Issues in Frequency Analysis of δ Scuti Stars I—HD 39641

Bill Rea
Richmond, New Zealand; rea.william@gmail.com

Received September 28, 2021; revised January 17, April 1, 2022; accepted May 3, 2022

Abstract We examine the consequences for the frequency analysis of δ Scuti stars of using pre-whitening if the assumptions of stationarity and sinusoidal light curves were violated. We show through numerical simulation that if the assumption of stationarity is violated then very large numbers of spurious frequencies may be generated which span the entire frequency range of the analysis. We also show that if the light curve is asymmetric, even when regularly spaced, highly significant but entirely spurious frequencies are generated. We apply the results of the numerical simulations to the δ Scuti HD 39641 and show that many statistically significant frequencies are artifacts of the data analysis process. Finally, we propose a method we call restricted range frequency analysis which uses existing tools, but aims to curb the worst features of frequency analysis when assumption violations are present in the light curve.

1. Introduction

This is the first of two papers which are intended to lay out for the amateur short period variable star observing community some common problems which arise in frequency analysis in non-mathematical terms and further illustrate these through the analysis of the δ Scuti HD 39641. There is a long literature addressing this set of problems, and readers with the necessary mathematical and statistical skills to study this literature would certainly benefit from doing so. A small sampling of these papers are Deeming (1975), Lomb (1976), Scargle (1982), Mary (2005), Balona (2014a), Pascual-Granado (2018), and VanderPlas (2018). Further references can be found within the citations contained in these papers.

Throughout the long history of the American Association of Variable Star Observers (AAVSO), amateurs have observed numerous varieties of pulsating stars ranging from long period variables, such as Miras with periods ranging from hundreds to thousands of days, through to very short period variables such as ZZ Ceti with periods as short as around two minutes. Among the tools provided by the AAVSO is the VStar (Benn 2012) light curve analysis software which is often cited in articles published in this journal. There are a number of other frequency analysis tools freely available such as FAMIAS (Zima 2008) and SigSpec (Reegen 2011), described in section 2, which are the two used in this paper.

For stars pulsating in only one or two modes, in general, few problems arise with frequency analysis. However, for stars pulsating in numerous modes, such as δ Scutis, serious problems may arise in interpreting the periodogram in terms of a model of the pulsations. Conceptually, the task is to distinguish between statistically significant peaks in the periodogram which correspond to physical pulsations and statistically significant peaks which arise from some combination of the method of analysis and the spacing of the observations but do not correspond to a physical pulsation.

This problem has been known for a long time. For example, Lomb (1976) pointed out that no more than one period or, equivalently, frequency can be estimated from any one calculation of the frequency spectrum. This necessitates successively identifying periodicities and subtracting them out of the data and then recalculating the frequency spectrum from the residuals, a process usually called pre-whitening. In their textbook on asteroseismology Aerts *et al.* (2010, p. 342) state that "A pre-whitening strategy thus has to be chosen to perform the frequency analysis."

The science of asteroseismology uses the observations of the light curve first to identify the pulsation frequencies, amplitudes, and modes and then to infer the internal structure of the star (Aerts *et al.* 2010). The so-called hybrid γ Doradus/δ Scuti pulsators are of particular asteroseismological interest because they exhibit both g- (or gravity) mode and p- (or pressure) mode pulsations. The g-mode pulsations originate and propagate deep in the interior, allowing the modelling of interior, perhaps even to the core. The p-mode pulsations originate and propagate in the outer convective layers, allowing that portion of the star also to be modelled. Thus, taken together, they may allow the asteroseismologist to model the star from its surface to its center (Goupil *et al.* 2005).

Pre-whitening, described above, when regarded as a statistical method, appears to have three underlying assumptions, namely: (1) the identified frequencies are generated by sinusoids; (2) they are stationary, in both amplitude and frequency; and (3) where multiple frequencies are present they are combined into the final light curve in an additive manner. Stationarity is defined precisely in many textbooks on time series analysis, see Bloomfield (2000, pp. 167–173) for one such definition. For our purposes it is the assumption that the frequencies, modes, and amplitudes of the pulsations do not change on the time scale spanned by the data.

While pre-whitening sounds a reasonable approach, even when all three assumptions are met, Balona (2014b) was able to show, through numerical experiments, that such a method can extract from the data large numbers of statistically significant frequencies but which do not correspond to any actual frequency in the data set. These are often referred to either as spurious or fictitious frequencies (Balona 2014b). The origins of these spurious or fictitious frequencies can be understood if we think of the pre-whitening process as adding a frequency to the data which is intended to interfere destructively with a particular frequency estimated from the data. Any errors in the estimation of the frequency, amplitude, and phase, together with the fact

that the light curve of the frequency identified may not be sinusoidal in shape, will result in the intended interference not being completely destructive, and so many further frequencies are added to the residuals, which are then used to estimate another significant frequency. This is particularly problematic in the analysis of δ Scutis because their multiple active frequencies require multiple cycles of pre-whitening.

Balona (2014b) constructed simulated time series (light curves) with known frequencies and amplitudes, and reported that often frequency analysis employing pre-whitening did not recover all of the real frequencies while extracting more spurious frequencies than there were real frequencies present in the data. Balona (2014b) used stationary sinusoids which he combined additively in his simulations. As we show below, the problem is potentially much worse if any of the three assumptions are violated.

The problem of non-stationary amplitudes is well-known in δ Scuti pulsators. A number of papers report significant changes in the amplitude of one or more pulsation frequencies. For example, Bowman *et al*. (2016) studied 983 δ Scuti stars observed by *Kepler*, and reported that 61.3% exhibited amplitude modulation. If stars with amplitude modulation were subjected to a frequency analysis using pre-whitening, the consequences of this type of assumption violation needs to be taken into account, but it is likely that many serious amateurs are not fully aware of the potential problems.

These findings pose a significant problem for the analysis of δ Scuti pulsators because they are expected to have large numbers of frequencies excited requiring many cycles of pre-whitening; see Uytterhoeven *et al.* (2011) and Poretti *et al.* (2009) for examples.

While the identification of high-amplitude pulsation peaks in a δ Scuti periodogram is not an issue—they can often be identified visually without the need for sophisticated statistical analysis—this alone does not mean they are physically meaningful. Kurtz *et al*. (2015) were able to show that for slowly pulsating B-stars (SPB) combination frequencies can have amplitudes larger than the physically meaningful base frequencies which generate them.

In δ Scutis the number of modes excited is expected to increase as their pulsation amplitudes decrease, making it challenging to distinguish between physically meaningful frequencies present in the data and spurious frequencies which are an artifact of the method(s) of data analysis.

Turning now to the assumptions that the pulsations generate sinusoidal light variations, with multiple pulsations generating multiple sinusoids which are summed to form the observed light curve, it is well known from the study of Fourier series that a non-sinusoidal periodic function will require multiple sinusoids of decreasing amplitude to model it. Thus a pulsation generating an asymmetric variation in light output will be split into multiple frequencies of the form f, 2f, 3f, and so on to the maximum allowable frequency specified by the method. As we show below, these frequencies can be highly significant. These types of frequency spacings have been reported in the literature, for example see Poretti *et al.* (2009) Table 3 where their f_1 also has significant frequencies at $2f_1$ and $3f_1$. If f_1 was an asymmetric pulsation, then frequencies $2f_1$ and $3f_1$ may well be artifacts of the data analysis method, each requiring a pre-whitening cycle to model and remove it. In such a case they may be useful in trying to quantify the asymmetry but have no other physical significance.

In this paper we lay out the effect that amplitude modulation and asymmetric light curves have on frequency analysis with pre-whitening and then apply this to an analysis of the δ Scuti star HD 39641. Towards the end of the paper we mention one possible way to avoid some of the problems with frequency analysis with pre-whitening which will be examined in more detail in Paper II.

The remainder of the paper is structured as follows: section 2 presents the numerical experiments and their results; section 3 discusses the observational data and data analysis issues; section 4 presents the results from the observational data; section 5 contains the discussion; section 6 contains our conclusions and suggestions for future research.

2. Numerical experiments

The purpose of this section is to generate some fairly simple synthetic light curves of known frequency composition and which resemble, in a somewhat idealized form, changes in light curve structure observed in some δ Scutis with references to the original paper which inspired the particular simulation. The particular violations of assumptions underlying pre-whitening are noted. From these we demonstrate that many more spurious frequencies are reported by the frequency analysis than were actually present in the simulated series.

2.1. Data generation

Seven simulated time series (light curves) were generated and analyzed which examined several possible sources of amplitude or frequency modulation and light curve asymmetry. While the units of the simulated series are arbitrary, they were chosen to mimic the 2-minute (120-second) cadence observation mode of the NASA Transiting Exoplanet Survey Satellite (TESS) (Ricker *et al.* 2014) of a δ Scuti with a main pulsation frequency of 11.25 cycles d^{-1}, or a period of 128 minutes, but without the data gaps. All simulated time series or light curves consisted of 131,072 (2^{17}) data points. The sampling time was once every two minutes, which yielded 64 data points or samples per pulsation cycle. The simulated series contained no added noise. The simulated series were as follows:

1. *Stationary* Two time series with two different frequencies were generated. The main frequency of 11.25 cycles d^{-1} with amplitude 1.0 and a second frequency of 11.206226 d^{-1} with amplitude 0.5. The two series were added, yielding a series with obvious beats between the two frequencies. The resulting time series was stationary and no analytical assumptions were violated.

2. *Non-linear* The main frequency was generated as in method 1, and a second frequency with a period 45.51 days and amplitude of 0.5 was generated. The final series was produced by multiplying the two series. This violates the assumption that frequencies are combined additively.

3. *Amplitude reduction* The main frequency was generated as in method 1 and its amplitude remained constant over

45.51 days (32,768 data points, 2^{15}). The amplitude was then linearly reduced to 0.5 over the next 45.51 days, after which it remained constant until the end of the series. The assumption of a stationary amplitude was violated. This experiment was motivated by Figure 1 of Bowman *et al.* (2016).

4. *Slow rise, rapid fall* The main frequency was generated as in method 1 and was modulated by a slow increase in amplitude from 0.5 to 1.5 over 12,288 data points (17.0667 days), followed by a rapid decrease over 4,096 data points (5.6889 days), giving eight complete cycles of amplitude changes. In this simulation there were three frequencies present; the two amplitude modulation frequencies were subject to regime switching so that only two frequencies were present at any one time. The assumption of stationary amplitude and frequency was violated. This experiment was motivated by Figure 2 of Barceló Forteza *et al.* (2015).

5. *Slowly changing frequency* In this simulation the central frequency was set to 11.25 cycles d^{-1} as with the other simulations but a second frequency was used to directly modify the main frequency, which caused it to move through eight cycles between 11.20 cycles d^{-1} and 11.30 cycles d^{-1}. The assumption of stationarity of frequency was violated.

6. *Asymmetric light curve* The light curve was generated with a slow rise followed by rapid decline somewhat akin to a mirror image of a classical Cepheid light curve. This was done with two frequencies with regime switching so that only one frequency was active at any point in the series. Two variations of this were tried, one with a ratio of 3:1 (rise:fall) and the other with a ratio of 9:7. Their periodograms were obviously different but the number of significant frequencies was identical for both cases. The assumption of sinusoidal light variation was violated.

Frequency analysis was primarily carried out using two packages:

1. *SigSpec (Significance Spectrum)* (Reegen 2011) is a batch process frequency analysis tool. The user supplies the input data file and sets a range of options and parameters in an initialization file. SigSpec reads both and completes the frequency analysis without further user intervention.

2. *FAMIAS (Frequency Analysis and Mode Identification in Asteroseismology)* (Zima 2008) was also used because it is fully interactive, meaning that the user must choose the data and analysis options at each step of the analysis process. FAMIAS also stores the output of the intermediate steps so it is easy to go back to re-examine earlier periodograms and pre-whitened data sets and rerun the analysis with different options.

SigSpec uses the Lomb-Scargle periodogram (Lomb 1976; Scargle 1982) as the basis for its frequency analysis; see VanderPlas (2018) for an informative paper on this particular periodogram. See section 3.4 for further details.

2.2. Results

The results of these experiments are presented in Table 1 and Figures 1 through 3.

Table 1 reports the number of significant frequencies (sig.≥4) extracted by SigSpec (Reegen 2011) for each of the six numerical experiments described in section 2.1. The third column reports the actual number of frequencies present in the data, noting that in the "slow rise, rapid fall" that two of the three frequencies undergo regime switching so that at any particular point only two frequencies are active.

Table 1. The number of statistically significant frequencies reported by SigSpec (Reegen 2011).

Simulation Number	*Simulation Type*	*Real Frequencies*	*Reported Frequencies*
1	Stationary	2	6
2	Non-linear	2	4
3	Amplitude Reduction	1	4668
4	Amplitude Slow Rise, Rapid Fall	3	1157
5	Slowly Changing Frequency	2	13
6	Asymmetric Light Curve	2	5

In Figure 1 panels a through d the periodograms correspond to simulations 1 through 4, respectively. The spectral window was identical in all cases and is plotted in red; the periodograms are plotted in black. The frequencies 0.15 cycle d^{-1} either side of the main 11.25-cycles d^{-1} frequency were plotted. Panel a shows the two frequencies of simulation 1 were clearly resolved. Panel b, of the non-linear amplitude modulation, shows that the periodogram mimics the mode splitting caused by stellar rotation for the l = 1 modes. Panel c, of the amplitude reduction, shows a significant amount of power was missing from the the central peak. This was widely spread throughout the periodogram. However, the fact that the power had been spread out is not the whole explanation for the 4,668 significant frequencies reported in Table 1; see the discussion of Figure 2.

Finally, panel d with the slow rise, rapid fall amplitude modulation, mimics the mode splitting caused by stellar rotation for the l = 2 modes.

Figure 2 presents the frequencies extracted by SigSpec (Reegen 2011) for simulations 3 and 4 in the order in which they were extracted. While the power in the periodograms was spread out, this is not the explanation for the pattern of frequencies seen in the Figure. This was due to the pre-whitening process creating a new frequency or frequencies with each pre-whitening cycle until the process was terminated by hitting the bounds of the frequency analysis.

Figure 3 presents the full periodograms (0 to 50 cycles d^{-1}) for the two asymmetric light curves. While SigSpec reported five significant frequencies for both simulated series, it is clear that the more asymmetric light curve gave much stronger signals at 2f, 3f, and 4f. The upper bound on the frequencies was 50 cycles d^{-1}. If the upper limit had been raised, further significant frequencies would have been reported, at least for the more asymmetric curve where raising the upper frequency limit to 100 cycles d^{-1} yielded nine significant frequencies.

3. Observational data, issues, and frequency analysis

3.1. HD 39641

HD 39641 is relatively bright, low amplitude δ Scuti variable star observed in TESS' continuous viewing zone. Basic details on the star are presented in Table 2.

3.2. Data

The raw data for this paper were downloaded from the TESS Asteroseismic Science Operations Center web site on 25

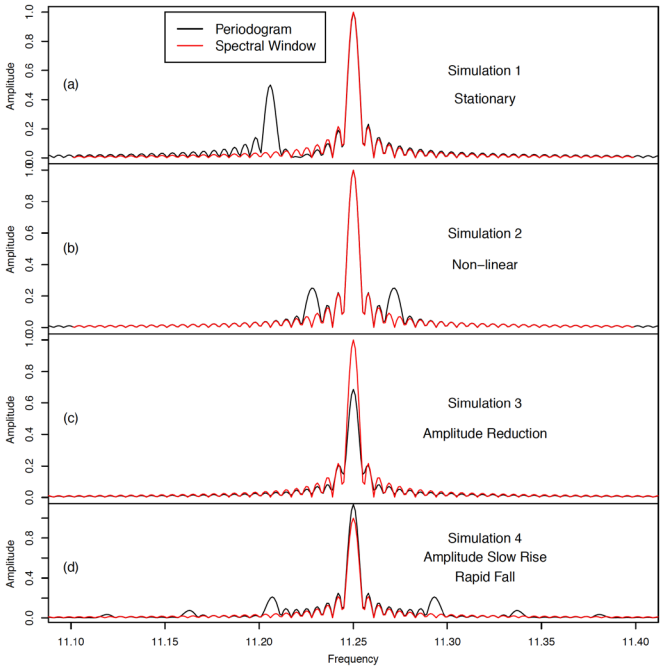

Figure 1. Panel (a) the spectral window and periodogram from simulation 1. Panel (b) the spectral window and periodogram from simulation 2. Panel (c) the spectral window and periodogram from simulation 3. Panel (d) the spectral window and periodogram from simulation 4.

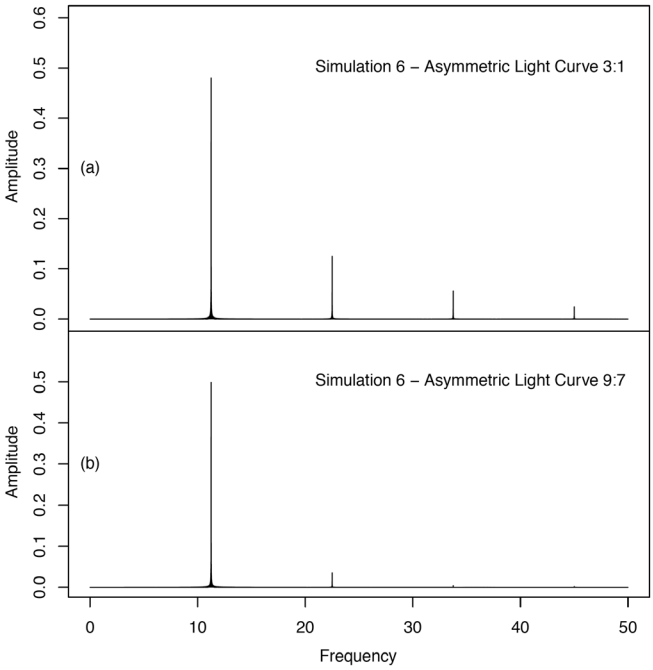

Figure 3. Panel (a) the periodogram for the 3:1 rise to fall ratio asymmetric light curve of simulation 6. Panel (b) the same for the simulation with the 9:7 rise to fall ratio asymmetric light curve.

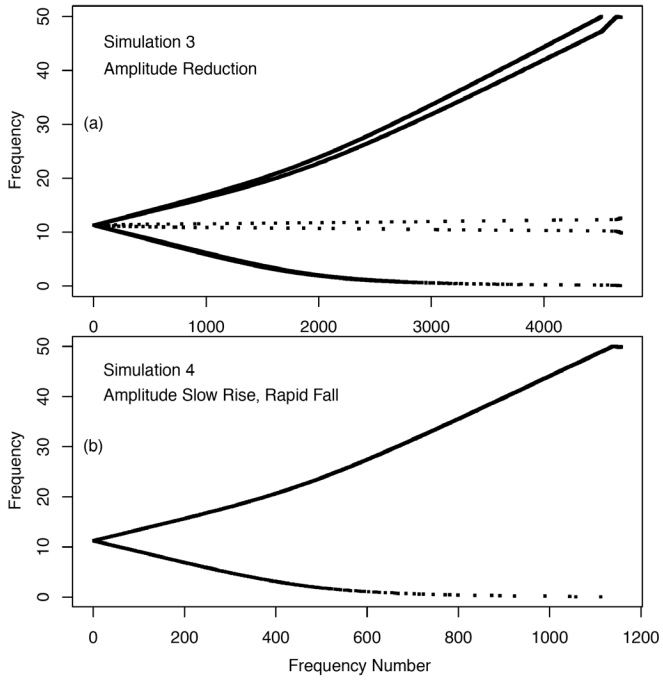

Figure 2. Panel (a) the statistically signicant frequencies reported by SigSpec (Reegen 2011) for simulation 3 in the order in which they were extracted. Panel (b) the same for simulation 4.

July 2020. HD 39641 was observed by TESS in its 120-second cadence mode with a total elapsed time of approximately 357 days, which yielded 211,932 usable data points. The observing runs had significant gaps on the order of one-and-a-half days, producing 26 segments of nearly continuous data which were analyzed both separately and as a combined data set. Subsequently, a further six segments of data became available from later observations and these were downloaded on 25 Dec 2020. These were analyzed separately from the first 26 segments.

The reported corrected flux was converted to magnitudes using the value for HD 39641's magnitude in the V band as reported in SIMBAD as the mean value for each observation run. Observations were discarded if the value in the Pixel Quality Field (PQF) was non-zero or either the date or the corrected flux was recorded as not-a-number (nan). The purpose of using the V band magnitude was to align the 26 data segments from HD 39641 to prevent statistically significant but spurious low frequencies appearing in the analysis of the combined data set. There are a range of other options available for aligning the data. An approximately five-day segment of data is presented in Figure 4.

3.3. The spectral windows

A TESS sector is observed continuously for two orbits except for an approximately day-long gap when the data are downloaded to the ground. These data gaps can give rise to the presence of aliases and/or strong side lobes which must be taken account of during any frequency analysis. Figure 5 presents two views of the spectral window of the complete (357-day) data set generated by FAMIAS (Zima 2008) and the spectral window for the data from the first orbit of the TESS mission. In

Table 2. Basic stellar data for HD 39641.

Property	*Value*
R.A.	$05^h\ 49^m\ 46.689^s$
Dec.	–60° 48' 19.939"
Spectral Type (SIMBAD)	A3mA7-A8
Spectral Type (VSX)	A5mF1
Period (VSX)	0.048869 days / 70.371 min
Distance	227.8 ± 1.25 pc
Mean Mag (V)	9.3 ± 0.02
Amplitude (V)	0.015
TIC	149630117

Note: The data was drawn from the TASOC web site (https://tasoc.dk/), SIMBAD (http://simbad.u-strasbg.fr/simbad/) (Wenger et al. *2000), and the Variable Star Index (VSX, https://www.aavso.org/vsx/). The distance is from the GAIA DR2 catalogcitep (Gaia Collab. 2018). TIC is the TESS Input Catalog number.*

Table 3. The frequencies and height of the main peak relative to the alias and each of the side lobes seen in Figure 5.

Alias Frequency cycles d^{-1}		*Relative Height to Main Peak*
± 0.07392		8.1689
Side Lobe Frequency cycles d^{-1}	*Frequency Step*	*Relative Height to Main Peak*
± 0.00392	—	4.4069
± 0.00672	0.00280	6.0144
± 0.00952	0.00280	9.0188
± 0.01288	0.00336	10.8069
± 0.01792	0.00504	13.5853

Table 4. Types of statistically significan frequencies in combined data set of HD 39641.

Frequency Type	*Number*
δ Scuti	1,166
γ Doradus	297
Others	49
Total	1,512

panel (a) the main peak is approximately 8.1 times the height of the alias. Inverting the frequency difference of the alias from the peak yields 13.52 days, which closely matches Tess's 13.7-day orbital period. In panel (b) of the Figure there are five noticeable side lobes; details of these are given in Table 3. The FWHM of the central peak was 0.00336 d^{-1}, meaning that frequencies separated by less than 0.00168 d^{-1} are unlikely to be resolved. Within the individual segments the FWHM varied from segment to segment, but for comparison purposes the frequency resolution the first segment was 0.045372 d^{-1}. Panel (c) presents the spectral window from the data from the first orbit of the Sector 1 observations. The much broader frequency peak, indicating lower frequency resolution, is because the time base line of the observations is much shorter than in the combined data set.

This paper presents the analysis both of the complete data set and of each individual TESS orbit. Details of the date range and number of data points from each orbit are given in Table 5.

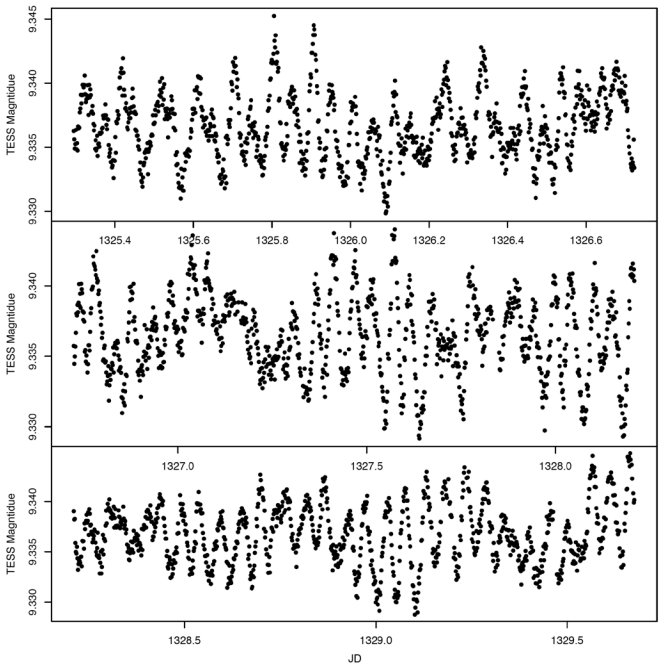

Figure 4. Approximately five days of HD 39641's light curve from the TESS data. The time axis is labelled JD and is the baricentric Julian day –2457000.

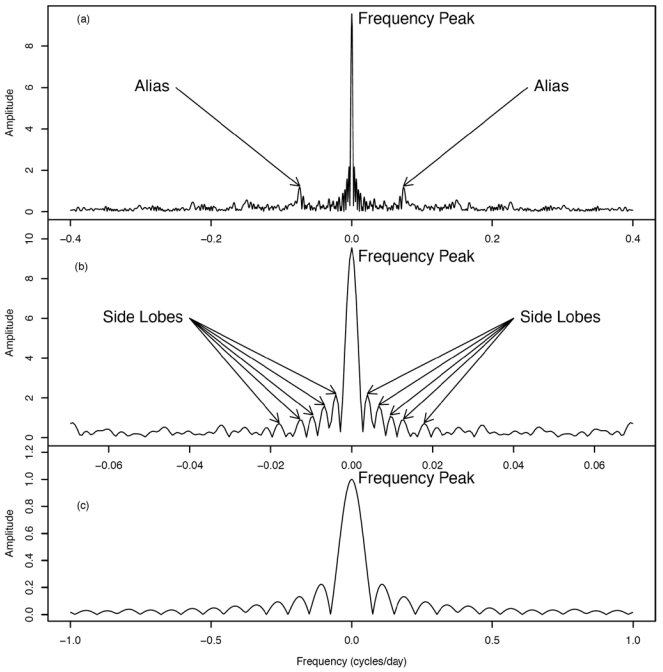

Figure 5. Panels (a) and (b) present the spectral window for the TESS data for the combined data set of HD 39641 as generated by FAMIAS (Zima 2008). Panel (a) shows the window to 0.4 cycles d^{-1} in which there is a single alias at 0:07392 cycles d^{-1}. Panel (b) shows a detail of the window to 0.07 cycles d^{-1} showing five side lobes at 0.00392, 0.00672, 0.00952, 0.01288, and 0.01792 cycles d^{-1}. Further details of the side lobes are presented in Table (3). Panel (c) presents the spectral window from the first orbit of Sector 1 of obsevations.

3.4. Frequency analysis

Similar to the simulated data, The frequency analysis was carried using FAMIAS and SigSpec, and with user-written code in Matlab (MathWorks 2019) or R (R Core Team 2019). FAMIAS had a limit of 100 frequencies which could be extracted from the data.

The default frequency range for both packages was 0 to 50 cycles d^{-1}. The minimum signal to noise ratio (SNR) used with FAMIAS was four. The minimum significance used with SigSpec was also four. These two measures are different, with an SNR of four being equivalent to a spectral significance of five.

We also carried what we are calling a "restricted range frequency analysis." While we are introducing this method in the current paper it will be the examined in some detail in a second paper. This used SigSpec's options to change the range of frequencies being analyzed and restrict this range to only encompass specific, relatively narrow, ranges. We only undertook two ranges. The first was 0.28 to 3.2 cycles d^{-1}, which is slightly wider than the traditional γ Doradus range; see Table 9.1 of Catelan and Smith (2015). This range was further subdivided into two halves, 0.28 to 1.74 and 1.74 to 3.2 cycles d^{-1}, and the results of the three analyses were combined to assess which frequencies were likely to be physically meaningful and which were like to be artifacts of the data analysis process.

The second was the 4.0 to 6.0 cycles d^{-1} range which appeared devoid of any significant pulsations in the periodogram from HD 39641's combined data set. Possible future directions for this method are discussed further in section 6.2 below.

4. Results

4.1. Observational data

Figure 6 presents a Fourier spectral analysis generated by FAMIAS (Zima 2008) of HD 39641's combined data set. The γ Doradus frequency range is marked. The vertical axis is the amplitude; this is different from the periodograms presented in Figures 8 through 11, in which the vertical axis is $\log_{10}$ (Amplitude).

An analysis of the combined data set was undertaken with SigSpec (Reegen 2011). 1,512 statistically significant frequencies were reported. A summary of their types is given in Table 4. Frequency of $f \geq 3$ cycles d^{-1} were classified as δ Scuti types, $0.3 \leq f < 3$ where classified as γ Doradus types and $f < 0.3$ cycles d^{-1} were classified as other.

Each observing run was also analyzed. Details of the date range, number of usable and unusable data points, numbers of statistically significant frequencies, and their types are reported in Table 5. It can be seen that some orbits within some sectors had a large number of usable data points. The row for JD 1410–1423 shows 2,073 unusable data points, a consequence of a continuous period of approximately 2.7 days in which all fields, apart from the PQF, were recorded as nan. The raw data for JD 1558.28–1568.47 had an approximately half-day period with non-zero PQF values.

Figure 7 presents a plot of the frequencies below 30 cycles d^{-1} in the order in which SigSpec (Reegen 2011) extracted them.

Table 5. A summary of the frequency analysis of the 32 sub-periods of the TESS data for HD 39641.

Date Range	*Usable Data Points*	*Unusable Data Points*	*Frequencies* δ *Sct*	γ *Dor*	*Others*	*Total*
1325.30–1338.52	9223	300	96	19	1	116
1339.66–1353.18	8875	857	94	21	2	117
1354.11–1367.15	9225	167	89	19	3	111
1368.20–1381.51	9090	207	93	18	2	113
1385.95–1395.44	6683	152	74	14	2	90
1396.64–1406.21	6767	127	78	14	1	93
1410.90–1423.51	7005	2073	69	13	3	85
1424.56–1436.84	8784	60	84	20	3	107
1437.99–1450.19	8746	35	84	17	2	103
1451.56–1463.94	8888	28	84	17	1	102
1468.27–1477.02	6263	37	67	14	1	82
1478.12–1490.04	8561	26	88	18	2	108
1491.63–1503.04	8190	23	90	19	2	111
1504.71–1516.09	8168	26	87	17	2	106
1517.40–1529.07	8379	29	87	16	3	106
1535.00–1542.00	5019	19	58	10	1	69
1544.88–1555.54	7653	22	81	16	0	97
1558.28–1568.47	6935	403	69	16	1	86
1571.06–1581.78	7695	28	77	16	2	95
1584.74–1595.68	7852	29	81	16	1	98
1599.85–1609.69	6989	100	74	15	2	91
1612.37–1623.89	8268	31	82	15	3	100
1624.96–1639.00	9894	213	90	20	2	112
1640.04–1652.89	9207	48	89	17	4	110
1653.92–1667.69	9767	148	96	20	4	120
1668.63–1682.36	9812	75	96	21	3	120
2036.28–2048.13	8521	11	88	18	2	108
2049.15–2060.64	8258	14	86	17	0	103
2061.85–2071.57	6988	16	78	12	2	92
2075.16–2085.54	7386	91	69	17	2	88
2115.89–2127.43	8295	18	80	19	2	101
2130.21–2141.82	8329	30	81	16	2	99

Note: The date range is reported to two decimal places for formatting reasons, TESS dates are precise to eight decimal places.

Figure 8 presents the evolution of the periodogram of HD 39641 over the 26 orbits of the combined data set in the γ Doradus frequency range. Plotting the periodograms in this manner allows us to examine change in the active frequencies over time. Some, such as the strong peak at approximately 2.71 cycles d^{-1}, are relatively stable, whereas others, such as the more complicated peak near 2.3 d^{-1}, may show changes in frequency, amplitude, and splitting of an apparent single frequency into two or more frequencies followed by their recombining. When looking for deviations from a sinusoid which may have given rise to a cascade of spurious frequencies, such as can be seen in Figure 7, the 2.3 d^{-1} frequency is a clear candidate.

Figure 9 presents the evolution of the periodogram of HD 39641 over the 26 orbits of the combined data set over the 9 to 12 cycles d^{-1} part of the δ Scuti frequency range. Similar to what was seen in Figure 8, the two strongest frequencies vary across time, sometimes splitting into two distinct frequencies. As observed in the numerical simulations, such behavior can give rise to many spurious frequencies during analysis.

Figures 10 and 11 present the evolution of the periodogram of HD 39641 over the 26 orbits of the combined data set over

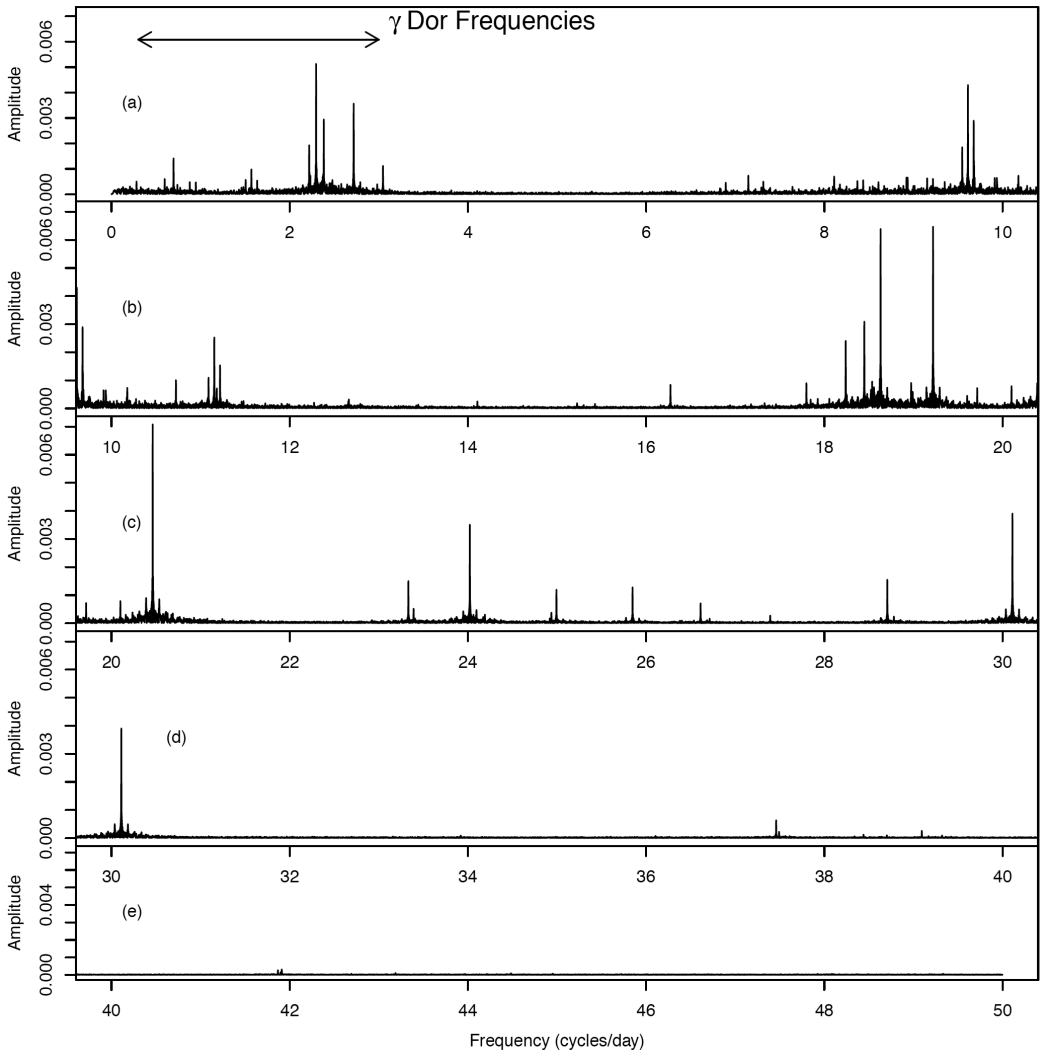

Figure 6. The Fourier spectral analysis of HD 39641's combined data set broken into five segments of frequencies. The range γ Doradus frequencies are marked in panel (a), the remainder of the frequencies in panel (a) and the frequencies in panels (b)–(e) are considered to be δ Scuti frequencies.

the 17.5 to 21 cycles d^{-1} and 23 to 31 cycles d^{-1} respectively. In contrast to both Figure 8 and Figure 9, the three strongest frequencies are stable both in amplitude and frequency. However, from this we cannot determine if these pulsations generate a sinusoidal light curve.

A search was made for highly significant frequencies (sig. ≥ 100), of which there were 84, which may be generating cascades of significant frequencies of the form $kf, k=1,2,...$, due to asymmetries in their contributions to the light curve. There were 13 such groups of frequencies and Table 6 reports the results of this search.

Figure 12 presents four details of the periodogram in Figure 6 where four sets of triplets were observed.

4.2. Restricted range frequency analysis

SigSpec (Reegen 2011), in common with many other frequency analysis tools, had the option of changing the range of frequencies being analyzed. Because it was clear from the results in the previous section that spurious frequencies were contaminating all of the frequency space, we restricted the range of frequencies as described above. We undertook two restricted range frequency analyses. The first was in the range 0.28 to 3.2 cycles d^{-1}, typical of the g-mode pulsations of γ Doradus stars. The periodogram of HD 39641's combined data set together with the periodogram of the residuals after the analysis was run are presented in Figure 13.

From the unrestricted analysis of the combined data set, there were 315 significant frequencies reported in the slightly enlarged γ Doradus range. This can be compared to the 48 significant frequencies from the restricted range frequency analysis.

In the second range, 4 to 6 cycles d^{-1}, the full frequency analysis yielded 59 significant frequencies, whereas the restricted range analysis yielded none. With no frequencies modelled and removed the periodogram was unchanged before and after the analysis.

5. Discussion

The results of the numerical experiments in section 2.2 clearly show that some types of non-stationarity pose a major problem for frequency analysis of δ Scuti stars using pre-

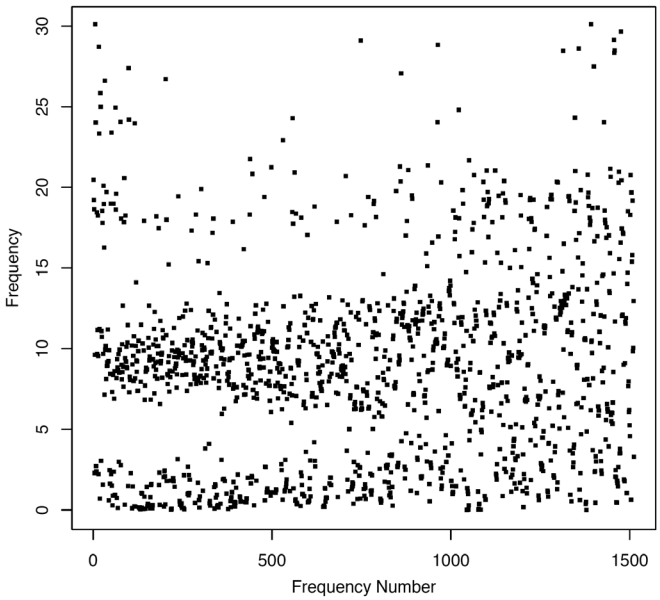

Figure 7. The statistically signicant frequencies as reported by SigSpec (Reegen 2011) in the order in which SigSpec extracted them.

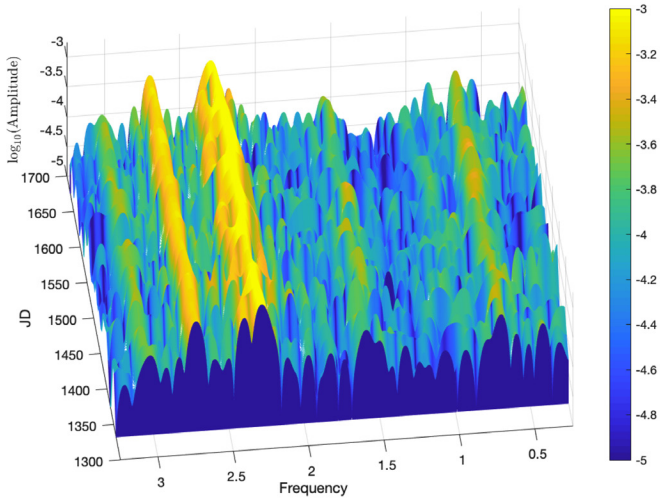

Figure 8. The evolution of the periodogram from the 26 data segments for the γ Doradus frequency range. The colorbar units are in $\log_{10}$ (Amplitude).

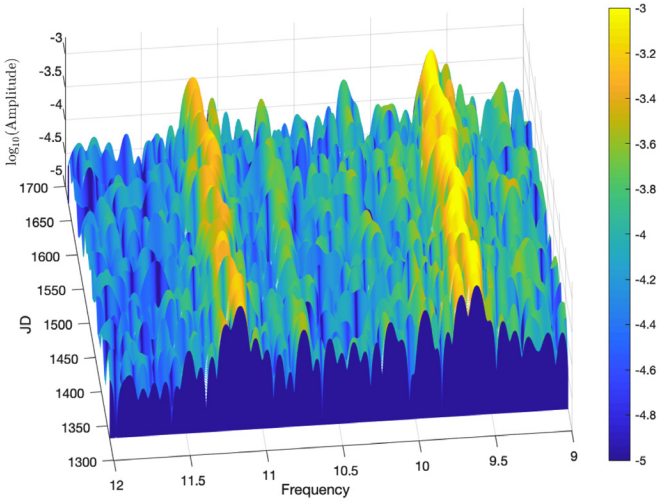

Figure 9. The evolution of the periodogram from the 26 data segments for part of the δ Scuti range of 9 to 12 cycles d^{-1}. The color bar units are in $\log_{10}$ (Amplitude).

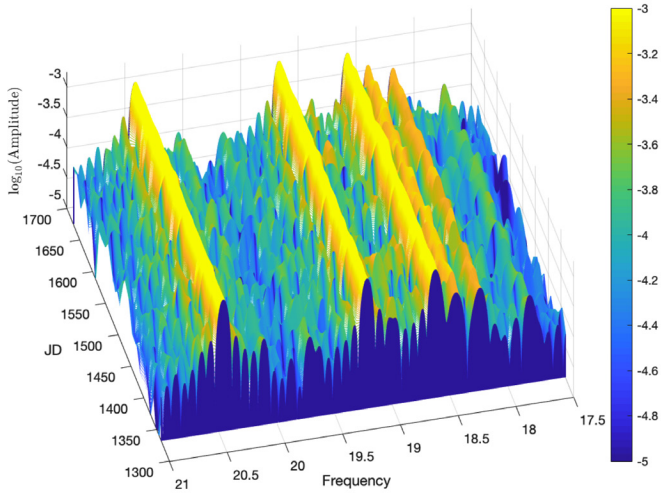

Figure 10. The evolution of the periodogram from the 26 data segments for the δ Scuti range of 17.5 to 21 cycles d^{-1}. The colorbar units are in $\log_{10}$ (Amplitude).

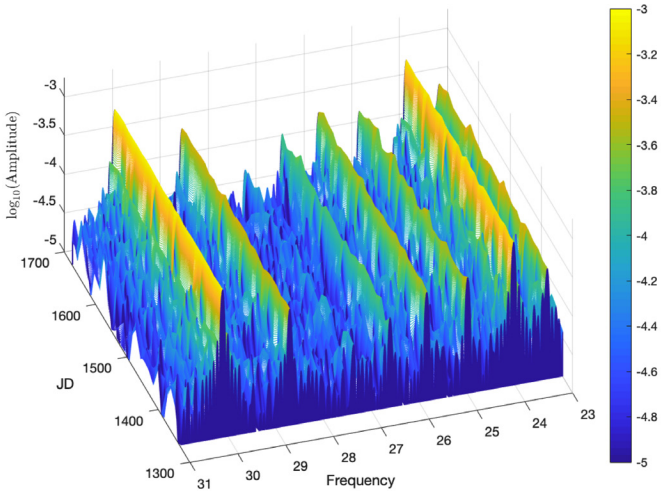

Figure 11. The evolution of the periodogram from the 26 data segments for the δ Scuti range of 23 to 31 cycles d^{-1}. The colorbar units are in $\log_{10}$ (Amplitude).

whitening. The two numerical experiments which yielded the largest number of statistically significant frequencies were motivated by results of the analysis of actual δ Scutis reported in the literature. If the reported amplitude modulation is not the result of very slow beating of otherwise stable frequencies, they have the potential to contaminate the entire frequency range with very large numbers of spurious frequencies.

If an apparent amplitude modulation is caused by two narrowly separated, but otherwise stable, frequencies then these frequencies will be able to be resolved with a long enough time series. However, similar to Balona (2014b), more spurious frequencies were reported than there were actual frequencies in the data in this case. Moreover, one of the spurious frequencies had a significance approximately equal to the two real frequencies (22,778.8 and 28,327.8 for the two real frequencies and 26,420.8 for the first spurious frequency). The likely consequence would be that in a frequency analysis of a real star we would be unable to conclude the frequency was an artifact of data analysis from checking the significances alone. In this case the periodogram was sufficiently simple, and hence sufficiently easy to interpret, that a researcher would be

able to correctly conclude that only two physically meaningful frequencies were present by examining the periodogram at the locations of the reported significant frequencies.

Grigahcène *et al.* (2010), in their Figure 2, showed that for hybrid γ Doradus/δ Scuti stars the γ Doradus and δ Scuti frequency ranges should not overlap. It is clear even from a cursory examination of the periodogram of HD 39641's data, see Figure 6, that there is nothing other than noise in the periodogram between approximately 3.5 and 6.5 cycles d^{-1}. This matches the theoretically predicted gap. A detailed examination of the 26 individual periodograms which compose the subsets of HD 39641's combined data set showed no stable peaks across time within this range. However, when a frequency analysis was run using SigSpec (Reegen 2011) using the frequency range of 0 to 50 cycles d^{-1}, 105 significant frequencies were reported within the 3.5–6.5 cycles d^{-1} range. As Poretti *et al.* (2009) note, this range of frequencies is lower than the fundamental radial mode, to which one must add that they are also above the theoretical range of g-mode frequencies. This leaves no immediately obvious physical explanation for their origins.

Figure 7 shows that the frequency gap did not become appreciably filled in until some time after the 500th frequency was extracted. The gradual "fanning out" of significant frequencies, initially from the range 6–12 cycles d^{-1} into both higher and lower frequencies, was similar to that seen in Figure 2 from the numerical experiments. Also, there was a "fanning out" of frequencies from the γ Doradus range into the gap. The implication is that at least some, if not all, of these statistically significant frequencies are simply artifacts of the data analysis process and do not represent physically meaningful pulsations.

The "fanning out" of the significant frequencies from the γ Doradus and 6–12 cycles d^{-1} ranges seen in Figure 7 stands in contrast to the 18–21 cycles d^{-1} or 23–31 cycles d^{-1} regions (see Figures 10 and 11) where there were a number of very significant frequencies, including the three most significant in HD 39641's combined data set. But the evidence for the gradual "fanning out" of significant frequencies like that seen in the two lower ranges is much less. This could be explained if one or more frequencies in the γ Doradus and 6–12 cycles d^{-1} ranges (see Figures 8 and 9) were unstable either in amplitude or frequency, or both, but the higher frequencies were stable. Put another way, if there were assumption violations in the two lower ranges, but not in the higher ranges, then this could account for both presence of the spurious frequencies and the manner in which they appear in Figure 7.

The Figures 8 through 11, just referred to, seek to answer this question. These are plots of the evolution of the periodogram over the 26 segments forming HD 39641's combined data set in four different frequency ranges. Although SigSpec reported 33 frequencies with significance ≥ 100 from the combined data set in the 6–12 cycles d^{-1} range, it was clear through examining the 3-d plot and using Matlab's (MathWorks 2019) rotation tool that all frequencies which could be identified visually were unstable in either frequency or amplitude or both. In Figure 9 the two strongest frequencies clearly change over time and occasionally split into two distinct frequencies, behaviors which will generate spurious frequencies during analysis.

A similar situation exists in the γ Doradus frequency region (Figure 8). It did have a very wide peak covering the 2.25 to 2.50 cycles d^{-1} region which clearly split into two distinct frequencies five times, each time rejoining. Because the number of usable data points varied by almost a factor of two in the 26 individual segments, we checked whether this was just an artifact of a larger number of data points giving better resolution. This does not appear to be an explanation for the observed behavior.

These observations of the lower frequency ranges were in contrast to the higher frequency ranges of Figures 10 and 11. In these Figures the frequencies which could be identified visually were noticeably more stable in both frequency and amplitude. In Figure 7 these two higher frequency regions did not appear to generate large numbers of frequencies which then "fan out" from their region of origin.

In light of the numerical experiments, it is a reasonable conclusion that there were sufficiently large violations of the underlying assumptions of the frequency analysis that the two lower frequency regions generated large numbers of spurious frequencies, hence filled in the clear gap between the top of the γ Doradus frequency range and the bottom of δ Scuti range.

Similar comments could be applied to the frequency range 12–16.5 cycles d^{-1}, where there are only eight frequency peaks in the periodogram of HD 39641's combined data set (Figure 6). A careful examination of the evolution of the 26 periodograms revealed only one stable structure at 16.27343 cycles d^{-1}. The remaining seven peaks were discernable, but all of them had periods in which their amplitudes dropped sufficiently low that they were not statistically significant. However, that frequency region was gradually filled in as frequencies "fanned out" from the top of the 6 to 12 cycles d^{-1} region, eventually giving 116 significant frequencies in this region. It was difficult to tell if there was any appreciable fanning out from the 18 cycles d^{-1} and higher region because, if there was, it did not begin to occur until somewhere about the 800th significant frequency. SigSpec reported 158 significant frequencies in this "gap," of which 120 had significances under 10 and hence were very likely to be spurious.

Table 6 presents a series of frequencies of the form kf, k=1,2,.... If one or more of these were a cascade of frequencies generated by an asymmetric light curve we would expect that there to be a monotonic decline in amplitude and significance. However, with 1,512 significant frequencies, the problem of either a chance alignment with another spurious frequency or one of the kf frequencies merging with a physically meaningful frequency the expected pattern would be disrupted. Of the frequencies listed only frequency 6, namely 0.946136 cycles d^{-1}, fully follows the expected pattern, concluding with a frequency (3.783453), which lies between the top of the γ Doradus and the bottom of the δ Scuti frequencies and has no known physical explanation.

A restricted range frequency analysis was performed. Restricting the range of frequencies to be modelled does two things: first, it prevents spurious frequencies from the parts of the frequency range outside the range being analyzed from contaminating the analysis, and secondly, it prevents spurious frequencies from the range being modelled from contaminating other frequency ranges. The basic problem of the violation the

Table 6. The highly significant frequencies (sig ≥ 100) which also had further significant frequencies at integer multiples of the base frequency.

N	*f*	*2f*	*3f*	*4f*	*5f*	*6f*	*7f*	*8f*
1	0.1303333	0.2616298	0.3888451	0.5227608	0.6477295	0.7813393	0.9148378	1.0436858
	(125.15)	(15.18)	(53.22)	(21.12)	(6.56)	(12.83)	(31.63)	(63.41)
2	0.2815179	0.5615032						
	(248.75)	(21.20)						
3	0.5978526	1.1946123	1.7930717	2.3907837	2.9911594	3.5832385		
	(338.17)	(28.75)	(39.98)	(41.58)	(7.26)	(5.09)		
4	0.7403535	1.4815043	2.2191678	2.9650469				
	(148.27)	(81.11)	(1850.16)	(7.93)				
5	0.8787684	1.7575015						
	(246.30)	(9.29)						
6	0.9461362	1.8932267	2.8406652	3.7834527				
	(215.54)	(7.02)	(6.87)	(4.70)				
7	2.9822269	5.9655588	8.9486721					
	(119.0364)	(7.67)	(12.96)					
8	3.0469668	6.0950020	9.1414396	12.1840934				
	(885.64)	(4.53)	(12.30)	(8.95)				
9	8.5106303	17.0202440						
	(120.28)	(4.61)						
10	8.9220928	17.8442241						
	(378.06)	(6.37)						
11	9.1540018	18.3077848						
	(385.72)	(18.27)						
12	9.6114526	19.2218557	28.8358362					
	(4671.35)	(6.13)	(7.51)					
13	19.2199717	38.4399206						
	(6589.14)	(20.07)						

Note: The figures in brackets below each frequency are the significance as reported by SigSpec (Reegen 2011).

underlying statistical assumptions remained, but the hope was that some of the grossest features created by these violations could be usefully reduced.

The results indicate this is a worthwhile strategy which is investigated further in a second paper. The complete elimination of all 59 spurious frequencies in the 4 to 6 cycle d^{-1} range was particularly notable. Within this range there is nothing but noise, so no assumption violations were present. The reduction from 315 to 48 in the number of statistically significant frequencies in the γ Doradus range and the fact that there appeared to be nothing but noise in the periodogram of the residuals (see Figure 13) was also encouraging.

6. Conclusions and future research

6.1. Conclusions

Much of the promise which the study of δ Scutis holds for asteroseismology had not been realized. The evidence presented here confirms existing literature which argues that part of the problem is that the methods of data analysis in common use include pre-whitening as an integral part of their analysis and hence yield large numbers of statistically significant but physically meaningless frequencies. We have confirmed the reported result of Balona (2014b) that even if all assumptions are met, pre-whitening still generates numerous spurious frequencies. Further, we have shown that if one or more assumptions are violated the problem can be dramatically worse.

Given that the spurious frequencies generated by pre-whitening are not necessarily of low significance, meaning the problem cannot be solved simply by raising the significance threshold, it is necessary to check all reported frequencies against the periodogram of a data set, to determine whether they are likely to be real or spurious. The use of restricted range frequency analysis has the potential to significantly reduce the number of frequencies which need to be checked against the periodogram and will be investigated in more detail in a second paper.

If a light curve spans sufficient time to meaningfully split the data into non-overlapping subperiods, closely examining the evolution of the periodogram over time can yield additional insights not obtainable from treating the data as a single data set.

The fact that there are four triplets with almost identical frequency differences in HD 39641's periodogram strongly suggests that they are the result of rotation. However, the fact that they were not always present in the periodograms of the subperiods needs further research. Other δ Scutis with reported triplets should provide a good source of data to work with.

6.2. Future research

It is particularly troubling that mode splitting by stellar rotation can be convincingly mimicked by amplitude modulation. Further research is needed to find a method to distinguish between these two cases.

A curious finding was that in the lower frequency regions, namely the γ Doradus range and the 6–12 cycles d^{-1} region of the δ Scuti range, most of the frequencies were unstable, whereas above about 16 cycles d^{-1} they were largely stable. Further δ Scutis need to be studied to determine if this is a common feature among this class of variables.

The use of restricted frequency range analysis still has the same underlying assumptions as doing an unrestricted frequency

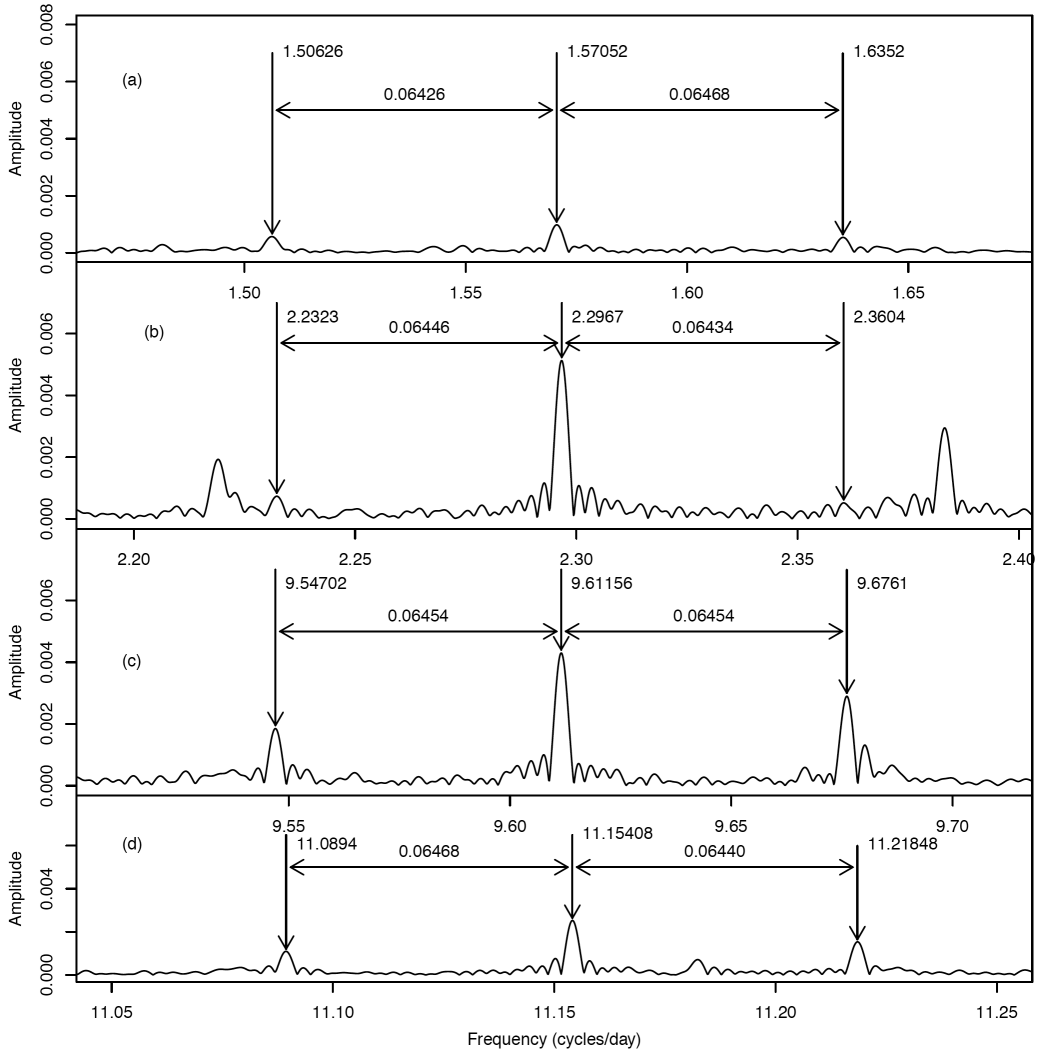

Figure 12. The Fourier spectral analysis of HD 39641's combined data revealed four sets of triplets. Panel (a) shows the rst set of triplets in the γ Doradus range. Panel (b) shows the second set of triplets in the γ Doradus range. Panel (c) which the first set of triplets in the lower part of the δ Scuti range while panel (d) also shows another set of triplets in the lower part of the δ Scuti range.

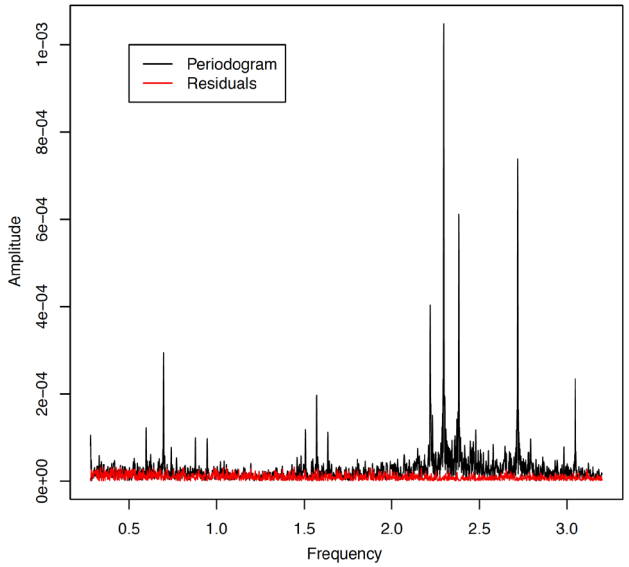

Figure 13. A detail of The Fourier spectral analysis of HD 39641's combined data set before (black) and the residuals after (red) a frequency analysis by SigSpec (Reegen 2011) with the frequency range restricted to 0.28 to 3.2 cycles d^{-1}.

analysis where the entire frequency analysis is done at one time. However, it has the potential to curb some of the worst consequences of assumption violations, such as the spreading of spurious frequencies across the entire frequency range being analyzed. This is addressed in greater detail in Paper II.

7. Acknowledgements

The author would like to thank the editor and an anonymous referee for valuable comments which helped to improve the paper.

The author would like to thank NASA for making the data from its Transiting Exoplanet Survey Satellite freely available and the TESS Asteroseismic Science Operations Center for operating their particular database (TASOC) of TESS photometry data.

The author would like to thank the American Association of Variable Star Observers for maintaining and making freely available the Variable Star Index (VSX).

This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

References

Aerts, C., Christensen-Dalsgaard, J., and Kurtz, D. 2010, *Asteroseismology*, Springer, Dordrecht.

Balona, L. A. 2014a, *Mon. Not. Roy. Astron. Soc.*, **437**, 1476.

Balona, L. A. 2014b, *Mon. Not. Roy. Astron. Soc.*, **439**, 3453.

Barceló Forteza, S., Michel, E., Roca Cortés, T., and García, R. A. 2015, *Astron. Astrophys.*, **579**, A133.

Benn, D. 2012, *J. Amer. Assoc. Var. Star Obs.*, **40**, 852.

Bloomfield, P. 2000, *Fourier Analysis of Time Series: An Introduction*, 2nd ed., Wiley, New York.

Bowman, D. M., Kurtz, D. W., Breger, M., Murphy, S. J., and Holdsworth, D. L. 2016, *Mon. Not. Roy. Astron. Soc.*, **460**, 1970.

Catelan, M., and Smith, H. A. 2015, *Pulsating Stars*, Wiley-VCH, Weinheim, Germany.

Deeming, T. 1975, *Astrophys. Space Sci.*, **36**, 137.

Gaia Collaboration, *et al.* 2018, *Astron. Astrophys.*, **616A**, 1.

Goupil, M.-J., Dupret, M. A., Samadi, R., Boehm, T., Alecian, E., Suarez, J. C., Lebtreton, Y., and Catala, C. 2005, *J. Astrophys. Astron.*, **26**, 249.

Grigahcène, A., *et al.* 2010, *Astrophys. J., Lett.*, **713**, L192.

Kurtz, D. W., Shibahashi, H., Murphy, S. J., Bedding, T. R., and Bowman, D. M. 2015, *Mon. Not. Roy. Astron. Soc.*, **450**, 3015.

Lomb, N. R. 1976, *Astrophys. Space Sci.*, **39**, 447.

Mary, D. L. 2005, *J. Astrophys. Astron.*, **26**, 283.

MathWorks. 2019, MATLAB programming and numeric computing platform, ver. 9.6.00, The MathWorks Inc., Natick, MA (https://www.mathworks.com/products/matlab.html).

Pascual-Granado, J., Suárez, J. C., Garrido, R., Moya, A., García Hernández, A., Rodón, J. R., and Lares-Martiz, M. 2018, *Astron. Astrophys.*, **614**, A40.

Poretti, E., *et al.* 2009, *Astron. Astrophys.*, **506**, 85.

R Core Team. 2019, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (https://www.R-project.org).

Reegen, P. 2011, *Commun. Asteroseismology*, **163**, 3.

Ricker, G. R. *et al.* 2014, *Proc. SPIE*, **9143**, id. 914320.

Scargle, J. D. 1982, *Astrophys. J.*, **263**, 835.

Uytterhoeven, K., *et al.* 2011, *Astron. Astrophys.*, **534**, A125.

VanderPlas, J. T. 2018, *Astrophys. J., Suppl. Ser.*, **236**, 16.

Wenger, M., *et al.*, 2000, *Astron. Astrophys., Suppl. Ser.*, **143**, 9.

Zima, W. 2008, *Commun. Asteroseismology*, **155**, 17.