Mon. Not. R. Astron. Soc. **376,** 233–247 (2007) doi:10.1111/j.1365-2966.2007.11425.x

Fitting sinusoids to observations of rotating spotted stars

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

Accepted 2006 December 18. Received 2006 December 18; in original form 2006 August 8

ABSTRACT

It is assumed that K blocks (e.g. seasons) of observations are available, and the parameters characterizing a fixed-frequency sinusoidal variability (mean light level, amplitude, phase) are constant within each of the blocks. The paper is concerned with estimation when any combination of these parameters varies between blocks of observations. This allows observations subject to changes in mean light level, spot sizes and/or spot locations to be modelled. Objective choices between competing models, and the calculation of the standard errors of model parameters, are also dealt with. Illustrative applications to simulated and real data are given.

Key words: methods: statistical – stars: spots.

1 INTRODUCTION

The presence of large dark (or bright) areas on rotating stars can be manifested by sinusoidal modulations of the stellar brightness. Such spots are not necessarily fixed features, but may change in size and/or position. Consequently, the amplitude and/or phase of the luminosity modulation could change over time. A further potential complicating factor is variability in the mean brightness level, due, for example, to the influence of circumstellar material.

It is assumed that the observations consist of K blocks of fairly closely spaced observations. Each block is short enough that the amplitude, phase and mean brightness level may be considered constant for the duration of the block. Individual measurements y_{jk} are then described by the model

$$y_{jk} = \mu_k + c_k \cos(\omega t_{jk} + \phi_k) + e_{jk}, \quad j = 1, 2, \ldots, N_k, \quad k = 1, 2, \ldots, K, \tag{1}$$

where k indexes the particular block of observations, and j indexes the measurements within each block. The amplitude, phase and mean brightness level associated with block k are c_k, ϕ_k and μ_k, respectively. The time points of observation are t_{jk}; e_{jk} are zero-mean measurement errors; and N_k is the number of observations making up block k. The angular frequency $\omega = 2\pi/P$, where P is the rotation period of the star, may also be unknown. For convenience, $t_{1,1} \equiv 0$ will be assumed below, i.e. phase is measured with respect to the time of the first observation.

It is assumed that the errors e_{jk} are uncorrelated and their statistical characteristics (in particular, their variances) are constant with time: these assumptions are not critical, and can be relaxed at the expense of a slightly more involved formulation. Note that the e_{jk} may include sources of variability other than measurement errors.

The paper is concerned with estimation of the unknown parameters and with calculating their approximate standard errors. Relatively simple solutions to the estimation problem are described in Sections 2 and 3: the focus is on reduction of dimensionality from $\sim K$ to the order of unity. This is accomplished by obtaining (effectively) explicit solutions for most parameters in terms of a few remaining unknowns.

Of course, there may be applications where some or all of the μ_k, c_k and ϕ_k do not depend on k, so that the model simplifies. It is therefore also necessary to consider techniques for selecting the appropriate submodel from the general formulation (equation 1). This is dealt with in Section 4. Section 5 contains a brief discussion of the calculation of standard errors of the estimated parameters. A few simulation results are detailed in Section 6, and a 'real-life' example analysed in Section 7. A brief discussion of possible extensions of this work concludes the paper (Section 8).

The eight different models considered in Sections 2 and 3 are numbered to facilitate reference; a summary is provided at the start of Section 6.

2 PARAMETER ESTIMATION IN THE MOST DETAILED CASE (MODEL 1)

The sum of squared errors is

$$Q = \sum_{k=1}^{K}\sum_{j=1}^{N_k}[y_{jk} - \mu_k - c_k\cos(\omega t_{jk} + \phi_k)]^2$$

$$\equiv \sum_{k=1}^{K}\sum_{j=1}^{N_k}[y_{jk} - \mu_k - a_k\cos(\omega t_{jk}) - b_k\sin(\omega t_{jk})]^2, \tag{2}$$

where

$$c_k = \sqrt{a_k^2 + b_k^2} \quad \phi_k = \tan^{-1}\left(-\frac{b_k}{a_k}\right). \tag{3}$$

Use of the transformation from the parameters (c_k, ϕ_k) to (a_k, b_k) can lead to substantial simplifications.

Non-parametric estimates are obtainable from equation (2) by minimizing Q with respect to the unknowns. This can be done by applying brute force global optimization methods, but much more efficient strategies can be devised, as is now demonstrated.

Consider initially the case in which the frequency ω is known. The minimum of the function S is then found by setting the partial derivatives with respect to the $3K$ remaining parameters equal to zero:

$$\frac{\partial Q}{\partial \mu_k} = \frac{\partial Q}{\partial a_k} = \frac{\partial Q}{\partial b_k} = 0 \tag{4}$$

$(k = 1, 2, \ldots, K)$. The result is a set of $3K$ simultaneous linear equations:

$$\sum_{j=1}^{N_k} y_{jk} - N_k\widehat{\mu}_k - \widehat{a}_k\sum_{j=1}^{N_k}\cos(\omega t_{jk}) - \widehat{b}_k\sum_{j=1}^{N_k}\sin(\omega t_{jk}) = 0$$

$$\sum_{j=1}^{N_k} y_{jk}\cos(\omega t_{jk}) - \widehat{\mu}_k\sum_{j=1}^{N_k}\cos(\omega t_{jk}) - \widehat{a}_k\sum_{j=1}^{N_k}\cos^2(\omega t_{jk}) - \widehat{b}_k\sum_{j=1}^{N_k}\cos(\omega t_{jk})\sin(\omega t_{jk}) = 0$$

$$\sum_{j=1}^{N_k} y_{jk}\sin(\omega t_{jk}) - \widehat{\mu}_k\sum_{j=1}^{N_k}\sin(\omega t_{jk}) - \widehat{a}_k\sum_{j=1}^{N_k}\sin(\omega t_{jk})\cos(\omega t_{jk}) - \widehat{b}_k\sum_{j=1}^{N_k}\sin^2(\omega t_{jk}) = 0, \tag{5}$$

where the notation $\widehat{x}$ indicates an estimate of the true value of the parameter x.

Inspection of equation (5) reveals that equations for a given value of k are decoupled from those for other values of k. Solutions for the different blocks of observations can therefore be obtained independently. In matrix form

$$\begin{bmatrix} N_k & C_k & S_k \\ C_k & CC_k & CS_k \\ S_k & SC_k & SS_k \end{bmatrix}\begin{bmatrix} \widehat{\mu}_k \\ \widehat{a}_k \\ \widehat{b}_k \end{bmatrix} = \begin{bmatrix} Y_k \\ YC_k \\ YS_k \end{bmatrix}, \tag{6}$$

where

$$C_k = \sum_{j=1}^{N_k}\cos(\omega t_{jk}), \qquad CC_k = \sum_{j=1}^{N_k}\cos^2(\omega t_{jk}), \qquad YC_k = \sum_{j=1}^{N_k} y_{jk}\cos(\omega t_{jk}),$$

$$S_k = \sum_{j=1}^{N_k}\sin(\omega t_{jk}), \qquad SS_k = \sum_{j=1}^{N_k}\sin^2(\omega t_{jk}), \qquad YS_k = \sum_{j=1}^{N_k} y_{jk}\sin(\omega t_{jk}),$$

$$Y_k = \sum_{j=1}^{N_k} y_{jk}, \qquad CS_k = SC_k = \sum_{j=1}^{N_k}\cos(\omega t_{jk})\sin(\omega t_{jk}). \tag{7}$$

Equation (6) can be written in the conveniently concise form

$$\mathbf{M}_k\boldsymbol{z}_k = \boldsymbol{d}_k, \tag{8}$$

where the entries in the matrices $\mathbf{M}_k$ and vectors $\boldsymbol{d}_k$ are known (for known ω), and $\boldsymbol{z}_k$ is a vector of unknowns.

Note that the simple linear form (equation 8) is a consequence of the transformation from (c_k, ϕ_k) to (a_k, b_k). Its solution is

$$\boldsymbol{z}_k = \mathbf{M}_k^{-1}\boldsymbol{d}_k. \tag{9}$$

The estimates $\boldsymbol{z}_k$ for the different k can be substituted into equation (2) to obtain the value of $Q_* = Q_*(\omega)$, i.e. the sum of squared errors is then a function of ω only. An optimal estimate of ω follows by minimizing Q_* with respect to ω.

In summary:

(i) choose a realistic range $[\omega_1, \omega_2]$ within which ω is expected to lie;
(ii) for each value of ω in this range, estimate (μ_k, a_k, b_k) $(k = 1, 2, \ldots, K)$ using equation (9);
(iii) substitute $(\widehat{\mu}_k, \widehat{a}_k, \widehat{b}_k)$ $(k = 1, 2, \ldots, K)$ from the previous step into the expression for the sum of squares (equation 2) to obtain $Q_* = Q_*(\omega)$;
(iv) it is convenient to plot Q_* against ω – this is a 'least-squares periodogram', from which the best estimate $\widehat{\omega}$ is easily determined.

We proceed to examine a number of cases in which some of the parameters are constant over all blocks, i.e. independent of k. Paradoxically, solutions are usually more difficult than in the case of the full complement of $3K + 1$ unknown parameters above.

3 SPECIAL CASES

3.1 Constant mean light level (Model 2)

In the event that $\mu_1 = \mu_2 = \cdots = \mu_K = \mu$, the condition

$$\frac{\partial Q}{\partial \mu} = 0$$

leads to the single equation

$$\sum_{k=1}^{K}\sum_{j=1}^{N_k} y_{jk} - N\widehat{\mu}_k - \sum_{k=1}^{K}\widehat{a}_k \sum_{j=1}^{N_k}\cos(\omega t_{jk}) - \sum_{k=1}^{K}\widehat{b}_k \sum_{j=1}^{N_k}\sin(\omega t_{jk}) = 0 \tag{10}$$

($N \equiv \sum_k N_k$) which replaces the K equations of the first form in equation (5).

If there is a sufficiently large number of observations, then the sums over sinusoids in equation (10) will be small compared to the other terms and

$$\widehat{\mu} \approx \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{N_k} y_{jk} \equiv \overline{y} \tag{11}$$

which is certainly a reasonable result. If this estimate is substituted into the remaining two equations in (5), two simultaneous equations for the two unknowns $\widehat{a}_k$ and $\widehat{b}_k$ result.

For smaller numbers of observations, the K blocks are no longer decoupled as in the case discussed in Section 2: instead, there are $2K + 1$ simultaneous equations which can be written in the matrix form as

$$\begin{bmatrix}
N & C_1 & C_2 & C_3 & \cdots & C_K & S_1 & S_2 & S_3 & \cdots & S_K \\
C_1 & CC_1 & 0 & 0 & \cdots & 0 & CS_1 & 0 & 0 & \cdots & 0 \\
C_2 & 0 & CC_2 & 0 & \cdots & 0 & 0 & CS_2 & 0 & \cdots & 0 \\
\vdots & \vdots & 0 & \ddots & & \vdots & \vdots & 0 & \ddots & & 0 \\
C_K & 0 & \vdots & 0 & \cdots & CC_K & 0 & \vdots & 0 & \cdots & CS_K \\
S_1 & SC_1 & 0 & 0 & \cdots & 0 & SS_1 & 0 & 0 & \cdots & 0 \\
S_2 & 0 & SC_2 & 0 & \cdots & 0 & 0 & SS_2 & 0 & \cdots & 0 \\
\vdots & \vdots & 0 & \ddots & & \vdots & \vdots & 0 & \ddots & & 0 \\
S_K & 0 & \vdots & 0 & \cdots & SC_K & 0 & \vdots & 0 & \cdots & SS_K
\end{bmatrix}
\begin{bmatrix}
\widehat{\mu} \\ \widehat{a}_1 \\ \vdots \\ \widehat{a}_K \\ \widehat{b}_1 \\ \vdots \\ \widehat{b}_K
\end{bmatrix}
=
\begin{bmatrix}
\sum_k Y_k \\ YC_1 \\ \vdots \\ YC_K \\ YS_1 \\ \vdots \\ YS_K
\end{bmatrix}. \tag{12}$$

The structure of the square matrix in equation (12) is simple: it is

$$\mathbf{M} = \begin{bmatrix} N & \boldsymbol{C}^{\mathrm{t}} & \boldsymbol{S}^{\mathrm{t}} \\ \boldsymbol{C} & \mathrm{diag}(CC) & \mathrm{diag}(CS) \\ \boldsymbol{S} & \mathrm{diag}(SC) & \mathrm{diag}(SS) \end{bmatrix}, \tag{13}$$

where $\boldsymbol{C}$ and $\boldsymbol{S}$ are column vectors with entries C_k and S_k; and diag (CC), diag (SS) and diag (SC) = diag (CS) are diagonal matrices with elements CC_k, SS_k and $CS_k \equiv SC_k$, respectively. Equation (12) is of the same form as equation (8):

$$\mathbf{M}\boldsymbol{z} = \boldsymbol{d}, \tag{14}$$

with solution

$$\boldsymbol{z} = \mathbf{M}^{-1}\boldsymbol{d}. \tag{15}$$

Given ω, equation (15) provides simultaneous estimates for all $2K + 1$ unknowns. As described in Section 2, these can be substituted into Q to give $Q_*(\omega)$. Estimation can then proceed exactly as for the case dealt with in Section 2, i.e. the recipe at the end of that section can be followed.

3.2 Constant amplitude and phase (Model 3)

If $c_1 = c_2 = \cdots = c_K = c$ and $\phi_1 = \phi_2 = \cdots = \phi_K = \phi$, then

$$\begin{bmatrix} \mathrm{diag}(N_k) & \boldsymbol{C} & \boldsymbol{S} \\ \boldsymbol{C}^{\mathrm{t}} & \sum_k CC_k & \sum_k CS \\ \boldsymbol{S}^{\mathrm{t}} & \sum_k SC_k & \sum_k SS \end{bmatrix} \begin{bmatrix} \widehat{\mu}_1 \\ \vdots \\ \widehat{\mu}_K \\ \widehat{a} \\ \widehat{b} \end{bmatrix} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_K \\ \sum_k YC_k \\ \sum_k YS_k \end{bmatrix}. \tag{16}$$

This is of the same form as equation (14), with a solution similar to equation (15).

3.3 Constant amplitude, phase and mean luminosity (Model 4)

The case $\mu_k = \mu$, $a_k = a$, $b_k = b$, for all k, is tantamount to having a single block of $N = N_1 + N_2 + \cdots + N_K$ observations. For given ω, estimates of the three parameters of interest follow from the equation

$$\begin{bmatrix} N & \sum_k C_k & \sum_k S_k \\ \sum_k C_k & \sum_k CC_k & \sum_k CS_k \\ \sum_k S_k & \sum_k SC_k & \sum_k SS_k \end{bmatrix} \begin{bmatrix} \widehat{\mu} \\ \widehat{a} \\ \widehat{b} \end{bmatrix} = \begin{bmatrix} \sum_k Y_k \\ \sum_k YC_k \\ \sum_k YS_k \end{bmatrix} \tag{17}$$

[cf. equation (6)].

3.4 Constant phase (Model 5)

The case $\phi_1 = \phi_2 = \cdots = \phi_K = \phi$ is dealt with in this section. The transformation $(c_k, \phi_k) \rightarrow (a_k, b_k)$ is a hindrance rather than a help if the phase is constant, hence we revert to the first form in equation (2). Assume initially that both ω and ϕ are given, i.e. parameters to be estimated are the $2K$ unknowns $\mu_1, \mu_2, \ldots, \mu_K$ and $c_1, c_2, \ldots, c_K$. Proceeding as for the two previous cases, the matrix equation

$$\begin{bmatrix} \mathrm{diag}(N_k) & \mathrm{diag}\left(C^{(1)}\right) \\ \mathrm{diag}\left(C^{(1)}\right) & \mathrm{diag}\left(CC^{(1)}\right) \end{bmatrix} \begin{bmatrix} \widehat{\mu}_1 \\ \vdots \\ \widehat{\mu}_K \\ \widehat{c}_1 \\ \vdots \\ \widehat{c}_K \end{bmatrix} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_K \\ YC_1^{(1)} \\ \vdots \\ YC_K^{(1)} \end{bmatrix} \tag{18}$$

is obtained, where the entries in the diagonal matrices are N_k and

$$C_k^{(1)} = \sum_{j=1}^{N_k} \cos(\omega t_{jk} + \phi), \quad CC_k^{(1)} = \sum_{j=1}^{N_k} \cos^2(\omega t_{jk} + \phi), \quad YC_k^{(1)} = \sum_{j=1}^{N_k} y_{jk} \cos(\omega t_{jk} + \phi). \tag{19}$$

Before giving a recipe for the full solution, two practical issues need to be pointed out. First, in order to obtain a unique solution for ϕ, it is necessary to impose a restriction on its domain: we use $\phi \in [-\pi, \pi]$. (The same applies to the remainder of the special cases dealt with below.) Secondly, all estimated amplitudes should be either positive or negative. In the latter case, all amplitudes are multiplied by -1, and either $+\pi$ or $-\pi$ added to the phase (such that the final value lies in the interval $[-\pi, \pi]$). Solutions with a mixture of positive and negative amplitudes are not physically realistic: simulation experiments show that these are obtained when the phase is in fact variable.

In general, for unknown ω and ϕ, we then proceed as follows.

(i) Choose a realistic range $[\omega_1, \omega_2]$ within which ω is expected to lie.

(ii) For each value of ω in this range, assume a starting value for ϕ. This is conveniently obtained by fitting the constant amplitude and constant phase model (Section 3.2) to the data.

(iii) Estimate (μ_k, c_k) $(k = 1, 2, \ldots, K)$ using equation (18).

(iv) Substitute the estimates from the previous step into the expression for the sum of squares (equation 2) to obtain $Q_{**} = Q_{**}(\omega, \phi)$.

(v) Repeat steps (iii) and (iv) for different ϕ, obtaining

$$Q_*(\omega) = \max_{\phi} Q_{**}(\omega, \phi).$$

(vi) Repeat steps (iii)–(v) for all ω.

(vii) Obtain $\widehat{\omega}$ as the frequency at which the least-squares periodogram $Q_*(\omega)$ reaches a minimum.
(viii) Ensure that all $\widehat{c}_k \leqslant 0$. If there is a mixture of positive and negative amplitudes, then the model is not appropriate.

3.5 Constant mean and phase (Model 6)

If $\mu_1 = \mu_2 = \cdots = \mu_K = \mu$ and $\phi_1 = \phi_2 = \cdots = \phi_K = \phi$, equation (18) reduces to

$$\begin{bmatrix} N & [\boldsymbol{C}^{(1)}]^{\mathrm{t}} \\ \boldsymbol{C}^{(1)} & \mathrm{diag}\left(CC^{(1)}\right) \end{bmatrix} \begin{bmatrix} \widehat{\mu} \\ \widehat{c}_1 \\ \vdots \\ \widehat{c}_K \end{bmatrix} = \begin{bmatrix} \sum_k Y_k \\ YC_1^{(1)} \\ \vdots \\ YC_K^{(1)} \end{bmatrix}. \tag{20}$$

The solution method is otherwise unchanged from that outlined in Section 3.4. As for the model of Section 3.4, a mixture of positive and negative amplitudes is obtained for simulated data in which the phase is variable.

3.6 Constant amplitude (Model 7)

If $c_1 = c_2 = \cdots = c_K = c$, the sum of squared errors is

$$Q = \sum_{k=1}^{K} \sum_{j=1}^{N_k} [y_{jk} - \mu_k - c\cos(\omega t_{jk} + \phi_k)]^2. \tag{21}$$

Setting partial derivatives with respect to $\mu_k (k = 1, 2, \ldots, K)$ and c equals to zero,

$$\begin{bmatrix} \mathrm{diag}(N_k) & \boldsymbol{C}^{(2)} \\ \boldsymbol{C}^{(2)\mathrm{t}} & \sum_k CC_k^{(2)} \end{bmatrix} \begin{bmatrix} \widehat{\mu}_1 \\ \vdots \\ \widehat{\mu}_K \\ \widehat{c} \end{bmatrix} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_K \\ \sum_k YC_k^{(2)} \end{bmatrix}, \tag{22}$$

where

$$C_k^{(2)} = \sum_{j=1}^{N_k} \cos(\omega t_{jk} + \phi_k), \quad CC_k^{(2)} = \sum_{j=1}^{N_k} \cos^2(\omega t_{jk} + \phi_k), \quad YC_k^{(2)} = \sum_{j=1}^{N_k} y_{jk}\cos(\omega t_{jk} + \phi_k). \tag{23}$$

Solution proceeds as for the case of constant phase (Section 3.4), but steps (ii)–(v) involve maximizing over K phase values instead of one. In step (ii), good starting values for the phases can be obtained by fitting the all-parameters-variable model (Section 2) to the data.

Since there is only one amplitude, the problem of unphysical solutions (Sections 3.4 and 3.5) does not arise: negative $\widehat{c}$ is simply multiplied by -1, with accompanying transformation of all phases.

3.7 Constant mean and amplitude (Model 8)

This is a special case of the situation dealt with in the previous section: now $\mu_1 = \mu_2 = \cdots = \mu_K = \mu$. Equation (22) reduces to

$$\begin{bmatrix} N & \sum_k C_k^{(2)} \\ \sum_k C_k^{(2)} & \sum_k CC_k^{(2)} \end{bmatrix} \begin{bmatrix} \widehat{\mu} \\ \widehat{c} \end{bmatrix} = \begin{bmatrix} \sum_k Y_k \\ \sum_k YC_k^{(2)} \end{bmatrix}. \tag{24}$$

If $\widehat{c} < 0$, it should be multiplied by -1, and all phases transformed accordingly.

4 WHICH MODEL IS APPROPRIATE?

The simplest, or null model, is that given in Section 3.3. If a specific alternative (H_1) to the null (H_0) is considered, then a hypothesis test can be performed. This can be done by comparing the Gaussian log likelihoods

$$\log L = -\frac{N}{2}\log 2\pi - N\log\sigma - \frac{1}{2\sigma^2}Q \tag{25}$$

calculated, respectively, under the null and alternative hypotheses. If the error variance σ^2 is known, then the likelihood ratio statistic is

$$\begin{aligned} \Lambda &= 2\{\max[\log L(H_1)] - \max[\log L(H_0)]\} \\ &= \frac{1}{\sigma^2}\{\min[Q(H_0)] - \min[Q(H_1)]\} \\ &\equiv \frac{1}{\sigma^2}[Q_0 - Q_1]. \end{aligned} \tag{26}$$

If σ^2 is unknown, it can be estimated by

$$\widehat{\sigma^2} = \frac{1}{N} Q; \tag{27}$$

substitution into equation (25) then gives

$$\log L \approx -\frac{N}{2}[\log 2\pi + \log \widehat{\sigma^2} + 1], \tag{28}$$

and the likelihood ratio statistic reduces to

$$\Lambda \approx N[\log Q_0 - \log Q_1]. \tag{29}$$

Inspection of equations (26) and (29) shows that Λ compares the values of Q obtained under the null and under the alternative hypotheses.

The likelihood ratio statistic has an asymptotic chi-squared distribution with degrees of freedom equal to the number of parameter constraints required by H_0 over that required by H_1. For example, in the models

$$\begin{aligned} H_0: &\quad \mu_k = \mu,\ c_k = c, \quad \phi_k = \phi, \qquad k = 1, 2, \ldots, K, \\ H_1: &\quad \mu_k = \mu,\ c_k = c, \quad \phi_k \text{ unspecified}, \quad k = 1, 2, \ldots, K, \end{aligned} \tag{30}$$

H_0 requires the $K - 1$ specifications

$$\phi_2 = \phi_1,\ \phi_3 = \phi_1, \ldots, \phi_k = \phi_1$$

over and above the restrictions imposed by H_1: in this case $\Lambda \sim \chi^2(K - 1)$.

In general, though, hypothesis testing cannot easily be used to choose between models, as they may be non-nested, i.e. one model is not necessarily a special case of the other. (To name but one example, the models of Sections 3.4 and 3.6 are non-nested.) Furthermore, it may be desirable to consider an array of models and to select one which is, in some sense, optimal. We turn therefore to information criteria: these are generally two-term expressions, one of which reflects the goodness-of-fit of the model to the data, while the other is a penalty term for the number of parameters needed to accomplish the fit. The two specific forms used in the paper are

$$\begin{aligned} \text{AIC} &= N \log Q + 2M + \frac{2M(M+1)}{N - M - 1}, \\ \text{BIC} &= N \log Q + M \log N, \end{aligned} \tag{31}$$

the Akaike and Bayes information criteria. The goodness-of-fit is described by Q, while the terms proportional to M, the number of fitted parameters, are the 'penalty'. The last term in the expression for the AIC is a bias correction (Burnham & Anderson 2004). Since small values of both Q and M are desirable, the 'best' model is that which minimizes the information criterion.

The numerical values of the information criteria are not amenable to direct interpretation, but

$$\begin{aligned} p_{\text{A}}(j) &= \frac{1}{W_{\text{A}}} \exp -\frac{1}{2}[\text{AIC}(j) - \text{AIC(min)}], \\ p_{\text{B}}(j) &= \frac{1}{W_{\text{B}}} \exp -\frac{1}{2}[\text{BIC}(j) - \text{BIC(min)}], \end{aligned} \tag{32}$$

can be interpreted as the probability of model j. The normalizing constants W_{A} and W_{B} in equation (32) are chosen such that $\sum_j p_{\text{A}}(j) = \sum_j p_{\text{B}}(j) = 1$.

5 STANDARD ERRORS OF THE ESTIMATES

Bootstrapping is a general computer-intensive method for obtaining standard errors and/or confidence intervals for estimated parameters (e.g. Efron & Tibshirani 1993; Davison & Hinkley 1997). The essence of the method is the production of many (typically several thousand) synthetic data sets which resemble the original observations. The same estimation procedure is carried out for each of the synthetic data sets, and the standard deviations of the estimates are used to approximate the required standard errors.

In the present context, the synthetic data sets are constructed by (i) fitting the model to the observations; (ii) estimating the N residuals by subtracting the fitted model from the observations; (iii) drawing N pseudo-residuals from the estimated residuals; and finally (iv) adding the pseudo-residuals to the fitted model. A distinguishing feature of the bootstrapping algorithm is that the pseudo-residuals are drawn from the original residuals with replacement, i.e. repeated values are allowed.

6 EXAMPLE ANALYSES OF SIMULATED DATA

For convenience, the different models are summarized in Table 1.

The simulation model was designed to resemble the type of data which may be encountered when observing very low mass stars with rapidly (time-scales of hours or days) evolving periodicities (e.g. Koen 2006). The model has $K = 3$, with $N_1 = 60$, $N_2 = 73$, $N_3 = 53$. The starting times for the blocks are 0.013, 2.054 and 3.039; the time interval between successive observations is 0.004 units. The frequency is 15 cycles per time unit. In the constant-parameter model amplitudes are unity, phases -2 and mean levels 7. The (Gaussian) error standard deviation is 0.3 units, i.e. the signal-to-noise ratio is 3.3. A typical simulated data set is plotted in Fig. 1.

Table 1. A summary of the models discussed in Sections 2 and 3

Model	Variable parameters
1	Level, amplitude, phase
2	Amplitude, phase
3	Level
4	None
5	Level, amplitude
6	Amplitude
7	Level, phase
8	Phase

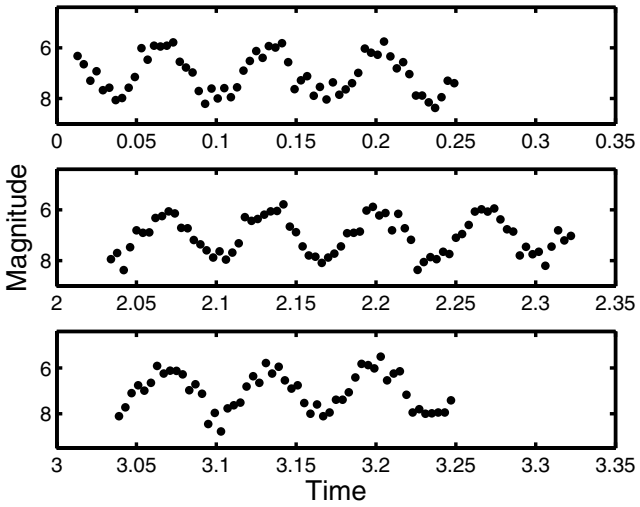

Figure 1. The basic synthetic data – see the text for a description of the parameters used in the simulation.

The sums of squares $Q_*(f)$, minimized over all parameters except the frequency f, are plotted in Figs 2–4 for various models. The most striking feature is the marked difference in frequency resolution between Models 1, 2, 7 and 8 (Fig. 2) on the one hand, and the remainder of the models on the other. Inspection of Table 1 shows that the distinguishing feature is whether the phase is considered variable or fixed.

This result may seem surprising at a first glance. However, a little thought shows that if the phases of the different blocks can be adjusted independently their combination does not provide any improvement in the frequency resolution. Instead, it is effectively only required that the frequency fits the data from each block reasonably well. The situation is therefore akin to having several replications of a time-series –

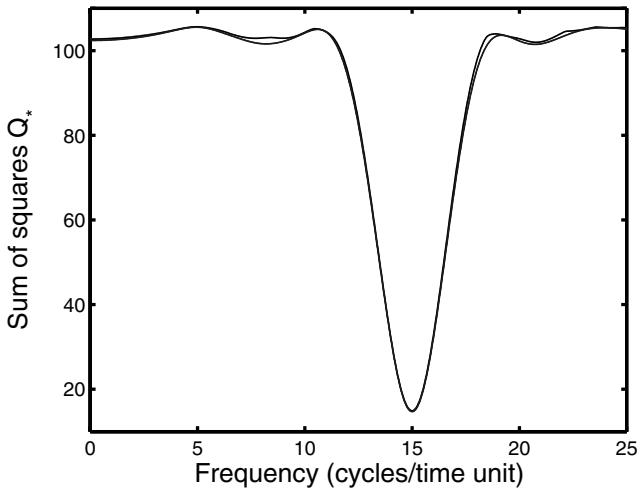

Figure 2. Least-squares spectra of the basic simulated data, plotted for the Models (1, 2, 7, 8) which allow for phase changes.

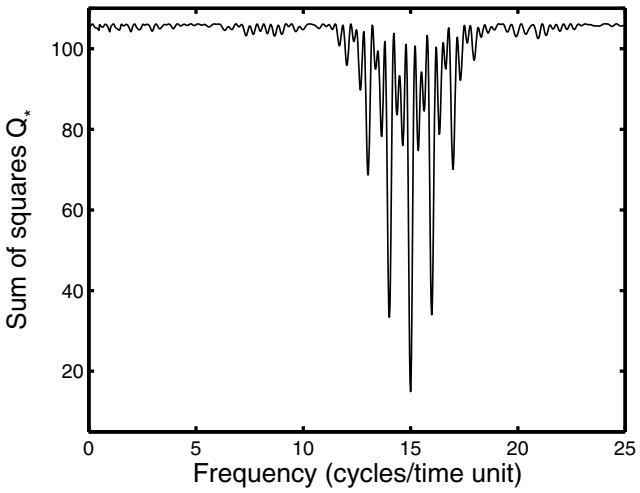

Figure 3. Least-squares spectrum of the basic simulated data for Model 3. The spectrum for Model 4 is virtually identical. Both models assume that the amplitudes and phases are constant.

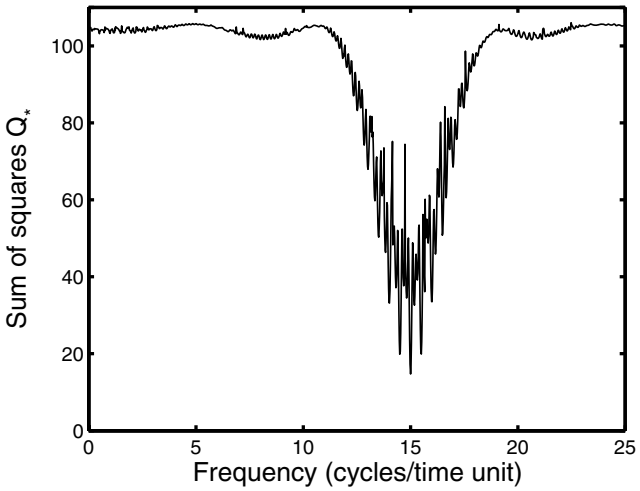

Figure 4. Least-squares spectrum of the basic simulated data for Model 5. The spectrum for Model 6 is virtually identical. Both models assume that the phases are constant, but allow for variability in the amplitude.

the improvement in the accuracy with which the frequency can be estimated is determined by the lengths and numbers of the blocks, not by the total time interval spanned.

All eight models in Table 1 were fitted to the data. Parameter estimates, and other relevant information, are given in Table 2. Both the AIC and the BIC assign the highest probability to Model 4, which is correct. The parameter estimates are also all very close to the true values.

Standard errors of the parameter estimates, calculated according to the prescription in Section 5, are given in Table 3. It can be seen that the upper and lower boundaries of the 95 per cent confidence intervals are close to ± 2 standard errors from the means of the 5000 bootstrap samples. Also, the mean values are close to the estimates in Table 2, suggesting that the estimation method does not suffer from any notable bias.

Simulation experiments were conducted to test the efficacy of the information criteria. Seven variations on the model described above were created by changing various combinations of $\mu_2(7 \rightarrow 6.6)$, c_2 $(1 \rightarrow 1.4)$ and $\phi_2(-2 \rightarrow -1)$. In all cases but two, the BIC model probabilities of the correct model are 0.97 or larger; the corresponding AIC probabilities range from 0.55 to 0.89. The two exceptions are the data sets defined by $(c_2 = 1.4, \phi_2 = -1)$ and $(\mu_2 = 6.6, c_2 = 1.4, \phi_2 = -1)$. Details of the estimated models for the latter data set are given in Table 4; the corresponding least-squares spectra are in Figs 5–7. Both the BIC and AIC select the correct model, but the associated probabilities are unusually low for the BIC, and high for the AIC. Inspection of the Table shows that the low value of $p_{\rm B}$ for Model 1 can be traced to competition from Model 5, which has $p_{\rm B} = 0.36$. However, Models 5 and 6 should be disregarded, since the estimated values of c_1 are negative. If these are discounted, then $p_{\rm B} \approx 1$ for Model 1. The high value of $p_{\rm A}$ for Model 1 has an entirely different origin: it is well known that the AIC is generally more generous than the BIC as regards model complexity, i.e. it allows for more model parameters. In Table 4, Model 1 is not only the most complex, but also the correct one, hence its high probability according to the AIC.

Table 2. Estimated parameters for each of the eight models fitted to the simulated data plotted in Fig. 1. The number of parameters M associated with each model is given in the second line of the table; $p_{\rm B}$ and $p_{\rm A}$ are the model probabilities according to the Bayes and Akaike criteria, respectively; f is the best-fitting frequency; and $\sigma_{\rm e}$ is the standard deviation of the measurement error.

Model	1	2	3	4	5	6	7	8
M	10	8	6	4	8	6	8	6
$p_{\rm A}$	0.00	0.03	0.07	0.56	0.03	0.24	0.01	0.07
$p_{\rm B}$	0.00	0.00	0.01	0.97	0.00	0.02	0.00	0.01
f	15.00	15.01	15.00	15.00	15.00	15.00	15.00	15.00
μ_1	6.97	6.99	6.97	6.99	6.97	6.99	6.97	6.99
μ_2	6.99	6.99	6.99	6.99	6.99	6.99	6.99	6.99
μ_3	6.99	6.99	6.99	6.99	6.99	6.99	6.99	6.99
c_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
c_2	0.95	0.95	1.00	1.00	0.95	0.95	1.00	1.00
c_3	1.07	1.07	1.00	1.00	1.06	1.07	1.00	1.00
ϕ_1	−2.04	−2.04	−2.04	−2.04	−2.04	−2.04	−2.04	−2.04
ϕ_2	−2.09	−2.12	−2.04	−2.04	−2.04	−2.04	−2.04	−2.07
ϕ_3	−2.08	−2.13	−2.04	−2.04	−2.04	−2.04	−2.01	−2.05
$\sigma_{\rm e}$	0.29	0.29	0.29	0.29	0.29	0.29	0.29	0.29

Table 3. The results of bootstrapping the optimal model from Table 2.

Parameter	Mean	95 per cent confidence interval	Standard error
f	15.0011	(14.9933, 15.0087)	0.0039
μ	6.986	(6.945, 7.027)	0.021
c	0.999	(0.942, 1.059)	0.030
ϕ	−2.042	(−2.145, −1.936)	0.053
$\sigma_{\rm e}$	0.283	(0.252, 0.314)	0.016

Table 4. Same as Table 2, but for the data set containing changes in mean level, amplitude and phase in the second block of observations. Note that the solutions for Models 5 and 6 are unphysical, and should therefore be disregarded.

Model	1	2	3	4	5	6	7	8
M	10	8	6	4	8	6	8	6
$p_{\rm A}$	0.64	0.00	0.00	0.00	0.36	0.00	0.00	0.00
$p_{\rm B}$	0.97	0.00	0.00	0.00	0.03	0.00	0.00	0.00
f	15.04	15.05	14.65	14.65	14.84	14.84	15.06	15.06
μ_1	6.97	6.83	6.97	6.83	6.97	6.83	6.95	6.82
μ_2	6.60	6.83	6.60	6.83	6.59	6.83	6.59	6.82
μ_3	6.99	6.83	7.00	6.83	7.00	6.83	6.98	6.82
c_1	1.00	1.02	1.06	1.07	−1.00	−1.02	1.22	1.24
c_2	1.51	1.53	1.06	1.07	1.51	1.52	1.22	1.24
c_3	1.07	1.08	1.06	1.07	1.06	1.07	1.22	1.24
ϕ_1	−2.07	−2.07	−2.06	−2.04	1.22	1.24	−2.07	−2.07
ϕ_2	−1.58	−1.66	−2.06	−2.04	1.22	1.24	−1.77	−1.77
ϕ_3	−2.87	−3.03	−2.06	−2.04	1.22	1.24	3.14	3.10
$\sigma_{\rm e}$	0.38	0.42	0.59	0.61	0.39	0.43	0.41	0.45

Results for the data set with amplitude and phase changes in the second block of observations ($c_2 = 1.4, \phi_2 = -1$) were comparable to those given in Tables 3 and 4. Model 2 was (correctly) selected by both information criteria ($p_{\rm A} = 0.87, p_{\rm B} = 0.65$). The nearest rival of Model 2, namely Model 6 ($p_{\rm B} = 0.34$), again had a negative amplitude solution.

The reader's attention is drawn to the apparently very poor frequency solutions obtained when fitting Models 3 and 4 to the data. The tabled results are somewhat deceptive, as these are due to the marginal selection of the incorrect alias – there are secondary minima at $f = 15.02$ in both instances.

Given that the optimal model in Table 4 has 10 parameters, whereas the best model in Table 2 only has four, it is hardly surprising that the standard errors of the estimated parameters in the former are generally about 2–3 times larger (compare Tables 5 and 3). None the less, the errors in the frequency, and in ϕ_2 and ϕ_3, seem exaggerated. The large frequency error is in fact due to the poor frequency resolution

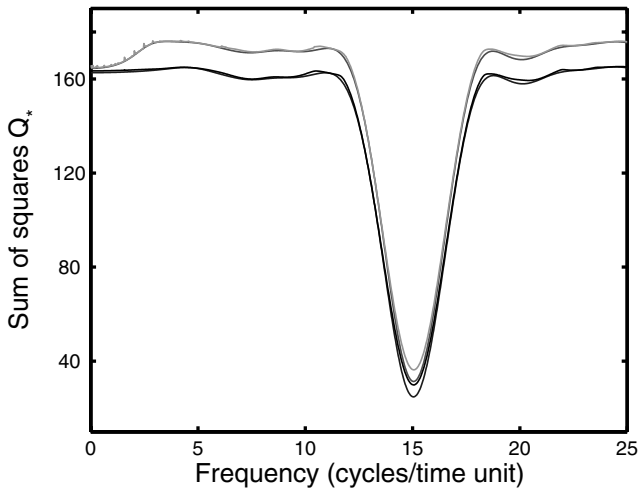

Figure 5. Least-squares spectra of the simulated data with variable parameters. The figure shows the results for Models 1, 7 (bottom two lines, blue and black) and 2, 8 (top two lines, red and green). The ragged appearance of the Model 8 spectrum at very low frequencies is due to the slow convergence of the phase solutions in that regime.

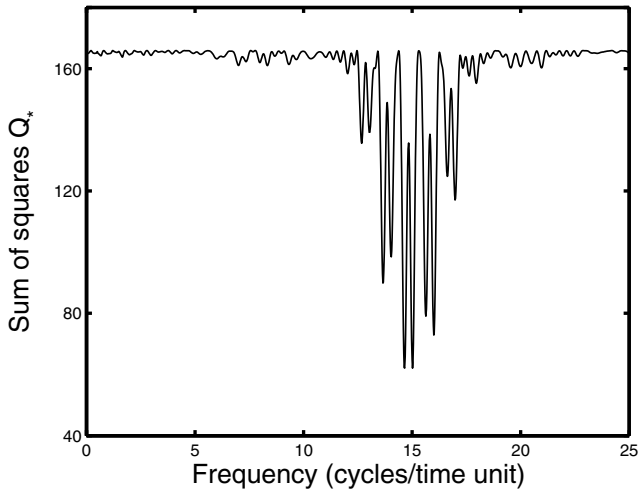

Figure 6. Least-squares spectrum of the simulated data with variable parameters for Model 3. The Model 4 spectrum is very similar, but is offset to slightly (∼10 units) larger sums of squares.

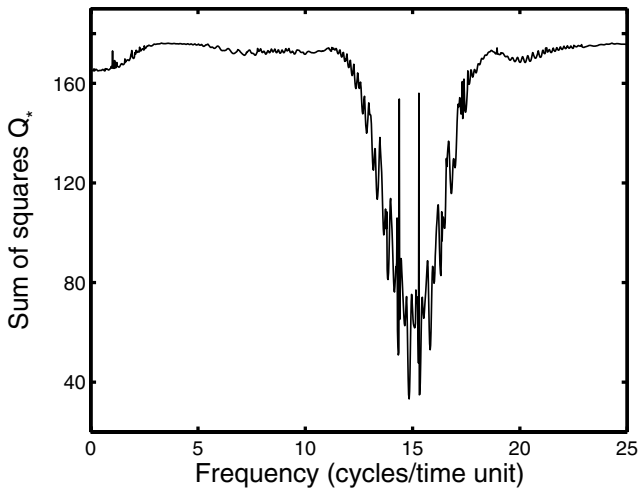

Figure 7. Least-squares spectrum of the simulated data with variable parameters, for Model 6. The Model 5 spectrum is similar over most of the frequency range, but is offset to lower sums of squares (6–11 units). Note though that the Model 5 solution is unphysical.

Table 5. The results of bootstrapping the optimal model from Table 4. Alternative confidence intervals and standard errors for ϕ_2 and ϕ_3, after correction for ambiguity in the definition of the phase, are also given – see the text for further discussion.

Parameter	Mean	95 per cent confidence interval	Standard error
f	15.044	(14.896, 15.190)	0.075
μ_1	6.972	(6.857, 7.082)	0.058
μ_2	6.595	(6.493, 6.696)	0.052
μ_3	6.991	(6.871, 7.107)	0.060
c_1	1.006	(0.847, 1.168)	0.082
c_2	1.515	(1.370, 1.659)	0.074
c_3	1.067	(0.905, 1.236)	0.085
ϕ_1	−2.067	(−2.26, −1.88)	0.097
ϕ_2	−1.184	(−2.98, 2.91)	1.34
		(−2.06, 2.00)	1.01
ϕ_3	−0.200	(−3.05, 3.05)	2.13
		(−2.71, 2.76)	1.39
σ_e	0.441	(0.390, 0.493)	0.027

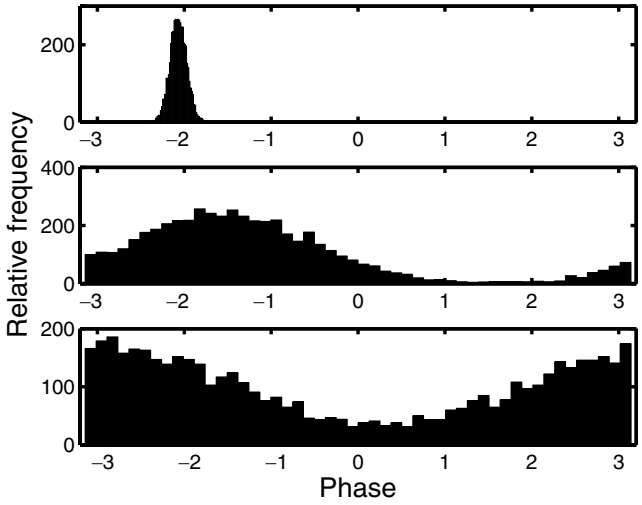

Figure 8. The distributions of the estimated phases for the optimal model (number 1) fitted to the simulated data with all parameters variable. Distributions of $\widehat{\phi}_1$, $\widehat{\phi}_2$ and $\widehat{\phi}_3$ are shown from top to bottom.

engendered by the phase change (see the discussion of Figs 2–4). In the case of the phase errors, there are two contributory factors, which are now discussed.

The distributions of the three sets of bootstrap phase estimates can be seen in Fig. 8. The spread in ϕ_2 is much larger than that in ϕ_1, and the spread in ϕ_3 is even more. Clearly, this is to some extent due to the ambiguity in the definition of the phase: Fig. 9 shows the scatter with respect to the parameter estimates ($\widehat{\phi}_2 = -1.58$, $\widehat{\phi}_3 = -2.87$, respectively), again restricted to the interval $[-\pi, \pi]$. The standard deviations are 1.01 and 1.39 rad, respectively; these constitute more realistic error estimates than the directly calculated standard errors. The corresponding 95 per cent confidence intervals are also narrower – see Table 5.

Although redefining the phases improved the standard errors of $\widehat{\phi}_2$ and $\widehat{\phi}_3$, the values are still considerably larger than the corresponding value for $\widehat{\phi}_1$. The explanation lies in the interrelation between the frequency and the phase: the quantity of importance is the argument

$$\Psi_{jk} = 2\pi f t_{jk} + \phi_k \tag{33}$$

of the cosine function in equation (1). A small error in the frequency can, to some extent, be compensated for by a change in the phase. Note though that the frequency is weighted by t_{jk}, which increases with k; this implies that the impact of small frequency errors on ϕ_k will increase with k.

Some support for these ideas can be found in the correlation coefficients of the frequency estimate $\widehat{f}$ with the phase estimates; the coefficients are −0.55, −0.20 and 0.45 for $\widehat{\phi}_1$, $\widehat{\phi}_2$ and $\widehat{\phi}_3$, respectively.

The 95 per cent confidence intervals for the various parameters are once again close to a ±2 standard error interval around the corresponding estimate given in Table 4. The exception is the error standard deviation: the difference between the entry in Table 4 ($\sigma_e = 0.38$) and that in Table 5 ($\sigma_e = 0.44$), which is about two standard errors in size, shows that $\widehat{\sigma}_e$ is biased.

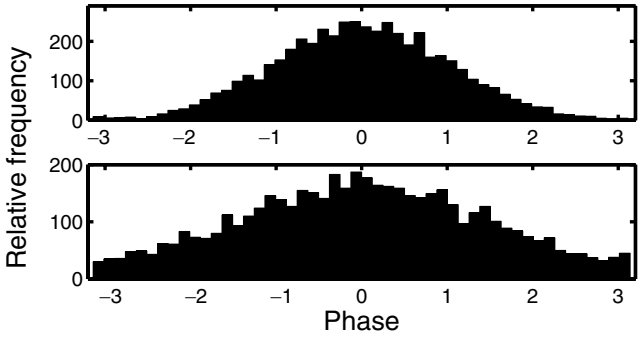

Figure 9. The distributions of $\widehat{\phi}_2$ and $\widehat{\phi}_3$ (see Fig. 8) adjusted for the ambiguity in the definition of phase.

It is noted in passing that in the case of the fixed-phase models, errors in the phases can be reduced by suitably redefining the time zero-point (see e.g. Montgomery & O'Donoghue 1999). As an example, if t_{jk} is replaced by $t_{jk} - \overline{t_{jk}}$ in the first example data set analysed above, the estimated phase is 2.336, with standard error 0.029. The width of the 95 per cent phase confidence interval is correspondingly reduced from 0.21 to 0.11 rad. Of course, the phase is defined with respect to $\overline{t}$, rather than the time of the first observation, which may not be convenient.

It also seems plausible that in the case of variable phases, the time zero-point for each block of observations could be set independently to minimize the errors in the estimated phases.

7 EXAMPLE ANALYSIS OF REAL DATA

The ultracool nature of the object 2MASS J06050196−2342270 (hereafter 2M 0605−2342) was discovered by Cruz et al. (2007). Its spectral classification is L0. Koen (2006) obtained a little over 11 h of I_C band photometry of 2M 0605−2342, spread over three consecutive nights (Fig. 10). The three blocks of observations consist of 33, 98 and 37 measurements, respectively. According to Koen (2006), there is a sinusoidal variation with frequency 9.8 d^{-1} in the data, the amplitude of which declined over the three nights of monitoring.

Figs 11–13 show the least-squares spectra of the data, for each of the eight models. Substantial aliasing can be seen in the latter two diagrams; different one cycle per day aliases are, in fact, selected by Models 3 and 4. The full model fitting results are given in Table 6: Models 3 (changes in mean levels only) and 1 (changes in mean levels, phases and amplitudes) are selected by the BIC and AIC, respectively.

Bootstrap estimates of 95 per cent confidence intervals and standard errors of the parameter estimates can be found in Tables 7 (Model 3) and 8 (Model 1). Perhaps the most noteworthy feature of Table 7 is the large standard error of the estimated phase. Redefining the phase to avoid $\pm 2\pi$ ambiguities reduces this from 0.44 to 0.18, which is also more in line with the width of the 95 per cent confidence interval.

As far as Table 8 is concerned, shifting the phase zero-points to the values in Table 6 and restriction to $[-\pi, \pi]$ lead to standard errors of 0.21, 1.68 and 1.82 for $\widehat{\phi}_1$, $\widehat{\phi}_2$ and $\widehat{\phi}_3$, respectively. The latter two values are very close to those given in Table 8: in the case of the second block this follows because $\widehat{\phi}_2$ is close to zero, so that few bootstrap solutions for ϕ_2 are ambiguous. In the case of $\widehat{\phi}_3$, the distribution of its bootstrap distribution is close to uniform, hence the assumed value of the centroid makes little difference to the spread.

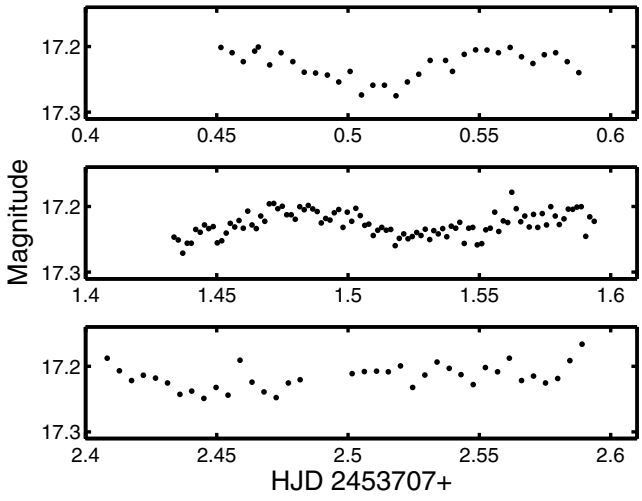

Figure 10. Photometry of the ultracool object 2M 0605−2342 obtained over three successive nights.

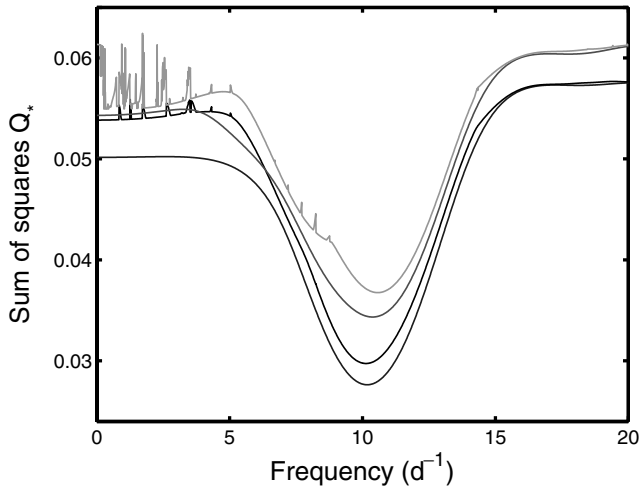

Figure 11. Same as Fig. 5, but for the observations of 2M 0605−2342 shown in Fig. 10.

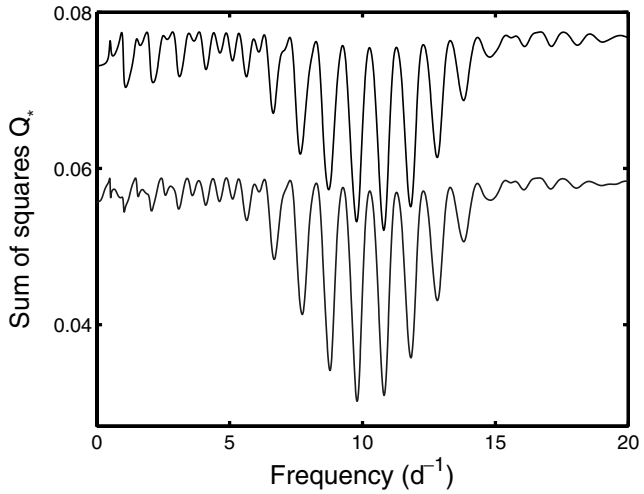

Figure 12. The sum of squares spectra for Models 3 (bottom curve) and 4 (top curve), fitted to the observations of 2M 0605−2342. The Model 4 curve has been offset by +0.015. Note that the different frequency aliases are selected by the two models.

Careful consideration of the results above leads to the conclusion that there is good evidence for a change in the mean brightness level; in particular, although $\widehat{\mu}_2 \approx \widehat{\mu}_1$, the star/brown dwarf was brighter on the third night. The situation as regards the amplitudes is, however, not clear-cut. The fact that the second-string model according to both information criteria is number 5, which involves both mean level and amplitude changes, constitutes some evidence in favour of amplitude changes.

8 CONCLUSIONS

It is not possible to draw general conclusions from the very limited studies presented in Sections 6 and 7. None the less, it seems that the information criteria can fruitfully be used for detecting large model parameter changes, at least in high signal-to-noise ratio data.

The author is grateful to the referee of the paper for drawing attention to the fact that the methodology is not restricted to rotating spotted stars, but can also be applied to other types of periodic variables. In particular, pulsation amplitudes and phases may be wavelength-dependent, and this should be taken into account when combining data blocks obtained using different photometric filters.

In the above it was assumed that the errors were white noise with constant variance. In principle, it is straightforward to extend this to the case where the errors within each block are stationary. The following scheme is suggested.

(i) First fit models as above, and obtain the residuals.

(ii) If the autocorrelation of the errors within a block is non-zero, fit a time-series model to those residuals. Otherwise, calculate the variance of the errors in the block.

(iii) Use the information from step (ii) to calculate the covariance matrices Σ_k of the errors in each block.

(iv) The sum of squares in equation (2) is replaced by

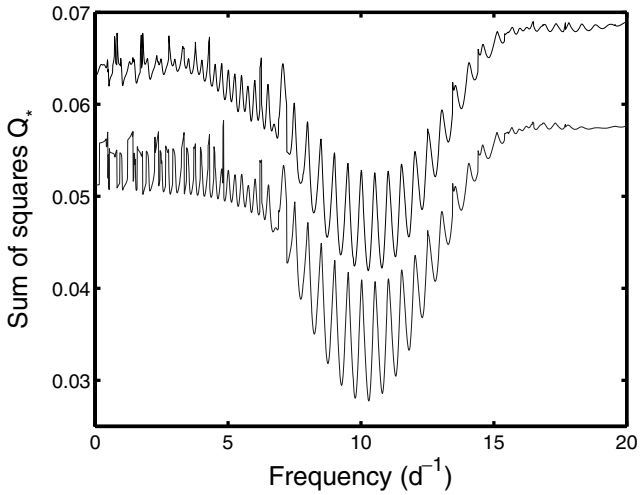

Figure 13. The sum of squares spectra for Models 5 (bottom curve) and 6 (top curve) fitted to the observations of 2M 0605−2342. The Model 6 curve has been offset by +0.0075.

Table 6. Estimated model parameters for the eight models, fitted to the three blocks of observations of 2M 0605−2342 (Fig. 10). The discrepancy between optimal models selected by the AIC (Model 1) and BIC (Model 3) is noteworthy.

Model	1	2	3	4	5	6	7	8
M	10	8	6	4	8	6	8	6
$p_{\rm A}$	0.6069	0.0000	0.0310	0.0000	0.3501	0.0000	0.0120	0.0000
$p_{\rm B}$	0.0350	0.0000	0.5947	0.0000	0.3580	0.0000	0.0123	0.0000
f	10.1751	10.3766	9.7967	10.8004	10.8037	10.7874	10.1295	10.5708
μ_1	17.2331	17.2259	17.2316	17.2256	17.2330	17.2259	17.2315	17.2254
μ_2	17.2281	17.2259	17.2284	17.2256	17.2279	17.2259	17.2282	17.2254
μ_3	17.2142	17.2259	17.2124	17.2256	17.2151	17.2259	17.2129	17.2254
c_1	0.0270	0.0243	0.0188	0.0172	0.0264	0.0238	0.0188	0.0173
c_2	0.0188	0.0184	0.0188	0.0172	0.0187	0.0184	0.0188	0.0173
c_3	0.0106	0.0074	0.0188	0.0172	0.0095	0.0077	0.0188	0.0173
ϕ_1	2.5740	2.5505	2.6641	2.2808	2.2582	2.3343	2.5952	2.4594
ϕ_2	0.0998	−1.2669	2.6641	2.2808	2.2582	2.3343	0.4006	−2.5593
ϕ_3	−2.0992	1.4890	2.6641	2.2808	2.2582	2.3343	−1.4715	−0.8872
$\sigma_{\rm e}$	0.0132	0.0147	0.0136	0.0150	0.0134	0.0146	0.0136	0.0151

Table 7. The results of bootstrapping the BIC optimal model from Table 6.

Parameter	Mean	95 per cent confidence interval	Standard deviation
f	9.80	(9.75, 9.84)	0.022
μ_1	17.232	(17.227, 17.237)	0.0026
μ_2	17.228	(17.225, 17.232)	0.0015
μ_3	17.212	(17.207, 17.217)	0.0025
c	0.019	(0.016, 0.022)	0.0016
ϕ	2.6372	(2.31, 3.00)	0.44
$\sigma_{\rm e}$	0.015	(0.013, 0.017)	0.0011

$$Q = \sum_{k=1}^{K} \sum_{i,j=1}^{N_k} \left(\Sigma_k^{-1}\right)_{ij} [y_{ik} - \mu_k - c_k \cos(\omega t_{ik} + \phi_k)][y_{jk} - \mu_k - c_k \cos(\omega t_{jk} + \phi_k)].$$

(v) Repeat the model fitting described in Sections 2 and/or 3, minimizing Q as defined in step (iv).

The methodology in Sections 2 and 3 does not lend itself to the case in which each block is characterized by a different period. In practice, it seems likely that large period changes will be accompanied by changes in other parameters, so that treating blocks independently is indicated in such cases.

Table 8. The results of bootstrapping the AIC optimal model from Table 6.

Parameter	Mean	95 per cent confidence interval	Standard deviation
f	10.18	(9.54, 10.82)	0.32
μ_1	17.233	(17.228, 17.238)	0.0027
μ_2	17.228	(17.225, 17.231)	0.0016
μ_3	17.214	(17.209, 17.219)	0.0025
c_1	0.027	(0.020, 0.035)	0.0038
c_2	0.019	(0.015, 0.023)	0.0021
c_3	0.011	(0.004, 0.018)	0.0035
ϕ_1	2.56	(2.16, 2.96)	0.39
ϕ_2	0.00	$(-2.96, 2.94)$	1.68
ϕ_3	0.01	$(-2.98, 2.99)$	1.81
σ_e	0.015	(0.013, 0.017)	0.0011

A MATLAB implementation of the algorithms is available from the author.

REFERENCES

Burnham K. P., Anderson D. R., 2004, Sociol. Methods Res., 33, 261
Cruz K. et al., 2007, AJ, 133, 439
Davison A. C., Hinkley D. V., 1997, Bootstrap Methods and their Application. Oxford Univ. Press, Oxford
Efron B., Tibshirani R. J., 1993, An Introduction to the Bootstrap. Chapman & Hall, London
Koen C., 2006, MNRAS, 367, 1735
Montgomery M. H., O'Donoghue D., 1999, Delta Scuti Newsl., 13, 28

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