SEARCH AND DETERMINATION OF PERIODICITY

S. FERRAZ MELLO

Instituto Astronômico e Geofísico, Universidade de São Paulo, Sao Paulo, SP, Brasil

and

G. R. QUAST

Laboratório Nacional de Astrofísica, Itajubá, MG, Brasil

The Problem

Search for the existence of a periodicity, given a series of data, taking into account the irregular distribution of dates, the small variations of the measured quantity, and the small number of available observations.

As an example we consider the dM1.5e variable star DT Vir, the magnitude of which has been measured with respect to another star during a survey made at the ITA Astronomical Observatory in 1971—1972 (see TORRES and FERRAZ-MELLO, 1973; TORRES *et al.*, 1983). Table I gives 20 measurements of the relative magnitude, ΔB, of this star made during a three-month period in 1972.

TABLE I
Differential observations of the B magnitude of DT Vir in 1972. The phases were computed using the period 2.847D derived from Figure 68.

Date JD 2441400+	ΔB	f	Phase
05.682	0.493	−0.023	0
29.630	0.528	+0.012	0.41
31.664	0.496	−0.020	0.13
33.651	0.527	+0.011	0.82
36.671	0.533	+0.017	0.88
51.596	0.481	−0.035	0.13
52.583	0.550	+0.034	0.47
56.511	0.503	−0.013	0.85
69.518	0.522	+0.006	0.42
77.499	0.506	−0.010	0.23
89.477	0.519	+0.003	0.43
90.458	0.514	−0.002	0.78
91.503	0.495	−0.021	0.14
93.468	0.525	+0.009	0.83
94.464	0.506	−0.010	0.18
95.443	0.532	+0.016	0.53
96.461	0.517	+0.001	0.89
97.479	0.510	−0.006	0.24
98.442	0.524	+0.008	0.58
101.469	0.532	+0.016	0.64

J. Kleczek (ed.), Exercises in Astronomy, 231—235.

Procedure

(1) Look at the dates. The spacing between consecutive dates ranges from 1 to 24 days. Note also the decimal parts. For neighbouring dates they are very close, but in general the values are diminishing. This is due to the fact that observations are made preferably when the star is close to the meridian. The spacing between observations is thus roughly a multiple of one sidereal day. The double of this natural interval of the time series is called its Nyquist period.

(2) Look at the relative magnitudes ΔB. The standard error of these measurements is estimated to be around 0.01 mag. The amplitude of the variation is significant but the existence of periodicity is not evident.

(3) Choice of method. The unevenness in the spacing between consecutive dates and the small size of the series makes use of the techniques normally used for uniformly spaced time series (e.g. annual crops, rainfall, etc.) very hazardous. Therefore a date-compensated technique must be adopted. The technique called Date-Compensated Discrete Fourier Transform (DCDFT), corresponds to a curve-fitting approach using a sinusoid-plus-constant model, and is summarized below. For each trial frequency ω, one coefficient of spectral correlation S is obtained by the following formulae (FERRAZ-MELLO, 1981):

$$a_0^{-2} = N$$

$$a_1^{-2} = \sum \cos^2 x_i - a_0^2 \left(\sum \cos x_i\right)^2$$

$$a_2^{-2} = \sum \sin^2 x_i - a_0^2 \left(\sum \sin x_i\right)^2 - a_1^2 M^2$$

where $M = \Sigma \cos x_i \sin x_i - a_0^2 (\Sigma \sin x_i)(\Sigma \cos x_i)$ and

$$c_1 = a_1 \sum f_i \cos x_i$$

$$c_2 = a_2 \sum f_i \sin x_i - a_1 a_2 c_1 M.$$

$$S = \frac{c_1^2 + c_2^2}{\sum f_1^2}\ .$$

N is the number of observations in the series, t_i are the observation dates, $x_i = 2\pi\omega t_i$ and f_i are the measures ΔB referred to their mean value, i.e., shifted in a way such that $\Sigma f_i = 0$. All other symbols are intrinsic quantities. The summations are made for $i = 1$ to $i = N$.

(4) Use of DCDFT. Choose the range of frequencies you wish to investigate. Usually we consider the range of frequencies from 0 to the Nyquist frequency. Since the Nyquist period of the series under consideration is 2 sidereal days, i.e. 1.99454 D, the Nyquist frequency is 0.50137 D^{-1} or, roughly, 0.5 D^{-1}. Now choose the step to be used in the scanning of the frequency range; the step must be sufficiently small to avoid missing an important peak of the spectrum; for instance, in the exercise under consideration the choice of a step 0.02 D^{-1} may lead to computing S at the frequencies 0.34 and 0.36 and thus miss the most important feature of the power

spectrum, located at the frequency 0.351. On the other hand the calculation time is proportional to the number of steps; thus a step too small must also be avoided.

Once the frequency range and the step of scanning are chosen, compute the value of S for each trial frequency, and plot the function $S(\omega)$.

(5) How do you decide if the peaks in your graph are significant or not? How can you be sure that the greater ones are not merely due to chance? We suggest the use of a test derived by G. R. Quast from some testing techniques used in statistics and from the study of many computer-simulated time series. Compute the following quantities

$$G = -\frac{N-3}{2}\ln(1-S)$$

$$H = \frac{N-4}{N-3}(G + e^{-G} - 1)$$

$$a = \frac{2(N-3)\Delta t \cdot \Delta\omega}{3(N-4)}$$

$$C = (1 - e^{-H})^{a}$$

where ΔT is the time interval covered by the observations and $\Delta\omega$ is the range of frequencies sampled. C is the confidence of the result. $(1 - C)$ may be interpreted as the probability of having the height of the highest peak by chance only. Note that 5% is considered to be a large value for this chance; thus, in order to believe that the highest peak corresponds to a true periodicity of the astronomical phenomena under study, a confidence greater than 95% is recomended. Compute the confidence of the greatest peak in the graph and compare it to the result shown in Figure 68.

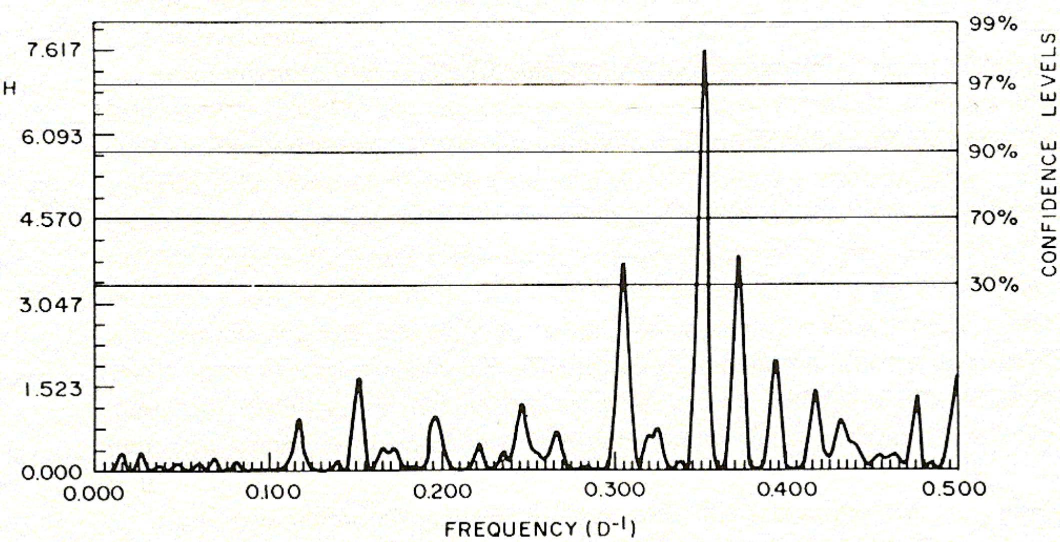

Fig. 68. Modified Periodogram or graph of the function $H(\omega)$. The confidence levels are also shown.

If you intend to write a complete program you may put together the two parts of this exercise to obtain $H(\omega)$ directly as shown in Figure 68.

You should keep in mind that the performance of statistical methods generally depends on the data distribution. Thus, if the observed data are strongly clustered (for instance because of many observations in one night), the computed values of C may indicate confidence levels higher than the real ones.

(6) From the value of ω for which S is a maximum, if its confidence is high enough, you have the period $T = 1/\omega$ of the variation of the starlight. Taking an arbitrary value, say zero, for the phase at the first date, calculate the phase at each date in Table I and plot ΔB as a function of the phase. The result is the light curve of Figure 69.

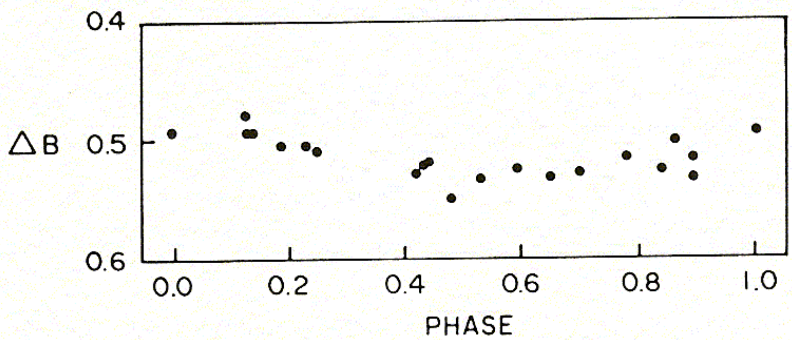

Fig. 69. Light curve of DT Vir (B magnitude vs. phase) obtained using the period derived from Figure 68.

(7) Consider 0.5 to 1.0 D^{-1} as a new interval of frequencies and proceed exactly as for the first considered interval. Compare the resulting graph with that obtained in the interval 0 to 0.5 D^{-1}, and discover why the Nyquist frequency is also called the folding frequency.

You may note that at the frequency 0.65 D^{-1} a new peak occurs, higher than the peak at $\omega = 0.35$ D^{-1}. In fact one peak is the alias of the other. The choice of one of the two frequencies is a generally difficult problem that may be solved taking into a account other factors. (In the case of DT Vir, the frequency 0.35 D^{-1} was chosen because the frequency 0.65 D^{-1} leads to a rotation period believed to be too short for this kind of star.) The reproduction of the spectrum in the range 0.5 to 1.0 D^{-1} is as much exact as the observations were made near the meridian. This phenomenon is called aliasing and you may get other folds going beyond 1 D^{-1}. The only way of avoiding an excessive aliasing is to take some precautions when the observations are made, for instance, by making further observations some hours after or before the meridian transit, and when possible, alternating these situations, or even making both.

(8) If you intend to gain experience with this technique you may consider the whole procedure applied to subsets of the data given in Table I. For instance you may omit some data at the end (or at the beginning) of the series. The results are

expected to be worse since you lessen the number and the time span of the observations. One further step may be to study simulated data (see FERRAZ-MELLO, 1981, Section IV).

References

BLOOMFIELD, P.: 1976, *Fourier Analysis of Time Series: An Introduction*, Wiley, New York.

FERRAZ-MELLO, S.: 1981, *Astron. J.* **86**, 619.

TORRES, C. A. O., BUSKO, I. C., and QUAST, G. R.: 1983, 'The ITA Survey on Red Dwarf Variable Stars. Final Report, in P. B. Byrne and M. Rodonò (eds.), *Activity in Red-Dwarf Stars*, D. Reidel Publ. Co. Dordrecht, Holland, p. 175.

TORRES, C. A. O. and FERRAZ-MELLO, S.: 1973, *Astron. Astrophys.* **27**, 231.