c of Equation (19) and z_c of Equation (24) and is the result given by *Press et al.* [1986] for the conventional Lomb-Scargle statistical significance test. Since the statistic in Equation (28c) and the test of Equation (29) are half the value of the relevant statistic and test in Equations (15b) and (24), the resultant significance tests are seen to be the same as the other two tests, albeit with a scale change. The results of the significance test just described are shown in Figure 4, indicating the equivalence of these two periodogram tests. It should be pointed out that this factor of 2 discrepancy applies strictly to the (now corrected) Lomb-Scargle statistic calculated using the expression of Equation (28c).

This approach does resolve the discrepancy between the conventional Lomb-Scargle and both the Schuster-Walker and Fisher significance tests illustrated in Figures 2 and 3. Note, however, that it has now become necessary to calculate the complete natural frequency periodogram in order to obtain the correct Lomb-Scargle results, which is not the case in the *Press et al.* [1986] approach. As noted earlier, the variance s_v^2 used in the conventional Lomb-Scargle expression of Equation (16b) will not necessarily have the same numerical value as σ_v^2 derived from the expectation value in Equation (21), because of sampling fluctuations. Therefore a simple correction to the *Press et al.* [1986] conventional Lomb-Scargle statistic is not the answer since, for typical small data samples, the Lomb-Scargle statistic calculated using the two different approaches will not necessarily have the same value. In addition, the Lomb-Scargle approach in the *Press et al.* [1986] recipe incorrectly implies that any number of periodogram frequencies are possible. It is for this reason that the appropriate (natural frequency) statistic of Equation (28c) has been used here, and it is strongly suggested for use.

Because of the near equality of the three statistical significance testing methods (Schuster-Walker, Fisher, and the corrected Lomb-Scargle), the more exact *Fisher* [1929] method will be used in the following text. It should be noted that regardless of which significance test is employed, the problem of true degrees of freedom, which define the actual number of independent frequencies possible, has to be faced.

5. Degrees of Freedom

As remarked earlier, the sample variance is used in the Schuster-Walker and conventional Lomb-Scargle method as a measure of the true variance in order to carry out the χ_2^2 test, while the Fisher test uses the power as an unbiased estimate for the variance. The sample variance is normally an optimistic estimate of the true variance, as it uses the number of data points (N) in the sample, rather than the number of degrees of freedom ν [*Priestley*, 1981]. Because of serial coherence in the original data, the number of points is seldom an appropriate estimate of the true number of degrees of freedom, as the samples may be drawn at time intervals too short to be independent [*Leith*, 1973; *Harrison and Larkin*, 1997]. Of course, as the sample size becomes large, the sample variance will asymptotically reach the

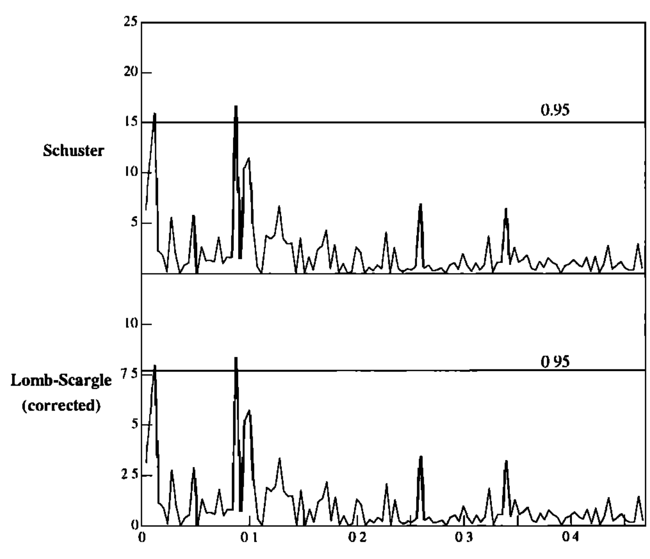

Figure 4. Statistical significance of the Schuster-Walker and the corrected Lomb-Scargle periodogram test methods. The statistic of Equation (28c) was employed for the Lomb-Scargle test. Note the equivalent results for the two methods. The natural frequencies are shown.

true variance. Thus, in particular for small samples, it becomes necessary to obtain an estimate of the number of the sample's degrees of freedom in order to have a proper estimate of the true variance. An estimate of the number of degrees of freedom can be obtained from the data by deriving the time between independent observations. This requires a knowledge of the autocorrelation of the data under consideration, which conveniently is available as the Fourier transform of the already determined power spectrum or periodogram. A simple Fourier transformation of the latter spectrum provides the desired information. Figure 5 shows the autocorrelation of the experimental data of Figure 1 near the first zero crossing.

The two common techniques to estimate the time between independent observations [*Harrison and Larkin*, 1997] are the use of the first zero crossing of the autocorrelation of the data and the *Leith* [1973] technique. The Leith method shows that a data series with variance σ^2, when filtered with a running mean containing M points, is expected to have a variance $\sigma_M^2 = \sigma^2 \nu^{-1}$. Here ν is the effective number of independent points which are averaged together. The expression given by *Leith* [1973] and reported by *Harrison and Larkin* [1997] for a sampled series is:

$$\sigma_M^2 = \sigma^2 \nu^{-1} = \sigma^2 M^{-1}\left\{1 + 2\sum_{n=1}^{M}[1-nM^{-1}]r_n\right\}, \quad (30)$$

where r_n are the elements of the autocorrelation of the data. The time between independent observations can then be derived [*Harrison and Larkin*, 1997]:

$$\tau = T\nu^{-1} = \left\{1 + 2\sum_{n=1}^{M}[1-nM^{-1}]r_n\right\}. \quad (31)$$

The number of degrees of freedom in the data is equal to the data length T divided by the time between independent observations. Estimation of the time between independent observations by the first zero crossing occurrence is an approximate, safe, and, usually, conservative overestimate. The *Leith* [1973] method is a more direct estimate and it also lends itself to a confidence interval calculation [*Harrison and Larkin*, 1997]. It should be noted that the autocorrelation of a time series consisting of random noise, i.e., having zero-mean amplitude with variance σ^2 and random phase, will have its zero crossing near the first lag. Then the degrees of freedom are therefore equal to N, the number of points in the sample.

A third method to estimate the number of degrees of freedom can also be directly derived from the power spectrum of the time series. Here one makes use of the property that a running mean operation is a convolution in the time plane, which in the frequency plane transforms into a product of a sinc function, i.e, $\mathrm{sinc}(x) = [\sin(\pi x)]/(\pi x)$ [*Bracewell*, 1965], with the amplitude function. This can be trivially extended to the periodogram ordinate. As given in Equation (21), the sum of the power is an unbiased estimate of the variance; thus the sum of the power obtained after the running mean averaging will be smaller by the number of independent points which are averaged together. Therefore, the ratio of the power before application of the running mean operation over the power after the running mean will be a measure of the number of these independent points in the *Leith* [1973] sense.

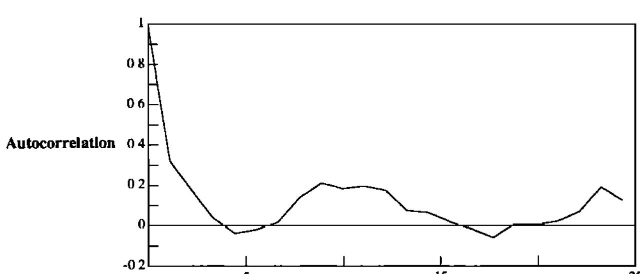

Figure 5. Autocorrelation plot of the time series given in Figure 1. Only the relevant information near the first zero crossing is shown.

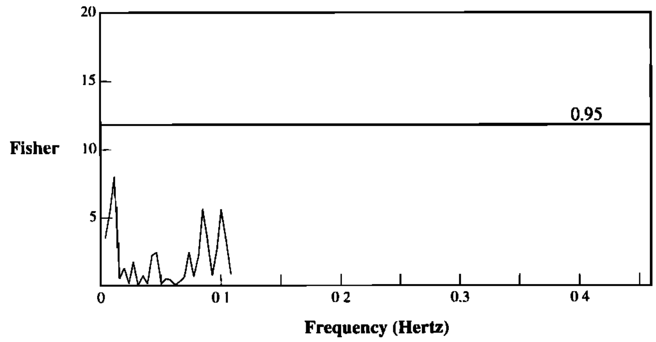

Figure 6. Periodogram for the natural frequencies available when only the true number of degrees of freedom of the time series of Figure 1 is taken into account. Same frequency scale as Figure 2 is used in order to emphasize the true information content of the time series. See text for details.

Application of these methods to estimate the time between independent observations in the present data set, using the autocorrelation results shown in Figure 5, gives coherence times of 3.5 s and 4.44 s for the zero crossing and *Leith* [1973] methods, respectively. These coherence times translate into 73 and 58 degrees of freedom, which are to be contrasted with the original 235 points in the data sample. Therefore the critical level used in the significance tests has been overestimated for all the methods. Moreover, the limited number of degrees of freedom has automatically limited the number of independent (natural) frequencies possible in the periodogram. This has also automatically lowered the value of the Nyquist frequency. This result shows the fundamental limitation presented by the existence of a finite serial coherence time, namely, that the highest independent periodicity observable in a time series is roughly twice this serial coherence time. This result is independent of the length and the fineness of the time grid of the time series. Any measurements made closer than the serial coherence time simply lead to oversampling, and no new information is acquired.

Figure 6 gives a new calculation of the periodogram for the real data of Figure 1 using the 29 frequencies actually available. Beside the coarseness of the spectrum, the natural frequencies no longer coincide with the spectral features, and the frequency range is considerably smaller than seen in Figure 2. This is a typical example of the limitations found in real world data, where the amount of information available for proper statistical testing is generally small. One should remark that the finite amount of this information is an intrinsic and fundamental property of the data, not of any process with which they are handled.

Earlier, we discussed the method of redefinition of the fundamental frequency as a means to include desired frequencies as harmonics of this fundamental period. Figure 7 shows the improved results obtained by doing such an adjustment. Notice

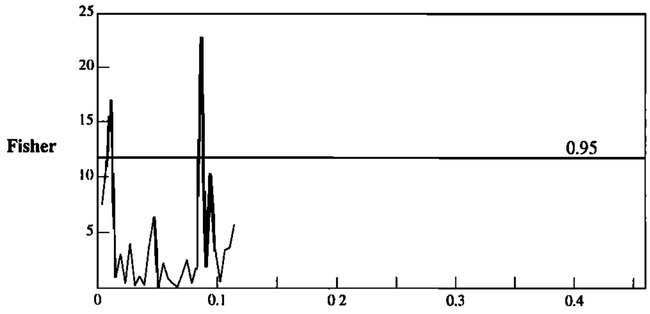

Figure 7. Periodogram with the number of natural frequencies allowed by the available degrees of freedom. The fundamental frequency has been redefined by cropping the original data 8.1% in its time length. Compare with the periodogram in Figure 6.

the change in the power near 0.1 Hz. These results are obtained by cropping the data by a small adjustment (8.1%) in their time length. Needless to say, this process can not be carried out more than once, as the resulting determined frequencies would no longer be independent. Again, note that the critical level of significance in Figure 7, relative to Figure 2, has decreased because of the smaller number of degrees of freedom involved.

6. Tapering

In Section 2 we mentioned the use of 'tapering' of a given time series in order to minimize edge effects of the limited length of the data series. This operation consists of reshaping the original time series with a smooth and symmetrical 'window' function which smoothly decreases, or tapers, from unity value at the middle to a zero value at the extremes of the data. In addition, the application of this symmetrical taper window enhances the cyclical behavior implicitly present in Fourier analysis. 'Prewhitening' or the controlled addition of noise in order to decrease the dynamic range of a data set [*Blackman and Tukey*, 1959; *Percival and Walden*, 1993], is often used in conjunction with tapering.

Tapering can be better understood in terms of Fourier transforms, where the tapering consists of the multiplication of the time series by the window function in the time plane. This multiplication translates into a convolution of the two transforms of these functions in the frequency plane [*Bracewell*, 1965]. The convolution process has the effect of broadening (and smoothing) the resultant frequency plane function. Therefore the oscillatory character of the transform of the time series function which has been sharply cropped at the edges (i.e., having a limited length) becomes spread over frequency, and less noticeable, by the convolution operation. Details on tapering are found in most statistical references [*Blackman and Tukey*, 1959; *Priestley*, 1981; *Percival and Walden*, 1993], where they are discussed in terms of the (original function) Fejer kernel, its manipulation, and the resultant effects in the analysis. Here we will examine the effects that the tapering process has on our data sample.

As expected, there exist a large number of window functions and figures of merit associated with them [see *Blackman and Tukey*, 1959; *Priestley*, 1981; *Press et al.*, 1986; *Percival and Walden*, 1993], and they all accomplish about the same result [*Press et al.*, 1986]. A recommended [*Press et al.*, 1986] practical window is the Welch window:

$$\tau_{x_i} = 1 - \left[(x_i - x_1 - x_c)/x_c\right]^2 \;;\; x_c = (x_N - x_1)/2, \quad (32)$$

where x_i is the abscissa and N is the number of points. Figure 8 shows the window and its product with the original data given in Figure 1. The decrease in amplitude at both ends of the data is noticeable. It must be remarked that diminishing the amplitude of the original information inevitably throws away some information.

Since the shape of a Welch window is fixed, it is useful to have a window whose slope and one-half power points are arbitrarily selectable. A hyper-Gauss window function is a good example of a variable shape window:

$$\tau_x = \exp(x/G)^{2M} \;;\quad G = x_{1/2}/(\ln 2)^{-1/2M}\,. \quad (33)$$

Here $x_{1/2}$ is the one-half power point and M is the (integer) number which defines the slope. The value of M is found by arbitrarily defining a value of x where the function has an arbitrary fractional transmission relative to the peak value. An example of this window is illustrated in Figure 9. As can be seen, by the appropriate choice of one-half power point and large values of M, this window will asymptotically approach a rectangular window. This hypothetical window becomes equivalent to the sampling window used to choose the sample series, i.e., no tapering. The hyper-Gauss window is useful since it has no ringing of its own to contribute to the final spectrum. By adjusting the one-half power width and the value of M so that ringing is minimal, one can optimize for a minimum loss of data when using this window.

The number of available degrees of freedom in the time series shown in Figures 8 and 9 is relatively unchanged from that of the original series, namely 57 and 56 for the Welch and hyper-Gauss windows, respectively. The periodograms derived from these data are given in Figures 10 and 11, respectively, where the time series length has been cropped so that the data frequencies match the periodogram natural frequencies. These periodograms are similar to the periodogram given in Figure 7, except for the apparent increased statistical significance at the lower frequency. This is specially obvious in the Welch window case. Closer examination, by calculating synthetic spectra, shows that the use of a tapering window changes the periodogram power distribution as a function of frequency and phase. This power distribution change is noticeable as an enhancement of the relative power in the first few harmonic terms, typically up to the tenth term for a Welch window. Although the use of these tapering windows has not noticeably altered the number of degrees of freedom, it has changed the distribution of power with frequency, leading to

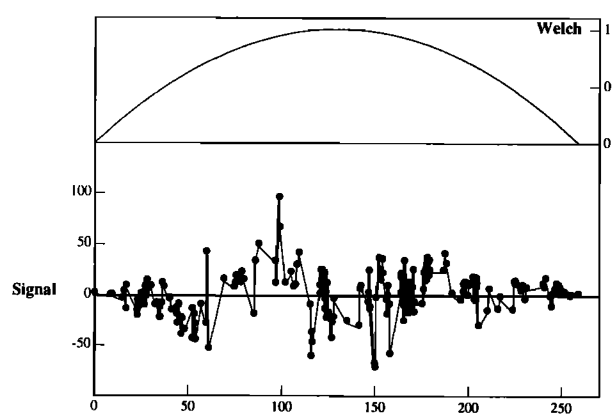

Figure 8. Welch window and the original time series data of Figure 1 after their multiplication. Note the loss of amplitude in the data as the edges are approached.

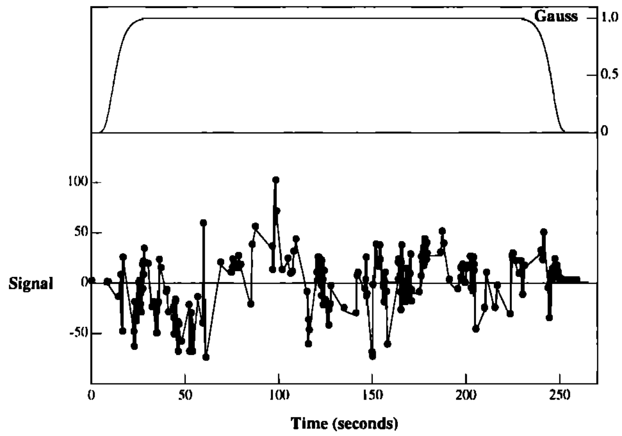

Figure 9. Hyper-Gauss window and the original time series data of Figure 1 after multiplication by this window.

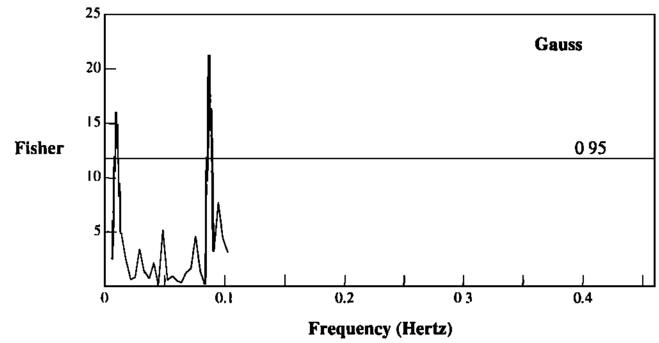

Figure 11. Periodogram using the number of natural frequencies allowed by the available degrees of freedom with a hyper-Gauss window. As in Figure 7, the fundamental frequency has been redefined by cropping the data length.

results that incorrectly appear either to contain more statistically significant features than were present in the original data or to enhance those at the lower frequency region.

Pre-whitening after data have been obtained is to mainly avoid difficulty with the minor lobes of the spectral windows [*Blackman and Tukey*, 1959]. Effectively, the desired result is to change the original process distribution so that it has a 'nearly white' spectral density function [*Priestley*, 1981]. This can be accomplished by suitably filtering the signal, which may include the addition of white noise to the original information. Filtering, in this context, is understood to be a change of the relative contributions of the different portions of the spectrum. Depending on the character of the (arbitrary) filtering applied to the data, pre-whitening can either reduce or increase the serial coherence of the data, when compared with the original sample. Associated with such manipulation is the consequent increase or decrease in the degrees of freedom. As will be seen in Section 7, such arbitrary changes to the character and magnitude of the sample variance lead to unpredictable results, which cannot be statistically supported.

A secondary, but interesting, property of strong and symmetrical taper windows, such as the Welch window, is that they tend to enhance the cyclical behavior implicit in Fourier analysis. By enhancing such behavior of a time series, a given realization can sometimes approach stationarity to at least order 2. This is a necessary requirement for the statistical investigation of the data.

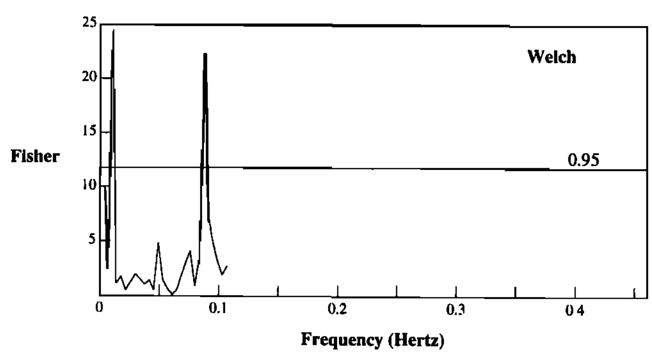

Figure 10. Periodogram using the number of natural frequencies allowed by the available degrees of freedom with a Welch window. Note, in particular, the relative enhancement of the power at the lower frequency, when compared to Figure 7. Here, as in Figure 7, the fundamental frequency has been redefined by cropping the data length.

However, the previously discussed difficulties with the use of taper windows remain.

7. Filtering and 'De-trending'

In the context of the periodogram material previously discussed, filtering amounts to the enhancement of a region of the spectrum relative to the rest of the spectrum. This filtering operation changes the relative power of a given feature in the spectrum and does not account for such a change. At the same time, filtering can change the serial coherence time of the data under consideration. As would be expected, the periodogram of such manipulated data would be quite different from that periodogram obtained from the original data. However, it is simply invalid to assign to the original data the results of significance tests performed on data whose power spectrum has been arbitrarily manipulated. In fewer words, the tests for statistical significance are no longer being made on the original data and therefore are not relevant to them.

A simple example of filtering is the use of a band-pass filter which selects a small region of a power spectrum and rejects all other information below and above the selected region, as shown in Figure 12. The selected region still only contains the same information and the same statistical significance it had when the full original spectrum was available, and no new information is gained by this filtering. In fact, if anything, both information and degrees of freedom would have been lost by the use of the filtering process. Filtering, such as the process described here, ignores the central thesis of significance testing; these tests clearly specify that the significance of any given feature's power must be measured relative to the rest of the power spectrum, as given in Equation (15b). Arbitrarily disposing of (or ignoring) the power associated with the rest of the spectrum as undesirable simply invalidates any significance tests performed on the filtered data. In other words, the statistical significance tests are not being performed on the original data and are no longer relevant to them. The results shown in Figure 12 can be obtained by either the product of the original spectrum and the filter transfer function or by only calculating the spectrum for the region of interest, which requires calculating the full power spectrum anyway, as given in Equation (15b). The results obtained by either of these methods are indistinguishable from each other.

Filtering is a concept which is more applicable to the scenario in which data are being obtained continuously and the signal-to-noise ratio is continuously increasing. After some time has elapsed, an arbitrary coherent frequency present in the informa-

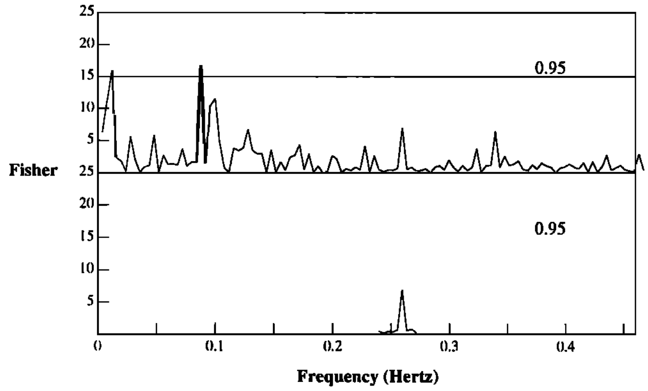

Figure 12. Periodograms of the data of Figure 1, before and after filtering with a square filter of 0.025 Hz bandwidth centered at 0.26 Hz. Note that the information available on the feature near 0.26 Hz has not changed after the arbitrary filtering operation. The 0.95 significance level for the upper periodogram for the natural frequencies, presuming all data points are independent, is also shown.

tion stream reaches a desired statistically significant level. This time is practically provided by the 'time constant' of the filter and/or the coherent integration of the signal. However, when one has a fixed total measurement interval, the information is frozen within the data and no further knowledge is available (unless new information is provided, usually in the form of assumptions, etc.).

Although de-trending data as part of statistical analysis is not normally considered a filtering operation, its effects on the results can be substantial and very similar in character to those occurring in filtering. A trend is normally described as an amplitude variation in the time series which has a periodicity much longer than the length of the available data, and de-trending consists of arbitrarily removing this signal [*Blackman and Tukey*, 1959]. Therefore de-trending becomes a specialized method of filtering, with all its associated pitfalls.

In principle, it is possible to separate a 'trend' from the underlying information as long as the form of the trend is known. There arises the immediate question as to how a statistical test of significance performed on de-trended data relates to the original data. Since the power associated with the trend has been removed, one is no longer measuring the significance of a given feature relative to the full spectrum. Thus the results no longer refer to the original data and are likely to be invalid when applied to it. In addition, the de-trending process may substantially change the serial coherence of the signal and the associated number of degrees of freedom.

The similarity between filtering and de-trending can be easily shown by subjecting a synthetic time series to a de-trending operation. For this example we will employ the equal-amplitude six-frequency time series used in the Appendix. The de-trending operation used here is to 'difference' the time series [*Priestley*, 1981; *Percival and Walden*, 1993]. Differencing amounts to the calculation of the first difference of the data, which in effect obtains its first derivative.

From Fourier transform theory [*Bracewell*, 1965] it is known that the derivative of a signal in the real plane translates in the transform plane into the product of the original signal transform times the imaginary frequency abscissa. Using the symbol $\supset$ to indicate the Fourier transform operation in which a real plane function $f(x)$ 'transforms into' $F(s)$ in the Fourier plane, we now have [*Bracewell*, 1965]:

$$f(x) \supset F(s) \quad , \tag{34}$$

$$f'(x) \supset i2\pi sF(s) \quad , \tag{35}$$

where $f'(x)$ denotes the first derivative of $f(x)$, $i = (-1)^{1/2}$ and s is the Fourier plane abscissa. The latter is identified with frequency when x is in time units.

Equation (35) clearly shows that a differencing operation depresses the amplitude of the low-frequency components and increases the amplitude of the high-frequency components of the original signal's spectrum. This operation conforms to the general definition of filtering as an enhancement of a region of a spectrum relative to the rest of the spectrum. The upper panel of Figure 13 shows the periodogram of a raw synthetic series. Although the raw synthetic time series consists of six equal-amplitude oscillations (see FigureA2(a)), the periodogram shows two of these frequencies as missing because they are outside the Nyquist limit determined from the serial coherence of the data (see Figure A2(b)), and the surviving four frequencies as having unequal power. This latter effect is caused by the mismatch between the frequencies of the oscillations and the natural frequencies of the periodogram. The lower panel of Figure 13 shows the periodogram of the differenced synthetic time series. As expected from the derivation of Equation (35) the low frequencies of the differenced periodogram have been depressed while its high frequencies have been enhanced. At the same time the number of degrees of freedom of the 'differenced' series has increased, as can be seen by the increased frequency range of the lower panel periodogram in Figure 13, and the appearance of the sixth frequency with exaggerated power and enhanced noise.

The results shown in Figure 13 clearly show the changes introduced by the difference procedure employed in de-trending. Three of the known frequencies known to be statistically significant in the original data have disappeared, while two new frequencies not supported by the serial coherence of the original time series have appeared because of the increased degrees of freedom created by the differencing operation. Clearly, the lower panel periodogram in Figure 13 bears little similarity to the original raw synthetic data periodogram and cannot be described as representing the spectrum of the original time series. Thus, the statistical results derived from the examination of a differenced signal no longer refer to the original signal and therefore are not relevant to it.

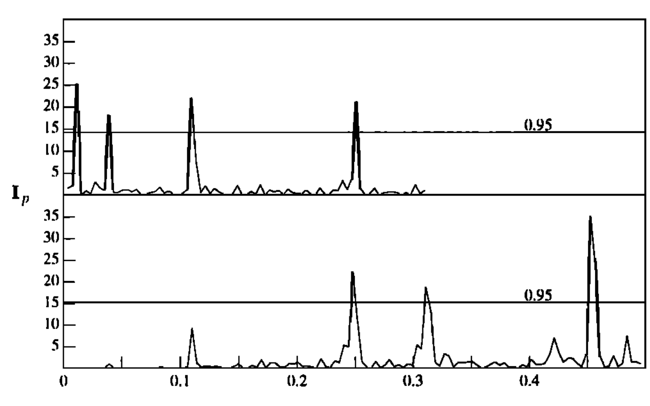

Figure 13. Periodograms before and after de-trending a synthetic time series by differencing. The obvious change in the spectral character and degrees of freedom is noticeable. The 0.95 significance level is shown.

The same difference operation can be applied to the MF radar data shown in Figure 1. The results of the analysis are given in Figure 14. The upper panel is the original data periodogram (see Figure 7), while the lower panel shows the periodogram of the differenced time series. The differencing process has eliminated the significant features existing in the original time series. These two examples shown in Figures 13 and 14 clearly show the perils inherent in the manipulation of data when attempting to de-trend a time series. Although differencing is only one of the many methods used in de-trending data, the other methods lead to questionable results similar to those obtained in the examples given here. The de-trending operation is basically a specialized filtering operation which will likely lead to incorrect results when applied to the original data.

Both filtering and de-trending are simple data manipulations, in which power is arbitrarily removed from, added to, or shifted within the spectrum of the original information and where no accounting is made for this change. Not surprisingly, the results after any of these manipulations can provide almost any desired answer, which cannot be supported with a critical statistical test performed on the original data. Clearly, any manipulation of the original data which arbitrarily changes the character and magnitude of the sample variance will lead to unpredictable results which cannot be statistically supported. Therefore, in general, the statistical results obtained from a data set which has been filtered and/or de-trended should be suspected of being unreliable.

8. Summary and Discussion

In this study we have investigated in detail the requirements for obtaining statistically significant information on coherent oscillations present in a time series by using periodogram methods. To accomplish this, a background study of the requirements and mathematical techniques utilized to obtain a periodogram from a time series has been investigated, coupled with a study of the statistical tests employed to extract statistically significant information from the periodograms. The basic statistical tests are applicable only for the largest periodogram ordinate, while the *Whittle* [1952] suggestion (see Equation (25) and our modified result of Equation (26) allow testing of the next largest ordinates and therefore make it possible to estimate the maximum number K of (statistically significant) periodic components with which to represent the time series.

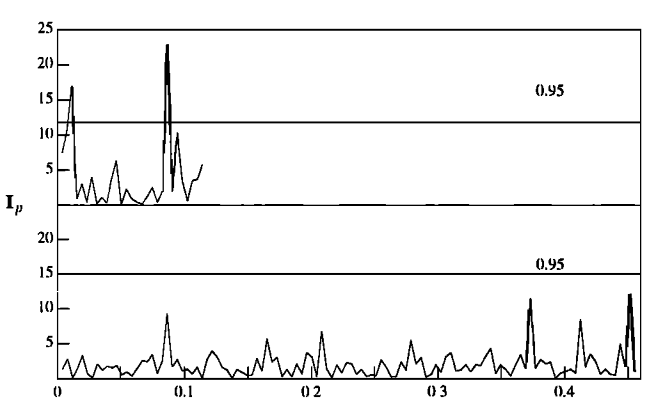

Figure 14. Periodograms before and after de-trending the time series of Fig. 1 by differencing. Again, the obvious change in the spectral character is noticeable. In the upper panel the fundamental frequency has been redefined by cropping the data length (see Figure 7).

For illustration, these mathematical methods have been applied to a real-world geophysical time series, but with the implicit presumption that all the individual measurements in the time series are independent. In this process a discrepancy between the Fisher and the Schuster-Walker significance tests and the conventional Lomb-Scargle [*Press et al.*, 1986] significance test has been found and resolved. It is reassuring that these three tests, based on nearly the same assumptions, now give essentially the same measure of statistical significance. However, the resolution of this discrepancy requires using the statistic given in Equation (28c) for the Lomb-Scargle test, rather than the *Press et al.* [1986] formulation for this statistic.

The presumption that individual measurements in a time series are indeed independent deserves to be closely investigated since the statistical significance tests are only applicable to those periodogram frequencies which are truly independent. Thus one must ascertain the actual number of independent samples, or number of degrees of freedom ν, present in the time series. The number of degrees of freedom is determined from the serial coherence, or 'memory', of individual data points about previous events in the time series. This information is available from the autocorrelation function of the time series at the time its first zero crossing occurs, which is called the serial coherence time of the process, and defines the time between independent observations. The number of degrees of freedom is simply given by the ratio of the time length of the series divided by the (determined) serial coherence time. The fundamental 'natural' frequency is defined as that frequency corresponding to the period associated with the length of the time series, and the harmonics are evenly spaced up to the Nyquist limit, as given by the $\nu/2$ possible frequencies.

Inherent to this discussion of natural or independent frequencies is the appearance of the concept of a critical sampling rate in the collection of data, defined by the serial coherence time. Data taken at sampling rates faster than this critical sampling rate are not independent and carry no new information with them. The parallel interpretation of a critical sampling rate is that a time series is intrinsically band limited as given by the Nyquist frequency associated with the serial coherence time. Finding this critical band-limiting frequency defines a central and necessary statistical requirement for the efficient design of an experiment. Obviously, the band-limiting frequency defines the shortest periodicity available for investigation in the time series. The above arguments are applicable when the number of degrees of freedom is smaller than the number of measurements in the time series; that is, the information is overdetermined by the sampling process represented by the time series. As an illustration of these deductions, the experiment from which our time series example was extracted would have been better served if the data rate had been halved and the gaps filled instead.

Once the appropriate natural frequencies of a time series are determined (based on the serial coherence time of the data) and the periodogram is obtained, the statistical significance tests can then be applied in order to find their statistical validity. The Schuster-Walker and the (corrected) Lomb-Scargle tests are asymptotic tests more applicable to large samples, while the Fisher test is useful for any sample size and is the preferred statistical test to use. Since one does not know a priori what frequencies may be present in the time series, it is possible that they may not be the same frequencies as the natural frequencies used in the periodogram. Redefinition of the fundamental frequency, by cropping of the length of the time series so that one (or more) of the natural frequencies coincide with the frequencies present in the time series, presents a solution for this frequency mismatch. The price paid for this solution is the loss of some information forever. Obviously, this cropping can only be done once, lest the required independence is lost. Although this cropping of the length of the time series to match frequencies appears to be arbitrary, it is no more arbitrary than the selection of the length in the original time series before it was to be investigated for statistical significance.

The tests used to search for statistical significance have a clear and strong message: the significance of any feature of the power spectrum derived from a time series is relative to the total power of the series. Any manipulation that changes, shifts, removes, or increases the power in one region of a power spectrum, relative to another region, will alter the significance of the results obtained. Basically, such manipulation arbitrarily changes the character and amplitude of the sample variance and, thus, the results. There are three mainstream procedures that perform such an operation: filtering, tapering (which includes pre-whitening) and detrending. As shown in the text, all of these procedures preferentially and arbitrarily change the power in a portion of a power spectrum, and yet give no account for this power manipulation. These manipulations usually also change the number of degrees of freedom from those available in the original sample. Carrying out statistical analysis on a periodogram from data which have been tampered with in this fashion will give almost any desired answer. In particular, if a feature is not found to be statistically significant in the original power spectrum, it should not acquire any further significance by frequency dependent manipulation of the power spectrum. To assign to the original data the results of significance tests performed on data whose power spectrum has been arbitrarily manipulated is simply invalid. These significance tests are no longer being made on the original data and are no longer relevant to them.

Although the use of periodogram techniques sacrifices the phase information in order to determine statistical significance, this information is still available for those frequencies whose power has tested to be statistically significant. If anything, the phase information is made more significant by its being associated with a frequency whose power has been determined (statistically) to exist.

In conclusion, the present study has rigorously investigated the process of determining the presence of meaningful, statistically significant, coherent oscillations in time series when using periodogram techniques. As part of this process, it has been shown there exist both fundamental limitations on the extent of available information present in a given time series and also pitfalls to be avoided. In particular, it has been shown that some of the commonly used data manipulation methods can lead to incorrect and misleading interpretations, since they can alter the information contained in the original data to such an extent that the final results obtained are no longer relevant to the original data. The methodology described here has been employed on a real sample of upper atmosphere wind measurements in order to illustrate the extent of extractable meaningful information, as well as to suggest a direction for the statistical analysis of the large existing collection of these measurements. This approach would also be applicable to many other available data collections such as incoherent radar information, satellite observations of atmospheric properties, meteorological records, solar plasma measurements, etc.

Besides the periodogram, there exist other methods to extract information from a time series. These methods and the criteria for their significance are being investigated, and the results from this study will be reported later.

Appendix

In this appendix we use synthetic data in order to illustrate further the properties and behavior of periodograms when employed in the search for statistically significant coherent oscillations in a time series. Examining a synthetic time series, whose properties are known in advance, makes it possible to show clearly some of the results derived in detail in the preceding main text.

The specific synthetic time series to be employed in the following text consists of six equal amplitude frequencies randomly distributed across the spectrum, with the arbitrary addition of Gaussian noise of known variance. The resultant time series looks, to the eye, very similar to the real experimental data which are found in the aeronomical studies with which this author is associated. Further, the amplitudes at these frequencies will be purposely changed to illustrate the sometimes profound differences that these variations in amplitude can have in the final results.

The starting specific synthetic series is 256 seconds long with a sampling rate of one per second, and consists of six oscillations of 80-, 25-, 9-, 4-, 3.2-, and 2.2-s periodicity. As previously stated, all these periodicities begin with equal amplitude, arbitrarily set equal to 12 units. Finally, Gaussian random noise of zero mean and 15 unit standard deviation noise has been added to the series made up from these six frequencies. The top panel of Figure A1 shows the original equal-amplitude series, while the lower two panels illustrate an arbitrarily increasing amplitude of oscillation of the 80-s periodicity. The applicable amplitude in the middle panel is given as a(80): 18, indicating that the 80-s periodicity oscillation amplitude is 18 units, increasing to 25 units in the lower panel.

Referring to Figure A1, the reader can see the obvious presence of a lower-frequency oscillation, as well as an increase in its amplitude. Figure A2 shows the periodograms for Figure A1.

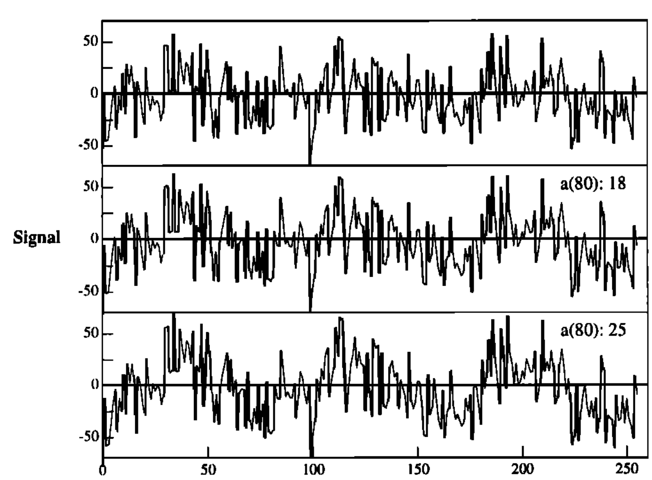

Figure A1. Synthetic series generated with six frequencies and the arbitrary addition of Gaussian noise of known variance. The upper panel has all six frequencies of equal amplitude (12 units), while the middle and lower panels have increasing amplitude for the oscillation with 80-s periodicity, as indicated. The added Gaussian noise has a zero mean and a standard deviation of 15 units.

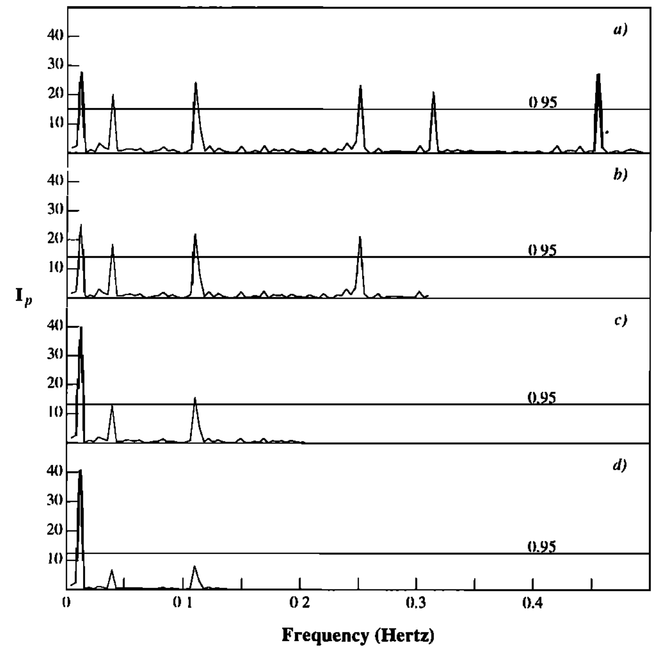

Figure A2. Periodograms for the time series given in Figure A1. Panels *a*) and *b*) refer to the equal-amplitude series of the upper panel of Figure A1. Panel *a*) presumes all the 'measurements' in the series are independent. *b*) uses the number (162) of degrees of freedom derived from the serial coherence of the data. Panels *c*) and *d*) apply to the middle and lower panels of Figure A1, where the amplitude of the lowest frequency is arbitrarily increasing. Note the decrease in the number (106 and 76) of the degrees of freedom available and consequent decrease in the spectral region available for examination. The Fisher 0.95 confidence significance tests are shown.

The Fisher statistical test for 0.95 significance is shown for the applicable number of degrees of freedom. Figure A2*a*) illustrates the periodogram when the assumption is made that all the 'measurements' making up the series are independent. The six frequencies making up the time series are clearly visible. However, their calculated power appear to be uneven. This is no surprise, since the periodogram-calculated frequencies do not exactly match those of the six oscillations present. Figure A2*b*) shows the limited spectral range which can be examined because of the limited number (162) of available degrees of freedom when the

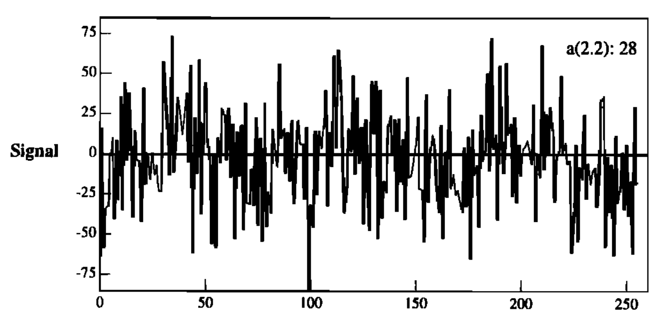

Figure A3. Six-frequency synthetic series, where the first five oscillations have equal (12 units) amplitude while the 2.2-s oscillation now has higher amplitude (28 units), as denoted. Gaussian noise with a zero mean and a standard deviation of 15 units has been added. Compare with the upper panel of Figure A1.

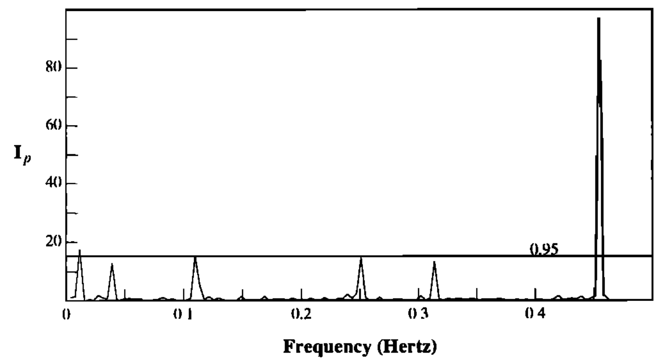

Figure A4. Periodogram of the time series of Figure A3. Note the increase in the number of degrees of freedom brought by the increase in amplitude of the highest frequency oscillation. The Fisher test is also shown.

serial coherence of the series is taken into account. Statistically, we can say nothing about the 2.2-s periodicity; it is not independent. Note that this periodogram has been used in Figure 13. Figure A2*c*) and A2*d*) correspond to the middle and lower panels of Figure A1, where the amplitude of the 80-s oscillation has been arbitrarily increased, and we note the effects of a further decrease in the number of degrees of freedom available, down to 106 and 76 respectively. Besides the loss in frequency range available for examination if a low-frequency oscillation amplitude increases, the relative power of the other surviving frequencies becomes a smaller portion of the total power, leading to a loss in significance of these now weaker power features.

Because of the strong effect that low frequencies in a time series have in the resultant periodogram, i.e., increasing the serial coherence time and thus lowering the number of degrees of freedom available, the question arises as to how much power the high-frequency spectrum oscillations should have before they can become amenable for examination. Figure A3 shows the six-frequency spectrum where highest frequency (2.2-s oscillation period) amplitude is arbitrarily increased until the number of degrees of freedom of the time series is large enough to allow examination of the region of the spectrum where it resides. Figure A4 gives the resultant periodogram showing the presence of this sixth oscillation, and clearly demonstrates that, in the presence of lower frequencies, higher-frequency oscillations need to have a higher amplitude (and power) in order to overcome the serial coherence of any of the lower frequency components present in a given spectrum. In general, one can conclude that the higher frequencies possibly present in a time series must have

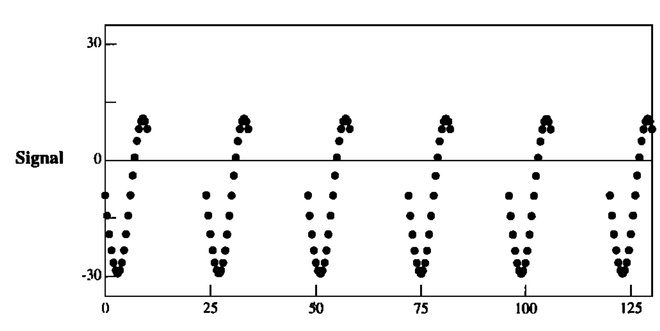

Figure A5. Synthetic time series with equally spaced data gaps.

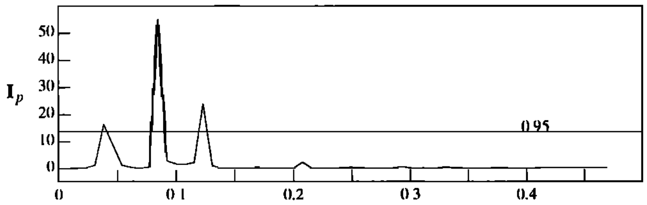

Figure A6. Periodogram of the time series of Figure A5. Note the modulation effect of the data gaps. The Fisher significance test is also shown.

much larger amplitude than the existing lower frequencies in order to be amenable for statistical significance examination. It should be noted that, incidental to this requirement of larger power for a given high frequency to be amenable for examination, this higher power would tend to diminish the significance of the rest of the lower-frequency spectrum. This can be seen by comparing Figure A2(*a*) with Figure A4.

In the Section 2 we discussed the need that the data from which a periodogram would be derived should ideally be equally spaced. However, with the *Lomb* [1976] approach it is possible to handle data gaps and unequally-spaced data. In both these cases of irregularity, the gaps and the unequally-spaced data should have nearly random separation, lest other problems appear. The randomness of the separation can be easily tested using standard techniques. Here we illustrate the difficulties encountered when data gaps in a time series are themselves periodic, as when measurements are only possible for part of the day or when missing data points have a recurrent pattern. For consistency with previous material in the text, we will use seconds in the abscissa as a proxy for hours. For the present synthetic investigation we will presume that the process to be observed consists of a 12-s oscillation, which, because of limitations in the measuring technique, can only be observed during the night. The results of these hypothetical observations are illustrated in Figure A5. The periodogram analysis of these data is given in Figure A6, where we have presumed that the individual 'measurements' are independent and the series is quasi-stationary. Although we know that the only oscillation of interest present in the measurements is a 12-s periodicity, the appearance of other significant frequencies is apparent. The results shown in Figure A6 are not surprising, since they simply show the amplitude modulation of the 12-s oscillation by the recurrent 24-s gaps. From Fourier transform theory one should expect that frequencies at $f_{12} \pm k\ f_{24}$, where k is an integer, should appear as can be seen in Figure A6. Should there exist other true frequencies in the original data series, their interpretation would be cross-contaminated by the sidebands of the other frequencies. This example is simply an ill posed problem for statistical significance analysis by periodogram techniques, but it is illustrative of the dangers lurking when the methodology is improperly applied.

Acknowledgments. The author would like to express his appreciation to G. J. Fraser for providing the measurements used in the data sample. Also, the author would like to thank K. C. Clark, G. J. Fraser, and R. W. Smith for stimulating discussions. This investigation was supported in part by grants ATM-9610200 and OPP-9615157 from the National Science Foundation.

Janet G. Luhmann thanks David Altadill and another referee for their assistance in evaluating this paper.

References

Blackman, R. B. and J. W Tukey, *The Measurement of Power Spectra From the Point of View of Communications Engineering*, Dover Publications, New York, 1959.

Bracewell, R., *The Fourier Transform and Its Applications*, McGraw-Hill, New York, 1965.

Fisher, R. A., Tests of significance in harmonic analysis, *Proc. R. Soc. London, Ser. A, 125*, 54-59, 1929.

Grenander, U., and M Rosenblatt, *Statistical Analysis of Stationary Time Series*, Chelsea, New York, 1957.

Hamilton, K., and R. R. Garcia, Theory and observations of the short-period normal mode oscillations of the atmosphere, *J. Geophys. Res., 91*, 11,867-11,875, 1986.

Harrison, D. E., and N. K. Larkin, Darwin sea level pressure, 1876-1996: Evidence for climate change?, *Geophys. Res. Lett., 24*, 1779-1782, 1997.

Hoel, P. G., *Introduction to Mathematical Statistics*, John Wiley, New York, 1954.

Kendall, M. G., *The Advanced Theory of Statistics*, vol. 2, Charles Griffin, London, 1948.

Leith, C. E., The standard error of time-average estimates of climatic means, *J. Appl. Meteorol., 12*, 1066-1069, 1973.

Little, R. J. A., and D. A. Rubin, *Statistical Analysis with Missing Data*, John Wiley, Inc., New York, 1987.

Lomb, N. R., Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci, 39*, 447-462, 1976.

Percival, D. B., and A. T. Walden, *Spectral Analysis for Physical Applications*, Cambridge Univ. Press, Cambridge, 1993.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge Univ. Press, London, 1986.

Priestley, M. B., *Spectral Analysis and Time Series*, Academic Press, London, 1981.

Scargle, J. D., Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data, *Astrophys. J., 263*, 835-853, 1982.

Schuster, A., On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, *Terr. Mag. Atmos. Elect., 3*, 13-41, 1898.

Walker, G. T., Correlation in seasonal variations of weather, III. On the criteria for the reality of relationships or periodicities, *Mem. Indian. Meteorol. Dep., 21*(9), 13-15, 1914.

Whittle, P., The simultaneous estimation of a time series harmonic components and covariance structure, *Trab. Estadistica, 3*, 43-57, 1952.

G. Hernandez, Graduate Program in Geophysics, Box 351650, University of Washington, Seattle, WA 98195-1650. (e-mail: hernandez@u.washington.edu)

(Received September 4, 1998; revised January 5, 1999; accepted January 5, 1999.)