Mon. Not. R. Astron. Soc. **392,** 190–204 (2009) doi:10.1111/j.1365-2966.2008.14017.x

Statistical tests for changes in the amplitude, frequency or phase of a sinusoidal variation

Chris Koen★

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

Accepted 2008 September 20. Received 2008 September 1; in original form 2007 November 5

ABSTRACT

The problem considered is that of testing for small changes over time in the properties of a sinusoidal signal contaminated by noise. Two test statistics are proposed. The first is a frequency domain statistic based on the spectrum of the data, while the second is a time domain statistic due to Nyblom. Power studies are used to show that the Nyblom statistic is generally the better. The form of the Nyblom statistic appropriate for phase-change tests also performs well in testing for frequency changes.

Key words: methods: statistical – stars: oscillations.

1 INTRODUCTION

Light curves of pulsating stars are commonly modelled as sums of sinusoids. Denoting the jth of N brightness measurements by y_j, and assuming the signal contains K sinusoids,

$$y_j = \mu + \sum_{k=1}^{K} C_k \cos(\omega_k t_j + \phi_k) + e_j \qquad j = 1, 2, \ldots, N, \tag{1}$$

where C_k, ω_k and ϕ_k are the amplitude, (angular) frequency and the phase associated with the kth sinusoid. The jth measurement is obtained at time t_j, and μ is the mean brightness level. It is assumed for the time being that the e_j are uncorrelated: this assumption is not essential, and a red noise example is treated in Section 4. The methodology presented below will work best if the observation times t_j are regularly spaced, i.e. $t_{j+1} - t_j$ is constant, but can be applied to irregularly spaced data with some possible loss of efficiency. The necessary modifications to tailor the test statistics to specific time spacings will be dealt with in a later paper. For convenience only constant time spacing examples will be dealt with below, the exception being the application in Section 7.

A question which arises in variable star research is whether the amplitudes and frequencies/phases of sinusoids in (1) are constant. Two statistical tests which address this question form the subject of this paper.

Nyblom (1989) derived a very general time domain statistic for changes in time series model parameters. The statistic is of the form of a likelihood ratio, which is commonly used for hypothesis testing (e.g. Mood, Graybill & Boes 1974). Properties of an alternative frequency domain statistic (FDS) are also investigated below: it is based on the spectrum of the observations. The FDS is introduced in the next section of the paper.

In Sections 3 and 4, we study the two simplest applications of the Nyblom (1989) statistic to parameter changes in (1), namely single sinusoids ($K = 1$) with respective changes only in the amplitude (Section 3) and only in the phase or frequency (Section 4). The performance of the Nyblom statistic is contrasted with that of the FDS. Throughout Sections 3 and 4, it is assumed that the noise e_j in (1) is uncorrelated, i.e. white. In practice, measurement errors are often correlated, particularly if the observations are obtained as unstandardized high-speed photometry. The generalization of the methodology to red noise measurement errors is presented in Section 5. The extension to multiple pulsation modes ($K > 1$), and multiple non-constant parameters, is dealt with in Section 6.

Simulated data are used to illustrate the theory, while an application to observations of the pulsating sdB star Feige 48 is given in Section 7. Conclusions follow in the last section of the paper.

It is assumed throughout that μ is constant – although testing for changes in the mean light level is easily incorporated into the formulation; this would be astrophysically distinct from changes in pulsation properties, and is therefore not of interest in the present context.

For the benefit of the readers less familiar with statistics, a few words regarding the test procedures are added here. This paper is concerned with hypothesis testing: in particular, the null hypothesis (commonly designated 'H0') considered throughout is that the amplitudes, phases and frequencies in (1) do not change over time. The alternative ('H1') is that one or more of the parameters are variable. In principle, there

★E-mail: ckoen@uwc.ac.za

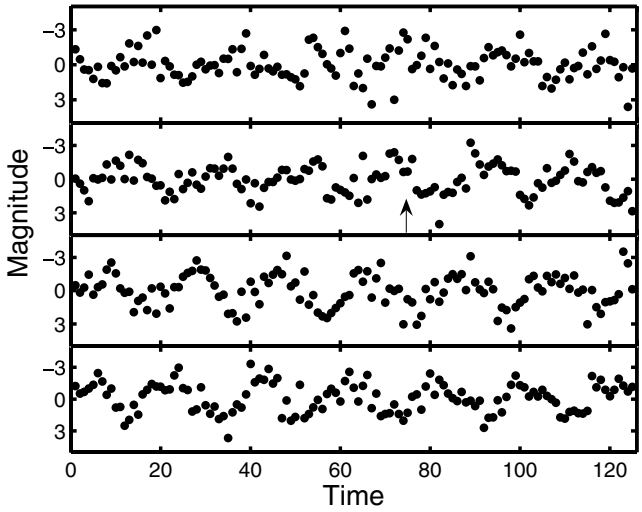

Figure 1. A simulated data set of length $N = 500$. At the time marked by the arrow ($t = 200$), the amplitude jumps from unity to 1.5. The starting phase has been randomly chosen, and the phase at the jump-point is the same as at $t = 0$ (since the period is 20 time units). The noise is white, with variance $\sigma_e^2 = 1$.

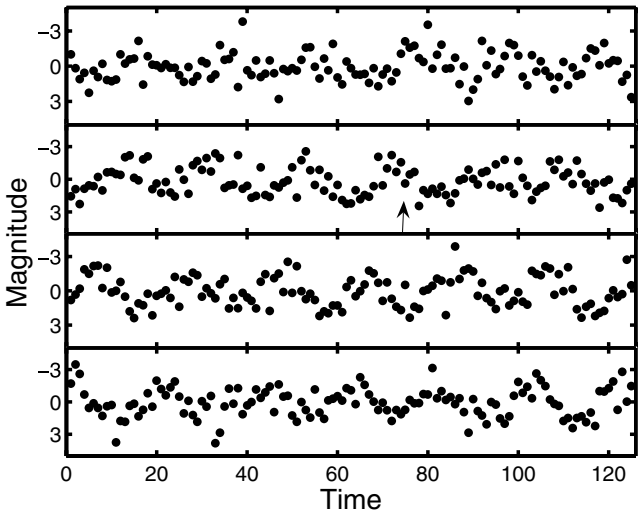

Figure 2. A simulated data set of length $N = 500$. At the time marked by the arrow ($t = 200$), the amplitude jumps from unity to 1.4. The starting phase has been randomly chosen, and the phase at the jump-point is the same as at $t = 0$ (since the period is 20 time units). The noise is white, with variance $\sigma_e^2 = 1$.

are many different statistics which can be used to compare the evidence for each of the two hypotheses, and to select the more plausible. Below, two such test procedures are contrasted.

Hypothesis tests are usually done conservatively: test statistics should lean so far in favour of H1 that there is only a very slim chance (usually 5 or 1 per cent) that H0 is correct. On the other hand, if H1 is true, then it is desirable that H0 be rejected. The proportion of rejections of H0, given that H1 is true, is referred to as the 'power' of the test. Clearly, the power will increase both with sample size (deviations from H0 are easier to detect if there are a lot of data) and with the size of the deviation from H0. Although the dependence of power on sample size is not studied below, its dependence on the size of the deviation is investigated. The results are presented in the form of power curves, i.e. plots of power against the size of the deviation from H0.

It is also worth mentioning that different test statistics may be most powerful depending on the exact form of H1.

2 A GENERAL-PURPOSE FREQUENCY DOMAIN STATISTIC

The fundamental ideas are more transparent in the case of a single sinusoid

$$y_j = \mu + C\cos(\omega t_j + \phi) + e_j \qquad j = 1, 2, \ldots, N; \tag{2}$$

more complicated models will be dealt with later in the paper. Two simulated example data sets are shown in Figs 1 and 2. The basic parameter set, which is also used in other simulations reported below, is a single sinusoid with frequency $\omega = 2\pi \times 0.05$ and a phase ϕ randomly chosen from $[0, 2\pi]$. (Starting phases of sinusoids are random throughout the paper: this is to avoid presenting results which depend on phase.)

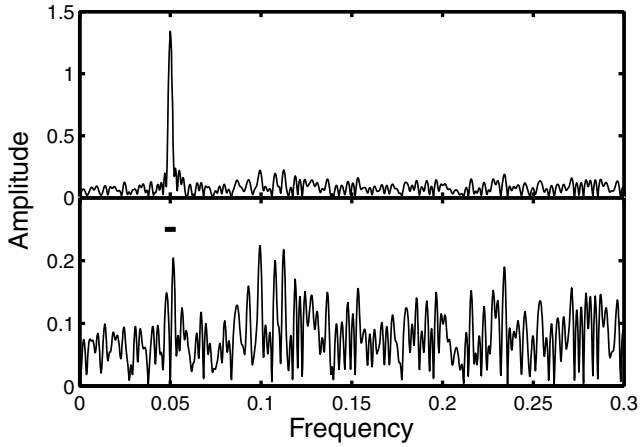

Figure 3. Top panel: the amplitude spectrum of the data in Fig. 1. Bottom panel: the amplitude spectrum of the residuals left after pre-whitening by the best-fitting sinusoid. Note the residual power near the pre-whitened frequency at 0.05. The short horizontal bar centred on $f_0 = 0.05$ shows the minimum frequency interval ($W = 1.5$) for the calculation of the FDS (see the text for details).

The 'observations' are taken regularly, i.e. $t_j = j (j = 1, 2, \ldots, 500)$, and are contaminated by Gaussian measurement error with variance $\sigma_e^2 = 1$. For the data sets of Figs 1 and 2, the amplitude of the sinusoid is initially $C = 1$, and then changes discontinuously to $C = 1 + \Delta C$ at $t = 200$ (indicated by the arrow). The amplitude changes are $\Delta C = 0.5, 0.4$ in Figs 1 and 2, respectively. Despite the fact that the changes are large, they are certainly not blatantly obvious to the human eye – particularly in the case of Fig. 2.

For convenience, the series mean μ is set to zero throughout the simulations. It is none the less explicitly estimated in the evaluation of the various statistics.

Rapid variability in the amplitude, frequency and/or phase of a single sinusoid can be modelled in the frequency domain by adding further sinusoids with closely similar frequencies. This is perhaps most easily seen in the case of amplitude variations, which can be described by 'beating' of sinusoids with slightly different frequencies. A FDS for changes in any of the parameters characterizing a sinusoid can then be constructed by checking for the presence of additional sinusoids with close frequencies.

Formally, such a test can be based on the periodogram

$$I_y(\omega) = \frac{1}{N}\left|\sum_{j=1}^{N}(y_j - \overline{y})\exp(-i\omega t_j)\right|^2 \tag{3}$$

of the time series. Denote by f_0 the frequency of the sinusoid which fits the data best. If the best-fitting sinusoid is pre-whitened from the data, then any substantial change in C, ω and/or ϕ will result in non-negligible residual power at one or more frequencies close to f_0. A possible frequency domain test statistic is then simply the maximum residual power in a narrow frequency range around f_0, standardized by the variance of the residuals.

The point is illustrated by Fig. 3, which shows the periodogram of the simulated data in Fig. 1, and also the periodogram of the residuals after removing a least-squares fitted sinusoid. The single sinusoid was clearly not an adequate model for the cyclical variability in the data, with the result that there is excess power near the frequency 0.05 after its removal. The amount of this excess, as measured by the peak power near $f = 0.05$, serves as the basic for the FDS.

The statistic is, of course, of not much use without an accompanying significance level. The latter can be obtained by bootstrapping. In essence, the method relies on producing many data sets which are statistically similar to the original, as specified by the null hypothesis (i.e. constant C, ω and ϕ, in the present context). The statistic of interest is then calculated for each of these bootstrap data sets. The observed value of the statistic can then be compared to the values generated by the bootstrapping procedure in order to determine its significance. The interested reader is referred to Efron & Tibshirani (1993) for a detailed description of the method, or Koen (2007) for a very brief summary. In the simulations reported in this paper, the number of bootstrap samples is 1000.

The full procedure is as follows.

(i) Starting with the original data set, obtain the frequency $\omega_0 = 2\pi f_0$, amplitude C_0 and phase ϕ_0 of the best-fitting sinusoid, as well as the mean series level μ_0. This is easily done by least squares, facilitated by a preliminary estimate of ω_0 from the periodogram I_y.
(ii) Pre-whiten the data by the sinusoid, i.e. calculate

$$\widehat{e}_j = y_j - \mu_0 - C_0\cos(\omega_0 t_j + \phi_0) \qquad j = 1, 2, \ldots, N.$$

(iii) Calculate the periodogram I_e of the $\widehat{e}_j$ over the interval

$$\left[f_0 - \frac{W}{T}, f_0 + \frac{W}{T}\right] \qquad f_0 = \omega_0/2\pi, \tag{4}$$

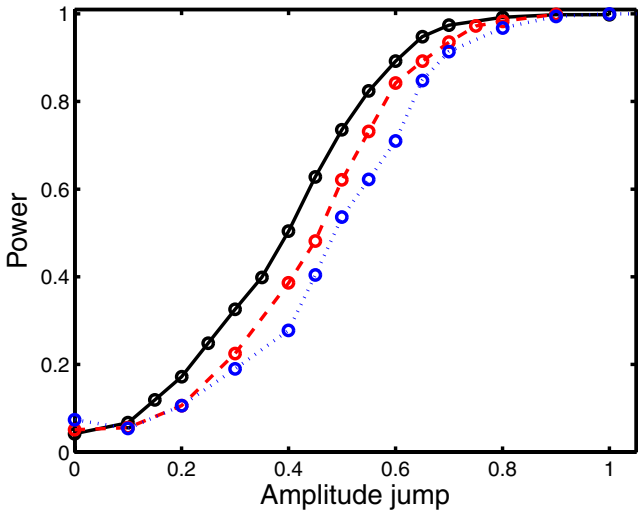

Figure 4. The power of the FDS for the case where the amplitude increases abruptly from unity, by the amount shown on the horizontal axis. The solid, broken and dotted lines, respectively, indicate results obtained with $W = 1.5$, 2.5 and 4 in equation (4). Hypothesis tests were performed at the 5 per cent level. The amplitude change point is at $t = 200$.

where T is the time interval spanned by the data: $1/T$ is the approximate frequency resolution of the periodogram (e.g. Loumos & Deeming 1978). The factor W determines the width of the interval; its specification is discussed below.

(iv) Note the peak value of the periodogram I_e, and standardize it by dividing by the variance

$$s^2 = \frac{1}{N-4}\sum_j \widehat{e}_j^2. \tag{5}$$

Denote the statistic by $\mathcal{F}_*$.

(v) Generate a bootstrap sample by first drawing a random sample $\{\epsilon_1^{(1)}, \epsilon_2^{(1)}, \ldots, \epsilon_N^{(1)}\}$ from the collection of $\widehat{e}_j$ calculated in Step (ii); some of the $\widehat{e}_j$ will occur more than once, while some values will be absent.

(vi) Construction of the bootstrap sample is completed by adding the $\epsilon_j^{(1)}$ to the estimated signal from Step (i), i.e.

$$y_j^{(1)} = \mu_0 + C_0\cos(\omega_0 t_j + \phi_0) + \epsilon_j^{(1)} \qquad j = 1, 2, \ldots, N.$$

(vii) Now repeat Steps (i)–(iv), treating the $y_j^{(1)}$ exactly as the observed data y_j. This gives a first bootstrap statistic $\mathcal{F}^{(1)}$.

(viii) Repeat Steps (v)–(vii) a large number M (typically 1000 or more) times, obtaining the collection of bootstrap statistics $\mathcal{F}^{(1)}, \mathcal{F}^{(2)}, \ldots, \mathcal{F}^{(M)}$.

(ix) The bootstrap significance level of $\mathcal{F}_*$ is found by determining its percentile level with respect to the set of bootstrap statistics.

Simulation experiments were conducted in order to gain some idea as to the efficacy of the FDS. In the first experiment, the power of the statistic against a single jump in the amplitude at $t = 200$ was determined. This was done by simulating 500 data sets, and counting the number of rejections of the null hypothesis H0: (C is constant). In order to eliminate any influence of the phase, ϕ is randomly chosen from $[0, 2\pi]$ for each of the 500 simulations. The test was performed at the 5 per cent level. The results are summarized by power curves, i.e. fractions of rejections of H0, plotted as a function of the size of the amplitude jump.

The width of the interval around f_0 to be scanned for residual power still needs to be specified. Fig. 4 compares power curves for three different widths, $W = 1.5$, 2.5, 4 in (4). Better power is obtained with $W = 1.5$. Further reducing W increases the risk that any relevant spectral peak is outside the frequency interval (4). In the remainder of the paper $W = 1.5$ is therefore used. In the bottom panel of Fig. 3, the interval $(f_0 - 1.5/T, f_0 + 1.5/T)$ is marked by the short horizontal bar centred on $f_0 = 0.05$ (similarly in Fig. 7, for the data plotted in Fig. 2).

If there is information available about the likely form of any parameter changes (or the coherence time-scale of the parameters), an appropriate W could be calculated. In practice, this will probably rarely be the case. It therefore seems prudent to choose W to be small: any parameter changes will inflate the power very close to f_0, and small W will limit the dilution by random features. Of course, if the excess power is spread over a wide frequency interval, then there will be a loss of efficiency if W is chosen too small.

Inspection of the $W = 1.5$ power curve in Fig. 4 shows that an amplitude jump needs to be quite substantial in order to be sure of its detection: for example, there is only a 50 per cent chance that H0 would be rejected at the 5 per cent level if the amplitude increases by 40 per cent.

The FDS will be used as a benchmark for the performance of the Nyblom (1989) statistics studied below.

3 NYBLOM STATISTIC FOR ONE SINUSOID WITH A POSSIBLE AMPLITUDE CHANGE

The simplest test is for changes in the amplitude C in (2). Both ω and ϕ in (2) are initially taken to be known constants: the consequences of having to estimate these two quantities are addressed below. Equation (2) is then a simple linear regression

$$y_j = \mu + z_j C + e_j, \tag{6}$$

where

$$z_j = \cos(\omega t_j + \phi) \tag{7}$$

are known constants.

The form of the Nyblom (1989) statistic appropriate to test for changes in the parameter C in (6) is

$$L(C) = S^{-1} \sum_{j=1}^{N} D_j^2 / \widehat{\sigma}_e^2, \tag{8}$$

where

$$S = \frac{1}{N} \sum_j z_j^2 = \frac{1}{N} \sum_j \cos^2(\omega t_j + \phi)$$

and the D_j are the partial sums

$$D_j = \sum_{i=j}^{N} \widehat{e}_i z_i = \sum_{i=j}^{N} (y_i - \overline{y} - z_i \widehat{C}) z_i \qquad j = 1, 2, \ldots, N\ . \tag{9}$$

In (9), $\overline{y}$, the mean of the observations, estimates the mean light level μ, and $\widehat{C}$ is the least-squares estimate of the amplitude. The residual variance is

$$\widehat{\sigma}_e^2 = \frac{1}{N-2} \sum_j \widehat{e}_j^2 = \frac{1}{N-2} \sum_j (y_j - \overline{y} - z_j \widehat{C})^2. \tag{10}$$

For the purpose of comparison with results of the next section, it is useful to combine some of the formulae above and to write

$$L(C) = \frac{N}{\widehat{\sigma}_e^2} \sum_{j=1}^{N} \left[\sum_{i=j}^{N} \widehat{e}_i \cos(\omega t_i + \phi) \right]^2 \Bigg/ \sum_{j=1}^{N} \cos^2(\omega t_j + \phi). \tag{11}$$

In practice, it is convenient to scale the statistic by N^{-2}, and to work with its logarithm:

$$L' = \log(L) - 2\log(N). \tag{12}$$

Nyblom (1989) provides asymptotic (i.e. large sample) percentage points for L/N^2; these have been converted to L' in Table 1. Only the first line of the table is relevant to this section and the next, as only changes in a single parameter are considered.

The conservative approach of determining the significance levels by bootstrapping is followed here.

The simulations described in Section 2, with a single amplitude jump, were repeated in order to study the power of $L(C)$. For each simulated data set, a preliminary estimate of the frequency is obtained by locating the periodogram maximum; the frequency is then refined by a non-linear least-squares method which simultaneously estimates the phase and amplitude. The null hypothesis of a constant amplitude was tested at both the 5 and 1 per cent levels, and the number of rejections noted.

A final remark before presenting the results concerns the fact that the z_j in (7) are, in practical applications, *not* known constants, since both ω and ϕ would need to be estimated from the data. Since significance levels are determined by bootstrapping, this additional uncertainty is automatically accommodated. However, the variance estimator (10) should be replaced by

$$\widehat{\sigma}_e^2 = \frac{1}{N-4} \sum_j \widehat{e}_j^2\ . \tag{13}$$

Power curves for testing at either the 5 or 1 per cent levels are shown in Fig. 5. For amplitude jumps of 0.7 units, there is virtual certainty of detecting the change, regardless of the level of testing. For smaller changes, the likelihood of rejection of H0 is, of course, greater when

Table 1. Percentage points of the transformed Nyblom statistic L', for large samples. Values have been calculated from Nyblom (1989, table 2). The first column is the number of parameters which may be subject to change.

Parameters	Significance levels			
	10 %	5 %	1 %	0.1 %
1	− 1.0584	− 0.7744	− 0.2971	0.1553
2	− 0.4992	− 0.2904	0.1020	0.4318
3	− 0.1732	0	0.3067	0.6195
4	0.0611	0.2127	0.4843	0.7650

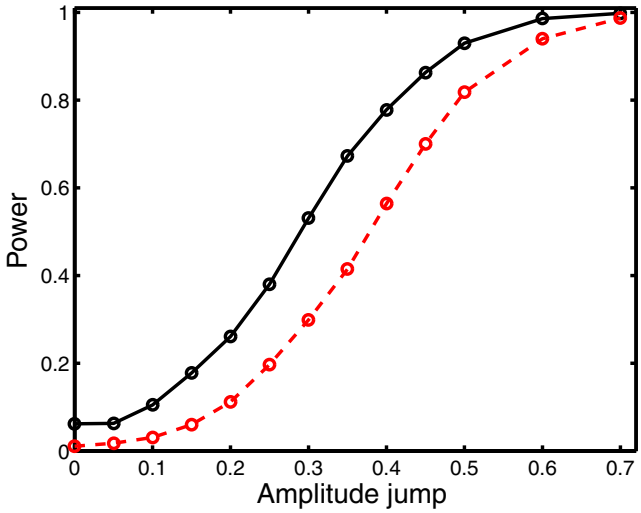

Figure 5. The power of the Nyblom statistic for the case where the amplitude increases abruptly from unity, by the amount shown on the horizontal axis. The solid line is the fraction of rejections, at the 5 per cent level, of the constant-amplitude hypothesis. The broken line is the equivalent result for testing at the 1 per cent level. The series length is $N = 500$, and the abrupt amplitude change is at $t = 200$.

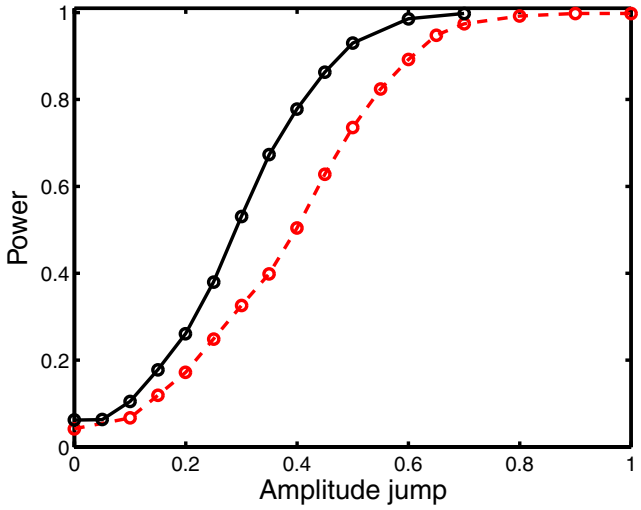

Figure 6. A comparison between the powers of the Nyblom (solid line) and the frequency domain (broken line) statistics for the case where the amplitude increases abruptly from unity, by the amount shown on the horizontal axis. Hypothesis tests were performed at the 5 per cent level. The amplitude change point is at $t = 200$.

using the less demanding 5 per cent test level: for example, there is an 80 per cent chance of rejecting H0 if the amplitude jumps by 0.4 units, which drops to 60 per cent if the test is performed at the 1 per cent level.

Two other interesting results are mentioned, without going into detail. First, the percentage points of L' obtained from bootstrapping agree quite well with those given in Table 1. Secondly, simulations with the jump-point set to $t = 300$ instead of $t = 200$ gave very similar power results to those shown in Fig. 5.

The powers of the Nyblom statistic $L(C)$ and the FDS are compared in Fig. 6. The Nyblom statistic is clearly substantially better than the FDS, particularly for amplitude changes in the range 0.3 to 0.5.

It is interesting to compare significance levels of the FDS and Nyblom statistic for the data sets in Figs 1 and 2. For the Fig. 1 data, these are 0.01 (FDS) and 0.003 (Nyblom), both highly significant. For the Fig. 2 data, on the other hand, the levels are 0.23 and 0.017, respectively, i.e. the FDS is not significant. A glance at the bottom panel of Fig. 7 shows the reason: there is little or no excess power near f_0 in the residual spectrum of the Fig. 2 data.

Fig. 8 compares powers of the two statistics for the case of a sustained linear increase of the amplitude over time. The power difference is perhaps slightly larger than in the case of the abrupt increase (again in favour of the Nyblom statistic).

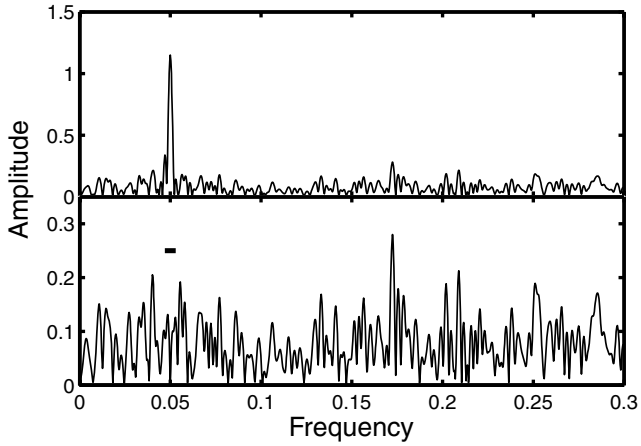

Figure 7. Top panel: the amplitude spectrum of the data in Fig. 2. Bottom panel: the amplitude of the residuals left after pre-whitening by the best-fitting sinusoid. The short horizontal bar centred on $f_0 = 0.05$ shows the minimum frequency interval ($W = 1.5$) for the calculation of the FDS (see the text for details).

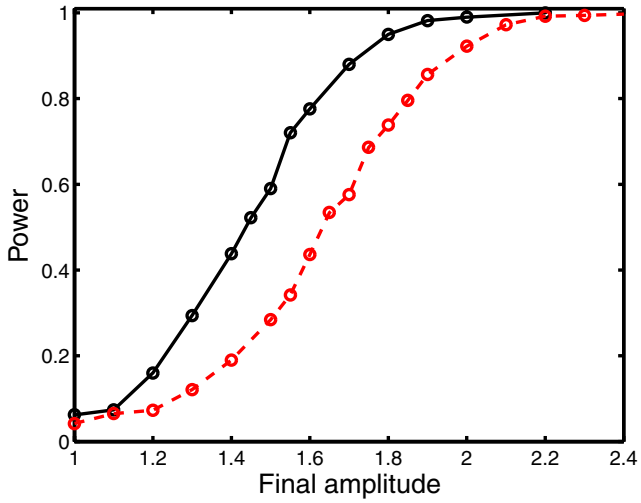

Figure 8. A comparison between the powers of the Nyblom (solid line) and the frequency domain (broken line) statistics for the case where the amplitude increases linearly from unity to the value shown on the horizontal axis. Hypothesis tests were performed at the 5 per cent level.

4 ONE SINUSOID WITH A POSSIBLE PHASE OR FREQUENCY CHANGE

Inspection of (1) shows that, in principle, only the arguments

$$\omega_k t_j + \phi_k$$

of the sinusoids are measurable, and not both the frequency and the phase. The implication is that possible changes in frequency and phase would be confounded. None the less, in practice, there may be instances in which frequency (phase) changes provide a more natural description than phase (frequency) changes. To give an example, modelling a single abrupt change in frequency by a continuously changing phase seems contrived. It is therefore worthwhile giving consideration to statistics for both phase and frequency changes.

Although the model equation, namely (2), is the same as that analysed in the previous section, it cannot be used in the convenient linear form (6) since the quantities of interest – the phase ϕ or the frequency ω – are non-linear regression parameters. It is necessary to resort to a more fundamental expression for the Nyblom statistic: a more general form (Nyblom 1989) is

$$\begin{aligned} L &= \sum_{j=1}^{N} D_j^2 / S \\ D_j &= \sum_{i=j}^{N} d_i \\ S &= \frac{1}{N} \sum_j \mathrm{E} d_j^2, \end{aligned} \tag{14}$$

where E is the expectation operator, and

$$d_i = \frac{\partial \log f(y_i)}{\partial \phi} \quad \text{or} \quad = \frac{\partial \log f(y_i)}{\partial \omega} \ , \tag{15}$$

$f(y_i)$ being the likelihood of the ith observation. (More correctly: the likelihood of the ith observation, conditional on all preceding observations.)

For zero-mean Gaussian measurement errors, the necessary likelihood is

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp -\frac{1}{2}\left[\frac{y_i - \mu - C\cos(\omega t_i + \phi)}{\sigma_e}\right]^2 . \tag{16}$$

Concentrating first on testing for phase changes,

$$d_i = \frac{-[y_i - \mu - C\cos(\omega t_i + \phi)]}{\sigma_e^2} C\sin(\omega t_i + \phi) \equiv -\frac{e_i}{\sigma_e^2} C\sin(\omega t_i + \phi).$$

Furthermore, since $\mathrm{E}e_i^2 = \sigma_e^2$,

$$S = \frac{C^2}{N\sigma_e^2} \sum_j \sin^2(\omega t_j + \phi).$$

Substitution into the first equation in (14) leads to

$$L(\phi) = \frac{N}{\sigma_e^2} \sum_{j=1}^{N} \left[\sum_{i=j}^{N} e_i \sin(\omega t_i + \phi)\right]^2 \Bigg/ \sum_{j=1}^{N} \sin^2(\omega t_j + \phi). \tag{17}$$

Replacing the e_i by $\widehat{e}_i = y_i - \widehat{C}\cos(\widehat{\omega} t_i + \widehat{\phi})$ and σ_e^2 by $\widehat{\sigma}_e^2$ (equation 13) gives a calculation formula for L, and hence L'.

It is quite instructive to compare (17) with (11): the two forms are remarkably similar, the only difference being the presence of sines in (17), and cosines in (11).

It is easy to derive the frequency change statistic from the formulae above, but it is not usable, as some of the technical assumptions made in the derivation of (14) are not satisfied. It is, in fact, easy to see that the statistic

$$L(\omega) = \frac{N}{\sigma_e^2} \sum_{j=1}^{N} \left[\sum_{i=j}^{N} e_i t_i \sin(\omega t_i + \phi)\right]^2 \Bigg/ \sum_{j=1}^{N} t_j^2 \sin^2(\omega t_j + \phi)$$

is wanting: although scale invariant, $L(\omega)$ is *not* translation invariant. Furthermore, the value of the statistic will obviously be dominated by terms containing large values of t_i. A single, very large, outlying value t_N would, in fact, render this statistic useless.

Power studies of $L(\phi)$ led to some surprising results.

(i) The phase change statistic (17) is quite efficient in detecting abrupt *frequency* changes, at least for the configurations investigated.
(ii) For the simulated data sets studied, the percentage points in Table 1 are such poor approximations that they cannot be used. This means that for $N = 500$, the large-sample approximation works for amplitude change tests, but not for phase/frequency change tests. We speculate that this is due to the fact that C is a linear parameter in (2), whereas ω and ϕ enter the equation non-linearly.
(iii) The power of (17) is a sensitive function of the position of the change point, in the case of an abrupt frequency change. For example, changes respectively near the beginning or end of the series are not detected with equal efficiencies.

Point (iii) is illustrated in Fig. 9: frequency jumps at $t = 300$ are detected much more readily than those at $t = 200$. This raises the question as to whether the same effect is seen in the case of amplitude or phase changes. Power studies similar to those summarized in Fig. 3, but with the amplitude change point at $t = 300$, showed only minor differences with the results in Fig. 3. Similarly, powers against phase changes at $t = 200$ and 300 are, within the errors, the same (see Fig. 10 for the case $t = 200$).

Why is power so strongly dependent on the location of the change point in the case of frequency changes? The reason lies in the fact that the $L(\phi)$ (and FDS – see below) detect not only the change in frequency, but also the jump, induced by the change in ω, in the argument $(\omega t + \phi)$ of the sinusoid in (2). This jump increases with the size of t, and is therefore more easily detected at $t = 300$ than at $t = 200$.

This is, of course, not satisfactory, as the power curves do not exactly reflect the ability of the statistic to detect pure frequency changes. Although the discontinuity in the argument of the sinusoid could be avoided by a suitable adjustment of the phase, this does not solve the problem – there would then be changes in both frequency and phase. These results recall the point made at the beginning of this section, namely that the argument $(\omega t + \phi)$, rather than the phase or frequency, is the primary quantity.

Exactly the same form of the FDS as described in Section 2 could be used to test for frequency or phase changes, because the statistic simply tests for excess residual power near the null hypothesis frequency f_0. Since describing a frequency or phase change would, in the frequency domain, require more than one sinusoid with frequencies close to f_0, the same principle as outlined for the amplitude change test applies. FDS power curves analogous to those in Fig. 9 are displayed in Fig. 11. It can be seen that the FDS statistic is not as powerful as $L(\phi)$ in detecting frequency changes, and that its power also depends on the position of the change point.

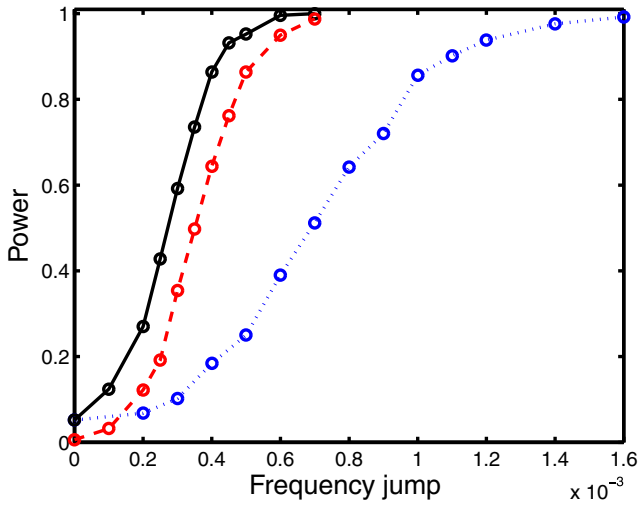

Figure 9. The power of the Nyblom phase-change statistic (17) for the detection of *frequency* changes. Solid and broken lines show the results of testing at the 5 and 1 per cent levels, respectively, when the frequency increases abruptly at $t = 300$, from the fiducial value 0.05. The dotted line shows the power when the change point is at $t = 200$ (5 per cent level results only).

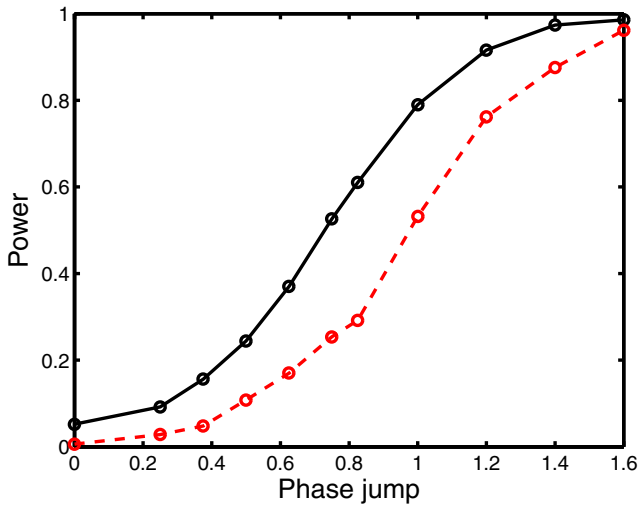

Figure 10. The power of the Nyblom phase-change statistic (17) for the detection of *phase* changes. Solid and broken lines show the results of testing at the 5 and 1 per cent levels, respectively, when the phase increases abruptly at $t = 200$. Results for the case where the abrupt change is at $t = 300$ are closely similar.

5 CORRELATED MEASUREMENT ERRORS

In order to keep the treatment in this section simple, the observations are explicitly taken to be regularly spaced, i.e. $t_j = j$. This is tantamount to treating time as an integer-valued, rather than continuous, variable. It is possible to do away with this assumption, but estimation of the parameters of the noise series is considerably more involved in the continuous-time case (e.g. Koen 2005).

For regularly spaced time series, the first-order autoregressive form

$$e_j = \alpha e_{j-1} + v_j \qquad 0 < \alpha < 1, \tag{18}$$

with v_j white noise, is an uncomplicated, but realistic, model for red noise. A simple (albeit slightly biased) estimator for α, given a set of observations e_j, is

$$\widehat{\alpha} = \sum_{j=1}^{N-1} e_j e_{j+1} \Bigg/ \sqrt{\sum_{j=1}^{N-1} e_j^2 \sum_{j=2}^{N} e_j^2}\,, \tag{19}$$

i.e. the correlation coefficient between e_j and e_{j+1}. A more common form is

$$\widehat{\alpha} = \sum_{j=1}^{N-1} e_j e_{j+1} \Bigg/ \sum_{j=1}^{N-1} e_j^2$$

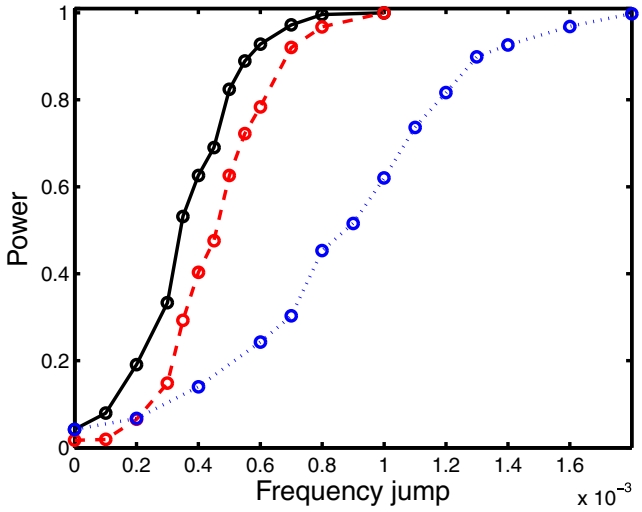

Figure 11. As for Fig. 9, but for the FDS.

(e.g. Chatfield 2003), but (19) is a little more convenient to calculate with computer packages which have built-in functions for correlation coefficients.

The frequency spectrum associated with (18) is

$$f_e(\omega) = \frac{\sigma_v^2}{1 + \alpha^2 - 2\alpha\cos\omega} = \frac{(1-\alpha^2)\sigma_e^2}{1 + \alpha^2 - 2\alpha\cos\omega} \tag{20}$$

(e.g. Chatfield 2003).

Standardization of the FDS discussed in Section 2 involved division by the variance $s^2 = \widehat{\sigma}_e^2$ [see Step (iv) of the bootstrap procedure]. The generalization to correlated noise involves division by the estimated spectral density, which is obtained by replacing α and σ_e^2 in (20) by their estimators $\widehat{\alpha}$ and s^2. For sufficiently small intervals around the frequency of interest ω_0, it may be assumed that f_e is approximately constant at $f_e(\omega_0)$.

In the case of the Nyblom statistic $L(C)$, (7) is supplemented by (18), and α is a nuisance parameter along with ω and ϕ. Once it has been estimated by, for example, (19), the white noise elements follow from

$$\widehat{v}_j = \widehat{e}_j - \widehat{\alpha}\widehat{e}_{j-1} \qquad j = 2, 3, \ldots, N. \tag{21}$$

Equations (8), (9) and (13) are modified to read

$$\begin{aligned} L(C) &= S^{-1}\sum_{j=2}^{N} D_j^2/\widehat{\sigma}_v^2 \\ D_j &= \sum_{i=j}^{N} \widehat{v}_i z_i \\ \widehat{\sigma}_v^2 &= \frac{1}{N-6}\sum_{j=2}^{N}\widehat{v}_j^2. \end{aligned} \tag{22}$$

Equation (11) is changed to

$$L(C) = \frac{N}{\widehat{\sigma}_v^2}\sum_{j=1}^{N}\left[\sum_{i=j}^{N}\widehat{v}_i\cos(\omega t_i + \phi)\right] \Bigg/ \sum_{j=1}^{N}\cos^2(\omega t_j + \phi)\ , \tag{23}$$

and there are similar changes to the expression (17) for $L(\phi)$.

A comment on the denominator of (22) is perhaps not amiss. First, there is a loss of one degree of freedom associated with the estimation of each of the five parameters μ, C, ω, ϕ and α. Furthermore, the effective length of the series is reduced from N to $N-1$, since v_1 cannot be estimated (see equation 21). This means that the degrees of freedom left for the estimation of the residual variance σ_v^2 is $N-6$.

As may have been anticipated, red noise undermines the power of the statistics: compare Fig. 12 with Fig. 5 (Nyblom statistic), and Fig. 13 with Fig. 4 (FDS).

6 MORE COMPLICATED MODELS

There are two different generalizations of the material presented in Sections 3 and 4. The first is the possibility of a number of *nuisance parameters*. By this is simply meant 'passive' parameters similar to μ in equation (1), which may need to be estimated, but are constant over

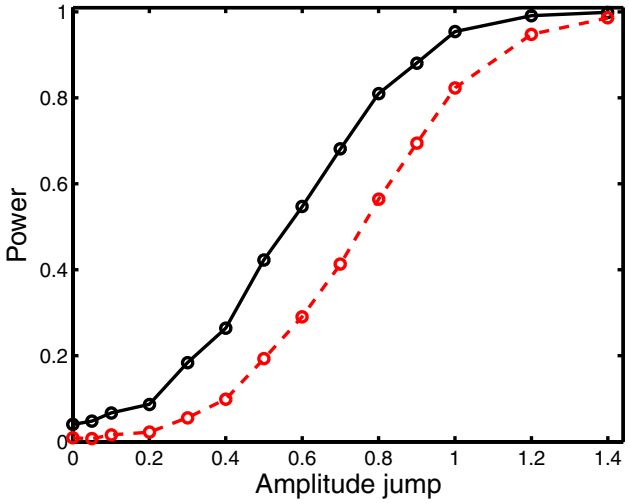

Figure 12. As for Fig. 5, but for the case of correlated (i.e. red) noise with $\alpha = 0.5$. Note that for a given amplitude jump, the power is considerably lower than in the white noise case.

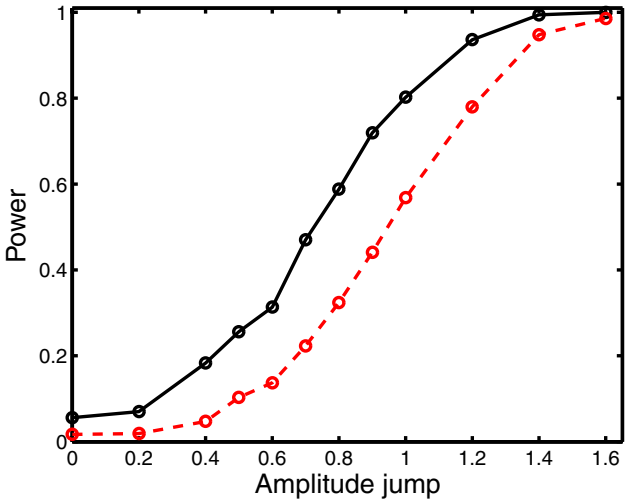

Figure 13. As for the $W = 1.5$ curve in Fig. 4, but for the case of correlated (i.e. red) noise with $\alpha = 0.5$. The solid line shows the power when testing at the 5 per cent level, while the broken line gives the 1 per cent level results. Note that for a given amplitude jump, the power is considerably lower than in the white noise case.

the span of the observations. Without loss of generality, let the parameter which may be subject to change be associated with the first term in the sum (1). Formally,

$$y_j = F(\boldsymbol{\Theta}) + C_1 \cos(\omega_1 t_j + \phi_1) + e_j$$

$$F(\boldsymbol{\Theta}) = \mu + \sum_{k=2}^{K} C_k \cos(\omega_k t_j + \phi_k)$$

$$\boldsymbol{\Theta} = \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \vdots \\ \Theta_K \\ \Theta_{K+1} \\ \vdots \\ \Theta_{2K-1} \\ \Theta_{2K} \\ \vdots \\ \Theta_{3K-2} \end{bmatrix} = \begin{bmatrix} \mu \\ \omega_2 \\ \vdots \\ \omega_K \\ C_2 \\ \vdots \\ C_K \\ \phi_2 \\ \vdots \\ \phi_K \end{bmatrix}. \tag{24}$$

The unknown parameters Θ_i are estimated under the null hypothesis, i.e. as if there are no changes in C_1 or ϕ_1: for small changes in C_1 or ϕ_1 this seems reasonable, and it is vindicated by simulation results. The only required modification to the treatment in Sections 3 is then to use

$$\widehat{e}_j = y_j - F(\widehat{\boldsymbol{\Theta}}) - \widehat{C}_1 \cos(\widehat{\omega}_1 t_j + \widehat{\phi}_1) \tag{25}$$

in (9)–(10), and also to change the denominator in (13) to $N - (3K + 1)$. In Section 4,

$$\begin{aligned} \widehat{e}_j &= y_j - F(\widehat{\boldsymbol{\Theta}}) - \widehat{C}_1 \cos(\widehat{\omega}_1 t_j + \widehat{\phi}_1) \\ \widehat{\sigma}_e^2 &= \frac{1}{N - 3K - 1} \sum_j \widehat{e}_j^2. \end{aligned} \tag{26}$$

The second generalization is to consider the possibility of a change in more than one parameter. Let the number of parameters subject to possible change be m amplitudes and n phases, say $C_{a1}, C_{a2}, \ldots, C_{am}$ and $\phi_{p1}, \phi_{p2}, \ldots, \phi_{pn}$. Denote these collectively by $\theta_1, \theta_2, \ldots, \theta_{m+n}$. In the completely general case, the Nyblom statistic L' is then calculated in the sequence of steps

$$\begin{aligned} \boldsymbol{d}_k &= \partial \log f_k(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \\ \boldsymbol{D}_j &= \sum_{k=j}^{N} \boldsymbol{d}_k \\ \boldsymbol{S} &= \frac{1}{N} \sum_k \mathrm{E}[\boldsymbol{d}_k \boldsymbol{d}_k'] \\ L &= \mathrm{trace} \left[\boldsymbol{S}^{-1} \sum_{j=1}^{N} \boldsymbol{D}_j \boldsymbol{D}_j' \right] \\ L' &= \log(L) - 2\log(N). \end{aligned} \tag{27}$$

The first line is obviously a generalization of (15) to the case of $m+n > 1$ parameters of interest: $\boldsymbol{d}_k$ is an $m + n$-vector of derivatives of the kth likelihood term. Generalizing (16), the likelihood factors are

$$f_k \equiv f(y_k) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp - \frac{1}{2\sigma_e^2} \left[y_k - \mu - \sum_{j=1}^{K} C_j \cos(\omega_j t_k + \phi_j) \right]^2$$

and the components of the vector $\boldsymbol{d}_k$ are

$$d_{k,i} = \frac{\partial \log f(y_k)}{\partial \theta_i},$$

which reduce to

$$\begin{aligned} d_{k,i} &= \frac{1}{\sigma_e^2} \left[y_k - \mu - \sum_{j=1}^{K} C_j \cos(\omega_j t_k + \phi_j) \right] \cos(\omega_{ai} t_k + \phi_{ai}) \\ &= \frac{e_k}{\sigma_e^2} \cos(\omega_{ai} t_k + \phi_{ai}) \quad i = 1, 2, \ldots, m \\ d_{k,i} &= -\frac{1}{\sigma_e^2} \left[y_k - \mu - \sum_{j=1}^{K} C_j \cos(\omega_j t_k + \phi_j) \right] C_{pi} \sin(\omega_{pi} t_k + \phi_{pi}) \\ &= -\frac{e_k}{\sigma_e^2} C_{pi} \sin(\omega_{pi} t_k + \phi_{pi}) \quad i = m+1, m+2, \ldots, m+n. \end{aligned}$$

In vector form

$$\boldsymbol{d}_k = \frac{e_k}{\sigma_e^2} \begin{bmatrix} \cos(\omega_{a1} t_k + \phi_{a1}) \\ \cos(\omega_{a2} t_k + \phi_{a2}) \\ \vdots \\ \cos(\omega_{am} t_k + \phi_{am}) \\ -C_{p1} \sin(\omega_{p1} t_k + \phi_{p1}) \\ \vdots \\ -C_{pn} \sin(\omega_{pn} t_k + \phi_{pn}) \end{bmatrix}. \tag{28}$$

It follows that

$$\boldsymbol{S} = \frac{1}{N\sigma_e^2} \begin{bmatrix} \boldsymbol{P} & \boldsymbol{R} \\ \boldsymbol{R}' & \boldsymbol{Q} \end{bmatrix} \equiv \frac{1}{N\sigma_e^2} \boldsymbol{S}_*, \tag{29}$$

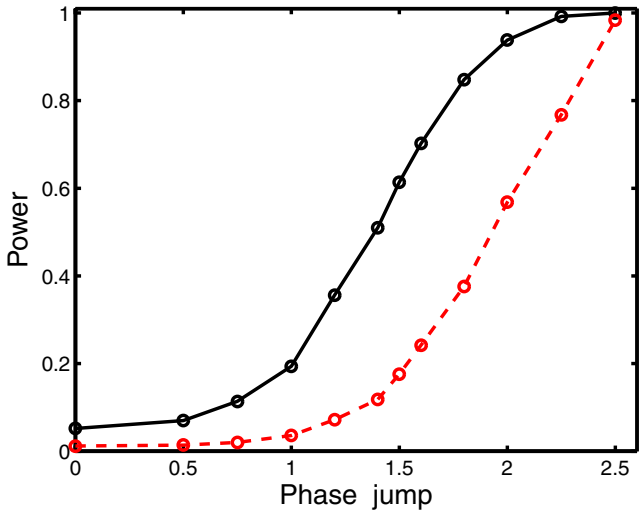

Figure 14. The power of the Nyblom statistic $L(C, \phi)$ (see equation 31) for the detection of *phase* changes. Solid and broken lines show the results of testing at the 5 and 1 per cent levels, respectively, when the phase increases abruptly at $t = 200$.

where the components of the submatrices are

$$\begin{aligned} P_{ij} &= \sum_{k=1}^{N} \cos(\omega_{ai} t_k + \phi_{ai}) \cos(\omega_{aj} t_k + \phi_{aj}) \qquad i, j = 1, 2, \ldots, m \\ Q_{ij} &= C_{pi} C_{pj} \sum_{k=1}^{N} \sin(\omega_{pi} t_k + \phi_{pi}) \sin(\omega_{pj} t_k + \phi_{pj}) \qquad i, j = 1, 2, \ldots, n \\ R_{ij} &= -C_{pj} \sum_{k=1}^{N} \cos(\omega_{ai} t_k + \phi_{ai}) \sin(\omega_{pj} t_k + \phi_{pj}) \qquad i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, n. \end{aligned} \tag{30}$$

As an example, if $m = n = 1$,

$$L(C, \phi) = \frac{N}{|\boldsymbol{S}_*| \sigma_e^2} \text{trace} \left\{ \begin{bmatrix} Q_{11} & -R_{11} \\ -R_{11} & P_{11} \end{bmatrix} \sum_{j=1}^{N} \sum_{r,s=j}^{N} e_r e_s \begin{bmatrix} \cos(\psi_{r,a1}) \cos(\psi_{s,a1}) & -C_{p1} \cos(\psi_{r,a1}) \sin(\psi_{s,p1}) \\ -C_{p1} \sin(\psi_{r,p1}) \cos(\psi_{s,a1}) & C_{p1}^2 \sin(\psi_{r,p1}) \sin(\psi_{s,p1}) \end{bmatrix} \right\} \tag{31}$$

where $\psi_{j,i} = \omega_i t_j + \phi_i$, and the determinant $|\boldsymbol{S}_*| = P_{11} Q_{11} - R_{11}^2$.

Surprisingly, the efficiency of $L(C, \phi)$ in the detection of pure amplitude changes is very similar to that of $L(C)$. As an illustration, significance levels of $L(C, \phi)$ for the data sets of Figs 1 and 2 are 0.002 and 0.019, which may be compared with the levels of 0.003 and 0.017 attained by $L(C)$. However, there is a loss of power against pure phase or frequency changes, as compared to the performance of $L(\phi)$ (see Figs 14 and 15, which should be compared with Figs 10 and 9, respectively).

7 AN APPLICATION

Pulsations in the subdwarf B star Feige 48 were discovered by Koen et al. (1998). An extensive study of the pulsation frequencies was later published by Reed et al. (2004): those authors conclude that there are five resolved periodicities. Here, the longest run (6.7 h) from Koen et al. (1998) is studied (see their fig. 1 for a light curve). The observation method was high-speed (10 s integrations) photo-electric photometry. The total number of usable measurements was $N = 2103$. An amplitude spectrum, plotted over the frequency range of prime interest, is plotted in the top panel of Fig. 16: the data have been de-trended by removing a local linear fit, hence filtering out very low frequency variations.

Attention is focussed on the well-resolved frequency near 2.64 mHz (Reed et al. 2004), hence the complex of frequencies near 2.9 mHz is first removed, to facilitate the analysis. The amplitude spectrum of the residuals is in the bottom panel of Fig. 16: it is dominated by the 2.0 mmag peak at a frequency 2.637 mHz. The constancy of the amplitude, phase and/or frequency of this periodicity is now addressed.

The p-values of the change statistics discussed above can be seen in the first line of Table 2. For the FDS, W in (4) was set at 1.5. Significance levels followed from 5000 bootstrap replications of the data. Both the Nyblom statistic $L(C, \phi)$ and the FDS are very highly significant; the specific tests indicate an amplitude change.

Since the time series is fairly long, it could be partitioned in order to get more detailed information. The last two lines of the table show that the parameter changes are concentrated over the first part of the data, and that there were both amplitude and phase/frequency changes.

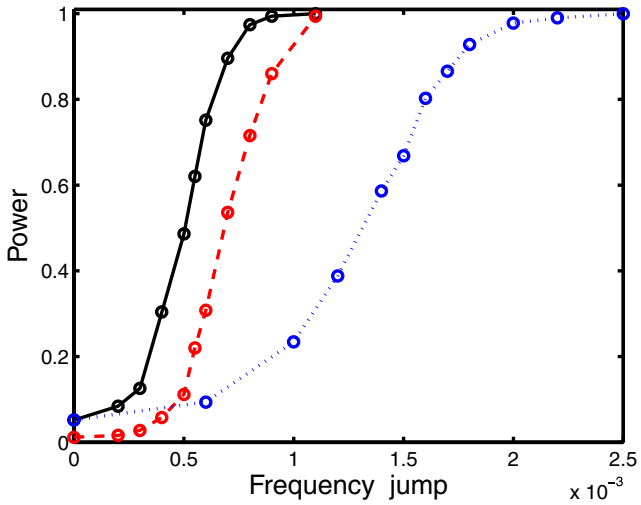

Figure 15. The power of the Nyblom statistic $L(C, \phi)$ (see equation 31) for the detection of *frequency* changes. Solid and broken lines show the results of testing at the 5 and 1 per cent levels, respectively, when the frequency increases abruptly at $t = 300$ from the fiducial value 0.05. The dotted line shows the power when the change point is at $t = 200$ (5 per cent level results only).

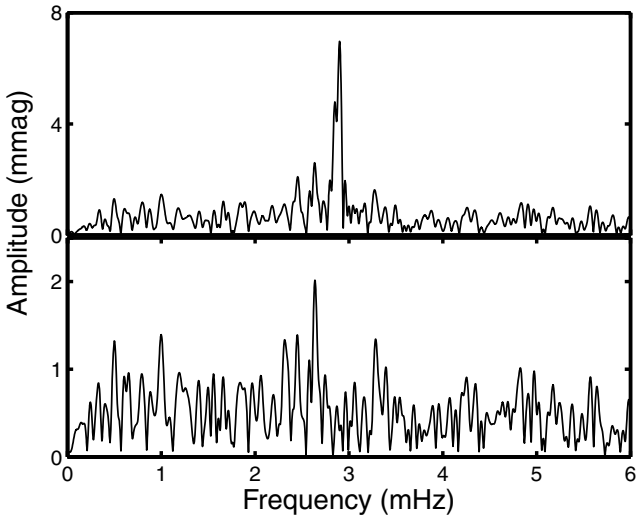

Figure 16. Top panel: the amplitude spectrum of photometric observations of the EC 14026 star Feige 48 (see text for details). Bottom panel: the amplitude of the residuals left after pre-whitening by the complex of frequencies near 2.9 mHz.

Table 2. The p-values of the various Nyblom (1989) statistics, and the FDS, for the Feige 48 data ($N = 2103$). Each significance level was calculated from 5000 bootstrap samples.

Data set	Significance levels of statistics:			
	$L(C, \phi)$	$L(C)$	$L(\phi)$	FDS
1-2103	0	0	0.19	0.002
1-1402	0	0	0.018	0.001
702-2103	0.46	0.41	0.46	0.90

8 CONCLUSIONS

A few points are worth remarking on.

(i) The Nyblom statistics $L(C)$ and $L(\phi)$, respectively, have good power against changes in the amplitude and phase of a sinusoid. The statistic $L(\phi)$ also has good power against frequency changes.

(ii) The general form of the Nyblom test statistic is optimal in the sense that it is the most powerful test for small parameter changes of certain common types (jumps and random walks being two instances). It is therefore not surprising that it outperforms the FDS. The

results in Fig. 8 (a linear amplitude increase) shows that the Nyblom statistic is also powerful for other types of deviations from the null hypothesis.

(iii) The FDS is an 'omnibus' statistic, in the sense that it really tests for a change in any of the three parameters (amplitude, frequency and/or phase) characterizing a sinusoid. In that sense, it should perhaps be compared to the time domain statistic in (31), rather than to $L(C)$ or $L(\phi)$. A disadvantage of working in the frequency domain is that it gives no direct information as to which parameter is subject to change.

(iv) Certain assumptions need to be satisfied in order for the Nyblom (1989) statistics to have the asymptotic distribution giving rise to the percentage points in Table 1. Furthermore, the sample sizes needed for convergence to the asymptotic limit may be very large. It therefore seems prudent to rely on bootstrapping to determine significance levels.

(v) Commonly, the noise variance of astronomical data is not constant – for example, it usually increases with increasing airmass. Examination of (11), (17) and (31) shows that the high-variance residuals will lend greater weight to the more noisy measurements. This could be counteracted by estimating a function $V(t) = \sigma_e^2(t)$ which describes the variations in the variance, and then replacing $\widehat{e}_j$ by the standardized value $\widehat{e}_j/\sqrt{V(t_j)}$. It seems likely that the FDS will be less affected by this problem.

(vi) It may seem more useful to test separately for amplitude and frequency/phase changes. However, significance levels are compromised by repeated hypothesis testing. Formally, the correct procedure is to apply a single test [such as (31)] for parameter changes: if the null hypothesis of 'all parameters are constant' is rejected, then more detailed testing is warranted in order to establish which parameter(s) is variable.

(vii) Due to the many factors involved – series length; time spacing of the measurements; signal-to-noise ratio; phase and frequency of the sinusoid(s); the character of the noise process and the nature of any parameter changes – it is not possible to draw completely general conclusions about the performance of the test statistics. None the less, the power studies presented in this paper can serve as a first introduction to the potential of the techniques covered.

ACKNOWLEDGMENTS

The author is grateful to Professor Fred Lombard (University of Johannesburg) for discussions of the topic of this paper. The referee's comments led to a substantial improvement in the paper.

REFERENCES

Chatfield C., 2003, The Analysis of Time Series: An Introduction, 6th edn. Chapman & Hall/CRC Press, Boca Raton
Efron B., Tibshirani R. J., 1993, An Introduction to the Bootstrap. Chapman & Hall, London
Koen C., 2005, MNRAS, 361, 887
Koen C., 2007, MNRAS, 376, 233
Koen C., O'Donoghue D., Pollacco D. L., Nitta A., 1998, MNRAS, 300, 1105
Loumos G. L., Deeming T. J., 1978, Ap&SS, 56, 285
Mood A. M., Graybill F. A., Boes D. C., 1974, Introduction to the Theory of Statistics, 3rd edn. McGraw-Hill, New York
Nyblom J., 1989, J. Am. Stat. Assoc., 84, 223
Reed M. D. et al., 2004, MNRAS, 348, 1164

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