Mon. Not. R. Astron. Soc. **388,** 1693–1707 (2008) doi:10.1111/j.1365-2966.2008.13499.x

Significance of periodogram peaks and a pulsation mode analysis of the Beta Cephei star V403 Car

F. A. M. Frescura,[1⋆] C. A. Engelbrecht[2⋆] and B. S. Frank[3]

[1] *Centre for Theoretical Physics, University of the Witwatersrand, 2050, South Africa*
[2] *Department of Physics, University of Johannesburg, 2006, South Africa*
[3] *School of Physics, University of the Witwatersrand, 2050, South Africa*

Accepted 2008 May 20. Received 2008 May 14; in original form 2008 February 11

ABSTRACT

We discuss some commonly used methods for determining the significance of peaks in the periodograms of time series. We review methods for constructing the classical significance tests, their corresponding false alarm probability functions and the role played in these by independent random variables and by empirical and theoretical cumulative distribution functions. We discuss the concepts of independent frequencies and oversampling in periodogram analysis. We then compare the results of new Monte Carlo simulations for evenly spaced time series with results obtained previously by other authors, and present the results of Monte Carlo simulations for a specific unevenly spaced time series obtained for V403 Car.

Key words: methods: data analysis – methods: statistical – stars: oscillations.

1 INTRODUCTION

Periodogram analysis is used to identify periodicities in oscillations of stars and is a vital ingredient of asteroseismology. Typically, the data analysed are noisy. As a result, spurious peaks arise in periodograms of the data, not because of any periodicity in the observed system, but because of the way that the noisy signal has been sampled. These spurious peaks can be surprisingly large. It is essential therefore to have reliable tests by which to determine the significance of periodogram peaks.

This topic has already received attention in the literature. Key classical papers include those of Deeming (1975), Lomb (1976), Scargle (1982) and Horne & Baliunas (1986, hereafter HB). We discuss pertinent aspects of these papers in the sections that follow. These papers were criticized in the work of Koen (1990) and Schwarzenberg-Czerny (1998), amongst others. Much of the criticism revolved around the appropriate means for attaching significance to peaks that arise in a calculated periodogram.

Significance tests for periodograms are hugely important to the asteroseismologist who relies on periodograms to deliver precise values for purported eigenfrequencies of pulsation. Comparison of the values of observationally determined eigenfrequencies with the values predicted by the latest theoretical models should, in principle, allow the identification of modes actually excited in real stars and, subsequently, allow for asteroseismological analysis of those stars.

Asteroseismology appears to be on the threshold of a golden age, as extensive surveys like All Sky Automated Survey (ASAS; Pojmanski 1998), and space missions in the mold of Convection Rotation and Planetary Transits (COROT; Baglin et al. 2002), hugely increase the number of known pulsating stars, as well as the time coverage available for their analysis. It is expected that periodogram analysis will continue to play a prominent role in asteroseismology. Hence, accurate interpretation of periodogram peaks is an issue of prime importance.

Various authors have recently presented fresh approaches to this problem. Reegen (2007) has introduced a new and unbiased reliability criterion that he calls the 'spectral significance', which depends on a simultaneous treatment of frequency and phase in determining significances. More recently, he also introduced (Reegen et al. 2008) the 'composed spectral significance' to identify spurious peaks in multiple spectra and remove instrumental or environmental artefacts in calculated spectral signatures. Baluev (2008) presented a treatment of the significance of periodogram peaks based on extreme value theory in which he proposed formulae in closed form for false alarm probabilities, suitable for the time series generated by systematic surveys. We will present a comparative treatment of these new methods in a future paper.

In this paper, we consider some of the well-established methods currently in use for assessing the relevance of periodogram peaks. Our aim is to explore their relative merits. The structure of this paper is as follows. We first discuss the construction of significance tests in general (Section 2), and of Scargle's significance test in particular (Section 3). We next consider the concept of 'independent frequencies' in periodogram analysis (Section 4), and comment on aspects of the work reported by Scargle (1982), HB and Schwarzenberg-Czerny (1998). We report our attempts at reproducing the results of HB by Monte Carlo simulation, and discuss our failure to reproduce their results in detail (Sections 5 to 9). Discrepancies between our results and theirs lead us to a number of important conclusions

⋆E-mail: fabio.frescura@wits.ac.za (FAMF); chrise@uj.ac.za (CAE)

regarding the concept of independent frequencies described by HB and the importance of Monte Carlo simulations in assessing the significance of periodogram peaks. These conclusions are worth noting, given that, far from being obsolete, the methods introduced by Scargle (1982) and HB are still in widespread use in many fields in astrophysics. The following list, though not comprehensive, illustrates the pervasive use of these methods: Benlloch et al. (2001), Enoch, Brown & Burgasser (2003), Falter et al. (2003), Tackett, Herbst & Williams (2003), Lamm et al. (2004), Yabushita (2004), Claudi et al. (2005), Wen et al. (2006), De Cat et al. (2007). We also consider the problem of oversampling periodograms (Section 8). Finally, we apply our conclusions regarding methods of significance testing to an analysis of the pulsation frequencies of the Beta Cephei star V403 Car. Our method allows the identification of as many as seven pulsation frequencies with significances of over 90 per cent (Section 10).

Definitions of the periodogram assumed in our discussion are given in the appendices of this paper. Detailed discussions of the phenomena of aliasing and spectral leakage, to which we refer in the text, may be found in Deeming (1975) and Scargle (1982).

2 SIGNIFICANCE TESTS

Noisy data produce noisy periodograms. Peaks in a periodogram may therefore not be due to the presence of any real periodic phenomenon at all. They may simply be random fluctuations in periodogram power caused by the presence of a noise component in the data. Peaks arising in this way are spurious: they are not due to any real periodicity in the observed phenomena, but are simply artefacts of chance events in the accompanying noise. Spurious peaks can be surprisingly large, so it is important to have reliable tests for detecting their presence.

In this section, we review the theory underlying a class of general, model-independent tests which are in common use. These determine the probability that the periodogram powers[1] observed in a data set might have arisen from pure noise[2] alone, with no other form of signal present. The basis for these tests is the single trial cumulative distribution function (CDF),

$$F_Z(z) = \Pr[Z \leq z]. \tag{1}$$

Here, the random variable $Z = P_X(\omega)$ is the periodogram power at frequency ω for the time series X, and z is a specified power threshold. The function $F_Z(z)$ gives the probability that, when the data X are pure noise, their periodogram power at the given frequency ω does not rise above power-level z. This CDF provides a test as follows. Suppose a model predicts an oscillation at frequency ω. Then, we expect $P_X(\omega)$ to be large at this frequency. However, pure noise by itself could also produce a large value of $P_X(\omega)$. CDF (1) provides an objective criterion for assessing whether the observed large value of $P_X(\omega)$ is due to noise. Were the data pure noise, the probability that $P_X(\omega) < z_0$ for given threshold z_0 is $p_0 = F_Z(z_0)$. Inverting this function, we obtain the threshold power-level z_0 for which a power value $Z \leq z_0$ has a probability p_0 of being due to pure noise alone. It is given by $z_0 = {F_Z}^{-1}(p_0)$. Equivalently, a power value $P_X(\omega) > z_0$ has probability $1 - p_0$ of being due to pure noise alone. This test is primitive and negative. It does not tell us that p_0 is the probability that our signal contains a periodic component of frequency ω, but only that p_0 is the probability that our signal is not pure noise.

In practice, we do not evaluate the periodogram power at a single frequency only, but at a selected set $\{\omega_\mu : \mu = 1, 2, \ldots, N\}$. We then plot $P_X(\omega_\mu)$ versus ω_μ and scan the plot for peaks. The conclusion we would like to draw is that a peak that rise substantially above all others is due to a genuine periodicity, but to justify this conclusion we must first rule out the possibility that the plot could have been produced by pure noise alone. We do this by calculating the probability that the entire observed periodogram profile could be produced by pure noise alone. Suppose the data are pure noise. To calculate the probability that all of the periodogram powers $\{P_X(\omega_\mu) : \mu = 1, 2, \ldots, N\}$ at the sampled frequencies $\{\omega_\mu\}$ fall below a specified power threshold z, define a new random variable,

$$Z_{\max} = \text{maximum}\,\{P_X(\omega_\mu) : \mu = 1, 2, \ldots, N\}. \tag{2}$$

Thus $Z_{\max}$ is the maximum periodogram power among the set of N sampled powers. Now, the power at each of the sampled values will fall below some specified threshold z if and only if $Z_{\max} \leq z$. We thus need to calculate the CDF:

$$F_{Z_{\max}}(z) = \Pr[Z_{\max} \leq z]. \tag{3}$$

The function $F_{Z_{\max}}(z)$ gives the probability that, when the data are pure noise, the periodogram power $P_X(\omega_\mu)$ does not rise above the threshold z at any of the sampled frequencies $\{\omega_\mu\}$. The second significance test is constructed as follows. Let z_0 be a specified power threshold. The probability that pure noise alone will produce periodogram powers $P_X(\omega_\mu)$ that do not exceed the threshold z_0 at any of the sampled frequencies $\{\omega_\mu\}$ is given by

$$p_0 = F_{Z_{\max}}(z_0). \tag{4}$$

Inverting this function,

$$z_0 = {F_{Z_{\max}}}^{-1}(p_0). \tag{5}$$

For given p_0, this inverse function defines a threshold power-level z_0 such that, if the periodogram power at each of the frequencies $\{\omega_\mu\}$ has value $Z \leq z_0$, then the observed periodogram profile has probability p_0 of being due to pure noise alone. This test is more sensitive than the first and reduces the probability of spurious detections.

3 SCARGLE'S SIGNIFICANCE TEST

If the data are Gaussian pure noise, the periodogram power $Z = P_X(\omega)$ at any given frequency ω of the sampled signal X_k is exponentially distributed with probability density function defined by (Scargle 1982, p. 848),

$$\begin{aligned} p_Z(z)\,\mathrm{d}z &= \Pr[z < Z < z + \mathrm{d}z] \\ &= \frac{1}{\sigma_X^2}\,\mathrm{e}^{-z/\sigma_X^2}\,\mathrm{d}z. \end{aligned} \tag{6}$$

The CDF is thus given by

$$\begin{aligned} P_Z(z) &= \Pr[Z < z] \\ &= \int_{\zeta=0}^{z} p_Z(\zeta)\,\mathrm{d}\zeta = 1 - \mathrm{e}^{-z/\sigma_X^2}. \end{aligned} \tag{7}$$

We are interested in the probability that the periodogram power at the given frequency is greater than a specified threshold z. This is given by

$$\Pr[Z > z] = 1 - P_Z(z) = \mathrm{e}^{-z/\sigma_X^2}. \tag{8}$$

[1] The word 'power' is not used here in its formal statistical sense of the probability of rejection of the null hypothesis given that the null hypothesis is false, but in its accepted physical sense. Thus, 'periodogram power at frequency ω' means $P_X(\omega)$ as defined in Appendices A and B.

[2] Pure noise is defined in Appendix C.

As the observed power z becomes larger, it becomes exponentially less likely that so high a power level (or higher) could be produced by pure noise alone, and correspondingly more likely that the observed power level is due to a genuine deterministic (i.e. non-noise) feature in the measured signal.

Note that the argument of the exponential in the CDF is not simply the observed power z, but the ratio z/σ_X^2, which is the ratio of the periodogram power to the total variance of the data (called total input signal power by some). This is an important point, worth emphasizing, as did HB. If the incorrect power ratio is used, then the statistical tests considered by Scargle will necessarily fail. Thus, normalization of the periodogram power by the number N_0 of data points used to calculate the periodogram (classical normalization), or by the residual power after a sine curve has been removed from the data, or by the variance of the observational uncertainty, all lead to completely different statistical distributions for the periodogram power and invalidate Scargle's analysis summarized in this paper. Of course, this does not make alternative normalizations 'wrong'. It does mean however that they must be accompanied by alternative statistical analyses (Schwarzenberg-Czerny 1998).

Suppose now that we evaluate the periodogram at frequencies $\{\omega_\mu : \mu = 1, 2, \ldots, N_i\}$. Denote the periodogram powers at these frequencies by $Z_\mu = P_X(\omega_\mu)$. If the data X are pure noise, then the Z_μ are random variables. We are interested in determining the probability that this entire set of observed periodogram powers could have been produced by pure noise alone. In general, no progress can be made theoretically in this respect unless we assume that the Z_μ are independent random variables. This is an essential ingredient in any theoretical derivation of a false alarm probability function. If the random variables considered are independent, a large body of theorems is available for use. In the absence of independence, we face very serious complications both in the reasoning and proofs of the needed results. We will therefore assume with Scargle that the frequencies ω_μ have been chosen such that the Z_μ are mutually independent. HB refer to a set of frequencies chosen in this way as 'independent frequencies'. This is an abuse of terminology, since it is not the frequencies that are 'independent', but the random variables Z_μ. However, this lack of precision leads to no ambiguity and so is tolerable. To calculate the probability that all the sampled periodogram powers are less than some specified threshold power z, define a new random variable

$$Z_{\max} = \text{maximum}\{Z_1, Z_2, \ldots, Z_{N_i}\}.$$

The probability that any given power Z_μ in this set falls below the threshold is

$$\Pr[Z_\mu < z] = 1 - \mathrm{e}^{-z/\sigma_X^2}.$$

Since the Z_μ are independent, the probability that they all fall below the threshold z is given by

$$\begin{aligned}\Pr[Z_1 < z \text{ and } Z_2 < z \text{ and} \ldots \text{and } Z_{N_i} < z] &= \Pr[Z_1 < z]\Pr[Z_2 < z]\ldots\Pr[Z_{N_i} < z] \\ &= \left[1 - \mathrm{e}^{-z/\sigma_X^2}\right]^{N_i}.\end{aligned}$$

The probability that not all the powers Z_μ are less than the threshold z, that is, the probability that at least one of the powers Z_μ is above the threshold z, is then

$$\Pr[Z_{\max} > z] = 1 - \left[1 - \mathrm{e}^{-z/\sigma_X^2}\right]^{N_i}. \quad (9)$$

This is the function that Scargle proposes as a false alarm probability function. This function is used as follows: choose a probability p_{A} that we regard as an acceptable level of risk for the false detection of real deterministic signals; then solve (9) for z to get a reference power threshold level z_{A}:

$$z_{\mathrm{A}} = -\sigma_X^2 \ln\left[1 - (1 - p_{\mathrm{A}})^{1/N_i}\right]. \quad (10)$$

If we claim a detection whenever the power level at one of the frequencies $\{\omega_\mu : \mu = 1, 2, \ldots, N_i\}$ exceeds the reference level z_{A}, the probability that we will be wrong is given by p_{A}.

4 INDEPENDENT FREQUENCIES

Scargle's test is constructed on the assumption that we can identify a set of frequencies at which the periodogram powers are independent random variables. In the case where the time-domain data are evenly spaced, we are guaranteed the existence of such a set. These are called the natural frequencies (Scargle 1982), or the standard frequencies (Priestley 1981). These are given by

$$\omega_k = \frac{2\pi k}{T}, \quad (11)$$

where T is the total time span of the data set, that is, $T = t_{N_0} - t_1$, and $k = 0, \ldots, [N_0/2]$, where $[N_0/2]$ signifies the integer part of $N_0/2$. The statistics of $P_X(\omega_k)$ with $k = 0$ are different from those with $k \neq 0$ (Priestley 1981). If we omit $P_X(\omega_0)$, this leaves us with at most $[N_0/2]$ independent frequencies. In practice, the omission of $\omega_0 = 0$ from the set of independent frequencies is of no consequence. This frequency corresponds to a DC component in the signal which is generally removed from the data before their periodogram is calculated. Thus, in the case of evenly spaced data, we can easily construct the Scargle false alarm probability function and apply it to determine the significance of high periodogram power levels at these 'independent frequencies'.

It is worth emphasizing that, since the false alarm probability function assumes independent powers at the examined frequencies, we can only use it to put a significance level on the values of the periodogram power at the chosen independent frequencies. Peaks found at other frequency values by oversampling the periodogram cannot be assessed in this way.

In the unevenly sampled case, the situation changes dramatically. The statistical analysis of the classical periodogram becomes intractable. The results are sampling-grid dependent, and no general analysis applicable to all cases has yet been produced. To simplify the statistical analysis, Scargle proposed that the definition of the periodogram be modified. His modified periodogram had already been used by Barning (1963), Vanicek (1969) and Lomb (1976). These authors did not view the modified periodogram as an attempt to estimate the Fourier power spectrum from unevenly sampled data, but as a spectral method for searching for the best-fitting harmonic function to their data. The novelty of Scargle's approach was that he generated the same spectral method as used by these authors by imposing simple constraints on a generalized form of the Fourier transform (FT): the modified periodogram should mimic as closely as possible the statistical properties of the classical periodogram, and the resulting spectral function should be insensitive to time translations of the data in the time domain.

The first demand was only partially successful. Forcing time translation invariance, and demanding that the statistics of the random variable $P_X(\omega)$ at a single selected frequency remain unchanged, that is, demanding that $P_X(\omega)$ be exponentially distributed, exhausts the free parameters in Scargle's modified FT and yields Lomb's spectral formula. In this way, Scargle was able to reproduce some properties of the periodogram for the evenly sampled case.

However, most other familiar properties of the evenly sampled case are lost. The most important loss is the existence of independent frequencies.

All relevant information about correlation and mutual dependence of the random variables $\{P_X(\omega)\}$ is contained in the window function, $G(\omega)$. (For a discussion of the window function, see Scargle 1982, appendix D, p. 850, and also his discussion on p. 840.) Thus, the coefficient of linear correlation between $P_X(\omega)$ and $P_X(\omega')$ is given by $G(\omega' - \omega)$ (Lomb 1976). For independence of $P_X(\omega)$ and $P_X(\omega')$, it is necessary (but not sufficient) that $G(\omega' - \omega) = 0$. Furthermore, for mutual independence of a set $\{P_X(\omega_k)\colon k = 1, 2, \ldots, r\}$ of more than two periodogram powers, it would also be necessary (but not sufficient) to have the ω_k evenly spaced. These are very difficult conditions to realize in practice. Koen (1990) searched numerically for such mutually uncorrelated sets in a variety of sampling schemes and failed to turn up more than two simultaneously uncorrelated frequencies.

For Scargle, this loss of independent frequencies is not debilitating. He says (p. 840, column 1) that '. . . if the frequency grid is well chosen, the degree of dependence between the powers at the different frequencies is usually small', and (p. 840, column 2) that, 'With a wide variety of sampling schemes $G(\omega)$ does have nulls, or relatively small minima, that are approximately evenly spaced. . . 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'. The implication, though not explicitly stated by Scargle, is that in spite of the loss of independence of the random variables $P(\omega)$ at the natural frequencies, the false alarm probability given by our equation (9) (equation 14 in Scargle 1982, p. 839) still provides a reliable significance test in the wide variety of sampling schemes that he considered.

It seems that Scargle's recommendation for the case of unevenly spaced data is as follows: evaluate the modified periodogram at the natural frequencies defined by the given data span, and use the false alarm probability calculated for the evenly spaced case to evaluate the significance of the periodogram peaks. He further recommends that, to improve the detection efficiency, we decrease the number of frequencies inspected (p. 842). The effect of this reduction is that we reduce the power threshold for a given significance level of peak heights.

The value of N_i is a critical ingredient in Scargle's false alarm probability function. There has been some debate concerning its correct value, as well as its meaning. HB appear to have been unsatisfied with the value $N_i = [N_0/2]$ and proposed to determine N_i by a method which we describe in the following section.

5 HORNE AND BALIUNAS DETERMINATION OF N_i

HB determined N_i by the following procedure. They simulated a large number of data sets, each consisting of pseudo-Gaussian noise. The periodogram of each data set was evaluated from $\omega = 2\pi/T$ to $\omega = \pi N_0/T$, where T is the total time interval. They then chose the highest peak in each periodogram, combined these and fitted the Scargle false alarm probability function to the peak distribution using N_i as the variable parameter.

The HB simulations investigated three major types of spacing in the time coordinate. In the first, the data were evenly spaced in time. In the second, each time followed the previous one by a random number between 0 and 1. In the third, the data were clumped in groups of three at each evenly spaced time interval.

In the case where the data are evenly spaced in time, theoretical statistical analysis provides us with a very clear, unambiguous picture of what to expect from the simulations: the random variables $\{P_X(\omega_k) : k = 1, \ldots, [N_0/2]\}$, where $\omega_k = 2\pi\, k/T$ and T is the total time interval covered by the data, are mutually independent; the window function, which contains all relevant information about dependencies and correlations of the random variables $P_X(\omega)$, shows that these are the only frequencies at which the periodogram powers are independent (Scargle 1982, p. 840 and 843); the listed frequencies ω_k contain maximal information about the power distribution of the sampled signal. This is seen from the fact that the discrete FT (DFT) evaluated at these frequencies contains exactly enough information to reconstruct completely the original data. So, from theory, we expect the total number N_i of independent frequencies in the case of evenly spaced time series consisting of zero mean pure noise to be exactly $[N_0/2]$. In practice, a simulated time series, generated from a zero mean distribution, will not have precisely zero mean. We must therefore remove its mean before finding its periodogram. Once this is done, theory guarantees that a simulated data set will have exactly $[N_0/2]$ independent frequencies. If the data are used also to estimate other parameters, this number will be reduced further. The value $[N_0/2]$ is therefore the maximum number of independent frequencies expected in a periodogram of evenly spaced data.

Surprisingly, the best fits obtained by HB consistently produced values of N_i which were substantially higher than this expected upper limit (HB, table 1, p. 759). In fact, their fitted values are consistently higher than N_0, with the exception of their two smallest data sets (10 and 15 points, respectively) where the fitted value of N_i is slightly less than N_0, but still about twice as large as expected.

These results are puzzling. Theory and simulations appear to be in conflict. Cumming, Marcy & Butler (1999) note that Baliunas has indicated typographical errors in the values listed in HB. Koen (1990) and Schwarzenberg-Czerny (1996) have also noted mistakes in HB. We have repeated the HB simulations for the case of even sampling in the time domain. We have also extended somewhat the scope of their investigations to consider the alternative false alarm probability function proposed by Schwarzenberg-Czerny (1998), as well as the effects of oversampling the periodogram. The results are interesting, and we report them in the corresponding sections below.

In our first set of simulations, we attempted to reproduce the results reported by HB in their table 1, p. 759, for the case of evenly spaced data. HB describe the method they followed in their simulations as follows: 'The periodogram of each data set was evaluated from $\omega = 2\pi/T$ to $\omega = \pi N_0/T$ . . . The highest peak was then chosen in each periodogram.' It was not clear to us whether they sampled the periodogram values $P_X(\omega)$ only at the natural frequencies $\omega_k = 2\pi k/T$, and then chose the highest periodogram power from this restricted sampled set, as prescribed by Scargle; or whether they followed the practice of a not insubstantial number of astronomers who search for the highest periodogram peak in the given range by grossly oversampling the periodogram, and then choose the maximum value obtained irrespective of whether it occurs at one of the natural frequencies ω_k. Accordingly, we ran two sets of simulations implementing both procedures. We fitted the Scargle false alarm function to our results by the method of least squares. All our periodograms were normalized using the sample variance of the simulated data, and not the variance of the distribution used to generate the sample. We failed to reproduce the HB results in detail. Sampling the periodogram at the natural frequencies only and choosing the highest value among these yielded values of N_i

Table 1. Results of Monte Carlo simulations.

N_0	HB value of N_i	Number of tests	Oversampled		Natural frequencies	
			Scargle function best-fitting N_i	SC function best-fitting N_i	Scargle function best-fitting N_i	SC function best-fitting N_i
10	9.70	1395	9.09	26.80	5.00	8.70
15	14.45	347	14.09	32.80	7.90	13.60
25	27.38	213	24.91	47.80	12.80	20.00
35	38.40	214	35.64	60.40	18.50	26.50
50	54.45	369	54.00	84.00	25.70	34.40
64	71.76	512	70.45	102.80	33.00	42.40
75	86.05	153	85.82	121.90	39.30	50.10
100	119.58	296	113.91	152.10	51.60	62.80
128	152.53	913	149.09	191.80	65.50	77.70
170	218.33	218	210.09	261.40	89.30	103.00
256	369.97	224	306.36	361.20	128.40	143.00
300	455.95	107	361.45	420.20	148.40	163.30
400	618.69	106	477.18	540.10	204.30	221.60

Note. Comparison of HB values of N_i with the results of our numerical simulations, fitting both Scargle and Schwarzenberg-Czerny (SC) false alarm functions to CDFs constructed from oversampled periodograms and from periodograms sampled at the natural frequencies. The corresponding best-fitting functions are displayed in Figs 2 and 3, together with the corresponding functions constructed with the correct value of $N_i = [N_0/2]$.

that were consistently lower than those obtained by HB. In fact, we obtained values very close to $[N_0/2]$, as expected theoretically, but in conflict with the results published by HB. Searching for the highest peak by oversampling also yielded values that were consistently lower than HB, but higher than sampling at the natural frequencies. More precisely, our results agree closely with those of HB for the smaller data sets up to 170 data points. This leads us to suspect that the HB table was constructed by gross oversampling. However, our results strongly deviate from theirs for the larger data sets with $N_0 > 170$, with our values being substantially lower. Plotting N_i versus N_0 (Fig. 1), we observe the following features. The values yielded by our simulations increase linearly with N_0, as expected. In contrast, the results published by HB in their table 1 appear to lie, not on a quadratic (as claimed by them), but on two straight lines of different slope, a sharp change in slope appearing for data sets with $N_0 > 170$. This seems to be indicative of a systematic error. Fitting a quadratic function to these data points, as was done by HB, may therefore be misleading and renders suspect its use in estimating the parameter N_i.

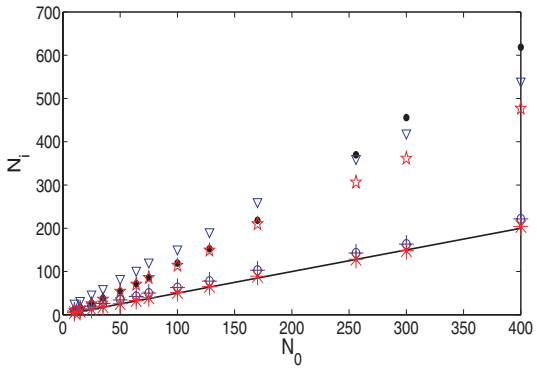

Figure 1. Plots of N_i versus N_0 of the data published by Horne & Baliunas (1986) in their table 1, p. 759, for the case of evenly spaced data, and of our simulations, fitting the Scargle and Schwarzenberg-Czerny false alarm functions to the empirical CDFs obtained by sampling at the natural frequencies and by oversampling. Solid dots: published HB values; asterisks: Scargle function fitted to empirical CDFs obtained by sampling at the natural frequencies; circled crosses: Schwarzenberg-Czerny function fitted to the same; stars: Scargle function fitted to empirical CDFs obtained by oversampling; triangles: Schwarzenberg-Czerny function fitted to the same. The solid line is the theoretically expected relationship $N_i = [N_0/2]$.

Note however that, in the case of oversampling, both our results and those of HB consistently yield values of N_i that are higher than the theoretically expected value of $[N_0/2]$. These values are thus apparently in conflict with the theory. The interpretation of N_i as the number of independent frequencies is therefore questionable in this context. The HB method for determining N_i is eminently practical and reasonable, but it only yields correct values when the periodogram is sampled at the natural frequencies. This means that, in the context of oversampling, we cannot assign to the parameter N_i the meaning that it had in its original derivation, namely the number of independent frequencies in the associated periodogram. Rather, we must treat N_i as nothing more than a floating parameter in a one-parameter family of candidate CDFs which we are attempting to fit to our data.

Another problem with the HB method should be noted. Inspection of a plot of the best-fitting Scargle false alarm probability function shows it to be a very uncomfortable fit to the experimentally obtained cumulative distributions of periodogram peak heights (see Fig. 2). This is true both in the case of sampling at the natural frequencies and of oversampling. Its general trend is good, but its detailed behaviour does not match that of the experimental curve. This mismatch is most pronounced for small data sets, and becomes less noticeable as the data sets increase in size. However, it never vanishes completely. The conclusion forced on us by our simulations is that the Scargle false alarm probability function fails to reproduce the detailed behaviour of the simulated data sets. This is both good news and bad news: good news because it shows that the Scargle function underestimates the significance of periodogram peaks; and bad news because it leaves us without a useable false alarm probability function.

In summary, our simulations indicate that (1) the Scargle probability function incorrectly describes the statistical behaviour of the periodogram (we discuss a possible reason for its failure in the next section), and (2) the HB method is not a way to assess the number of independent frequencies in a periodogram, but is a way for

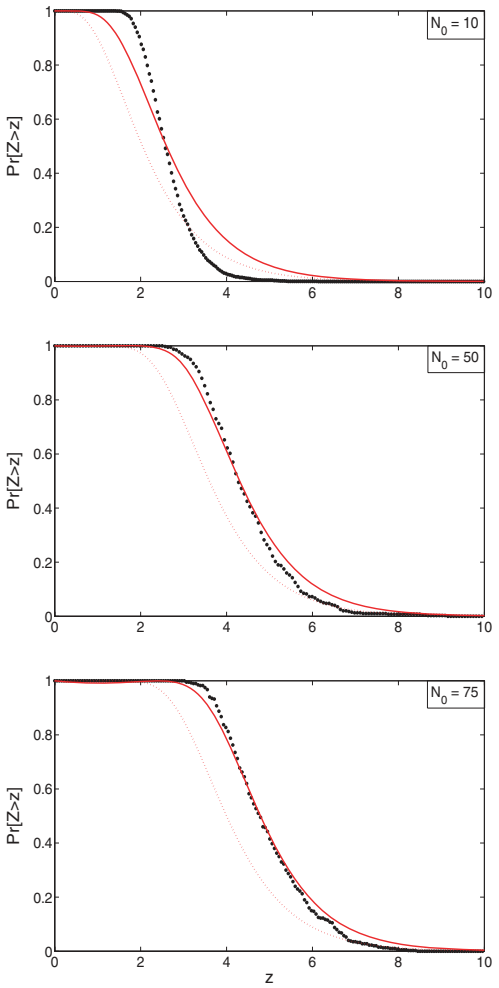

Figure 2. Empirical CDFs (heavy dotted line) constructed by oversampling the periodogram, with best-fitting Scargle false alarm probability function (solid line) for (a) $N_0 = 10$, (b) $N_0 = 50$ and (c) $N_0 = 75$ data points. The fit improves with increasing N_0. The corresponding best-fitting values of N_i are (a) $N_i = 9.09$, (b) $N_i = 54.00$ and (c) $N_i = 85.82$. The light dashed line shows the Scargle function for $N_i = [N_0/2]$. In all cases the best-fitting value of N_i exceeds $[N_0/2]$.

estimating the best-fitting value of N_i in an ill-fitting class of candidate CDFs.

6 THE SCHWARZENBERG-CZERNY FALSE ALARM FUNCTION

Koen (1990) pointed out an important implicit assumption in Scargle's derivation of his false alarm probability function. Scargle assumed that the variance σ_X^2 of the data X_k is known a priori. There are situations in which this condition is true, but it is satisfied neither in the case of real astronomical data nor in that of the HB simulations. In the simulations, pseudo-data are generated using a pre-selected variance and mean (chosen to be zero), but the variance and mean of the generated sample will differ in general from those used in their generation. Thus both variance and mean need to be estimated from the data.

This changes the statistical analysis significantly. Schwarzenberg-Czerny (1998), in a particularly clear and thorough exposition of the issues involved, has shown that the CDF of maximum peak heights appropriate to the Lomb–Scargle periodogram and calculated from a finite sample of Gaussian pure noise is the (regularized) incomplete beta function:

$$I_{1-z/[N_0/2]}([N_0/2], 1) = \left(1 - \frac{z}{[N_0/2]}\right)^{[N_0/2]}. \quad (12)$$

To construct the corresponding false alarm probability function, we need to use this distribution in place of the exponential distribution used above. If in our periodogram we can identify a set of frequencies at which the periodogram powers are mutually independent, then the probability that the power at at least one of these frequencies rises above given threshold power z is given by

$$\Pr[Z_{\max} > z] = 1 - \left[1 - \left(1 - \frac{z/\sigma_X^2}{[N_0/2]}\right)^{[N_0/2]}\right]^{N_i}, \quad (13)$$

where N_i is the number of mutually independent frequencies inspected, and $Z_{\max} = \text{maximum}\{Z_1, Z_2, \ldots, Z_{N_i}\}$ is the maximum power among the mutually independent powers Z_μ. In our discussion, we shall call equation (13) the Schwarzenberg-Czerny false alarm probability function. In passing, note that Schwarzenberg-Czerny (1998) provides a number of alternative distributions and test statistics appropriate to other methods of data analysis and is able thereby to resolve extant disputes about the 'correct' normalization procedure for periodograms.

In the limit $N_0 \to \infty$, the distribution in equation (12) becomes exponential and coincides with that used by Scargle. Accordingly, in the same limit, the associated false alarm probability function in equation (13) reduces to the Scargle false alarm function. A Q–Q plot of the Schwarzenberg-Czerny versus Scargle false alarm functions (see Schwarzenberg-Czerny 1998, fig. 1, p. 835) shows that, while the agreement between them is good for large N_0, they differ substantially for small data sets, with Schwarzenberg-Czerny's false alarm function yielding consistently smaller false alarm probabilities than Scargle's. According to this analysis, therefore, for given N_i, the Scargle false alarm function consistently underestimates the statistical significance of periodogram peaks.

One reason for the failure of the Scargle function to reproduce the behaviour of our empirical CDFs may be its implicit assumption that the variance σ_X^2 is known a priori. To correct this error, we replaced the Scargle function by Schwarzenberg-Czerny's and repeated the HB simulations for equally spaced data. Using their method for determining N_i, we fitted the Schwarzenberg-Czerny false alarm function to our empirical CDFs. We found very good, but not perfect, agreement between the best-fitting theoretical curves and the corresponding empirical ones, with the greatest deviations occurring for small data sets (see Fig. 3). For these, the theoretical best-fitting curves consistently yield values that are lower than those of the empirical curves, thus overestimating the significance of peaks. For the larger values of N_0, the deviations of the fitted from the empirical curves may be understood in the context of order statistics.

In spite of the excellent nature of these fits, there is nevertheless an interesting feature in these results that is worth noting. For the CDFs of periodogram powers sampled at the natural frequencies, the best-fitting values of N_i are consistently larger than the theoretically expected number of independent frequencies, which is at most $[N_0/2]$ (see Table 1). Correspondingly, a plot of the Schwarzenberg-Czerny function for the value $[N_0/2]$ of independent frequencies yields a curve that deviates badly from the corresponding empirical CDF and which leads to a severe overestimation of the significance of periodogram peaks. We have no option but

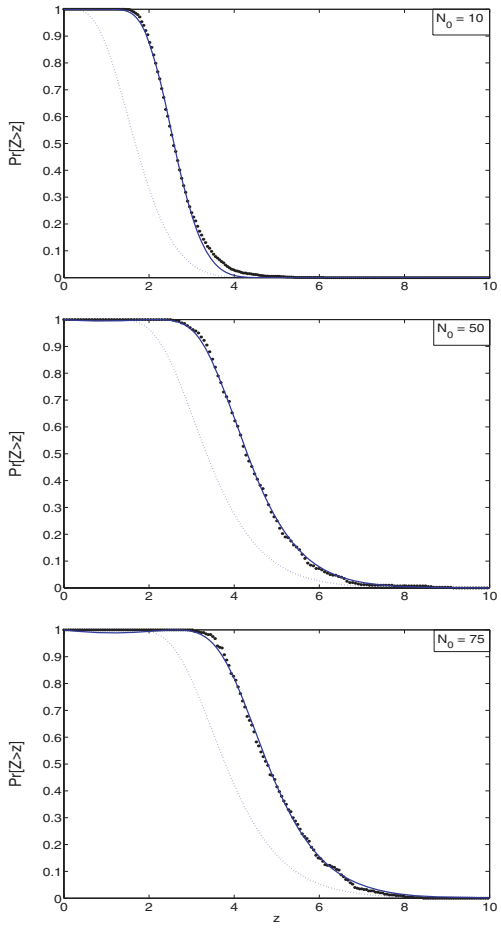

Figure 3. Empirical CDFs (heavy dotted line) constructed by oversampling the periodogram, with best-fitting Schwarzenberg-Czerny false alarm probability function for (a) $N_0 = 10$, (b) $N_0 = 50$ and (c) $N_0 = 75$ data points. The fits are significantly better than the corresponding ones for Scargle's function. However, for low N_0, Schwarzenberg-Czerny's distribution is still significantly different from the empirical one and overestimates the significance of high peaks. The light dashed line shows the corresponding Schwarzenberg-Czerny false alarm function for $N_i = [N_0/2]$. In all cases, the best-fitting value of N_i again exceeds $[N_0/2]$.

to conclude from these results that, like the Scargle false alarm function, the Schwarzenberg-Czerny false alarm function, given by equation (13), appears not to describe the CDFs of our simulations. Note also from Table 1 that the best-fitting values of N_i for CDFs constructed from oversampled periodograms are higher than those for the CDFs obtained by sampling at the natural frequencies. This is consistent with our previous results for the Scargle false alarm function. The results of our simulations again appear to be at variance with the theory. For evenly spaced data, the theory (which seems unassailable) predicts unambiguously the existence of at most $[N_0/2]$ independent frequencies, with the CDF for periodogram powers sampled at these frequencies given by the Schwarzenberg-Czerny false alarm function. Our empirical CDFs differ substantially from those predicted by this theory, with HB best fits occurring at values of N_i that are consistently higher than expected. We are thus forced to the same conclusion as in the previous section: the HB best-fitting value of N_i does not provide an estimate of the number of independent frequencies in a periodogram. It is a floating parameter in a one-parameter family of candidate CDFs. Further, as candidate CDFs, the Schwarzenberg-Czerny false alarm function consistently performs better than Scargle's.

7 FALSE ALARM FUNCTIONS FOR UNEVENLY SPACED DATA

The principal difficulty encountered when searching for a theoretical false alarm function when the data are unevenly spaced is the loss of the independent frequencies. It is not that they are present, but difficult to identify. They are simply not there at all, and their absence makes the search for a theoretical false alarm function intractable. Our only alternative therefore is to use Monte Carlo simulations.

Schwarzenberg-Czerny (1998) expresses a distinct lack of confidence in this approach. To assess the statistical significance of periodogram peaks, we use principally the high-peak tail of the false alarm function. Construction of it by simulations relies on rare events of low probability. Schwarzenberg-Czerny argues that the accuracy of random number generators, and the accuracy of periodogram algorithms, is not well tested in this domain.

On the other hand, observers need a reliable method for assessing candidate peaks. They clearly cannot rely on the current generation of theoretical distributions. These are all based on the assumption of independent frequencies, and all require a value of N_i to be selected before they can be used. In the case of evenly spaced data, for periodograms evaluated at the standard frequencies, it might be argued that the correct value for N_i is $[N_0/2]$, suitably reduced by the number of parameters already estimated from the data. For the general case, however, even were we to believe the conjecture that independent frequencies exist, there appears to be no clear a priori theoretical criterion for choosing the value of N_i. The only practical method available for estimating N_i is that of HB, which requires us first to construct the CDF by simulation, and then to fit it with some chosen one-parameter family of theoretical distributions. Like it or not, we are therefore left with no option but to use Monte Carlo methods.

This option is not as bleak as it might at first appear. Schwarzenberg-Czerny's opinion regarding random number generators is not unwarranted. However, their performance is continually being improved, and there is every reason to believe that existing problems with them will eventually be resolved, if they have not been resolved already. In contrast, the problem of the lack of independent frequencies is permanent.

As regards the use of theoretical false alarm probability functions, we do not really need them. The empirically generated CDFs contain all the information that we need, whether or not we have a closed-form formula for them, and can be used to determine significance thresholds. A closed-form formula would be useful to facilitate calculation of the thresholds, but is not essential. If one is needed, we can resort to fitting the empirical CDF as closely as possible by any suitable form of trial function. In fact, we do not need even to fit the entire CDF. We are interested only in the high-peak tail above a certain minimum confidence threshold and so need only obtain a good fit in that region. (See Section 10 of this paper for an example.) Should formulae be needed for other regions, we can resort to multiple fits that together cover the entire CDF.

8 THE PROBLEM OF OVERSAMPLING

Theoretical false alarm probability functions are based on the assumption of the existence of independent frequencies and contain the number N of frequencies inspected as a parameter. When the

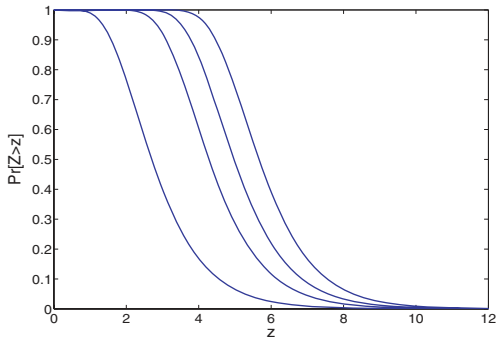

Figure 4. Scargle false alarm probability function as a function of N for values $N = 10, 50, 100, 200$. As N increases, the probability of finding a peak above any given threshold value increases. This illustrates Scargle's 'statistical penalty': if many independent frequencies are inspected for a spectral peak, we should expect to find a large peak even when no signal is present. As N increases, the CDF moves progressively to larger peak height values without limit.

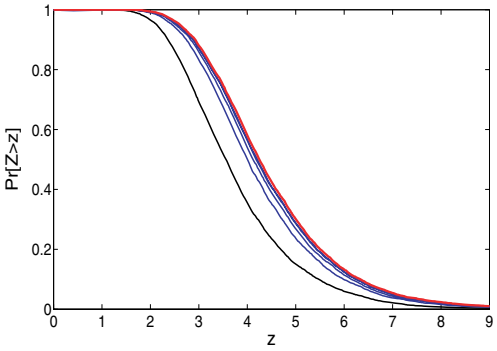

Figure 5. Empirical CDFs as a function of oversampling. The figure shows the CDFs corresponding to sampling at 1, 2, 3, 4, 5 and 10 times the Scargle sampling rate. The corresponding CDFs converge rapidly to a limiting CDF. The limiting CDF coincides almost perfectly with the CDF for an oversampling factor of 10.

periodogram is inspected at the maximum number N_i of independent frequencies, $N = N_i$. For example, Scargle's function is given by

$$\Pr[Z > z] = 1 - F_{Z_{\max}}(z) = 1 - (1 - \mathrm{e}^{-z})^N, \tag{14}$$

when N is the number of independent frequencies inspected.

From Fig. 4, it is seen that, for given z, this probability increases as the number N of sampled independent frequencies is increased. Scargle (1982, p. 839) describes this property as the statistical penalty that we must pay for inspecting a large number of frequencies. He also notes that the expected value of the maximum power $Z_{\max}$ of a white noise spectrum over a set of N frequencies at which the power is independent is given by

$$\langle Z_{\max} \rangle = \sum_{k=1}^{N} \frac{1}{k} \tag{15}$$

which diverges logarithmically with N. These comments seem to suggest that prodigious sampling of the periodogram at the independent frequencies might lead eventually to the dismissal of all periodogram peaks as spurious. They also appear strongly to discourage oversampling of the periodogram. Were these conclusions correct, periodogram methods would be severely compromised.

To understand Scargle's comments correctly, we need first to note his argument is based on the assumption that we are able to identify N independent frequencies ω_k. For an evenly sampled time series with N_0 data points, there are at most $N = [N_0/2]$ of these. There is correspondingly also a maximum value for $\langle Z_{\max} \rangle$. So, in practice, there is no logarithmic divergence to fear.

Second, if we oversample the periodogram, the powers at the sampled frequencies are no longer independent, and so equation (9) is no longer correct. This lack of independence, however, is no obstacle to the construction of a CDF by numerical experiment. The results of our numerical simulations show that successive oversamplings have progressively smaller effects on the CDF, until the CDF eventually converges to a limiting function beyond which no further refinement of the sampling grid changes the result (see Fig. 5). With hindsight, we should have expected this. The original time domain data contain a finite amount of information. There is therefore a limit to how much information they can be forced to yield. Based on the numerical experiments described in this paper, we would therefore like to refine Scargle's 'lesson', drawn from a consideration of statistical penalties. The (gloomy) lesson he drew was 'If many frequencies are inspected for a spectral peak, expect to find a large peak power even if no signal is present' (Scargle 1982, p. 840, column 1). Our revision of Scargle's lesson is this: if many frequencies are inspected for a spectral peak, expect to find a large peak power even if no signal is present – but the total number of independent frequencies present in any given time series is limited, so do not expect the number of large peaks produced by white noise to increase without limit. More importantly, oversampling the periodogram does not dramatically increase the number of large peaks expected.

9 A PRACTICAL METHOD FOR DETERMINING FALSE ALARM PROBABILITIES

The theoretical false alarm probability functions extant in the literature all rely for their validity on the existence of a set of independent frequencies. Such a set is guaranteed for evenly spaced data, but not for data that are unevenly spaced. Even in the case where the data are evenly spaced, we may wish to inspect the periodogram at frequencies that do not coincide with Scargle's natural ones. Such is the case when a pronounced peak occurs at an intermediate frequency. How do we assess the significance of periodogram peaks in these cases?

Based on our investigations described above, we suggest the following method is the only one that is viable in the case of unevenly spaced data. In fact, based on our analysis above, we suggest that it is also the only viable method even in the special case of evenly spaced data.

(i) Using the sampling times of the actual data set to be analysed, construct a large number of pseudo-Gaussian random time series (i.e. construct a large number of data sets consisting of pseudo-Gaussian noise).

(ii) Select a convenient grid of frequencies that cover the frequency range in the periodogram that is to be inspected. (We discuss how to choose these frequencies in the next paragraph. For the moment, assume that they have been selected.)

(iii) Construct the periodogram for each pseudo-random time series, sampling it at each of the selected frequencies.

(iv) In each periodogram, identify the highest periodogram power that occurs at the pre-selected frequencies only, and use these highest values to construct the CDF of these highest power values.

The CDF thus obtained is an empirically generated graphical representation of the probability function Pr $[Z_{\max} \leq z]$. It gives the probability that pure noise alone could have produced power values less than or equal to a given threshold value z at each of the selected sampling frequencies.

The plot of $1 - \Pr[Z_{\max} \leq z]$ is thus the required false alarm probability function. It gives the probability that pure noise alone could produce a peak at the inspected frequencies of value higher than the threshold z.

How do we choose the frequencies at which to sample the periodogram? In a sense, it makes little difference how we choose them since, once chosen, we generate an empirical false alarm probability function that is tailor-made for our particular choice. However, for each choice, there is a price to be paid, and the final decision on how to choose the sampling frequencies is determined by what we consider to be the best compromise between the price paid and the advantage gained. For a given false alarm probability p_A, the denser the sampling, the higher the associated threshold z, with the heaviest penalty being paid for oversampling sufficiently dense as to produce a fully resolved periodogram curve. In our simulations, this occurred at approximately five times the Scargle sampling rate, that is, using $\Delta\,\omega = (1/5)(\pi\, N_0/T)$ (see Fig. 6). The sampling rate sufficient to guarantee convergence to the limiting CDF must be established individually for each data set. This can be done using plots like those shown in Figs 6 and 7.

If we are interested in pinpointing precisely the frequency of a peak (as we are in asteroseismology), then gross oversampling may be the route to follow. However, there is a limit to the amount of information contained in the periodogram of a finite time series. There is therefore also a limit to how finely the frequency axis should be subdivided. This limit is given by $\Delta\omega_{\min} = \pi/T$, which is the smallest frequency interval that can reasonably be resolved by the data set. Dense oversampling in pursuit of the convergence limit of the CDF may lead to a choice of $\Delta\omega$ smaller than this interval. If the limiting CDF differs substantially from that obtained from $\Delta\omega_{\min}$, then limiting the sampling interval to $\Delta\omega_{\min}$ may be a better option.

10 APPLICATION TO V403 Car

Rapid advances in detailed modelling of the interiors of various classes of pulsating star (see for example Pamyatnykh 2003 and Smolec & Moskalik 2007) have made urgent the accurate and complete identification of pulsation modes in real stars.

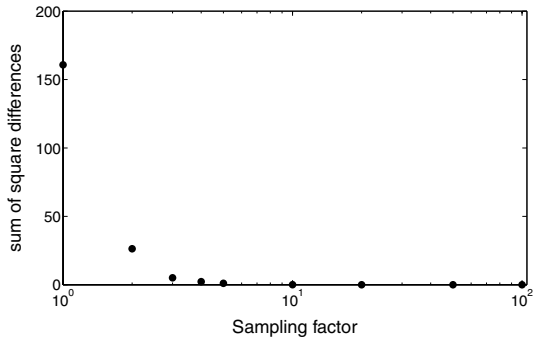

Figure 6. Logarithmic plot of the sum of square deviations (from the limiting CDF) of the CDF for ν times oversampling versus the oversampling factor ν. The convergence to the limiting curve is seen to be very rapid. For the data set used in this simulation, the convergence occurs approximately at an oversampling factor of 5. The convergence shows up in this plot as a sharp levelling off of the graph.

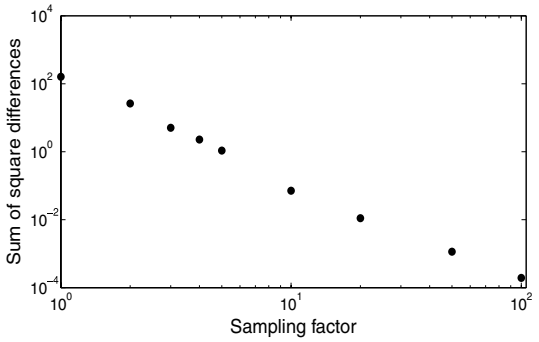

Figure 7. Log–log plot of the sum of square deviations (from the limiting CDF) of the CDF for ν times oversampling versus the oversampling rate ν.

In this paper, we have considered and evaluated common methods currently in use for estimating the statistical significance of periodogram peaks. The conclusion forced on us is that the estimates of significance provided by the available theoretical probability functions are unrealistic and, in many cases, positively misleading. This makes the estimation of the relevant probability functions by Monte Carlo simulation more attractive. The simulation method we have described allows us to generate a unique CDF for any given real data set irrespective of whether the measurement times of that data set are equally or unequally spaced. The generated CDF can then be used to determine quantitatively the significance of candidate peaks in the periodogram for that data set. For the purpose of asteroseismology, the theorist may then be guided by such quantified significances when constructing models for respective stars. Specifically, potential pulsation modes identified in a star may be included or excluded from a hypothetical model, with differing amounts of uncertainty.

As an example of what may be achieved, we illustrate the application of these methods to a real data set, previously analysed in Engelbrecht (1986). The data set consists of 530 observations of V403 Car (star no. 16 in the designation of Feast 1958 and Turner et al. 1980 for NGC 3293) in the Johnson B band, obtained with the 1.0-m telescope at the Sutherland station of the South African Astronomical Observatory (SAAO) over a period of 70 d in 1984. The Lomb–Scargle periodogram of the data, normalized by the variance of the data, is shown in Fig. 8. This periodogram has been oversampled by a factor of approximately 25.

The limiting empirical CDF computed for the time data, as described in preceding sections of this paper, is shown in Fig. 9. An enlargement of the critical region of this CDF is shown in Fig. 10. The

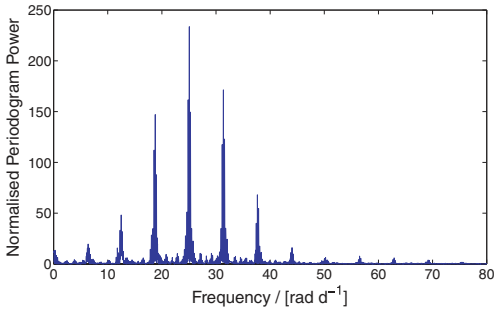

Figure 8. Oversampled Lomb–Scargle periodogram for V403 Car.

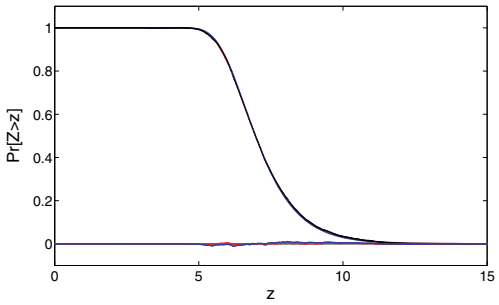

Figure 9. Limiting empirical CDF for the V403 Car data set, displayed with the best-fitting Scargle and Schwarzenberg-Czerny false alarm functions and their corresponding residuals. The best-fitting Scargle and Schwarzenberg-Czerny false alarm functions are in good agreement with each other, and also with the empirical CDF, as may be seen from the residuals. However, the best-fitting values of N_i are 733 and 804, respectively, and bear no resemblance to any value that may reasonably have been anticipated for this parameter.

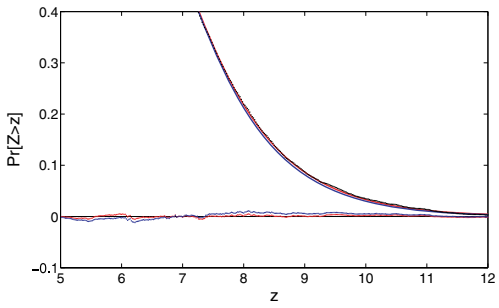

Figure 10. Enlargement of critical region of the CDF for V403 Car showing the empirical CDF for this data set, together with best-fitting Scargle and Schwarzenberg-Czerny false alarm functions. The corresponding residuals for the fits are also displayed.

best-fitting Scargle and Schwarzenberg-Czerny false alarm functions, together with their residuals, are also displayed in Figs 9 and 10. Both functions are in good agreement with each other. For this particular data set, they are also in good agreement with the associated empirical CDF. Note however that the best-fitting values of the parameter N_i are 733 and 804, respectively, and bear no resemblance to any value that may reasonably have been anticipated for this parameter. It should be emphasized that the good agreement of the theoretical false alarm functions with the empirically generated one is a property of the particular data set used in this example and should not be expected in general. For some data sets, particularly those with small value of N_0, the disagreement may be quite severe.

A further enlargement of the critical region is displayed in Fig. 11. At this scale, differences between the Scargle and Schwarzenberg-Czerny best-fitting functions become visible, with the Scargle function providing a closer fit to the empirical CDF than the Schwarzenberg-Czerny function. The smoothness of the empirical CDF can be improved by increasing the number of simulations used in its construction by one or two orders of magnitude. Since the Scargle function follows the empirical CDF in this region so closely, we have preferred in this illustrative example to adopt a different procedure: we have fitted both the Scargle and the Schwarzenberg-Czerny function to the tail of the CDF in the critical region only. These best fits are also displayed in Fig. 11. These follow the empirical CDF more closely still, with the Scargle function providing

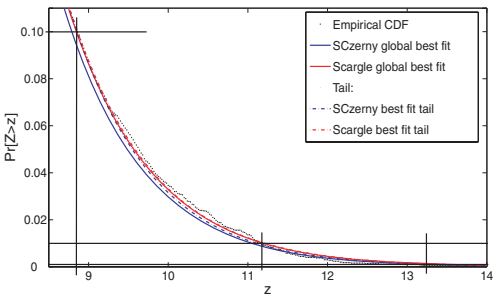

Figure 11. Further enlargement of critical region of the CDF for V403 Car, showing the 90, 99 and 99.9 per cent significance levels. The corresponding normalized periodogram power levels can be read off on the abscissa, as indicated by the short vertical lines. Also displayed are the best-fitting Scargle and Schwarzenberg-Czerny false alarm functions for the entire data set (solid lines), together with the best-fitting Scargle and Schwarzenberg-Czerny false alarm functions for the tail region alone (dashed lines). In this example, the best-fitting Scargle function provides the better estimator for the empirical CDF. Interestingly, this enlargement also shows that the theoretical false alarm functions have a shape different from that of the empirical CDF. In this example, the agreement is so close that this shape difference is insignificant, but it need not always be so.

Table 2. Significance levels for the V403 Car periodogram.

Significance	Periodogram power level		
	Empirical CDF	Scargle function	SC function
90 per cent	8.86	8.85	8.79
99 per cent	11.11	11.17	11.06
99.9 per cent	13.31	13.24	13.25

the better fit. In a case like this, it is more efficient to substitute the best-fitting tail for the empirical CDF rather than to attempt to improve the smoothness of the empirical CDF which would increase considerably the computational time required to achieve greater smoothness. The loss of accuracy suffered in the present example by this substitution is minimal. Note that this option is not available when the discrepancy between theoretical and empirical CDFs is substantial. In such a case, there is no alternative to improvement of the empirical CDF in the rare-event tail.

In Fig. 11, we have also displayed horizontal lines indicating false alarm probabilities of 10, 1 and 0.1 per cent, respectively (i.e. peak significances of 90, 99 and 99.9 per cent, respectively). The periodogram peak values associated with the above significances were read off the abscissa of the best-fitting Scargle false alarm function for the tail region only. The power levels corresponding to the three significance levels mentioned above are displayed in Table 2.

We now determine the significances of peaks appearing in the periodogram of V403 Car using the best fit in the critical region of the empirical CDF. The oversampled, normalized Lomb–Scargle periodogram shown in Fig. 8 was subjected to a standard pre-whitening procedure, as follows:

(i) determine the frequency at which the highest peak occurs in the periodogram;

(ii) determine the best-fitting amplitude and phase of a sinusoid with this frequency by least-squares comparison with the data;

(iii) subtract this sinusoid from the original data;

Table 3. Results of pre-whitening analysis for V403 Car.

	Frequency/[rad d^{-1}]	Power	Significance
f_1	25.07	234	>99.99999
f_2	6.34	68.3	>99.99999
f_3	2.56	30.9	>99.99999
f_4	1.37	22	>99.99999
f_5	25.75	22.6	>99.99999
f_6	30.71	19.3	99.9998
f_7	28.91	10.3	97.7
f_8	32.65	9.1	92.2
f_9	25.18	8.8	89.5
f_{10}	37.58	8.8	89.5

(iv) recalculate the oversampled, normalized Lomb–Scargle periodogram for the modified data;

(v) repeat step (i) above;

(vi) determine the respective best-fitting amplitudes and phases of two sinusoids with the frequencies determined in steps (i) and (v) above;

(vii) iterate steps (iii) to (vi) above, recalculating the respective best-fitting amplitudes and phases for the entire set of determined frequencies each time.

The day-to-day and week-to-week spacings in the observations of V403 Car generate substantial noise in the low-frequency part of the periodogram. Once the second, third and fourth strongest peaks (determined through the pre-whitening procedure outlined above) had been removed, further pre-whitening was limited to frequencies above 20 rad d^{-1} (3.18 cycle d^{-1}), since no pulsation frequencies are expected to lie below this threshold. Pre-whitening was continued until the 90 per cent significance level (as established by the best fit to the empirical CDF) of the data set had been reached. The results of the pre-whitening procedure appear in Table 3, where the following values are displayed: frequency with highest power in the periodogram (limited to values above 20 rad d^{-1}, from frequency 5 onwards); amplitude of the peak associated with this frequency (in units of normalized power); significance of the periodogram peak, established in this instance by the best fit to the empirical CDF. The appearance of the periodogram after pre-whitening with 4, 6, 8 and 10 frequencies, respectively, is displayed in Fig. 12. Many practitioners have been using a signal-to-noise amplitude ratio of 4 (or sometimes 3.5) in the periodogram (Breger et al. 1993) as a dividing line separating significant peaks from the rest. The determination of the noise level is to some degree subjective. Consequently, it is not possible to estimate with precision the level of periodogram power that corresponds to a signal-to-noise ratio of 4. However, using the empirical CDF, one can explicitly quantify the significances of periodogram peaks.

As explained above, the second, third and fourth frequencies appearing in Table 3 are likely to be due to the spacing of the V403 Car data in the time domain. This conjecture could be tested by an appropriate multisite campaign on V403 Car. The remaining seven frequencies are represented in Fig. 13. Frequencies f_6, f_7 and f_8 constitute an almost equally spaced triplet of frequencies, which raises the question of rotational splitting of eigenfrequencies. Frequencies f_6 and f_7 are separated by 1.80 rad d^{-1}, while f_6 and f_8 are separated by 1.94 rad d^{-1}. Heynderickx, Waelkens & Smeyers (1994) determined a value of $\log T_{\rm eff} = 4.398$ for V403 Car, and a mass of 13.51 $M_\odot$. Balona (1975) measured a projected rotation

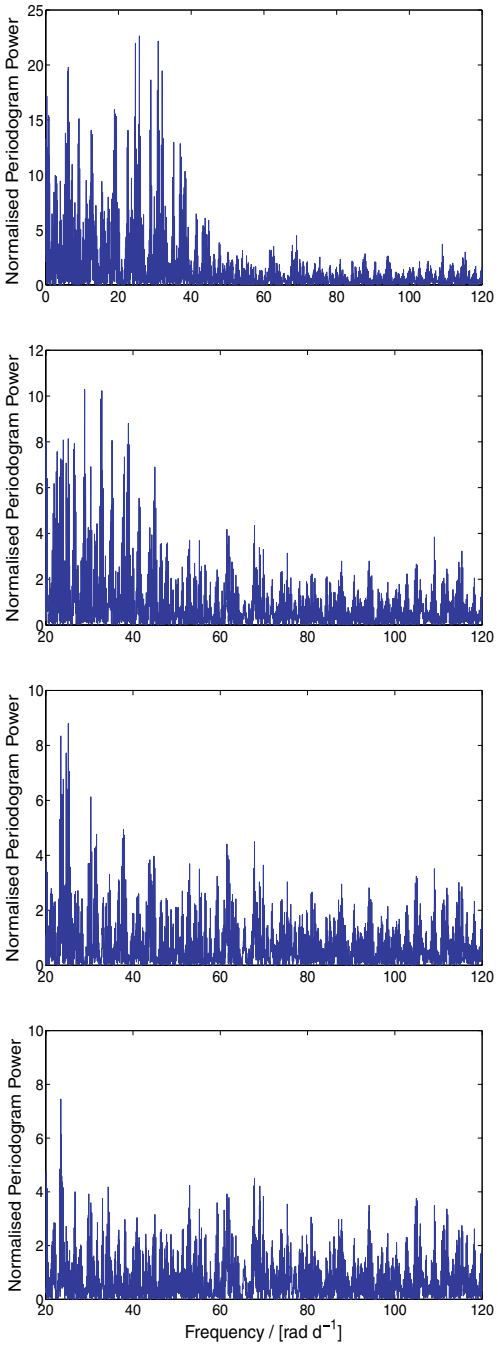

Figure 12. Periodograms calculated after successive pre-whitenings, as discussed in the text.

velocity of 50 km s^{-1} for V403 Car, while Saio (1981) published coefficients for rotational splitting in a pulsating star modelled as a polytrope. Reese, Lignieres & Rieutord (2006) recently found results very similar to those of Saio. Pamyatnykh (2003) found seven or eight excited pulsation modes in his analysis of a 12 $M_\odot$ model Beta Cephei star at the same effective temperature as V403 Car, at least among the radial, dipole and quadrupole modes. Our observed triplet is very unlikely to be due to rotational splitting. In the first place, if one assumes (m, $-m$) splitting (for example, $m = 1$ and $m = -1$, respectively), then the asymmetry in our splitting is the inverse of what it should be: the second-order rotational splitting

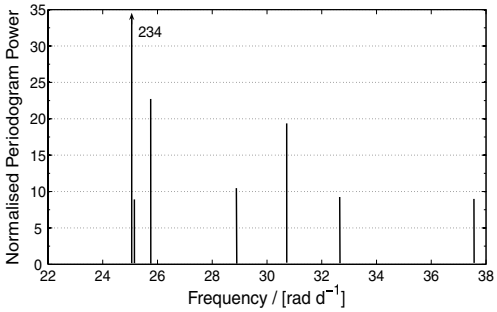

Figure 13. Frequencies for V403 Car, and their corresponding powers, found by the pre-whitening process described in the text.

term should split the $m = 0$ and $m = -1$ modes less than the $m = 0$ and $m = 1$ modes, which is the opposite of what we observe. Secondly, for the physical parameters of V403 Car, and using Saio's values for dipole and quadrupole modes predicted by Pamyatnykh, the splitting between the $m = 0$ and $|m| = 1$ modes should lie in the range of approximately 0.6–0.7 rad d^{-1} for the lowest possible ($\sin i = 1$) rotation velocity of 50 km s^{-1}. To accommodate the observed separation of 1.8–1.9 rad d^{-1} in our observed triplet, one requires an actual rotation velocity of about 150 km s^{-1} for $|m| = 1$, which would correspond to a larger asymmetry in the triplet due to second-order terms than is actually observed. The second-order terms will be smaller if the triplet consists of the $m = 0$ and $|m| = 2$ modes, but there is no compelling reason for the $|m| = 1$ modes to be missing in that case.

We conclude that our observed triplet is not due to rotational splitting. Table 3 includes three potential pulsation frequencies (f_1, f_5 and f_6) with overwhelmingly large significance. Frequencies f_7 through to f_{10} are more doubtful, although they still have significances of 90 per cent or higher. These frequencies could potentially be reconciled with seven of the excited modes in Pamyatnykh's model. However, more detailed modelling, and better observational analysis, of V403 Car will be necessary to establish a precise identification of the excited pulsation modes in this star.

11 SUMMARY AND CONCLUSIONS

Currently available theoretical false alarm probability functions are all derived from what appear to be reasonable assumptions about the data to be analysed and about the periodograms that they yield. Their validity, reliability and usefulness therefore strongly depend on how well these assumptions are met in practice.

A key assumption made by all authors is that the frequency range inspected in the periodogram contains a set of N_i frequencies $\omega_k : k = 1, \ldots, N_i$ at which the periodogram powers $P_X(\omega_k)$ are mutually independent. In theoretical statistical analysis, we have little hope of obtaining a false alarm probability function in the absence of this assumption. Without independence, very few general statistical results are available, and none is relevant to the problem at hand. The assumption of the existence of independent frequencies is therefore necessary in any theoretical discussion of the problem of significance of periodogram peaks and poses the first and most important obstruction to its resolution.

The existence of independent frequencies is guaranteed when the data are evenly spaced. We should therefore be able to test the validity of proposed false alarm probability functions for this case against the results of Monte Carlo simulations. Reasonable requirements on candidate functions include a good fit to the empirical CDFs, and their ability to predict correctly the number of independent frequencies known to exist from the theory.

Though they seem not to have viewed their work in this light, HB effectively performed this test for Scargle's false alarm function. They constructed the empirical CDF for periodogram peak heights produced by a pure noise time series consisting of N_0 evenly spaced data points, and fitted the Scargle false alarm function to it by least squares using the number N_i of independent frequencies as the fitting parameter. According to the theory, they should have obtained $N_i \leq [N_0/2]$. However, their results consistently yielded $N_i > N_0$. HB did not comment on this anomaly.

We have repeated their simulations, obtaining results similar to theirs only for gross oversampling of the periodogram, and only for data sets with $N_0 \leq 170$ data points. For gross oversampling and data sets with $N_0 > 170$, we were unable to reproduce their results. The values of N_i obtained by HB are consistently and systematically larger than ours. In our simulations, the best-fitting value of N_i increases linearly with N_0, in conflict with the quadratic dependence claimed by HB. Inspection of a plot of the values published in HB appears to indicate that their points lie on two straight lines, with a disjunction of slope at $N_0 = 170$ data points. We conjecture from these results that HB constructed their empirical CDFs by gross oversampling of the periodogram. This might explain why they consistently obtained $N_i > N_0$. We also conjecture that the sharp disjunction in slope at $N_0 = 170$, which is not observed in our simulations, is due to a systematic error in theirs. If so, the quadratic dependence of N_i on N_0, sometimes exploited by astronomers in the analysis of their data, might not be a real feature of real astronomical data but rather a spurious artefact of the HB simulations.

Given the assumption of independence that lies at the heart of Scargle's derivation of his false alarm probability function, it seemed unreasonable to suppose that it would provide an adequate description of the empirical CDFs obtained by oversampling the periodograms. Accordingly, we initially ran the HB simulations for CDFs constructed by sampling the periodograms only at the natural frequencies. The best-fitting values of N_i were very close to the theoretically expected value of $[N_0/2]$. The best-fitting Scargle functions were very poor fits to the empirical CDFs, displaying large deviations from the empirical CDFs in the domain of most interest when assessing the significance of periodogram peaks. The theoretical false alarm functions were consistently substantially higher in value than the empirical CDFs, leading to severe underestimation of peak significance. This same behaviour was observed for the best-fitting curves to the CDFs constructed by oversampling the periodograms. Researchers using the Scargle false alarm function, with or without the HB algorithm, are thus at significant risk of rejecting peaks that reflect real periodicities in their data.

A flaw in Scargle's derivation of his false alarm function was pointed out by Koen (1990) and by Schwarzenberg-Czerny (1996): Scargle assumes that the variance σ_X^2 of the noise is known a priori. This condition is not satisfied either in the simulations (where we sample pure pseudo-Gaussian noise), or in real data sets (where the data variance must be estimated from the data themselves). Both Koen and Schwarzenberg-Czerny correct this error in their respective treatments of the problem. Koen (1990) concludes that Scargle's false alarm function should be replaced by the Fisher (or Fisher–Snedecor) distribution. Schwarzenberg-Czerny (1998) pointed out that the Fisher distribution is applicable only for ratios of independent random variables. With ratios of random variables that are not independent, the Fisher distribution must be replaced by the incomplete β-function. He also showed that, in the case of the Lomb–Scargle periodogram, the correct distribution is given by

the incomplete β-function. On the strength of the work of these authors, we tested Schwarzenberg-Czerny's proposed function on CDFs constructed by oversampling periodograms and also on CDFs constructed by sampling only at the natural frequencies. In both cases, we have found the best-fitting Schwarzenberg-Czerny function, obtained by the HB algorithm, consistently to fit the empirical CDFs far more closely than Scargle's function, with impressively good agreement on all but the smallest data sets, where the theoretical function deviates only slightly from the empirical CDFs.

In spite of the excellent fits provided by the Schwarzenberg-Czerny false alarm function, our simulations display an alarming feature: the best-fitting values of N_i that yield such excellent agreement with the empirical CDFs are all consistently higher than the theoretically expected value of $[N_0/2]$. This is not unexpected for CDFs constructed by oversampling. In the case of CDFs constructed by sampling only at the natural frequencies, however, this result is in conflict with the theory. This means that, as in the case of the Scargle function, we cannot interpret the best-fitting value of the parameter N_i as the number of independent frequencies. It must be regarded rather as a fitting parameter in a one-parameter family of candidate CDFs that fit the empirical CDFs better than Scargle's candidate functions. Note that the Schwarzenberg-Czerny false alarm functions, when constructed independently of simulations and by using a priori theoretical values rather than fitted values for N_i, badly overestimate the significance of periodogram peaks and may result in the acceptance of spurious peaks as genuine. Unqualified confidence in analytical single-trial probability distributions in the construction of false alarm probability functions thus seems to be misplaced.

Ultimately, our principal interest is in the case of unevenly spaced data, not data that are evenly spaced. The loss of independence of the variables $P_X(\omega)$ in this case calls into question the validity and the expediency of searching for a formula in closed form for a false alarm probability function. All formulae proposed hitherto are based on the assumption of the existence of a set of mutually independent periodogram powers. This assumption is not realistic in uneven sampling schemes, as shown by Koen (1990). Thus, theoretical probability distributions provide no predictive power in determining false alarm criteria appropriate to a given data set which is independent of the empirical CDFs generated by simulations.

In the final analysis, what we need is a reliable false alarm probability function. Though we do not possess this function as a closed-form formula, we nevertheless have a numerical plot of it in the form of the CDF of maximum peak heights. This plot can be used as easily as any formula to get the answers that we want. If we insist on having such a formula to facilitate significance estimation, the empirical CDF can be fitted in the region of interest by any number of candidate fitting functions. This obviates the need for a theoretical formula.

We have applied the method of significance testing using an empirical CDF, as discussed in earlier sections of this paper, to the Beta Cephei star V403 Car. To date, only one (Stankov & Handler 2005) or two (Engelbrecht 1986) pulsation frequencies have been identified in this star. The analysis presented in this paper indicates that as many as seven pulsation frequencies can be attributed with a significance above 90 per cent.

ACKNOWLEDGMENTS

We sincerely thank Mike Gaylard, Chris Koen and Melvyn Varughese for their critical reading of draft versions of the manuscript. We also thank the South African SKA Office in Johannesburg for the use of their facilities.

REFERENCES

Baglin A., Auvergne M., Barge P., Buey J.-T., Catala C., Michel E., Weiss W., COROT Team, 2002, in Battrick B., Favata F., Roxburgh I. W., Galadi D., eds, ESA SP-485, Proc. First Eddington Workshop on Stellar Structure and Habitable Planet Finding. ESA, Noordwijk, p. 17
Balona L. A., 1975, Mem. R. Astron. Soc., 78, 51
Baluev R. V., 2008, MNRAS, 385, 1279
Barning F. J. M., 1963, Bull. Astron. Inst. Neth., 17, 22
Benlloch S., Wilms J., Edelson R., Yaqoob T., Staubert R., 2001, ApJ, 562, L121
Breger M. et al., 1993, A&A, 271, 482
Claudi R. U. et al., 2005, A&A, 429, L17
Cumming A., Marcy G. W., Butler R. P., 1999, ApJ, 526, 890
De Cat P. et al., 2007, A&A, 463, 243
Deeming T. J., 1975, Ap&SS, 36, 137
Engelbrecht C. A., 1986, MNRAS, 223, 189
Enoch M. L., Brown M. E., Burgasser A. J., 2003, AJ, 126, 1006
Falter S., Heber U., Dreizler S., Schuh S. L., Cordes O., Edelmann H., 2003, A&A, 401, 289
Feast M. W., 1958, MNRAS, 118, 618
Heynderickx D., Waelkens C., Smeyers P., 1994, A&AS, 105, 447
Horne J. H., Baliunas S. L., 1986, ApJ, 302, 757 (HB)
Koen C., 1990, ApJ, 348, 700
Lamm M. H., Bailer-Jones C. A. L., Mundt R., Herbst W., Scholz A., 2004, A&A, 417, 557
Lomb N. R., 1976, Ap&SS, 39, 447
Pamyatnykh A. A., 2003, Ap&SS, 284, 97
Pojmanski G., 1998, Acta Astron., 48, 35
Priestley M. B., 1981, Spectral Analysis and Time Series, Vols 1 and 2. Elsevier Academic Press, London
Reegen P., 2007, A&A, 467, 1353
Reegen P., Gruberbauer M., Schneider L., Weiss W. W., 2008, A&A, 484, 601
Reese D., Lignieres F., Rieutord M., 2006, A&A, 455, 621
Saio H., 1981, ApJ, 244, 299
Scargle J. D., 1982, ApJ, 263, 835
Schuster A., 1898, Terr. Magn., 3, 13
Schuster A., 1906, Phil. Trans. R. Soc. Lond., 206, 69
Schwarzenberg-Czerny A., 1996, ApJ, 460, L107
Schwarzenberg-Czerny A., 1998, MNRAS, 301, 831
Smolec R., Moskalik P., 2007, MNRAS, 377, 645
Stankov A., Handler G., 2005, ApJS, 158, 193
Tackett S., Herbst W., Williams E., 2003, AJ, 126, 348
Thompson R. O. R. Y., 1971, IEEE Trans. Geosci. Electron., GE-9, 107
Turner D. G., Grieve G. R., Herbst W., Harris W. E., 1980, AJ, 85, 1193
Vanicek P., 1969, Ap&SS, 4, 387
Wen L., Levine A. M., Corbet R. H. D., Bradt H. V., 2006, ApJS, 163, 372
Yabushita S., 2004, MNRAS, 355, 51

APPENDIX A: CLASSICAL PERIODOGRAM

As originally conceived, the classical periodogram may be regarded essentially as a Fourier power spectrum estimator for an infinite continuous-time signal $X(t)$ that has been discretely sampled for a finite time at equally spaced time intervals. The data for this estimator form a finite discrete time series consisting of N values $X_i = X(t_i)$, $i = 1, \ldots, N$, of the physical parameter X at times $t_i = t_0, t_0 + \Delta t, t_0 + 2\Delta t, \ldots, t_0 + (N-1)\Delta t$. The DFT, DFT $_X(\omega)$, of this time series X_i, which is defined by

$$\mathrm{DFT}_X(\omega) = \sum_{r=1}^{N} X(t_r)\, \mathrm{e}^{-\mathrm{i}\omega t_r}, \tag{A1}$$

may be regarded as an estimator of the FT $\mathrm{FT}_X(t)$ of $X(t)$. The power spectral density of the signal may then be estimated by the function

$$|\mathrm{DFT}_X(\omega)|^2 ,$$

with some suitably chosen normalizing coefficient. A commonly used normalization is

$$\mathrm{CP}_X(\omega) = \frac{1}{N} |\mathrm{DFT}_X(\omega)|^2 = \frac{1}{N} \left| \sum_{r=1}^{N} X(t_r) \, \mathrm{e}^{-\mathrm{i}\omega t_r} \right|^2 . \tag{A2}$$

A simple calculation then yields the formula

$$\mathrm{CP}_X(\omega) = \frac{1}{N} \left[\left(\sum_{r=1}^{N} X(t_r) \cos \omega t_r \right)^2 + \left(\sum_{r=1}^{N} X(t_r) \sin \omega t_r \right)^2 \right] . \tag{A3}$$

Following Scargle (1982), we call this function the classical periodogram. This definition agrees with that given originally by Schuster in Schuster (1898), but not with that in his later publications. It also agrees with the definitions used in Thompson (1971) and Deeming (1975), and differs by a factor of 2 from that used by Priestley (1981).

It is easy to see from equation (A2) why the classical periodogram is useful in identifying the frequencies of harmonic components in the signal X. Suppose X contains an harmonic component of frequency $\tilde{\omega}$. Then, when ω is very different from $\tilde{\omega}$, $X(t)$ and $\mathrm{e}^{-\mathrm{i}\omega t}$ are out of phase, and the product $X(t)\,\mathrm{e}^{-\mathrm{i}\omega t}$ oscillates rapidly. The sum of the products $X(t_r)\,\mathrm{e}^{-\mathrm{i}\omega t_r}$, which is a discrete estimator of the integral $\int X(t)\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t$ will thus have a value close to zero, albeit masked by whatever other signal is present in $X(t)$. As ω approaches the value of $\tilde{\omega}$, the factors $X(t)$ and $\mathrm{e}^{-\mathrm{i}\omega t}$ get closer in phase, so the product $X(t)\,\mathrm{e}^{-\mathrm{i}\omega t}$ oscillates more slowly. The value of the sum of the products $X(t_r)\,\mathrm{e}^{-\mathrm{i}\omega t_r}$ will thus rise, reaching a maximum at $\omega = \tilde{\omega}$. The presence of a harmonic signal of frequency $\tilde{\omega}$ thus produces a peak in the periodogram with maximum at $\tilde{\omega}$.

The converse however is not true. A peak in the periodogram does not necessarily reflect the presence of an harmonic component in the signal X. Peaks might be produced by other effects. Thus, the presence of measurement error, signal noise or random physical processes in the observed system might, by a spurious random fluctuation, also produce a peak. Peaks may also be produced by aliasing and/or spectral leakage, and the observing window. The potential for producing peaks that are not due to harmonic components in the observed signal makes the interpretation of peaks in the periodogram very difficult and presents many hazards and pitfalls for the unwary. The dangers posed by these effects were already noted by Schuster as early as 1906, '. . . it has generally been assumed that each maximum in the amplitude of a harmonic term corresponded to a true periodicity. The extent to which this fallacious reasoning has been made use of would surprise anyone not familiar with the literature of the subject.' (Schuster 1906, p. 71, 72). Strangely, his warning has often been ignored, and sometimes even disdainfully brushed aside.

APPENDIX B: LOMB–SCARGLE PERIODOGRAM

Following Scargle (1982, appendix B), we define the Lomb–Scargle periodogram by the formula

$$P_X(\omega) = \frac{1}{2} \left\{ \frac{\left[\sum_{i=1}^{N} x_i \cos \omega(t_i - \tau) \right]^2}{\sum_{i=1}^{N} \cos^2 \omega(t_i - \tau)} + \frac{\left[\sum_{i=1}^{N} x_i \sin \omega(t_i - \tau) \right]^2}{\sum_{i=1}^{N} \sin^2 \omega(t_i - \tau)} \right\} , \tag{B1}$$

where the epoch translation parameter $\tau(\omega)$ is defined implicitly by the formula

$$\tan(2\omega\tau) = \frac{\sum_{i=1}^{N} \sin(2\omega t_i)}{\sum_{i=1}^{N} \cos(2\omega t_i)} . \tag{B2}$$

The data used to calculate $P_X(\omega)$ form a finite discrete time series consisting of N values $X_i = X(t_i)$, $i = 1, \ldots, N$, of the physical parameter X measured at times $\{t_i \mid i = 1, 2, \ldots, N\}$ which are arbitrarily spaced in time. Lomb (1976), following Barning (1963) and Vanicek (1969), arrived at this formula via a least-squares fitting procedure in which sampled values $X(t_i)$ are fitted with an harmonic signal of frequency ω. For these three authors, therefore, $P_X(\omega)$ does not represent an attempt at estimating the Fourier power spectrum of any continuous time physical signal $X(t)$, but is rather a spectral best-fitting parameter that displays how closely the data may be fitted with a single harmonic function of frequency ω. The larger the value of $P_X(\omega)$, the better the fit.

In contrast with these authors, Scargle (1982) arrived at the same spectral function by first relaxing the definition of the DFT for application to the case of unevenly spaced data (Scargle 1982, appendix A), and then imposing two demands on the periodogram (which he calls the modified, or generalized periodogram) calculated from this relaxed definition: the statistical distribution of the generalized periodogram will be made as closely as possible the same as it is in the evenly spaced case; and, the generalized periodogram will be made invariant to translations in time. These two requirements yield uniquely the formulae in equations (B1) and (B2). Arguably, this restores the interpretation of the modified periodogram as an estimator of the power spectrum of the physical signal $X(t)$ in the case where the signal is unevenly sampled in time. However, it is probably more accurate to regard it as a spectral goodness-of-fit parameter. This view also enables one better to understand a variety of other, alternative, periodogram formulae currently offered in the literature.

APPENDIX C: PURE NOISE

A random process $X(t)$ is said to be a purely random process, pure noise or white noise, if it consists of a sequence of uncorrelated random variables. This means that, for all $t' \neq t$,

$$\mathrm{cov}[X(t), X(t')] = 0.$$

Pure noise is the simplest of all random process models. It corresponds to a case where the process has 'no memory' in the sense that the value of the random variable $X(t)$ at time t has no relation whatever to its value $X(t')$ at any other time t', no matter how close or distant t and t' are to each other. In this sense, $X(t)$ neither remembers its past, nor is aware of its future. Knowing the value of $X(t_0)$ at any time t_0 therefore provides no way at all, other than

by the probability distribution $p_{X(t)} = p(x, t)$, of predicting within reasonable limits and uncertainties the value of $X(t)$ at time t. This is to be contrasted with correlated noise where the values $X(t)$ and $X(t')$ are in general related or 'correlated'. In this case, we can do better in predicting the value of $X(t + \tau)$ from $X(t)$ than in the case of uncorrelated, or pure, noise. From a knowledge of the value $X(t)$, we can set narrower limits on the probable values of $X(t + \tau)$ than is possible from the distribution $p_{X(t+\tau)}$ alone (Priestley 1981, p. 114).

Pure noise is said to be Gaussian if the random variables $X(t)$ are jointly normally distributed. Noise of this kind is often called Gaussian white noise. In this case, the random variables $\{X(t)\}$ are also mutually independent.

Note that some authors define pure noise more stringently. For them, a random process $X(t)$ is pure noise if the random variables $\{X(t)\}$ are independent, and identically distributed with zero mean.

In this paper, a data set $\{X_k \mid k = 1, 2, \ldots, N_0\}$ is said to be pure noise if the values X_k are independent, and identically distributed random variables with zero mean. For simplicity, we assume also that the X_k are each normally distributed. Denote their common variance by σ_X^2. Since the X_k have zero mean, their covariance matrix is given by

$$C_{jk} = E[(X_j - \mu_{X_j})(X_k - \mu_{X_k})] = E[X_j X_k] = \sigma_X^2 \delta_{jk}. \qquad \text{(C1)}$$

This paper has been typeset from a TEX/LATEX file prepared by the author.