THE ASTROPHYSICAL JOURNAL, 224:953–960, 1978 September 15

PERIOD DETERMINATION USING PHASE DISPERSION MINIMIZATION

R. F. STELLINGWERF
Department of Physics and Astronomy, Rutgers University
Received 1978 February 6; accepted 1978 March 22

ABSTRACT

We derive a period determination technique that is well suited to the case of nonsinusoidal time variation covered by only a few irregularly spaced observations. A detailed statistical analysis allows comparison with other techniques and indicates the optimum choice of parameters for a given problem. Application to the double-mode Cepheid BK Cen demonstrates the applicability of these methods to difficult cases. Using 49 photoelectric points, we obtain the two primary oscillatory components as well as the principal mode-interaction term; the derived periods are in agreement with previous estimates.

Subject headings: stars: Cepheids — stars: individual — stars: pulsation

I. INTRODUCTION

In connection with a survey of RR Lyrae stars, Lafler and Kinman (1965, hereafter LK) proposed a technique of period determination designed for use on an electronic computer. This technique is simply an automated version of the classical method of distinguishing between possible periods, in which the period producing the least observational scatter about the mean (derived) light curve is chosen. Apart from its historical flavor, this approach is appealing for a number of reasons. It is well suited to cases in which only a few observations are available over a limited period of time, especially if the light curve is highly nonsinusoidal. An optimum light-curve shape is obtained, which can be subtracted from the data to allow a search for other periods to be made (Stobie 1972; Stobie and Hawarden 1972). Also, the computation is very straightforward, allowing complete automation of the period search. This technique has recently been applied to the double-mode Cepheid U TrA by Henden (1976) and to an analysis of theoretical variability by Cox, Hodson, and King (1978).

Fourier techniques have also been applied to period determination in variable stars (Wehlau and Leung 1964; Fitch 1967). They are particularly useful in complicated cases, especially if deconvolution is needed or if spectral information is required over a continuum of frequencies. If a power spectrum has been obtained and indicates the presence of discrete lines, then the LK approach can be used to obtain accurate periods and nonsinusoidal light curves for the component oscillations (Ostriker and Hesser 1968; Hesser, Ostriker, and Lawrence 1969). In the case of a very complicated spectrum, the maximum entropy technique (Percy 1977) should be considered.

In this paper a generalization of the LK technique is presented. This generalization greatly increases the utility of the algorithm, allows an arbitrary degree of smoothing, and provides complete statistical information. A discussion of statistical significance is given by LK, but their results are model-dependent and essentially empirical. The need for smoothing can be seen by comparing the spectral curve in LK (Fig. 3) with those computed by Stellingwerf (1975, Fig. 1), who used a smooth theoretical statistic. The smoothing is accomplished in the present paper in a natural way, by increasing the number of degrees of freedom of the algorithm.

II. CHARACTERISTICS OF THE PHASE DISPERSION MINIMIZATION METHOD

a) *Definitions*

A discrete set of observations can be represented by two vectors, the magnitudes $\boldsymbol{x}$ and the observation times $\boldsymbol{t}$, where the ith observation is given by (x_i, t_i) and there are N points in all ($i = 1, N$). Let σ^2 be the variance of $\boldsymbol{x}$, given by

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}, \qquad (1)$$

where $\bar{x}$ is the mean; $\bar{x} = \sum x_i/N$. For any subset of x_i we define the sample variance s^2 exactly as in equation (1). Suppose we have chosen M distinct samples, having variances s_j^2 ($j = 1, M$) and containing n_j data points. The overall variance for all the samples is then given by

$$s^2 = \frac{\sum (n_j - 1)s_j^2}{\sum n_j - M} \qquad (2)$$

as a consequence of equation (1).

We wish to minimize the variance of the data with respect to the mean light curve. Let Π be a trial period, and compute a phase vector $\boldsymbol{\phi}$: $\phi_i = t_i/\Pi - [t_i/\Pi]$; here brackets indicate the integer part. Equivalently, $\boldsymbol{\phi} = \boldsymbol{t} \bmod (\Pi)$. We now pick M samples from $\boldsymbol{x}$ using the criterion that all the members of sample j have similar ϕ_i. Usually the full phase interval (0, 1) is divided into fixed bins, but the samples may be chosen in any way that satisfies the criterion. All points need not be picked, or, alternatively, a point can belong to many samples. The variance of these samples gives a measure of the scatter around the mean light curve defined by the means of the x_i in each sample, considered as a function of ϕ. We define the statistic

$$\Theta = s^2/\sigma^2 , \tag{3}$$

where s^2 is given by equation (2) and σ^2 is given by equation (1). If Π is not a true period, then $s^2 \approx \sigma^2$ and $\Theta \approx 1$, whereas if Π is a correct period, Θ will reach a local minimum compared with neighboring periods, hopefully near zero.

Since this technique seeks to minimize the dispersion of the data at constant phase, we will refer to it as "phase dispersion minimization" (PDM). Mathematically, this is a least-squares fitting technique, but rather than a fit to a given curve (such as a Fourier component), the fit is relative to the *mean curve* as defined by the means of each bin. We simultaneously obtain the best least-squares light curve and the best period. The PDM technique is thus a "Fouriergram" method (as discussed by Faulkner 1977*a*) of infinite order, since all harmonics are included in the fitted function. The Fourier series technique, a least-squares fit to a truncated series with variable amplitudes and phase, often requires additional constraints and rather high orders for nonsinusoidal variations (Lucy 1976).

In an interesting, related method discussed by Whittaker and Robinson (1926, hereafter WR), one seeks the maximum variance of the bin means (as opposed to the mean of the bin variances). If s_m^2 is the variance of the bin means, define $\Theta_{\rm WR} = 1 - s_m^2/\sigma^2$ for comparison purposes. In general, $\Theta_{\rm WR}$ will vary between 0 and $1 - \langle 1/n_j \rangle$, where $\langle n_j \rangle$ is the mean number of points per bin. At a true period, $s_m^2 \approx \sigma^2$ and $\Theta_{\rm WR} = 0$. Here we seek periods at which the amplitude of the mean curve is a maximum, which in most cases will correspond to minimum phase dispersion. The calculation of σ_m^2 is easier than the calculation of s^2 (eq. [2]), but the much lower number of degrees of freedom suggests less sensitivity (see § II*c*). We will show below that, in most cases, this supposition is true, although for one range of parameters the WR technique may be preferable. The WR method is discussed in detail by Chapman and Bartels (1940) under the title "persistence analysis."

Although the individual samples may be chosen in many ways, it is convenient to define a standard bin structure. We divide the unit interval into N_b bins of length $1/N_b$, and we take N_c "covers" of N_b bins, each cover offset in phase by $1/(N_b N_c)$ from the previous cover, using periodic boundary conditions on the unit interval to obtain a uniform covering. We thus obtain $M = N_b N_c$ bins, each of length $1/N_b$, and whose midpoints are uniformly spaced along the unit interval at a distance of $1/(N_b N_c)$. Clearly, each data point will fall in exactly N_c bins. Denote a given bin structure by (N_b, N_c).

It is easily shown that the Θ statistic defined in LK is exactly twice that given by equation (3), provided that a bin structure of $(N/2, 2)$ is used and that bin widths are adjusted to always include exactly two data points per bin.

b) *Computation Time*

Each data point is included in N_c bins; so the running time per trial frequency is proportional to NN_c. Let N_f be the number of frequency points desired. N_f will depend on the total time base and the frequency range desired (see § II*d*). For large N the overall time will be roughly $5\mu N_f N_c N$, where μ is the multiplication time for the machine. No trigonometric lookups are required; so, if the s_j^2 computation is optimized, the calculation can be done very quickly.

c) *Statistical Properties*

Here we assume that the total "population" of possible observations of an object $\boldsymbol{X}$ is approximately normally distributed. Actually, the true distribution is a convolution of a nearly uniform distribution due to the time variation and a Gaussian noise component, but the statistical analysis is relatively insensitive to such subtleties.

To compute Θ, we take the ratio of variances of the two subsets of $\boldsymbol{X}$, that of the actual observations $\boldsymbol{x}$, and that of the bins. Therefore, Θ has a probability density given by an F distribution with $\sum n_j - M$ and $N - 1$ degrees of freedom. It is convenient to define F as a number greater than unity; so $F \equiv \Theta^{-1}$. The probability P that a given value of Θ is due to random fluctuations (also called the "significance") is twice the area of the F distribution above Θ^{-1} (two-sided test). This probability approaches unity as $\Theta \to 1$. Thus, for significance P, we compute

$$F_{(P/2, N_{1f}, N_{2f})} = 1/\Theta , \quad N_{1f} = N - 1 , \quad N_{2f} = \sum n_j - M . \tag{4}$$

P may then be obtained by reference to an F table or by using an approximation to $P(F, N_{1f}, N_{2f})$ (see § 26 of Abramowitz and Segun 1965).

If N is large (>100), we may take $\sigma^2(\boldsymbol{x}) \approx \sigma^2(\boldsymbol{X})$. In this case we may use the somewhat simpler χ^2 test:

$$\chi^2{}_{(P/2, N_f)} = N_f \Theta , \quad N_f = \sum n_j - M . \tag{5}$$

Here P is twice the area of the χ^2 distribution *below* $N_f \Theta$.

These results may be applied to the analysis of AX Vel by Stobie and Hawarden (1972). These authors suppressed the annual sidelobes in an initial frequency scan by computing s^2 separately for each year's observations and combining them according to equation (2) above. Their Figures 2 and 3 show $\Theta_{\rm LK}$ computed in this way, using data spanning 4 years and $N = 60$. The LK technique, with $(N/2, 2)$ bin structure, yields $N_f = N$ degrees of freedom, which then must be decreased by the number of annual means estimated. For an F test we therefore have $N_{1f} = 59$, $N_{2f} = 56$. Including the factor of 2 (§ II*a*), the three Θ minima shown in Figure 3 of Stobie and Hawarden (1972) yield $\Theta_{\min} \approx 0.60$, 0.40, and 0.63. Consulting an F table, we find significances of 0.05, less than 0.01, and 0.10. The middle period is thus highly significant; the others are marginally significant. The first minimum occurs at half the frequency of the second (most significant) minimum and is undoubtedly the first subharmonic. It will be shown below that this spurious minimum could have been suppressed through the use of a larger bin width.

A statistical criterion may also be derived for the WR technique. The distribution of bin means is normal if M is large with $\sigma_m^2 = \sigma^2/n_j$. Since M is generally *not* large, however, a t distribution with $M - 1$ degrees of freedom will actually be obtained, with $\sigma_m^2 \approx (M-1)/(M-3)\sigma^2/n_j$. We may always select bins with $n_j = N/N_b$, for all j. Note that normal constant-phase bins will produce binomially distributed n_j and will increase σ_m^2, since a harmonic mean of n_j appears in σ_m^2. We wish to test whether the observed s_m^2 is significantly larger than σ_m^2. Evidently s_m^2/σ_m^2 follows an F distribution with $M - 1$ and $N - 1$ degrees of freedom. Using the definition of Θ_{WR}, we have

$$F_{(M-1,N-1)} \approx \frac{(M-3)}{(M-1)}\frac{N}{N_b}(1-\Theta_{\mathrm{WR}}) . \tag{6}$$

d) *Line Characteristics*

The character of Θ near a minimum (i.e., a spectral line) is easily obtained. An oscillation with amplitude A will produce a maximum variance in x of $\sigma_0^2 = A^2/12$, since the resulting distribution is approximately uniform. In addition, suppose x contains a variance due to "noise" (observational errors, other periods, etc.) given by σ_N^2. We then have signal-to-noise ratio $\epsilon = \sigma_0/\sigma_N$ and overall variance $\sigma^2 = \sigma_0^2 + \sigma_N^2$.

The bin variance s^2 will depend on the distance from line center ($\Delta\Pi$ in trial period or Δf in trial frequency) and will also contain contributions from noise and bin width. Let $g(\phi)$ represent the mean light curve as a function of phase, with average slope $\langle g' \rangle \approx \alpha A$. Here α is a parameter depending on curve shape; in general, $\alpha \approx 2$. Numerical tests give $\alpha = 1.7$ for a sine wave, $\alpha = 2$ for a picket-fence function, and $\alpha = 2.7$ for a sawtooth mean curve, all relative to a (5, 2) bin structure. We obtain $\langle g' \rangle \approx (12)^{1/2}\alpha\sigma_0$. Now a change in trial period of $\Delta\Pi$ will induce a scatter in phase over a range $\Delta\phi = 0 \rightarrow (\Delta\Pi/\Pi)(T/\Pi)$, where $T = t_N - t_1$ is the time base. We thus obtain for the induced variance

$$s_\phi^2 = \frac{1}{12}\left(\frac{T}{\Pi}\frac{\Delta\Pi}{\Pi}\right)^2 \langle g' \rangle^2 = \left(\frac{T}{\Pi}\frac{\Delta\Pi}{\Pi}\right)^2 \alpha^2\sigma_0^2$$
$$= (T\Delta f)^2\alpha^2\sigma_0^2 , \tag{7}$$

using $\Delta\Pi/\Pi = \Delta f/f$ for $f = 1/\Pi$. Since each bin width is $1/N_b$, the variance caused by variation of g across a bin is

$$s_b^2 = \frac{1}{12}\left(\frac{g'}{N_b}\right)^2 = \frac{\alpha^2\sigma_0^2}{N_b^2} , \tag{8}$$

and the total bin variance is given by

$$s^2 = s_\phi^2 + s_b^2 + \sigma_N^2 .$$

We therefore find

$$\Theta_{\mathrm{line}} = \frac{s_\phi^2 + s_b^2 + \sigma_N^2}{\sigma_0^2 + \sigma_N^2}$$
$$= \frac{1 + \epsilon^2\alpha^2[1/N_b^2 + (T\Delta f)^2]}{1 + \epsilon^2} . \tag{9}$$

The line shape near minimum is parabolic. If we extend this parabola to the $\Theta = 1$ level, we obtain the half-width for the line:

$$\Delta f_{1/2} = \frac{1}{T}\left(\frac{1}{\alpha^2} - \frac{1}{N_b^2}\right)^{1/2} . \tag{10}$$

For $N_b \geq 5$ we may ignore N_b in equation (10), which yields $f_{1/2} \approx 1/(2T)$, in agreement with the results of Stobie and Hawarden (1972). The half-width in terms of trial period is $\Delta\Pi_{1/2} \approx \Pi^2/(2T)$. This strong period dependence is clearly visible in the rather formidable aliased period shown in Figure 3 of LK. The true minimum and the spurious minimum would show similar widths if plotted versus frequency rather than period.

The PDM technique finds all periodic components; so *subharmonics* ($f_n = f_1/n$, where f_1 is the principal frequency) will also be found. For the nth subharmonic, however, $\alpha_n \approx 2n$ and the half-width is $(\Delta f_n)_{1/2} \approx 1/(2Tn)$, which implies that subharmonics should be identifiable by their narrow line widths (as well as by a doubly periodic mean curve). Note that *harmonics* ($f^n = nf_1$), if detected at all, will have n-valued mean curves (n-peaked distribution in each bin) and line widths given by equation (10).

The statistical significance of the line will be given by the value of Θ_{line} at $\Delta f = 0$; equation (9) becomes

$$\Theta_{\min} = \frac{1 + \epsilon^2\alpha^2/N_b^2}{1 + \epsilon^2} . \tag{11}$$

For observations with a low noise level, ϵ is large and equation (11) becomes $\Theta_{\min} = \alpha^2/N_b^2 \approx 4/N_b^2$ for the principal minimum and $\Theta_{\min} \approx 4n^2/N_b^2$ for subharmonics. Coarse bin size can therefore be used to suppress subharmonics. In fact, f_n with $n \gtrsim N_b/2$ will not appear in the spectrum, since the variation of g becomes A within each bin.

These remarks are illustrated by Figure 1, which shows the computed Θ transforms for (Fig. 1*a*) a sine-wave [$\sin(2\pi ft)$] and (Fig. 1*b*) a "saw-tooth" function (fractional part of ft). These functions have been selected because they represent the two extremes found in variable-star light curves. In each case x consisted of 201 data points evenly distributed over 10 periods with $f = 1$, $T = 10$. A bin structure of $(N_b, N_c) = (5, 2)$ was used. The main line at $f = 1$ is virtually identical to the Fourier power spectrum line shape $[\sin(x)/x]^2$. The half-width is twice that given by equation (9), but a parabolic fit near $\Theta_{\min}$ does

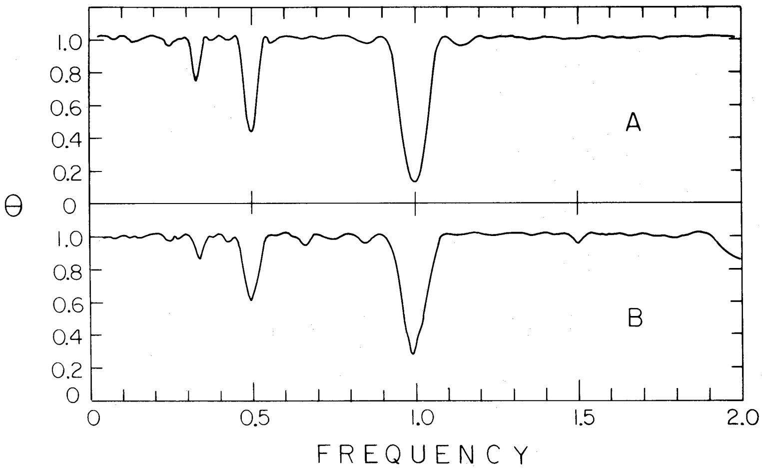

FIG. 1.—The Θ statistic versus frequency for the two test cases described in the text. (*a*) Sine-wave transform; (*b*) sawtooth function transform.

give the correct width when continued to $\Theta = 1$ (in practice, the upper regions of the line will be obscured by noise, and so a width based on a fit near $\Theta_{\min}$ is most practical). Two subharmonics are present (the third, just visible at $f = \frac{1}{4}$, is below the cutoff $n = N_b/\alpha \approx 3$ and is not significantly different from 1). The great similarity of the two curves shown in Figure 1 is, of course, the strong point of the PDM technique. Highly nonsinusoidal variations are handled very efficiently. The presence of subharmonic response could be a disadvantage if oscillations with widely spaced frequencies are present. We have shown, however, that, in practice, subharmonics can be distinguished in at least three ways: (1) light-curve shape, (2) narrow line widths, and (3) reduced significance with increasing bin size. If initial scans of the full frequency range use broad bin sizes [(5, 2), say], subharmonics should pose no problems.

The transforms in Figure 1 could be severely affected by nonuniform time coverage, as in the Fourier theory. In individual cases it is sometimes useful to compute a transform of $\sin(ft)$ with the given amplitude, period, and t values; this will indicate the magnitude and spacing of sidelobes and aliased frequencies. The estimated noise level can also be included in such a test (see Lucy 1976).

The Θ_{WR} transforms were also computed and found to be virtually identical to those of Figure 1. The minima were actually slightly deeper. We now compute the expected value of $(\Theta_{\mathrm{WR}})_{\min}$. Clearly, we have $\sigma^2 = \sigma_0^2 + \sigma_N^2$, as above. Also, at a true minimum, $s_m^2 \approx \sigma_0^2 + \sigma_N^2/n_j$. The effect of finite bin width here will be to smooth out the peaks of the light curve. If the light curve is roughly parabolic at maximum and minimum light, the amplitude of the mean light curve will be $A[1 - (\beta/N_b)^2]$, where β is a constant depending on the sharpness of the peaks. For a sine wave, $\beta = 1.28$; a sharp maximum requires $\beta \approx (\alpha N_b/2)^{1/2}$. We take $\beta \approx 2$, appropriate to a pure parabolic wave. If we put $\delta = (\beta/N_b)^2$, we then have

$$s_m^2 = \sigma_0^2(1 - \delta)^2 + \sigma_N^2 N_b/N\,.$$

We therefore find

$$(\Theta_{\mathrm{WR}})_{\min} = \frac{1 + N_b/N + \epsilon^2\delta(2 - \delta)}{1 + \epsilon^2}\,. \qquad (12)$$

For low noise, $(\Theta_{\mathrm{WR}})_{\min} \approx \delta(2 - \delta)$, slightly lower than equation (11) for smooth variations. We see from equation (12) that the WR approach is likely to run into trouble in cases with small N (small n_j) in the presence of noise.

e) *Period Determination with Small N*

If a large number of observations are available, statistical significance is usually not a consideration. Indeed, exactly the opposite problem is almost certain to arise: many possible periods resulting from aliasing and sidelobes, all of which are well above the noise level. For small N, however, statistical effects will be extremely important. We may ask, For a given instrumental sensitivity and number of observations N, what is the minimum amplitude that will yield a significant period? Or, if the amplitude of the object is known, how many observations will be required to determine a period?

The significance of a local Θ minimum may be obtained by equating $\Theta_{\min}$ in equation (10) to $1/F$ as given by equation (5). Solving the resulting expression for ϵ, we obtain

$$\epsilon = \left[\frac{F - 1}{1 - (\alpha/N_b)^2 F}\right]^{1/2}, \qquad (13)$$

F here is $F_{(P/2, N_{1f}, N_{2f})}$, where $N_{1f} = N - 1$ and $N_{2f} = N_c(N - N_b)$. Since F is a function of N, equation (13) relates N, P, and ϵ for a given bin structure.

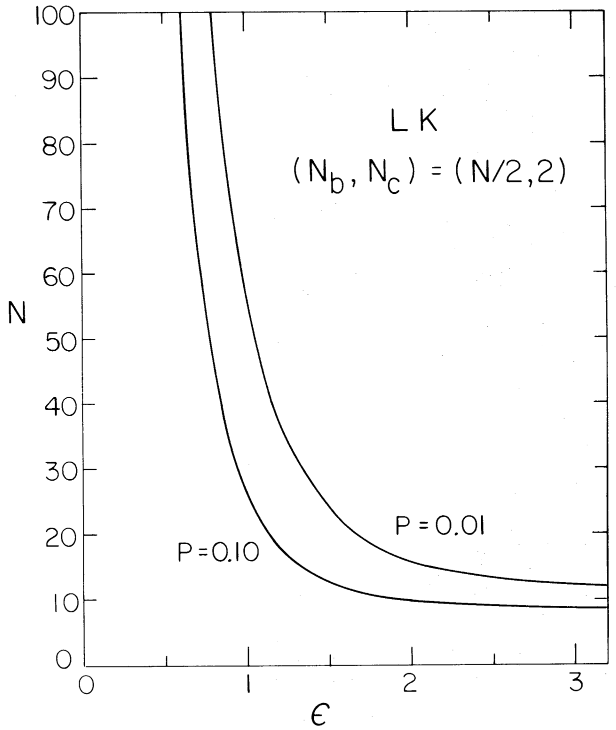

FIG. 2.—Statistical characteristics of the LK algorithm. Lines of constant significance (P) of a Θ minimum (spectral line) are shown on the (N, ϵ)-plane. Here N is the number of data points and ϵ is the signal-to-noise ratio.

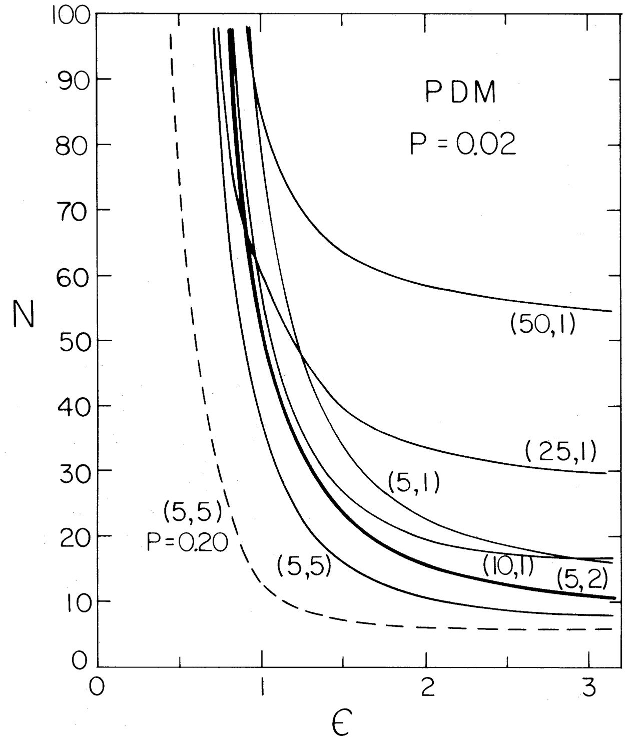

FIG. 3.—Statistical characteristics of the PDM method. Notation as in Fig. 2, with various bin structures (N_b, N_c) as indicated. Solid lines refer to P = 0.02; dashed line is P = 0.20.

Consider first the LK bin structure: $(N_b, N_c) = (N/2, 2)$ with two points per bin. Using equation (13), we have computed the relationship of ϵ and N for significance levels 0.10 (barely above noise level) and 0.01 (rather well determined) for the representative case $\alpha = 2$ (e.g., an RR*ab* variable of moderate amplitude). The results are shown in Figure 2. This figure is directly comparable to Figure 1 of LK, in which a semiempirical significance criterion is given. Note that, since $\sigma_0^2 \approx A^2/12$ and $\sigma_N = 0.15$ mag for the Lick survey, signal-to-noise ratios in LK are given by $\epsilon_{\rm LK} = A/0.52$ mag; so $\epsilon_{\rm LK}$ ranges from 1 to 3 in LK's Figure 1. If one considers just the $P = 0.01$ curve of Figure 2, the present results are slightly more optimistic than the LK results at high ϵ, indicating a minimum of 12 rather than 14 observations needed for $\epsilon = 3$. The $P = 0.10$ curve suggests that, with luck, 10 points might be sufficient. This assumes that adequate time base and phase coverage are present, and that the aliasing problem has somehow been avoided (e.g., an approximate period may be available). On the other hand, at small ϵ Figure 2 is more conservative than LK, calling for $N = 55$ at $\epsilon = 1$ compared to $N \approx 25$ according to Figure 1 of LK. This result is in good agreement with the actual distribution shown in Figure 2 of LK, however, which clearly shows that the LK relation is overly optimistic at small amplitudes (eq. [11] provides an excellent fit to the data of Fig. 2 of LK if used with α = 2–3, $\epsilon = A/0.52$, and doubled to match $\Theta_{\rm LK}$).

The effects of varying bin structures are shown in Figure 3, in which all solid curves represent a significance of 2%. It is found that $(N_b, N_c) = (5, 2)$ has sensitivity nearly identical to the LK $(N/2, 2)$ bin structure. At very small N ($N < 20$), Figure 3 indicates that coarse bin structure ($N_b < 10$) is essential. A comparison of the (5, 1), (5, 2), and (5, 5) curves clearly shows the advantage to be gained from multiple bin coverings at small N. Figure 3 indicates that a significant period may be detected with as few as 10 observations if $\epsilon > 2.5$ and if a (5, 5) bin structure is used. At larger N ($N > 100$), the sensitivity is determined by the bin number $M = N_b N_c$; so, if subharmonics are not a problem, the $N_c = 1$ choice is computationally most efficient and will be preferable. It is clearly very difficult to obtain periods if the signal-to-noise ratio falls below $\epsilon \approx 0.8$, although, since $F \to 1$ at large N, equation (13) indicates that $\epsilon \to 0$ in all cases, and *any* signal-to-noise ratio can be overcome with sufficiently large N. As an indication of the "theoretical limit" of sensitivity, the dashed curve in Figure 3 shows the (5, 5) result at $P = 0.20$ significance. Along this line there is a one-in-five chance that the signal will be lost in noise.

The WR technique performed well on the test cases in § II*d*; its statistical properties are therefore also of interest. Using equation (12) in equation (6), we may

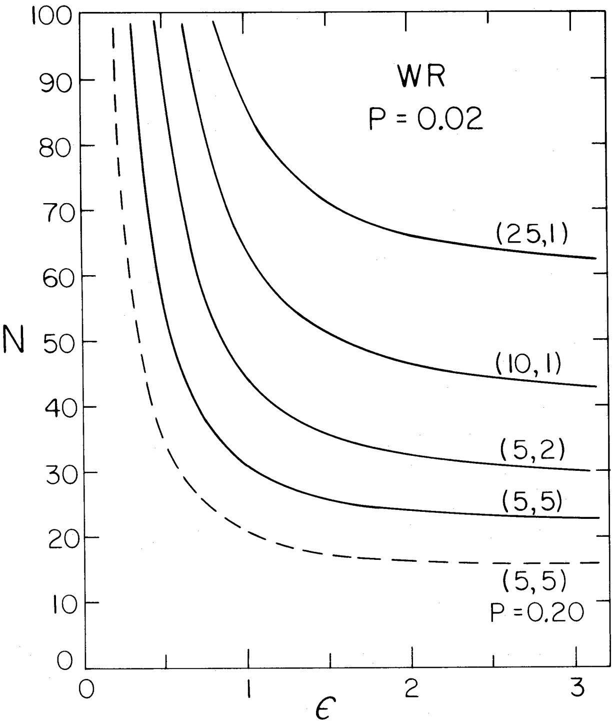

FIG. 4.—Statistical characteristics of the WR method. Notation as in Fig. 3.

solve for ϵ to obtain the WR version of equation (13):

$$\epsilon_{\mathrm{WR}} = \left[\frac{F' - 1}{N(1 - \delta)^2/N_b - F'}\right]^{1/2}, \qquad (14)$$

where $F' = (M - 1)F/(M - 3)$ and δ is given in connection with equation (12). In this case, F is $F_{P/2, M-1, N-1}$. Solutions of equation (14) for various bin structures are shown in Figure 4. Comparison with Figure 3 indicates that the WR approach is inferior for the small N case ($N < 50$). It is also apparent that the WR technique does better for cases with severe noise if $N > 50$. If $N = 100$, for example, signal-to-noise ratios as low as $\epsilon = 0.3$ should yield highly significant periods. For the Lick survey in LK, for example, amplitudes of 0.16 mag would be detectable with 100 points and of 0.26 mag with 50 points—numbers substantially lower than those actually found. If the signal-to-noise ratios is greater than unity, the PDM method is preferred, since larger N_b results in better resolution of the mean curve.

f) *Summary*

We summarize the various possible computational requirements and how they can be achieved within the PDM technique.

a) *Run time* can be minimized by using small N_f and N_c. Techniques that reduce the effective time base (Fitch 1967; Stobie and Hawarden 1972), thus widening the lines, are also very useful (see "segmental averaging" in Brault and White 1971).

b) *Resolution* of the mean curve increases with $M = N_b N_c$, but accuracy is lost if $N_b \gtrsim N/5$.

c) *Accuracy* of the period depends primarily on the total time base T, provided N and N_f are sufficiently large.

d) *Subharmonics* are suppressed for small N_b.

e) *Statistical significance* depends on N, ϵ, and bin structure, as shown in Figure 3.

f) *Aliased frequencies* can be minimized only through careful planning of the observing schedule.

A very good general scheme would be to use a (5, 2) bin structure for a "rough cut" scan of the data. Once the main frequencies have been identified, finer bin structure and frequency steps should be used to obtain an accurate period and light curve. If multiple periods are suspected, each oscillation should be removed from the data before one searches for weaker modes. Mode interactions can be treated in the same way. Improved primary periods can sometimes be obtained with interaction terms removed, and this procedure can be iterated to obtain optimum results.

In cases with very severe noise levels ($\epsilon < 1$), the WR algorithm may give acceptable results.

III. MULTIPLE PERIODS OF BK CENTAURI

To illustrate the use of these methods, we briefly present an analysis of the light variation of the double-mode Cepheid BK Centauri. For this purpose we will use only the 49 photoelectric measurements made by C. J. van Houten in 1965 for which Leotta-Janin (1967) has published the ΔV values. This author comments that "their number is too small to allow a reliable determination of the beat period" and so uses 25 years of plate estimates as well as knowledge of other Cepheid period ratios to obtain $\Pi_0 = 3.17389$ days and $\Pi_1 = 2.2366$ days. We have found that the photoelectric points alone unambiguously confirm these results and that they provide information on mode interaction as well.

The 49 measurements span 137 days. Line widths will therefore be about 0.007 in frequency (eq. [10]); so the frequency step size was taken to be 0.0015 for moderate resolution. Twelve nights have two observations, providing minimal alias discrimination. Although the measurements are quite accurate, we nonetheless estimate $\epsilon \approx 1.5$ due to the secondary oscillation. Figures 3 and 4 indicate that the PDM and the WR techniques should be about equally good for this case, provided $N_b < 10$. We choose a (5, 2) bin structure to maintain about 10 points per bin.

Figure 5*a* shows the Θ transform of the raw data. The primary period is clearly evident (minimum *e*), with minima *a* and *b* forming the standard subharmonic sequence found in Figure 1. Minimum *f* is narrow and showed a two-cycle mean curve; it is therefore the first subharmonic of a sizable minimum off the scale, at $f \approx 0.7$. This is the main alias of *e*, occurring at $1 - f_0$. Comparison of the subharmonics *b* and *f* indicates that *e* is the principal frequency. The

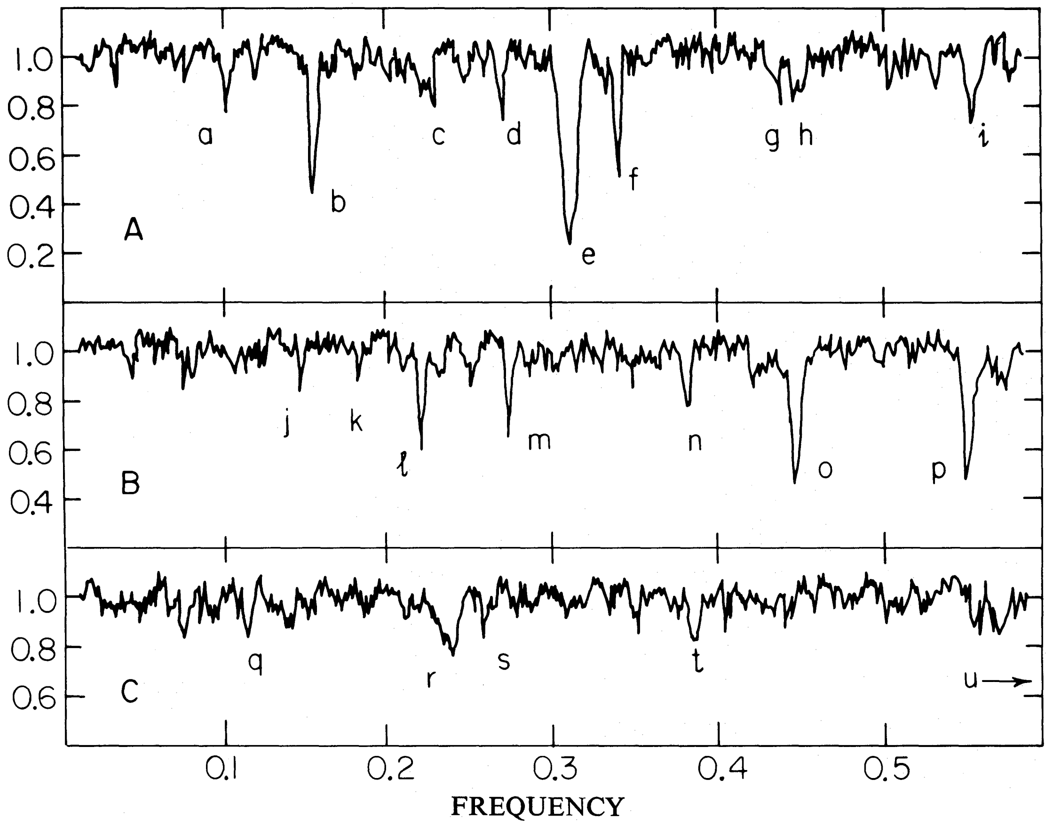

FIG. 5.—The Θ transforms of the BK Cen data, plotted versus frequency in cycles per day. (*a*) Transform of the raw data. (*b*) Transform of the data after removal of the fundamental oscillation. (*c*) Transform of the data after removal of the fundamental and first-overtone oscillations. Labeled minima are identified in Table 1.

TABLE 1

IDENTIFICATION OF FEATURES IN FIGURE 5

Feature	Identification	Frequency (d^{-1})	Significance
a......	$f_0/3$	0.105	0.13
b......	$f_0/2$	0.158	0.00
c......	$(1 - f_0)/3$	0.226	0.39
d......	$(1 - f_1)/2$	0.276	0.09
e......	f_0	0.316	0.00
f......	$(1 - f_0)/2$	0.344	0.00
g......	$(1 + f_0)/3$	0.439	0.24
h......	f_1	0.447	0.27
i......	$1 - f_1$	0.553	0.05
j......	$f_1/3$	0.148	0.50
k......	$(1 - f_1)/3$	0.183	0.55
l......	$f_1/2$	0.224	0.04
m......	$(1 - f_1)/2$	0.276	0.08
n......	$(f_0 + f_1)/2$	0.385	0.33
o......	f_1	0.447	0.00
p......	$1 - f_1$	0.553	0.00
q......	$[1 - (f_0 + f_1)]/2$	0.116	0.60
r......	$1 - (f_0 + f_1)$	0.235	0.27
s......	$(f_0 + f_1)/3$	0.253	0.55
t......	$(f_0 + f_1)/2$	0.383	0.43
u......	$f_0 + f_1$	0.764	0.11

identifications of the minima, together with their approximate frequencies and significances, are given in Table 1. Among the barely significant group, we identify *g* as the second subharmonic of the second alias of *e*, at $1 + f_0$, while *d*, *h*, and *i* are first-overtone features.

We see here a distinct *advantage* of the subharmonic response of this method: Any other major frequency in the range $0.6 < f < 1.2$ would show a subharmonic on Figure 5*a* comparable to *b* or *f*. Any minimum in the range $1.2 < f < 1.8$ would show a second subharmonic comparable to *a* or *g*. Since all features have been identified, no high-frequency components exist. This "look-ahead" feature can be extended by increasing N_b, at the expense of further complicating the spectrum.

Having identified the principal frequency (f_0), we then removed this component from the data (using linear interpolation between bin means). The transform of the reduced data is shown in Figure 5*b*. Here *j*, *l*, *o* are due to the first overtone, while *k*, *m*, and *p* represent the aliases. Note that each alias feature is slightly less significant than the corresponding "real" feature. A mode interaction feature (*n*) has also

TABLE 2

SUMMARY OF PRINCIPAL COMPONENTS

Mode	Frequency (d^{-1})	Period (d)	Amplitude (mag)	Adjusted Amplitude (mag)	Phase*
f_0...........	0.3153	3.172	0.544	0.598	0.40
f_1...........	0.4473	2.236	0.222	0.244	0.56
$f_0 + f_1$........	0.7682	1.302	0.151	0.166	0.16

* Phase is $(t_{\max} - t_1)/\Pi_0$, where t_1 = JD 2,438,813.48.

appeared. We can again conclude that no more significant modes exist to $f = 1.8$, since no subharmonics appear.

Figure 5*c* shows the transform of the doubly reduced data. No significant feature appears, but many marginal features are related to the mode interaction term $f_0 + f_1$. An extension of the transform to higher frequencies showed that this mode did indeed appear at the $P = 0.11$ level. This is the only remaining significant mode. This particular interaction term $(f_0 + f_1)$ has also been found to be excited in several other studies of similar objects (Fitch 1967; Henden 1976; Faulkner 1977*a*, *b*; Fitch and Szeidl 1976). In particular, this term invariably dominates the *beat* frequency $(f_1 - f_0)$ component. At present, the meaning of this is not clear.

The periods and amplitudes of the three components were determined using a finer frequency scan and are given in Table 2. When the bin size is allowed for, the actual amplitudes should be about 10% larger (i.e., eq. [12]); this is shown in the corrected amplitudes. The sum of these estimates is 1.1 mag, in agreement with the range of the photoelectric measures. (Obtaining the final amplitudes using smaller bin widths is impractical here because of the small data set.) The standard deviation of the residuals (triply reduced data) was $\sigma_{\text{fit}} = 0.048$ mag. Since the amplitudes are based on means of two bins, $\sigma_{\text{amp}} \approx \sigma_{\text{fit}}(2N_b/N)^{1/2} \approx 0.022$ mag.

The frequency labeled $f_0 + f_1$ in Table 2 is listed as derived and differs from the sum of the principal components. The discrepancy may not be real, since it amounts to about $1/T$—suggestive of a sidelobe problem.

IV. CONCLUSIONS

We have presented detailed analyses of two time-domain period determination techniques. The PDM method incorporates all the data directly into the test statistic and is thus well suited to small data sets. The WR method calculates an initial average that reduces its sensitivity, but it produces excellent noise suppression for sufficiently large N. Both methods are easy and efficient to use, work well on irregularly spaced data, and are ideally suited to highly nonsinusoidal time variations.

An analysis of the double-mode Cepheid BK Cen shows that a straightforward analysis of the three primary component variations is possible with 49 observed points (for a review of multiple-mode variables, see Stobie 1977). The availability of reliable statistics forms an important part of such an analysis.

This research was supported by National Science Foundation grant AST 77-26993 through Rutgers University, Department of Physics and Astronomy. This work is a consequence of discussions with A. Cox and S. Hodson during a visit to Los Alamos Scientific Laboratory, whose hospitality the author is happy to acknowledge.

REFERENCES

Abramowitz, M., and Segun, I. A. 1965, *Handbook of Mathematical Functions* (New York: Dover).
Brault, J. W., and White, O. R. 1971, *Astr. Ap.*, **13**, 169.
Chapman, S., and Bartels, J. 1940, *Geomagnetism*, Vol. **2** (London: Oxford University Press), p. 545.
Cox, A. N., Hodson, S. W., and King, D. S. 1978, *Ap. J.*, **220**, 996.
Faulkner, D. J. 1977*a*, *Ap. J.*, **216**, 49.
———. 1977*b*, *Ap. J.*, **218**, 209.
Fitch, W. S. 1967, *Ap. J.*, **148**, 481.
Fitch, W. S., and Szeidl, B. 1976, *Ap. J.*, **203**, 616.
Henden, A. A. 1976, *Proc. Los Alamos Solar and Stellar Pulsation Conf.*, ed. A. N. Cox and R. G. Deupree (LA-6544-C), p. 76.
Hesser, J. E., Ostriker, J. P., and Lawrence, G. M. 1969, *Ap. J.*, **155**, 919.
Lafler, J., and Kinman, T. D. 1965, *Ap. J. Suppl.*, **11**, 216 (LK).
Leotta-Janin, C. 1967, *Bull. Astr. Inst. Netherlands*, **19**, 169.
Lucy, L. B. 1976, *Ap. J.*, **206**, 499.
Ostriker, J. P., and Hesser, J. E. 1968, *Ap. J. (Letters)*, **153**, L151.
Percy, J. R. 1977, *M.N.R.A.S.*, **181**, 647.
Stellingwerf, R. F. 1975, *Ap. J.*, **199**, 705.
Stobie, R. S. 1972, *M.N.R.A.S.*, **157**, 167.
———. 1977, *M.N.R.A.S.*, **180**, 631.
Stobie, R. S., and Hawarden, T. 1972, *M.N.R.A.S.*, **157**, 157.
Wehlau, W., and Leung, K. 1964, *Ap. J.*, **139**, 843.
Whittaker, E. T., and Robinson, G. 1926, *The Calculus of Observations* (London: Blackie and Son) (WR).

Note added in proof.—A fully documented Fortran package implementing the procedures discussed in this paper is now available from the author upon request.

ROBERT F. STELLINGWERF: Department of Physics and Astronomy, Rutgers University, Frelinghuysen Road, Piscataway, NJ 08854