DFT, LS or GLS? A 'road test' of three period-finding techniques applied to stellar photometry

C A Engelbrecht[1,3] **& F A M Frescura**[2]
[1]Department of Physics, University of Johannesburg, Kingsway, Auckland Park, Johannesburg 2092, South Africa
[2]School of Physics, University of the Witwatersrand, Private Bag 3, WITS 2050, Johannesburg, South Africa

E-mail: chrise@uj.ac.za

Abstract. The mathematical properties of harmonic functions have made Fourier-based algorithms very popular in searches for periodic behaviour in stellar light curves and other astronomical data. The Discrete Fourier Transform and the Lomb-Scargle periodogram have been in common use for many decades, as methods particularly suitable for the non-equally-spaced time data that are typical of astronomical measurements. The "Generalised Lomb-Scargle" or GLS method is a recent refinement of the Lomb-Scargle method. Recent tests of these methods have emphasised the strengths and deficiencies of each. These test results demonstrate that certain methods are distinctly unsuitable for certain types of data and could lead to erroneous conclusions about the types of periodic behaviour present in measured data. Given the rapidly growing body of time-domain data in astronomy and the considerable importance of some of the conclusions that have been made on the basis of these data, the recent developments in the study of period-finding algorithms are significant. This paper reports on a comparative study of the above-mentioned methods applied to simulated stellar photometry compiled for three observing scenarios: space-based observation, ground-based observer-driven campaigns, and ground-based survey programs.

1. Introduction

In recent years, asteroseismology has made enormous contributions to the depth and scope of our understanding of stellar structure and evolution. For example, a very powerful correlation between the spacing of g-modes in the frequency domain and the dominant nuclear energy source (i.e. helium-core or hydrogen-shell fusion) in red giants was reported in [1]. Accurate determination of pulsation frequencies, amplitudes and phases is an essential component of such breakthroughs in asteroseismology. Two widely applied 'workhorses' of period determination are the Discrete Fourier Transform (DFT) adapted for non-equally-spaced time series, and the Lomb-Scargle (LS) periodogram. A number of diverse new approaches for determining the features of periodic behaviour in stellar magnitudes have been reported in recent years. One such approach is the "Generalised Lomb-Scargle" (or GLS) method [2]. This reference will be referred to as ZK09 in what follows. Following on earlier work by [3], who introduced a floating mean into the LS periodogram calculation, ZK09 added a weighting procedure to the floating mean procedure to accommodate observations that do not all have exactly the same precision. They concluded that, compared with the

[3] To whom any correspondence should be addressed.

classical LS periodogram, the GLS periodogram provides a more accurate frequency determination, is less susceptible to aliasing, and gives a much better determination of the spectral intensity. A recent, comprehensive comparison of various period-finding methods by [4] provides strong endorsement for the accuracy of the GLS algorithm.

2. Simulated light curves

In order to explore the relative merits of the DFT, LS and GLS methods, three different observing scenarios were simulated. The following parameters are common to the three simulated time series: the total time-span of the observations is 110 days and a total of 1000 observations are taken during this time-span. These choices were informed by the authors' research interest in pulsating B stars, for which the chosen values are typical of actual observing scenarios that produce useful asteroseismological results for such stars.

The following three realistic observing scenarios were chosen:

i) Scenario 1: A strictly equally-spaced time sampling over the 110 days, as would be possible to schedule on a space-based telescope capable of 24-hour observation of the same object. The observing protocol of the original mission design of the *Kepler* space telescope is similar to this scenario (although *Kepler* – in its 'Long-Cadence' observing mode - would obtain approximately 5200 equally-spaced observations over a 110-day period, instead of the chosen value of 1000 in this simulation). With the chosen numbers, the equal time-spacing between successive observations is 0.11 days, or 2.64 hours. This implies a Nyquist frequency of just over 4.5 cycles per day (c/d). All of the calculated periodograms in this study were therefore terminated at a frequency value of 4.5 c/d. This equally-spaced observing scenario is perfectly suited to the DFT method and we expected the DFT periodograms to perform just as well as the LS and GLS periodograms in this scenario.

ii) Scenario 2: A typical ground-based observing campaign, consisting of five full weeks (typical of the small-telescope observing allocations of the South African Astronomical Observatory at its Sutherland site in the Northern Cape) spread over the 110-day time-span and consisting of varying numbers of photometric nights per week. In the simulation, a total of 17 photometric nights were chosen among the 35 days of allocated observing time. Again, this is typical of actual conditions at Sutherland. The actual sampling times were adapted to account for the advance of sidereal time over the 110 days of calendar time, which accounts for the earlier rising of target stars at the end of the observing campaign than at the beginning. Within a night, the observations were taken at equal time intervals of 7.49 minutes. This number was determined by requiring 1000 observations to be taken in 17 photometric nights. Again, this time-spacing is typical of ground-based, observer-driven observations of pulsating B stars.

iii) Scenario 3: A ground-based, pre-programmed survey project with no observer input, observing the target star a few times each day (and moving to other survey locations in between). A random component was introduced in selecting the actual times of observation on any particular day. On most nights, the total number of observations taken per night was 8, 9 or 10. The spacing between successive observations was *not equal* and varied by more than an order of magnitude.

The simulated light curves shown in figure 1 illustrate the nature of the sampling regimes for the three scenarios.

Scenarios 1, 2 and 3 determined the *time* vectors in the simulated light curves. The *signal* vectors were calculated as follows: As a first step, three sinusoids with amplitudes, frequencies and phase

differences typical of pulsating B stars were added together. The value of this cumulative 'clean' signal was evaluated at the calculated time values obtained in each of the respective scenarios. Secondly, an underlying mean signal level as well as a random error (different for each individual signal value (i.e. for each individual 'observation') were added to these 'clean' signal values. The magnitudes of the random errors were chosen to scatter around typical mean values for small-telescope photometry at Sutherland.

The calculated light curves were then fed into the mathematical algorithms for calculating DFT, LS and GLS periodograms and the results were compared. These results are discussed in the following section.

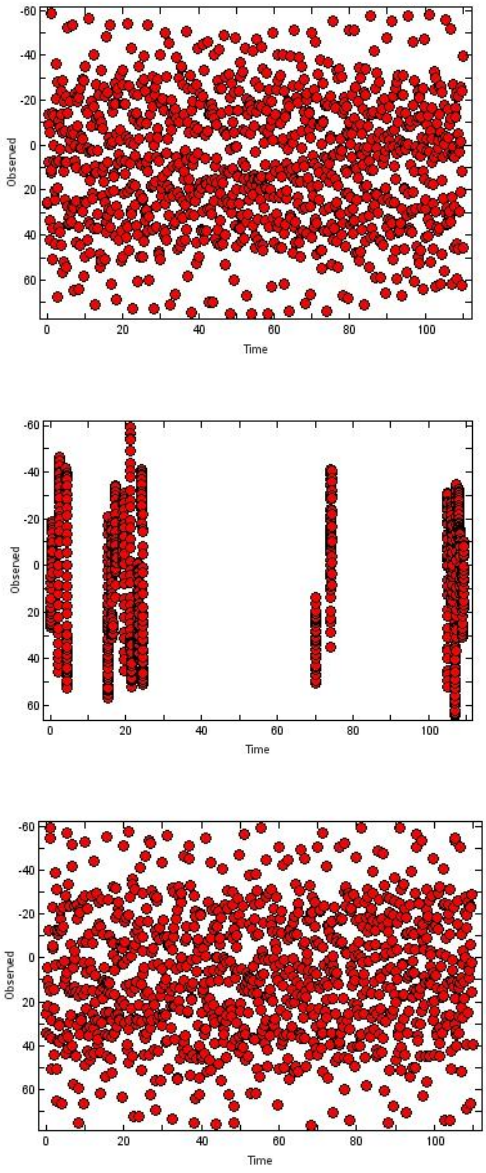

Figure 1. A sample of light curves generated for Scenarios 1, 2 and 3 (from top to bottom) respectively. Since the signal has a random noise component, the light curves will appear slightly differently every time they are calculated.

3. Retrieving periods

The ability of the DFT, LS and GLS periodograms to retrieve the input parameters of the periodic variations present in the simulated light curves was explored for various signal-to-noise (SNR) ratios and differing degrees of homogeneity of the signal errors. A brief summary of results is presented in this section.

Result A: Firstly, we considered typical SNR values for ground-based asteroseismology of B stars and a homogeneous error distribution across the entire observing epoch. As expected, the three methods all succeeded in retrieving the input frequencies and amplitudes of harmonic variation. Initial periodograms (i.e. before prewhitening) for Result A are shown in figure 2.

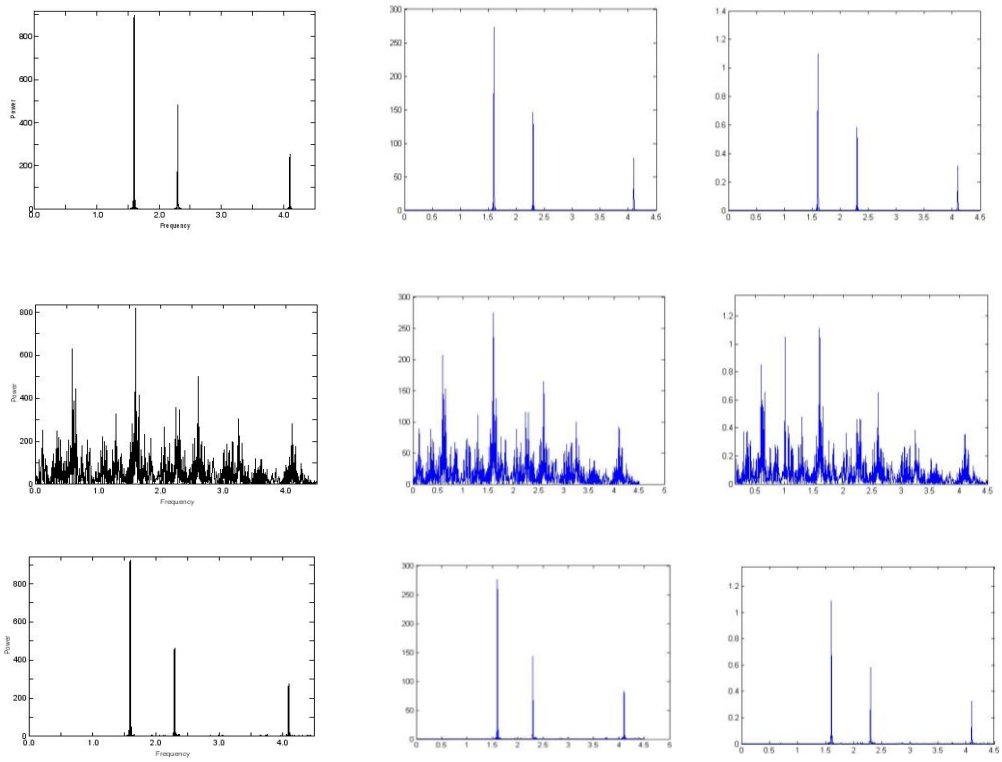

Figure 2. Periodograms obtained with DFT, LS and GLS algorithms respectively, for result A. First row: scenario 1 (equal time spacing); second row: scenario 2 (ground-based observer); third row: scenario 3 (ground-based survey). In each row, the DFT, LS and GLS periodograms appear from left to right respectively. In the second row, the daily gap aliasing is clearly seen. There are (mostly) subtle differences between the periodograms in this row, but all three input frequencies are successfully retrieved. One strange feature is the 1 c/d peak (with no aliases) in the GLS periodogram in the second row.

Result B: As one of the stated advantages of the GLS periodogram is its diminished vulnerability to aliasing, we deliberately constructed a time series that produces severe aliasing in the DFT periodogram by simply removing every second day's data from the equally-spaced time series (scenario 1). The time series for the other two scenarios were also reduced by 50% in a quasi-regular way. Initial periodograms (i.e. before prewhitening) for Result B are shown in figure 3.

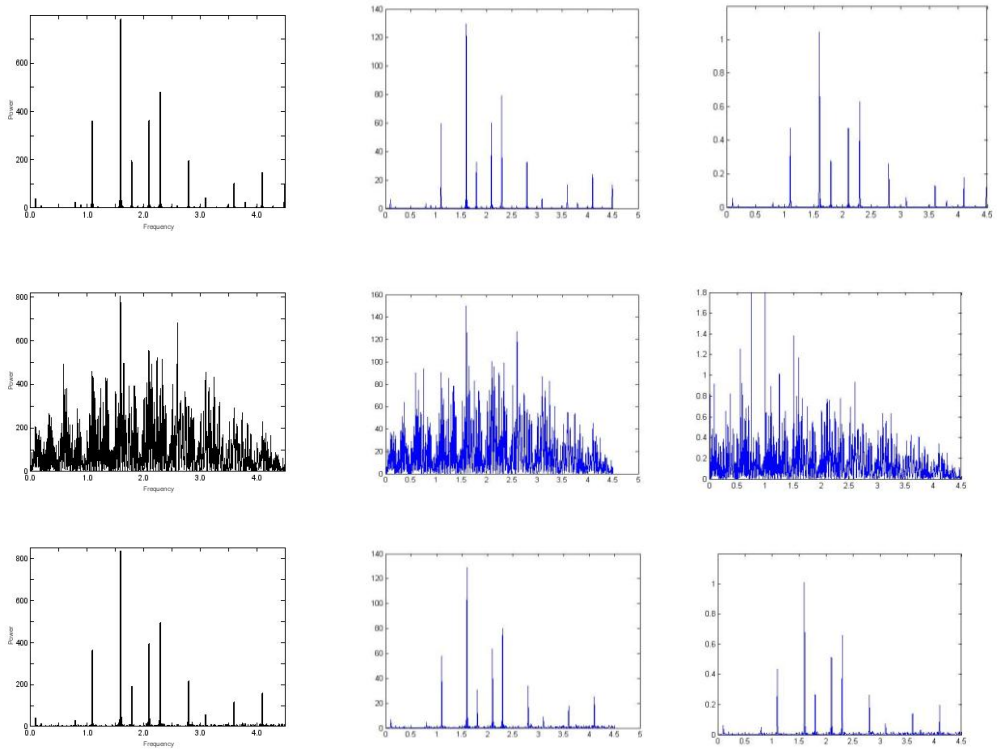

Figure 3. Periodograms obtained with DFT, LS and GLS algorithms respectively, for Result B. First row: scenario 1 (equal time spacing); second row: scenario 2 (ground-based observer); third row: scenario 3 (ground-based survey). In each row, the DFT, LS and GLS periodograms appear from left to right respectively. In the first (DFT) plot in the first row, we do indeed see a lot of aliasing, as expected (eight very prominent peaks appear in the periodogram, whereas only three periodic functions are used to construct the time series). However, the LS and GLS periodograms alongside are equally badly affected. In this instance, the GLS algorithm does not appear to avoid aliasing any better than the DFT and LS periodograms do. The same argument applies to the second and third rows. The GLS periodogram in row 2 actually appears more affected by aliasing than the DFT and LS periodograms are.

Result C: As the GLS periodogram is specifically designed to deal with varying amounts of signal noise, a third experiment was run with the time series divided into three blocks - with the mean noise levels chosen to lie in the ratio 4:2:1 among the blocks. The highest noise level was chosen to be larger than two of the three input amplitudes of the harmonic signal in the simulated time series. Initial periodograms (i.e. before prewhitening) for Result C are shown in figure 4.

4. Discussion and conclusions

The three experiments described as Result A, Result B and Result C in the previous section have not supported the strong claims made in favour of the GLS algorithm by ZK09. For the various time series subjected to the tests reported here, there is no advantage in choosing the increased processing time involved in calculating GLS periodograms, as far as period detection is concerned. In fact, in the experiment which was expected to favour the GLS periodogram (Result C), it was of no practical use. It is accepted that there might be time series formats for which GLS does offer a distinct advantage. The test protocol described in this paper may be adapted to other time series constructs and tested in similar fashion.

It will be of value to the practitioners of asteroseismology to be informed of the relative successes of the methods considered here in other scenarios.

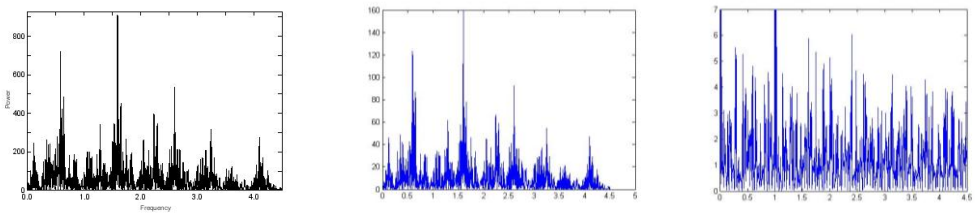

Figure 4. Periodograms obtained with DFT, LS and GLS algorithms respectively, for Result C. Only scenario 2 is shown, as the other scenarios add nothing new. Whereas the DFT and LS periodograms retrieve the input signal amongst some aliasing, the GLS periodogram is effectively of no use in this scenario.

References

[1] Bedding T S et al. 2011 *Nature* **471** 608
[2] Zechmeister M. and Kuerster M 2009 *A&A* **496** 577
[3] Cumming A, Marcy G W and Butler R P 1999 *ApJ* **526** 890
[4] Graham M J, Drake A J, Djorgovski S G, Mahabal A A, Donalek C, Duan V and Maker A 2013 *MNRAS* **434** 3423