Mon. Not. R. Astron. Soc. **263**, 287–308 (1993)

The analysis of indexed astronomical time series – I. Basic methods

Chris Koen[1] and Fred Lombard[2]

[1] *South African Astronomical Observatory, PO Box 9, Observatory, Cape 7935, South Africa*
[2] *Department of Statistics, Rand Afrikaans University, PO Box 524, Johannesburg 2000, South Africa*

Accepted 1992 December 1. Received 1992 October 5

ABSTRACT

The application of autoregressive moving-average (ARMA) modelling to discrete time or indexed data is discussed. An essential tool is the autocorrelation function, and its utility is demonstrated with some examples. Extensions of the ARMA models to include non-stationarity are described; variability in both the series mean and the ARMA coefficients can be dealt with by recursively calculated or 'state space' models. A number of tests for non-stationarity of the series mean are described in detail. Particular attention is devoted to the effects of serial dependence of the data. A problem with the O − C (observed − calculated) method as traditionally used in astronomy is pointed out. Non-stationarity in the data variance and the identification of outlying observations are dealt with at some length. Tests designed to discriminate between different outlier types are described. Brief mention is made of a few other important topics, particularly non-linearity and data transformation.

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

1 INTRODUCTION

In this series of papers, the authors will be dealing with the analysis of indexed or regularly spaced astronomical time series. Typical examples of such time series are the successive cycle lengths of pulsating stars, the lengths of the quiescent periods between eruptions of dwarf novae, and regularly spaced high-speed photometric observations. The series of Figs 1 and 2 are representative of the first two examples. It is evident that the first series has a more or less well-defined mean value of 332.3 d, while the mean of the VW Hydri data appears to shift up and down. There is considerable scatter around the mean values. It is of course the stochastic or noisy nature of these series which poses a challenge to the modeller. The particular aim of this first paper in the series is to give a description of some of the statistical methods for the analysis and modelling of such series. As will become clear, some of the methods are well known in the context of astronomy; some are well known to statisticians and engineers but have not been used by astronomers; and some are altogether new techniques.

The next section of the paper is concerned with the modelling of stationary time series. By this is roughly meant that the nature of the series is constant over the span of the observations. The series mean, or the mean amplitude of the noise, for example, remains unchanged. It will be evident that the serial correlation structure of the time series is of fundamental importance in constructing models. The recursive modelling of time series is described in Section 3. This methodology provides a flexible means of dealing with non-stationarity in the series mean. Two recursive filters are presented: the first, due to Ledolter (1979), is able to model changes in the series correlation structure, while the second (Jones 1980) is designed to cope with gaps in the observed series. The use of O − C diagrams for identification of non-stationarity of the mean of indexed astronomical time series is well known. A related method, the CUSUM chart, is introduced in Section 4, and significance tests based on this procedure are described. Attention is paid in particular to the modifications of standard tests necessitated by serial dependence of data. Section 5 describes tests for the constancy of the noise variance of series. Heterogeneity of the variance (referred to as 'heteroscedasticity' in the statistics literature) may arise, for example, from different levels of activity of the system under study. Improved measurement techniques could also lead to decreasing variance in the reported observations. A number of other important topics related to the contents of this paper are touched on in Section 6. These include non-linear and chaotic time series, where the aim of the discussion is to highlight the relationship of such series to the linear models extensively dealt with

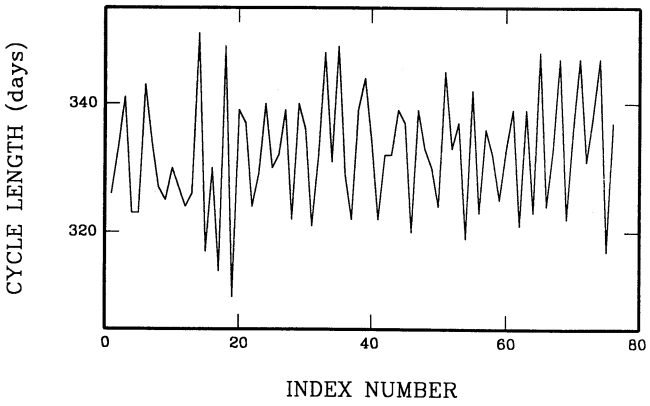

Figure 1. Seventy-six successive times between light maxima for the long-period pulsating star Omicron Ceti (Mira) (Campbell 1955; Mattei, Mayall & Waagen 1990: these data are also available from the AAVSO in machine-readable form; elsewhere in the paper this source will be referred to as AAVSO 1990). The mean period is 332.2 d, and the apparent standard deviation of the scatter is 9.3 d.

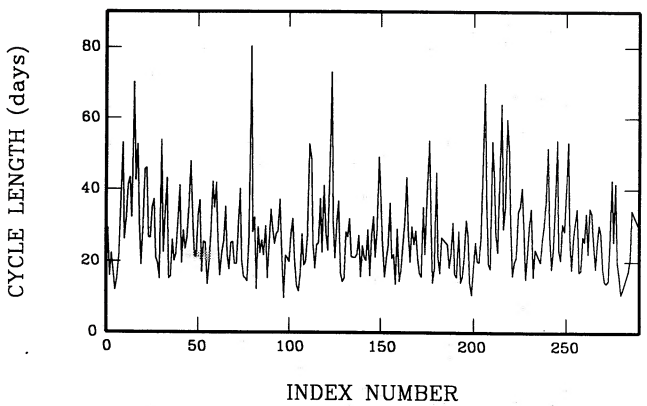

Figure 2. Times between the onset of eruptions of the dwarf nova VW Hydri ($N=281$ values; see Bateson 1977). The series mean is 27.7 d, and the apparent standard deviation of scatter is 11.6 d.

here. Also discussed are the identification of outlying observations, data transformations and the topic of missing observations. The closing paragraphs of the paper are devoted to a summary of a recommended strategy for the analysis of indexed time series.

It is perhaps worth remarking on the philosophy underlying this series of papers. It is well known that the best model obtained for a time series by use of a particular statistical approach is not necessarily physically meaningful. An oft-quoted statement is that 'all models are wrong but some are useful' (Box 1979). This points to one way in which results obtained by the methods below (and indeed all other statistical methods) may be interpreted: statistical models provide *concise summaries of certain features* of the data. Any physical model of the astronomical system must, of course, be capable of reproducing these features.

Sampling errors are responsible for a related problem. Often, very large sets of observations are required to identify 'correctly' an underlying statistical model, and to estimate its parameters accurately. It may then be very useful to analyse a number of related data sets, e.g. the distribution of quiescent intervals of *all* well-observed dwarf novae, in order to formulate a rather precise model for the class of observations. Use will be made of this approach in later papers in this series.

2 STATIONARY SERIES

A point of extreme importance regarding indexed time series is that successive observations are usually related. The autocorrelation function (acf) of a stationary series summarizes the structure of its short time-scale dependence. The acf is defined by

$$r_k=\frac{c(k)}{c(0)}, \qquad c(j)=\frac{1}{N}\sum_{t=1}^{N-j}(y_t-\bar{y})(y_{t+j}-\bar{y}), \qquad j=0,1,2,\ldots,J, \tag{1}$$

where y_t is the series value at index or epoch t and $\bar{y}$ is the mean over all N observations. The maximum lag J to which autocorrelations are calculated is usually about $N/4$ (Box & Jenkins 1976). The empirical power spectrum (or periodogram),

$$I(\omega)=\frac{1}{N}\left|\sum_{t=1}^{N}(y_t-\bar{y})\,\mathrm{e}^{-2\pi\mathrm{i}\omega}\right|^2=\sum_{j=-N+1}^{N-1}c(j)\,\mathrm{e}^{-2\pi\mathrm{i}j\omega}, \tag{2}$$

is a closely related quantity, defined in the frequency domain. In this paper, $0<\omega<1$ and, because $I(\omega)$ is symmetrical about $\omega=1/2$, it will only be shown over the domain $0<\omega<1/2$ in plots. Inversion of (2) gives, for $j=0,1,\ldots,N-1$,

$$c(j)=\int_0^1 \mathrm{e}^{2\pi\mathrm{i}j\omega}I(\omega)\,\mathrm{d}\omega=\frac{1}{N}\sum_{k=0}^{N-1}\mathrm{e}^{2\pi\mathrm{i}jk/N}I(k/N) \tag{3}$$

(see Priestley 1981). This last form for $c(j)$ is particularly useful for numerical computation using a fast Fourier transform algorithm.

We digress briefly to note that $c(j)$ defined in (1) is known in statistical parlance as the sample or estimated covariance at lag k. The 'true' or 'ensemble' covariance is $\lim_{N\to\infty}c(j)=\gamma(j)$. The corresponding analogue of r_k is $\rho_k=\gamma(k)/\gamma(0)$. It is also common

practice in statistics to denote $\gamma(0)$ by σ^2 (the ensemble variance) and its estimator $c(0)$ by s^2 (the sample variance). The ensemble power spectrum is

$$S(\omega)=\sum_{j=-\infty}^{\infty}\gamma(j)\,e^{-2\pi i j\omega}.$$

It should be pointed out, however, that $I(\omega)$ does not tend to $S(\omega)$ as the sample size N increases. [In statistical language, $I(\omega)$ is not a 'consistent' estimator of $S(\omega)$.] Some local smoothing of $I(\omega)$ must be performed in order to obtain a consistent estimate of $S(\omega)$; see Priestley (1981). For later reference, we also note that, for uncorrelated data [$\gamma(j)=0$ for $j\neq 0$],

$$S(\omega)=\sigma^2 \qquad \text{for all } \omega.$$

In this case, then, $c(0)$ provides a consistent estimate of $S(\omega)$.

Fig. 3 shows the acfs of the data of Figs 1 and 2. Also shown are the 95 per cent confidence bands for no correlation. Both series have highly significant autocorrelations at a lag of one time unit, negative for the Mira data (a long pulsation cycle is preferentially followed by a short cycle, and vice versa), and positive for the VW Hydri data (a long quiescent period tends to be followed by another long period, and a shorter-than-average rest period by another short period). The standard deviation of the autocorrelation is given, to good approximation, by $N^{-1/2}$. Note, though, that these confidence limits are applicable for testing of the significance of the autocorrelation at any pre-selected lag. If the acf has been calculated at several lags, the probability of finding at least one correlation to be significant is of course increased. In order to take account of this fact, an overall or portmanteau statistic,

$$Q=N\sum_{k=1}^{J} r_k^2,$$

can be used. Q is distributed approximately as χ^2 with J degrees of freedom. Non-significant values of Q then suggest that the series has no substantial correlation structure. The Q-statistics for the first $J=10$ autocorrelations of the data of Figs 3(a) and (b) are significant at the 2.5 per cent and better than 0.5 per cent levels respectively. (The high significance levels are of course due to the large correlations at lag 1.)

As a further example of the use of the acf, Fig. 4 shows the acfs of the successive time intervals between fadings of four R Cr B stars. None of the correlations is significant, and it is therefore not surprising that the Q-statistics are also insignificant: significance levels are 0.71 for S Aps ($N=20$, $J=5$), 0.24 for RY Sgr ($N=14$, $J=4$), 0.80 for R Cr B ($N=43$, $J=11$), and 0.37 for SU Tau ($N=27$, $J=7$). It may be concluded that there is no evidence for correlation structure in these series. The result is of some interest. Howarth (1976, 1977, 1978) showed that the fadings have Poisson distributions. His analysis implicitly assumed that successive intervals between fades are independent (see also Rozenbush 1985). Detre (1969) had earlier expressed doubts about the validity of this assumption. The acfs calculated here indicate that independence may in fact be a reasonable assumption; however, see also the discussion below.

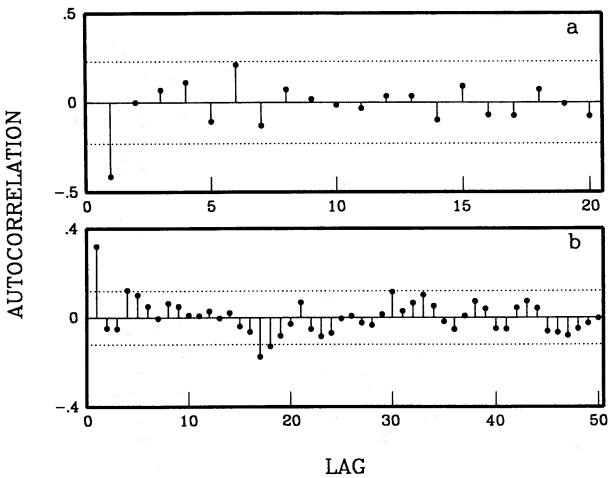

Figure 3. (a) The acf of the Mira data in Fig. 1. Note the large negative autocorrelation at lag 1. (b) The acf of the VW Hydri data in Fig. 2. Note the large positive autocorrelation at lag 1.

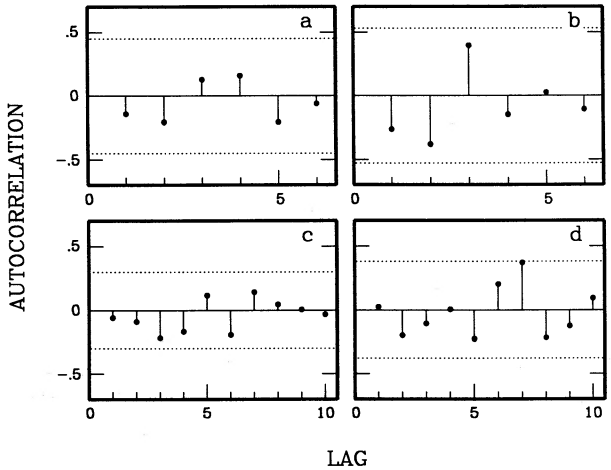

Figure 4. The acf of the intervals between fades of four R Coronae Borealis stars: (a) S Apodis (Howarth 1976); (b) RY Sagittarii (Howarth 1978); (c) R Coronae Borealis (Howarth 1977); (d) SU Tauri (Howarth 1977).

In the case of data with no correlation structure, such as those of the intervals between fadings of the aforementioned R Cr B stars, an appropriate time series model is

$$y_t = \varepsilon_t,$$

where ε_t is used throughout this series of papers to denote a stationary random process. In the case of data with significant correlation structure, such as those in Figs 1 and 2, use can be made of very general linear models extensively discussed by Box & Jenkins (1976). In their scheme, the basic building blocks are autoregressive processes of order p:

$$z_t = \sum_{i=1}^{p} \alpha_i z_{t-i} + \varepsilon_t,$$

and moving-average processes of order q:

$$z_t = \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t.$$

These are conveniently referred to as AR(p) and MA(q) processes. In these definitions, $z_t = y_t - \mu$, μ being the series mean. In other words, the processes are defined for zero-mean series z_t, obtainable by subtracting the series mean from each observation y_t. One may think of an AR(p) process as one in which the observation at epoch t is conditioned by the series values up to p periods in the past. An MA(q) process, on the other hand, can be thought of as one in which random 'shocks' or 'innovations' ε_t which enter the series at epoch t continue to exert an influence over the following q time periods. The coefficients α_k and β_k specify the impact of z_t and ε_t at epoch $t+k$.

Even more generally, the two types of model may be combined to give an autoregressive moving-average model of order (p, q), or ARMA(p,q) model:

$$z_t = \sum_{i=1}^{p} \alpha_i z_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t. \tag{4}$$

Note that all these specifications contain a completely random element ε_t, uncorrelated with any other $\varepsilon_j (j \neq t)$.

The immense popularity which ARMA models have enjoyed amongst time-series analysts can in part be ascribed to a theorem due to Wold (1938) [this has been discussed in the astronomical literature by Scargle (1981); see also Priestley (1981, section 10.1.5)]. Wold's decomposition theorem states that *any* stationary time series can be decomposed into two parts: a predictable or deterministic process (i.e. one whose future values can be predicted without error from information on its past values), and an MA process. This may be supplemented by the observation that any MA process also has an AR representation, which may be of infinite order (e.g. Box & Jenkins 1976). It is therefore of course also possible to use a mixture of AR and MA processes to model the stochastic component of stationary time series. Series can usually be adequately approximated by finite-order ARMA processes. It is of course convenient to choose a representation with as few ARMA parameters as possible. In fact, one of the guiding principles in time-series model selection is that of parameter parsimony. This idea has been given concrete form by the introduction of so-called information criteria, which are essentially residual sums of squares with penalties introduced for the number of estimated parameters. In these papers, use will be made of the Akaike information criterion

$$AIC = \frac{2P}{N} + 2\log(s),$$

and the Bayes information criterion

$$BIC = \frac{P \log N}{N} + 2\log(s),$$

where P is the number of fitted parameters and s is the residual standard deviation.

It is of some interest to contrast the conventional ARMA modelling procedure used in the present series of papers with that proposed by Scargle (1981). Scargle generalized the conventional ARMA model theory to include acausal or two-sided processes, e.g.

$$z_t = \sum_{j=-r}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t,$$

where in general $r>0$. It is notable that future innovations, i.e. ε_{t+1}, $\varepsilon_{t+2},\ldots,\varepsilon_{t+r}$, also contribute to the current z_t. Scargle interpreted MA processes as filters which act on delta function spikes (the innovations), spreading them out over time. In his acausal processes, the filters thus anticipate the innovations. The advantage of his method is that better statistical descriptions of data may be obtainable, as the parameter space is less restricted. The disadvantage is that physical interpretation of the results is more problematical. The authors hasten to point out, though, that to obtain even a causal model which is excellent from a statistical point of view is no guarantee that one is accurately modelling any underlying physical process. Scargle's second major generalization was that processes or filters are not necessarily taken to be of 'minimum delay' form. 'Minimum delay' means that $\sum_{j=1}^{q}\beta_j^2$ is a minimum compared to all other MA processes which have the same autocorrelation structure as the given data. The sizes of the MA coefficients in minimum delay processes thus decay monotonically with increasing lag. This will be of importance when the physical system giving rise to the data has a delayed full response, so that its maximum reaction to innovations happens some lags after an initially small response. The data analysed in the current series of papers have very short memory spans, so it is not necessary to introduce this generalization.

Finally, Scargle (1981) did away with the usual assumption that innovations are Gaussian. Up to this point of the paper, no distributional assumptions about the ε_t have been made; it has merely been stated that they are uncorrelated. As Scargle pointed out, there may be some phase information still contained in the set of innovations, even though they are uncorrelated. This is because, in general, 'uncorrelated' is not equivalent to 'independent', so that data may still be dependent even though the acf has no significant features – the dependence will manifest itself in the higher order moments of the data. For this reason, Scargle concentrated on criteria for innovation independence rather than on uncorrelatedness; see also Chan & Tran (1992). Fortunately, for Gaussian data the issue does not arise – lack of correlation implies independence in this very special case. In much of what follows, it is implicitly assumed that the data under scrutiny are Gaussian; when this assumption is doubtful, it will be pointed out. The reader will now realise that independence of the intervals between fades of the R Cr B stars has not been definitely shown: the intervals are not normally distributed, so the lack of correlation established above does not necessarily imply independence. The issue will not be pursued any further here, being somewhat outside the scope of this paper.

The process of finding an appropriate ARMA model for a time series is referred to as 'model identification' in the statistics literature. The acf is one of the principal tools used in model identification. The interested reader is referred to Box & Jenkins (1976) or any modern text on time-series analysis for details. The second phase in the model-fitting sequence is the estimation of the parameters of the model, i.e. the values of the α_i and β_j, and of the innovations ε_t. This can be conveniently done using commercially available software such as GENSTAT, BMDP, CSS, STATGRAPHICS, etc. Finally, the fitted model must be evaluated statistically ('diagnostic checking'). This entails, amongst other things, testing the significance of the fitted parameters and calculating the acf of the estimated ε_t to establish whether uncorrelatedness is a reasonable assumption. The information from the diagnostic checking stage may lead one to apply the above sequence of steps to alternative models.

The modelling techniques discussed above were applied to the data of Figs 1 and 2. The data of Fig. 1 are fitted about equally well by an AR(1) model with $\alpha_1=-0.41$ and an MA(1) model with $\beta_1=-0.42$. The acf of the residuals shows no significant features, and the Q-statistic is not significant either. It may be concluded that there is no information on the correlation structure of the data remaining in the residuals. The VW Hydri data can be fitted by an AR(4) model with $\alpha_1=0.38$, $\alpha_2=-0.15$, $\alpha_3=-0.03$ and $\alpha_4=0.15$, with all coefficients except α_3 significant at better than the 2 per cent level. A more parsimonious fit is given by an MA(1) model with $\beta_1=0.36$; although the acf of the residuals is rather large at small lags, no individual value is significant at the usual 5 per cent level, and the Q-statistic for the first 10 autocorrelations is only significant at the 24 per cent level. The residual standard deviations are 10.88 and 10.78 respectively for the two models, giving information criteria $AIC=4.78$ for both models [$P=1$ for the MA(1) model, $P=3$ for the AR(4) model with α_3 set to zero]. The Bayes information criterion is 4.79 for the MA(1) model and 4.82 for the AR(4) model. These results confirm that the MA(1) model may be preferable, particularly if it is borne in mind that the AIC tends to encourage overparametrization (Harvey 1989, p. 80).

The next three sections of this paper are devoted to non-stationary time series.

3 RECURSIVE MODELLING

Koen (1992) suggested the recursive updating of the statistical model as a means of dealing with non-stationarity. This has the advantage that not only the mean, but also the correlation (ARMA) structure of the model, may be tracked over time. There has been considerable recent interest in this type of modelling in the statistics literature, where it is variously referred to as dynamic linear modelling, structural modelling, Kalman filtering and state-space modelling. The term state-space model (SSM) will be used in this paper.

At the heart of the SSM are two equations, the observation equation

$$y_t=\boldsymbol{F}_t\boldsymbol{\theta}_t+\varepsilon_t \tag{5}$$

and the system or state equation

$$\boldsymbol{\theta}_t=\mathbf{G}\boldsymbol{\theta}_{t-1}+\boldsymbol{\eta}_t. \tag{6}$$

The observation equation is simply a statistical model of one of the previously encountered forms, e.g. equation (4). In general, the elements of the vector $\boldsymbol{F}_t$ are variables which are known at the current epoch t. This vector may thus, for example, contain $y_{t-1}, y_{t-2}, \ldots$. The matrix $\mathbf{G}$ contains only constants. The observation equation is a generalization of previous model equations, in that the parameters of the model are permitted to be time-dependent. The possible time dependence of the parameters is described by the system equation. The vector $\boldsymbol{\theta}_t$ contains the current values of the model parameters, which are related deterministically to parameter values at the previous time period through the matrix $\mathbf{G}$ which has fixed entries. It is notable that the elements of $\mathbf{G}$ may be unknown and hence need to be estimated. The vector $\boldsymbol{\eta}_t$ is added to encompass the possibility of stochastic variations in $\boldsymbol{\theta}$.

As an example, an autoregressive model in which the AR coefficients may vary stochastically can be cast into the SSM form

$$y_t = \sum_{i=1}^{p} \alpha_{it} y_{t-i} + \varepsilon_t = [y_{t-1}\ y_{t-2}\ \ldots\ y_{t-p}] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}_t + \varepsilon_t,$$

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}_t = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}_{t-1} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_p \end{bmatrix}_t .$$

This particular form of the equation applies to series with zero means; a more general form of the observation equation with $\mu \neq 0$ is

$$y_t - \mu = \sum_{i=1}^{p} \alpha_{it}(y_{t-i} - \mu) + \varepsilon_t = [(y_{t-1} - \mu)\ (y_{t-2} - \mu)\ \ldots\ (y_{t-p} - \mu)] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}_t + \varepsilon_t. \tag{7}$$

Of particular interest is the case where the series mean may be variable. Unfortunately, the model

$$y_t - \mu_t = \sum_{i=1}^{p} \alpha_{it}(y_{t-i} - \mu_{t-i}) + \varepsilon_t,$$

where μ_t is subject to unknown variation, cannot be written in the form (5). The same is true of models with variable MA coefficients. Such models are in fact non-linear as far as their SSM representations are concerned, and must be linearized if they are to be written in standard SSM form. Ledolter (1979) provides the relevant linearization; his results are reproduced below.

$$y_t \approx \boldsymbol{F}_t(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})_t + \varepsilon_t = [\hat{y}\ u\ v_1\ \ldots\ v_p\ w_1\ \ldots\ w_q] \begin{bmatrix} 1 \\ \mu - \hat{\mu} \\ \alpha_1 - \hat{\alpha}_1 \\ \vdots \\ \alpha_p - \hat{\alpha}_p \\ \beta_1 - \hat{\beta}_1 \\ \vdots \\ \beta_q - \hat{\beta}_q \end{bmatrix}_t + \varepsilon_t, \tag{8}$$

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})_t = \mathbf{G}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})_{t-1} + \boldsymbol{\eta}_t,$$

where $\hat{x}$ denotes an estimate of the parameter x at epoch t based on the observations up to epoch $t-1$. The values of $\hat{\mu}$ and the ARMA coefficients $\hat{\alpha}_i, \hat{\beta}_j$ are the recursively calculated values at time $t-1$, while

$$\hat{y}_t = \hat{\mu} + \sum_{i=1}^{p} \hat{\alpha}_i y_{t-1} + \sum_{j=1}^{q} \hat{\beta}_j \hat{\varepsilon}_{t-j}.$$

The residuals are estimated by

$$\hat{\varepsilon}_{t-j} = y_{t-j} - \hat{y}_{t-j}.$$

Defining relations for other elements in $\boldsymbol{F}$ are

$$u=\left(1-\sum_{i=1}^{p}\hat{\alpha}_i\right)\left(1+\sum_{j=1}^{q}\hat{\beta}_j\right)^{-1},$$

$$(v_k)_t=\sum_{i=1}^{p}\hat{\alpha}_i(v_k)_{t-i}+\hat{\varepsilon}_{t-k},$$

$$(\omega_k)_t=-\sum_{j=1}^{q}\hat{\beta}_j(\omega_k)_{t-j}-\hat{\varepsilon}_{t-k}.$$

Having specified the observation and system equations of the SSM, it remains to describe how these are actually used to calculate the model parameters $\boldsymbol{\theta}_t$ at each index t. The following distributional assumptions are made:

$$\varepsilon_t\sim N(0,\sigma^2),\qquad \boldsymbol{\eta}_t\sim N(\boldsymbol{0},\sigma^2\mathbf{V}),$$

i.e. the $\varepsilon_t (t=1,2,\ldots,N)$ are identically and independently distributed as normal (Gaussian) variates with zero mean and variance σ^2; likewise for the $\boldsymbol{\eta}_t$, where the variance of the ith component $(\eta_i)_t$ is $\sigma^2 V_{ii}$. It is further assumed that the distribution of the system vector $\boldsymbol{\theta}$ is known when observations commence: $\boldsymbol{\theta}_0\sim N(\boldsymbol{m},\mathbf{C})$. Estimates of $\boldsymbol{\theta}$ can be updated recursively by a set of equations known as the Kalman filter:

$$\hat{\boldsymbol{\theta}}(t|t-1)=\mathbf{G}\hat{\boldsymbol{\theta}}(t-1|t-1),$$

$$\mathbf{P}(t|t-1)=\mathbf{G}\mathbf{P}(t-1|t-1)\mathbf{G}'+\mathbf{V},$$

$$\hat{\boldsymbol{\theta}}(t|t)=\hat{\boldsymbol{\theta}}(t|t-1)+\boldsymbol{k}(t-1)[y_t-\boldsymbol{F}_t\boldsymbol{\theta}(t|t-1)],$$

$$\mathbf{P}(t|t)=\mathbf{P}(t|t-1)-\boldsymbol{k}(t-1)\boldsymbol{F}_t'\mathbf{P}(t|t-1),$$

$$\boldsymbol{k}(t-1)=\mathbf{P}(t|t-1)\boldsymbol{F}_t[1+\boldsymbol{F}_t'\mathbf{P}(t|t-1)\boldsymbol{F}_t]^{-1}\equiv\mathbf{P}(t|t-1)\boldsymbol{F}_t f_t^{-1},$$

where $\hat{\boldsymbol{\theta}}(0|0)=\boldsymbol{m}$ and $\mathbf{P}(0|0)=\mathbf{C}$. The notation $\hat{x}(i|j)$ denotes an estimate of x at epoch i, given observations at times up to and including j. The vectors $\hat{\boldsymbol{\theta}}_t$ are the estimates of the system parameters, while $\sigma^2\mathbf{P}(t|t)$ is the covariance matrix of $\hat{\boldsymbol{\theta}}$. The vector $\hat{\boldsymbol{\theta}}(t|t)$ is made up of the forecast $\hat{\boldsymbol{\theta}}_t(t|t-1)$ and the modifying term

$$\boldsymbol{k}(t-1)[y_t-\boldsymbol{F}_t\boldsymbol{\theta}(t|t-1)]=\boldsymbol{k}(t-1)(y_t-\hat{y}_t)=\boldsymbol{k}(t-1)\nu_t. \tag{9}$$

The vector $\boldsymbol{k}$, known as the Kalman gain vector, thus weights the forecast error ν_t. It is notable that $\nu_t\sim N(0,\sigma^2 f_t)$. The Kalman filter equations have been derived in a number of different ways: see Kalman (1960), Duncan & Horn (1972) and West & Harrison (1989).

It is of course rather unlikely that the parameters σ^2, $\mathbf{V}$, $\boldsymbol{m}$ and $\mathbf{C}$ will be known. Furthermore, the matrix $\mathbf{G}$ may in general also contain unknowns associated with deterministic changes in $\boldsymbol{\theta}$. Usually, the specific functional form of such changes (e.g. long-term trends or abrupt changes in the mean or ARMA coefficients of the series) will not be known. Fortunately, it is possible to model deterministic changes in $\boldsymbol{\theta}$ as a succession of stochastic changes (see, for example, the simulation results of Koen 1992). This approach allows one to treat $\mathbf{G}$ as a matrix of constants. In particular, for Ledolter's form of the SSM, $\mathbf{G}$ is the identity matrix. The parameters σ^2 and $\mathbf{V}$ pertain to all observations taken over the time-span of the series, and therefore must be estimated rather accurately. By contrast, the values of $\boldsymbol{m}$ and $\mathbf{C}$ only have a bearing on the first few values of $\boldsymbol{\theta}$. Although it is possible to deal with unknown $\boldsymbol{m}$ and $\mathbf{C}$ formally by the so-called diffuse Kalman filter (e.g. De Jong 1991), it is simpler (and for present purposes sufficient) to use 'reasonable' values and to discard the first few calculated values of $\boldsymbol{\theta}$ which will be sensitive to the arbitrary $\boldsymbol{m}$ and $\mathbf{C}$. For further discussion of this point, see Harvey (1984) and Ledolter (1981). This procedure is referred to as 'initialization' of the Kalman filter.

A maximum likelihood method can be used to determine σ^2 and the elements of $\mathbf{V}$. The analytic expression

$$\hat{\sigma}^2=\frac{1}{N-\ell}\sum_{t=\ell+1}^{N}\nu_t^2/f_t \tag{10}$$

is obtained for the estimator of the innovation variance, with ν_t and f_t as defined above. It has been assumed that the first ℓ observations are used to initialize the Kalman filter. The elements of the matrix $\mathbf{V}$ are found by maximization of the function

$$L=-(N-\ell)\log\hat{\sigma}-\frac{1}{2}\sum_{t=\ell+1}^{N}\log f_t \tag{11}$$

(Ledolter 1979). Ledolter (1981) also points out that the statistic $Q' = 2(L - L_0)$, where L_0 is the function L evaluated under the assumption that the covariance matrix $\mathbf{V}$ above is zero, may be used to test the significance of the elements of $\mathbf{V}$. A conservative test is obtained by comparing Q' to percentiles of the χ^2_k distribution, where k is the number of variable components of $\boldsymbol{\theta}$.

It is possible to re-assess the recursively calculated values $\hat{\boldsymbol{\theta}}_t$ in the light of later observations at epochs $t+1, t+2, \ldots, N$. The calculation of the revised values of the $\hat{\boldsymbol{\theta}}_t$ is referred to as 'smoothing'. Harvey (1984) gives the following scheme for smoothing backwards from epoch $t = N$:

$$\hat{\boldsymbol{\theta}}(t|N) = \hat{\boldsymbol{\theta}}(t|t) + \mathbf{A}(t)[\hat{\boldsymbol{\theta}}(t+1|N) - \mathbf{G}\hat{\boldsymbol{\theta}}(t|t)] = \hat{\boldsymbol{\theta}}(t|t) + \mathbf{A}(t)[\hat{\boldsymbol{\theta}}(t+1|N) - \hat{\boldsymbol{\theta}}(t|t)],$$

$$\mathbf{P}(t|N) = \mathbf{P}(t|t) + \mathbf{A}(t)[\mathbf{P}(t+1|N) - \mathbf{P}(t+1|t)]\mathbf{A}'(t),$$

$$\hat{\boldsymbol{\theta}}(t|t) = \hat{\boldsymbol{\theta}}(t|t-1) + \boldsymbol{k}(t-1)[y_t - \boldsymbol{F}_t\hat{\boldsymbol{\theta}}(t|t-1)],$$

$$\mathbf{A}(t) \equiv \mathbf{P}(t|t)\mathbf{G}'\mathbf{P}^{-1}(t+1|t) = \mathbf{P}(t|t)\mathbf{P}^{-1}(t+1|t).$$

The diagonal elements of the smoothed covariance matrix $\mathbf{P}(t|N)$ can be used to calculate approximate confidence intervals for the components of $\boldsymbol{\theta}$: the standard deviation of the jth component of $\boldsymbol{\theta}$ at epoch t is

$$\text{standard deviation } (\theta_j)_t \approx \hat{\sigma}[\mathbf{P}(t|N)]_{jj}^{1/2}$$

(Visser & Molenaar 1988). 95 per cent confidence limits for $(\theta_j)_t$ are then simply given by

$$(\theta_j)_t \pm 1.96\hat{\sigma}[\mathbf{P}(t|N)]_{jj}^{1/2}.$$

Up to this point in the paper, it has been implicitly assumed that the data sets are complete, i.e. that there are no missing observation values. However, in practice many astronomical time series are incomplete. The model specified by equation (8) can only be fitted to complete data, so that it is necessary to fill in the gaps by estimation of the missing values. This is conveniently done by using the model under consideration to predict missing values, i.e. to estimate missing y_t by $\hat{y}_t$ as defined above. A different approach is possible in cases where the correlation structure of the series may be assumed to be fixed over time. The model given by

$$y_t = \mu_t + x_t,$$

$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t, \tag{12}$$

$$\mu_t = \mu_{t-1} + \eta_t$$

is of the 'unobserved components' form, i.e. neither μ_t nor x_t is observable – only their sum y_t can be measured. In this model μ_t is the mean, while x_t is a zero-mean process which describes the correlation structure of the data. Jones (1980) has described an SSM form of this model for zero-mean processes ($\mu_t \equiv 0$). The pertinent formulae from Jones' paper are generalized to non-zero-mean processes and presented below.

The state of the process is defined by

$$\boldsymbol{\theta}_t = [\mu_t\ x(t|t)\ x(t+1|t) \ldots x(t+m-1|t)]',$$

$$m \equiv \max(p, q+1),$$

where $x(t+j|t)$ is the predicted value of the unobserved x at epoch $t+j$, based on the observations obtained prior to and including epoch t. The appropriate system equation is then

$$\boldsymbol{\theta}_t = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & \ldots & & & 1 & \\ 0 & \alpha_m & \alpha_{m-1} & \ldots & & \alpha_2 & \alpha_1 \end{bmatrix} \boldsymbol{\theta}_{t-1} + \begin{bmatrix} 1 & 0 \\ 0 & g_1 \\ 0 & g_2 \\ \vdots & \vdots \\ 0 & g_m \end{bmatrix} \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix},$$

where the g_i are found recursively from

$$g_1 = 1,$$

$$g_j = \beta_{j-1} + \sum_{k=1}^{j-1} \alpha_k g_{j-k}, \qquad j = 2, \ldots, m.$$

In these formulae, $\beta_k = 0$ for $k > q$ and $\alpha_k = 0$ for $k > p$. The observation equation can be written in matrix form as

$$y_t = \mu_t + x_t = [1\ \ 1\ \ 0\ \ 0\ \ \ldots\ \ 0] \begin{bmatrix} \mu_t \\ x(t|t) \\ x(t+1|t) \\ \vdots \\ x(t+m-1|t) \end{bmatrix}.$$

The fixed unknowns are the α_i, β_j and the variances σ_ε^2 and σ_η^2. The innovation variance σ_ε^2 follows from equation (10) as before, while the other parameters are obtained by maximization of the log-likelihood function (11). If the observation y_t at epoch t is missing, the Kalman gain vector in (9) is simply set to zero. Jones (1980) also gives a procedure for calculating an estimate of $\mathbf{P}(0|0) = \mathbf{C}$.

As an illustration, consider the plot in Fig. 5 of the times between successive light minima of the long-period variable RV Cassiopeiae (AAVSO 1990: see caption to fig. 1). The data can apparently be successfully modelled by a stationary MA(1) process, but reductions in the *AIC* and *BIC* criteria are obtainable by allowing the mean to be variable. Applying the SSM of equation (8), both *AIC* and *BIC* criteria select an MA(1) process with constant coefficient $\beta_1 \approx -0.6$ and trend in the mean μ. The statistic $Q' = 12.2$ may be compared to a χ_1^2 distribution; it is significant at better than the 0.5 per cent level. The results are confirmed by the SSM equation (12), which finds $\beta_1 = -0.64$. The raw and smoothed means, recursively calculated according to the Jones (1980) method, are shown in Fig. 6; the SSM of equation (8) gives a quantitatively similar result. Koen (1992) has given another example of the efficacy of the SSM approach for the analysis of this type of data.

4 THE DETECTION OF CHANGES IN THE SERIES MEAN

The main concern with data such as those in Fig. 5 is often the determination of whether the mean value is changing over time. With a representation of the form

$$y_t = f(t) + \varepsilon_t, \tag{13}$$

this is equivalent to the determination of whether or not f is constant as a function of time.

4.1 A simple trend test

Fig. 7 shows the times between light minima of the long-period variable X Ophiuchi (AAVSO 1990). A simple linear regression of the period y_t on epoch number t gives $y_t = 336.602 - 0.072t$, with the slope significant at the 6.5 per cent level. This is obviously a rather crude way of determining whether there may be a trend in the pulsation period. Brillinger (1989) has proposed the statistic

$$B = \frac{\sum_{t=1}^{N} c_t y_t}{\left[\hat{S}_\varepsilon(0) \sum_{t=1}^{N} c_t^2\right]^{1/2}}, \qquad c_t = \left[(t-1)\left(1 - \frac{t-1}{N}\right)\right]^{1/2} - \left[t\left(1 - \frac{t}{N}\right)\right]^{1/2}, \tag{14}$$

to test for monotone trends. Here, $\hat{S}_\varepsilon(0)$ is an estimate of the spectral density function S_ε (see Section 2) of the ε_t in (13), evaluated at zero frequency. The statistic strongly contrasts the beginning and end of any data set, and is therefore eminently suitable for testing for monotone trends. Provided that the number of observations N is fairly large, B has a normal distribution with zero mean and unit standard deviation, i.e. $B \sim N(0, 1)$. For smaller samples, a randomization test based on the unstandardized statistic $B' = \sum c_t y_t$ can be used. The interested reader is referred to Cox & Hinkley (1974) and Edgington (1980) for information and references on randomization testing.

The calculation of the estimate $\hat{S}_\varepsilon(0)$ for correlated data is discussed in Section 4.4. For the moment, however, we consider the special case of uncorrelated ε_t, an assumption which the X Ophiuchi data seem to satisfy. Then $\hat{S}_\varepsilon(0) = \sigma_\varepsilon^2$, as pointed out in Section 2. It should be noted that this quantity refers to the *unobserved* ε_t and *not* to the observed y_t. Thus one cannot use

$$c_y(0) = \frac{1}{N} \sum_{t=1}^{N} (y_t - \bar{y})^2$$

to estimate σ_ε^2, because this is corrupted by the unknown trend function $f(t)$. A number of techniques for the estimation of σ_ε^2 have been suggested.

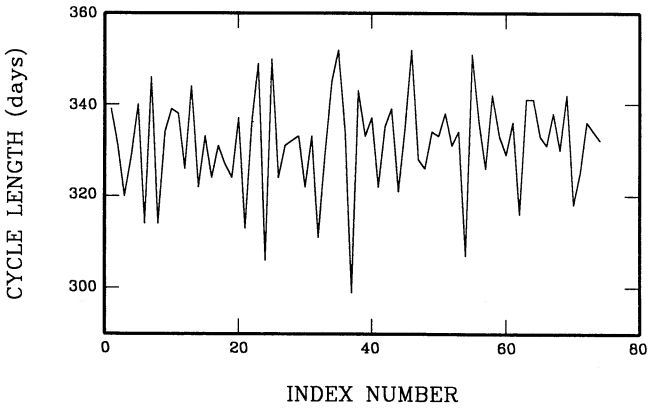

Figure 5. Intervals between successive minima of the long-period variable RV Cassiopeiae (AAVSO 1990).

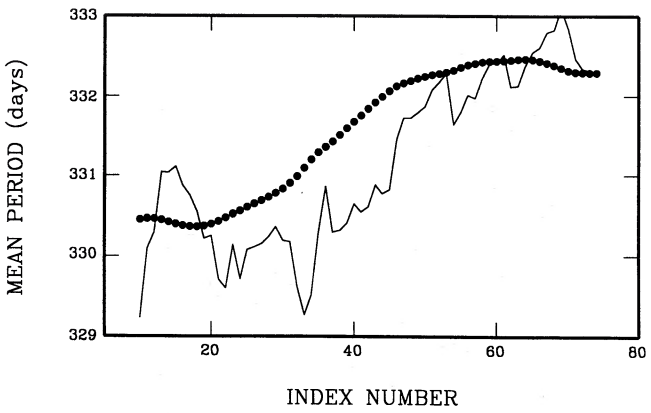

Figure 6. State-space modelling of the mean of the data in Fig. 5. The continuous line is plotted between raw estimates; the dotted line shows the retrospectively smoothed estimates.

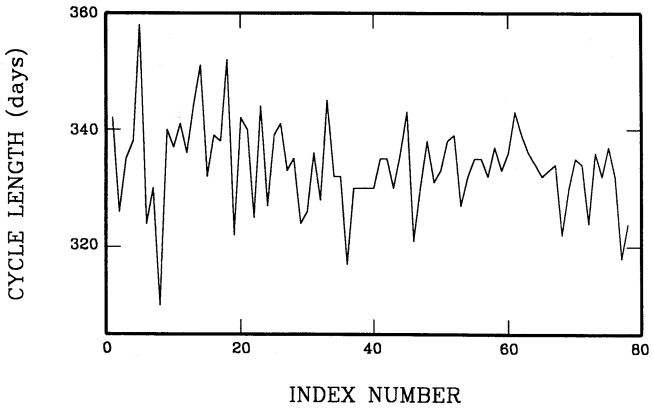

Figure 7. Intervals between successive minima of the long-period variable X Ophiuchi (AAVSO 1990).

(i) Conceptually, the simplest method is to estimate the ε_t from (13) as

$$\hat{\varepsilon}_t = y_t - \hat{f}(t),$$

where $\hat{f}(t)$ is an estimate of the trend in the data. The function $\hat{f}(t)$ may be found by, for example, smoothing of the data by a moving average or by using the SSM methods described above. One can then use the estimate

$$\sigma_\varepsilon^2 = \frac{1}{L-1} \sum_{j=1}^{L} \hat{\varepsilon}_j^2,$$

where L is the number of estimated ε.

(ii) A computationally simpler method, which is applicable when the trend is approximately linear, is to base the estimate on data differences:

$$\sigma_\varepsilon^2 = \frac{1}{2(N-1)} \sum_{j=1}^{N-1} (y_{j+1} - y_j)^2.$$

(iii) Another approach uses the facts that the periodogram can be written

$$I_y(\omega) \approx I_f(\omega) + I_\varepsilon(\omega)$$

[neglecting a term which involves the cross-spectrum of $f(t)$ and ε_t], and that for smooth functions $f(t)$ its spectrum has virtually all its power concentrated at frequencies $0 < \omega < L/N$ and $(1 - L/N) < \omega < 1$ for L some small integer close to unity. Then, for all but these frequencies,

$$I_y(\omega) \approx I_\varepsilon(\omega)$$

so that, from (3),

$$\sigma_\varepsilon^2 \approx \frac{1}{N}\sum_{k=L}^{N-1} I_y(k/N).$$

An appropriate value of L can usually be found by comparison of the low-frequency behaviours of I_y and of the periodogram of an estimate of f [such as that described under (i) above].

Application of the three methods detailed above to the data of Fig. 7 leads to standard deviations $\sigma = 7.32$, 7.90 and 7.66 for methods (i), (ii) and (iii) respectively. [A moving average of length 7 was used in method (i), and the periodogram was averaged over abscissae $L \geq 3$ in method (iii). Neither of these two values is critical.] The corresponding values of Brillinger's statistic are $B = -2.32$, -2.15, -2.22. Testing the null hypothesis of no trend against the one-sided alternative of a decreasing mean period, the significance level is 1.6 per cent or better; against the two-sided alternative of any trend the significance level is then of course 3.2 per cent or better. The significance levels were also checked by a randomization test based on 5000 permutations of the observations: probabilities were 1.5 and 3.3 per cent for one- and two-sided alternatives respectively, quite close to the normal probability values found before. Comparison with the 6.5 per cent significance level of the linear least-squares fit clearly demonstrates the power of Brillinger's statistic.

There are, however, situations in which Brillinger's statistic shoud be applied with extreme caution. Fig. 8, showing times between pulsation minima of the long-period variable SV Scorpii (AAVSO 1990), is a case in point. Values of B ranging from 3.82 to 4.07 are found, depending on which variance estimator is used. The corresponding significance level for the two-sided alternative is about 0.013 per cent. Should one decide, though, that the first observation is an outlier (see Section 6.2) and discard it, values of B of between 1.35 and 1.43 are obtained, with highest significance 15 per cent (two-sided)! The level found by randomization (5000 trials) is 18.2 per cent. The problem is of course due to the extreme weighting of the observations at the beginning and end of the data series by the values of c_t given in (14) (see Brillinger 1989 for a graph of c_t against t). In fact, it may be more accurate to typify Brillinger's test as one which is sensitive to differences between the initial and final levels of the data, rather than a trend test per se. Deletion of the first observation leaves a series in which the first three observations are higher than the last, and Brillinger's test fails. The situation is easily remedied by introducing the weights $c_t = 2t - N$ in place of the c_t above, giving a new statistic B_*. Values of B_* for the full data set of Fig. 8 range from 2.84 to 3.03 (two-sided significance level at least 0.23 per cent), and from 2.08 to 2.18 for the reduced set (significance level better than 1.9 per cent). The corresponding randomization results from 5000 trials are 0.7 and 2.9 per cent.

In conclusion, the statistic B_* based on linear weights is probably in general to be preferred to B because of its lower sensitivity to outlying observations at the ends of observed series.

A number of complications associated with simple tests for trend deserve mention. Trends are of course not necessarily monotonic. This bedevils a number of tests for trends which assume monotonicity, albeit implicitly in some cases. The Brillinger (1989) method, for example, is not able to identify quadratic changes in the mean if the initial and final levels of the data do not differ substantially. In addition, almost all tests for either gradual or abrupt changes in the mean assume that the time-series values are serially independent. D'Astous, Hipel & McLeod (1979) have given a beautiful illustration of how a stationary autocorrelated process may appear to have a changing mean. The implication is that special attention should be paid to checks of whether or not the time-series values exhibit autocorrelation. In the presence of such correlation, one of the estimates of $S_\varepsilon(0)$ discussed in Section 4.4 should replace the $c_\varepsilon(0)$.

It is noted in passing that the SSM methodology is to some extent impervious to the three problems discussed above. However, it is an estimation rather than a test procedure, and as such may not be very sensitive, particularly to small abrupt

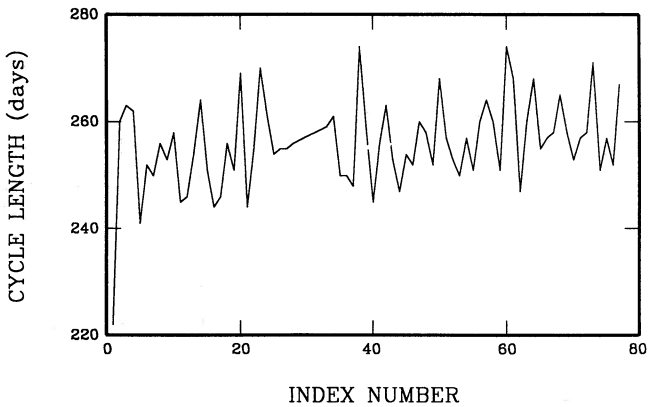

Figure 8. Intervals between successive minima of the long-period variable SV Scorpii (AAVSO 1990).

changes in the mean. We turn now to a useful general technique for studying changes in the function f of equation (13) which may be non-monotonic.

4.2 CUSUM analyses

If f were indeed constant, its value could be estimated by $\bar{y}$, the data mean. A plot of

$$C_k = \sum_{t=1}^{k} (y_t - \bar{y}), \qquad k = 1, 2, \ldots, N-1$$

against k is known as a cumulative sum (CUSUM) chart. A systematic tent-shaped deviation from the horizontal of sufficiently large magnitude suggests a non-constant function f.

Fig. 9 shows times between successive light maxima in the pulsation cycle of the Mira star T Centauri (AAVSO 1990). The corresponding CUSUM chart is shown in Fig. 10. The question, of course, is whether the deviation from the line $C_k = 0$ is 'statistically significant'. In order to answer this, one must first standardize the CUSUM values to accommodate the intrinsic variability in the stationary process ε_t. This is given by

$$\tau_N^2 = \mathrm{var}\left(\sum_{t=1}^{N} \varepsilon_t\right) \approx N S_\varepsilon(0), \qquad S_\varepsilon(0) = \sum_{j=0}^{\infty} \gamma_\varepsilon(j), \tag{15}$$

where $\gamma_\varepsilon(j)$ is the covariance of the ε_t at lag j and $S_\varepsilon(0)$ is the spectral density function of the ε_t at zero frequency. Estimation of this quantity from the y_t data will be discussed in Section 4.4. The standardized CUSUMs are given by

$$C'_k = C_k / \tau_N,$$

and an intuitively obvious overall measure of the deviation from the horizontal line $C_k \equiv 0$ is

$$D_N = \max_k |C'_k|. \tag{16}$$

The distribution of D_N is (for large N) the same as that of the well-known Kolmogorov–Smirnov statistic, namely

$$\mathrm{Prob}(D_N > d) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2 d^2) \approx 2 \exp(-2d^2).$$

For the T Cen data one finds $\tau_N = 26.78$ (see Section 4.4 below) and $D_N \equiv d = 2.61$, yielding a significance level of 2×10^{-6}. This is decisively in favour of a changing mean. The importance of taking account of serial correlation in the data cannot be overemphasized. For uncorrelated data, $\gamma(j) = 0$ for all $j \neq 0$ and

$$S_\varepsilon(0) = \sigma_\varepsilon^2,$$

the variance of the ε_t process. For the T Cen data $\hat{\sigma}_\varepsilon^2 = 29$. Thus, if the serial correlation is not accounted for, $\tau_N = 70.6$ and $D_N = 0.99$ are obtained, yielding a significance level of 0.28. This leads to a conclusion contrary to that reached above.

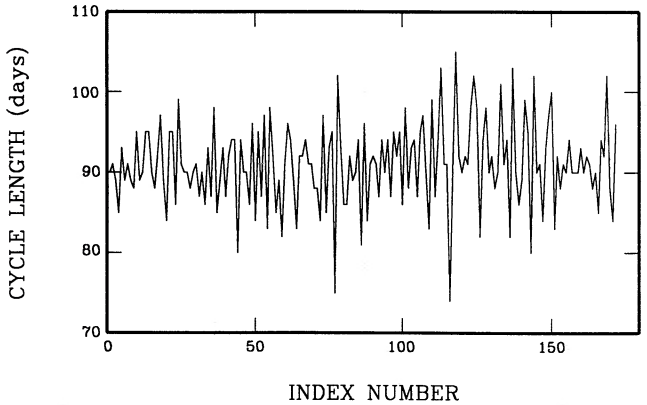

Figure 9. Intervals between successive minima of the long-period variable T Centauri (AAVSO 1990).

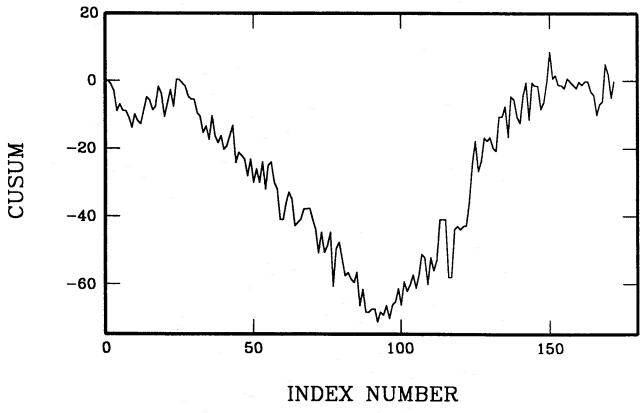

Figure 10. The CUSUM chart of the T Cen data shown in Fig. 9.

The reader must now be alerted to the fact that (16) is but one of a plethora of quantities that have been designed to detect almost any conceivable type of deviation of the series mean from constancy. However, (16) is widely known and able to cope with many types of deviation from constancy.

4.3 O − C diagrams

It is appropriate at this point to remark on the relationship between the CUSUM methodology set out above and the O − C methodology prevalent in the astronomy literature. The symbol y_t, as used in this paper, refers to the *intervals* between events (such as pulsation maxima or minima, or fades of an R Cr B star). Denote the *times* of the events by Y_0, Y_1, Y_2,.... Then

$$Y_t = Y_0 + \sum_{j=1}^{t} y_j = Y_0 + t\bar{y} + C_t, \tag{17}$$

where $\bar{y}$ is the mean of the intervals y_t and C_t is the CUSUM as defined above. The equation is in the form of a linear regression model. The traditional O − C analysis indeed postulates a regression model such as

$$Y_t = a + bt + e_t,$$

and aims to establish whether there is any deviation from this form. On the basis of some deviation in the O − C diagram, it may be decided to fit a quadratic form to the Y_k:

$$Y_t = a + bt + ct^2 + e_t. \tag{18}$$

The more scrupulous will at this point probably perform a standard statistical test in order to decide whether the coefficient c of the quadratic term is significant. These tests require that the error series e_t have a white noise character. Suppose, however, that c is in fact zero. Comparison of (18) with (17) then shows that $a \approx Y_0$, $b \approx \bar{y}$ and $e_k \approx C_k$, the latter usually being a non-stationary and highly correlated series which does not have the required white noise character. This invalidates the usual test of significance of the coefficient c in (18). Failure to recognize this fact can easily lead to the (unjustified) conclusion that, for example, c in (18) is non-zero and, consequently, that the series of intervals is non-stationary.

4.4 Estimation of variability

Estimation of the quantity $S_\varepsilon(0)$ in (15) must now be considered. If $f(t)$ in (13) were constant, this would be simple. An estimate of $S_\varepsilon(0)$, the spectral density at zero frequency, could be obtained by averaging periodogram ordinates in the immediate vicinity of the zero frequency:

$$\hat{S}_\varepsilon(0) \approx \frac{1}{L} \sum_{j=1}^{L} I_y(j/N),$$

where the periodogram $I_y(\omega)$ was defined in Section 1 of the paper and L is small compared to N; see Priestley (1981, p. 581). If $f(t)$ is non-constant but varies smoothly, its spectrum shows high power at low frequencies, the power diminishing rapidly with increasing frequency. The periodogram of the y_t-data, being essentially a superposition of the periodograms of ε_t and of $f(t)$, will therefore be close to the ε_t periodogram at all but the low frequencies (see also Section 4.1). This suggests that $S_\varepsilon(0)$ may be estimated by

$$\hat{S}_\varepsilon(0) \approx \frac{1}{L} \sum_{j=K}^{K+L} I_y(j/2N), \tag{19}$$

for K an appropriate integer close to unity. K is to be chosen so that the power of $f(t)$ is mostly contained in the frequency range $\omega = 2\pi j/N$; $j = 1, 2, \ldots, K-1$. Note, however, that this method will usually tend either to over- or to underestimate $S_\varepsilon(0)$, depending on the magnitude of the slope of $S_\varepsilon(\omega)$ for ω near zero.

Fig. 11 contains a plot of the periodogram for the T Cen data. The large power at the two lowest frequencies is evidently due to trend; the latter is shown by the 17-point moving average in Fig. 12. Fig. 13 clearly demonstrates that the lowest frequency part of the spectrum is indeed due to the data trend. A choice of $K = 3$ seems appropriate. The quantity (19) varies between 3.61 and 4.52 for L between 1 and 8. The larger of these values was chosen for the calculations in Section 4.2. A very conservative procedure would be to take $K = 1$ in all cases. This makes no allowance for the generally decreasing nature of the periodogram of the ε_t as the frequency tends towards zero. For $L = 8$, one then arrives at an estimate $\hat{S}_\varepsilon(0) = 13.0$. If this is used in the calculations in Section 4.2, one finds $D_N = 1.48$, yielding a significance level of 0.025.

The method just discussed is non-parametric, no specific assumptions having been made regarding any particular form that the theoretical power spectrum of the ε_t-process may have. [Various refinements of the method have been discussed in the

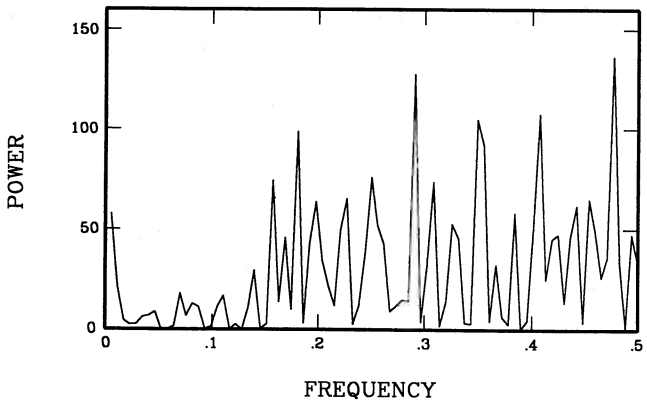

Figure 11. The periodogram of the T Cen data shown in Fig. 9.

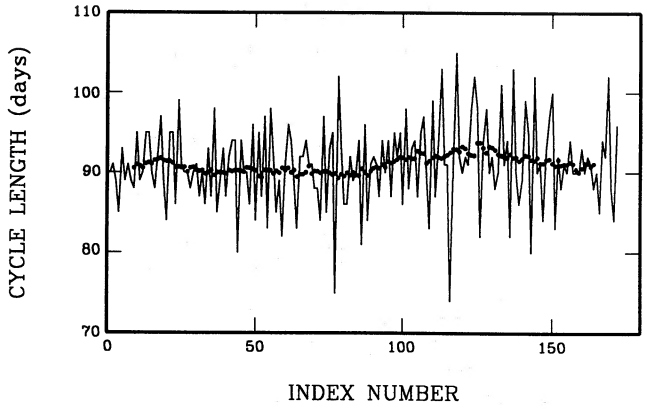

Figure 12. A 17-point moving average of the T Cen data shown in Fig. 9.

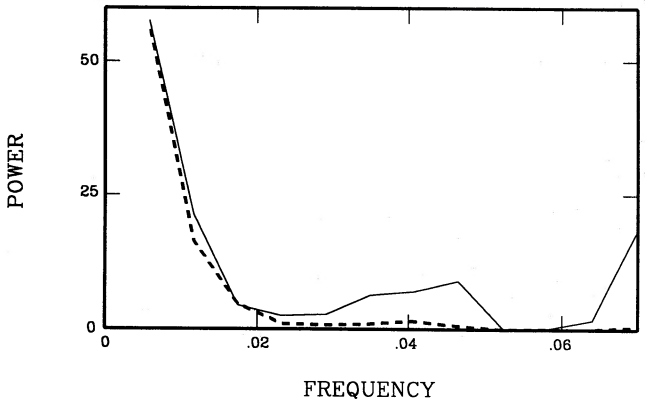

Figure 13. The low-frequency part of the periodogram in Fig. 11. Also shown is the low-frequency periodogram of the moving average in Fig. 12.

statistical literature; see e.g. Brillinger (1989) or Hart (1989).] It is also possible to use parametric methods such as those described in Section 3. Application of the SSM method to the T Cen data suggests that the ε_t have an MA(1) structure:

$$\varepsilon_t = -0.61 e_{t-1} + e_t,$$

where e_t is a white noise process with variance $\sigma^2 = 21.72$. The power spectrum of an MA(1) process is (see Priestley 1981)

$$S_\varepsilon(\omega) = \sigma^2(1 + \beta_1^2 + 2\beta_1 \cos \omega),$$

so that, for the case in hand,

$$\hat{S}_\varepsilon(\omega) = 21.72(1.36 - 1.22 \cos \omega),$$

whence $\hat{S}_\varepsilon(0) = 3.04$, which is of the same order of magnitude as the first of the two estimates computed above.

Finally, it is worth remarking that the general results presented in this section of the paper do not require the data to be Gaussian or linear.

5 HETEROSCEDASTICITY

It has been implicitly assumed in the foregoing that the variance σ^2 of the innovations is constant. Deviation from this state is referred to as 'heteroscedasticity'. The reader's indulgence is asked for what may be considered a rather lengthy discussion of this topic – it is usually ignored in any discussion of variability of astronomical time series, attention being exclusively focused on possible changes of the series mean.

In what follows, it is assumed that a time-series model has been fitted to the data and that the innovations ε_t have been estimated. Implicit in this is the assumption that the ε_t are uncorrelated (if this were not so, the time-series model fitted would have been judged unsatisfactory). Since σ^2 is estimated by $\hat{\sigma}^2 = N^{-1}\sum \varepsilon_t^2$, the evaluation of heteroscedasticity is based on consideration of the squared innovations ε_t^2. Two situations may be distinguished: either the ε_t^2 are uncorrelated, or they are not. In the latter case, the process generating the data may be of the so-called ARCH (autoregressive conditional heteroscedasticity)

type: the innovation variance at series index number t is determined by the innovation variances at preceding epochs. Symbolically,

$$E(\varepsilon_t^2|y_{t-1}, y_{t-2},\ldots)=\gamma_0+\sum_{i=1}^{r}\gamma_i\varepsilon_{t-i}^2,$$

i.e. the expected value (ensemble average) of the innovation variance at epoch t, given the prior observations y_{t-j}, can be written in the form shown. Note, though, that the expected value of the *unconditional* innovation variance is constant if the time series is stationary.

The following analogy may illuminate the ARCH concept further. Consider, for simplicity, a zero-mean AR(1) process $y_t=\alpha_1 y_{t-1}+\varepsilon_t$. If y_t is a stationary process, then $E(y_t)=E(y_{t-1})=\ldots 0$. However, the conditional expected value of the y_t is $E(y_t|y_{t-1}, y_{t-1},\ldots)=E(y_t|y_{t-1})=\alpha_1 y_{t-1}$. The conditional mean or expected value of the AR(1) process is thus variable. Identification of the behaviour of the conditional mean of a series is in fact the essence of ARMA modelling. One may think of the conditional parameters (mean and variance) of the series as describing its *local* behaviour, while the unconditional parameters typify the series *globally*.

The ARCH model is one of many types of non-linear time-series model. Non-linearity is briefly discussed in the next section of the paper, but in the interest of brevity it has been decided not to deal with specific models. The reader is referred to Engle (1982) and Weiss (1984) for more information on ARCH models. It is noted in passing that other types of non-linear series (e.g. bilinear series) may show significant features in the acf of the squared residuals of an ARMA fit (Maravell 1983; Li 1984). In the remainder of this section of the paper, it will be assumed that the squared innovations ε_t^2 are uncorrelated, and hence that any heteroscedasticity present in the data is a manifestation of non-stationarity.

A summary of a number of statistical tests for heteroscedasticity can be found in Judge et al. (1980); in the present context these tests involve either partitioning of the data and comparison of variances of the subsets, or regression of the squared innovations on the epoch number of the time series. Perhaps the best known of the tests requiring data partitioning is that due to Bartlett (1937): it is described in many standard statistical texts and will therefore not be repeated here. Application of Bartlett's test requires the ε_t to be Gaussian. For non-Gaussian innovations, a rank test of the partitioned data may be attempted (e.g. Sprent 1989, p. 117).

Clearly, many of the available tests for non-stationarity of the observations y_t may also be applied to the squared innovations. A complication is that, for ε_t Gaussian, ε_t^2 will have a χ^2 distribution, so that tests based on the assumption of normally distributed data will require the evaluation of new significance values. Brown, Durbin & Evans (1975) have proposed a graphical technique which could be used to identify shifts in the variance – it is essentially a standardized CUSUM of the squared innovations:

$$W_t=\frac{\sum_{j=1}^{t}(\varepsilon_j^2-\hat{\sigma}^2)}{N\hat{\sigma}^2}.$$

Tables for testing of the significance of large values of $|W_t|$ can be found in Durbin (1969) and Harvey (1981). The tables are appropriate for Gaussian innovations. For non-Gaussian ε_t, a randomization test based on large absolute values of $W_t'=\sum_{j=1}^{t}(\varepsilon_j^2-\hat{\sigma}^2)$ can be used. Hsu (1977) has given a simple test statistic for a single abrupt change in variance for Gaussian ε_t:

$$H=(T-0.5)/\sigma_T,$$

where

$$T=\frac{\sum_{t=1}^{N}(t-1)\varepsilon_t^2}{(N-1)\sum_{t=1}^{N}\varepsilon_t^2}, \qquad \sigma_T^2=\frac{M+1}{6(M-1)(M+2)}.$$

Significance levels for H can be found in Hsu's (1977) paper; Hsu remarks that the distribution of H is close to that of a standard normal variate for N larger than approximately 25.

The use of some of the above statistics can be illustrated by reference to the observations of the long-period variable V Librae (AAVSO 1990) given in Fig. 14. Fig. 15 shows the corresponding squared innovations $\varepsilon_t^2=(y_t-\bar{y})^2$. The maximum value of $|W_t|$ is 0.28 (significant at better than 0.5 per cent) and the Hsu (1977) statistic is $H=-3.08$ (significance level 0.1 per cent). The standardized CUSUM chart of Fig. 16 confirms the impression given by Fig. 15: an appropriate statistical model may be an abrupt change in the innovation variance at $t\approx 53$.

The significance levels were also checked by performing a randomization test. Random permutations of the ε_t were performed and the statistic W_t' determined. The process was repeated 5000 times and the percentile of the observed value of W_t'

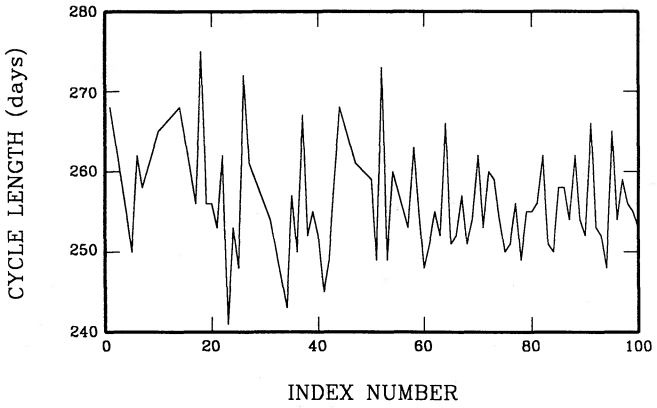

Figure 14. Intervals between successive maxima of the long-period variable V Librae (AAVSO 1990).

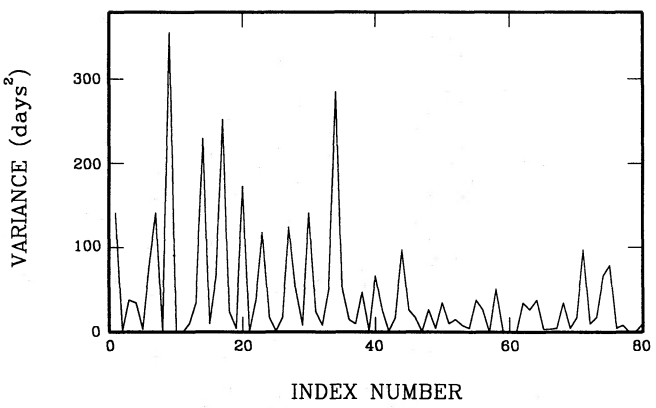

Figure 15. The squared innovations (i.e. local variance) of the V Librae data in Fig. 14.

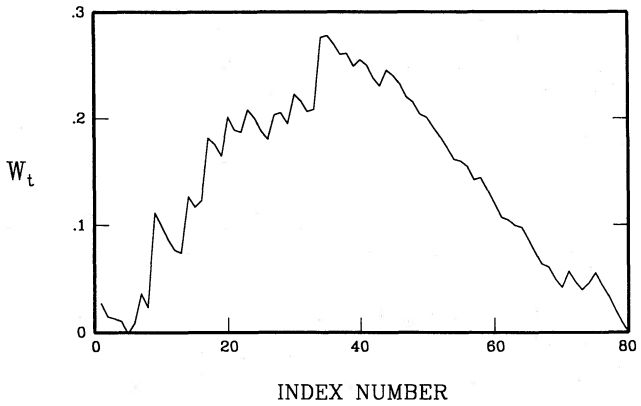

Figure 16. The CUSUM W_t of the squared innovations in Fig. 15.

with respect to the permutation values was calculated. This procedure yielded a significance level of approximately 0.3 per cent, similar to those found above.

6 NUMBER OF OTHER IMPORTANT CONSIDERATIONS

6.1 Transformations and normality

There are several reasons why transformation of time series may be considered. For example, it may sometimes be more appropriate to study $\log(y_t)$ rather than y_t itself, as when the interest is in proportional rather than absolute evolution of the series (e.g. Box & Jenkins 1976, p. 94). Heteroscedasticity may sometimes be removed by appropriate transformation. Lastly, the innovations from the transformed series may be closer to Normal (Gaussian) than those from the original series. This is advantageous both because many statistical procedures (e.g. maximum likelihood estimation of ARMA parameters) assume Normal innovations, and because one may be more confident that all meaningful information has been extracted from the data in the modelling process [see the discussion of Scargle (1981) in Section 2]. A standard reference on time-series transformations is Box & Cox (1964).

As an illustration, the innovations from the MA(1) model fit of Section 2 to the quiescent intervals of VW Hydri are not Gaussian; a chi-square test gives the significance of the deviation as better than 0.0001 per cent. Not only does an MA(1) model fit the logarithmically transformed series marginally better (smaller Q-statistic of the residuals), but the innovations are Gaussian. It is perhaps worth remarking on the fact that in the light of these results the logarithms of the quiescent intervals constitute a more 'natural' representation of the series than do the original y_t.

Statistical software packages such as those referred to in Section 2 have facilities for testing data for normality. Alternatively, the usual chi-squared or Kolmogorov–Smirnov goodness-of-fit tests can be performed. It is important to bear in mind that the residuals, and not the series itself, must be tested. The reason is, of course, that, if there is autocorrelation in the series, the assumption of independence of the y_t made by all the tests is no longer valid. Lomnicki (1961) has given a test for Gaussianity which can be applied to the y_t directly, but its small-sample performance appears to be unknown.

6.2 Outliers

Briefly, outliers are atypical observations. Outlier identification is important in the present context for at least two reasons. First, as a number of authors have pointed out, the estimates of model parameters may be severely distorted by outliers. As an example, consider the intervals between successive light minima of the long-period variable RX Lyrae shown in Fig. 17 (AAVSO 1990). Analysis of these data leads to an MA(1) model with $\beta_1 = -0.54$ (standard error 0.09) and an innovation standard deviation $\sigma = 9.32$. Leaving out the observations at epochs 62 and 63 (i.e. treating them as missing observations), one finds instead $\beta_1 = -0.34$ and $\sigma = 5.81$. A decrease in σ is of course to be expected if widely discordant observations are removed, but the two-standard-error change in the MA(1) coefficient comes as more of a surprise. The bulk of the observations are obviously better represented by the second model fitted.

The second important consequence of the presence of outliers is that they are potential further sources of information about the underlying physical system. It may be noted, for example, that the observation of the light minimum common to the two extreme intervals is given a high reliability weight by the AAVSO (Mattei, Mayall & Waagen 1990). The implication is that the outliers are probably caused by unusual behaviour of the star, rather than being due to recording errors. Further light may be shed on the situation if one is able to discriminate between different types of outlier, and pertinent definitions are now introduced.

Denote by x_t the observed series, which may contain outliers, and by z_t underlying 'normal' series, so that

$$z_t = \sum_{i=1}^{p} \alpha_i z_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t, \qquad x_t = z_t$$

ordinarily. An *additive outlier* at epoch T is defined by

$$x_t = z_t + \xi_t, \qquad \xi_t = \begin{cases} 0 & t \neq T, \\ \omega_T & t = T. \end{cases}$$

By contrast, an *innovation outlier* at $t = T$ is defined by

$$x_t = z_t, \qquad \varepsilon_t \rightarrow \varepsilon_t + \xi_t,$$

with ξ_t as above. Additive outliers may thus be thought of as disturbances in the *observations*, while innovation outliers are large deviations from the norm in the *driving noise* of the observed system. Note that, whereas an additive outlier impacts on the observed series at a single epoch T, the innovation outlier can make itself felt for many epochs after T through disturbance of z_T and ε_T directly. It is of obvious importance to distinguish between the two types of outlier.

Statistics designed to discriminate between outlier types have been suggested by Tsay (1988), Chang, Tiao & Chen (1988) and Muirhead (1986), amongst others. The statistics proposed in these papers are based on the estimated innovations, i.e. the residuals after fitting an ARMA model. Tsay (1988) and Chang et al. (1988) define

$$\lambda_{I,t} = \frac{\hat{\varepsilon}_t}{\sigma}, \qquad \lambda_{A,t} = \frac{\rho_t}{\sigma}\left(\hat{\varepsilon}_t - \sum_{i=1}^{N-t} \pi_i \hat{\varepsilon}_{t+i}\right),$$

where

$$\rho_t = \left(1 + \sum_{j=1}^{N-t} \pi_j^2\right).$$

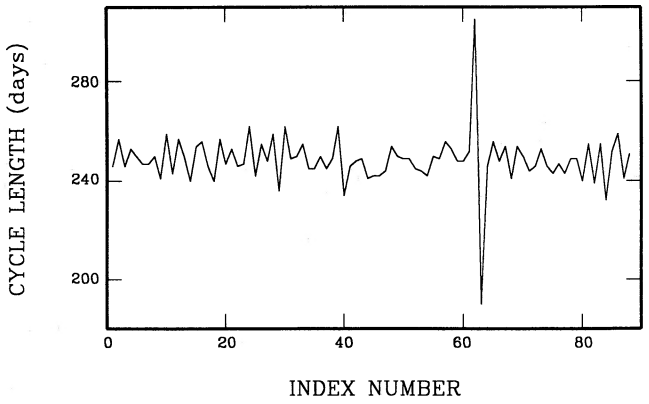

Figure 17. Intervals between successive minima of the long-period variable RX Lyrae (AAVSO 1990).

The π_j are the so-called π-weights of the ARMA process, i.e. the coefficients in an infinite-order pure AR representation of the ARMA process:

$$z_t = \sum_{i=1}^{p} \alpha_i z_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t = \sum_{i=1}^{\infty} \pi_i z_{t-i}.$$

[As an example, $\pi_j = (-1)^{j+1}\beta_1^j$ for an MA(1) process.] Observation x_t is designated an outlier if $|\lambda_{I,t}|$ or $|\lambda_{A,t}|$ is larger than 3 or 4 (depending on the required sensitivity of tests), and innovation type is decided on the basis of the larger (in absolute value) of the two statistics. The size of the disturbance ω_T as defined above is estimated variously as $\sigma\lambda_{I,T}$ or $\sigma\rho_T\lambda_{A,T}$. This may be subtracted from the observation x_T and the ARMA model refitted. Alternatively, once an observation has been identified as outlying it may simply be treated as missing, as was done above.

Muirhead's (1986) procedure is quite similar, although the initial identification of outliers is more formal: the approximate significance level $2N\Pr\{|\hat{\varepsilon}_t/\sigma| > t_{N-2}\}$ of the largest residual is calculated. (The notation t_{N-2} refers to a t-distribution with $N-2$ degrees of freedom.) Classification of outliers as innovation or additive is based on the larger of $S_{I,t}$ and $S_{A,t}$:

$$S_{I,t} = \rho_t^2 \left(\gamma \hat{\varepsilon}_t + \frac{1}{\gamma} \sum_{i=1}^{N-t} \pi_i \hat{\varepsilon}_{t+i} \right)^2, \qquad S_{A,t} = \left(\frac{1}{\gamma} \sum_{i=1}^{N-t} \pi_i \hat{\varepsilon}_{t+i} \right)^2,$$

$$\gamma = \left(\sum_{i=1}^{N-t} \pi_i \hat{\varepsilon}_{t+i} \right)^{1/2} = (1-\rho_t^2)^{1/2}.$$

A small adjustment for a bias towards classification of outliers as additive is also suggested, namely requiring that

$$S_{A,t} - S_{I,t} > \sigma^2 \ln(1+\gamma^2).$$

In this case, the disturbance ω_T is estimated by $\hat{\varepsilon}_T$ for both outlier types.

The procedures may be illustrated by reference to the successive intervals between light minima of the long-period variable RY Ophiuchi (AAVSO 1990, see Fig. 18). An MA(1) model with $\beta_1 = -0.37$ and innovation standard deviation $\sigma = 4.42$ is a good fit to the model. However, $\hat{\varepsilon}_{53}/\sigma = 4.69$, which is significant at the 0.1 per cent level according to the Muirhead (1986) test. The relevant statistics for deciding whether x_{53} is an innovation or an additive outlier are $\lambda_I = 4.69$, $\lambda_A = 4.94$, $S_I = 1.36$ and $S_A = 49.42$, all evaluated at epoch $T = 53$. The correction factor for comparison of S_I and S_A is $\sigma^2 \ln(1+\gamma^2) = 2.9$. The conclusion is that the observation at index number 53 is an additive outlier. An MA(1) model with $\beta_1 = -0.43$ and $\sigma = 3.99$ fits the data, with x_{53} removed, rather well. The largest residual (in absolute value) is at $t = 3$, with $\hat{\varepsilon}_3/\sigma = -3.99$. The Muirhead (1986) t-distribution test finds this to be significant at the 1.5 per cent level. The relevant statistics for deciding outlier type are $\lambda_I = -3.99$, $\lambda_A = -3.72$, $S_I = 34.26$ and $S_A = 1.28$. The outlier is of the innovation type. Dropping both the outliers, one finds a good fit with an MA(1) model with $\beta_1 = -0.40$ and $\sigma = 3.79$. Muirhead's t-test finds the largest residual (standardized value 2.91) to be significant at only the 56 per cent level. For this data set, the outliers do not have a profound influence on the parameter estimates; none the less, the identification of the outlier types may have implications for physical models of the stellar pulsations.

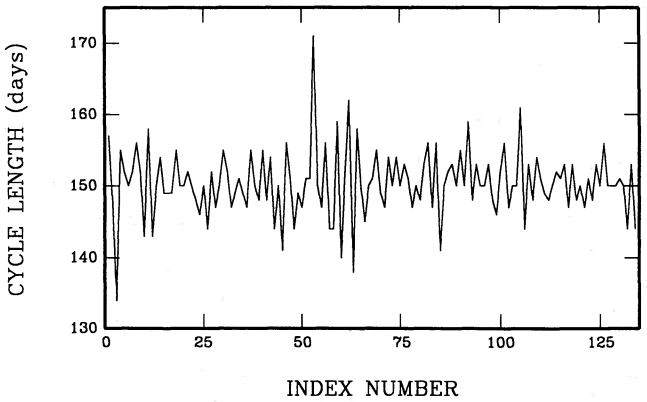

Figure 18. Intervals between successive minima of the long-period variable RY Ophiuchi (AAVSO 1990).

The statistics described above are intended for use on isolated outliers. In practice, outliers often occur in patches. Bruce & Martin (1989) have proposed the statistic

$$DV(k, t)=\frac{N}{2}\left(\frac{\sigma^2}{\sigma^2_{k,t}}-1\right)^2$$

for diagnosis of a patch of k successive outliers. Here $\sigma^2_{k,t}$ is the innovation variance estimated from the reduced data set obtained by treating the patch of k observations centred on epoch t as missing. The value of $DV(k, t)$ is compared to the 50th percentile of a χ^2_1 distribution, i.e. a critical value of 0.455. For the RX Lyrae data of Fig. 17 one finds $DV(2, 62)=108.37$, which is convincingly significant.

Implicit in much of the above discussion has been the assumption that the distribution of the innovations is Gaussian. The importance of the innovation distribution for the identification of outliers is well illustrated by the VW Hydri data of Fig. 2: the largest value of $|\hat{\varepsilon}_t/\sigma|$ after fitting an MA(1) model to the raw data is 4.75 (significant at the 0.1 per cent level according to Muirhead's test), while an MA(1) model fit to the $\log(x_t)$ leads to a maximum absolute standardized innovation of 2.95, significant at only the 97 per cent level. The difference is of course due to the distinctly non-Gaussian form of the raw data already referred to in Section 6.1.

6.3 Some remarks on the effects of missing observations

It often happens that some values in the sequence $y_1,\ldots,y_N$ are missing, i.e. are not observed. The periodogram is calculated as before by simply ignoring the missing values. One has to be careful, though, to keep track of the original indices of the observed values. Formally,

$$I_y(\omega)=\frac{1}{N'}\left|\sum(y_t-\bar{y})\exp(2\pi i t\omega)\right|^2,$$

where N' is the number of observed y_t, $\bar{y}$ is the mean of the observed y_t and the sum is over indices of observed y_t. Since this methodology is well known to astronomers, it will not be elaborated upon here.

It is perhaps worth remarking that the CUSUM method discussed in Section 4.2 requires only minor changes when missing values are present: the variance factor τ_N should be replaced by an estimate calculated on the basis of the observed spacing of the data and the CUSUMs $\sum_{t=1}^{k}(y_t-\bar{y})$ should be taken only over t-indices corresponding to observed y_t. The second paper in this series (Lombard & Koen 1993) gives a detailed CUSUM analysis of data containing missing values.

6.4 Non-linearity

This subsection presents a very brief introduction to the theory underlying non-linear time series. It is important to bear in mind that not all series can be described by the linear ARMA models discussed above. Tong (1990) presents a summary of a number of tests for non-linearity of time series.

In Section 2 it was pointed out that any stationary process can be written as the sum of a deterministic series and an infinite moving-average process:

$$z_t=\varepsilon_t+\sum_{j=1}^{\infty}\beta_j\varepsilon_{t-j},$$

where the ε_t are all uncorrelated (Wold's decomposition theorem). It was also mentioned that further information about the process may be extracted if the ε_t are dependent, although they are uncorrelated (Scargle 1981). This will be the case if the innovations are non-Gaussian or if the process is inherently non-linear. The non-Gaussian or non-linear nature of the series (or of the innovations ε_t) can be revealed by studying the higher order properties of the time series, the acf being a second-order diagnostic only. As an example, Subba Rao & Gabr (1980) proceeded from the assumption that the series may be expanded as

$$z_t=\sum_{j=-\infty}^{\infty}\beta_j\varepsilon_j,$$

i.e. a doubly infinite MA process, where the ε_j are all independent (with, as usual, zero mean). The authors showed that, under these assumptions,

$$X_{ij}\equiv\frac{|f(\omega_i,\omega_j)|^2}{f(\omega_i)f(\omega_j)f(\omega_i+\omega_j)}=\frac{\mu_3^2}{2\pi\sigma^2}$$

for all frequencies ω_i, ω_j. In this equation, $f(\omega)$ is the spectral density function, $f(\omega_i, \omega_j)$ is the bispectral density function (see e.g. Brillinger 1965), and σ^2 and μ_3 are the second and third moments of the ε_t respectively. Three distinct situations obtain. If z_t is Gaussian, then X_{ij} is identically zero. If X_{ij} is constant for all frequencies, but non-zero, the process is linear but non-Gaussian. Finally, if X_{ij} varies with frequency, the assumed form must be incorrect, i.e. z_t cannot be written as the sum of independent linear terms. The series must therefore be non-linear.

To take the discussion further, it is noted that Nisio (1960) has proved the following: for any stationary series, a sequence of series which converges to the observed series can be constructed. Each element of the constructed series is a polynomial sum of independent Gaussian variates. Symbolically,

$$y_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{i+t} + \sum_{i,j=-\infty}^{\infty} a_{ij} \varepsilon_{i+t} \varepsilon_{j+t} + \ldots.$$

If y_t is Gaussian, only the first (linear) sum of the sequence is required. Nisio's non-linear representation is more basic than Wold's linear representation, in that the ε_t in Nisio's (1960) theorem are independent, while they are only uncorrelated in Wold's (1938) theorem.

A large number of distinct non-linear forms for time series have been proposed; the interested reader is referred to Tong (1990) for a survey of some of these (see also Priestley 1988). An astrophysical application of one particular model can be found in Vio et al. (1991).

6.5 Chaos

It is important to be able to discriminate between stochastic and chaotic time series, as this provides information regarding the physical system giving rise to the observations. The necessity to keep this paper to a reasonable length does not allow the authors to do justice to the topic of testing for chaos, and it is therefore left for a later paper in the series. The interested reader is referred to Vol. 54 (2) of the *Journal of the Royal Statistical Society*, which contains a number of papers (with numerous references to previous work) dealing with tests for chaos. Attention is drawn in particular to the paper by Casdagli (1991), which shows the progression from chaotic, through non-linear stochastic, to linear stochastic series, in terms of the number of phase-space points necessary to calculate predictions from the series.

7 BASIC STEPS IN A TIME-SERIES ANALYSIS

Here, we briefly outline an approach to the initial stages of indexed time-series data analysis which we have found useful. First, a simple plot of the data against their time indices often reveals noteworthy features such as trends, possible outliers and the presence of autocorrelation. In this respect, a solid-line plot is usually more revealing than a scatterplot. Fig. 9, for instance, clearly reveals the possibility of a trend in the data, while the systematic up–down succession of values is typical of negatively autocorrelated data. Recall that the presence of both these features has been confirmed by the use of more sophisticated statistical methods. However, neither of these features stands out in Fig. 19, which is the scatterplot corresponding to Fig. 9; this underlines the necessity to experiment with different graphical representations. Secondly, a periodogram will reveal the presence of more or less strict periodicities in the data, in the form of rather distinct peaks at certain frequencies. In addition, a monotone trend in the data shows up as a distinct peak (or peaks) at the very lowest frequencies in the periodogram. Such a trend can be

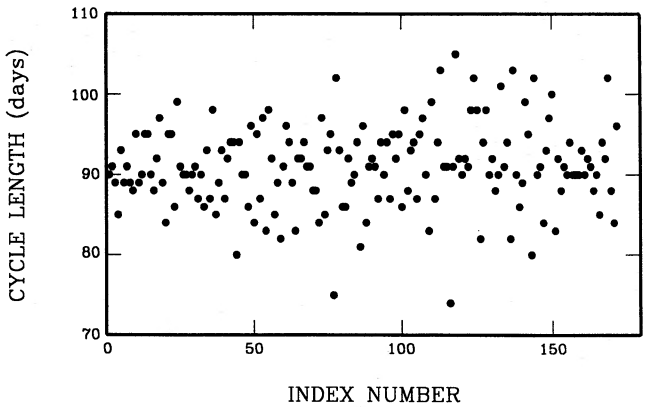

Figure 19. As in Fig. 9, but graphed in scatterplot format rather than by connecting succeeding points by solid lines.

estimated via a moving average of the data. Subtraction of the estimated trend from the data yields residuals that would be the object of further study. Trend effects can also be removed by employing a differencing procedure (see e.g. Box & Jenkins 1976). For example, first-differencing,

$$y'_t \equiv \nabla y_t = y_t - y_{t-1}, \qquad t = 2, 3, \ldots, N,$$

of a series y_t subject to smooth shifts in the series mean gives a series y'_t which is stationary. Second-differencing,

$$y''_t \equiv \nabla^2 y_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_{t-2}, \qquad t = 3, 4, \ldots, N,$$

produces a stationary series y''_t if y_t is subject to smooth changes in both level and slope. The periodograms of the series of differences and of the innovations ε_t are related through so-called transfer functions. For instance, the relationship between the periodograms of y''_t and ε_t is given by

$$I_\varepsilon(\omega) \approx [4 \sin^2(\omega/2)]^{-2} I_{y''}(\omega)$$

for $\omega \neq 0$ (Priestley 1981).

If the innovations ε_t are Gaussian, the CUSUM or Brillinger tests for trend, applied to the *periodogram ordinates* of the ε_t, i.e. to $I_\varepsilon(k/N)$ $(k = K, K+1, \ldots, M)$, become tests of whether the data are autocorrelated. This follows because the power spectrum values $S_\varepsilon(k/N)$ are constant if and only if the εs are uncorrelated. (The smallest index $k = K$ is to be chosen in such a way that any remaining trend in the residuals is manifested at frequencies below $\omega = 2\pi K/N$, while $2\pi M/N$ is the Nyquist frequency.) This provides us with useful *global* tests for autocorrelation, in contrast to the tests considered in Section 2 which are restricted to consideration of only a small number of lags.

Finally, a plot of the autocorrelations of residuals at lags $1, 2, \ldots, N/4$ (also known as the correlogram) provides some clues to the autocorrelation structure in the data that may be helpful in the model-building process.

ACKNOWLEDGMENTS

The authors are grateful to the American Association of Variable Star Observers (AAVSO) for providing the machine-readable form of their long-period variable star data base. CK is indebted to Professor Walter Zucchini of the Statistics Department at the University of Cape Town for his generosity in terms of time and advice. FL's work was supported by a grant from the Foundation for Research Development.

REFERENCES

AAVSO, 1990, Maxima and Minima of Long Period Variables. Am. Assoc. Variable Star Obs., Cambridge, MA (machine-readable data)
Bartlett M. S., 1937, Proc. R. Soc. London A, 160, 268
Bateson F. M., 1977, N.Z. J. Sci., 20, 73
Box G. E. P., 1979, in Launer R. L., Wilkinson G. N., eds, Robustness in Statistics. Academic Press, New York, p. 202
Box G. E. P., Cox D. R., 1964, J. R. Statist. Soc. B, 26, 211
Box G. E. P., Jenkins G. M., 1976, Time Series Analysis: Forecasting and Control. Holden-Day, Oakland, CA
Brillinger D. R., 1965, Ann. Math. Statist., 36, 1351
Brillinger D. R., 1989, Biometrika, 76, 23
Brown R. L., Durbin J., Evans J. M., 1975, J. R. Statist. Soc. B, 37, 149
Bruce A. G., Martin R. D., 1989, J. R. Statist. Soc. B, 51, 363
Campbell L., 1955, Studies of Long Period Variables. Am. Assoc. Variable Star Obs., Cambridge, MA
Casdagli M., 1991, J. R. Statist. Soc., 54, 303
Chan N. H., Tran L. T., 1992, J. Time Ser. Anal., 13, 19
Chang I., Tiao G. C., Chen C., 1988, Technometrics, 30, 193
Cox D. R., Hinkley D. V., 1974, Theoretical Statistics. Chapman & Hall, London
D'Astous F., Hipel K. W., McLeod A. I., 1979, Water Resources Research, 15, 501
De Jong P., 1991, J. Time Ser. Anal., 12, 143
Detre L., 1969, in Detre L., ed., Non-Periodic Phenomena in Variable Stars. Academic Press, Budapest, p. 3
Duncan D. B., Horn S. D., 1972, J. Am. Statist. Assoc., 67, 815
Durbin J., 1969, Biometrika, 56, 1
Edgington E. S., 1980, Randomization Tests. Marcel Dekker Inc., New York
Engle R. F., 1982, Econometrica, 50, 987
Hart J. D., 1989, Stochastic Processes Appl., 31, 251
Harvey A. C., 1981, The Econometric Analysis of Time Series. Philip Allan, Oxford
Harvey A. C., 1984, J. Forecasting, 3, 245
Harvey A. C., 1989, Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge Univ. Press, Cambridge
Howarth I. D., 1976, Publ. Variable Star Sect. R. Astron. Soc. N.Z., 4, 4
Howarth I. D., 1977, Acta Astron., 27, 5
Howarth I. D., 1978, J. Br. Astron. Assoc., 88, 145
Hsu D. A., 1977, Appl. Statist., 26, 279
Jones R. H., 1980, Technometrics, 22, 389
Judge G. G., Griffiths W. E., Hill R. C., Lee T.-C., 1980, The Theory and Practice of Econometrics. John Wiley & Sons, New York
Kalman R., 1960, Trans. ASME J. Basic Eng., 82, 35
Koen C., 1992, in Warner B., ed., ASP Conf. Ser. 30, Variable Stars and Galaxies. Astron. Soc. Pac., San Francisco
Ledolter J., 1979, Commun. Statist. Theor. Meth., A8, 1227
Ledolter J., 1981, in Findley D. F., ed., Applied Time Series Analysis II. Academic Press, London, p. 449
Li W. K., 1984, J. Time Ser. Anal., 5, 173
Lombard F., Koen C., 1993, MNRAS, 263, 309 (this issue)
Lomnicki Z. A., 1961, Metrika, 4, 37
Maravell A., 1983, J. Business Econ. Statist., 1, 66
Mattei J. A., Mayall M. W., Waagen E. O., 1990, Maxima and Minima of Long-Period Variables 1949–1975. Am. Assoc. Variable Star Obs., Cambridge, Mass
Muirhead C. R., 1986, J. R. Statist. Soc. B, 48, 39

Nisio M., 1960, J. Math. Soc. Japan, 12, 207
Priestley M. B., 1981, Spectral Analysis and Time Series, Vol. 1. Academic Press, London
Priestley M. B., 1988, Non-Linear and Non-Stationary Time Series. Academic Press, London
Rozenbush A. E., 1985, Kinematika Fiz. Nebesn. Tel., 1 (5), 10
Scargle J. D., 1981, ApJS, 45, 1
Sprent P., 1989, Applied Nonparametric Statistical Methods. Chapman & Hall, London
Subba Rao T., Gabr M. M., 1980, J. Time Ser. Anal., 1, 145
Tong H., 1990, Non-Linear Time Series: A Dynamical System Approach. Oxford Univ. Press, Oxford
Tsay R. S., 1988, J. Forecasting, 7, 1
Vio R., Cristiani S., Lessi O., Salvadori L., 1991, ApJ, 380, 351
Visser H., Molenaar J., 1988, Biometrics, 44, 929
Weiss A. A., 1984, J. Time Ser. Anal., 5, 129
West M., Harrison J., 1989, Bayesian Forecasting and Dynamic Models. Springer-Verlag, New York
Wold H., 1938, A Study in the Analysis of Stationary Time Series. Almquist & Wiksell, Uppsala