A&A 614, A40 (2018)
https://doi.org/10.1051/0004-6361/201732431

Astronomy & Astrophysics

Impact of gaps in the asteroseismic characterization of pulsating stars

I. The efficiency of pre-whitening[⋆]

J. Pascual-Granado[1], J. C. Suárez[2,1,⋆⋆], R. Garrido[1], A. Moya[2,3,4], A. García Hernández[2],
J. R. Rodón[1], and M. Lares-Martiz[1]

[1] Instituto de Astrofísica de Andalucía, Glorieta de la Astronomía s/n, 18008 Granada, Spain
e-mail: javier@iaa.es
[2] Departamento de Física Teórica y del Cosmos, Universidad de Granada, Campus de Fuentenueva, 18071 Granada, Spain
[3] Departamento de Astrofísica, Centro de Astrobiología (CAB/INTA-CSIC), 28850 Torrejón de Ardoz, Madrid, Spain
[4] School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK

Received 7 December 2017 / Accepted 6 February 2018

ABSTRACT

Context. It is known that the observed distribution of frequencies in CoRoT and *Kepler* δ Scuti stars has no parallelism with any theoretical model. Pre-whitening is a widespread technique in the analysis of time series with gaps from pulsating stars located in the classical instability strip, such as δ Scuti stars. However, some studies have pointed out that this technique might introduce biases in the results of the frequency analysis.

Aims. This work aims at studying the biases that can result from pre-whitening in asteroseismology. The results will depend on the intrinsic range and distribution of frequencies of the stars. The periodic nature of the gaps in CoRoT observations, only in the range of the pulsational frequency content of the δ Scuti stars, is shown to be crucial to determining their oscillation frequencies, the first step in performing asteroseismology of these objects. Hence, here we focus on the impact of pre-whitening on the asteroseismic characterization of δ Scuti stars.

Methods. We select a sample of 15 δ Scuti stars observed by the CoRoT satellite, for which ultra-high-quality photometric data have been obtained by its seismic channel. In order to study the impact on the asteroseismic characterization of δ Scuti stars we perform the pre-whitening procedure on three datasets: gapped data, linearly interpolated data, and data with gaps interpolated using Autoregressive and Moving Average models (ARMA).

Results. The different results obtained show that at least in some cases pre-whitening is not an efficient procedure for the deconvolution of the spectral window. Therefore, in order to reduce the effect of the spectral window to a minimum, in addition to performing a pre-whitening of the data, it is necessary to interpolate with an algorithm that is aimed to preserve the original frequency content.

Key words. asteroseismology – methods: data analysis – stars: oscillations – stars: variables: δ Scuti

1. Introduction

Even when ultra-precise space satellites like CoRoT (Baglin et al. 2006) or *Kepler* (Gilliland et al. 2010) are used to observe the variation in the brightness of stars, these time series can contain gaps due to operational procedures or environmental effects. Gaps introduce correlation between the frequency bins of a periodogram (Gabriel 1994) and may have a significant impact on seismic studies which are based on frequency analyses. An unbiased estimation of the power spectrum is necessary for seismic studies in order to obtain realistic models of the stars.

A number of studies exist in the literature focussing on gaps in time series from the Sun, solar-like, or red giant stars (e.g., Brown & Christensen-Dalsgaard 1990; Fossat et al. 1999; Stahn & Gizon 2008; García et al. 2014) or, more generally, on unevenly sampled data (e.g., Deeming 1975; Scargle 1982) and how to diminish its contribution to the spectral window (Foster 1995).

A new approach to fill the gaps was proposed in Pascual-Granado et al. (2015, hereafter PG15), using a method based on Autoregressive and Moving Average models models (ARMA), which aims to preserve the original frequency content of the signal. This technique has been used successfully in several studies (e.g., García Hernández et al. 2015, 2017). However, in asteroseismic studies of δ Scuti stars the usual procedure to deal with the effects of the spectral window is to perform a pre-whitening (e.g., García Hernández et al. 2009, 2013; Poretti et al. 2009; Barceló Forteza et al. 2015), which is similar to the CLEAN method introduced by Hogbom (1974). However, this technique is not without its challenges (Balona 2014; Mary 2005).

Here we want to check the consistency of pre-whitening as a technique for spectral window deconvolution of light curves of δ Scuti stars. The test is based on a comparison of the frequency content of gapped data and data interpolated with an algorithm that is aimed to preserve the information. In order to perform the interpolation we make use of the gap-filling algorithm

[⋆] Tables 5–49 are only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/vol/page
[⋆⋆] Associate researcher at institute (1).

MIARMA. The subsequent consistency check consists in verifying the compatibility between both distributions.

For this study we have chosen a sample of light curves of δ Scuti stars from CoRoT Seismofield (with a sampling of 32 s). CoRoT has a duty cycle of approximately 90% with a 10% loss of data due to the passing through the South Atlantic Anomaly (SAA). Therefore, roughly a 10% loss in amplitude is usually expected for periods shorter than the duration of the passing through the SAA, which is about 9 min ($\nu > 160$ d^{-1}), and a minor contribution is expected for periods longer than 18 min ($\nu < 80$ d^{-1}) (see Appourchaux et al. 2008). Periods of 9–18 min contribute the most to the spectral window but for δ Scuti stars periods no shorter than 18 min are expected. The correction originally implemented in CoRoT pipeline was a simple linear interpolation which was expected to be enough to avoid the contribution of the gaps to the spectral window. In PG15, however, it was shown that linear interpolation introduces spurious frequencies that might affect seismic studies of these stars.

We perform the frequency analysis of light curves with gaps that are filled with two different techniques: the MIARMA algorithm, which is aimed to preserve the frequency content, and linear interpolation, which is not. Linear interpolation is not so often used in frequency analyses of δ Scuti stars as in solar-like studies (e.g., García et al. 2009; Zwintz et al. 2011; Benomar et al. 2009) but by introducing an interpolation method in the consistency check that is aimed to maximize computational efficiency and not to preserve the original frequency content, we can study the impact of the interpolation on the distribution of frequencies. Our results show that the pre-whitening process is not efficient enough in the deconvolution of the spectral window, and that a linear interpolation can, in some cases, give a similar distribution of frequencies to gapped data.

The paper has the following structure: we first present in Sect. 2 the sample of δ Scuti stars analyzed and the characteristics of the observations. In Sect. 3 we describe the methodology followed in this study, we outline the gap-filling methods, the frequency analysis procedure, and a cleaning procedure for the frequency combinations and nonlinear interactions performed to avoid biases in the analysis. In Sect. 4 we show the results of the gap-filling, the frequency spectra and the histogram of detected frequencies through the methodology described in the previous section. Section 5 is devoted to a discussion of the results and, finally, in Sect. 6 we present conclusions and plans for a future publication concerning the impact on periodicity studies performed over the periodogram in order to find patterns that help in the modal identification (see García Hernández et al. 2009 for a detailed explanation).

2. CoRoT data

CoRoT had two channels: one for the study of pulsating stars (Seismofield) and another for exoplanet detection (Exofield). The observations in the Seismofield are of greater precision since the targets are brighter (i.e., lower noise levels) and the cadence is higher (32 s). We have chosen Seismofield light curves since they have a better signal-to-noise ratio (S/N) caused by the reduced readout noise due to the large number of pixels involved, and this is interesting in order to compare the distribution of frequencies minimizing the contribution of noise to the results.

The orbital frequency of the satellite is on average 13.972 d^{-1} and the passage through the SAA occurs twice a day (Auvergne et al. 2009). This has a direct impact on the observed stellar power spectra in the form of spurious peaks with their multiples, as well as in the form of combinations with frequencies of the intrinsic stellar oscillations produced by the convolution with the spectral window (Deeming 1975). In the case of δ Scuti stars, pulsations are excited in the range between 10 and 80 d^{-1} (see Moya et al. 2017, for evidence of this). As a consequence the light curves of δ Scuti stars are affected the most by these spurious frequencies.

The main criterion for the selection of stars was the type of variable; therefore, the sample is composed of 15 stars observed in the Seismofield during the Initial Run (IR), Short Run (SR) and the Long Run (LR) both in galactic center and anticenter directions. Fourteen δ Scuti stars and one γ Dor present excited frequencies in the δ Scuti stars regime (Chapellier et al. 2011). Most of these stars (eleven) are A-type stars, and four are F-type stars (see Table 1 for the specific characteristics of the runs and stars).

As mentioned in the introduction, the most frequent gap duration in the light curves is 9 min but some gaps can last much longer. One example of this is HD 170699, which is observed in the run LRc0506 (i.e., 5th/6th Long Run with galaxy center orientation), where a gap of about 2 h is found in the light curve. This is the only star observed in LRc0506 in the sample of stars we have selected but in HD 174589 there is a gap of 5 h in duration. Both HD 174589 and HD 174532 have been observed in SRc02 (i.e., 2nd Short Run with galactic center orientation) so they have the same spectral window. In these stars, the contribution of gaps to the spectral window is much greater; see Table 2 for a complete statistical characterization of the gap distribution for the selected CoRoT dataset.

3. Methodology

Since we are interested in studying the impact of gaps in the usual harmonic analysis performed in asteroseismology, we follow what can be considered a standard workflow. The only deviation from a standard procedure consists in filling the gaps in the light curves with two different methods. We do this with the objective of comparing the results with gapped data. In summary the workflow is:

1. Correction for the instrumental drift (Auvergne et al. 2009) by performing a polynomial fit to the light curves.
2. Gap-filling with ARMA and linear interpolation.
3. Pre-whitening analysis of the interpolated and gapped data.
4. Detection and removal of frequency combinations due to nonlinear interactions between independent modes.

3.1. Gap-filling

Linear or polynomial interpolation is still widely used for filling the gaps present in CoRoT light curves (see e.g., Appourchaux et al. 2008; Gutiérrez-Soto et al. 2009; Kallinger & Matthews 2010) but also for datasets from other missions like SOHO (Seleznyov et al. 2011). Re-sampling of irregular data with such analytic methods can be justified in a few cases because they are more robust (de Waele & Broersen 2000) than more sophisticated interpolation methods that may have divergence issues, but the variance is erroneously estimated and the reconstruction of the original signal can be poor. Indeed, linear interpolation does not preserve the information (see, e.g., Figs. B.1, B.2, B.9–B.13). In Sect. 4, we show that in some cases linear interpolation alters the frequency content in a similar way as no interpolation.

Here we use the MIARMA algorithm (see PG15) which is aimed to preserve the original frequency content of the signal, thereby minimizing (and even avoiding) the contribution to the spectral window by the gaps that causes spurious variations

Table 1. Selected sample of δ Scuti stars observed by CoRoT.

Run	Star ID	CoRoT ID	SpT	m_v	log $T_{\rm eff}$ (K)	M_V	$v \sin i$ (km s^{-1})	Obs. time (d)
IRa01	HD 50844[1]	123	A2	9.1	3.88	1.31	64	57.713
SRc01	HD 174936[2]	7613	A2	8.58	3.9	1.88	170	27.194
SRc01	HD 174966[3]	7528	A3	7.72	3.88	1.95	125	27.197
LRc01	HD 181555[⋆,4]	8669	A5 V	7.52	3.85	−0.72	200	156.645
LRa01	HD 49434[5]	100	F1 V	5.74	3.86	2.74	87	136.890
LRc02	HD 172189[6,7]	8170	A2	8.73	3.89	1.04	78	149.013
SRc02	HD 174532[⋆,8]	7655	A2	6.90	3.86	1.38	32	26.239
SRc02	HD 174589[⋆]	7663	F2 III	6.09	3.85	1.45	100	26.168
LRa02	HD 51722[⋆]	1022	A5	7.53	3.86	1.13	127	117.375
LRa02	HD 51359[⋆]	1320	A5	8.50	3.9	0.89	–	117.41
LRa02	HD 50870[9]	546	F0	8.88	3.88	1.67	17	114.413
LRc0506	HD 170699[9]	8301	A2	6.95	3.88	1.49	270	89.282
IRLRa04	GSC 00144–03031[1,10]	21960	A8	10.1	3.89	–	–	79.133
IRLRa05	HD 41641[⋆]	5685	A5	7.9	3.882	1.92	28	94.432
SRa05	HD 48784[⋆]	3619	F0	6.66	3.84	1.87	108	25.305

Notes. From *left* to *right*: observing run, HD number, CoRoT number, spectral type, visual magnitude, effective temperature, absolute magnitude, rotational velocity and observation interval.

References. [(1)]Poretti et al. (2009), [(2)]García Hernández et al. (2009), [(3)]García Hernández et al. (2013), [(4)]Reese et al. (2013), [(5)]Brunsden et al. (2015), [(6)]Martín-Ruiz et al. (2005), [(7)]Creevey et al. (2009), [(8)]Fox Machado et al. (2010), [(9)]Mantegazza et al. (2012), Poretti et al. (2005), [(10)]Rainer et al. (2016), [(⋆)]CoRoT archive (`http://idoc-corot.ias.u-psud.fr/sitools/client-user/COROT_N2_PUBLIC_DATA/project-index.html`) accessed through the Seismic Plus portal (`http://voparis-spaceinn.obspm.fr/seismic-plus/`).

Table 2. Statistical parameters of the gap distribution for the light curves observed by CoRoT used in this paper.

Star ID	Duty cycle (%)	$\tau_g^{\rm max}$ (h)	$\bar{\tau_g}$ (h)	$\sigma(\tau_g)$ (h)	Mode (h)
HD 50844	88.69	1.751	0.173	0.126	0.266
HD 174936	88.83	0.409	0.162	0.116	0.267
HD 174966	89.55	0.409	0.175	0.099	0.267
HD 181555	88.95	2.516	0.175	0.125	0.267
HD 49434	88.44	13.858	0.182	0.359	0.258
HD 172189	89.07	7.031	0.079	0.162	0.009
HD 174532	87.18	5.280	0.072	0.265	0.009
HD 174589	87.22	5.289	0.072	0.266	0.009
HD 51722	89.30	3.476	0.106	0.150	0.009
HD 51359	88.90	3.476	0.094	0.138	0.009
HD 50870	89.78	3.476	0.102	0.142	0.009
HD 170699	88.46	4.044	0.067	0.139	0.009
GSC 00144–03031	88.46	7.636	0.066	0.227	0.009
HD 41641	78.78	144.249	0.119	2.552	0.009
HD 48784	88.55	1.582	0.161	0.137	0.267

Notes. In columns 3–5, τ_g refers to the duration of the gaps in hours, where column 3 is the maximum, column 4 is the mean duration, column 5 is the standard deviation of the gap distribution, and the last column is the most frequent gap duration.

in amplitudes and phases. The order of the ARMA model is selected through the Akaike Information Criterion (AIC, Akaike 1974) and the coefficients of the model are obtained through an optimization algorithm. In contrast to analytic methods, MIARMA guarantees that no bias due to ad-hoc hypothesis about the signal is introduced when filling the gaps (see full details in PG15).

Some previous studies have assumed that all the spurious frequencies produced by the convolution of the signal with the spectral window (Deeming 1975) are mitigated during the pre-whitening cascade. Therefore, gap-filling was unnecessary. We have checked this assumption by comparing the pre-whitening of data with gaps that are filled with two opposed gap-filling techniques: one is aimed at preserving information (MIARMA) and the other one is not (linear interpolation).

3.2. Pre-whitening

The classical pre-whitening technique (Ponman 1981) was modified by Reegen (2007) in order to analyze CoRoT data using a rigorous statistical treatment of how to determine the significance of a peak in a periodogram.

Light curves pre-processed as described above were subsequently analyzed using SIGSPEC. This algorithm has been extensively used by the asteroseismic community (see e.g., Paparó et al. 2016; Weiss et al. 2016; Molnár et al. 2017; Zwintz et al. 2017). It is based on an iterative sequence of frequency detection, least squares fitting, and a pre-whitening cascade. The iterative sequence stops when a significance threshold is reached (by default sig = 5.0, that is $\approx S/N = 4$).

The range of frequencies explored by the algorithm is chosen taking into account several considerations. First, the lowest frequencies saturate the power spectrum and, though a deeper study of the selected stars would require including mixed and g modes, to simplify our study we focus here only on the fundamental radial mode and frequencies above (p modes). On the other hand, δ Scuti stars have their main frequency content below 80 d^{-1}. Therefore, we have used a frequency interval of 2–100 d^{-1} for the frequency detection.

With these parameters SIGSPEC calculations lasted for several months in some cases (e.g., HD 181555) to release a list of significant frequencies. Such lists were then cleaned for spurious frequencies through the process explained below.

3.3. Nonlinear interactions

In order to detect and remove nonlinear interactions and spurious frequencies from the oscillation spectra, we followed the heuristic method outlined below.

We first estimate independent frequencies. We calculated the set of independent frequencies, f_i, in the range $2 \geq f_i \geq 100$ using the SIGSPEC addon COMBINE (Reegen 2007). The set of independent frequencies is arbitrarily truncated to 12 frequencies. When less than 12 frequencies were found we appended frequencies to the set as sorted by their amplitude. Frequencies below 2 d^{-1} were excluded because most of the signal there comes from instrumental effects giving rise to trends rather than pure harmonic components.

Then, we search for harmonics and combinations up to third order within a bin with the size of the Rayleigh dispersion ($f_R = 1/T_{obs}$).

We differentiate spurious frequencies from combinations that appear due to the nonlinear response of the physical system. Thus, spurious frequencies are identified with the orbital frequency, aliases of 1 d^{-1}, and the harmonics and combinations of these with independent frequencies.

In the second part our algorithm, perform calculations for each independent frequency found to obtain the following combinations:

1. Harmonics up to the fifth order.
2. First-order combinations including absolute value differences.
3. Second- and third-order combinations, that is:

$$2f_1 \pm f_2, \ldots, 2f_1 \pm 2f_2, \ldots,$$
$$3f_1 \pm f_2, \ldots, 3f_1 \pm 2f_2, \ldots, 3f_1 \pm 3f_2, \ldots$$

4. Combinations with the frequency of the satellite's orbit ($f_s = 13.972\ \mathrm{d}^{-1}$) and their harmonics up to fourth order and $f_s/2$, $f_s/3$, $f_s/4$.
5. Combinations of the harmonics of f_s up to the third order with the first four frequencies (only for the first order).
6. 1 d^{-1} aliases around f_s and their harmonics up to the fourth order, and the harmonics of the aliases up to the fifth order, that is,

$$\begin{aligned} f_s &\pm 1, \pm 2, \pm 3, \pm 4, \pm 5, \\ 2f_s &\pm 1, \ldots, \pm 5, \\ 3f_s &\pm 1, \ldots, \pm 5, \\ 4f_s &\pm 1, \ldots, \pm 5. \end{aligned}$$

7. Aliases up to the fifth order of the combinations calculated in step 5.
8. The two highest amplitude frequencies (f_b) in the range 0–2 d^{-1} are used to calculate interactions with the satellite orbit.
9. Combinations up to the fifth order between f_s and f_b (first order).
10. 1 d^{-1} aliases around combinations calculated in step 8, up to the fifth order.
11. Harmonics of the main frequency up to the fourteenth order.

Since the passing of the satellite through the SAA is twice each sidereal day we include 1 d^{-1} aliases at step 5. We highlight that, though we excluded frequencies below 2 d^{-1} from the calculation of independent frequencies, since we are considering only *p* modes, we included this range in step 8 as there might be interactions between low frequencies and *p* modes.

Due to the density of significant frequencies found in δ Scuti stars, we follow a conservative approach and identify a reduced set of harmonics and combinations with the orbital frequency in order to guarantee the unambiguous and robust identification of spurious frequencies. It is possible that extending the order of harmonics and combinations would increase these numbers, but this is beyond the scope of this paper since the aim is to test the impact of the gap-filling on the classical procedure for pre-whitening and cleaning used in asteroseismology.

In summary, the algorithm calculates for each star a set of 703 frequencies of potential nonlinear interactions (including spurious frequencies) that are searched for in the list of significant frequencies obtained by SIGSPEC. When coincidences of the order of the Rayleigh dispersion are found, those frequencies are removed from the list. Cleaned frequencies for each star along their corresponding parameters obtained following the procedure explained above are listed in Tables 5–19 for ARMA interpolated data, Tables 20–34 for linearly interpolated data, and Tables 35–49 for gapped data.

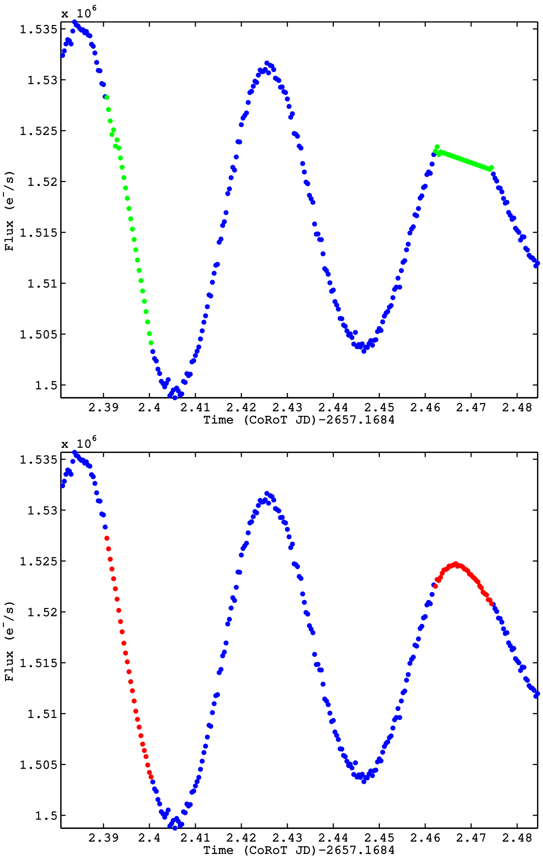

Fig. 1. Illustration of a gap in HD 174966 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

4. Results

In this section we present the results of gap-filling, power spectrum, and statistical characterization for the selected set of stars. Figures corresponding to HD 174966 are shown here as an illustrative case, the rest of the figures for the other 14 stars are shown in the Appendices.

In order to work optimally, radix-2 FFT algorithms require that the number of datapoints is a power of 2, so we have truncated *N* to the power of 2 closer to the original number of datapoints. Additionally, light curves are statistically normalized, and therefore the amplitudes are dimensionless.

Periodograms are plotted in log-scale in amplitudes and d^{-1} in frequencies in the range 2–100 d^{-1}. Power spectra of gapped and ARMA-interpolated data are compared in Appendix C together with power spectra of linearly interpolated data.

In Fig. 1 we show a typical gap in the light curve of HD 174966 that has been interpolated linearly and with MIARMA.

In this case it is clear that a linear interpolation does not preserve the frequency content of the signal and as a consequence, spurious peaks will appear in the power spectrum

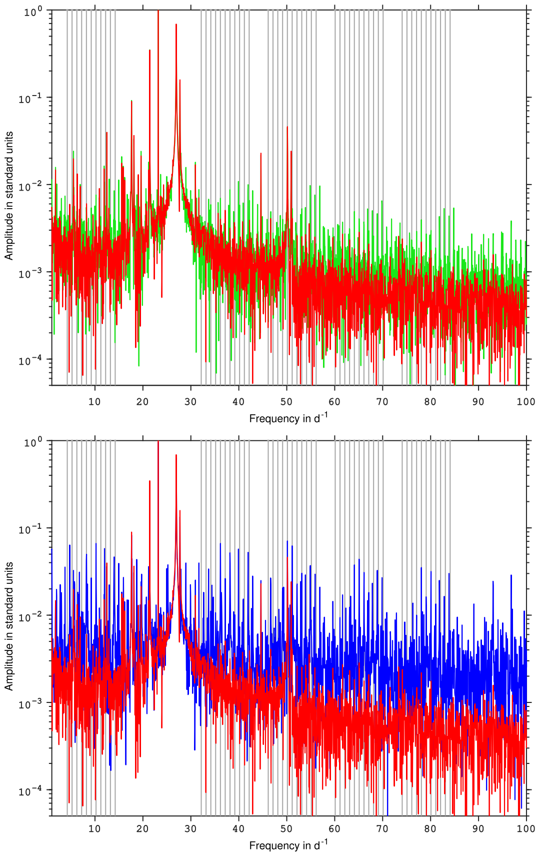

Fig. 2. Power spectra of the light curves from HD 174966: *lower panel* shows gapped data in blue and ARMA-interpolated data in red, *upper panel* shows linearly interpolated data in green and ARMA-interpolated data in red. Vertical gray lines show the main peaks of the spectral window.

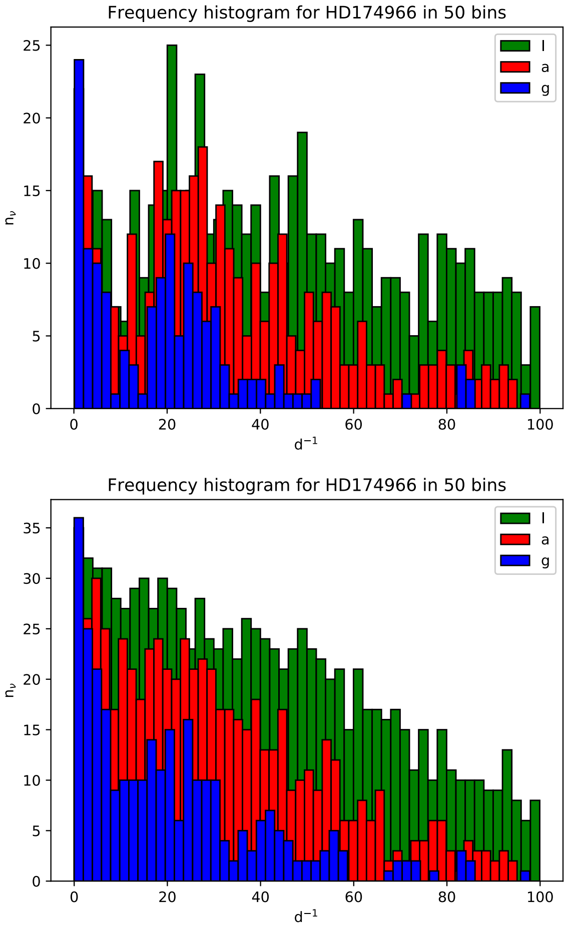

Fig. 3. Histograms of frequencies detected in the light curves of HD 174966. *Upper panel*: cleaned frequencies, whereas *lower panel*: frequencies before the cleaning is applied. Blue bars correspond to gapped data (g), red bars to ARMA-interpolated data (a), and green bars to linearly interpolated data (l).

that might introduce a bias in the parameter estimation of the harmonic components and its related physics. Whether a pre-whitening process permits us to overcome these difficulties in the frequency analysis or not is something that we discuss in the following section.

Some stars shown in the Appendices (see, e.g., Figs. B.3, B.4, B.14) present no apparent differences in the light curves interpolated by MIARMA and linearly, and between their corresponding power spectra (see, Figs. C.3, C.4, C.14). This is in contrast to other stars (see e.g., Figs. B.7, B.10, B.12) where ARMA models appear to preserve the signal and linear interpolation not (see also their corresponding power spectra in Figs. C.7, C.10, C.12). If we compare the first set of cases with the last one, we can see that linear interpolation appears to work better when the signal shows longer scale variations inside the gaps (i.e., derivatives do not change significantly). This can give us an idea of when interpolating linearly presents a risk that needs to be taken into account.

In Fig. 2 we show these plots for HD 174966. Both power spectra of gapped data and linearly interpolated data, calculated through a Lomb-Scargle periodogram (Scargle 1982), clearly show the effect of the satellite orbital modulation and SAA, that is, patterns of frequencies around 13.97 d^{-1} and their multiples (vertical gray lines). These patterns do not appear in the power spectrum of ARMA interpolated data.

In order to perform a statistical characterization we apply the cleaning procedure described in the previous section.

In Fig. 3 we show the histogram of the number of cleaned frequencies for each light curve (ARMA, linear, gapped) of HD 174966 resulting after applying the cleaning procedure (upper panel) and the same for the frequencies obtained before the cleaning (lower panel). The cleaning procedure changes the distribution and the differences between histogram bins considerably. Therefore, the cleaning is necessary to avoid biases in the analysis. We henceforth refer only to cleaned frequencies.

We note that the number of cleaned frequencies is lower in gapped data than in ARMA-interpolated data. One possible explanation is that, due to the lower duty-cycle of gapped data, some of the harmonic components detected in ARMA-interpolated data have less significance in gapped data.

On the other hand, the number of cleaned frequencies is lower in ARMA-interpolated data than in linearly interpolated data. Although the cleaning procedure is aimed to remove spurious and combination frequencies, due to the conservative limits used for the harmonic number and possible interactions, it is very probable that some spurious frequencies are not removed, as mentioned above. Subsequently, the higher number of frequencies detected in linearly interpolated data might be caused by a higher number of spurious frequencies appearing in the power spectrum. This is compatible with the results shown in Table 3 that are discussed thoroughly in the following section for HD 174966 and shown for the other stars in the Appendices.

5. Discussion

Whether the frequency content is preserved or not when filling gaps with interpolated data is strictly dependent on the algorithm used and the bandwidth of frequencies present in the signal.

Gaps in CoRoT data cause a reduced amplitude of the peaks when non-interpolated data are analyzed but, as can be seen in the figures included in the Appendices, linear interpolation also changes the original frequency content of the signal introducing spurious peaks that leak the power and change the original phases that might be crucial for modal identification due to the correlation between frequency bins (see Gabriel 1994, and Appendix A).

We note that not all data collected during the passing through the SAA are corrected. The algorithm for data correction implemented in the pipeline of CoRoT N2 level (version 0.2/1.3) is conservative and only interpolates those data points with a certain deviation from the mean. This means that in some light curves, incorrect data points appear that are not interpolated (see e.g., Figs. B.5 and B.6) which causes an additional spurious contribution to the power spectrum. However, this is a marginal effect in our sample.

The impact of the gaps depends on the spectral window associated to the gap distribution and sizes. In Table 2 we present the statistical parameters characterizing this distribution. The most frequent gap appearing in the light curves of nine stars is 9×10^{-3} h (32 s), which is just one data point, meaning that they are dominated by single outliers. The rest have their most frequent gap at 0.267 h (16.02 min), but their median is around 0.168 (∼10 min) which is more coherent with the duration of the pass through the SAA. However, some stars have much longer gaps (e.g., a gap of ∼6 d in HD 41641 or one of 13.86 h in the light curve of HD 49434). The contribution of these gaps to the spectral window is much greater.

The contribution of the interpolation to the power spectra shown in Figs. 2, C.2, C.6–C.8, C.10 is very similar. This is consistent with the values of the statistical parameters shown in Table 2 since these three groups of stars were each observed during the same run of observations: HD 174936, and HD 174966 (SRc01), HD 174532 and HD 174589 (SRc02), HD 51722 & HD 50870 (LRa02). Only HD 51359 (see Fig. C.9), also observed during LRa02, shows a different behavior since it appears to have only low-frequency variability.

The numbers of frequencies classified as spurious (NS), combinations (NC) and independent (NI) for each star and each light curve (gapped, linearly interpolated and ARMA interpolated) are collected in Table 3, where $NI + NS + NC$ amounts to the total number of frequencies detected using SIGSPEC.

We note that the numbers of spurious frequencies detected are a small fraction of the total number of frequencies in most cases. This might be due to the reduced set of harmonics and combinations used.

After cleaning spurious frequencies and linear combinations (produced by nonlinear interactions) of independent frequencies, the resulting number of frequency components (NI) is still of the order of approximately 1000 in many cases. We note that there may still be spurious and combination frequencies hidden in the final list that cannot be resolved unambiguously.

No frequency component has been detected beyond the photometric precision of CoRoT Seismofield cameras which is 0.6–4 ppm (Auvergne et al. 2009). Therefore the high number of frequencies detected cannot be explained by a lack of precision.

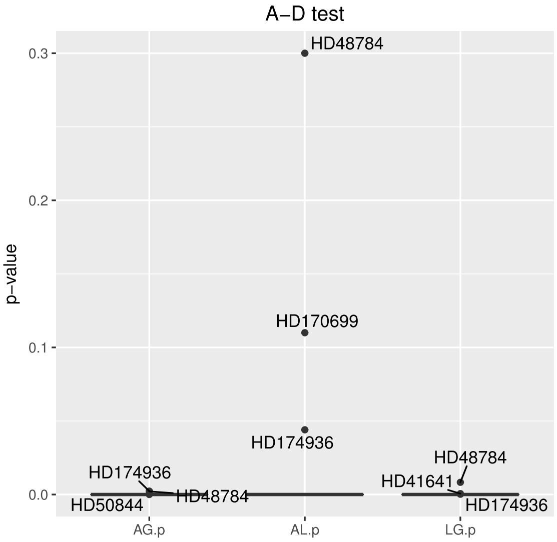

Fig. 4. Box plot of the p-values calculated in Table 3. We note that most of the stars present a p-value of $\sim 10^{-5}$. The outliers of the sample are labeled with their HD number.

Further analysis is required to clarify the origin of the frequency components.

In order to have a quantitative validation of the statistical results we have performed a k-samples Anderson–Darling test (Scholz & Stephens 1987) to the histograms of each case comparing them by pairs. This test evaluates the hypothesis that two time series with n independent samples arise from a common unspecified distribution. The choice of this test is motivated by the fact that it does not require any specific probability distribution.

In the last column of Table 3 we show the corresponding p-values (false-alarm probability) for each pair. The null hypothesis is that compared frequency values originate from a common distribution, meaning that when a small p-value is found (here below 0.01) it can be interpreted that the frequencies are not compatible with the same distribution. Only the pair AL for HD 170699, HD 48784, and HD 174936 show non-negligible p-values avoiding the rejection of the null hypothesis. This reinforces the conclusion that gap-filling methods have an impact on the estimation of the frequency content of our sample (i.e., the pre-whitening is not unbiased).

We also performed a box-plot of the p-values to evaluate differences in the distributions of frequencies (Fig. 4). The Anderson–Darling test gives p-values of the order of 10^{-5} in most cases. On the other hand, three stars (HD 48784, HD 170699, HD 174936 denoted by points) have much higher p-values in the comparison between ARMA and linearly interpolated data. This means that ARMA and linearly interpolated data present similar frequency distributions in these cases. This could be, in fact, an effect produced by the sample selection related to the analyticity of the intervals that are interpolated (see, e.g., Figs. B.13 and B.2). Also, other factors that could affect our sample, such as the rotational velocity and the visibility of the modes, cannot be assessed here. In any case, these latter three stars can be considered outliers in our sample that are not affecting the statistical characterization.

Finally, in order to confirm that the bias we have found originates in the pre-whitening process we have performed an additional AD test applied to the histogram densities of the (A, L, G) power spectrum before any pre-whitening process is applied (see Table 4).

Table 3. Number of independent frequencies detected after cleaning for nonlinear interactions and spurious frequencies (NI); number of frequency combinations found and removed (NC), and number of spurious frequencies found and removed (NS).

		NI			NC			NS			A–D test	
Star ID	A	L	G	A	L	G	A	L	G	AG	AL	LG
HD 50844	1189	1124	1001	237	214	168	320	319	320	1.9×10^{-5}	7.4×10^{-6}	4.6×10^{-10}
HD 174936	520	550	348	82	81	68	268	284	164	2.3×10^{-3}	4.4×10^{-2}	2.1×10^{-4}
HD 174966	364	583	152	42	21	8	239	412	136	3.4×10^{-15}	7.1×10^{-12}	1.4×10^{-22}
HD 181555	1782	1786	2054	171	211	196	151	157	250	3.5×10^{-30}	5.5×10^{-11}	3.1×10^{-47}
HD 49434	1232	1284	2243	287	161	308	93	96	199	2.9×10^{-48}	4.7×10^{-10}	7.8×10^{-41}
HD 172189	1360	1275	1790	114	284	158	150	177	151	3.3×10^{-40}	9.6×10^{-29}	1.1×10^{-26}
HD 174532	516	532	405	107	132	89	328	377	284	1.5×10^{-9}	2.9×10^{-8}	3.3×10^{-16}
HD 174589	237	246	116	58	80	35	208	260	100	5.9×10^{-25}	4.3×10^{-5}	4.1×10^{-34}
HD 51722	1461	2913	1379	210	53	178	179	284	193	5.9×10^{-23}	5.3×10^{-34}	3.1×10^{-35}
HD 51359	1692	1681	2634	144	244	175	174	229	306	5.7×10^{-28}	9.1×10^{-10}	6.3×10^{-29}
HD 50870	2041	3294	2150	204	219	207	239	290	260	3.3×10^{-21}	6.3×10^{-28}	2.0×10^{-18}
HD 170699	2878	2993	2042	238	150	198	326	392	310	1.4×10^{-12}	1.1×10^{-1}	1.5×10^{-10}
GSC 00144–03031	2971	2174	1735	73	304	157	444	429	285	6.2×10^{-12}	8.8×10^{-11}	1.5×10^{-5}
HD 41641	2017	2833	2301	182	249	238	351	419	389	1.6×10^{-17}	7.7×10^{-14}	6.1×10^{-4}
HD 48784	183	171	141	28	32	43	253	222	139	2.1×10^{-3}	3.0×10^{-1}	8.3×10^{-3}

Notes. Columns (A, L, G) correspond, respectively, to the frequencies detected in ARMA interpolated, linearly interpolated, and gapped light curves. The last three columns represent the Anderson–Darling (AD) test applied to the histogram densities of each (A, L, G) case compared by pairs: A with G, A with L, and L with G, respectively. These values correspond to the p-value of the AD test. The null hypothesis is that compared samples arose from a common distribution.

We check the consistency of the analysis (unbiasedness) comparing the AD test of Tables 3 and 4 and verifying if the differences in the frequency distribution of A, L, and G are always less after the pre-whitening process than before. We call this the consistency condition.

The possibilities are:

A – The frequency distributions are initially similar, and are similar at the end.

B – The frequency distributions are initially similar, but are different at the end.

C – The frequency distributions are initially different, but are similar at the end.

D – The frequency distributions are initially different, and are different at the end.

Cases A and C imply that the pre-whitening process fulfills the consistency condition. Case B shows an inconsistency in the analysis, i.e., the pre-whitening process biases the results. Finally, though case D fails the consistency condition, its origin is indeterminate. Now, assuming as significance criterion $p > 10^{-2}$ and comparing Tables 3 and 4, we can affirm the following statements relative to each comparison:

- AG - HD 174936 is case D, and the rest are B cases. No star fulfills the consistency condition.
- AL - HD 170699 and HD 48784 fulfill the consistency condition (case A) as does HD 174936 (case C). HD 50870 and HD 41641 are D cases and the rest are B cases.
- LG - HD 50870 and HD 41641 are D cases and the rest are B cases. No stars fulfill the consistency condition.

Since the intersection of the sets AL, AG, and LG fulfilling the consistency condition is null we conclude unequivocally that there is an inconsistency in the pre-whitening process.

Additionally, most of the tests are B cases showing that during the pre-whitening process spurious residual frequencies appear as a consequence of a bad fitting. This is consistent with the results that Balona (2014) obtained for HD 50844.

Table 4. Anderson–Darling (AD) test applied to the histogram densities of the (A,L,G) oscillation spectrum before any pre-whitening process was applied.

		A–D test	
Star	AG	AL	LG
HD 50844	(0, 1)	(0, 1)	(0, 1)
HD 174936	$(82.7, 1.04 \times 10^{-45})$	$(82{,}7, 1.04 \times 10^{-45})$	(0,1)
HD 174966	(0, 1)	(0, 1)	(0, 1)
HD 181555	(0, 1)	(0, 1)	(0, 1)
HD 49434	(0, 1)	(0, 1)	(0, 1)
HD 172189	(0, 1)	(0, 1)	(0, 1)
HD 174532	(0, 1)	(0, 1)	(0, 1)
HD 174589	(0, 1)	(0, 1)	(0, 1)
HD 51722	(0, 1)	(0, 1)	(0, 1)
HD 51359	(0, 1)	(0, 1)	(0, 1)
HD 50870	(0, 1)	$(114{,}3.45 \times 10^{-63})$	$(114, 3.45 \times 10^{-63})$
HD 170699	(0, 1)	(0, 1)	(0, 1)
GSC 00144-03031	(0, 1)	(0, 1)	(0, 1)
HD 41641	(0, 1)	$(112, 1.29 \times 10^{-61})$	$(112, 1.29 \times 10^{-61})$
HD 48784	(0, 1)	(0, 1)	(0, 1)

Notes. The pairs AG, AL, and LG represent the compared distributions for the three gap-filling methods considered in this work. The numbers in brackets are the values of the AD test and their corresponding p-values. The null hypothesis is that compared samples arise from a common distribution.

6. Conclusions

In PG15 it was shown that, for filling gaps in time series, it is essential to use interpolation techniques like MIARMA, which are aimed to preserve the original frequency content of the time series. Otherwise, the periodogram cannot be an unbiased estimator of the pulsational content of the stars.

Here we have used that result to investigate whether or not the classical pre-whitening procedure can lead to biased results when analyzing δ Scuti stars. To do this we have applied the widely used algorithm SIGSPEC to a set of 15 δ Scuti stars from the Seismofield of CoRoT to perform a frequency analysis of linearly interpolated, ARMA interpolated, and gapped light curves. This allowed us to test the efficiency of the pre-whitening cascade applied by this program in suppressing the contribution of the spectral window to the power spectrum.

The differences found between the results of the analyses show that, at least for δ Scuti stars, the pre-whitening cascade is not sufficiently capable of removing the spurious frequencies caused by the presence of gaps. This is a novel result that highlights the importance of using a gap-filling method aimed at preserving the information.

These results might have a significant impact on asteroseismic studies. In particular for δ Scuti stars, the study of quasi-periodicities is being used to constrain the internal structure of the stars (see e.g., García Hernández et al. 2009, 2015; Suárez et al. 2014). These periodicities are highly sensitive to the distribution of frequencies in the periodogram. Any variation due to an incorrect pre-whitening of the light curves might introduce a bias in these periodicities. We evaluate this in a future paper (Suárez et al. in prep.) using the same sample of δ Scuti stars observed by CoRoT satellite.

Acknowledgements. JPG acknowledge support from the "Junta de Andalucía" local government under project 2012-P12-TIC-2469. JCS and AGH acknowledge funding support from Spanish public funds for research under project ESP2015-65712-C5-5-R (MINECO/FEDER), and from project RYC-2012-09913 under the "Ramón y Cajal" programme of the Spanish MINECO. JCS also acknowledge support by the European FP7 project "SPACEINN" (FP7-SPACE-2011-1). AM acknowledges support by the Spanish MINECO National Space Program through project ESP2015-65712-C5-1-R and "Ramon y Cajal" program RYC-2012-09913. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 749962. The CoRoT space mission, launched on 2006 December 27, was developed and is operated by the CNES, with participation of the Science Programs of ESA, ESA's RSSD, Austria, Belgium, Brazil, Germany and Spain.

References

Akaike, H. 1974, IEEE Trans. Autom. Control, AC-19, 716

Appourchaux, T., Michel, E., Auvergne, M., et al. 2008, A&A, 488, 705

Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411

Baglin, A., Auvergne, M., Barge, P., et al. 2006, ESA SP, 1306, 33

Balona, L. A. 2014, MNRAS, 439, 3453

Barceló Forteza, S., Michel, E., Roca Cortés, T., & García, R. A. 2015, A&A, 579, A133

Benomar, O., Baudin, F., Campante, T. L., et al. 2009, A&A, 507, L13

Brown, T. M., & Christensen-Dalsgaard, J. 1990, ApJ, 349, 667

Brunsden, E., Pollard, K. R., Cottrell, P. L., et al. 2015, MNRAS, 447, 2970

Chapellier, E., Rodríguez, E., Auvergne, M., et al. 2011, A&A, 525, A23

Creevey, O. L., Uytterhoeven, K., Martín-Ruiz, S., et al. 2009, A&A, 507, 901

de Waele, S., & Broersen, P. M. T. 2000, IEEE Trans. Instrum. Meas., 49, 216

Deeming, T. J. 1975, Ap&SS, 36, 137

Fossat, E., Kholikov, S., Gelly, B., et al. 1999, A&A, 343, 608

Foster, G. 1995, AJ, 109, 1889

Fox Machado, L., Alvarez, M., Michel, R., et al. 2010, New Astron., 15, 397

Gabriel, M. 1994, A&A, 287, 685

García, R. A., Régulo, C., Samadi, R., et al. 2009, A&A, 506, 41

García, R. A., Mathur, S., Pires, S., et al. 2014, A&A, 568, A10

García Hernández, A., Moya, A., Michel, E., et al. 2009, A&A, 506, 79

García Hernández, A., Moya, A., Michel, E., et al. 2013, A&A, 559, A63

García Hernández, A., Martín-Ruiz, S., Monteiro, M. J. P. F. G., et al. 2015, ApJ, 811, L29

García Hernández, A., Suárez, J. C., Moya, A., et al. 2017, MNRAS, 471, L140

Gilliland, R. L., Brown, T. M., Christensen-Dalsgaard, J., et al. 2010, Pub. Astron. Soc. Pac., 122, 131

Gutiérrez-Soto, J., Floquet, M., Samadi, R., et al. 2009, A&A, 506, 133

Hogbom, J. A. 1974, A&AS, 15, 417

Kallinger, T., & Matthews, J. M. 2010, ApJ, 711, L35

Mantegazza, L., Poretti, E., Michel, E., et al. 2012, A&A, 542, A24

Martín-Ruiz, S., Amado, P. J., Suárez, J. C., et al. 2005, A&A, 440, 711

Mary, D. L. 2005, JApA, 26, 283

Molnár, L., Derekas, A., Szabó, R., et al. 2017, MNRAS, 466, 4009

Moya, A., Suárez, J. C., García Hernández, A., & Mendoza, M. A. 2017, MNRAS, 471, 2491

Paparó, M., Benkő, J. M., Hareter, M., & Guzik, J. A. 2016, ApJS, 224, 41

Pascual-Granado, J., Garrido, R., & Suárez, J. C. 2015, A&A, 575, A78

Ponman, T. 1981, MNRAS, 196, 583

Poretti, E., Alonso, R., Amado, P. J., et al. 2005, AJ, 129, 2461

Poretti, E., Michel, E., Garrido, R., et al. 2009, A&A, 506, 85

Rainer, M., Poretti, E., Mistò, A., et al. 2016, AJ, 152, 207

Reegen, P. 2007, A&A, 467, 1353

Reese, D. R., Prat, V., Barban, C., van 't Veer-Menneret, C., & MacGregor, K. B. 2013, A&A, 550, A77

Sargent, T. J. 1987, Macroeconomic Theory, 2nd edn. (Cambridge, MA: Academic Press)

Scargle, J. D. 1981, ApJS, 45, 1

Scargle, J. D. 1982, ApJ, 263, 835

Scholz, F. W., & Stephens, M. A. 1987, J. Am. Stat. Assoc., 82, 918

Seleznyov, A. D., Solanki, S. K., & Krivova, N. A. 2011, A&A, 532, A108

Stahn, T., & Gizon, L. 2008, Sol. Phys., 251, 31

Suárez, J. C., García Hernández, A., Moya, A., et al. 2014, A&A, 563, A7

Weiss, W. W., Fröhlich, H.-E., Pigulski, A., et al. 2016, A&A, 588, A54

Zwintz, K., Kallinger, T., Guenther, D. B., et al. 2011, ApJ, 729, 20

Zwintz, K., Moravveji, E., Pápics, P. I., et al. 2017, A&A, 601, A101

Appendix A: Spectral response function of linear interpolation

According to the definition given by Deeming (1975) the expected value of the classical periodogram is obtained as the convolution of the true power spectrum with the spectral window associated to the sampling used. The aim of gap-filling with any interpolation method is to recover the original regular sampling. Therefore, in this sense, the spectral window is just a sinc function. More important is to determine the spectral response function of the model used for the interpolation. Here we discuss the properties of the spectral response function of linear interpolation and its relation to a rectangular window (i.e., no interpolation).

In order to interpolate linearly inside gaps, the algorithm may use just two data points, one at the beginning and one at the end of the gap. Then, a linear interpolation can be considered as simply an oversampling of the gap interval.

Assuming that without loss of generality the gap interval is the unity and ξ is a number between 0 and 1, then a linearly interpolated value at an inner point is:

$$\hat{X}(n+\xi) = (1-\xi)\cdot X(n) + \xi\cdot X(n+1), \tag{A.1}$$

where $X(n)$ and $X(n+1)$ are the data points used for the interpolation. It is easy to see that for $\xi = 0$ and $\xi = 1$ the linearly interpolated value coincides with the data points $X(n)$ and $X(n+1)$ and for $\xi = 1/2$ it is just the mean between both values.

The interpolation error of the ξ datapoint

$$\epsilon_\xi = |\hat{X}(n+\xi) - X(n+\xi)| \tag{A.2}$$

is nonzero between the data points but it can be ≈0 if X is a linear function between n and $n+1$ (not necessarily in other intervals).

Now, for simplicity we assume that $n = 0$, then

$$\hat{X}(\xi) = (1-\xi)\cdot X(0) + \xi\cdot X(1).$$

If we interpolate L points inside a gap interval $(0,1)$, of unit sampling, we can say that this interval will be oversampled to $1/L$. Then $\xi = k/L$ with $0 \le k < L$. Now the equation above can be expressed as:

$$\hat{X}(k/L) = (1-k/L)\cdot X(0) + (k/L)\cdot X(1),$$

and defining $h(k) = 1 - k/L$ and $h(k-L) = k/L$ we have finally

$$\hat{X}(k/L) = h(k)\cdot X(0) + h(k-L)\cdot X(1). \tag{A.3}$$

In this way we have defined linear interpolation through the convolution of a triangular filter h of length $2L-1$ with impulse response:

$$\begin{aligned} h(k) &= 1 - |k|/L, \qquad |k| < L \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{A.4}$$

If we want to interpolate four points, for example, $L = 5$ and the impulse response h has a duration of $N = 2\cdot L - 1 = 2\cdot 5 - 1 = 9$ but only two nonzero samples are coincident with h.

A triangular filter can be defined as the convolution of two rectangular functions:

$$\mathrm{tri}(k) = \mathrm{rect}(k)\times\mathrm{rect}(k).$$

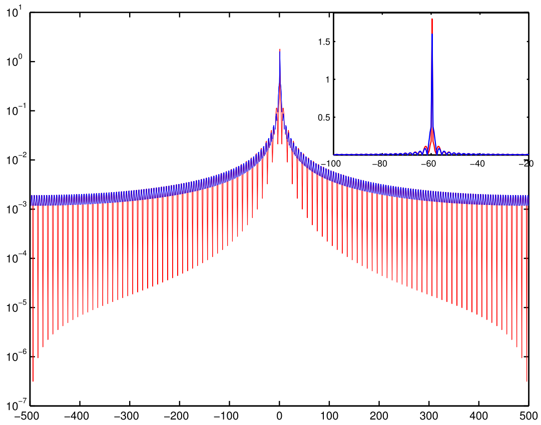

Fig. A.1. Spectral response function in log scale associated to gapped data (in blue) and linearly interpolated data (in red). See the inset for a zoom of the central peak in linear scale.

Therefore the frequency response of the linear interpolation is a product of $\sin c$ functions:

$$P(\nu) = N^2 \sin c^2(\nu/N). \tag{A.5}$$

In Fig. A.1 we compare the spectral response function caused by a gap and the one corresponding to linearly interpolated data. We note that, though linear interpolation concentrates the power at the central peak and reduces the power at the sidelobes, the shape of the function is very similar.

We note now that expression (A.1) looks like an autoregressive model. Indeed, this could explain that, in some cases, linear and ARMA interpolation give similar results since, by serendipity, the signal might have low variability near the gap and the models fitted by MIARMA to local data would have low order. However, in spite of the similarity between expression (A.1) and an AR model, the properties of the spectral response function cannot be extrapolated and some serious difficulties appear when trying to identify both methods, showing that their spectral response functions are in general very different.

First of all, MIARMA use only a causal representation of the data, either in forward or backward extrapolation (see PG15 for more details). However, in Eq. (A.1) the interpolated value $\hat{X} + \xi$ is obtained using $X(n+1)$, so linear interpolation can be considered a mixed (causal + acausal) representation. Using mixed representations is not an issue as demonstrated in Scargle (1981), but we should compare the spectral response function with an AR(1,1) model with coefficients $(1-\xi, 1, \xi)$ and this is formally different to what MIARMA does.

Now, for the sake of clarity we could evaluate the spectral response function of the most similar AR representation to Eq. (A.1) that is used by MIARMA which is, in fact, an AR(1) model

$$X(n+1) = a\cdot X(n) + E(n), \tag{A.6}$$

where $E(n)$ is an uncorrelated random time series representing the input of the model, and a is a constant coefficient with $|a| < 1$. $E(n)$ can be understood as random pulses feeding the model; this makes the AR model intrinsically different to Eq. (A.1) where there is no random component.

Furthermore, the mathematical form of Eq. (A.1) introduces some additional restrictions: the same model Eq. (A.6) should be fitted for both data segments around the gap, and the weighting between forward and backward extrapolation should be a

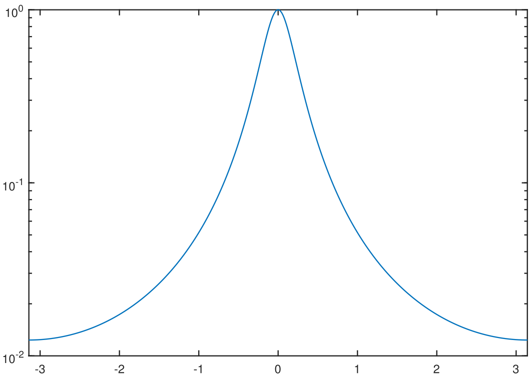

Fig. A.2. Spectral response function in log scale associated to an AR(1) model with $a = 0.8$.

triangular function. The first one is quite reasonable since it assumes stationarity which is guaranteed in most cases for a short time interval. The weighing function can be freely chosen in MIARMA but the default is a triangular function too, so there is no problem with any of these assumptions. Now, if we follow the same reasoning as before, Eq. (A.1) is now:

$$\hat{X}(n+\xi) = (1-\xi)\cdot\hat{X}^f(n+\xi) + \xi\cdot\hat{X}^b(n+\xi), \tag{A.7}$$

where $\hat{X}^f(n+\xi)$ and $\hat{X}^b(n+\xi)$ are the forward and backward extrapolations at $n+\xi$. These can be obtained from Eq. (A.6) and with some simple calculations it is easy to see that

$$\begin{aligned}\hat{X}(k/L) &= ah(k)X(k-1/L) + (1/a)h(k-L)X(k+1/L)\\ &\quad + ah(k)E(k-1/L) - (1/a)h(k-L)E(k+1/L).\end{aligned} \tag{A.8}$$

The differences between Eqs. (A.3) and (A.8) are clear. In addition to the random input $E(1/L)$ and the coefficients a and $1/a$, we note that for each partition k/L, the function X is evaluated in $k-1/L$ and $k+1/L$ and not in 1 and 0. That is, MIARMA fits causal models to the data bracketing the gap but each interpolated value depends on the previous and consecutive data points and not on fixed values $X(1), X(0)$.

It is not so simple to calculate the spectral response function of AR(1) as it is for linear interpolation but with a little more effort it can be shown (see, e.g., p. 261 of Sargent 1987) that it is:

$$P(\nu) = \frac{\sigma^2}{1-2a\cdot\cos(\nu)+a^2}, \tag{A.9}$$

where σ^2 is the variance. In Fig. A.2 we plot the spectral response function of an AR(1) model with $a = 0.8$ to illustrate the differences between the frequency responses of Eqs. (A.3) and (A.8).

We have studied in this appendix the spectral response function of an interpolation with a linear fitting and with an AR(1) model. Only a single gap has been considered for this study and the full spectral response depends, of course, on the gap distribution. The full spectral response will be, in general, more complex, but we can gain insight into the effect of the interpolation on spectral analysis with the calculations presented here.

Appendix B: Gap-filling features

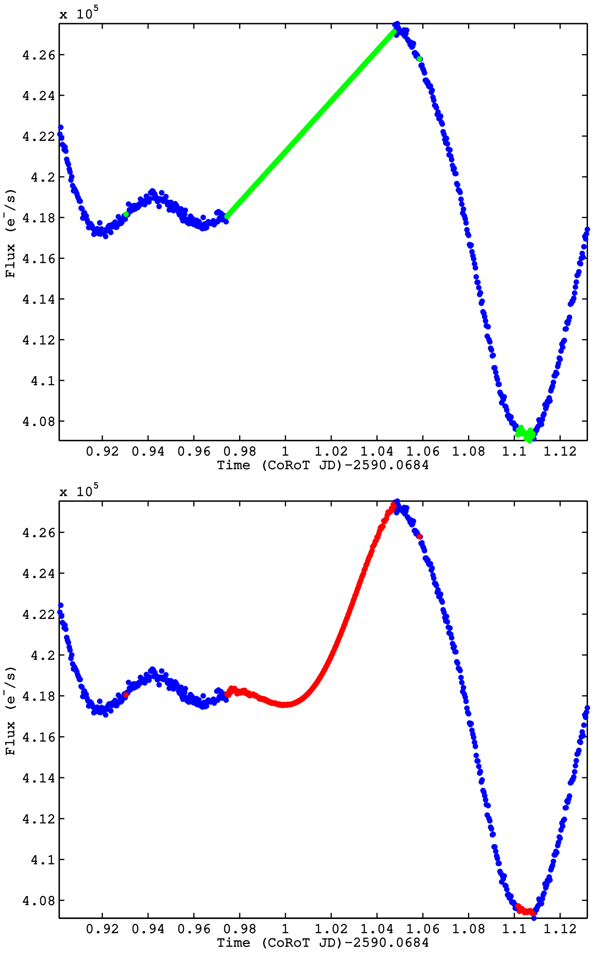

Fig. B.1. Illustration of a gap in HD 50844 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

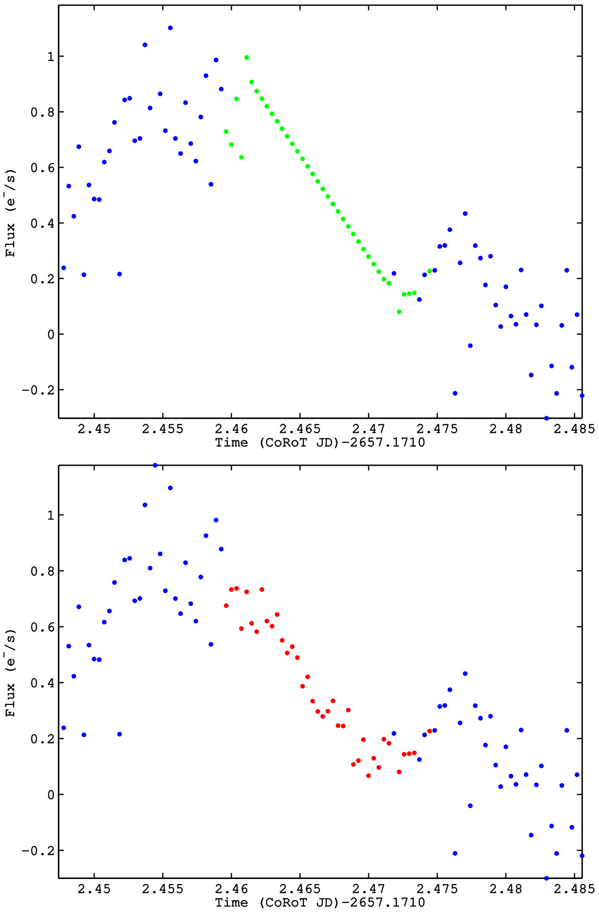

Fig. B.2. Illustration of a gap in HD 174936 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

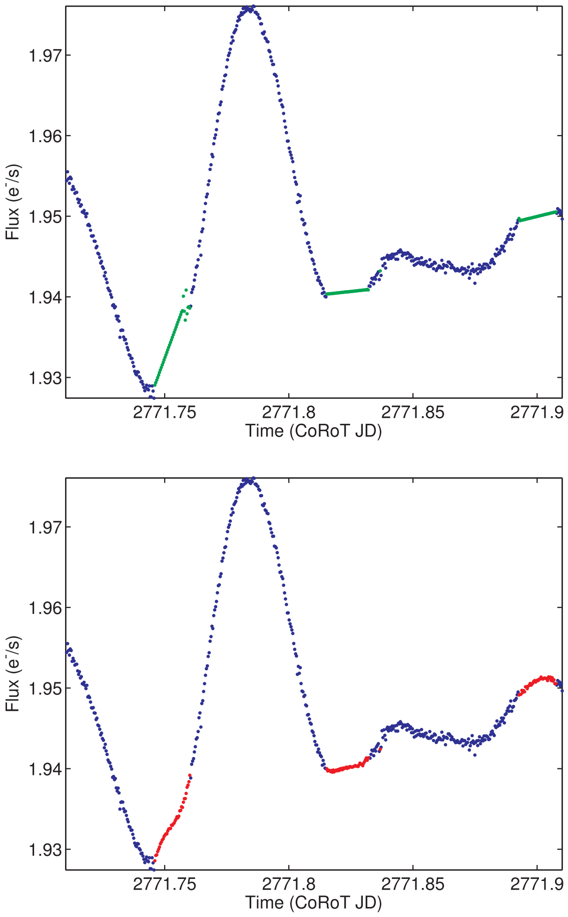

Fig. B.3. Illustration of a gap in HD 181555 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

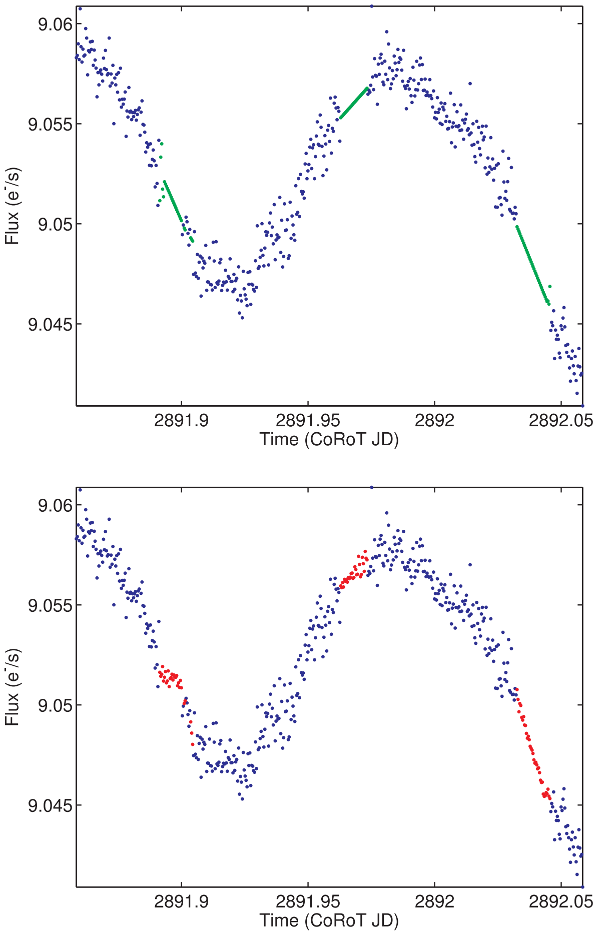

Fig. B.4. Illustration of a gap in HD 49434 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

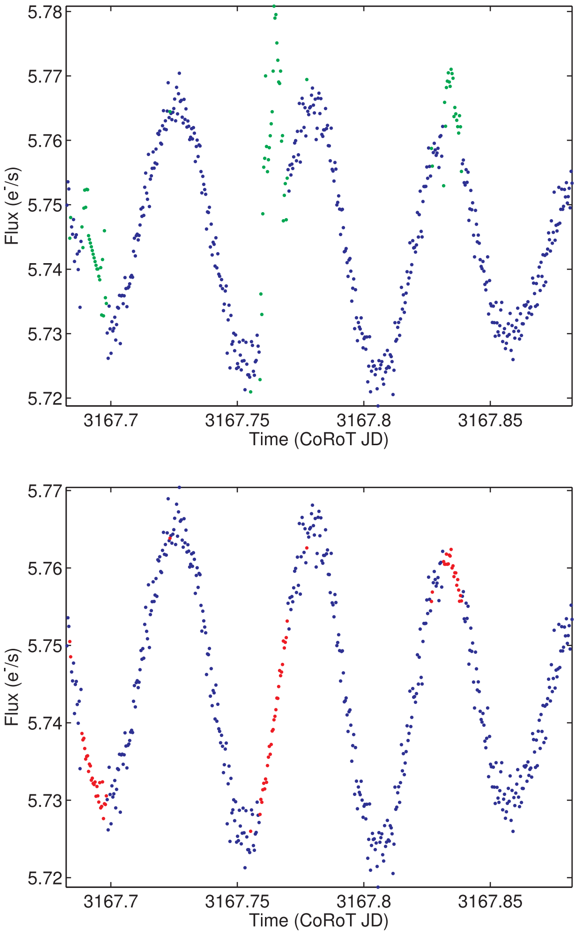

Fig. B.5. Illustration of a gap in HD 172189 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

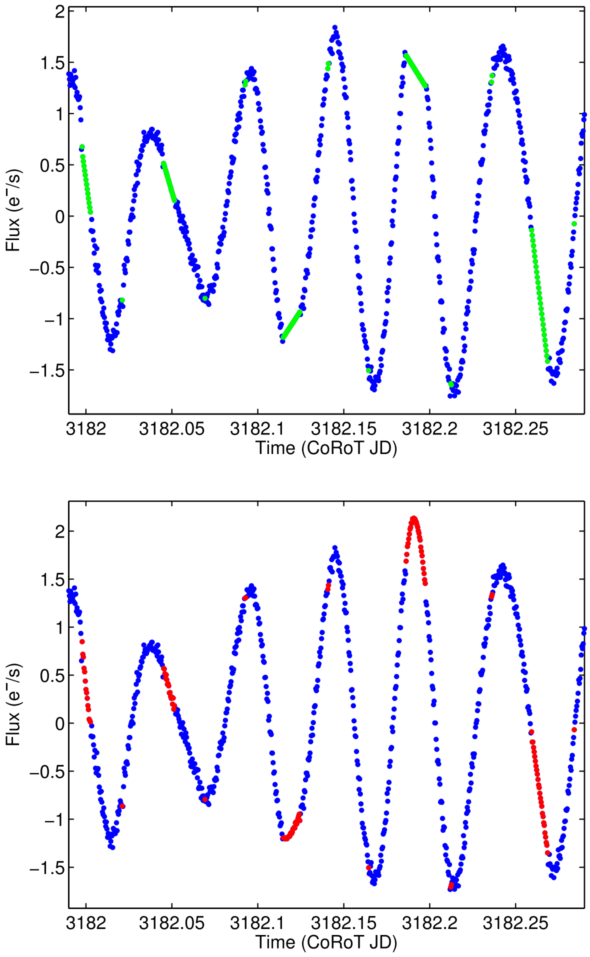

Fig. B.6. Illustration of gaps in HD 174532 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

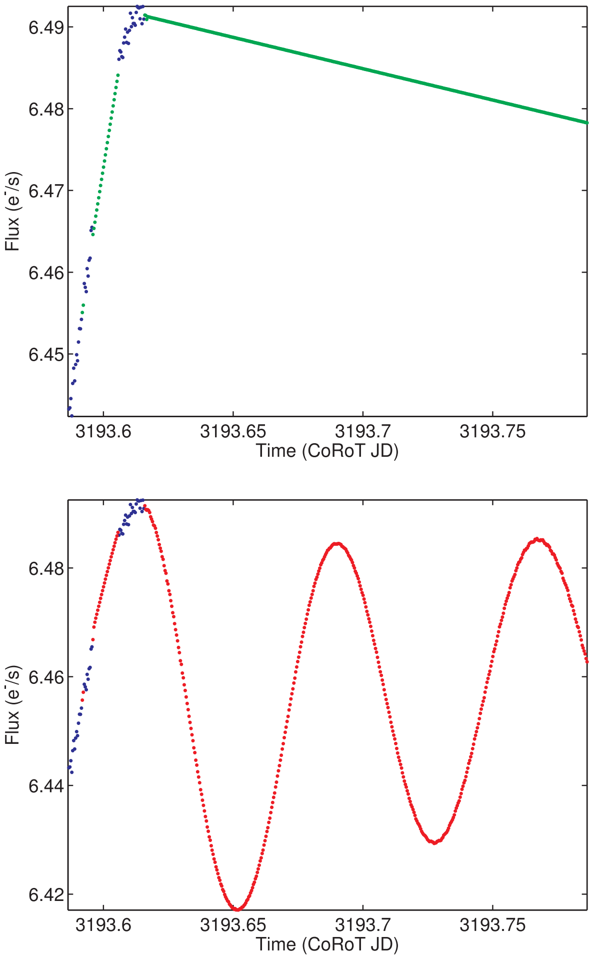

Fig. B.7. Illustration of a gap in HD 174589 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

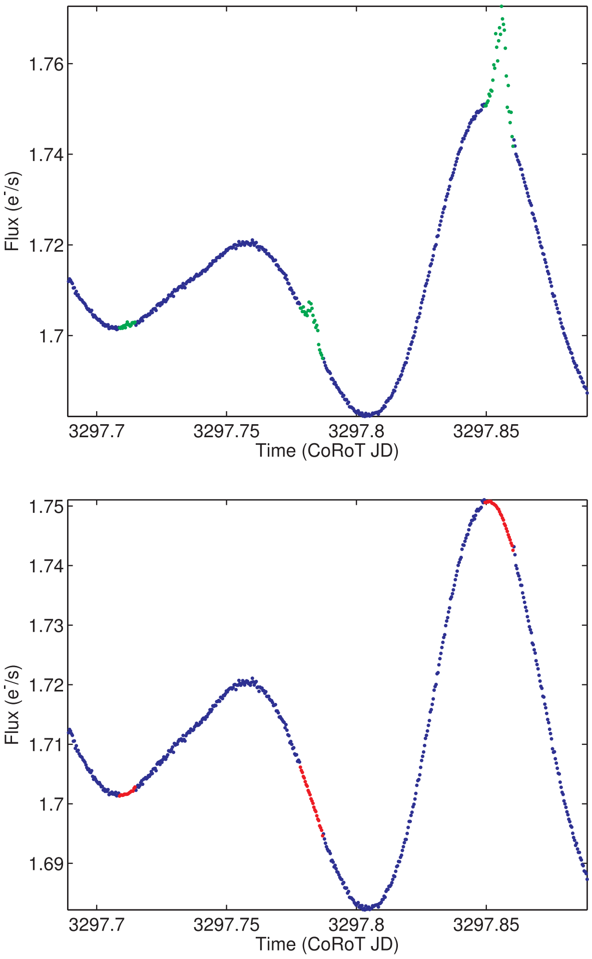

Fig. B.8. Illustration of a gap in HD 51722 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

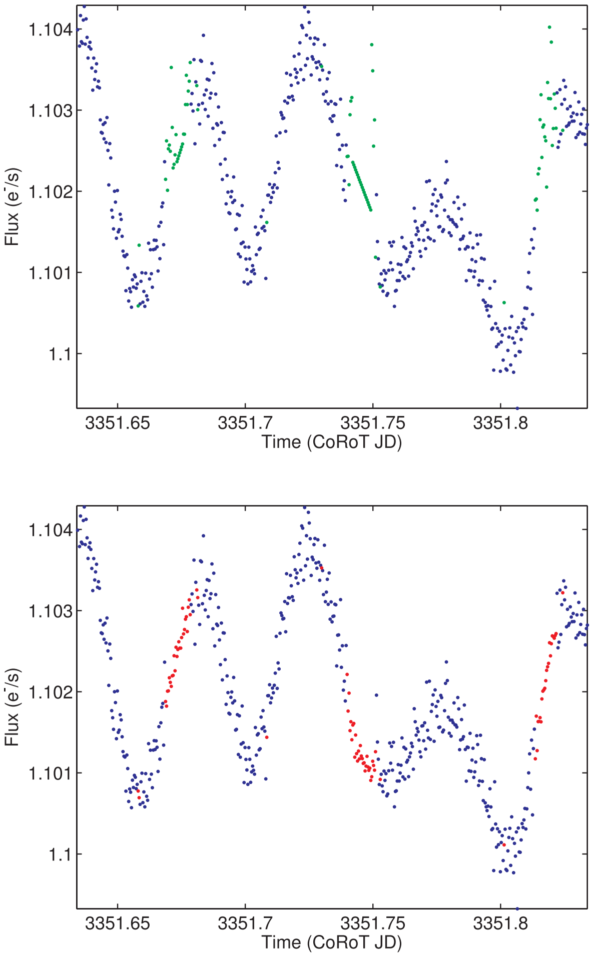

Fig. B.9. Illustration of a gap in HD 51359 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

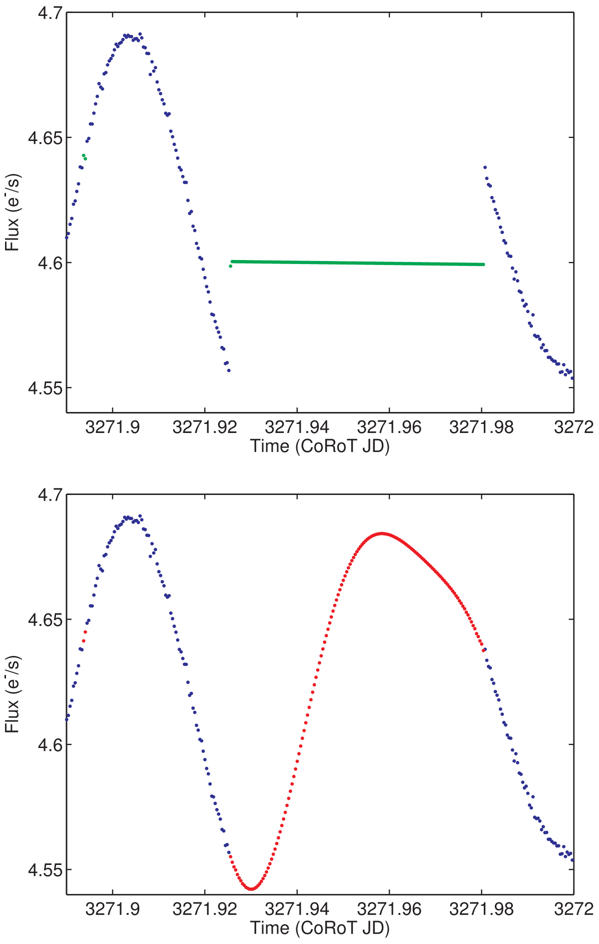

Fig. B.10. Illustration of a gap in HD 50870 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

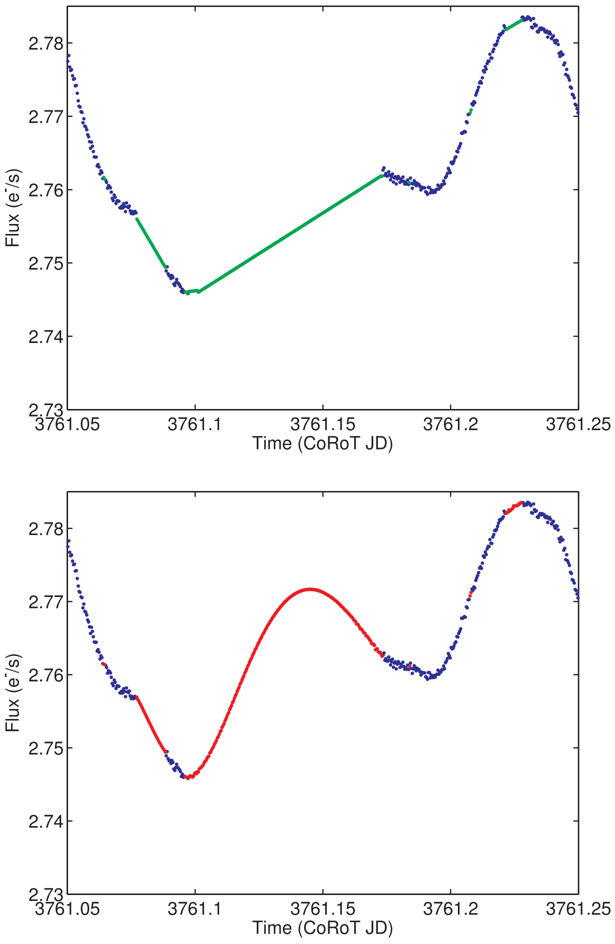

Fig. B.11. Illustration of a gap in HD 170699 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

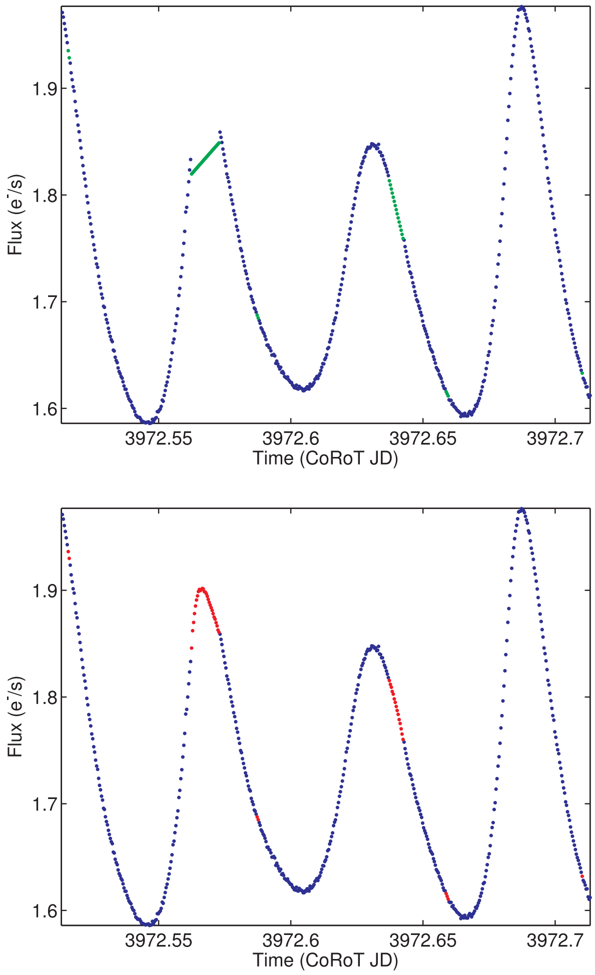

Fig. B.12. Illustration of a gap in GSC00244-03031 that has been interpolated linearly (*upper panel*, in green), and with MIARMA (*lower panel*, in red).

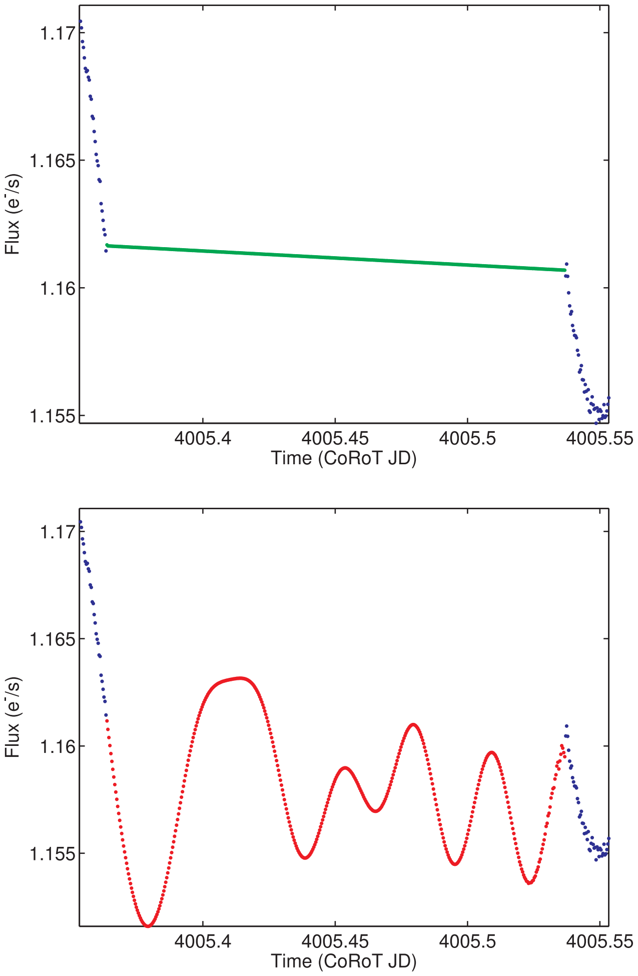

Fig. B.13. Illustration of gaps in HD 41641 that have been interpolated linearly (*upper panel*, in green, and with MIARMA (*lower panel*, in red).

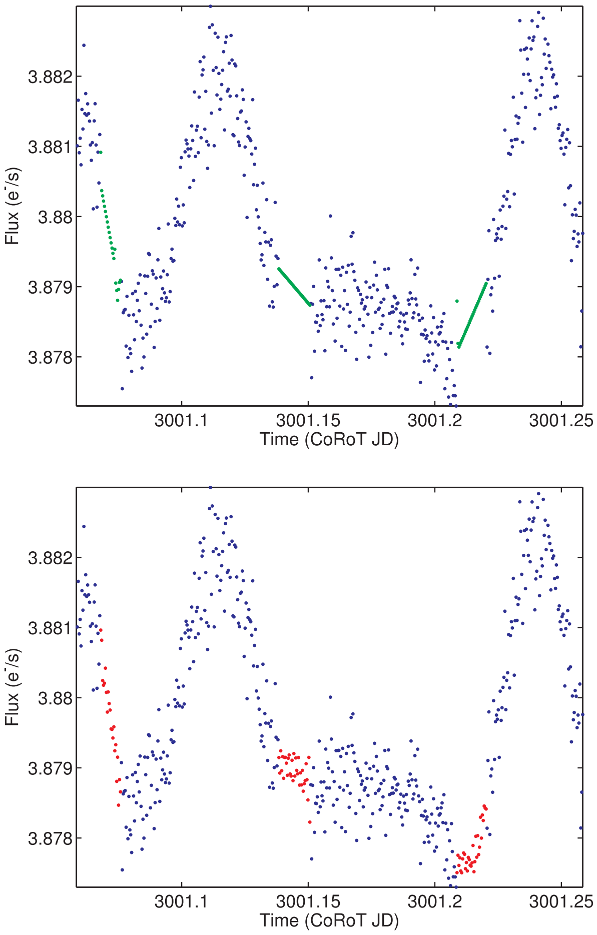

Fig. B.14. Illustration of gaps in HD 48784 that have been interpolated linearly (*upper panel*, in green, and with MIARMA (*lower panel*, in red).

Appendix C: Power spectra

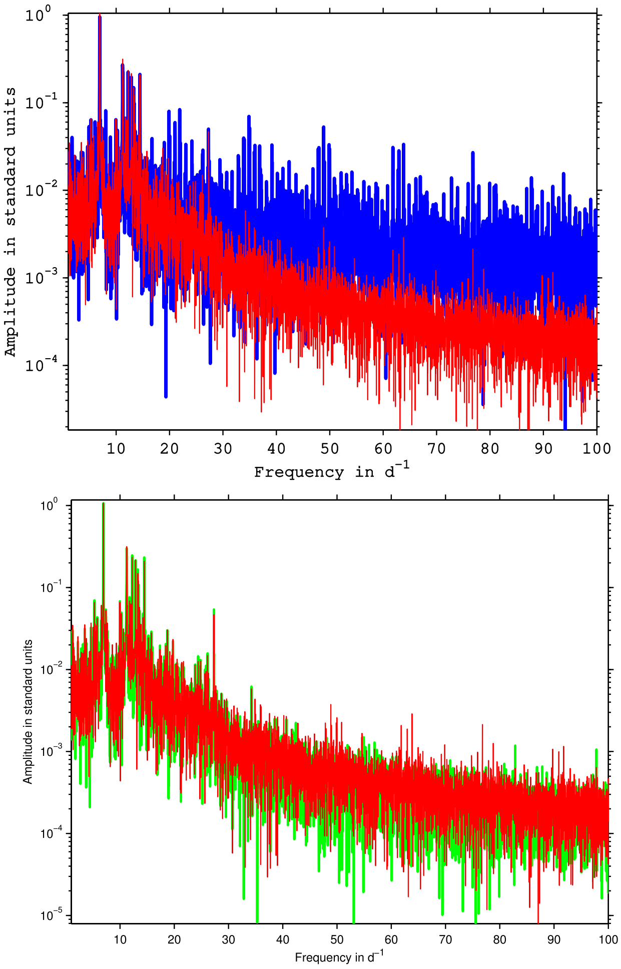

Fig. C.1. Power spectra of the light curves from HD 50844: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

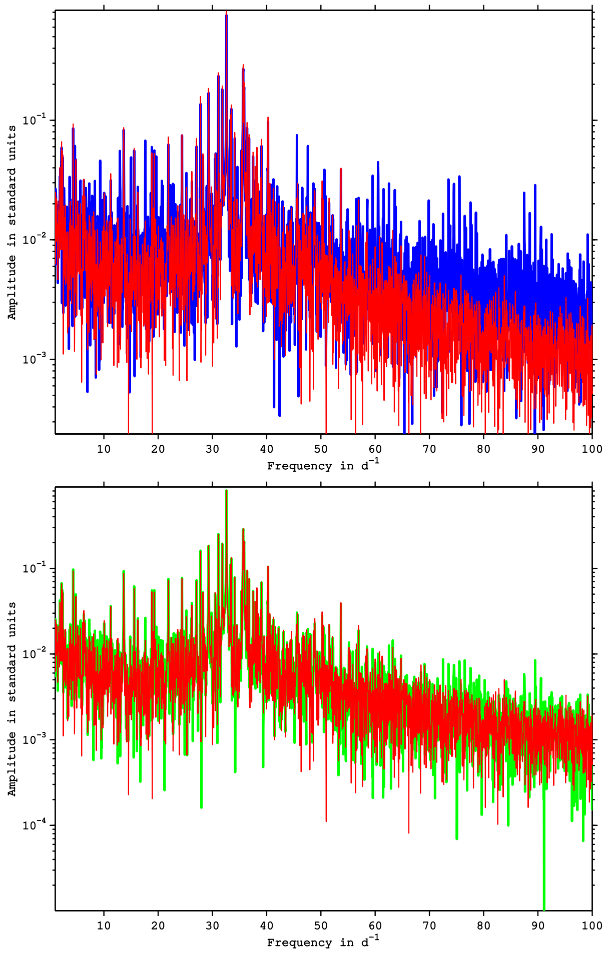

Fig. C.2. Power spectra of the light curves from HD 174936: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

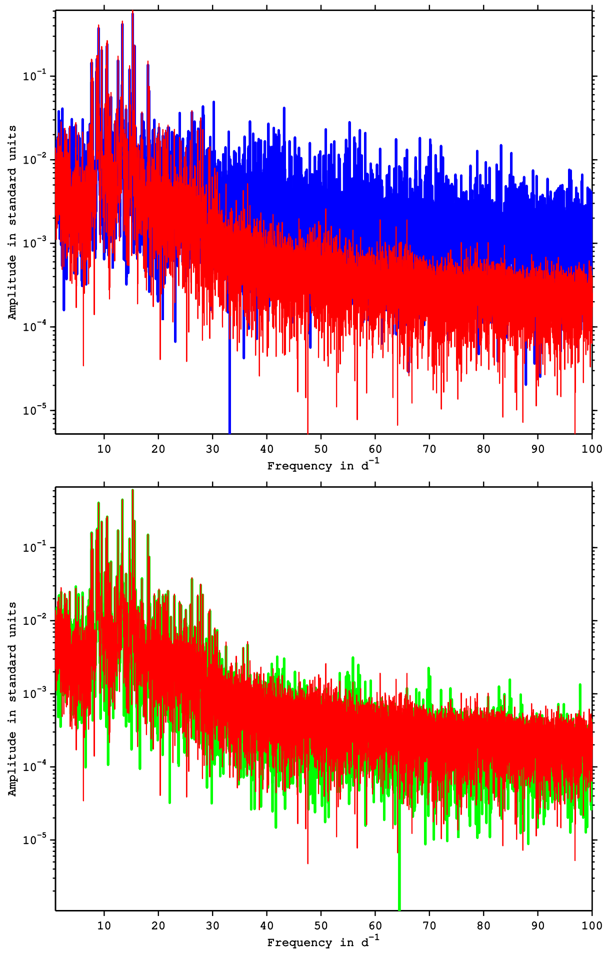

Fig. C.3. Power spectra of the light curves from HD 181555: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

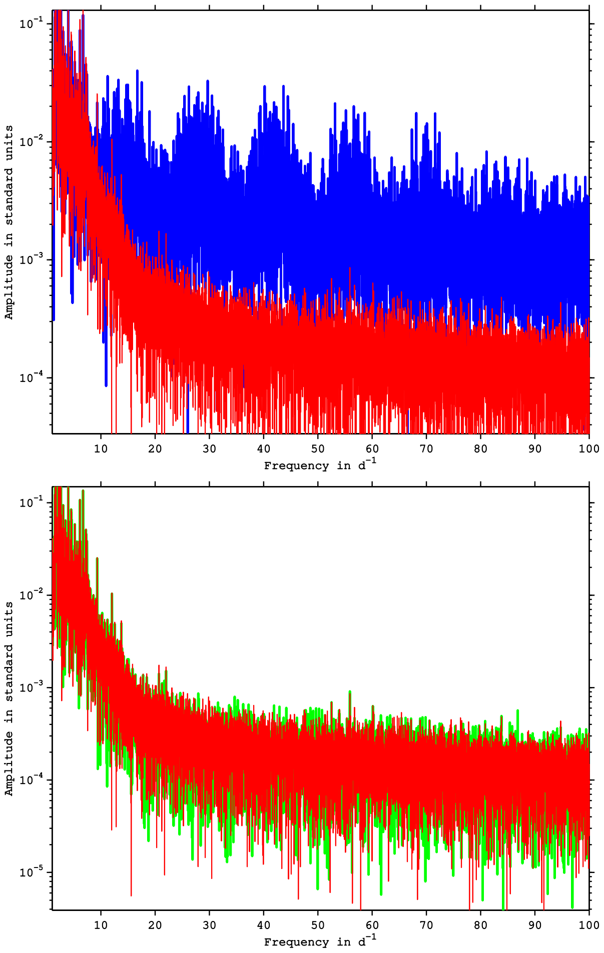

Fig. C.4. Power spectra of the light curves from HD 49434: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

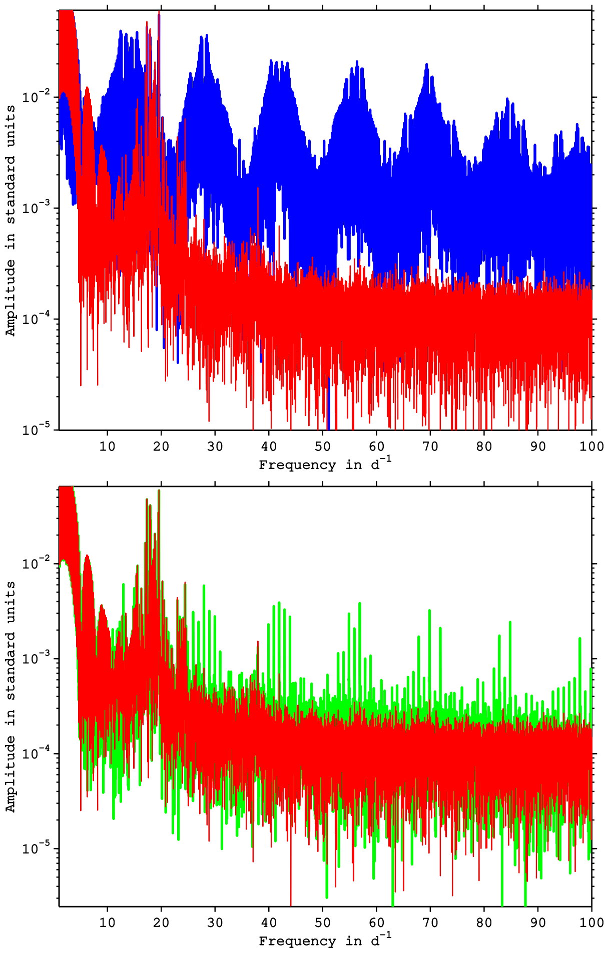

Fig. C.5. Power spectra of the light curves from HD 172189: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

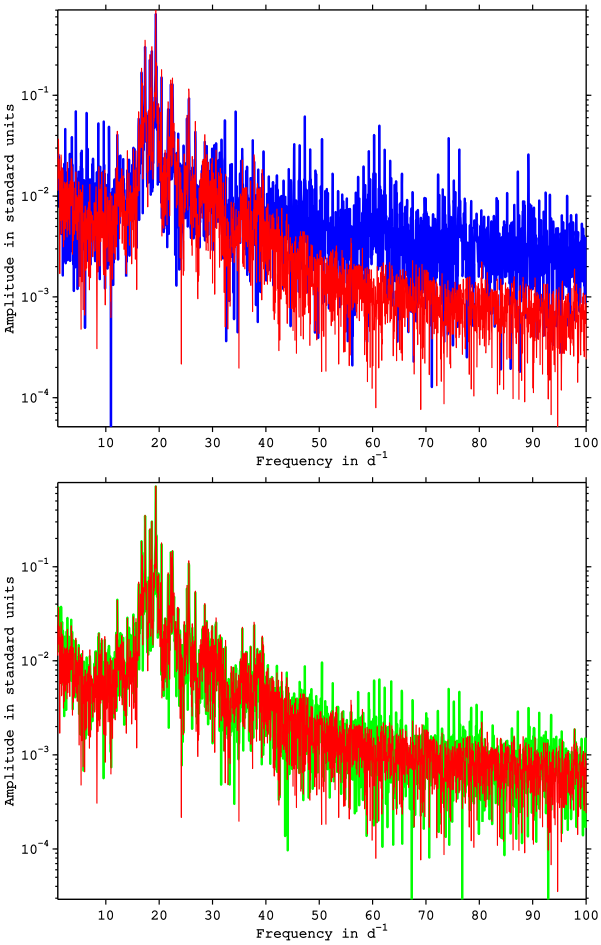

Fig. C.6. Power spectra of the light curves from HD 174532: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

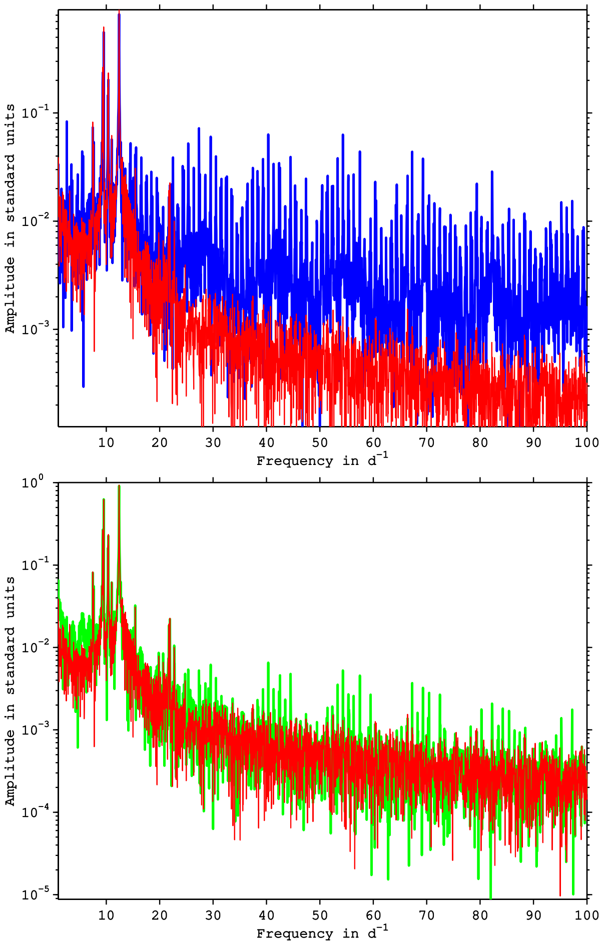

Fig. C.7. Power spectra of the light curves from HD 174589: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

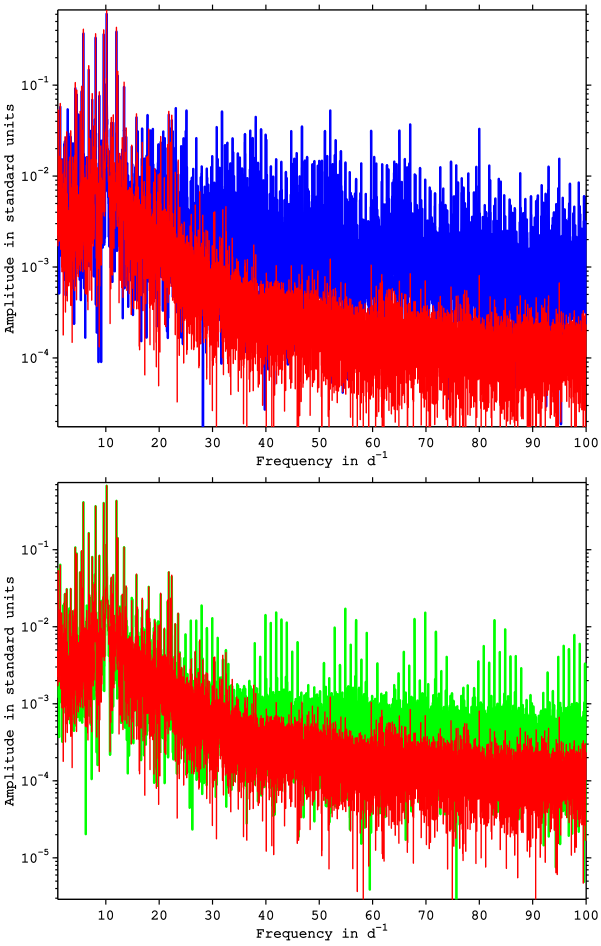

Fig. C.8. Power spectra of the light curves from HD 51722: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

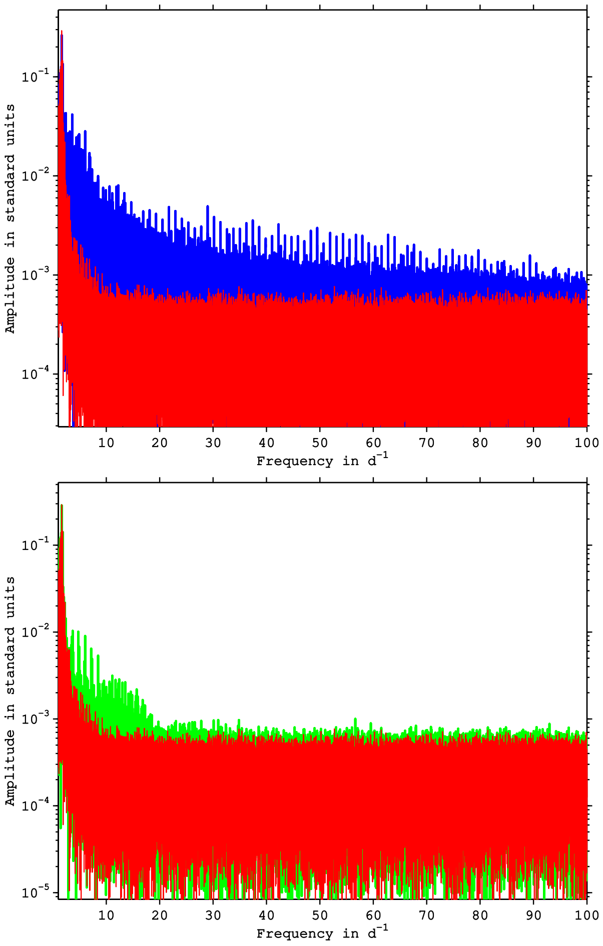

Fig. C.9. Power spectra of the light curves from HD 51359: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

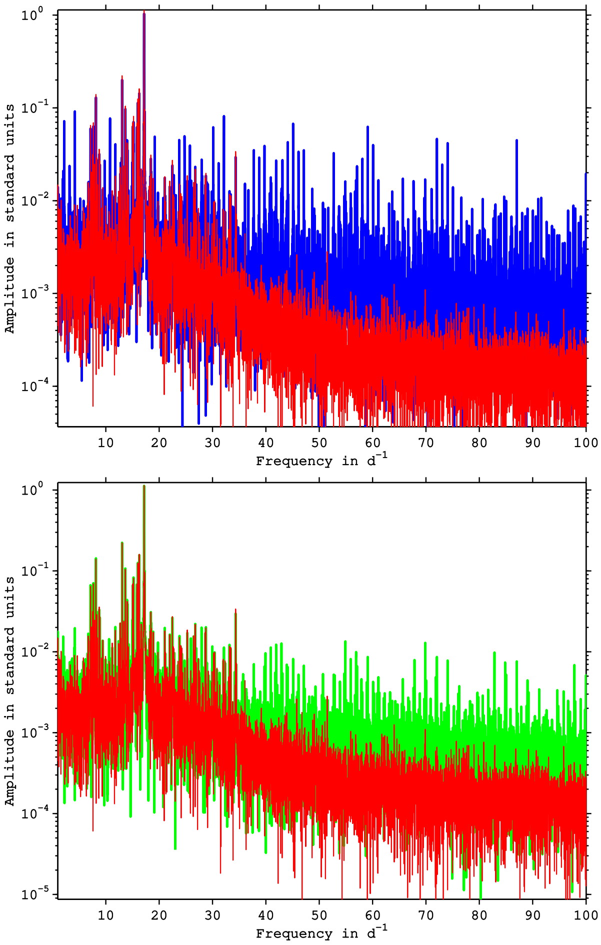

Fig. C.10. Power spectra of the light curves from HD 50870: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

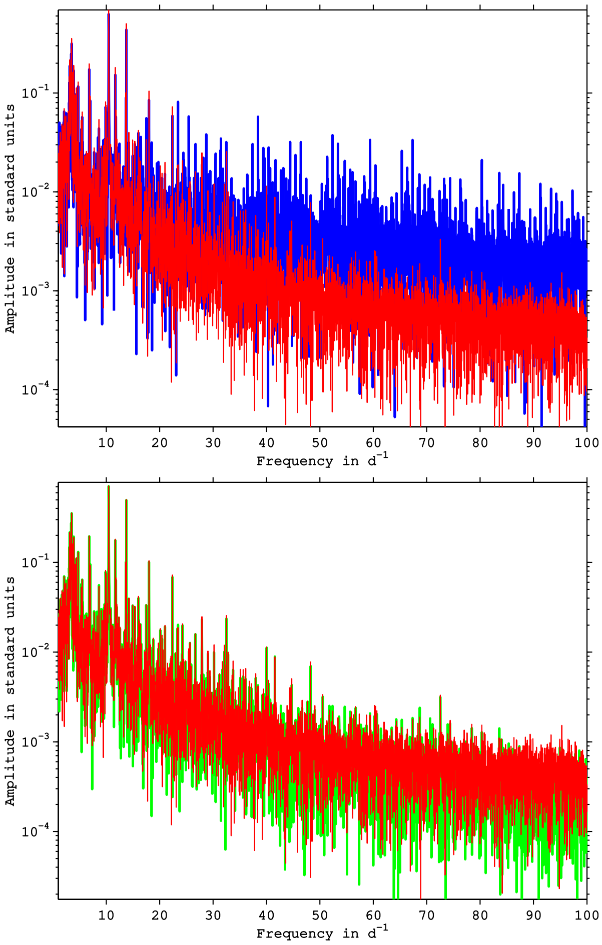

Fig. C.11. Power spectra of the light curves from HD 170699: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

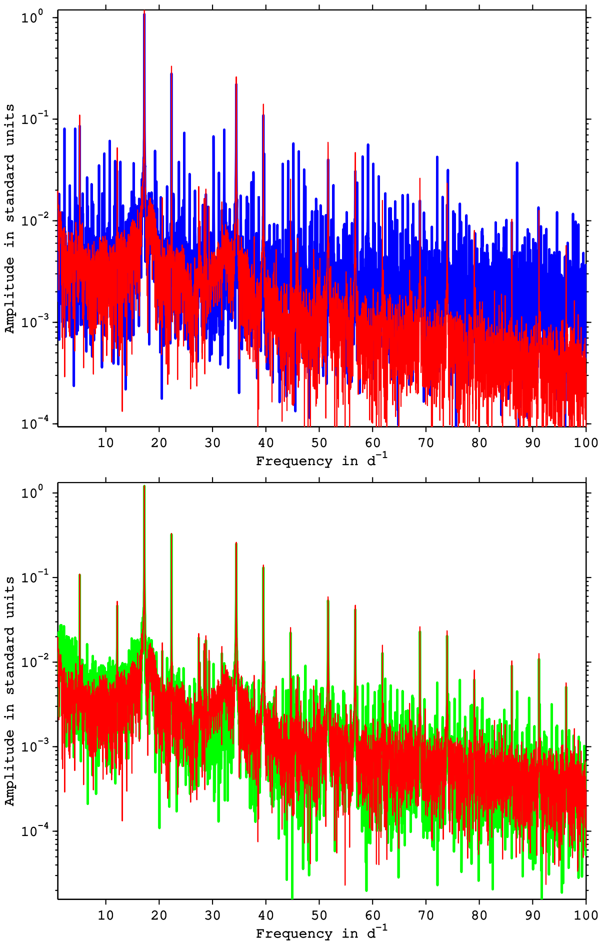

Fig. C.12. Power spectra of the light curves from GSC 00144-03031: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

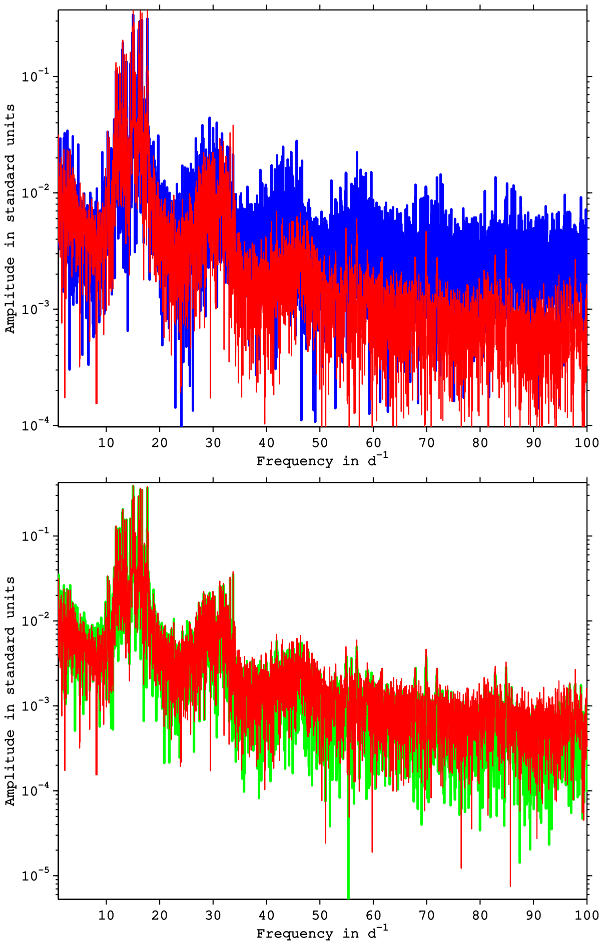

Fig. C.13. Power spectra of the light curves from HD 41641: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

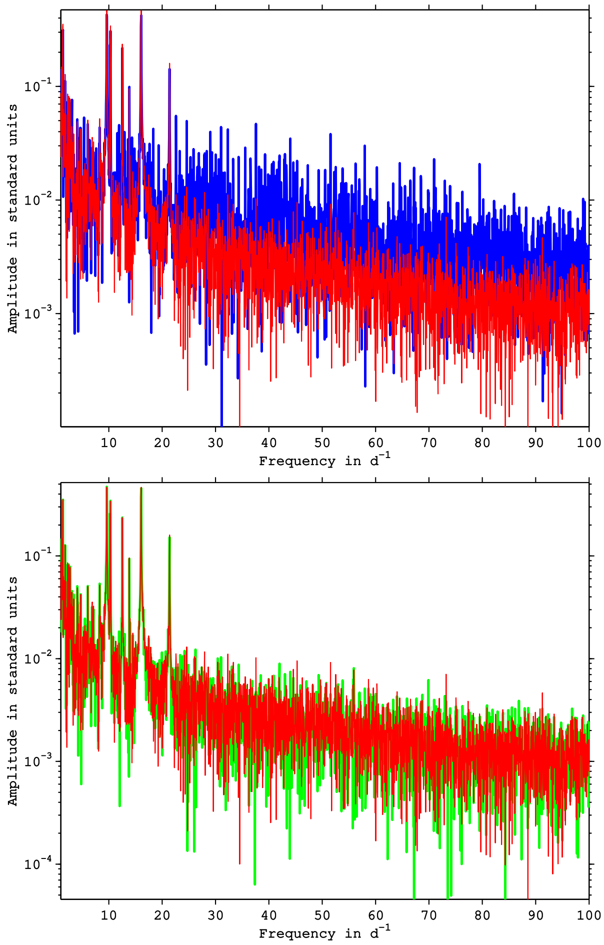

Fig. C.14. Power spectra of the light curves from HD 48784: *upper panel* shows gapped data in blue and ARMA interpolated data in red, *lower panel* shows linearly interpolated data in green and ARMA also in red.

Appendix D: Histograms

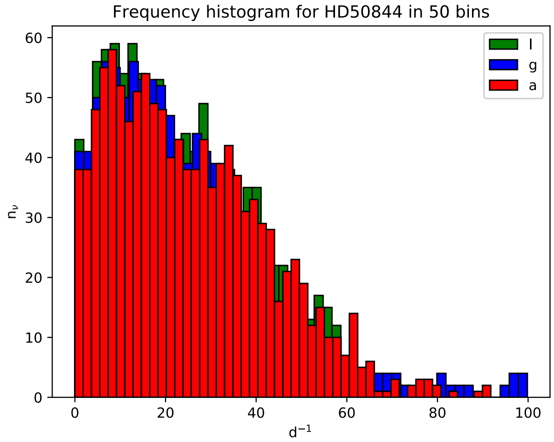

Fig. D.1. Histograms of detected frequencies in the light curves of HD 50844. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

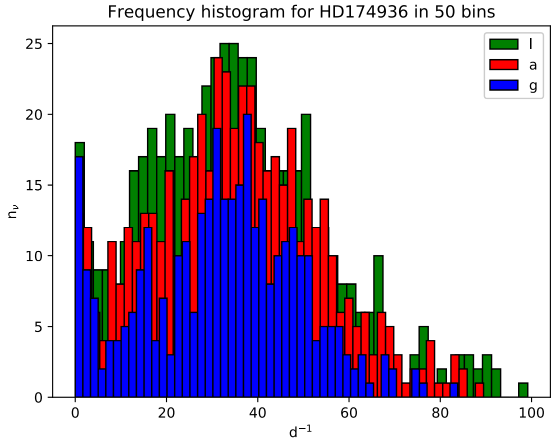

Fig. D.2. Histograms of detected frequencies in the light curves of HD 174936. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

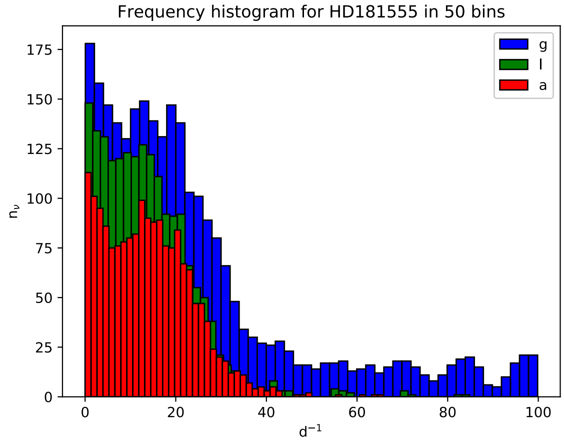

Fig. D.3. Histograms of detected frequencies in the light curves of HD 181555. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

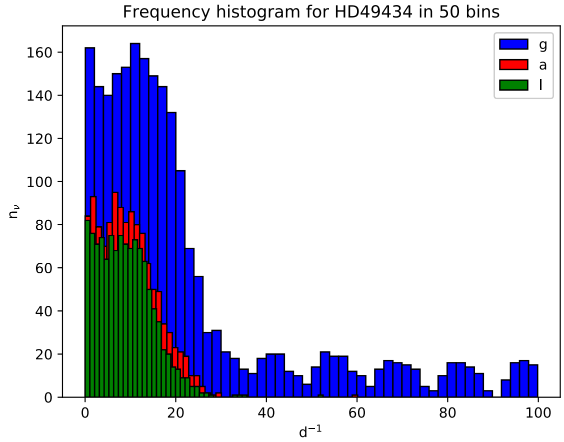

Fig. D.4. Histograms of detected frequencies in the light curves of HD 49434. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

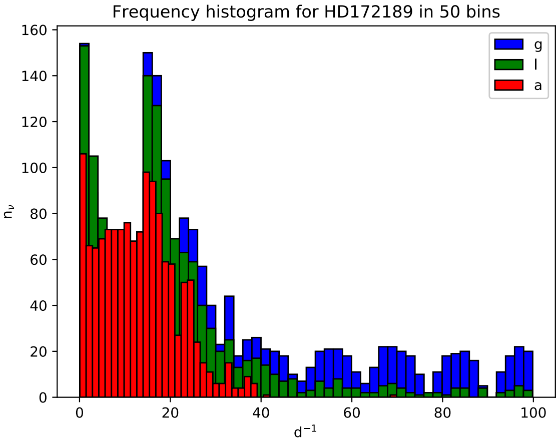

Fig. D.5. Histograms of detected frequencies in the light curves of HD 172189. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

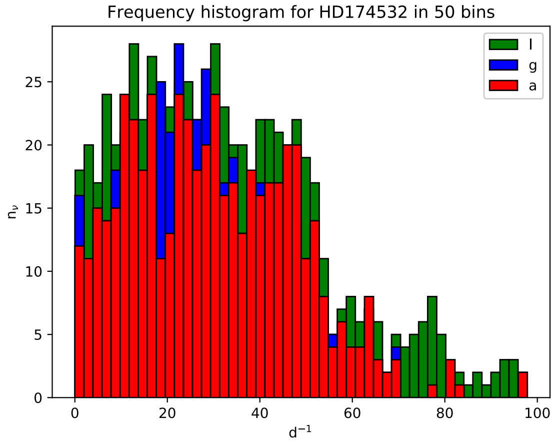

Fig. D.6. Histograms of detected frequencies in the light curves of HD 174532. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

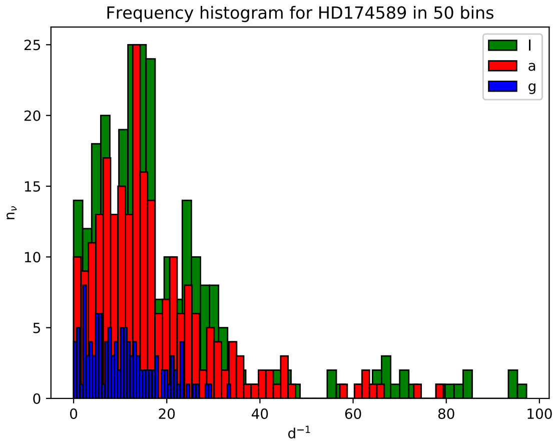

Fig. D.7. Histograms of detected frequencies in the light curves of HD 174589. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

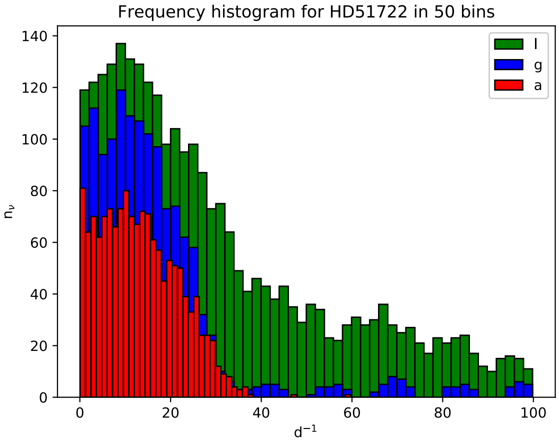

Fig. D.8. Histograms of detected frequencies in the light curves of HD 51722. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

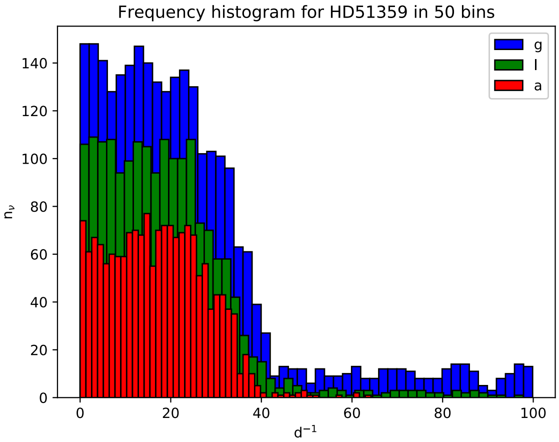

Fig. D.9. Histograms of detected frequencies in the light curves of HD 51359. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

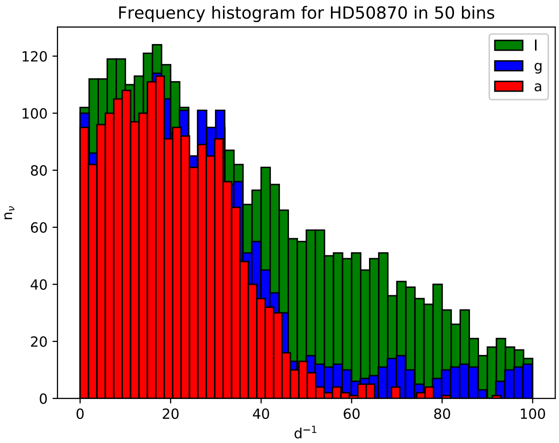

Fig. D.10. Histograms of detected frequencies in the light curves of HD 50870. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

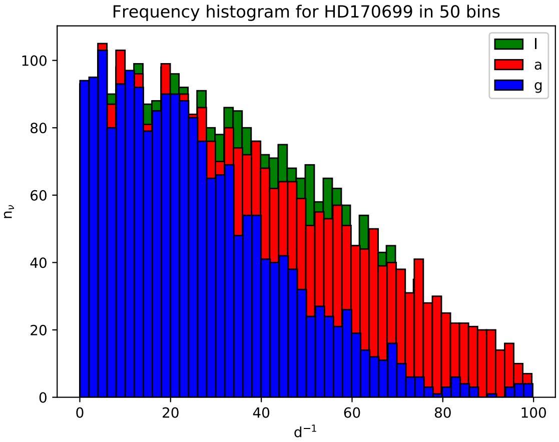

Fig. D.11. Histograms of detected frequencies in the light curves of HD 170699. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

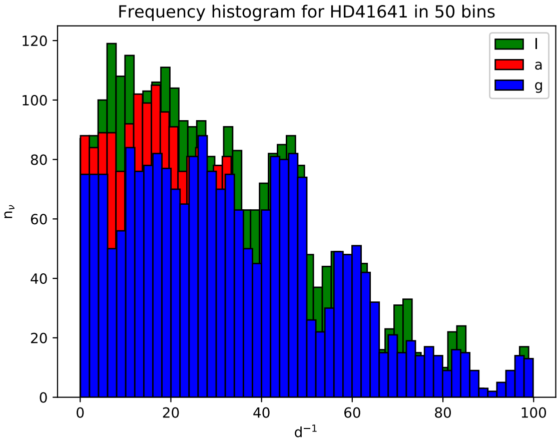

Fig. D.13. Histograms of detected frequencies in the light curves of HD 41641. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

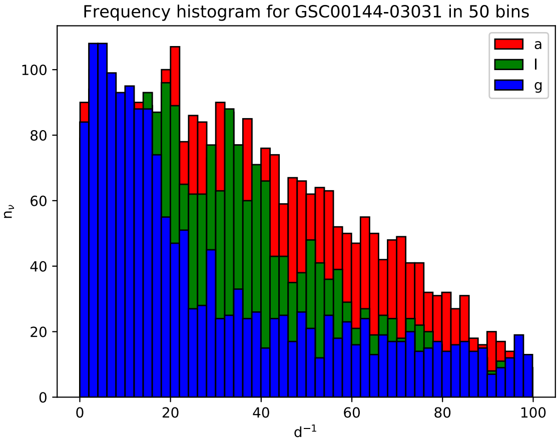

Fig. D.12. Histograms of detected frequencies in the light curves of GSC00144-03031. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.

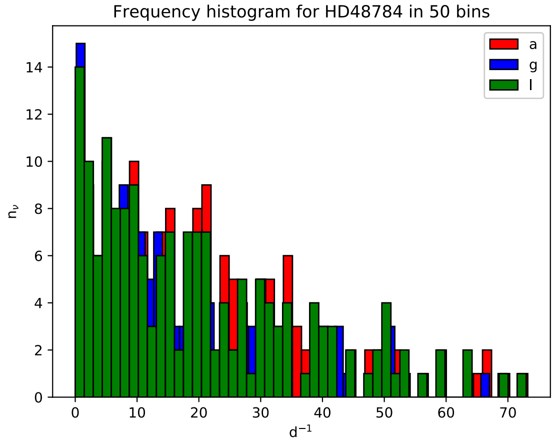

Fig. D.14. Histograms of detected frequencies in the light curves of HD 48784. Blue bars correspond to gapped data, red bars to ARMA interpolated data, and green bars to linearly interpolated data.