Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **444,** 1486–1495 (2014) doi:10.1093/mnras/stu1514

An O−C (and light travel time) method suitable for application to large photometric databases

Chris Koen★

Department of Statistics, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape, South Africa

Accepted 2014 July 25. Received 2014 July 11; in original form 2014 June 15

ABSTRACT

The standard method of studying period changes in variable stars is to study the timing residuals or O−C values of light-curve maxima or minima. The advent of photometric surveys for variability, covering large parts of the sky and stretching over years, has made available measurements of probably hundreds of thousands of variable stars, observed at random phases. Simple methodology is described which can be used to quickly check such measurements of a star for indications of period changes. Effectively, the low-frequency periodogram of a first-order estimate of the O−C function is calculated. In the case of light travel time (LTT) effects, the projected orbital amplitude follows by robust regression of a sinusoid on the O−C. The results can be used as input into a full non-linear least-squares regression directly on the observations. Extensive simulations of LTT configurations are used to explore the sensitivity of results to various parameter values (period of the variable star and signal to noise of measurements; orbital period and amplitude; number and time baseline of observations). The methodology is applied to observations of three previously studied stars.

Key words: methods: statistical – surveys – stars: individual: 1SWASP J234401.81-212229.1, HD 181469, ASAS 120036-3915.6.

1 INTRODUCTION

Traditional searches for period changes in variable stars are based on the analysis of timing residuals of specific light-curve phases. Typically, in eclipsing binaries the observed times of mid-eclipses are compared to predicted times, whereas for large amplitude pulsating stars such as Miras, Cepheids or RR Lyrae, observed times of maxima are compared to predicted times. For this reason many observed times of maxima or minima of variable stars are published annually. The methodology is generally referred to as 'O−C' (Observed minus Calculated) – see e.g. Sterken (2005), and papers therein. In the case of pulsating stars in binary orbits, or short-period binaries in orbit in triple systems, the O−C plots may reflect periodic behaviour, due to the light travel time (LTT) effect.

Recent years have seen the creation of a number of very large data bases of time series photometry of millions of stars [e.g. MACHO (Alcock et al. 1999), OGLE (Szymański 2005), ASAS (Pojmanski 1997), and SuperWASP (Pollacco et al. 2006), to mention only a few]. The nature of the time series collected by these projects is fundamentally different from those obtained by space missions dedicated to wide-field photometry, such as *Kepler* (e.g. Basri et al. 2011). Quite aside from the superior photometric accuracy achievable from space, the time sampling of the light curves measured from the ground is often sporadic, rather than continuous. The implication is that although there may be an abundance of observations for a given variable star, few (if any) determinations of light maxima or minima may be available. The methodology in this paper, by contrast with the standard approach, makes use of measurements at all phases at which the light curve F is not flat, i.e. the time derivative $F' \neq 0$. The reason for the qualification is not difficult to see; consider, for example, a long-period eclipsing binary star: the extended flat sections of the light curve outside of eclipses do not carry any information about possible period changes.

The approach below relies primarily on time domain calculations. Alternative frequency domain techniques were introduced by Shibahashi & Kurtz (2012, hereafter SK2012), and illustrated with applications to *Kepler* observations. It may be speculated that the frequency domain methodology, which relies on the identification of frequency multiplets in power spectra, may be less efficient for very sparse data.

2 METHOD

The following assumptions are made.

(i) The mean period (or periods, in the case of a multiperiodic pulsating star) is known accurately enough that there are no cycle count ambiguities across any gaps in the data.

(ii) Period changes are small, in a sense to be quantified below.

★ E-mail: ckoen@uwc.ac.za

(iii) The signal to noise of observations is large, also to be quantified below.

(iv) Although there may be slow systematic changes in the periods, there is no appreciable random cycle-to-cycle period jitter (as is particularly evident in long-period pulsating variables). Lombard & Koen (1993) demonstrated that the presence of small random period variations pose problems which require special analysis approaches (see also Sterne 1934, Koen 1996).

(v) The methodology works best on continuously changing light curves (pulsating stars, contact binaries). It is applicable, but less efficient, if the light curve is constant over much of the cycle. Of course, this will be true of any method, since a constant light curve is uninformative as regards period variations.

Denote the O−C at the time t of observation by Δ_t, and the predicted light curve by $F(\mathbf{\Theta}, t)$, where $\mathbf{\Theta}$ is a vector of descriptive parameters (zero-point, frequencies and amplitudes):

$$F(\mathbf{\Theta}, t) = \mu + \sum_{k=1}^{K} [A_k \cos(\omega_k t) + B_k \sin(\omega_k t)] \,. \tag{1}$$

Since the 'calculated' time C is $t - \Delta_t$, it follows that the observed light curve is

$$y_t = F(\mathbf{\Theta}, t - \Delta_t) + e_t, \tag{2}$$

where e_t represents measurement noise with variance σ_e^2. A first-order Taylor expansion gives

$$y_t \approx F(\mathbf{\Theta}, t) - \Delta_t F'(\mathbf{\Theta}, t) + e_t \tag{3}$$

and

$$\Delta_t \approx -r_t / F'(\mathbf{\Theta}, t),$$

where r_t is the residual

$$r_t = y_t - F(\mathbf{\Theta}) + e_t,$$

which can of course be estimated by

$$r_t \approx y_t - F(\mathbf{\Theta}, t) \,.$$

It follows that

$$\Delta_t \approx -[y_t - F(\mathbf{\Theta}, t)] / F'(\mathbf{\Theta}, t) \,. \tag{4}$$

Equation (4) is the crux of the proposed methodology. The estimated O−C values Δ_t are trivially calculated from the model fit residuals, and from the derivative of the fitted model. Δ_t can be analysed by e.g. checking for periodicities (indicative of apsidal motion, or a light-time effect), or by noting trends (indicative of systematic period changes). The approximation is accurate provided higher order terms in the Taylor series expansion are unimportant. This will typically be the case if

$$|F' \Delta_t / F| \ll 1,$$

i.e.

$$\omega_* |\Delta_t| \ll 1, \tag{5}$$

where ω_* is the frequency associated with the dominant periodicity in (1).

The residuals r_t in equation (4) consist of two terms:

$$r_t = [y_t - F(\mathbf{\Theta}, t)] + e_t \,,$$

the first being due to 'model misspecification' and the second due to measurement errors. In the present context, the model misspecification is of course due to the non-inclusion in equation (1) of period changes or light-time effects. In order to be able to identify the misspecified term at all,

$$\frac{1}{N} \sum_{t=1}^{N} [y_t - F(\mathbf{\Theta}, t + \Delta_t)]^2 \approx \frac{1}{N} \sum_{t=1}^{N} [\Delta_t F'(\mathbf{\Theta}, t)]^2 \gtrsim \sigma_e^2$$

is clearly necessary. In terms of ω_*,

$$\frac{C^2 \omega_*^2}{2N} \sum_{t=1}^{N} (\Delta_t)^2 \gtrsim \sigma_e^2$$

is required, where C is the amplitude associated with ω_*. This can also be written as

$$\mathrm{RMS}(\Delta_t) \gtrsim \frac{1.4 \sigma_e}{C \omega_*} \approx \frac{1.4}{\omega_* R}, \tag{6}$$

where $R = C/\sigma_e$ is the signal-to-noise ratio of the observations y_t, and $\mathrm{RMS}(\Delta_t)$ is the root mean square of Δ_t.

3 ESTIMATION IN THE CASE OF THE LTT EFFECT

The relevant formulae are:

$$\Delta_t = A \left[\frac{1 - e^2}{1 + e \cos\theta} \sin(\theta + \varpi) + e \sin \varpi \right],$$

$$\theta = 2 \arctan \left[\sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2} \right],$$

$$E - e \sin E = \frac{2\pi}{P} (t - t_0) \tag{7}$$

(e.g. Irwin 1952). In these formulae e is the eccentricity, E and θ the eccentric and true anomalies, ϖ the argument of periapsis, t_0 gives the time of zero eccentric anomaly, and P is the orbital period of the variable star around the system centre of mass. The amplitude $A = a \sin i / c$, with a the length of the semimajor axis, and i the inclination of the light-time orbit with respect to the line of sight.

For circular orbits

$$\Delta_t = A \sin \left[\frac{2\pi}{P} (t - t_0) \right] . \tag{8}$$

From equation (8),

$$\overline{(\Delta_t)^2} \sim A^2/2,$$

and hence equation (6) implies

$$A \gtrsim \frac{2}{\omega_* R} \,. \tag{9}$$

As an example, in the simulation experiments below in which $\omega_* = 2\pi \times 3\ \mathrm{d}^{-1}$ and $R = 40$, orbit sizes in excess of ∼230 light-second (0.46 au) are in principle detectable. Equation (5) requires

$$A \omega_* \ll 1$$

which is easily satisfied in all the examples which follow.

Simulated data are used to illustrate the estimation of the various parameters appearing in equations (7) or (8). Consider $N = 720$ measurements of a single sinusoid with frequency 3 d^{-1}, spread over a time base of $T = 3600$ d. This represents an average of about one measurement every 5 d, of a contact binary or short-period pulsator such as a δ Scuti or β Cephei star, observed over the course of 10 yr. The signal-to-noise ratio is taken as $R = 20$, and the LTT amplitude $A = 300$ s and period $P = 750$ d (frequency $f = 0.001\,33\ \mathrm{d}^{-1}$). The eccentricity is $e = 0$.

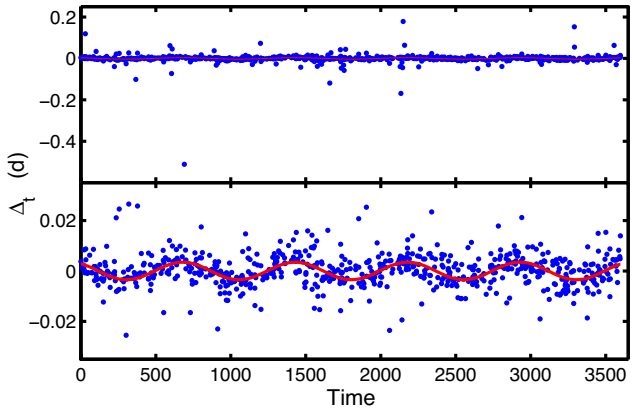

Figure 1. Top panel: O−C values estimated using equation (4), for a simulated data set with parameters given in the text. The outlying points are due to near-zero derivates of the light curve in the denominator of equation (4). Bottom panel: the result of trimming the largest 5 per cent (in absolute value) of the points in the upper panel. The solid curve is the true Δ_t.

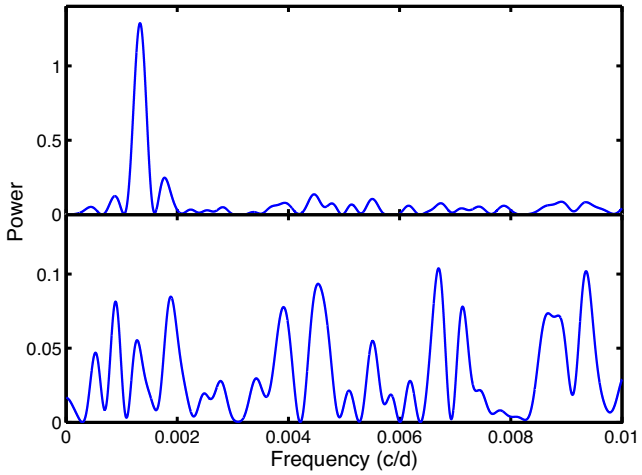

Figure 2. Top panel: the periodogram of the data in the bottom panel of Fig. 1. Bottom panel: the periodogram of the residuals left after pre-whitening the data by the sinusoid which best (by least squares) fits the Fig. 1 (bottom panel) data. Note the difference in the vertical scales on the two panels.

The estimated residuals from a simulation are plotted in the top panel of Fig. 1. A prominent feature is the sprinkling of outliers. The origin of these points is not hard to find: near-zero denominators in equation (4). Trimming away the largest 5 per cent of absolute values leaves the data points plotted in the lower panel, in which the actual modulating function Δ_t is also plotted.

The periodogram of Δ_t in the lower panel of Fig. 1 is shown in the top panel of Fig. 2. There is a strong peak at a frequency corresponding to a period of 751.6 d. Pre-whitening by this period results in the residual spectrum in the bottom panel of the figure: clearly, there are no further features of note, at least in the frequency range plotted.

The sinusoid which minimizes the residual sum of squares has an amplitude of 244 s, i.e. considerably lower than the true amplitude. This is a result of further outlying data visible in the bottom panel of Fig. 1: although not as extreme as the trimmed outliers, the presence of these data points none the less leads to underestimation of the amplitude. Robust regression, rather than ordinary least-squares fitting, was therefore applied. The particular method used was 'iteratively reweighted least squares', i.e. minimizing a weighted sum of squares, with weights adjusted iteratively (e.g. Holland & Welsch 1977). The weight function chosen was the bisquare form of Beaton & Tukey (1974). The results of the robust non-linear regression of the model (8) on the data Δ_t are an estimated amplitude $A = 290$ s, LTT period of $P = 753$ d, and $t_0 = 95$ d (true value 117 d).

There is a hitch in the case of elliptical orbits. Points in Fig. 3 show the estimated LTT shifts Δ_t, with data simulated in accordance with the same configuration as above, except that $e = 0.7$ and the signal-to-noise ratio is 100. The lower (blue) line shows the true Δ_t, as given by equation (7): the data clearly do not follow it very well. Instead, the data cluster around the upper (red) line, which represents $\Delta_t - \overline{\Delta_t}$.

The point is that Δ_t as defined in equation (4) estimates the zero mean function $\Delta_t - \overline{\Delta_t}$, rather than Δ_t. Comparison of equations (7) and (8) shows that this is not important if $e = 0$ (provided the total time baseline T spans a few orbital cycles), but may be quite important if $e \neq 0$, depending on the value of ϖ. For the data set plotted in Fig. 3, the estimated values of $A = 234$ s and $e = 0.23$ are badly biased, as are the estimates of $t_0 = 138$ (true value 70) and $\varpi = 2.61$ (true value 2.11).

The simplest solution is to fit the model $\Delta_t + \delta$ to the observed Δ_t. Estimated parameters obtained in this way are $A = 276$ s, $e = 0.62$, $t_0 = 64$ and $\varpi = 2.02$ – a considerable improvement. These values also compare favourably with those obtained by fitting the full model described by equations (1), (2) and (7) directly to the simulated observations ($A = 279$, $e = 0.63$, $t_0 = 67$ and $\varpi = 2.05$).

Based on the experience above, the recommended LTT model-fitting strategy is the following.

(i) Calculate Δ_t and trim the largest 5–10 per cent of values (5 per cent was used throughout this paper).

(ii) Calculate the periodogram of Δ_t. If there is no power excess at low frequencies, then there is no evidence for a variable period. Otherwise, if there is an excess at a single frequency, a model with $e = 0$ is indicated; if there are excesses at harmonically related frequencies, a model with $e > 0$ is indicated; and if there are multiple, unrelated, low frequencies with power excesses, the period is an aperiodic function of time. The period of maximum power will be denoted by $P^{(1)}$.

(A caveat: if there are substantial gaps in the data, aliasing will introduce additional structure in the periodogram.)

(iii) An estimate $A^{(1)}$ of the amplitude A follows by ordinary least-squares fitting of a sinusoid with period $P^{(1)}$ to Δ_t.

(iv) These estimates can be improved if the sinusoid is fitted to Δ_t by a robust procedure, giving $P^{(2)}$ and $A^{(2)}$.

(v) Fit the model (7) or (8) to Δ_t, using robust non-linear regression. The estimates $P^{(2)}$ and $A^{(2)}$ are used as starting guesses, and the iteratively improved estimates are $P^{(3)}$ and $A^{(3)}$. If $e \neq 0$ is assumed, then $e^{(1)}$ is also obtained.

(vi) The estimated parameters from (v) can now serve as initial estimates in a direct fit of the model (2) to the observations, by an ordinary least-squares algorithm. The final estimated parameter values are $P^{(4)}$, $A^{(4)}$ and $e^{(2)}$.

4 RESULTS OF SIMULATION STUDIES

The estimation procedure developed in the previous section was applied to a large number of simulated data sets, in order to assess its efficacy. The base parameter set was $N = 720$ observations of a $f = 3$ d^{-1} sinusoid, uniformly distributed over a time baseline

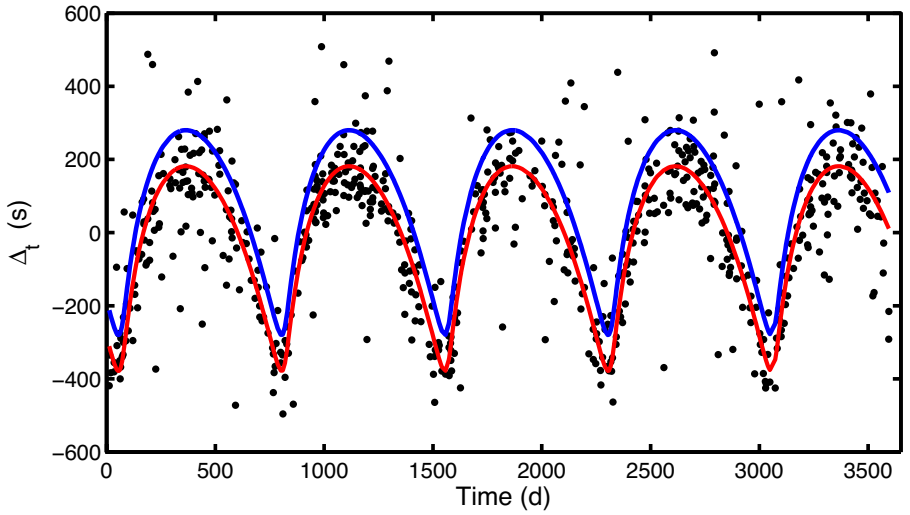

Figure 3. Points represent estimated (O–C) values, for a simulated data set. The assumed parameter values are the same as for the data in Fig. 1, except that $e = 0.7$ and the signal-to-noise ratio is higher at $R = 100$. The top (blue) line is the true underlying function Δ_t, while the bottom (red) line is the mean-adjusted function $\Delta_t - \overline{\Delta_t}$.

$T = 3600$ d, measured with a signal to noise of 20. The LTT effect was simulated by assuming $P = 750$ d and an amplitude of $A = 300$ s. These parameter values were varied one at a time, in order to see the effects on the bias and dispersion of the estimates.

Assuming uniform time spacing has the advantage that the simulation results do not depend on the specifics of the times of observation. The disadvantage is that uniform coverage is not representative of most current large-scale surveys. Results are therefore also presented for simulations using the base parameter values, but with time spacing designed to resemble those typical of surveys conducted from a single site. This was accomplished by subdividing the 3600 d time baseline into 10 blocks of length 360 d each, consisting of 260 d with uniformly distributed measurements, and a 100 d seasonal gap with no observations.

For each configuration, 2000 simulations were performed, with t_0 and ϖ randomized over $[0, P]$ and $[0, 2\pi]$, respectively. The exercise was performed for eccentricities $e = 0, 0.3, 0.7$: for the zero eccentricity case, $e = 0$ was assumed in the parameter estimation, i.e. equation (8), rather than (7), was fitted to the data.

The biases and standard errors of some of the parameter estimates are presented in Tables 1– 3. Some remarks on the information in the tables, and other aspects of the results, follow.

(1) For circular orbits, the distributions of $\widehat{P}$ and $\widehat{A}$ are generally roughly Gaussian, as judged from quantile–quantile plots. Deviations from the normal are usually heavy tailed. Table 1 shows that $A^{(1)}$ and $\widehat{\sigma}_e$ aside, biases are negligible. As mentioned above, $A^{(1)}$ is biased by the presence of outlying values of Δ_t, and clearly the robust regression used to determine $A^{(2)}$ and $A^{(3)}$ fixes the problem.

(2) Fig. 4 demonstrates that the upward bias in estimated $\widehat{\sigma}_e$ is due to the presence of an extended tail to large values. Interestingly, this bias is significantly reduced if the sample size is increased (model 5), but *increased* if the simulated measurement error variance σ_e^2 is decreased.

(3) Although the estimated values of σ_e are poorest at the highest signal to noise, the second half of Table 1 demonstrates that, as expected, all other estimates are more accurate. The substantial reduction in standard errors for model 3, with $f = 6$ is noteworthy: the more rapid the brightness changes of the variable star, the better. Adding a second, unrelated, frequency (model 9) also benefits the final estimates $P^{(4)}$ and $A^{(4)}$, while adding a subharmonic (model 8) makes little difference. Interestingly, the orbital period is estimated more accurately if A is increased, but the estimates of A itself are not affected much (model 6).

(4) For $e = 0.7$, a small number of grossly outlying solutions are obtained (0–9 of the 2000 simulations, depending on the model). These have been ignored in the discussion which follows.

(5) Bias in estimated values of P remains small for $e \neq 0$ (Tables 2 and 3, first halves). The same cannot be said for the estimated amplitudes A. The fact that $A^{(1)}$ and $A^{(2)}$ underestimate the true amplitude is easy to understand: these estimates are obtained by fitting sinusoids to data which have more complicated forms (particularly for larger e). The positive bias of $A^{(3)}$ and $A^{(4)}$, based on the correct functional forms of Δ_t, arises similarly to the bias in $\widehat{\sigma}_e$ – through an extended distribution tail towards large values (see point 2 above).

(6) The bias in the estimates of A is reduced by high signal to noise (model 7), or a denser spacing of observations (model 5). A higher variability frequency (model 3) also reduces bias in $A^{(3)}$ and $A^{(4)}$ if the eccentricity is moderate. The model dependence of the standard errors of $A^{(3)}$ and $A^{(4)}$ parallels that of the bias.

(7) Both bias and standard errors of $e^{(2)}$ are generally smaller than those of $e^{(1)}$. The distribution of estimates becomes bimodal with increasing eccentricity – see Fig. 5. The diagram also demonstrates the origins of the positive bias and large standard errors of the estimated eccentricities. Interestingly, bias and standard errors are smaller for model 6 ($A = 450$ instead of the usual 300), in addition to models 5 (more observations) and 7 (high signal to noise). There is also some improvement in the estimates if the variability is more rapid (model 3), but, as before, only in the case of moderate eccentricity.

(8) Fig. 6 illustrates the interdependence between estimates of A and e. It is clear that the distributions of the amplitude values have long tails to large values: this accounts for the very large standard errors of these estimates (Tables 2 and 3). By contrast, for most parameter sets in the study the distribution of the estimated periods

Table 1. Simulation results for the case $e = 0$. The base parameter set is $P = 750$ d, $A = 300$ s, $N = 720$, $T = 3600$ d, and assumes that the star varies sinusoidally with a frequency $f = 3$ d^{-1}. In each of models 2–7 one of these parameters is changed, in order to study its influence on the results. In model 8, a subharmonic is added to the base model, to mimic the light curve of a contact binary. In model 9, a second sinusoidal variation with $f = 6.7$ is added to the model 1 parameter set, to represent multiperiodicity. Signal-to-noise ratios of the additional periodicities are set at 30 per cent of those of the primary periodicity. The time spacing in model 10 incorporates seasonal gaps – see the text for details. The first half of the table reports biases, and the second standard errors of estimates. The entries for a given model are based on the estimates for each of 2000 simulated data sets, assuming Gaussian measurement errors, and values of t_0 uniformly distributed over $[0, P]$.

	1 Base	2 $P = 1000$	3 $f = 6$	4 $T = 1800$	5 $N = 1500$	6 $A = 450$	7 $\sigma_e = 0.025$	8 $f_2 = 1.5$	9 $f_2 = 6.6$	10 Gaps
					Bias					
$P^{(1)}$	− 0.23	0.14	− 0.42	2.19	− 0.24	− 0.31	− 0.44	− 0.71	0.14	− 0.57
$P^{(2)}$	− 0.43	− 0.62	− 0.40	2.02	− 0.19	− 0.23	− 0.34	− 0.54	− 0.17	− 0.08
$P^{(3)}$	− 0.43	− 0.60	− 0.42	2.02	− 0.18	− 0.23	− 0.35	− 0.53	− 0.16	− 0.10
$P^{(4)}$	− 0.14	0.01	− 0.11	0.70	0.08	0.08	− 0.01	− 0.13	− 0.03	0.22
$A^{(1)}$	− 14.57	− 19.18	− 16.53	− 19.09	− 15.65	− 23.70	− 17.06	− 15.87	− 14.83	− 14.74
$A^{(2)}$	0.24	− 3.32	− 0.66	− 3.80	− 0.26	− 1.15	− 1.01	− 0.46	− 0.27	0.53
$A^{(3)}$	0.18	− 3.33	− 0.72	− 3.91	− 0.30	− 1.25	− 1.05	− 0.52	− 0.40	0.40
$A^{(4)}$	0.60	0.13	0.07	1.48	0.30	0.72	0.62	1.26	1.06	2.27
$\widehat{\sigma}_e$	0.0062	0.0059	0.0060	0.0057	0.0031	0.0061	0.0097	0.0068	0.0066	0.0060
					Standard errors					
$P^{(1)}$	9.56	18.54	5.27	22.92	6.50	6.39	5.38	9.40	10.84	15.16
$P^{(2)}$	6.31	11.80	3.14	13.47	4.33	4.28	3.10	6.32	6.50	6.74
$P^{(3)}$	6.28	11.74	3.13	13.42	4.30	4.27	3.10	6.29	6.49	6.74
$P^{(4)}$	5.63	10.20	2.84	11.17	3.69	3.81	3.67	5.68	4.75	5.94
$A^{(1)}$	29.18	28.29	15.01	29.25	20.20	29.22	14.52	28.61	33.00	32.88
$A^{(2)}$	21.79	21.55	10.86	21.84	14.87	21.65	10.44	20.86	21.24	22.95
$A^{(3)}$	21.70	21.42	10.85	21.80	14.83	21.59	10.40	20.85	21.16	22.93
$A^{(4)}$	19.27	19.40	9.67	19.21	12.60	20.26	12.70	19.07	16.24	20.64
$\widehat{\sigma}_e$	0.0079	0.0075	0.0078	0.0075	0.0041	0.0079	0.0110	0.0086	0.0082	0.0077

Table 2. As for Table 1, but for $e = 0.3$. For each simulated data set ϖ was generated from a uniform distribution on $[0, 2\pi]$.

	1 Base	2 $P = 1000$	3 $f = 6$	4 $T = 1800$	5 $N = 1500$	6 $A = 450$	7 $\sigma_e = 0.025$	8 $f_2 = 1.5$	9 $f_2 = 6.6$	10 Gaps
					Bias					
$P^{(1)}$	− 0.65	− 0.35	− 0.20	1.02	− 0.15	− 0.33	− 0.36	− 0.36	− 0.01	0.02
$P^{(2)}$	− 0.61	− 0.71	− 0.20	1.62	− 0.06	− 0.24	− 0.25	− 0.27	− 0.23	− 0.41
$P^{(3)}$	− 0.31	− 0.32	0.17	0.31	0.19	0.06	0.02	0.00	0.14	0.003
$P^{(4)}$	− 0.28	− 0.08	0.19	0.13	0.07	− 0.01	0.14	0.16	0.01	0.085
$A^{(1)}$	− 28.42	− 29.08	− 29.36	− 30.76	− 28.58	− 43.16	− 29.73	− 28.98	− 27.98	− 27.88
$A^{(2)}$	− 13.01	− 14.59	− 14.22	− 16.60	− 13.48	− 20.73	− 14.53	− 14.13	− 13.51	− 13.24
$A^{(3)}$	17.48	13.86	2.48	16.72	5.22	7.18	1.81	13.87	16.09	18.87
$A^{(4)}$	13.26	13.20	2.61	14.04	4.12	6.44	3.64	14.16	7.18	17.64
$e^{(1)}$	0.065	0.059	0.013	0.063	0.024	0.023	0.008	0.060	0.062	0.071
$e^{(2)}$	0.048	0.048	0.011	0.052	0.017	0.016	0.013	0.049	0.034	0.056
$\widehat{\sigma}_e$	0.0060	0.0061	0.0058	0.0057	0.0031	0.0060	0.0096	0.0064	0.0062	0.0061
					Standard errors					
$P^{(1)}$	9.57	18.79	5.53	23.50	6.90	6.95	5.51	9.95	11.32	15.22
$P^{(2)}$	6.52	11.77	3.49	14.08	4.64	4.50	3.50	6.62	6.75	6.95
$P^{(3)}$	6.55	11.56	3.25	13.20	4.33	4.24	3.21	6.55	6.60	6.90
$P^{(4)}$	5.64	10.28	2.88	12.05	3.63	3.88	3.80	5.93	4.89	6.16
$A^{(1)}$	29.39	29.04	15.46	29.10	20.23	30.05	15.30	28.11	32.65	23.34
$A^{(2)}$	21.53	21.90	11.34	22.29	15.23	22.56	11.38	21.08	21.78	30.91
$A^{(3)}$	54.31	36.76	12.82	50.40	18.85	26.69	12.89	43.15	51.40	70.50
$A^{(4)}$	52.00	47.08	11.45	56.59	16.11	23.79	15.86	48.44	22.47	60.26
$e^{(1)}$	0.16	0.16	0.08	0.16	0.11	0.10	0.08	0.16	0.17	0.17
$e^{(2)}$	0.15	0.15	0.07	0.15	0.09	0.09	0.09	0.15	0.12	0.16
$\widehat{\sigma}_e$	0.0076	0.0077	0.0075	0.0074	0.0044	0.0080	0.0110	0.0082	0.0077	0.0079

Table 3. As for Table 1, but for $e = 0.7$. For each simulated data set ϖ was generated from a uniform distribution on $[0, 2\pi]$.

	1 Base	2 $P = 1000$	3 $f = 6$	4 $T = 1800$	5 $N = 1500$	6 $A = 450$	7 $\sigma_e = 0.025$	8 $f_2 = 1.5$	9 $f_2 = 6.6$	10 Gaps
					Bias					
$P^{(1)}$	0.14	0.33	− 0.15	1.49	− 0.02	− 0.39	− 0.03	− 0.28	− 0.26	− 0.31
$P^{(2)}$	− 0.16	− 0.01	− 0.18	1.94	− 0.11	− 0.46	− 0.17	− 0.30	− 0.33	− 0.21
$P^{(3)}$	− 0.03	0.13	0.07	0.50	0.16	− 0.02	− 0.03	0.01	− 0.02	− 0.18
$P^{(4)}$	0.12	0.13	− 0.05	0.26	0.11	0.04	− 0.04	0.07	− 0.01	− 0.49
$A^{(1)}$	− 85.12	− 85.30	− 83.73	− 87.63	− 85.33	− 131.8	− 88.11	− 85.51	− 84.26	− 83.05
$A^{(2)}$	− 74.95	− 74.90	− 74.28	− 77.14	− 74.34	− 115.6	− 76.98	− 74.98	− 75.82	− 72.46
$A^{(3)}$	155.1	143.6	150.0	150.3	58.63	90.79	23.39	151.4	145.8	150.51
$A^{(4)}$	97.04	98.03	98.60	98.37	30.96	60.85	36.63	92.89	70.19	100.98
$e^{(1)}$	0.077	0.074	0.080	0.076	0.043	0.038	0.024	0.076	0.072	0.069
$e^{(2)}$	0.061	0.059	0.065	0.065	0.028	0.032	0.030	0.059	0.048	0.061
$\widehat{\sigma}_e$	0.0058	0.0062	0.0058	0.0059	0.0031	0.0060	0.0095	0.0061	0.0064	0.0060
					Standard errors					
$P^{(1)}$	12.79	23.47	13.11	28.16	8.82	9.00	6.99	12.81	14.56	17.09
$P^{(2)}$	8.76	15.65	8.91	18.85	6.11	6.23	4.91	8.84	8.90	9.55
$P^{(3)}$	7.56	13.09	7.28	15.72	4.72	4.75	3.31	7.41	7.36	7.99
$P^{(4)}$	6.43	11.11	6.41	13.80	3.81	4.08	4.04	6.54	5.36	6.94
$A^{(1)}$	31.67	33.00	33.24	31.87	25.20	37.26	21.18	33.05	35.4	28.40
$A^{(2)}$	27.03	27.17	27.12	27.45	22.13	32.74	19.80	26.77	26.89	34.31
$A^{(3)}$	252.7	237.0	260.28	252.8	136.8	225.6	73.16	252.3	250.2	260.52
$A^{(4)}$	172.8	177.4	174.6	173.2	82.71	162.7	102.4	171.0	146.1	180.76
$e^{(1)}$	0.20	0.19	0.19	0.19	0.14	0.14	0.10	0.19	0.20	0.20
$e^{(2)}$	0.18	0.18	0.17	0.18	0.12	0.13	0.12	0.18	0.15	0.18
$\widehat{\sigma}_e$	0.0076	0.0080	0.0074	0.0079	0.0043	0.0073	0.0106	0.0078	0.0084	0.0077

P are roughly Gaussian, albeit with the occasional outlying value or heavy tails.

(9) The estimates $A^{(4)}$ and $P^{(4)}$ are consistently better than $A^{(3)}$ and $P^{(3)}$, which implies that fitting of the full non-linear model (2) is worthwhile.

(10) It comes as no surprise that aliasing is encountered in the simulations in which the time spacing has seasonal gaps (model 10). What is surprising is the low level of its presence: 6, 3 and 42 cases (out of 2000) in the simulations with $e = 0$, 0.3, 0.7, respectively. Most of the aliases of the LTT period occurred near 245 d. These, together with two outliers ($P^{(1)} > 1000$ d) in the case $e = 0.7$, were excluded from the calculations reported in Tables 1– 3.

(11) Continuing with model 10, the biases and standard errors of the estimates of the eccentricities and σ_e are similar to those of the base model. The situation with the estimated LTT periods and amplitudes is less clear-cut, although generally the biases and standard errors are somewhat larger in the model 10 simulations than in the case of model 1.

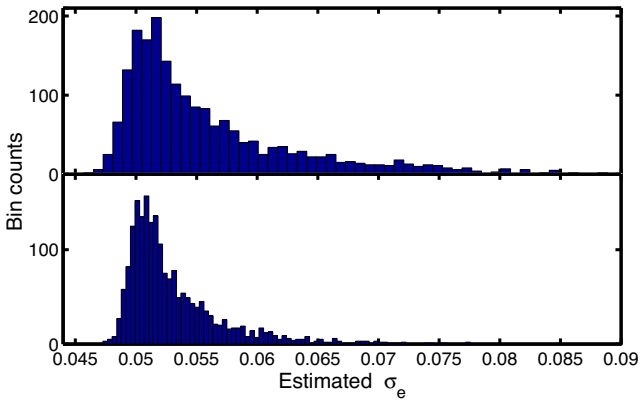

Figure 4. The distribution of the estimated measurement error standard deviations $\widehat{\sigma}_e$ for models 1 (top panel) and 5 (bottom panel) in the circular orbit case. Nine values in the range (0.09, 0.11) have been excluded from the top plot.

5 THREE APPLICATIONS TO REAL DATA

5.1 1SWASP J234401.81-212229.1

Variability in 1SWASP J234401.81-212229.1 (hereafter SWASP 2344-2122) was discovered by the SuperWASP enterprise (Pollacco et al. 2006): the star is an eclipsing binary, with a period of 0.213 676 d ($f = 4.679\,984$ d^{-1}). Traditional (O−C) studies of SWASP 2344-2122 were published by Lohr et al. (2013a,b). The authors based their analyses on 267 times of minima, estimated from measurements spread over 6.6 yr. Koen (2014) fitted the model (2) directly to 21540 SuperWASP measurements covering the first 5.2 yr of operation of the telescope. Here, the latter data set is re-analysed, using the methodology of this paper.

The Δ_t are plotted in the top panel of Fig. 7, and the 5 per cent trimmed values in the bottom panel. The periodogram of the trimmed data is shown in the top panel of Fig. 8. The multiple peaks suggest that the estimated (O–C) values are multiperiodic, but pre-whitening by a sinusoid with a frequency of 7.85×10^{-4} d^{-1} ($P = 1274$ d), leads to the residual periodogram plotted in the bottom panel of Fig. 8. The mean level of the power in this residual spectrum is about two orders of magnitude lower than the peak power in the top panel.

The seasonal gaps in the data in Fig. 7 induce aliasing, as could be demonstrated by plotting spectral window functions. For the reader not familiar with such time series analysis tools, an equivalent, but more transparent, demonstration follows.

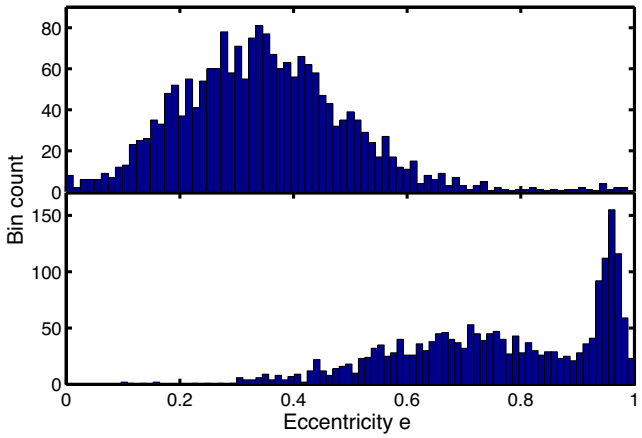

Figure 5. The distributions of the estimated eccentricities $e^{(2)}$ for model 1, for $e = 0.3$ (top panel) and $e = 0.7$ (bottom panel).

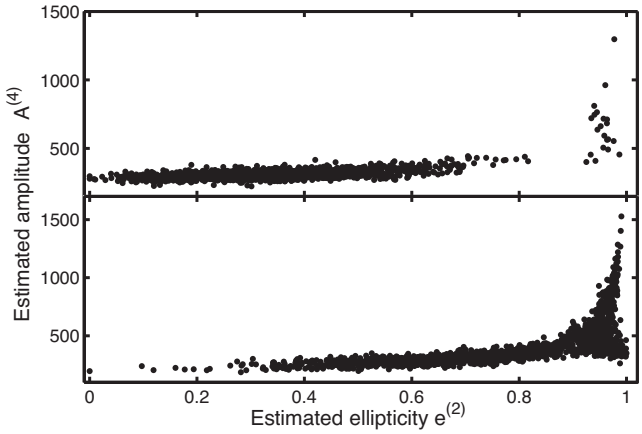

Figure 6. Estimated eccentricities and amplitudes A for 2000 data sets simulated using model 7 parameters with $e = 0.3$ (top panel) and $e = 0.7$ (bottom panel).

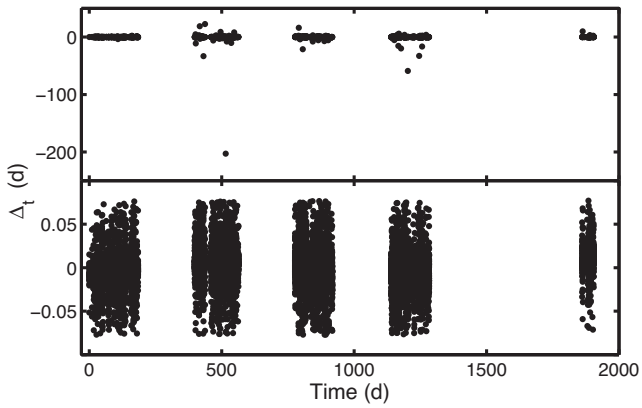

Figure 7. As for Fig. 1, but for the SuperWASP measurements of SWASP 2344-2122.

Fig. 9 shows that the large peaks near frequencies 0.002 and 0.0035 in the top panel of Fig. 8 (reproduced in the top panel of Fig. 9) are artefacts. A test data set was produced as follows.

(i) The frequency of the largest periodogram peak in the top panel of Fig. 8 was used as starting guess in a robust iterative fit of a sinusoid to Δ_t (i.e. $P^{(2)}$ and $A^{(2)}$ were estimated).

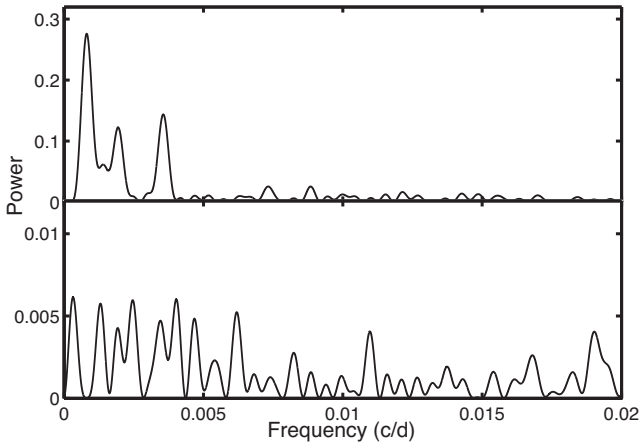

Figure 8. Top panel: the periodogram of the data in the bottom panel of Fig. 7. Bottom panel: the periodogram of the residuals after pre-whitening the highest-peak frequency selected from the periodogram in the top panel. Note the markedly different scales on the two vertical axes.

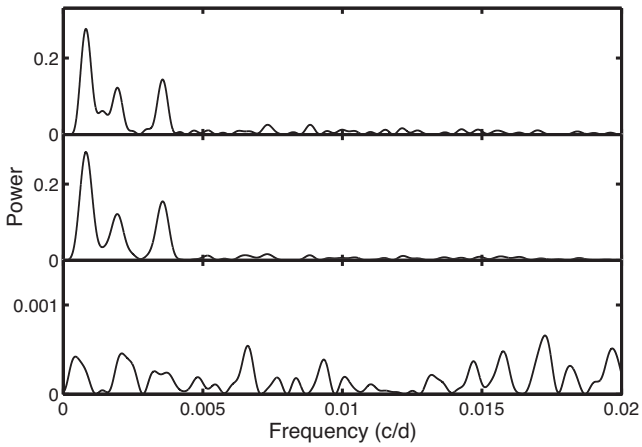

Figure 9. The top panel contains the periodogram of the data in Fig. 7, while the middle panel is the periodogram of a single sinusoid with added white noise, with the same time sampling as in the bottom panel of Fig. 7. If the frequency corresponding to the largest peak in the middle panel is pre-whitened from the simulated data, the periodogram in the bottom panel results.

(ii) The fitted sinusoid was subtracted from the data, leaving a set of residuals.

(iii) A random permutation of the residuals were performed in order to generate an artificial noise process with the same statistical distribution as that of the residuals.

(iv) A sinusoidal 'signal'

$$A^{(2)} \cos(2\pi t / P^{(2)} + \psi)$$

(where ψ is a random phase) was added to the noise generated in step (iii).

The spectrum of the test data set, which is of course known to contain a single sinusoid, is plotted in the middle panel of Fig. 9. The remarkable similarity to the spectrum in the top panel confirms that the two prominent additional low-frequency peaks in the spectra are aliases, induced by the large seasonal gaps visible in Fig. 7.

It is also instructive to compare the bottom panels in Figs 8 and 9. Both are spectra of residuals, but the level in the former is much higher than in the latter. By its construction, the noise process used

Table 4. Estimated circular orbital parameters for SWASP 2344-2122. The frequencies $f_1 = 4.679\,984$ and $f_2 = 4.679\,972$ d^{-1} are from Lohr et al. (2013b) and Koen (2014), respectively.

	$2 \times f_1$	$(1–6) \times f_1$	$2 \times f_2$	$(1–6) \times f_2$
t_0	882	880	859	857
$P^{(4)}$	1272	1271	1340	1334
$A^{(4)}$	510	511	536	538
$\widehat{\sigma}_e$	0.023	0.022	0.023	0.022

to generate the lower two panels of Fig. 9 was white, hence it may be concluded that the spectrum in the bottom panel of Fig. 8 does not reflect white noise. This is also evident from the frequency dependence of the residual spectrum in Fig. 8: there is an excess of power over a range [roughly (0, 0.006) d^{-1}] of low frequencies. The implication is that there is some low-level aperiodic low-frequency variability in the (O–C) values in addition to the overt sinusoidal variability.

As a final word on the spectra, it is noted that there is no sign of any excess power at the first harmonic of the peak at 7.85×10^{-4} d^{-1}. This implies that these data do not show any evidence for an eccentricity different from zero. As a consequence, only models with $e = 0$ were fitted to Δ_t. Initial guesses for the unknown parameters (P, A and t_0) were generated as described in steps (i)–(v) in Section 3. These were used in the non-linear least-squares fitting of the models (2) and (8). Table 4 reports the results. Since the light curve of SWASP 2344-2122 is dominated by the first harmonic, models with this single frequency were fitted to the data, using two different values from the literature. Additionally, since Koen (2014) showed that the first six harmonics could be identified in spectra of *all* the observations of SWASP 2344-2122 (i.e. including some of his own photometry), sets of models using all these frequencies were also fitted. Comparison of the different entries in Table 4 shows non-negligible, but not alarming, differences between the different models.

There are also 610 high-quality ASAS measurements of the star, over an interval of 8.9 yr. Only the first harmonic of the binary frequency is prominent in the spectrum of the data, at $2f_3 = 9.359\,8388 = 2 \times 4.679\,9194$ d^{-1}. This differs considerably from the frequencies derived from the SuperWASP observations. Although the time baseline is longer than that of the SuperWASP data, the measurements are far fewer, and also more noisy. Only the two frequencies f_1 and f_2 (see Table 4) are therefore used below.

Periodograms of the 5 per cent trimmed values of Δ_t, as derived from the ASAS observations, are plotted in Fig. 10. Results in the top two panels are based on the assumption that the signal in the observations was a sinusoid with frequency $2f_1$, while calculations resulting in the bottom panels used $2f_2$. The highest peaks in panels (a) and (b) are at frequencies of 1.6×10^{-4} and 1.7×10^{-4} d^{-1}, respectively (i.e. periods of around 17 yr). This indicates a systematic trend in Δ_t. Pre-whitening leads to the residual periodograms in panels (c) and (d), in which the second-highest peaks in panels (a) and (b) persist, at frequencies 7.7×10^{-4} and 7.8×10^{-4} d^{-1}, respectively (periods of 1300 and 1280 d).

The point here is that the much less voluminous and more noisy ASAS data also show evidence for an orbital period of order 1300 d. Additionally, there is evidence for a systematic trend in the period.

5.2 HD 181469

SK2012 performed a frequency domain LTT analysis of *Kepler* observations of a δ Scuti pulsator in a multiple star system (HD 181469 = KIC 4150611). The pulsator has four prominent frequencies, listed in their table 3, and used in the present analysis. Professor Kurtz kindly supplied the author with 33 078 *Kepler* observations

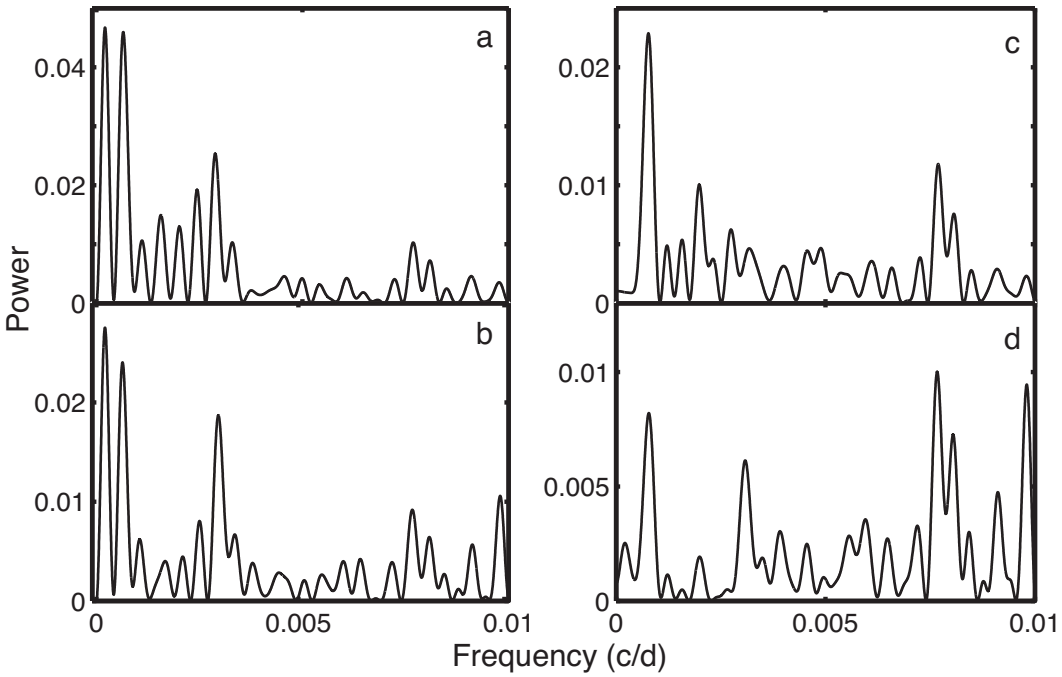

Figure 10. (a) The periodogram of Δ_t, estimated from the ASAS observations of SWASP 2344-2122, and derived assuming a binary period of $0.213\,6759 = 1/f_1$ d. (b) As in (a), but derived assuming a binary period of $0.213\,6765 = 1/f_2$ d. (c) The periodogram of the residuals after pre-whitening by the highest peak in (a). (d) The periodogram of the residuals after pre-whitening by the highest peak in (b). Note the different scales on the different vertical axes.

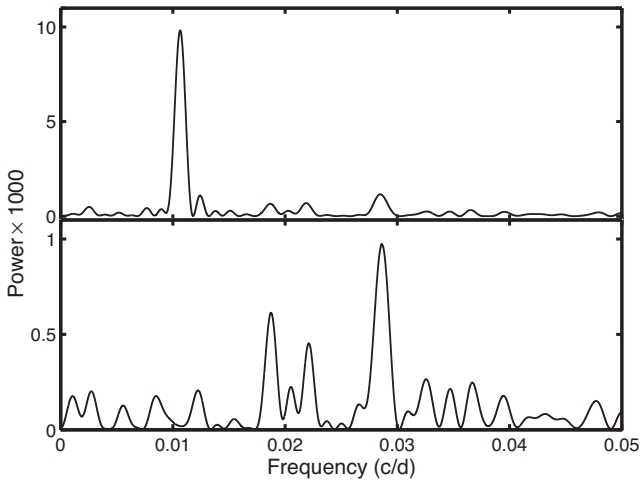

Figure 11. The top panel is the periodogram of Δ_t, estimated from measurements of the δ Scuti star HD 181469 (KIC 4150611). The bottom panel is the periodogram of the residuals left after pre-whitening Δ_t by the highest-peak frequency identified from the spectrum in the top panel.

of the star, in which eclipses had been masked and low-frequency variations removed.

The periodogram of the 5 per cent trimmed Δ_t is in the top panel of Fig. 11: the prominent peak is at a frequency of 0.010 65 d^{-1}. The corresponding period of 93.9 d is close to the known 94.1 orbital period. There is no power excess at the harmonic frequencies, which is confirmed by the periodogram of the residuals after pre-whitening the dominant frequency. This agrees with the conclusion reached by SK2012 that the star's orbit is circular.

Estimates from robust fitting of a sinusoid to the trimmed Δ_t are $P^{(3)} = 94.0$ d and $A^{(3)} = 100.7$ s. Direct fitting of equations (1), (2) and (8) gave $P^{(4)} = 94.05$ d and $A^{(4)} = 99.9$ s. The former value agrees very well with the results in SK2012.

The value of $A^{(4)}$ can be used to roughly estimate the inclination of the binary orbit. Since the LTT amplitude as used in this paper is

$$A = a_1 \sin i/c$$

(where a_1 is the length of the semimajor axis, and i the orbital inclination), $A^{(4)} = 99.9$ s implies $a_1 \sin i = 0.20$ au. [This is in excellent agreement with the entries in table 2 of Shibahashi & Kurtz (2014), which contains corrections of the erroneous values in table 4 of SK2012.] A convenient form of Kepler's law is

$$P^2 = a^3/(m_1 + m_2),$$

with the orbital period P in yr, the mean distance between the stars a in au, and masses in $M_\odot$. Substituting $P = 94$ d,

$$a = 0.4048(m_1 + m_2)^{1/3}$$

is obtained. SK2012 put the mass of the δ Scuti star at $m_1 = 1.7$, while the secondary is itself a binary consisting of two K stars. Assuming a combined mass of 1.2 $M_\odot$ for the K stars, $a = 0.58$ au, which is probably quite accurate given the weak dependence on mass (the corresponding numbers for total system masses of 2 and 3.5 $M_\odot$ are 0.51 and 0.61 au). This leads to $a_1 = 0.24$ au so that $a_1 \sin i = 0.20$ implies $i = 56°$.

Strictly speaking some sort of goodness-of-fit test should be performed for each final model fitted to data. A nod in that direction is shown in Fig. 12, in the form of amplitude spectra of the original HD 181469 data (top panel) and the residuals of the models (1), (2) and (8). Note that the contrast between the two spectra would have been greater if power, rather than amplitude, had been plotted, but in this particular instance the direct physical interpretation of the amplitude scale is useful. The highest peak in the residual spectrum is at a frequency of 20.2145 d (i.e. unrelated to the four frequencies used in the model fitted), and has an amplitude of only 0.16 mmag. All other peaks in the residual spectrum have amplitudes below 0.09 mmag. Clearly, the model represents most of the signal information in the data.

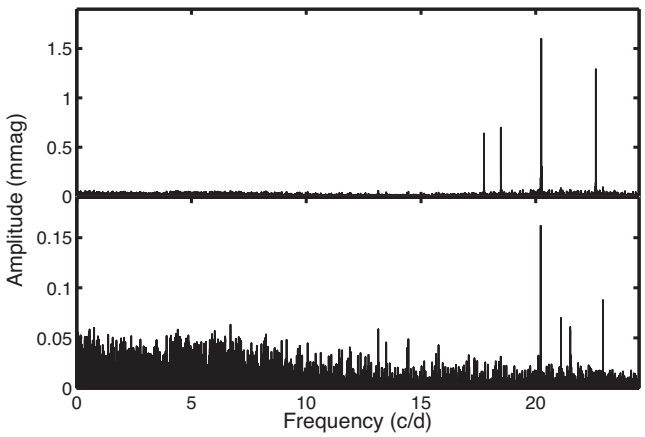

Figure 12. The top panel is an amplitude spectrum of the processed *Kepler* observations of HD 181469. The bottom panel is the spectrum of the residuals left after pre-whitening the final model [equations (1), (2) and (8)] fitted to the data. Note that the scale on the vertical axis in the top panel is a factor of 10 larger than the scale in the bottom panel.

5.3 ASAS 120036-3915.6

As a final exercise, the (O–C) function of the star ASAS 120036-3915.6 is calculated. A total of 630 high-quality ASAS measurements, spread over 9 yr, were used for the purpose. Skelton & Smits (2009) demonstrated a period change in this overcontact binary by partitioning the data and studying the scatter in light curves folded with respect to different periods. They give the overall best-fitting period as 0.292 672 d ($f_0 = 3.416\,79$ d^{-1}) and this figure is used in the analysis which follows.

Examination of the periodogram of the ASAS 120036-3915.6 data shows peaks above the noise level only at the first three even multiples of f_0, i.e. equation (1) is

$$F(\mathbf{\Theta}, t) = \mu + \sum_{k=1}^{3} [A_k \cos(4k\pi f_0 t) + B_k \sin(4k\pi f_0 t)] \qquad (10)$$

in this instance. The periodogram of Δ_t can be seen in Fig. 13: it is dominated by a peak at a frequency 2.3×10^{-4} d^{-1} (period 11.9 yr). Such a low frequency should not be taken literally – it is more likely to be indicative of a long-term trend in Δ_t.

The 5 per cent trimmed values of Δ_t are plotted in Fig. 14, together with a weighted non-parametric regression fit. The fit is locally linear, and only 30 per cent of the closest data are included in the estimate of each fitted value. The data weighting function is of the 'loess' form – see Cleveland & Devlin (1988), and Cleveland, Devlin & Grosse (1988).

The regression function can largely be described as made up of two extended linear sections, separated by the interval HJD 245 3100-3500. This could be interpreted as a fairly rapid change from one constant value of the binary period to another. The total range of

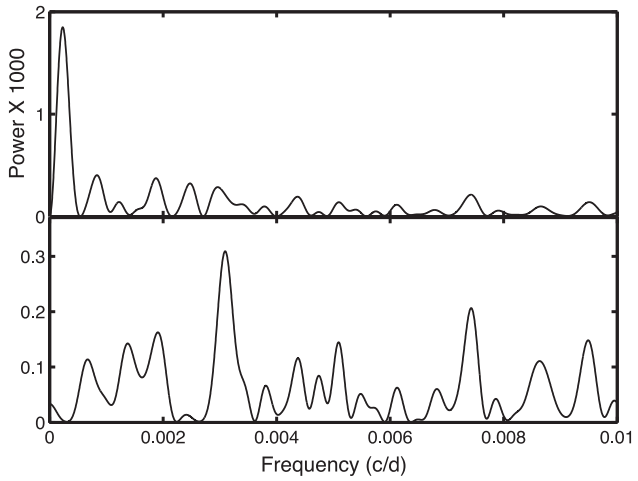

Figure 13. The top panel is the periodogram of Δ_t, estimated from measurements of ASAS 120036-3915.6. The bottom panel is the periodogram of the residuals left after pre-whitening Δ_t by the highest-peak frequency identified from the spectrum in the top panel.

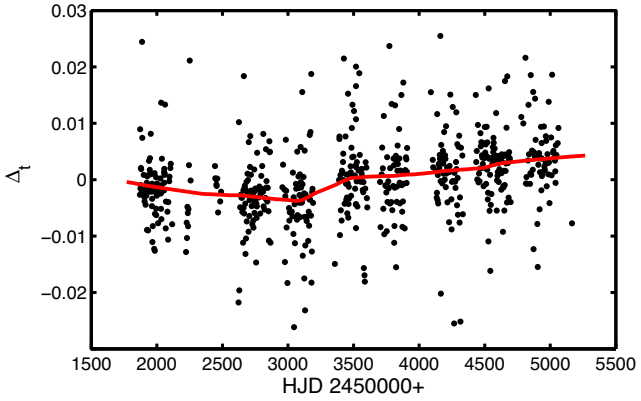

Figure 14. Points represent 5 per cent trimmed values of Δ_t, calculated from observations of ASAS 120036-3915.6, assuming the light curve (10). The line is a non-parametric (loess) locally linear regression function fitted to the points. A window function width of 0.3 was used (see text for details).

the non-parametric estimates of Δ_t over the ∼4100 cycles covering HJD 2451871-3070 is −0.0030 d; over the interval HJD 2453480-5166 there are ∼5760 cycles, and Δ_t has an estimated range of +0.0038 d. This implies misspecifications of the binary period by $+7.1 \times 10^{-7}$ and -6.6×10^{-7} d over the respective intervals, i.e. the period changed from 0.292 6713 to 0.292 6727 d over the course of roughly a year centred on HJD 2453270.

Periodograms were calculated over each of the two intervals HJD 2451871-3070 and HJD 2453480-5166, and the highest-peak frequencies were refined by least squares. The results were best-fitting values of 6.833 611 and 6.833 566 d^{-1}; these correspond to periods of 0.292 6710/2 and 0.292 6729/2, in good agreement with the values estimated from Fig. 14.

It is perhaps worth stressing that, the interpretation of Fig. 14 aside, the study of the ASAS 120036-3915.6 data is fully non-parametric – no specific model for the period change was posited. Instead, the analysis was completely data driven.

6 CONCLUSIONS

The analyses in the previous sections demonstrate that the methodology of Sections 2 and 3 can be usefully applied to the type of large-scale survey data acquired by modest-sized terrestrial telescopes. Furthermore, for higher quality photometric data the techniques also have the potential to reveal time-resolved detail in (O−C) behaviour which has hitherto not been explored.

ACKNOWLEDGEMENTS

The author is grateful for the public availability of All Sky Automated Survey data (Pojmanski 1997), and to Professor Don Kurtz for supplying the processed *Kepler* data analysed in Section 5.2. This research was partially funded by a South African National Research Foundation grant.

REFERENCES

Alcock C. et al., 1999, PASP, 111, 1539
Basri G. et al., 2011, AJ, 141, 20
Beaton A. E., Tukey J. W., 1974, Technometrics, 16, 147
Cleveland W. S., Devlin S. J., 1988, J. Am. Stat. Assoc., 83, 596
Cleveland W. S., Devlin S. J., Grosse E., 1988, J. Econometrics, 37, 87
Holland P. W., Welsch R. E., 1977, Commun. Stat. - Theory Methods, A6, 813
Irwin J. B., 1952, ApJ, 116, 211
Koen C., 1996, MNRAS, 283, 471
Koen C., 2014, MNRAS, 441, 3075
Lohr M. E., Norton A. J., Kolb U. C., Maxted P. F. L., Todd I., West R. G., 2013a, A&A, 549, A86
Lohr M. E., Norton A. J., Kolb U. C., Boyd D. R. S., 2013b, A&A, 558, A71
Lombard F., Koen C., 1993, MNRAS, 263, 309
Pojmanski G., 1997, Acta Astron., 47, 467
Pollacco D. L. et al., 2006, PASP, 118, 1407
Shibahashi H., Kurtz D. W., 2012, MNRAS, 422, 738 (SK2012)
Shibahashi H., Kurtz D. W., 2014, in Shibahashi H., ed., ASP Conf. Ser. Vol. 479, Progress in the Physics of the Sun and Stars. Astron. Soc. Pac., San Francisco, p. 503
Skelton P. A., Smits D. P., 2009, South Afr. J. Sci., 105, 120
Sterken C., ed., 2005, ASP Conf. Ser. Vol. 335, The Light-Time Effect in Astrophysics: Causes and Cures of the O−C Diagram. Astron. Soc. Pac., San Francisco
Sterne T. E., 1934, Harv. Circ. 386
Szymański M. K., 2005, Acta Astron., 55, 43

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