Mon. Not. R. Astron. Soc. **325,** 1124–1132 (2001)

The analysis of indexed astronomical time series – VII. Simultaneous use of times of maxima and minima to test for period changes in long-period variables

Chris Koen[1] and Fred Lombard[2]★

[1]*South African Astronomical Observatory, PO Box 9, Observatory, 7935 Cape, South Africa*

[2]*Dept of Mathematics and Statistics, Rand Afrikaans University, PO Box 524, Auckland Park 2006, South Africa*

Accepted 2001 March 16. Received 2001 March 7; in original form 2000 May 22

ABSTRACT

A statistical model is formulated that enables one to analyse jointly the times between maxima and minima in the light curves of monoperiodic pulsating stars. It is shown that the combination of both sets of data into one leads to analyses that are more sensitive. Illustrative applications to the American Association of Variable Star Observers data for a number of long-period variables demonstrate the usefulness of the approach.

Key words: methods: data analysis – methods: statistical – stars: variables: other.

1 INTRODUCTION

A great number of amateur astronomers around the world contribute observations of long-period pulsating stars (LPVs) to data bases maintained by the American Association of Variable Star Observers (AAVSO). These data are occasionally published in condensed form, giving amongst other things the times of maximum and minimum light (Campbell 1926, 1955; Mattei, Mayall & Waagen 1990). The present authors, like many before them and many after them, have used either the times of maxima or the times of minima to test for changes in the mean pulsation periods of these stars (e.g. Koen & Lombard 1993; Lombard 1998a). The present paper is concerned with a combination of the two sets of information in order that more sensitive tests can be performed.

Intuition leads us to believe that the more data we utilize, the more sensitive are the statistical tests based on them bound to be. On the other hand, since times between successive maxima and minima overlap, the two sets of information are obviously not independent. This dependence needs to be fully accounted for in any statistical analysis. The next section is concerned with the formulation of a statistical model for the intervals between timings. Section 3 deals with a proposed period change test statistic while Section 4 presents a number of example analyses. Conclusions are given in Section 5.

2 A STATISTICAL MODEL FOR THE DATA

It is assumed initially, and entirely as a matter of convenience in the exposition, that the first timing is that of a pulsation maximum. The observed times of successive maxima and minima are denoted by T_0, t_0, T_1, t_1, …. The observed times between light maxima and light minima are then $M_j = T_j - T_{j-1}$ and $m_j = t_j - t_{j-1}$, respectively. Instead of analysing the sequences M_j and m_j separately we propose to analyse the interleaved sequence

$$Y \equiv y_1, y_2, y_3, y_4, \ldots = M_1, m_1, M_2, m_2, \ldots$$

The components of the latter sequence can be represented as

$$\begin{aligned} y_{2j-1} &= M_j = P + \alpha_j + \delta_j - \delta_{j-1}, \\ y_{2j} &= m_j = P + \beta_j + \epsilon_j - \epsilon_{j-1}, \end{aligned} \tag{1}$$

where P denotes the mean period of the star (assumed constant for the moment), δ_j and ϵ_j denote the measurement errors in the jth times of the light maximum and minimum, respectively, while α_j and β_j are the random intrinsic period variations in the intervals (T_{j-1}, T_j] and (t_{j-1}, t_j], respectively. The latter model was first proposed by Sterne (1934) and was studied further by Lombard & Koen (1993), Koen & Lombard (1995) and by Lombard (1998a). The measurement errors are statistically independent of one another, as are the terms α_1, α_2, … and the terms β_1, β_2, …. However, because of the overlap [t_{j-1}, T_j] (the jth light curve rise time) between the intervals [T_{j-1}, T_j] and [t_{j-1}, t_j] and the overlap [T_j, t_j] (the jth light curve fall time) between the intervals [T_j, T_{j+1}] and [t_{j-1}, t_j] the terms α_j and β_j are correlated, as are the terms β_j and α_{j+1}. To make this more explicit, let d_j ('descending') and a_j ('ascending') denote, the contributions of the jth fall and rise times to the value of α_j respectively. Then

$$\alpha_j = d_j + a_j, \quad \beta_j = a_j + d_{j+1} \tag{2}$$

and it follows from the latter relations that $\mathrm{cov}(\alpha_j, \beta_j) = \sigma_a^2$ and $\mathrm{cov}(\beta_j, \alpha_{j+1}) = \sigma_d^2$. Thus, unless both σ_δ and σ_ϵ are zero, i.e. unless the Mira exhibits no intrinsic period scatter, at least one of the pairs (α_j, β_j) or (β_j, α_{j+1}) will be correlated.

★E-mail: fred@na.rau.ac.za

The implication from equations (1) and (2) and the preceding discussion is that the Y sequence has a correlation memory of at most two lags. This property agrees with the results of Isles & Saw (1987, 1989a,b,c) who, in a sample of 24 LPVs, found significant correlation between non-contiguous rise and fall times in only one instance. The results of exploratory studies of many data sets from Mattei et al. (1990) by the present authors are also in accord with the results of Isles and Saw. We note in passing that the results to be derived below are not overly sensitive to the particular correlation model chosen to represent the data. For instance, a general moving average model of order two may be used instead of a model with the specific structure postulated above.

If the first observation is a time of light minimum, then the data are

$$Y' \equiv y_2, y_3, y_4, y_5, \ldots = m_1, M_2, m_2, M_3, \ldots$$

It is convenient, for purposes of the following analysis, to regard Y and Y' as realizations of a single time series $Z \equiv z_1, z_2, \ldots$ defined by

$$Z = \xi Y + (1 - \xi) Y'$$

where $\xi = 0$ or 1 with a probability of 1/2 each. Then straightforward calculations based on equations (1) and (2) give

$$\gamma_Z(l) \equiv \mathrm{cov}(z_j, z_{j+l}) = \begin{cases} \sigma_a^2 + \sigma_d^2 + \sigma_\epsilon^2 + \sigma_\delta^2 & l = 0 \\ (\sigma_a^2 + \sigma_d^2)/2 & l = \pm 1 \\ -(\sigma_\epsilon^2 + \sigma_\delta^2)/2 & l = \pm 2 \end{cases} \tag{3}$$

and $E(z_j) = P$ for $j = 1, 2, \ldots$. It is thus seen that Z defines a stationary time series with a mean P, the period of an idealized, deterministic, (no measurement error or intrinsic period scatter) light curve.

By introducing the series Z we randomize the observational 'phase' of the data between 'the first observation is a maximum' and 'the first observation is a minimum'. This is akin to the phase randomization which is often introduced when analysing a function statistically, say a sinusoid $\cos(\theta + \omega t)$, for which the time origin is unspecified. If θ is regarded as a constant the sinusoid defines a non-stationary time series. If θ is regarded as a uniformly distributed quantity in the interval $[0, 2\pi)$, however, then the series becomes stationary, phase effects are annihilated and the time origin is irrelevant; see Priestley (1981). In our particular use of this approach the 'phase' annihilation is justified, by the requirement that the results of the analysis should not depend on whether the first observation is a maximum or a minimum.

3 A TEST FOR CONSTANCY OF THE MEAN PERIOD

Thus far we have assumed that the period of the LPV is constant, except for small random (i.e. non-systematic) perturbations from cycle to cycle. Thus, if X_j denotes the observed interleaved sequence of times between maxima and minima, then $X_j = Z_j$ and $E(X_j) = P$. Our purpose is to establish a method that is capable of detecting systematic deviations from such a stationary pattern. There are various possible types of systematic deviation that might be considered. Suppose, for instance, the idealized light curve (i.e. measured without error and containing no intrinsic scatter) has a period P_j. The case $P_j \equiv P$ corresponds to a constant period. In general, we wish to uncover situations in which P_j varies with j. For instance, if $P_j = P + \rho j$ the pulsation period is lengthening ($\rho > 0$), or shortening ($\rho < 0$), in a linear fashion. On the other hand, the pulsation period may be constant ($P_j = P$) up to cycle $j = j_1$, increase or decrease $[P_j = P + \beta(j - j_1)]$ between cycle $j = j_1$ and $j = j_2$, and then remain constant $[P_j = P + \beta(j_2 - j_1)]$ subsequent to cycle $j = j_2$. In the preceding two examples the non-constancy of the period was of a deterministic nature. A stochastic form of non-constancy can also be postulated. Suppose $P_j = P + U_j$ where U_j is a random walk with zero mean and variance $\sigma^2 \times j$, i.e. $U_j = U_{j-1} + u_j$ where u_j is a white noise with variance σ^2, then in such a case the condition $\sigma^2 = 0$ corresponds to the period being constant.

In reality, of course, the idealized light curve is not observable and, hence, neither is P_j. What is observed is a contaminated (signal plus noise) version of the latter, namely

$$X_j = (P_j - P) + Z_j, \quad j = 1, \ldots, N \tag{4}$$

where Z_j is the noise process defined above in Section 2. The test to be described below is designed to detect situations such as those described above in which P_j varies with j, i.e. $P_j - P$ is not identically zero, regardless of whether P_j is deterministic or stochastic, smooth or non-smooth. The sole condition imposed upon the 'signal' P_j is that its power resides primarily at low frequencies and that the power should be relatively negligible compared to the noise power at other frequencies. This is true, for example, in all three instances just discussed.

Lombard (1998a,b) proposed a class of period-change test statistics capable of detecting low-frequency signals. We shall use the particular form

$$T_N = \sum_{k=1}^{K-1} \frac{1}{k} \left[\frac{I_X(\omega_k)}{\hat{S}(\omega_k)} - 1 \right] \tag{5}$$

where K is the integer part of $N/2$,

$$I_X(\omega_k) = \frac{1}{N} \left| \sum_{j=1}^{N} X_j \exp(-\mathrm{i} j \omega_k) \right|^2$$

is the periodogram of the observed data X_j and $\hat{S}_Z(\omega_k)$ is an estimate of the theoretical spectrum $S_Z(\omega_k)$ of the noise process Z_j. Here and elsewhere, $\omega_k = 2\pi k/N$. Note that the test applies to *any* data that have the structure given by equation (4). When X_j represents the interleaved sequence of times between the maxima and minima we find from equation (3) that

$$S_Z(\omega_k) = \sum_{|l|<\infty} \gamma_Z(l) \exp(\mathrm{i}\omega_k l)$$

$$= B_1[1 + \cos(\omega_k)] + B_2[1 - \cos(2\omega_k)] \tag{6}$$

where

$$B_1 = \sigma_a^2 + \sigma_d^2, \quad B_2 = \sigma_\epsilon^2 + \sigma_\delta^2.$$

On the other hand, if $X_j = M_j$ represents the sequence of times between maxima (considered in isolation) then equation (1) gives, after some calculation,

$$S_Z(\omega_k) = A_1 - 2A_2[1 - \cos(\omega_k)] \tag{7}$$

where

$$A_1 = \sigma_a^2 + \sigma_d^2, \quad A_2 = \sigma_\delta^2.$$

When $X_j = m_j$ the same result holds when replacing σ_δ by σ_ϵ.

The rationale behind the test statistic is two-fold. First, if there is no period change, then since $X_j = Z_j$, the $I_X(\omega_k)$ will statistically fluctuate around the theoretical value $S_Z(\omega_k)$.

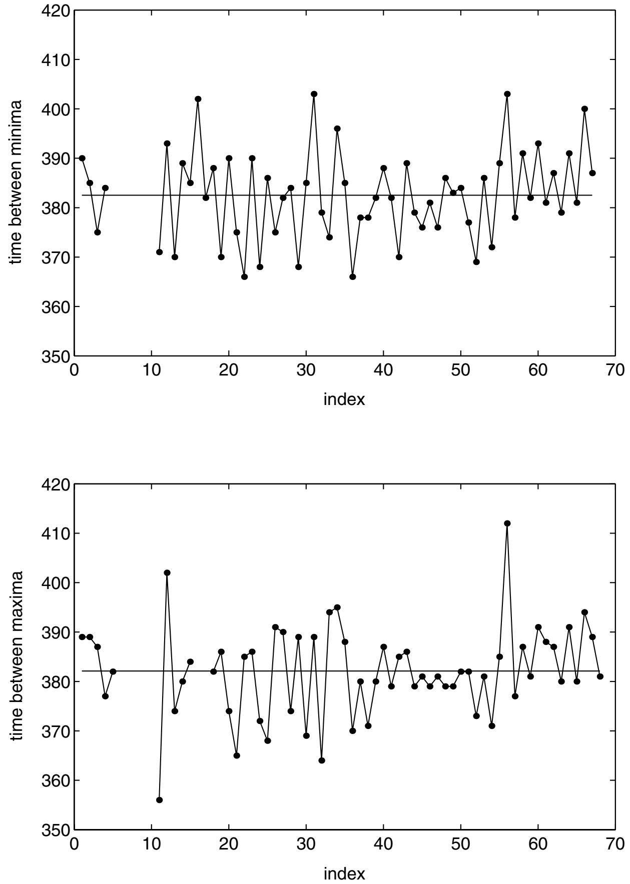

Figure 1. Time series plots of the times between minima (upper panel) and between maxima (lower panel) of the light curve of Y Aqr.

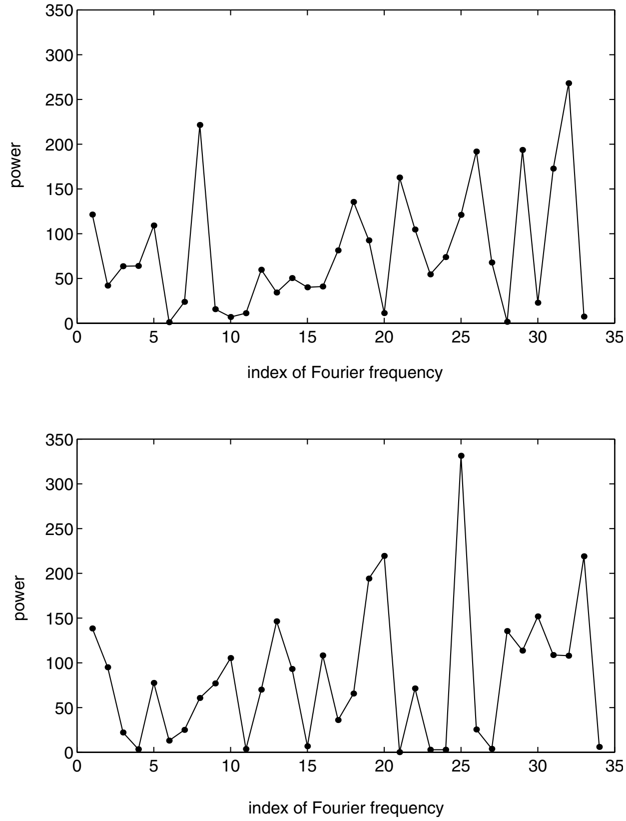

Figure 2. Periodograms of the times between minima (upper panel) and between maxima (lower panel) of the light curve of Y Aqr.

Secondly, if low-frequency trends in the period were present, systematically larger $I_X(\omega_k)$ values would occur at 'small' values of k because of power excesses at low frequencies. Subtraction of the constant one inside the bracket of equation (5) stabilizes the distribution of T_N as $N \rightarrow \infty$. It may be shown, see Lombard (1998b), that for large values of N,

$$\Pr\{T_N > c\} \approx 1 - \exp[-\exp(-c - \gamma)]$$

where $\gamma = 0.5772157\ldots$ (constant of Euler). The parameters B_1 and B_2 can be estimated by least squares from the approximate relation

$$I_X(\omega_k) \approx S_Z(\omega_k) = B_1[1 + \cos(\omega_k)] + B_2[1 - \cos(2\omega_k)],$$
$$k = L+1, \ldots, K-1 \tag{8}$$

which, in accordance with the discussion above, is valid if a small number $L \geq 0$ of the periodogram values at the lowest frequencies are excluded; see also Rice (1979) and Lombard (1998a).

Lombard (1998a) remarked on the slow convergence of a test statistic such as that in equation (5) to its asymptotic distribution, especially in the tails. For small-sample sizes realistic p values can be determined by simulating data sets which have $\hat{S}_Z(\omega)$ as their theoretical spectrum.

4 EXAMPLES

4.1 Y Aquarii

Time series plots of the two sequences M_j and m_j for this LPV are shown in Fig. 1. The immediate visual impression is that the period exhibits a moderate upward trend. The corresponding periodograms, shown in Fig. 2, are both consonant with the form given by equation (7) with perhaps more than just a hint of excess power at the lowest frequencies. Considered in isolation, however, neither sequence provides significant evidence of a change in the mean period – the test statistic given by equation (5), based upon the theoretical spectrum of equation (7), yields significance levels of 0.14 and 0.12.

Table 1. Relevant statistics for the minima and maxima of Y Aquarii.

	Minima	Maxima
L	0	0
$\hat{A}_1$	51.24	56.11
$\hat{A}_2$	14.59	13.35
T_N	1.346	1.255
p value	0.14	0.14

The relevant statistics for the minima and maxima are given in Table 1. Fig. 3 shows a time series plot of the sequence of 134 interleaved times between the maxima and minima, respectively together with the corresponding periodogram. The periodogram is consonant with the form given by equation (6) except for the obvious excess of power at the lowest frequencies. Application of the T_N test to these data yields a p value of 0.021, confirming what the latter two figures suggested. The p value was found from 10^4 Monte Carlo trials using the least-squares estimates of $\hat{B}_1 = 43.64$

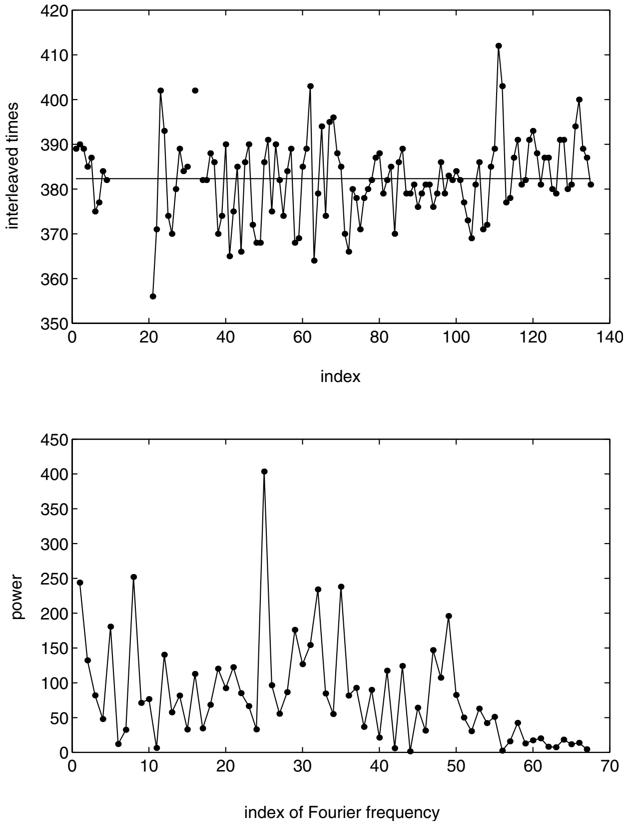

Figure 3. Time series plot of the interleaved sequence of times between the maxima and minima of the light curve of Y Aqr (upper panel), and their periodogram (lower panel).

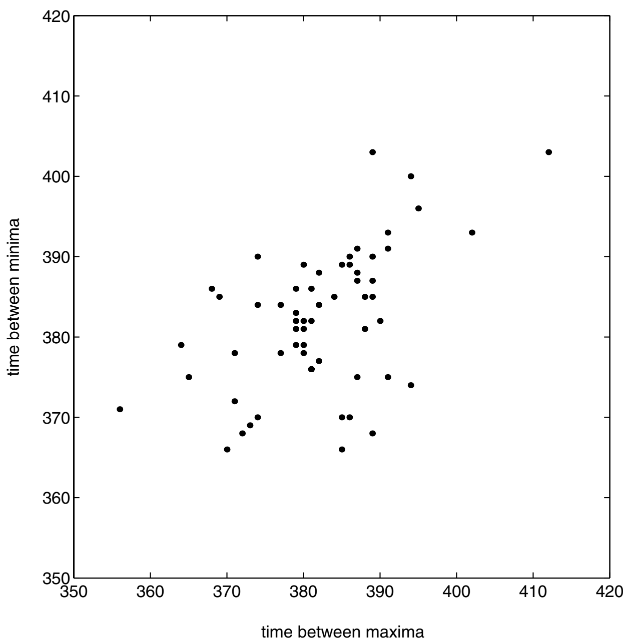

Figure 4. Scatter plot of the times between maxima against corresponding times between minima of the light curve of Y Aqr.

and $\hat{B}_2 = 35.68$. In obtaining the latter estimates we used the conservative choice of $L = 0$ in equation (8).

In this particular instance, therefore, use of the interleaved data set led to a conclusion substantially different from that ensuing

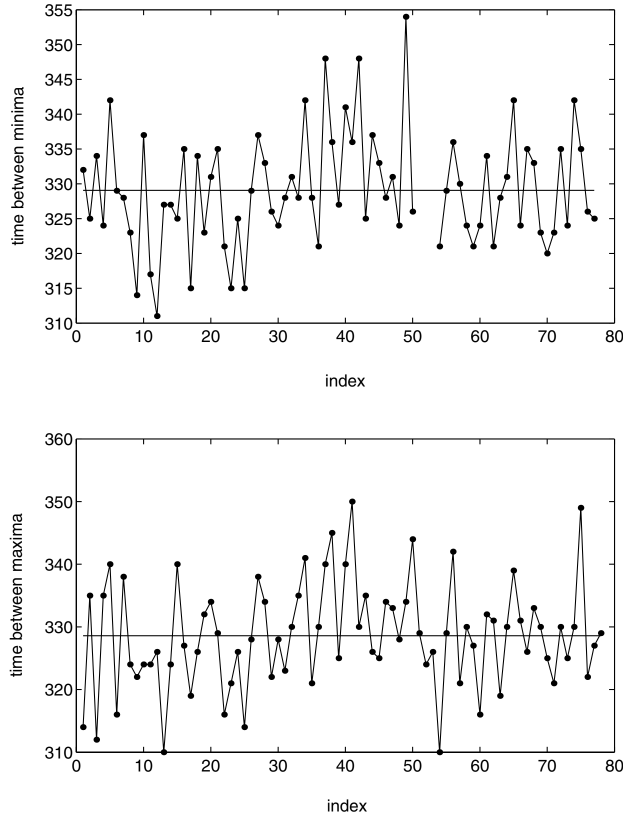

Figure 5. Time series plots of the times between minima (upper panel) and between maxima (lower panel) of the light curve of R CVn.

from a consideration of the times between maxima or between minima alone. The reason for this becomes even more transparent upon inspecting Fig. 4 which plots the pairs (M_j, m_j), $j = 1, \ldots, 134$. As was to be expected the pairs are not perfectly correlated – the Pearson correlation coefficient equals 0.51. This means that the information content about the period in the sequence M_1, m_1, M_2, m_2, … is *greater* than that in either of the subsequences M_1, M_2, … or m_1, m_2, … considered in isolation. Of course, if the sequences M_1, M_2, … and m_1, m_2, … were perfectly correlated then they would have the same information content and no advantage would arise by considering them jointly.

Inspection of Fig. 1 reveals that the Y Aqr data are incomplete, especially near the beginning of the series. The T_N test has no difficulty in dealing with such data. For relatively small fractions of missing values (of the order of 10 per cent for the Y Aqr data), the theory above applies to a good approximation. Partial verification of this statement can be obtained in the present instance by observing what happens when the analysis is restricted to the series that begins at the eleventh maximum. The interleaved sequence then consists of $N = 114$ values, and the new observed value $T_N = 2.42$ has a p value of 0.031 associated with it. The conclusion is virtually identical to that reached via analysis of the full data set. Lombard (1998a) and Sterken, Broens & Koen (1999) discuss the use of the frequency domain period change statistic in cases where there are many missing values.

4.2 R Canum Venaticorum

Time series plots of the two sequences are shown in Fig. 5, and the corresponding periodograms are shown in Fig. 6. The relevant statistics are given in Table 2.

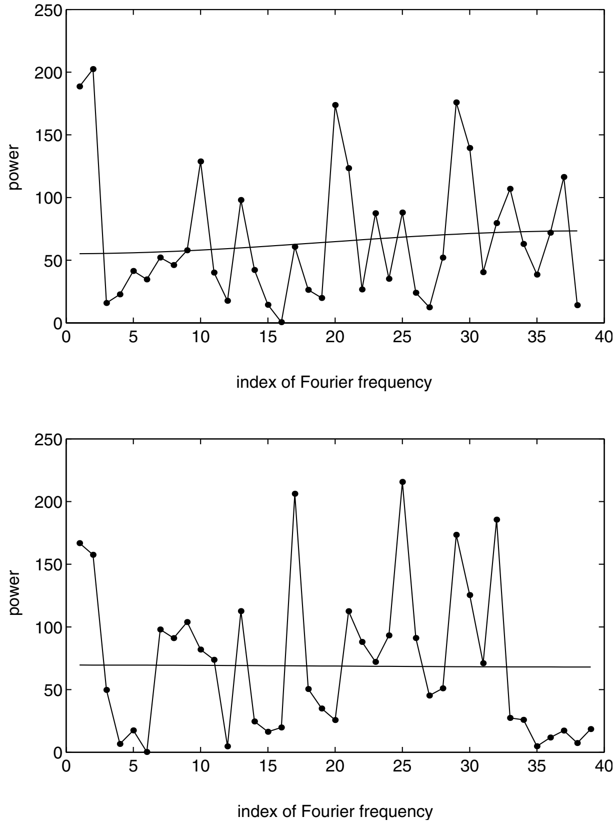

Figure 6. Periodograms of the times between minima (upper panel) and between maxima (lower panel) of the light curve of R CVn.

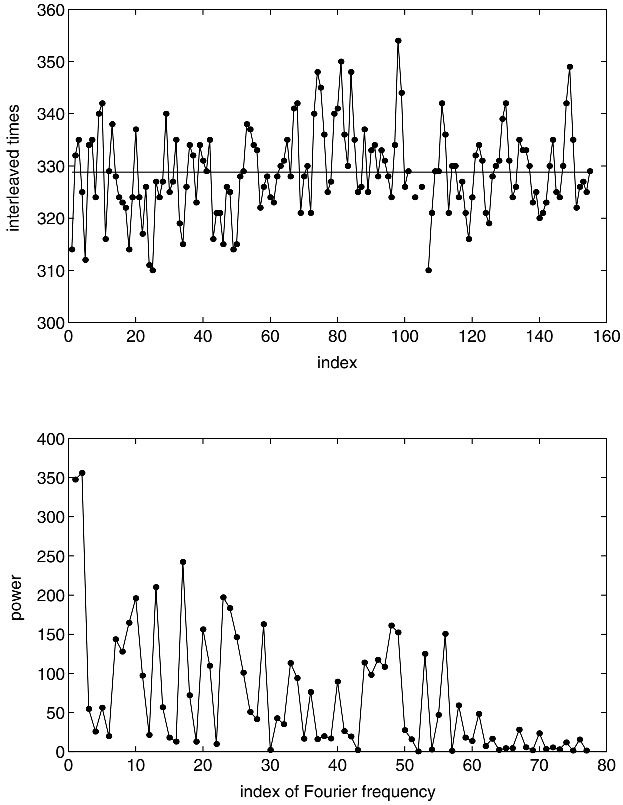

Figure 7. Time series plot of the interleaved sequence of times between the maxima and minima of the light curve of R CVn (upper panel), and their periodogram (lower panel).

Table 2. Relevant statistics for the minima and maxima of R Canum Venaticorum.

	Minima	Maxima
L	2	2
$\hat{A}_1$	41.11	61.35
$\hat{A}_2$	9.22	2.33
T_N	5.33	2.197
p value	.003	0.060

Table 3. Relevant statistics for the minima and maxima of R Bootis.

	Minima	Maxima
L	0	0
$\hat{A}_1$	0	16.00
$\hat{A}_2$	18.13	9.64
T_N	279.81	−0.78
p value	0.000	0.705

Fig. 7 shows a time series plot of the sequence of 134 interleaved times between the maxima and minima, respectively (the sequence starts and ends with a minimum), together with the corresponding periodogram. The separate analyses of the minima and maxima give widely differing results: the p value for change from the minima is an order of magnitude smaller than that for a change from the maxima. Application of the T_N test to these data yields a p value of 0.008. The latter was found from 2×10^4 Monte Carlo trials using the least-squares estimates of $\hat{B}_1 = 45.54$ and $\hat{B}_2 = 15.48$. In obtaining the latter estimates we made a conservative choice of $L = 2$ in equation (8). Thus, the T_N test resolves the difference between the p values from the separate analyses by making use of all the available information. The result indicates strongly that the period of R CVn has indeed changed.

4.3 R Bootis

Time series plots of the two sequences and their respective periodograms are shown in Figs 8 and 9. The relevant statistics for the minima and maxima are given in Table 3.

There is a major difference between the p values obtained from the two sequences. Observe, however, that the estimate $\hat{A}_1$ of intrinsic scatter variance from the minima is zero while that from the maxima is positive. Thus the estimated theoretical spectrum of the minima is exceedingly close to zero at, e.g. the lowest three frequencies, and division by such small numbers is a likely cause of the extraordinarily large T_N value. There is no evidence in Fig. 8 of a trend that would result in a T_N value of such magnitude. It is also clear on a closer inspection of the periodogram of the minima that the model given by equation (7) does not fit those particular data particularly well. There is nothing in the periodogram to suggest that the theoretical spectrum decreases towards the lower frequencies, as would be the case if equation (7) was indeed the theoretical spectrum. In fact, the periodogram seems to hover more or less around a constant value in the range of frequencies ω_k with indices $1 \leq k \leq 35$. The fitting of a (possibly) inappropriate model to the periodogram of the minima leads to the zero estimate of

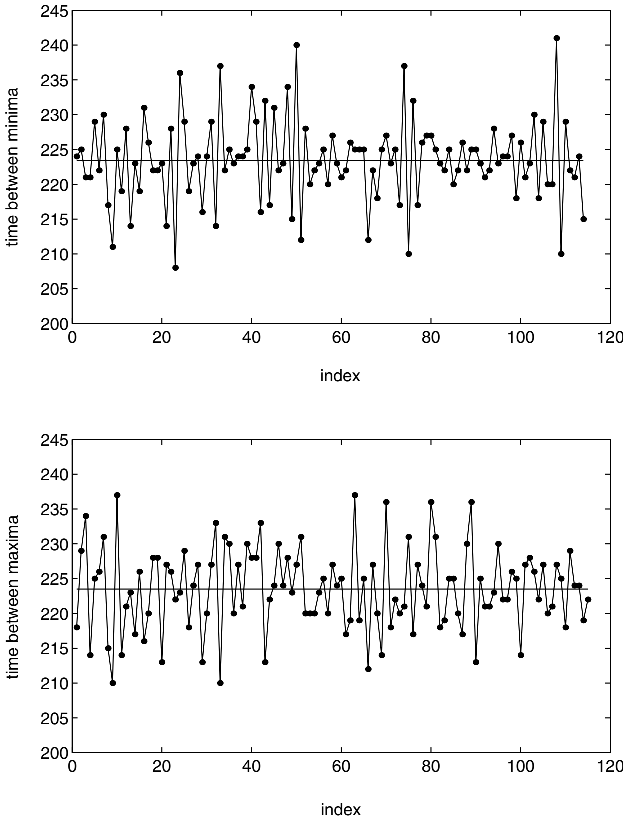

Figure 8. Time series plots of the times between minima (upper panel) and between maxima (lower panel) of the light curve of R Boo.

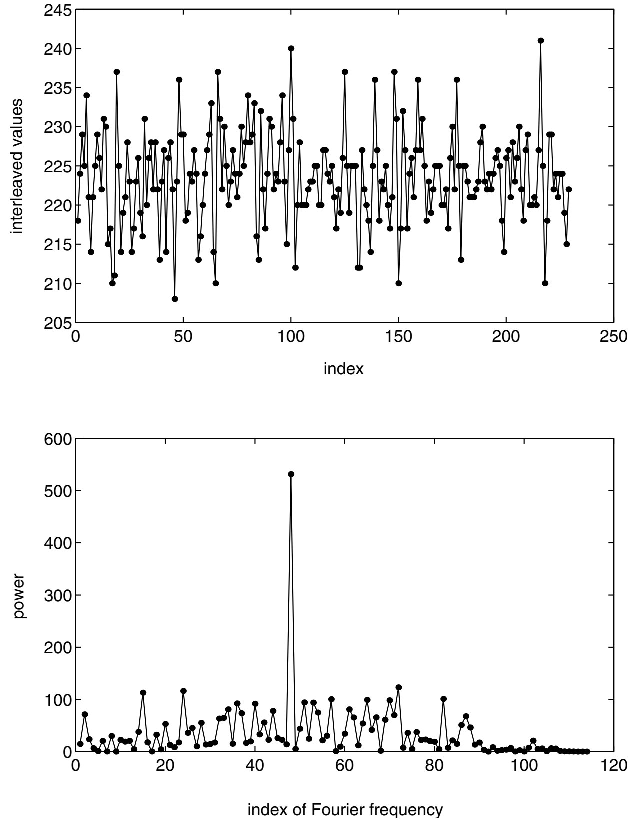

Figure 10. Time series plot of the interleaved sequence of times between the maxima and minima of the light curve of R Boo (upper panel), and their periodogram (lower panel).

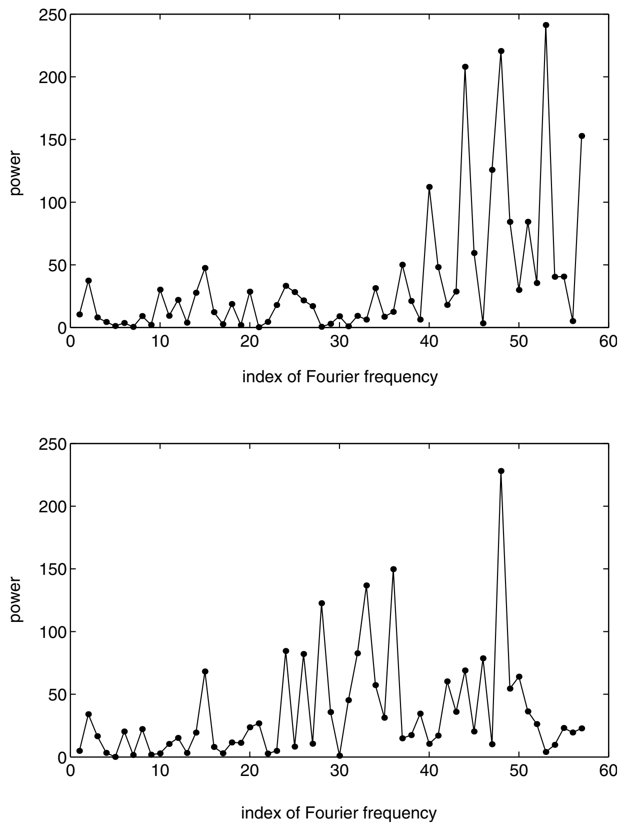

Figure 9. Periodograms of the times between minima (upper panel) and between maxima (lower panel) of the light curve of R Boo.

intrinsic scatter. The periodogram of the maxima, on the other hand, seems to be quite consonant with equation (7).

Fig. 10 shows plots of the interleaved sequence of $N = 229$ time intervals and of the associated periodogram. The following results are obtained for $L = 0$:

$$\hat{B}_1 = 10.18, \quad \hat{B}_2 = 20.94 \text{ and } T_N = 0.526,$$

the latter with an associated p value of 0.282. The conclusion is, therefore, that the period of R Bootis is a constant.

Inspection of Fig. 10 reveals that, with the exception of a powerful intermediate frequency component at the 48th Fourier frequency, the periodogram of the interleaved sequence is consonant with the form expected from equation (6). The deviation of the periodogram of the minima from its expected form (discussed above) therefore does not have a deleterious effect on the frequency domain characteristics of the interleaved sequence. The exceptional component at ω_{48} in the spectrum of the interleaved values also does not have a material effect on the result of the T_N test. The latter assigns a weight of 1/48 to this component, compared to the weights 1, 1/2, … assigned to the low-frequency components. Indeed, if the frequency ω_{48} is excluded from the estimation in equation (8) a p value of 0.322 (compared to 0.282 when it is not excluded) results.

Fig. 11 shows scatter plots of the pairs (M_j, m_j) and (m_j, M_{j+1}), respectively. The Pearson correlation coefficients are 0.053 and 0.281, respectively. This suggests that $\sigma_a = 0$ and that $\sigma_d > 0$; see the remarks just after equation (2). In other words, there is evidence that the intrinsic scatter in the period is largely confined to the descending branch of the light curve (the fall time). The rise times

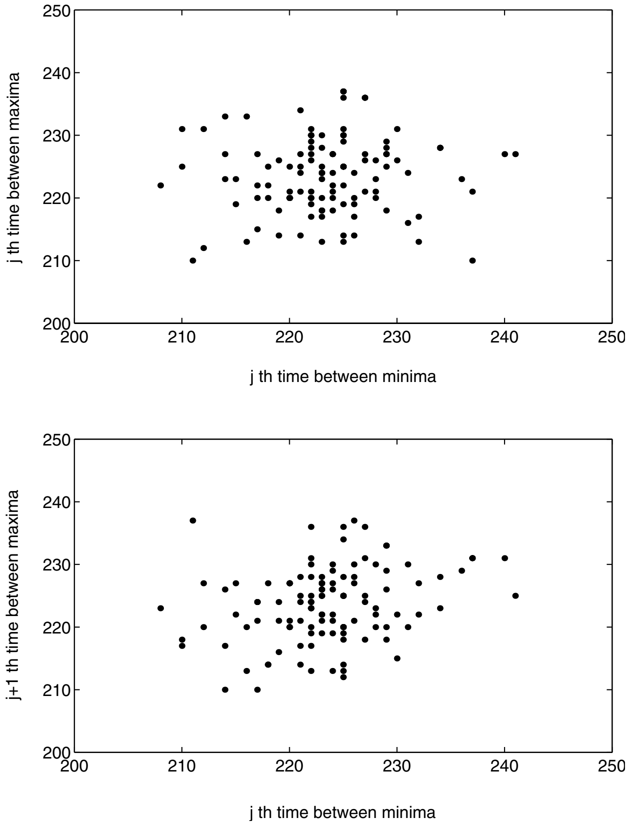

Figure 11. Scatter plot of the time between maxima against the time between following minima (upper panel), and of time between minima against the time between following maxima (lower panel) of the light curve of R Boo.

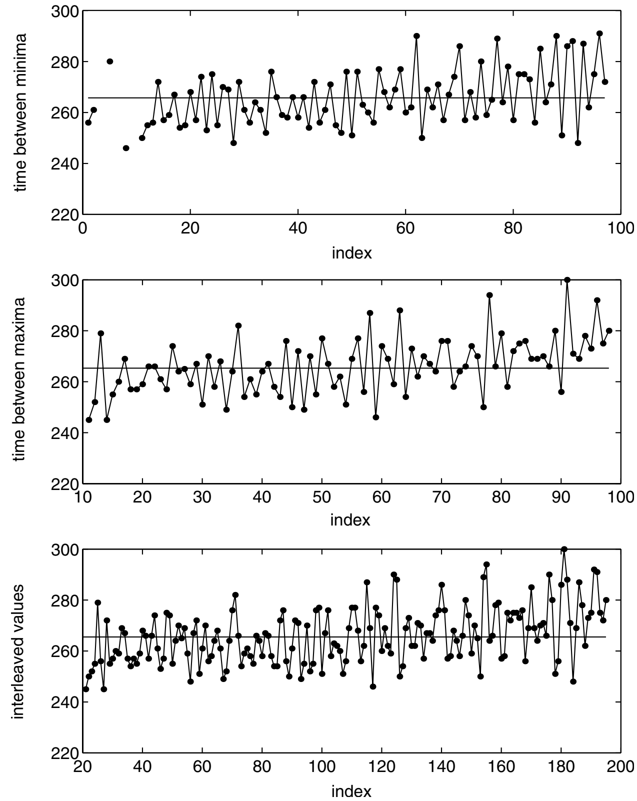

Figure 12. Time series plots of the times between minima (upper panel) and between maxima (centre panel) of the light curve of W Dra, and of the interleaved sequence of times (lower panel).

exhibit negligible intrinsic scatter. This is a surprising finding and we are not aware of whether such a phenomenon has been noticed before.

The value of the (constant) period of R Bootis can be estimated by the average of the interleaved sequence of 229 intervals. This is $\hat{P} = 223.47$ d. To find the standard error of the estimate it is necessary to take account of the correlation between the individual estimates. This is accomplished by allowing $\hat{S}_Z(0)$ to fulfil the role of the usual sample variance; see Priestley (1981). Thus, the estimated standard error of $\hat{P}$ is

$$\sqrt{\frac{\hat{S}_Z(0)}{229}} = \sqrt{\frac{2 \times 10.18}{229}} = 0.31\,\mathrm{d}.$$

4.4 W Draconis

It is clear from the time series plots in Fig. 12 that the period of this LPV is increasing at what seems to be a linear rate. The result of applying the T_N test ($N = 175$) confirms the change of period. The formal p value is 0.0007. A question that arises, however, is whether there is perhaps an additional long-term trend superposed upon the linear increase. A straight line fit of $P_j = P + \beta \times j + Z_j$ to the interleaved data gives the slope estimate $\hat{\beta} = 0.096$ with a standard error of 0.0086. Fig. 13 shows a time series plot of the residuals Z_j remaining after fitting the straight line, together with the corresponding periodogram. The T_N test applied to these residuals using $L = 4$ gives a p value of 0.102. It seems, therefore, that there are no significant trends in the period of W Draconis in addition to the linear one.

In calculating the standard error of the estimate $\hat{\beta}$, the quantity $\hat{S}_Z(0)$ takes the place of the usual residual variance $\frac{1}{N-4}\sum(X_j - \hat{P}_j)^2$, the latter being appropriate only when the Z_j are uncorrelated. In the present instance the estimated spectrum of the residual noise process Z_j in equation (8) is

$$\hat{S}_Z(\omega) = 16.467[1 + \cos(\omega_k)] + 87.107[1 - \cos(2\omega_k)],$$

so that $\hat{S}_Z(0) = 32.934$. The latter value is therefore the 'effective' residual variance which takes account of the correlation in the data.An estimate of the rate of period increase can be made. The units of $\hat{\beta}$ are d per index. There are 175 index values that index a total of 46 553 d, i.e. an average of 266.017 d per index. Hence the estimated rate of period increase is $0.192/266.017 = 7.2 \times 10^{-4}\,\mathrm{d}^{-1}$ with a standard error of $0.0086/266.017 = 0.3 \times 10^{-4}\,\mathrm{d}^{-1}$.

4.5 R Aquilae

It is clear from the time series plot of the $N = 172$ interleaved values in Fig. 14 that the period of this LPV is decreasing, again at what seems to be a linear rate. The question whether the linear decrease fully describes the extent of the period change must again be asked. The top panel in Fig. 15 shows a time series plot of the residuals remaining after fitting a straight line

$$P_j = P + \beta \times j + Z_j^{(1)} \tag{9}$$

to the sequence of interleaved times between maxima and between minima. There is visual evidence of low-frequency sinusoidal variation, an impression that is strengthened upon inspection of the bottom panel in Fig. 15, which shows the corresponding

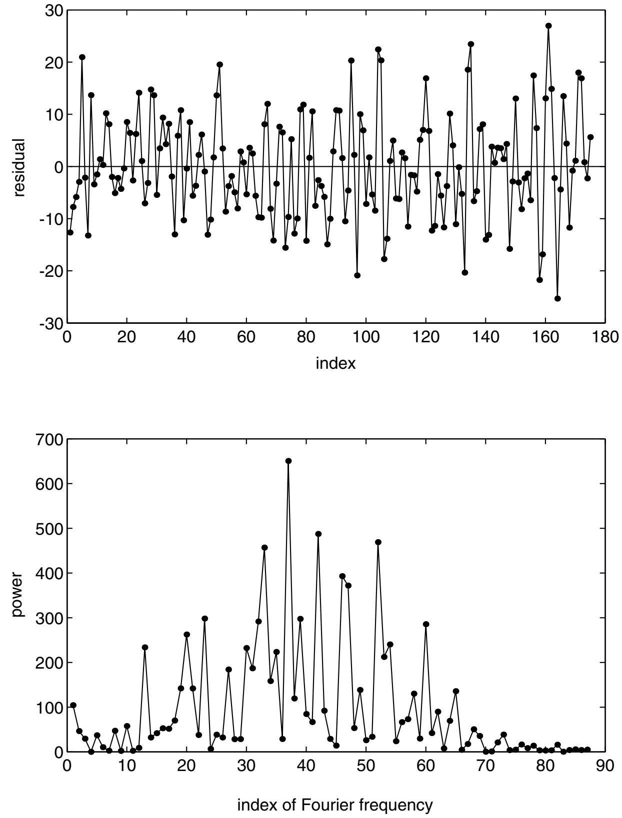

Figure 13. Time series plot of the residuals from a straight line fit to the interleaved sequence of times between the maxima and minima of the light curve of W Dra (upper panel), and their corresponding periodogram (lower panel).

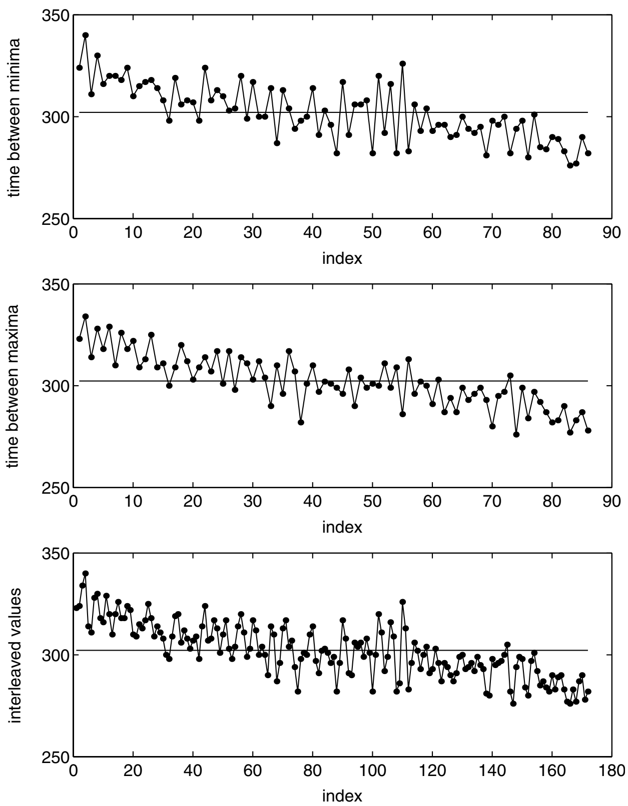

Figure 14. Time series plot of the interleaved sequence of times between the maxima and minima of the light curve of R Aql.

periodogram. Observe that the form of the latter is entirely consonant with the postulated form of equation (6) except, of course, at the very lowest frequencies. The T_N test based upon these residuals, using the very conservative value of $L = 0$ in equation (8), gives a p value of 0.028. The latter result confirms the existence of period variation in addition to a linear decrease. The precise specification and estimation of the latter additional variation will not be attempted here.

It follows that the least-squares estimate of $\hat{\beta} = -0.207$ obtained from the straight line fit is not to be trusted as a basis from which to estimate the rate of linear period decrease. The estimate can be expected to be biased since it was obtained from fitting an incorrect model to the data. The additional period variation has been absorbed into $Z_j^{(1)}$ in equation (9) and the latter is therefore not a genuine stationary noise process. It is, nevertheless, possible to estimate reasonably accurately the rate of linear period increase without specifying precisely the nature or form of the additional period variation. All that is required is the specification of a functional form that will capture (and not misrepresent) *most* of the additional variation. Doing this will remove to a large extent the bias mentioned above. In this spirit one can postulate the following tentative form for P_j in equation (4), namely

$$P_j = P + \beta_0 \times j + \beta_1 \sin(2\pi j/N) + \beta_2 \cos(2\pi j/N) + Z_j^{(2)}. \qquad (10)$$

There is no implication in this that the additional variation actually has the form specified by the sine and cosine terms. The term $Z_j^{(2)}$ in equation (10) will contain that part of the additional period variation that is not encapsulated by the first four terms on the

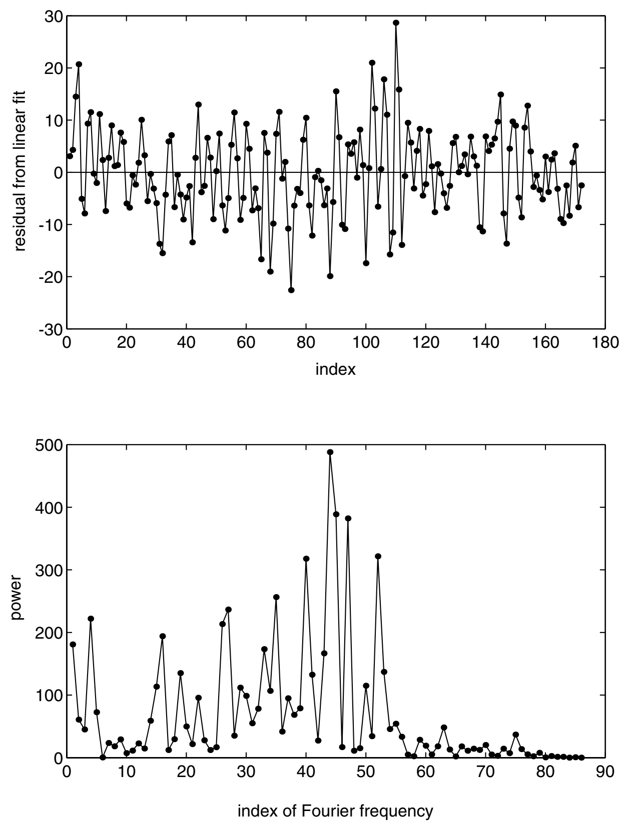

Figure 15. Time series plot of the residuals from a straight line fit to the interleaved sequence of times between maxima and minima of the light curve of R Aql (upper panel), and their corresponding periodogram (lower panel).

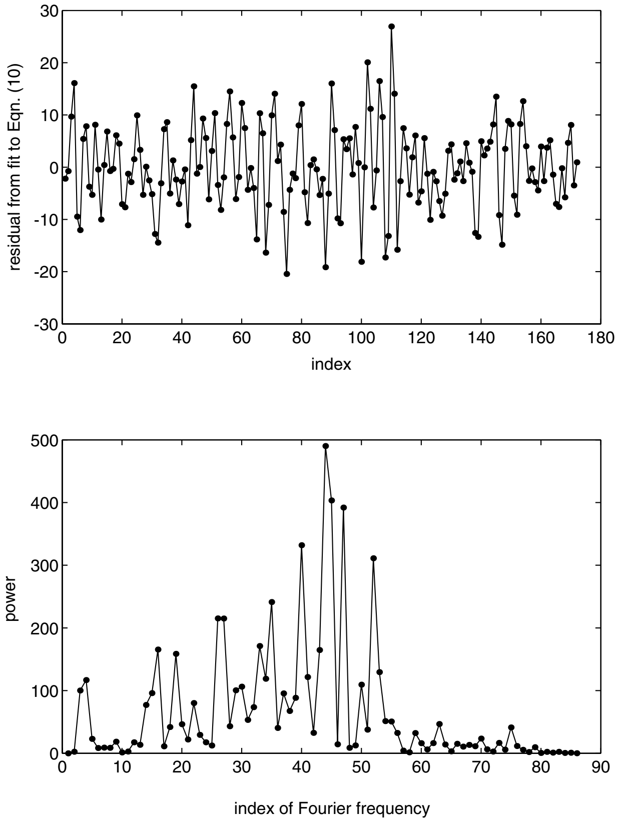

Figure 16. Periodogram of the residuals after fitting equation (10) to the interleaved sequence of times between the maxima and minima of the light curve of R Aql.

right-hand side. The least squares estimate of β_0 is $\hat{\beta}_0 = -0.259$ with a standard error of 0.010. The latter estimate differs substantially from the estimate $\hat{\beta} = 0.207$ obtained when the existence of additional period variation was not taken into account. Fig. 16 shows the periodogram of the residuals from the fit of equation (10). A comparison of this periodogram with that in Fig. 15 reveals that $Z_j^{(2)}$ has much lower power at low frequencies than $Z_j^{(1)}$. Indeed, the T_N test applied to the residuals from equation (10) gives a p value of 0.257.

In other words, the additional period variation that is not encapsulated by equation (10) is not statistically significant. This is not to say that equation (10) is the correct model. It certainly is not. The implication is merely that the difference between equation (10) and the true model is such that it will not affect materially the estimate of the slope β_0 made above. Hence, there are good grounds for assuming that the bias in the estimate $\hat{\beta}_0 = -0.259$ will be negligible, whatever be the true nature of the additional period variation. There are 172 index values that index a total of 51 980 d, i.e. an average of 302.209 d per index. Hence the estimated rate of linear period decrease is $0.259/302.209 = 8.6 \times 10^{-4}\,\mathrm{d}^{-1}$ with a standard error of $0.010/302.209 = 0.3 \times 10^{-4}\,\mathrm{d}^{-1}$.

5 CONCLUDING REMARKS

Two important general points are illustrated by the examples of Section 4. First, the increase in the number of data available when times of maxima and of minima are considered jointly can enable the analyst to draw far more definite conclusions than would otherwise be the case. Secondly, the outcomes of separate analyses of times of maxima and times of minima are never identical and often disagree substantially. The results in Section 4 show how simultaneous analysis of all data reconciles the information content of the two data sets, by taking account of the correlation between the two data sets.

We conclude with a reminder that it should not be expected that tests and estimates based on the simultaneous use of times of maxima and minima will be *twice* as powerful or precise as those based only on either of the two sets of measurements. The reason is, of course, that the two data sets are correlated, i.e. they contain 'some' of the same information. Nevertheless, as the discussion of the previous paragraph highlights, there is certainly much to be gained by pooling all available observations.

REFERENCES

Campbell L., 1926, Harvard Annals, 79, 87
Campbell L., 1955, Studies of Long Period Variables. AAVSO, Cambridge, Mass
Isles J. E., Saw D. R. B., 1987, J. Br. Astron. Assoc., 97, 106
Isles J. E., Saw D. R. B., 1989a, J. Br. Astron. Assoc., 99, 121
Isles J. E., Saw D. R. B., 1989b, J. Br. Astron. Assoc., 99, 165
Isles J. E., Saw D. R. B., 1989c, J. Br. Astron. Assoc., 99, 275
Koen C., Lombard F., 1993, MNRAS, 263, 287 (Paper I)
Koen C., Lombard F., 1995, MNRAS, 274, 821 (Paper III)
Lombard F., 1998a, MNRAS, 294, 657
Lombard F., 1998b, in Zsyszcowicz B., ed., Asymptotic Methods in Probability and Statistics. Elsevier, Amsterdam, p. 585
Lombard F., Koen C., 1993, MNRAS, 263, 309 (Paper II)
Lombard F., Koen C., Potgieter D., 2001, MNRAS, 325, 1133 (this issue)
Mattei J. A., Mayall M. W., Waagen E. O., 1990, Maxima and Minima of Long Period Variables. 1975 AAVSO, Cambridge, MA
Priestley M. B., 1981, Spectral Analysis and Time Series. Academic Press, New York
Rice J., 1979, J. Multivariate Anal., 9, 378
Sterken C., Broens E., Koen C., 1999, A&A, 342, 167
Sterne T. E., 1934, Harvard College Observatory, Circular 386, 1

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