The analysis of indexed astronomical time series – VI. Covariances of amplitude ratios and phase differences estimated from multicolour photometry of a pulsating star

Chris Koen
South African Astronomical Observatory, PO Box 9, Observatory, 7935 Cape, South Africa

Accepted 2000 March 2. Received 2000 February 23; in original form 1999 October 15

ABSTRACT
One of the tools used to identify the pulsation modes of stars is a comparison of the amplitudes and phases as observed photometrically at different wavelengths. Proper application of the method requires that the errors on the measured quantities, and the correlations between them, be known (or at least estimated). It is assumed that contemporaneous measurements of the light intensity of a pulsating star are obtained in several wavebands. It is also assumed that the measurements are regularly spaced in time, although there may be missing observations. The amplitude and phase of the pulsation are estimated separately for each of the wavebands, and amplitude ratios and phase differences are calculated. A general scheme for estimating the covariance matrix of the amplitude ratios and phase differences is described. The first step is to fit a time series to the residuals after pre-whitening the observations by the best-fitting sinusoid. The residuals are then cross-correlated to study the interdependence between the errors in the different wavebands. Once the multivariate time-series structure can be modelled, the covariance matrix can be found by bootstrapping. An illustrative application is described in detail.

Key words: methods: data analysis – methods: statistical – stars: oscillations.

1 INTRODUCTION
One of the methods commonly used to identify the oscillation modes in pulsating stars is the use of photometric observations obtained in different wavebands. Pulsation theory predicts certain relationships between the amplitudes (or phases) of the pulsation light curves, as seen in the different wavebands. The theory often contains undetermined parameters, and the predictions are commonly presented as regions in two-dimensional plots of, e.g., the ratio of the pulsation amplitudes in two different wavelengths, versus the phase difference between the light curves seen through the two different filters. The observed amplitude ratio and phase difference then provide a single point in the diagram, which hopefully lies in a region in the plane that corresponds to some particular pulsation mode.

In principle it is obviously advantageous to obtain light curves in as many wavebands as possible. There are, however, potential problems with applying the method described above if observations are made through three or more filters: for three filters, there are nine possible combinations of amplitude ratios and phase differences; for four and five filters the numbers grow to 36 and 100. Not only is it not feasible to study so many plots, but it is also unclear how conflicting results ought to be reconciled. Clearly what is needed is a way of combining the many parameters in some way so as to obtain a single solution that is optimal in some sense.

Koen et al. (1999; hereafter KRWM99) proposed minimizing the sums of squares of the differences between the observed and theoretical quantities. [The same fundamental approach was taken by Balona & Evers (1999), although the details of their method differ considerably from those presented below.] The minimization is with respect to the unknown pulsation mode of the star, as well as any other unknown parameters (e.g. those measuring the departure of the oscillations from adiabaticity). Clearly such a minimization needs to incorporate knowledge of the relative accuracies of the observed quantities, as well as correlations amongst these. Such knowledge is summarized in the covariance matrix \( \Sigma \) of the observables. This paper is devoted to a method for the estimation of \( \Sigma \).

If the standard errors of the estimated quantities (and their correlations) were known, the problem would be relatively simple (KRWM99). One of the complicating factors is the estimation of the standard errors of estimated amplitudes and phases, in situations where correlated noise is superposed on the signal of interest (Schwarzenberg-Czerny 1991; Montgomery & O’Donoghue 1999; Koen 1999, hereafter Paper V). In the present context there is the added complication of correlations between noise processes measured through different filters.
The complexity of the problem does not encourage an analytical approach; therefore the method proposed is bootstrapping, also
known as computer resampling (e.g. Efron & Tibshirani 1993). The various steps in implementing the bootstrap are as follows.

(1) The observations are pre-whitened by the sinusoid(s) arising from the pulsations. This leaves a noise process for each filter.
(2) A time-series model is fitted to the noise process associated with each of the filters. The model residuals, or ‘innovations’, should be free of autocorrelation.
(3) The innovation sequences associated with the different wavebands are cross-correlated, and the results are used to construct a covariance matrix for all the innovations.
(4) The innovations are transformed into a new set of residuals, which are free of both autocorrelation and cross-correlations. Since these new residuals do not have any sequential coherence at all, any permutation of them is statistically completely equivalent.
(5) The final residuals are used to re-construct sets of observations that are statistically equivalent to the originals. The parameters of interest (amplitude ratios, phase differences) are estimated for the synthetic data sets.
(6) Step 5 is repeated a large number of times, and the results are used to establish the distributional properties of the parameters of interest.

The general strategy is described in some detail in Section 2, followed by a complete analysis of a ‘real life’ example. Section 4 contains a few remarks.

2 METHODOLOGY

The situation considered is as follows: contemporaneous observations in several wavebands are obtained of a star that is varying sinusoidally with a known angular frequency \( \omega_0 \). (Extension to the case of several different periodicities is straightforward, but the notation more complicated.) The observations are described by

\[
y_j(k) = C(k) \cos(\omega_0 t + \phi(k)) + \varepsilon_j(k), \quad k = 1, 2, \ldots, N(k),
\]

where \( k \) indexes the filter and \( j \) is the time of observation. The notations \( C(k) \) and \( \phi(k) \) are used for the best-fitting (in the least-squares sense) amplitude and phase of the variations in waveband \( k \). The residuals left after subtracting the best-fitting sinusoid from the observations \( y_j(k) \) are denoted by \( \varepsilon_j(k) \). The observables of direct interest are the amplitude ratios

\[
u(1) = C(2)/C(1), \quad u(2) = C(3)/C(2), \ldots,
\]

\[
u(K - 1) = C(K)/C(K - 1),
\]

and the phase differences

\[
\nu(1) = \phi(2) - \phi(1), \quad \nu(2) = \phi(3) - \phi(2), \ldots,
\]

\[
\nu(K - 1) = \phi(K) - \phi(K - 1).
\]

It is convenient to stack the \( u(j) \) and \( \nu(j) \) in a single vector \( \mathbf{z} \), with its \( M = 2K - 2 \) components given by

\[
z_1 = u(1), \quad z_2 = u(2), \ldots, z_{K-1} = u(K - 1),
\]

\[
z_K = \nu(1), \quad z_{K+1} = \nu(2), \ldots, z_M = \nu(K - 1).
\]

Of interest here is the covariance matrix \( \mathbf{\Sigma} \) of the vector \( \mathbf{z} \).

Two important remarks need to be made: first, if the \( K \) noise sequences \( \{ \varepsilon_1(k), \varepsilon_2(k), \ldots, \varepsilon_K(k) \} \) are white noise processes which are also uncorrelated with one another, calculation of \( \mathbf{\Sigma} \) is fairly straightforward. The usual regression covariance matrix of the \( C(k) \) and \( \phi(k) \) is reliable in this case, and may be used to calculate a good approximation for \( \mathbf{\Sigma} \) (KRWM99). In what follows it is assumed that the \( K \) individual noise processes may be autocorrelated, as well as cross-correlated with one another. This invalidates use of the ordinary regression covariance results.

The second point concerns the assumption that the observations are made at regularly spaced discrete time intervals. This is obviously quite restrictive: it typically applies only to multi-channel high-speed photometry. Another observational configuration that falls within this ambit is that of continuously cycling through a filter sequence at a set speed, as when doing CCD photometry. In such cases the fundamental time unit is that required to complete one filter sequence.

We propose to calculate \( \mathbf{\Sigma} \) either by simulation, or by a technique called ‘bootstrapping’ (see below, and e.g. Efron & Tibshirani 1993). Both methods require statistical models for the individual noise processes \( \{ \varepsilon(k) \} \), as well as for the interrelationships amongst them. The next several paragraphs are devoted to a description of how to accomplish this.

A simplifying, but non-essential, assumption is that the sequences of residuals \( \{ \varepsilon(k) \} \) are statistically stationary, i.e. have properties (such as the mean value and variance) that do not change with time. (The principles involved do not change if the series are non-stationary, but additional treatment of the sequences may be required. The example observations analysed in Section 3 are in fact non-stationary in variance, but the data can be divided into segments that are stationary.) This innocuous assumption has the important consequence that the sequences of residuals may be modelled as autoregressive moving average (ARMA) processes — see Koen & Lombard (1993, Paper I in this series). The most general form of an ARMA process is

\[
\varepsilon_j(k) = \sum_{i=1}^{p} \alpha_i(k) \varepsilon_{j-i}(k) + \sum_{n=1}^{q} \beta_n(k) \eta_{j-n}(k) + \eta_j(k),
\]

where the sequences \( \{ \eta_1(k), \eta_2(k), \ldots, \eta_q(k) \} \) are white noise. The notation ‘ARMA\((p,q)\)’ is used, indicating that the process has autoregressive and moving average orders of respectively \( p \) and \( q \).

In practice, the great majority of processes are of forms with \( 0 \leq p, q \leq 2 \).

It would be out of place to describe the identification of the orders \( p \) and \( q \) in detail here; the interested reader is referred to the classical text by Box & Jenkins (1976), or modern treatments such as that by Brockwell & Davis (1991). The actual fitting of models is also dealt with in these texts, as well as in almost any other modern book on time-series analysis. A large number of commercial software packages capable of full ARMA analyses are available, and the accompanying manuals often contain brief descriptions of the procedures necessary to identify, estimate and test the optimal model.

A possible complication is that there may be ‘missing values’, i.e. observations are not available at all \( N \) time points. (For example, data streams might be interrupted in order to measure sky background brightness, or because of poor weather or instrument failures.) Commercial software is rarely equipped to deal with such data: often missing values are simply filled in with the overall mean values. This may not be serious if time series are almost complete, but may pose a substantial problem if there are many missing observations.

© 2000 RAS, MNRAS 316, 613–620
Fortunately, there are statistical methods available for ARMA analysis of large data sets from which substantial numbers of values are missing: Paper I contains a description of a technique due to Jones (1980). Jones’s (1980) algorithm is efficient in the sense that it based on maximizing the likelihood of the data. It avoids having to deal with the full covariance matrices of the data by drawing on a Kalman filter algorithm. Again, it would take us too far afield to deal with this in detail; the interested reader is referred to the papers mentioned.

Once the optimal ARMA model has been selected, it is necessary to test whether the model residuals are indeed uncorrelated. This is easily done by calculating the appropriate autocorrelation function, and checking that it is consistent with the residuals (estimated $\eta$) being white noise. If the residuals are correlated, a different model — often suggested by the pattern of the residual autocorrelations — is fitted. The process is continued until satisfactory models have been fitted to the $K$ individual series of observations. Note that missing observations require that some care be taken with the calculation of the autocorrelation function: equation (24) in Paper V is an appropriate formula.

The next step in the analysis is to enquire whether the individual series are related. It might be thought that this could have been done very simply at the outset by calculation of cross-correlation functions between the different sets of residuals defined in (1). This happens to be fraught with danger: cross-correlation of two completely unrelated series, which are themselves autocorrelated, can often lead to spurious evidence of strong correlation between them (see e.g. Jenkins 1979; Koen 1994). What is required is first to remove the autocorrelation from the individual series. Once appropriate ARMA models have been found for the individual series, this is trivial: the ARMA model is subtracted from the original sequence of observations (a procedure referred to as ‘pre-whitening’), leaving the correlation-free set of $\{\eta_j(k)\}$. Cross-correlating the $\{\eta_j(k)\}$ for different values of $k$ will then allow conclusions to be drawn about the relatedness of the error processes associated with the different filters.

Only the case of limited-span cross-correlation is considered in this paper: by this is meant that the different $\eta(k)$ are only correlated if they are close to being contemporaneous. Symbolically,

$$\text{cov}[\eta_j(k_1), \eta_{j+\ell}(k_2)] = \begin{cases} 0 & |\ell| \leq L \\ \neq 0 & |\ell| > L \end{cases}$$

(6)

for finite $L$. Typically $L$ will have values of 1 or 2. It is useful to define the vectors

$$w_j = [\eta_j(1)\eta_j(2)\ldots\eta_j(K)]'$$

and

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}.$$

(7)

We also define

$$A_\ell = \text{cov}(w_j, w_{j+\ell}).$$

(8)

Note that $A_\ell$ is independent of the index $j$, by the assumption of statistical stationarity. Also, $A_{-\ell} = A_\ell$.

Using the notation above, the covariance matrix of all the $\eta_j(k)$ can be written quite concisely; for the case $L = 1$ it is

$$S = \text{cov}(W, W') = \begin{bmatrix} A_0 & A_1 & 0 & \cdots & 0 \\ A_1^\top & A_0 & A_1 & \cdots & 0 \\ \vdots & A_1^\top & A_0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & A_1^\top \\ 0 & \cdots & \cdots & \cdots & A_0 \end{bmatrix}.\tag{9}$$

The form for larger $L$ has a similar band diagonal structure. The matrix $S$ is a covariance matrix, and hence it is symmetric and positive definite; this implies that it can be factorized as

$$S = LL^\top.$$

(10)

where $L$ is a lower triangular matrix (Cholesky decomposition – see e.g. Searle 1971). Define

$$Z = L^{-1}W;$$

(11)

then, from (10),

$$\text{cov}(Z, Z') = I,$$

where $I$ is the identity matrix. The last equation implies that the components of the vector $Z$ are uncorrelated, and constitute the basic building blocks from which the observed noise time series can be reconstructed.

To summarize: the cross-correlation function is used to establish the value $L$ of the maximum lag, and to estimate the entries in the matrices $A_\ell$; the matrix $S$ is constructed from the $A_\ell$ (equation 9); the matrix $L$ follows from (10). The vector $Z$ follows from (11); its components are fully uncorrelated.

It remains to discuss the generation of the synthetic data sets to be used in the bootstrapping procedure. The first step is to generate sets of vectors that are statistically equivalent to $Z$. This is easy, as the components of $Z$ are uncorrelated; the equivalents $Z'$ are formed by randomly drawing (with replacement) the necessary number of components from $Z$. The equation

$$W' = LZ'$$

(12)

then gives vectors $W'$ which are statistically similar to $W$, i.e.

$$\text{cov}(W', (W')') = S.$$

By (7), we then have a set of $N$ $K$-vectors $w'_j$, from which the full $K$ bootstrap noise processes $\{e'_j(k)\}$ may be created – see (5). The simulated observations follow immediately from equation (1).

As mentioned in the Introduction, the procedure described in the previous paragraph is repeated a very large number of times, and the amplitudes $C(k)$ and phases $\phi(k)$ estimated for each bootstrap sample. The covariances of the amplitude ratios and phase differences are then easily calculated from the collection of simulation results.

3 AN EXAMPLE

KRWM99 published the results of $UBVRI$ photometry of the $\delta$ Scuti star HD 95321. The star is apparently monoperiodic, with $P = 0.213658d$. The observations were obtained by CCD photometry, continuously cycling through the set of filters. One cycle of $UBVRI$ observations lasted 248.44 s, and this is used as the unit of time in discretizing the time variable. The observing run spanned six nights, one of which was not usable; the individual observing sessions ranged from 4.5 to 5.3 h. The number of useful observations ranged from 225 in the $I$ band to

© 2000 RAS, MNRAS 316, 613–620

Indexed astronomical time series – VI 615
304 in the $U$ band. An example of the noise processes in the different wavebands is shown in Fig. 1: these are the residuals after subtracting the best-fitting sinusoid with $P = 0.213656\text{d}$ from each of the sets of differential photometry obtained on a particular night.

There are of course large time lapses between the nightly observing runs, and these gaps are not integer multiples of 248.44 s. There is thus some ambiguity in the ‘cycle count’ during the gaps. This is of little consequence in dealing with the identification and simulation of the noise processes, as these are essentially independent from night to night. Fitting a short-period sinusoid over the entire set of times is a different matter, but need not be of great concern: we simply define a new time series on the estimated integer times, with a period equal to that of HD 95321. Although this is not an exact representation of the observed situation, we are only interested in the variability of the amplitude and phase estimates (i.e. relative, rather than absolute, values): the newly defined time series will clearly be an adequate proxy for the real thing as far as this aim is concerned.

We proceed with an analysis of the residuals left after pre-whitening the observations by the best-fitting sinusoid with the appropriate period. The first task is to fit models of the form (5) to these data. If the residuals from each night are modelled individually, a total of 25 models are required (five nights of observations, for each of the five filters used). There are several reasons why, if it is at all viable, it is desirable to fit a single model to all the residuals for a given filter. These include simplicity, reduction in the modelling effort, and a gain in accuracy through fitting to a larger data set. Note that this assumption is also implicit in the description, and notation, used in Section 2 above.

The statistical structure of the noise in Fig. 1 is typical of all the residuals. Clues such as the obviously long correlation memory suggests an ARMA$(1,0)$ model as a first attempt. Examination of the set of innovations $\{\eta_j(k)\}$ [see (5)] showed this model to be inadequate: innovations were strongly autocorrelated at a lag of one time unit. Such a limited-length autocorrelation structure is typical for an ARMA$(0,1)$ time series, which, combined with the original model, implies that an ARMA$(1,1)$ model may fit the noise well. This turned out to be so: the autocorrelation functions of the innovations of ARMA$(1,1)$ model fits were found to be consistent with white noise, implying that the models gave good representations of the entire correlation structure in the $K$ individual noise processes $\{e_j(k)\}$. It is reassuring that an ARMA$(1,1)$ is the optimal model (according to the Bayes information criterion – see Paper I) for the data for all of the filters: it suggests that this model may be describing common underlying noise processes. (See also Paper V, in which similar conclusions were reached for a large number of photometric noise processes.) The models are summarized in Table 1.

An examination of plots of the innovations against time shows that the scale of the scatter is different during different nights. Since the innovations are not necessarily Gaussian (see below), care has to be taken in applying standard tests for differences in the variances over different nights. For example, the standard Bartlett’s test (Bartlett 1937) is known to be sensitive to deviations from normality, and should not be used. Fortunately, the innovations are uncorrelated, and hence use can be made of distribution-free permutation tests. We use the simple statistic (Hartley 1950)

$$H(k) = \max_j \frac{\sigma_j^2(k)}{\min_j \sigma_j^2(k)},$$

where $j$ indexes the different nights, and $\sigma_j^2(k)$ is the variance of the innovations during night $j$ in waveband $k$. The statistic $H(k)$ is first calculated for the original set of innovations. The innovations

![Figure 1. An example of the residuals after pre-whitening a night’s observations of HD 95321 by a sinusoid with a period of 0.213656d. The run length was 5.1h. Starting at the top, the sequence of filters is $U$, $B$, $V$, $R$ and $I$.](image)

Table 1. The results of fitting ARMA$(1,1)$ models to the data for each of the filters. All nightly data for a given filter have been combined. The standard deviation $\sigma_\eta$ of the innovations is given in millimagnitudes. The first three columns are the initial estimates; the other columns contain estimated values after correcting the data for heteroscedasticity. The last column shows the minimum standard deviation of the innovations, followed by scalefactors for each of the five nights.

| Filter | $\alpha_1$ | $\beta_1$ | $\sigma_\eta$ | $\alpha_1$ | $\beta_1$ | $\sigma_\eta$ |
|--------|------------|------------|----------------|------------|------------|----------------|
| $U$    | 0.95       | -0.55      | 7.1            | 0.95       | -0.57      | 6.3 (1.11 1.00 1.03 1.24 1.24) |
| $B$    | 0.89       | -0.49      | 8.0            | 0.91       | -0.54      | 5.8 (1.00 1.29 1.19 1.83 1.46) |
| $V$    | 0.82       | -0.51      | 13.6           | 0.85       | -0.49      | 6.6 (1.00 2.65 1.83 2.32 2.03) |
| $R$    | 0.94       | -0.65      | 11.5           | 0.95       | -0.65      | 8.3 (1.00 1.47 1.07 1.34 1.81) |
| $I$    | 0.83       | -0.45      | 11.5           | 0.85       | -0.48      | 7.9 (1.00 1.05 1.48 2.11 1.33) |
It was found that the $B$, $V$ and $R$ data sets each had one outlying innovation ($T$ statistics of 4.87, 4.78 and 4.18; significance levels of 0.0004, 0.0004 and 0.005); after removal of these, all $T$-values have significance levels larger than 0.08.

An assumption of the outlier test is that the innovations have a Gaussian distribution. Since this is also a point of general interest, formal significance tests for deviation from normality were carried out. Three so-called ‘empirical distribution function’ statistics were calculated for each of the five sets of innovations: the Anderson–Darling $A^2$, Cramer–von Mises $W^2$ and Watson $U^2$ statistics [see D’Agostino & Stephens (1986) for definitions, significance level tables, etc.]. The $V$-band innovations deviate from normality more significantly than the 5 per cent level; the $I$-band innovations are marginally non-normal ($U^2$ is significant at the 5 per cent level; the other two statistics are non-significant); and the innovations associated with the other three wavebands have distributions consistent with Gaussians. The results imply that some caution needs to be exercised when basing deductions on the assumption of normality, at least for the $V$-band data.

A point regarding the tests for outliers and for normality worth remarking on is that the mean values of the sets of innovations were assumed known. This has a marked influence on the test results: the tests are a lot less stringent than they would be if sample means were used. However, as the innovations are effectively regression residuals, the assumption that the mean values are equal to zero is certainly plausible.

The results of cross-correlating the innovations associated with the various filters are given in Table 2. The standard errors of the cross-correlations are estimated as $M^{-1/2}$, where $M$ is the number of terms summed in calculating the cross-correlation. All correlations larger than two standard errors, and with lags $|\ell| \leq 5$, are listed. As expected, the strongest correlations occur at lags of zero and one time units. There are two reasons for doubting correlations at lags of 2 units or more. First, it seems implausible on physical grounds that there could be strong correlations at a larger lag, without significant correlation also occurring at smaller lags. Secondly, with only two exceptions, the correlations at lags larger than one unit are only marginally significant. In what follows, only correlations with $|\ell| \leq 1$ are considered meaningful, i.e. it is assumed that $L = 1$ in (6).

The two non-zero matrices $A_e$ (with magnitudes in mmag) are

$$ A_0 = \begin{bmatrix} 39.9 & 7.7 & 0 & 0 & 0 \\ 7.7 & 31.1 & 5.1 & 0 & 0 \\ 0 & 5.1 & 40.5 & 8.1 & 0 \\ 0 & 0 & 8.1 & 64.2 & 18.3 \\ 0 & 0 & 0 & 18.3 & 62.1 \end{bmatrix}, \quad \text{(15)} $$

$$ A_1 = \begin{bmatrix} 4.4 & 0 & 0 & 0 \\ 0 & 6.0 & 8.6 & 8.5 \\ 0 & 0 & 0 & 11.0 \\ 0 & 10.0 & 0 & 0 \\ 9.4 & 0 & 0 & 0 \end{bmatrix}. \quad \text{(16)} $$

There are a number of important practical issues in the calculation of the matrix $L$ in (10). First, the data are incomplete, since some of the components of $w_i$ are sometimes unavailable. In
such cases, the relevant rows and columns of \( A_0 \) and \( A_1 \) are deleted. There are also instances in which entire vectors \( w \) are missing (as, for example, in the gaps between observing runs). In these cases successive \( w \) are not correlated, and hence the matrix \( A_j \) must be replaced by the null matrix. The second issue of some importance is the size of the matrix \( S \); in the present instance the total number of elements in \( W \) is 1347, giving more than 1.8 million entries in \( S \). Of course, only half this number needs to be stored, as \( S \) is symmetrical, but the computer memory requirement is still enormous. One possible way of dealing with the problem is to make use of the sparsity of \( S \) and \( L \), and to keep careful track of the matrix positions of the non-zero elements. There is also a much simpler approach which can be used in the present case: the longest uninterrupted stretch of data is about 40 time units long, and each such block of residuals is uncorrelated, or very weakly correlated, with the rest. This means that \( S \) can be treated as consisting of a number of relatively small uncorrelated blocks, each of which can be independently Cholesky-decomposed as required in (10). For each independent block, the relevant part of the vector \( Z \) follows from (11).

A slightly different strategy is followed in implementing (12), as the ultimate aim is to simulate the noise processes \( e_j(k) \) described by (5). This requires having values of \( e_j'(k) \), i.e. noise values from time periods \( t = j + 1 \) preceding the epoch \( t = j \). We deal with this by calculating matrices \( S' \) as in (9), but (i) padding each of the independent blocks of which \( S \) is composed by an extra 10 rows of the form in (9), and (ii) ignoring the possibility of missing observations. Thus, if the block of observations is of length \( n \), we simulate the covariance matrix \( S' \) of \( n + 10 \) observations with the required covariances. The corresponding \( W' \) follows from (12), and the required \( \{ e_j'(k) \} \) from (5). Finally, the \( e_j'(k) \) corresponding to unobserved \( e_j(k) \) are ignored.

Results based on 5000 bootstrap replications of the original data are now discussed. Table 3 compares the estimates of the standard errors from the ordinary linear least-squares (LLS) theory and those obtained from bootstrapping, for the amplitudes and phases of the sinusoid observed in each of the five wavebands. It appears that the LLS amplitude standard errors underestimate the true errors by a factor of about 2, while the LLS phase standard errors underestimate by a factor of about 3. Defining the vectors

\[
C = \begin{bmatrix} C(U) \\ C(B) \\ C(V) \\ C(R) \\ C(I) \end{bmatrix}, \quad \Phi = \begin{bmatrix} \phi(U) \\ \phi(B) \\ \phi(V) \\ \phi(R) \\ \phi(I) \end{bmatrix},
\]

of estimated amplitudes and phases, the correlation matrices

\[
corr(C, C') = \begin{bmatrix} 1.00 & 0.33 & 0.02 & 0.0007 & 0.16 \\ 1.00 & 0.30 & 0.38 & 0.17 \\ 1.00 & 0.16 & 0.20 \\ 1.00 & 0.25 \\ 1.00 \end{bmatrix}
\]

and

\[
corr(\Phi, \Phi') = \begin{bmatrix} 1.00 & 0.31 & 0.0007 & -0.05 & 0.18 \\ 1.00 & 0.31 & 0.34 & 0.19 \\ 1.00 & 0.13 & 0.20 \\ 1.00 & 0.25 \\ 1.00 \end{bmatrix}
\]

are also of interest. Some of these correlations are clearly non-negligible, and will play a role when comparing theoretical and observed amplitude ratios and phase differences. Correlations between the amplitude and phase estimates are quite small: the largest absolute value is less than 0.1, implying that the quantities in equations (2) and (3) are essentially decoupled from one another.

The final results are given in Table 4, and in the correlation matrices (20) and (21). Perhaps the most notable result in Table 4 is that, with the now more dependable errors, none of the phase differences is significantly non-zero. In view of the results in Table 3, it is not surprising that the LLS errors quoted in Table 4 underestimate the true errors by factors of 2 to 3.

Defining vectors \( U \) and \( V \) with components \( u(i) \) and \( v(j) \) as in equations (2) and (3), the amplitude ratio and phase difference correlation matrices are

\[
corr(U, U') = \begin{bmatrix} 1.00 & -0.11 & 0.13 & -0.25 \\ 1.00 & -0.53 & 0.16 \\ 1.00 & -0.56 \\ 1.00 \end{bmatrix}
\]

and

\[
corr(V, V') = \begin{bmatrix} 1.00 & -0.08 & 0.16 & -0.27 \\ 1.00 & -0.47 & 0.14 \\ 1.00 & -0.66 \\ 1.00 \end{bmatrix}
\]

Table 3. LLS estimates of the amplitudes and phases of the 4.680412 cycle d\(^{-1}\) variation of HD 95321, as determined by KRWM99. The formal LLS standard errors and bootstrap standard errors are shown.

| Filter | Amplitude | LLS error | Bootstrap error | Phase | LLS error | Bootstrap error |
|--------|-----------|-----------|-----------------|-------|-----------|-----------------|
| U      | 27.3      | 0.97      | 2.7             | 1.69  | 0.035     | 0.11            |
| B      | 30.9      | 0.90      | 2.3             | 1.71  | 0.028     | 0.079           |
| V      | 29.0      | 1.3       | 2.7             | 1.85  | 0.044     | 0.11            |
| R      | 20.9      | 1.4       | 3.1             | 1.62  | 0.063     | 0.21            |
| I      | 16.5      | 1.3       | 2.6             | 1.87  | 0.077     | 0.18            |

© 2000 RAS, MNRAS 316, 613–620
There are several requirements that the statistical relationships between the innovations: it is certainly desirable to fit a complete multivariate time-series model to the data simultaneously, rather than use a piecemeal strategy such as done here. It may be surmised that a general line of attack on the problem, which should work in most cases; and thirdly, to develop specific results which should provide some sense of the difference in the final answers between a full solution and the simple-minded approach used by KRWM99.

The scheme used above was to some extent dictated by the availability of software, rather than sound theoretical considerations: it is certainly desirable to fit a complete multivariate time-series model to all the data simultaneously, rather than use a piecemeal strategy such as done here. It may be surmised that a VARMA(1, 1) model [i.e. a multivariate ARMA(1, 1) – see e.g. Lütkepohl 1993] would furnish a good description of the data. However, development of the software necessary to fit such models to data with missing values would be non-trivial.

A considerable part of the labour involved in calculating (20), (21) and the bootstrap standard errors in Table 4 stems from the requirement that the statistical relationships between the innovation processes associated with the different filters be faithfully reproduced in the bootstrapping machinery. If the processes in the different wavebands were unrelated, $A_0$ would be a diagonal matrix, $A$, would be null for $j \neq 0$, and $S$ would be diagonal; the triangular matrix $L$ would reduce to diagonal form, with elements equal to the square root of the elements of $S$. Clearly, the computer memory requirements would be very modest indeed, and the programming considerably simplified. It therefore seemed worthwhile to re-calculate standard errors and correlation matrices of the amplitude ratios and phase differences, ignoring the inter-related nature of the five series of observations. Interestingly, the standard errors are increased by 5–18 per cent, a typical figure being 15 per cent. Comparison of (22) and (23) with (20) and (21) shows that some of the correlations are quite strongly affected, so that neglecting the inter-dependence of the observed series may be risky:

$$\text{corr}(U, U') = \begin{bmatrix} 1.00 & -0.34 & -0.02 & 0.002 \\ 1.00 & -0.42 & -0.002 \\ 1.00 & -0.58 & 1.00 \end{bmatrix},$$

(22)

$$\text{corr}(V, V') = \begin{bmatrix} 1.00 & -0.35 & -0.009 & 0.03 \\ 1.00 & -0.37 & -0.01 \\ 1.00 & -0.67 & 1.00 \end{bmatrix}.$$  

(23)

As expected, the correlations between the $u(i)$ and $v(j)$ are small in absolute value, the largest being 0.06.

### 4 A FEW REMARKS

The casual reader may be left with a sense of disappointment that the author has not provided a simple, clear-cut, universally applicable solution to the problem discussed above. It may therefore be useful to state what the author’s intentions were: first, to attempt to demonstrate that, at least in principle, the required error covariance matrices can be calculated; secondly, to map out a general line of attack on the problem, which should work in most cases; and thirdly, to develop specific results which should provide some sense of the difference in the final answers between a full solution and the simple-minded approach used by KRWM99.

The last observation has an important consequence for meaningful mode discrimination. The Gaussian log likelihood of $z$ is

$$L = -4 \ln 2 \pi - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \left| z - z_p \right| \Sigma^{-1} \left| z - z_p \right|,$$

(24)

where $z$ and $\Sigma$ are defined in Section 2 above, and $z_p$ is the theoretically predicted value of the vector $z$. Table 5 compares the results for covariance matrices $\Sigma$ based on ordinary least-squares (KRWM99) and the bootstrapping technique of this paper. The ordering of the likelihood of the modes in terms of $S$ is the same in the two cases, but the contrasts are much reduced in the case of $\Sigma$ based on bootstrapping.

The last observation has an important consequence for meaningful mode discrimination. The Gaussian log likelihood of $z$ is

$$L = -4 \ln 2 \pi - \frac{1}{2} \ln |\Sigma| - \frac{1}{2} \left| z - z_p \right| \Sigma^{-1} \left| z - z_p \right|,$$

(25)

This means that the likelihood ratio statistics for comparing the results for different $\ell$ reduce to

$$\lambda(\ell_1, \ell_2) = 2 \left[ \ln L(\ell_1) - \ln L(\ell_2) \right] = SS(\ell_2) - SS(\ell_1).$$

(26)

In order to gain some insight into whether the results for the different modes in Table 5 differ meaningfully, we use the fact that the $\lambda(\ell_1, \ell_2)$ with $\ell_1$ corresponding to the optimal solution is asymptotically chi-squared-distributed with three degrees of freedom. (The three arise from the fact that not only $\ell$ but also the two non-adiabaticity parameter $R$ and $\Psi$ are estimated for each

---

**Table 4.** The amplitude ratios $u$ and phase differences $v$ for the different wavebands in which HD 95321 was observed, as given in KRWM99. The indices of the amplitudes and phases in (2) and (3) are $U = 1$, $B = 2$, $V = 3$, $R = 4$ and $I = 5$. Standard errors in KRWM99 and those calculated from 5000 bootstrap samples are compared.

|     | Observed Amplitude ratios | LLS error | Bootstrap error | Observed Phase differences | LLS error | Bootstrap error |
|-----|---------------------------|-----------|-----------------|---------------------------|-----------|-----------------|
| $u(1)$ | 1.13                      | 0.052     | 0.12            | $v(1)$          | 0.025     | 0.045           |
| $u(2)$ | 0.94                      | 0.051     | 0.094           | $v(2)$          | 0.14      | 0.052           |
| $u(3)$ | 0.72                      | 0.059     | 0.12            | $v(3)$          | -0.23     | 0.076           |
| $u(4)$ | 0.79                      | 0.084     | 0.15            | $v(4)$          | 0.26      | 0.099           |

**Table 5.** The weighted sums of squares (24) as calculated for ordinary LLS error estimates, and for error estimates based on bootstrapping. Likelihood ratio statistics $\lambda$ are also given.

|     | LLS | Bootstrapping |     |     |
|-----|-----|---------------|-----|-----|
| $\ell$ | 0 23.0 | 1.3 | 4.11 | 0.55 |
|     | 1 28.7 | 7.0 | 5.47 | 1.91 |
|     | 2 21.7 | 3.56 | 4.72 |
|     | 3 47.5 | 25.8 | 8.28 |
|     | 4 42.5 | 20.8 | 8.22 | 4.66 |

---

© 2000 RAS, MNRAS 316, 613–620
solution.) The interested reader is referred to any general introductory text on statistics (e.g. Mood, Graybill & Boes 1974) for more information.

The likelihood ratio statistics are also shown in Table 5. In the case of the ordinary least-squares entries, solutions for $\ell \geq 3$ can be rejected at a confidence level far better than 1 per cent; for the bootstrapping entries the worst and best solutions differ at only the 20 per cent level. The implication is that, with realistic error estimates, clearcut identification of the pulsation mode of HD 95321 from the KRWM99 multicolour photometry may be impossible. The KRWM99 data are rather more extensive, in terms of both wavelength coverage and time coverage, than those that have often used for mode identification. A repetition of the exercise of this paper for other stars is called for, in order to establish whether mode identification from multicolour photometry alone is currently viable.

REFERENCES

Balona L. A., Evers E. A., 1999, MNRAS, 302, 349
Barnett V., Lewis T., 1994, Outliers in Statistical Data, 3rd edn. John Wiley & Sons, Chichester
Bartlett M. S., 1937, Proc. R. Soc. London A, 160, 268
Box G. E. P., Jenkins G. M., 1976, Time Series Analysis, Forecasting and Control. Holden-Day, Oakland, CA

This paper has been typeset from a \TeX\ file prepared by the author.