Determining global parameters of the oscillations of solar-like stars

S. Mathur[1,2], R. A. García[1,3], C. Régulo[4,5], O. L. Creevey[4,5], J. Ballot[6], D. Salabert[4,5], T. Arentoft[7], P.-O. Quirion[7], W. J. Chaplin[8], and H. Kjeldsen[7]

[1] Laboratoire AIM, CEA/DSM – CNRS - Université Paris Diderot – IRFU/SAp, 91191 Gif-sur-Yvette Cedex, France
[2] Indian Institute of Astrophysics, Koramangala, Bangalore 560034, India
[3] GEPI, Observatoire de Paris, CNRS, Université Paris Diderot; 5 place Jules Janssen, 92190 Meudon, France
[4] Universidad de La Laguna, Dpto de Astrofísica, 38206, Tenerife, Spain
[5] Instituto de Astrofísica de Canarias, 38205, La Laguna, Tenerife, Spain
[6] Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, CNRS, F-31400, Toulouse, France
[7] Danish AsteroSeismology Centre (DASC), Department of Physics and Astronomy, Aarhus University, 8000 Aarhus C, Denmark
[8] School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

ABSTRACT

Context. Helioseismology has enabled us to better understand the solar interior, while also allowing us to better constrain solar models. But now is a tremendous epoch for asteroseismology as space missions dedicated to studying stellar oscillations have been launched within the last years (MOST and CoRoT). CoRoT has already proved valuable results for many types of stars, while Kepler, which was launched in March 2009, will provide us with a huge number of seismic data very soon. This is an opportunity to better constrain stellar models and to finally understand stellar structure and evolution.
Aims. The goal of this research work is to estimate the global parameters of any solar-like oscillating target in an automatic manner. We want to determine the global parameters of the acoustic modes (large separation, range of excited pressure modes, maximum amplitude, and its corresponding frequency), retrieve the surface rotation period of the star and use these results to estimate the global parameters of the star (radius and mass).
Methods. To prepare for the arrival and the analysis of hundreds of solar-like oscillating stars, we have developed a robust and automatic pipeline, which was partially adapted from helioseismic methods. The pipeline consists of data analysis techniques, such as Fast Fourier Transform, wavelets, autocorrelation, as well as the application of minimisation algorithms for stellar-modelling.
Results. We apply our pipeline to some simulated lightcurves from the asteroFLAG team and the Aarhus-asteroFLAG simulator, and obtain results that are consistent with the input data to the simulations. Our strategy gives correct results for stars with magnitudes below 11 with only a few 10% of bad determinations among the reliable results. We then apply the pipeline to the Sun and three CoRoT targets. In particular we determine the large separation and radius of the Sun, HD49933, HD181906, and HD181420.

Key words. Methods: data analysis – Stars: oscillations

1. Introduction

During the past 30 years, helioseismology has proven its ability to study the structure and dynamics of the solar interior in a stratified way. These seismic tools allow us to infer some physical quantities as a function of the radius: the sound-speed (e.g. Basu et al. 1997), the density (e.g. Turck-Chièze et al. 2001), the rotation profile in the convective zone (e.g. Thompson et al. 1996), or in the radiative zone (e.g. García et al. 2008) are some well known examples. But helioseismology has also provided other quantities such as the position of the base of the convection zone (Christensen-Dalsgaard et al. 1985; Ballot et al. 2004), and the photospheric Helium abundances (Vorontsov et al. 1991; Basu & Antia 1995). These observational constraints improved the standard solar models significantly. For instance, with the solar sound speed (e. g. Mathur et al. 2007, and references therein) the agreement between theory and seismic observations is better than a few parts per hundred (using the recent revision of solar abundances from Asplund et al. 2005) or even a few parts per thousand (with the old solar abundances of Grevesse et al. 1993).

The Sun is a fundamental calibrator of stellar evolution (e.g. Christensen-Dalsgaard 2009), but observations of many other stars – covering the HR diagram through asteroseismology – will allow testing stellar evolution and dynamo theories under many different conditions (Chaplin et al. 2008b). However, compared to the Sun, less information will be available from asteroseismology as only low-degree modes (those with a small number of nodal lines at the surface of the star) will be accessible due to the absence of spatial resolution in the observations. This reduces the accuracy and spatial resolution of the inversions. On the other hand, some stars offer the possibility of observing gravity modes, thus giving access to the structure and dynamics of their radiative interiors (e. g. Mathur et al. 2008a).

Nowadays, we are approaching a crucial period of asteroseismic observations. Several campaigns of ground-based observations have been organised (e. g. Bedding & Kjeldsen 2003; Arentoft et al. 2008). Besides, several space missions have already been observing – for more than 3 years – solar-like oscillating stars (whose acoustic oscillations are excited stochastically by convection): the American satellite WIRE (Wide-Field Infrared Explorer, Bruntt et al. 2005), the Canadian satellite MOST (Microvariability and Oscillations of STars, Walker et al. 2003), and the French-European-Brazilian mission CoRoT

(Convection, Rotation and planetary Transits, see e.g. Michel et al. 2008). With the nominal observation cycle of CoRoT (changing the pointing field every six months), only a very few solar-like oscillating stars are observed each year (see for example: Appourchaux et al. 2008; García et al. 2009; Barban et al. 2009), allowing us to analyse each individual target very carefully. The launch of Kepler (Borucki et al. 2009) on March 7, 2009, heralds a new era for asteroseismology, in that it will increase the number of solar-like stars observed for asteroseismology by two orders of magnitude. Kepler, which will search for habitable exoplanets, also has a wide asteroseismic programme, the Kepler Asteroseismology Investigation (KAI, Christensen-Dalsgaard et al. 2008). Because it will observe a large number of stars for asteroseismology, Kepler presents many challenges for the efficient analysis of the data. The nominal Kepler lifetime is expected to be 3.5 years allowing very accurate asteroseismic studies of the same field all or the entire mission and the possibility of measuring stellar cycles (Metcalfe et al. 2007; Karoff et al. 2009).

Concerning the asteroseismic part, a survey phase will take place over the first 10 months of Kepler nominal operations (e. g. Chaplin et al. 2010) The main goal of this phase is to select the best stars that will be targeted during the rest of the nominal mission for, at least, two and a half years. Each star will be observed for one month at a time, putting constraints on the way how to extract the seismic parameters. For brighter targets we will be able to get individual frequencies, but for the majority of stars (fainter ones) this will not be the case and one must rely on extracting mean global parameters (see for example the analysis of a 26 days of CoRoT observations on a solar-like target by Mosser et al. 2009).

About a thousand stars are going to be observed during the survey phase at a one-minute cadence. Consequently, people working in the Kepler Asteroseismic Science Consortium (KASC) will have to analyse several hundreds of stars every month. This is the reason why we need to change our single-star analysis approach to an automatized pipeline that will provide global-seismic parameters as well as other first estimations of structural parameters like their radius or mass. Some teams have developed their own pipeline (Hekker et al. 2009a; Huber et al. 2009) within the AsteroFLAG group to be ready to analyse the Kepler data.

The objective of this paper is to describe the pipeline we have developed to achieve this goal. The global parameters of the oscillations that are some of the output of our pipeline can be used by modelers to constrain their stellar models (e. g. Metcalfe et al. 2009; Stello et al. 2009b)

Sect. 2 describes the goals and methods of the different packages of our pipeline dedicated to the determination of the p-mode global parameters. Then in Sect. 3, we explain how we retrieve the radius of the stars from these parameters. To illustrate that our pipeline works properly as well as its limits, we applied it to some simulated data by asteroFLAG team (Chaplin et al. 2008a) and by a combined Aarhus/asteroFLAG simulator, to the Sun, and to three of the CoRoT targets in Sect. 4. Finally, we discuss the results of our pipeline and the strategy adopted.

2. Pipeline package description

The automatic pipeline that we have developed is called *A2Z*. It is divided into three main parts (see Fig. 1 for a detailed view of the workflow diagram). The first one is dedicated to the study of the global parameters of solar-like oscillating stars finishing by computing a first estimation of the mass and the radius of

the star deduced from empirical scaling laws. The second part corresponds to the extraction of p-mode properties through the fitting of individual modes. The third part uses a pre-calculated grid of evolutionary models where we give as input previously determined parameters such as the large separation and the frequency range of the p-mode excess to obtain a better estimation of the radius and mass.
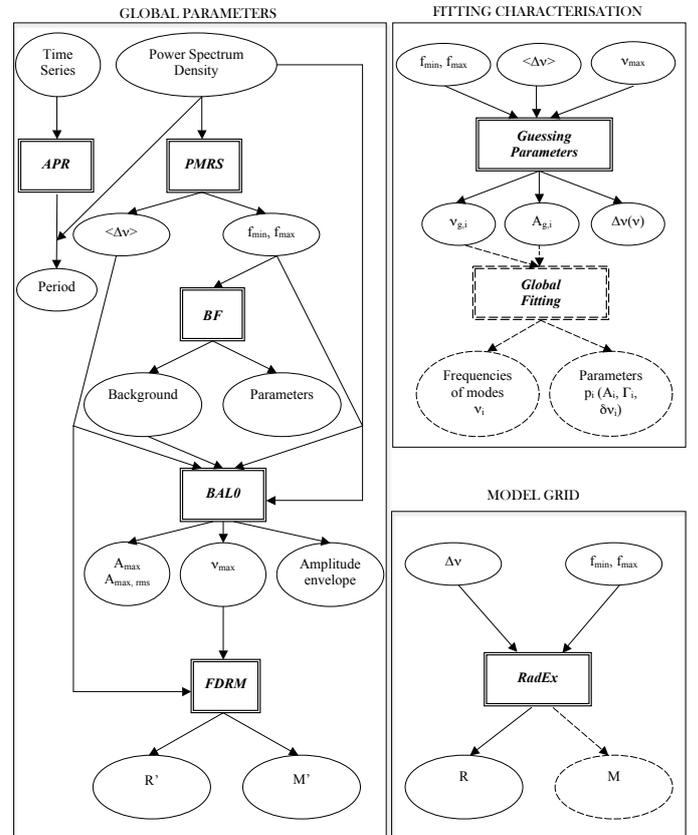

Fig. 1. Workflow diagram of the *A2Z* pipeline to extract the global parameters of stellar oscillations and the characteristics of the modes and to determine of the mass and radius through the comparison with a pre-calculated grid of models. The rectangles represent the code packages and the input/output are in the circles (each output has its corresponding error bar). The dashed lines correspond to the packages that are not described in this paper.

The starting point of our *A2Z* pipeline is either the time series or the power spectrum density (PSD) expressed in $ppm^2/\mu Hz$. A Lomb-Scargle algorithm is used depending on the regularity of the sampling rate of the time series.

The package APR (Sect. 2.1, average photospheric rotation) gives an estimation of the surface rotation period of the star analysing either the low-frequency part of the PSD or looking for the time-frequency signature of this periodicity using wavelets.

PMRS (Sect. 2.2, p-mode range search), uses two different methods to estimate the mean large separation ($\langle \Delta \nu \rangle$) and gives the range of the p-mode excess power (f_{min} and f_{max}). These latter values are then used to fit the background in the PSD with the package BF (Sect. 2.3, background fitting). Then, the third package, BAL0 (Sect. 2.4, bolometric amplitude per $\ell=0$), calculates the amplitude envelope and gives the maximum bolometric amplitude per radial mode (A_{max}) as well as the corresponding fre-

quency ($\nu_{\max}$). Finally, with these asteroseismic parameters, we can estimate the radius (R') and the mass (M') of the stars using empirical scaling laws with the package FDRM (Sect. 2.5, First Determination of the Radius and the Mass).

For the package dedicated to the fitting of the modes, we need a table with the guesses of the mode frequencies. The package Guessing Parameters (Sect. 2.6) uses the global parameters of the modes obtained previously to calculate a table of the guesses for the frequencies ($\nu_{g,i}$) and the amplitudes ($A_{g,i}$) (where g is used for the guesses and i is the mode index) as well as a first estimation of the variation of the large separation with frequency ($\Delta\nu(\nu)$).

Another package RaDex (Sect. 3, radius extractor) based on a grid of stellar models also estimates the radius of the stars using $\langle\Delta\nu\rangle$, $f_{\min}$, and $f_{\max}$. All the outputs of the different boxes are calculated with their uncertainties, ϵ.

If the signal-to-noise ratio (SNR) is high enough, the global parameters and the guesses are used by the maximum-likelihood Global Fitting code to infer the characteristics of each mode: the central frequency ν_i, the amplitude A_i, the width Γ_i, and the rotational splittings $\delta\nu_i$. This global fitting code will be used only on a few stars during the survey phase of Kepler. However, the characterisation of individual p modes is out of the scope of this paper. It has already been tested on the Sun a large number of times and applied to several CoRoT targets (e. g. Appourchaux et al. 2008; Barban et al. 2009; García et al. 2009).

In the following sections, we describe the different packages and apply them to solar-like stars simulated under typical asteroseismic conditions. In some packages, we use different methods to cross check the results.

2.1. Average Photospheric Rotation

One of the first things that we want to study is the surface rotation period of the star. Thanks to seismology, we can independently determined the inclination from the relative amplitude of mode components inside multiplets (Gizon & Solanki 2003). Unfortunately, the seismic estimates of the rotational splitting and of the inclination i –simultaneously obtained by fitting observed oscillation spectra– are strongly correlated for many of solar-like stars due to their short mode lifetime (for more details and discussion, see Ballot et al. 2006, 2008). Thus, an independent determination of one of these two parameters (rotation or inclination angle) could help during the fitting procedure of the p-mode parameters to establish, for example, some a priori constraints (e.g. Benomar 2008; Benomar et al. 2009). However, it is important to remember that the rotational splitting of the mode is an average value of rotation inside the cavity in which the mode propagates and can be different from the surface rotation if there is strong differential rotation in the interior of the star.

To infer the rotation period of the stellar surface several methods can be used: studying the PSD or using the wavelets. The classical method consists in analysing the PSD at very low frequency. We look at the highest peaks below 5 μHz (a value that can be changed to take into account more rapid rotators). However, with this method it is difficult to distinguish between the fundamental peak of the stellar surface rotation period and its first harmonics.

The other method that we use in this work consists of applying a time-frequency analysis using the wavelets (Torrence & Compo 1998). Before applying the wavelet tools, we reduce the number of points by rebinning the data by a boxcar function to speed up the calculation. We also filter the data to have only the low-frequency region. Then we use the Morlet wavelet, which is

the product of a Gaussian and a sinusoid, thus a finite pattern and change the scale of the wavelet, i.e. its frequency (or period). By sliding it along the time series we obtain the wavelet power spectrum (WPS), which is the result of the correlation between the wavelet and the data. This method has the advantage of lifting the uncertainty on the rotation period (see Sect. 4.2). This tool has also been extensively used to study solar and stellar activity (Mathur et al. 2008b). Besides, a very conservative approach has been followed by defining the cone of influence (COI) to determine the limit of a reliable result: a periodicity should be seen at least four times to have more confidence on the value.

An example of the WPS is plotted on Fig. 13 using solar data. We calculate the sum of the WPS along time for each period (or frequency) of the wavelet. The result, which is the global wavelet power spectrum (GWPS), is represented as a function of the period (or frequency of the wavelet) (see Fig.13, right panel). We overplotted as well the threshold for a confidence level of 95% (chosen for our study). Any peak that is above this threshold has more than 95% probability not to be due to noise. The surface rotation period of the star is easily identified with this confidence level.

2.2. P-Mode Range Search and large separation estimation

We study the stellar oscillations spectrum by searching for the frequency range where the p modes present most of their power in the PSD. This search begins with the estimation of the mean large separation, $\langle\Delta\nu\rangle$.

It is important to have two independent methods to obtain the spacing, to cross-check the results, and to be able to have reliable results.

2.2.1. Method 1

As the modes are equally spaced in frequency in the PSD, we compute the fast Fourier transform of a slice between 100 and 10000 μHz in the PSD, (called hereafter PS2) and look for the highest peak (see Fig. 2). The large separation is defined as:

$$\Delta\nu = \nu_{n+1,\ell} - \nu_{n,\ell}, \tag{1}$$

where ν is the frequency, ℓ the degree, and n the radial order of a mode. Between two consecutive orders of a given degree ℓ of a mode, we have the $\ell+1$ mode, which is approximately positioned at a distance $\Delta\nu/2$; therefore the periodicity of $\Delta\nu/2$ occurs more often in the PSD. Thus, we assume that the highest peak in the PS2 will correspond to half of the large separation.

Fig. 2 shows an example of a PS2 for a simulated star. We find the highest peak at 34.5 μHz, which gives $\langle\Delta\nu\rangle = 69$ μHz. The signal-to-noise ratio of the amplitude of this peak is 6.22 σ. Then, in the PSD, we take a box of 600 μHz starting at a frequency of 100 μHz, calculate the PS2 normalised by σ, where σ is the standard deviation of the PS2 and we look for the peak around $\langle\Delta\nu\rangle/2$. For this search, we look for the highest peak in the PS2 in the range $\langle\Delta\nu\rangle/2 \pm 10$ μHz. We repeat this by shifting the box in the PSD by 60 μHz until 10000 μHz. Then, we plot the power normalised by σ of the PS2 of the highest peak (or maximum relative power) found in this range as a function of the central frequency of each box.

An example of this kind of plot is in Fig. 3. The horizontal error bars correspond to the size of the box of 600 μHz. To put some statistical constraints on this search, we calculate the threshold, s, corresponding to the probability of finding a peak

in the PS2 using the formula (Chaplin et al. 2002; Gabriel et al. 2002):

$$s = \ln(\Delta T) - \ln P, \qquad (2)$$

where Δ is the size of the box, T is the length of the time-series, and P is the probability. In Fig. 3, we have plotted the threshold for a probability of 95%. So we can say that the frequency range where most of the p-mode power is found goes from 937 μHz to 2474 μHz, which corresponds to the boxes where the relative power of the peak is above the given threshold. Finally, we calculate the mean value of the $\langle \Delta \nu \rangle$ found in each independent box between f_{min} and f_{max} as well as the dispersion giving us the uncertainty on our determination of $\langle \Delta \nu \rangle$.
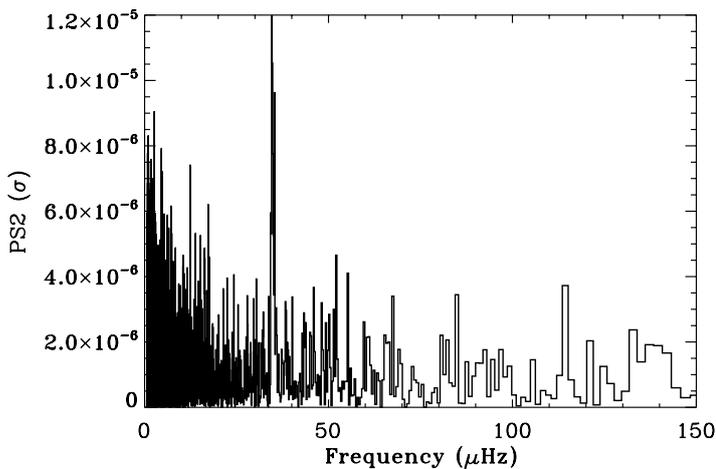

Fig. 3. Typical maximum relative power of the highest peak in the PS2 around half of the large separation, as a function of the central frequency of the sliding box taken in the PSD for one simulated star (Pancho). The dashed line represents the threshold corresponding to a 95% confidence level. The error bars represent the size of the boxes taken in the PSD (600 μHz in this case). Thus, the range of the p-mode power excess is [940, 2440] μHz.

Fig. 2. Power spectrum of the power spectrum (PS2) between 100 and 10000 μHz as a function of the frequency for one simulated star (Pancho) and normalised by σ of the PS2.

But as we can have very low SNR stars, we allow the threshold to go down to as low as a probability of 70% if the higher probabilities are not found.

When we do not find signal above 70% of confidence level, we change the size of the box of 600 μHz and try two different values, 900 and 300 μHz, and start the search again. We use a flag (*box*) to record the size of the box.

We have found in some simulated stars that two dissociated regions can be above the threshold, thus two bumps: the correct one and another one due to noise, which gives us a very wide range for the p-mode region. Thus, to determine the correct region, we use again a box of 900 or 300 μHz and if we have only one bump in the new analysis, we select the common bump of both analyses. By doing so, we have improved a few results. But if we still have two bumps, we use a flag (*dblpk*) to inform us about the presence of this double bump.

With the simulated data, we have noticed that the highest peak was not always the one corresponding to half of the large separation. So the first estimation of $\langle \Delta \nu \rangle$ might not be correct and might be due to noise. That is the reason why, if we do not find any signal above 70% of confidence level in the Maximum relative power, we look for the second highest peak and the third highest peak in the PS2 and do the whole search again.
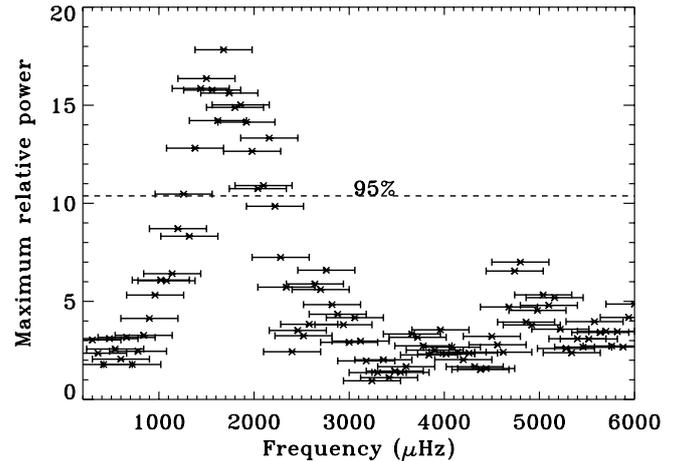
We also have a first estimation of ν_{max}, the frequency of maximum amplitude, by taking the frequency of the highest Maximum relative power.

Once we have determined $\langle \Delta \nu \rangle$ and we have an estimation of ν_{max}, we verify our results by using scaling laws (Stello et al. 2009a), which establish the following relation:

$$\langle \Delta \nu \rangle = 135 \times \left(\frac{\nu_{max}}{3050} \right)^{0.8}. \qquad (3)$$

If the $\langle \Delta \nu \rangle$ found with our method does not fall within an uncertainty of 20% around the theoretical value, we use a flag.

As an alternative, we can run this code a second time with an estimation of $\langle \Delta \nu \rangle$ as an input and by constraining our search for a peak in the PS2 around this theoretical value $\pm 15 \, \mu$Hz.

The output of this part of the pipeline is the value of the large separation and the frequency range of the p-mode bump with a given confidence level.

2.2.2. Method 2

The second method to look for the large separation is based on the estimation of some parameters with spectroscopic observations. From the fundamental stellar parameters: parallax, effective temperature, apparent magnitude, and log g or as in the Kepler stars: effective temperature, log g, and radius, given in the KIC (Kepler Input Catalog, Latham et al. 2005), we use the scaling laws from Kjeldsen & Bedding (1995) to estimate the range in frequency for the p-mode excess power and the large separation. With these initial values, we look for $\langle \Delta \nu \rangle$ in the power spectrum (PS) of the data as well as in some average spectra by taking several subseries in the data. We can choose the number of average spectra.

The search is done by computing the PS of a short slice (typically 900 μHz) of the PS where the p-mode excess power ap-

pears (Régulo & Roca Cortés 2002). The search for the spacing is done iteratively by trying different values in a 50 μHz-range around the estimated spacing and with a step of 0.01 μHz. We try to find if there is any signal above 1.5 times the rms of the power spectrum at any of the bins spaced by $\langle \Delta \nu \rangle$. To evaluate the significance of the peaks found and to avoid binning effects, this procedure is repeated 50 times and by continuously shortening the length of the 900 μHz-slice, down to $1\times\langle \Delta \nu \rangle$. The coincidences of spacing among the peaks found in the 50 trials are then registered (See Fig. 4).
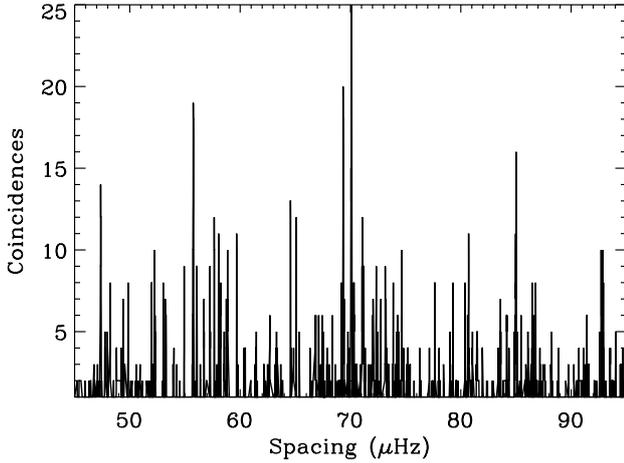

Fig. 4. Result for the search for the spacing in one month of data for one of the simulated stars (Pancho). The highest peak corresponds to the large separation present in the data.

From this search we obtain for the whole spectrum as well as for the averaged spectra a possible value of $\langle \Delta \nu \rangle$. We also fit the histogram of coincidences with a Gaussian profile. The width of the latter gives the error bar (σ) associated to the value obtained for the spacing. Then, we take the mean value of all the $\langle \Delta \nu \rangle$ found in each data series, typically 5 for a month of data. The values that are 3 σ above the mean value are rejected and a new mean value with its error is estimated. This process is repeated 3 times and the result with its error is taken as the large separation for the analysed data set. However, if the data have a poor SNR or if the fundamental stellar parameters are badly determined, it can lead to the wrong spacing, and it is very difficult to identify this result as incorrect.

In the possible case for which the fundamental stellar parameters are ill-determined, we have the possibility to run all the pipeline a second time using as an estimation for the p-mode frequency range and for the large separation, the results obtained blindly by Method 1 described above.

2.2.3. Cross-checking the two methods

To cross-check the large separations obtained with the two methods, we follow three steps:

First step: We run both methods, Method 1 based on the blind estimation of the initial parameters, and Method 2, with the initial parameters obtained from the scaling laws. We obtain some

results that are compatible between both methods and some that are different.

To cover the possibility to have selected a wrong region for the p-mode range in the blind way of finding it in Method 1 or the possibility to have a bad determination of the fundamental parameters in Method 2, we implement two more steps.

Second step: Method 2 is run using the blind estimations from Method 1 and the new results of Method 2 are compared with the results from Method 1.

Third step: Method 1 is run using the scaling law estimations from Method 2 and the new results of Method 1 are compared with the results from Method 2.

As explained in Sect. 2.2.1, we calculate the threshold for detecting a peak in the PS2 for a given confidence level. We say that we are very confident when we are above a threshold of 85% (confidence level of 2), less confident between 70 and 85% (confidence level of 1), and not confident at all below 70% (confidence level of 0).

To select the most reliable results, we dismiss the step in which we obtain the highest number of common stars with a confidence level of 0. Among the 2 remaining steps in which we remove the stars with the confidence level of 0, we choose the step where we have the highest number of coincidences.

As a fourth step, to try to recover the spacing of more stars, we look at the common stars in the two comparisons in which we have obtained the lower number of stars with zero confidence level. For example if among 300 stars we have 170 stars with common results with step 1 (with a confidence level of 1 and 2) and in step 2 there are 195 of them, we check if the first 170 common stars from step 1 are included in the 195 from step 2. If only 150 stars are the same between step 1 and 2, we accept the other 20 as correct results and we add them to the 195 obtained from step 2. Finally, we will have 215 common stars.

2.3. Background Fitting

The background of the spectrum is modeled with three components:

1. white noise to model the photon noise (W);
2. Harvey's model which reproduces the convective contribution to the background (Harvey 1985);
3. a power law modelling activity effects and low-frequency trends.

Our model has then 6 free parameters:

$$B(\nu) = W + \frac{A}{1 + (2\nu/\nu_c)^{\alpha}} + a\nu^{-b}. \tag{4}$$

All of the parameters have to be positive, which is a constrain in the fitting code. The model is fitted to the spectrum over the domain $[\nu_1, f_{\min}] \cup [f_{\max}, \nu_{\mathrm{cut}}]$ using a classical maximum-likelihood estimator.

The parameters $f_{\min}$ and $f_{\max}$ are the bounds of the interval where p-mode power excess has been detected and returned as an output of the package PMRS. ν_{cut} is the cut-off frequency and ν_1 the lowest considered frequency (in practice, we consider $\nu_1 = 1 \mu$Hz). In other words, we simultaneously fit the full spectrum after having excluded the p-mode region.

We use as an initial guess for W the average of the power spectrum around the cut-off frequency, where the photon noise is expected to dominate the other components. To find guesses for a and b, we perform a log-log regression in the low-frequency domain. The only input guess requested by this package is ν_c.

The value of the spectrum around the frequency ν_c is then used to extrapolate a guess for A. We have typically used values in the range 100–1000 μHz for the guess of ν_c. By using this set of guesses, a first fit is performed where α is kept fixed with a value of 2. The fitted values are then used as guesses for a second fit where all of the parameters have been let free.

We have noticed with our benchmark of tests that this routine is robust and stable. The final results depend indeed only weakly on the exact value used for the input guess of ν_c.

2.4. Characterisation of the P-Mode Envelope (Bolometric Amplitude per ℓ=0)

Once we have fit the background with the package BF, we subtract it from the PSD as we need to estimate the amplitude of the modes without the contribution of the stellar noise. Then, we smooth the new PSD using a sliding window of $q \times \langle\Delta\nu\rangle$, where q=1, 2 or 3. We obtain the envelope of the power spectrum. We fit the result with a Gaussian function, which gives the maximum power, $P_{\max}$ in ppm$^2\mu$Hz^{-1} and the frequency position, $\nu_{\max}$. We also recover the error estimate from this fitting, which gives us the standard deviation related to both coefficient, $P_{\max}$ and $\nu_{\max}$. To convert this power into amplitude, $Ampl_{max}$, we use the following formula as we have a single-sided PSD:

$$Ampl_{\max} = \sqrt{2 \times P_{\max} \times q \times \langle\Delta\nu\rangle}. \qquad (5)$$

Then, we want to calculate the rms maximum amplitude per radial mode or the bolometric amplitude, $A_{\max}$. As we use the method described in Kjeldsen et al. (2008), the formula for the amplitude is the following:

$$A_{\max} = Ampl_{\max}/\sqrt{2 \times q \times c} = \sqrt{(P_{\max} \times \langle\Delta\nu\rangle)/c}, \qquad (6)$$

where c is related to the spatial response to the observations of the modes of degree $\ell = 0$ to 4 and to the type of observations (intensity or velocity). For the simulated stars, we took c=3.16 as given in Kjeldsen et al. (2008). However, the correction to obtain the bolometric amplitude per radial mode depends both on the instrument and the star ($T_{\rm eff}$), thus it should be calculated for each star and for a given instrument. This has been done by Michel et al. (2009) for the CoRoT mission and should be done for the Kepler mission as well. Finally, we can plot the bolometric amplitude per radial mode in the frequency range found in the package PRMS, which is shown in Fig. 5.

From now on, when we talk about the maximum amplitude of the mode, we will refer to the rms maximum amplitude.

2.5. First Determination of the Mass and the Radius

We assume here that we do not have any information on the radius and the mass of the star beforehand, which is usually the case. Instead of using the scaling laws (Kjeldsen & Bedding 1995; Bedding & Kjeldsen 2003) to estimate $\langle\Delta\nu\rangle$ and $\nu_{\max}$, we use them as input. We solve the two equations with two unknown variables: the radius of the star R' and its mass M'. We can calculate them according to formulae (9) and (10) from (Kjeldsen & Bedding 1995).

$$\frac{R'}{R_\odot} = \left(\frac{135}{\langle\Delta\nu\rangle}\right)^2 \left(\frac{\nu_{\max}}{3050}\right) \left(\frac{T_{\rm eff}}{5777}\right)^{1/2} \qquad (7)$$

$$\frac{M'}{M_\odot} = \left(\frac{135}{\langle\Delta\nu\rangle}\right)^4 \left(\frac{\nu_{\max}}{3050}\right)^3 \left(\frac{T_{\rm eff}}{5777}\right)^{3/2}. \qquad (8)$$
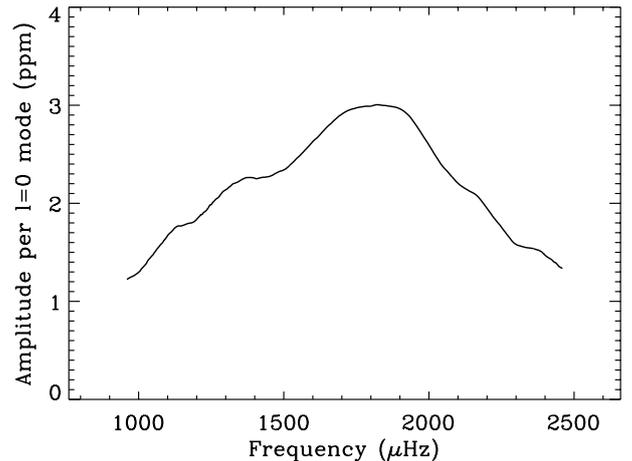

Fig. 5. Typical bolometric amplitude of the mode $\ell = 0$ as a function of the frequency for one simulated star.

Concerning the uncertainty on these values, we have to take into account the error bars on $\langle\Delta\nu\rangle$, $\nu_{\max}$. This leads us to uncertainties of about 10% and 20% respectively for the radius and the mass. These results depend on the validity of the scaling laws. It has been recently shown, thanks to the CoRoT mission, that they are closely followed for red giants (Hekker et al. 2009b; de Ridder et al. 2009) and one of the first goals of the asteroseismic investigation of the Kepler mission will be to check their validity.

This package gives a first estimation of the mass and radius while the package described in the next section to estimate the radius is more accurate but more time consuming.

2.6. Guessing Parameters and variation of the large separation with frequency

In this section, we describe how the guess-parameter table is calculated, which is needed to perform the peak fitting of the individual modes. The guess frequencies thus obtained provide a first estimation of the frequency of each (ℓ, n) mode and the large-separation variation with frequency.

The power spectrum is corrected from the background obtained by the package BF and smoothed by a factor of 2% of the estimated mean large separation $\langle\Delta\nu\rangle$ (Sect. 2.2). The guess table is constructed over the frequency range $[f_{\min}, f_{\max}]$ where p-mode power excess was found (see Sect. 2.2). The frequency bin of the highest peak in the power spectrum around the estimated $\nu_{\max}$ (Sect. 2.4) is found. Then, the consecutive mode frequencies $\nu_{g,i\pm1}$ are determined by searching for the highest peaks within frequency windows of $(\nu_{g,i} \pm \Delta\nu_i) \pm 0.15\Delta\nu_i$, where the large separation $\Delta\nu_i$ is continuously updated based on the previously measured frequencies.

The guesses of the Lorentzian-mode amplitudes are calculated from the maximum power corresponding to each guess frequency.

A first estimate of the frequency dependence of the large separation $\Delta\nu(\nu)$ is thus obtained. An example for 30 days of a simulated asteroFLAG star is shown in Fig. 6.
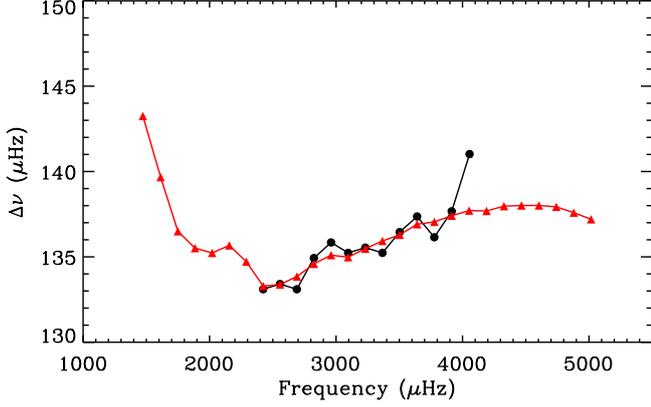
Fig. 6. Frequency dependence of the large separation of the group of even modes $l = 0 - 2$ measured from 30 days of one simulated asteroFLAG star (Boris). The triangles represent the input values, and the circles the results from the pipeline.

Fig. 7. The fitted masses and ages for a simulated star. The filled circles show the masses and ages derived from our fitting procedure starting from eight different initial guess values. The input value is denoted by the square.

3. Radius from a grid of stellar models

Once we have robust estimates of $\langle\Delta\nu\rangle$, $f_{\min}$ and $f_{\max}$ (see Sect. 2.2), we couple this information with the available atmospheric parameters to estimate the radius of the star using stellar models. In the specific case of Kepler data, the atmospheric parameters that we use and that are available to us from the KIC are $\log g$, $[M/H]$, and T_{eff}. To determine the global parameters, we compare the observations ($\langle\Delta\nu\rangle$, $\log g$, $[M/H]$, and T_{eff}) with model observables from stellar evolution and pulsation codes. We use the Aarhus Stellar Evolution Code (ASTEC) coupled to an adiabatic pulsation code (ADIPLS) (Christensen-Dalsgaard 2008b,a). These codes need as input the stellar parameters of mass, age, chemical composition and mixing-length parameter, and return the stellar model interior profiles, global parameters, such as radius and effective temperature, and the frequencies of the oscillation modes.

The parameters that best describe the observables are obtained by minimising a χ^2 function;

$$\chi^2 = \sum_{i=1}^{M} \left(\frac{y_i - B_i}{\epsilon_i}\right)^2, \qquad (9)$$

where y and B are the $i = 1, 2, ..., M$ observations and model observables. The Levenberg-Marquardt algorithm is used for the optimization, and this incorporates derivative information to guess the next set of parameters that will reduce the value of χ^2. Naturally, an initial guess of the parameters is needed and these are obtained from a small grid of stellar evolution tracks (T_{eff}, $\log g$) where we assume $\langle\Delta\nu\rangle$ is proportional to the mean density of the star.

Because there are few observations and just as many parameters, there are inherent correlations between mass, age, and metallicity. To help avoid local minima problems, we minimise the χ^2 function beginning at several initial guesses of the parameters (mainly varying in mass and age), and these initial guesses are estimated from the grids. For each star, we therefore obtain several sets of parameters with a corresponding χ^2 value that match the observations as best as possible. To highlight the correlations we obtain, Fig. 7 shows several values of mass and age that adequately match the observations of a simulated star. The square symbol highlights the true input values of these models.
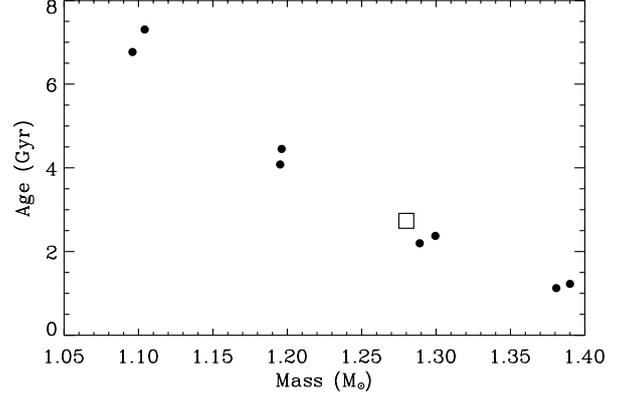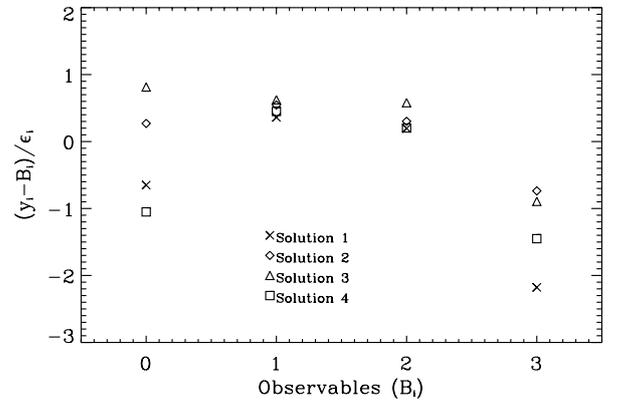

Fig. 8. The residuals between the observations and model observables for four of the models from Fig.7. From left to right, the observables are: T_{eff}, $\log g$, $[M/H]$, and $\Delta\nu$.

Fig. 8 shows the residuals (observation - observable) from 4 of the best-fitting models. It can be understood quickly from these figures, that there are difficulties with extracting both the mass and the age of the stars.

Luckily, the radius parameter is highly dependent on the set of observations. This means that although we obtain quite a scatter in mass and age for each of the simulated stars, we do obtain robust estimates of the radius.

The radius is calculated as the average value from the best-fitted models where the χ^2 value is below 3.9^2. The error is defined as one sixth of the difference between the maximum and minimum radius. The value of six stems from the "error" being defined as half of the scatter in values, and then we divide by 3 to produce an estimate of the $1\,\sigma$ uncertainty, assuming that the range of values spans $3.9\,\sigma$. The choice of dividing by 3 instead of 3.9 is to account for the low number of models that we are using to estimate the paramaters.

4. Results

4.1. Simulated data

In this work, we have used simulated data generated within the AsteroFLAG team for hare-and-hounds exercises (Chaplin et al. 2008a) dedicated to prepare the Kepler mission. We have also applied our pipeline to data from the Aarhus simulator (Stello et al. 2004). These simulated stars are based on the information available in the KIC of the stars that will be first observed by Kepler during the survey phase.

4.1.1. AsteroFLAG hare and hounds

We have decided to work on three stars (or "cats") simulated by the AsteroFLAG team: Arthur, Boris, and Pancho. For each of them, we have 2 different rotation rates (low and high), 3 angles of inclination (0, 30, and 60°), and 4 magnitudes (from 9 to 12). From these 72 simulated stars with different fundamental parameters, 2592 independent spectra of one month have been obtained (36 for each star). These 2592 spectra have been analysed independently with the methods described above for the package PMRS and the results have been compared following our cross-checking methodology (Sect. 2.2.3) for estimating $\langle \Delta \nu \rangle$.

Table 1 shows the statistics of the common results found for all the 2592 spectra according to the type of star. We can see that by exchanging our results in steps 2 and 3, the number of common results increases. For a given step, if we have a high number of common stars with a confidence level of 0, it means that the common results could be incorrect. This is due to the fact that if there is some signal (real or not) and if both methods search around the same region, they should find almost similar results.

4.1.2. Aarhus/asteroFLAG simulator stars

We have also tested our pipeline on 176 stars simulated with the Aarhus/asteroFLAG simulator. These stars have magnitudes from 7 to 11 and they are one-month time-series.

Once again we have analysed the common results for the different steps of our methodology of the PMRS package (Table 1). For step 1, we have 33 results in common with 3 results with level 0. For step 2, we have 75 results in common and 4 level 0. Finally, for step 3, we have 78 common results with 36 level 0.

4.1.3. Strategy to select reliable results

The strategy adopted in this work is to look at the number of common results in each step described previously and among them, to take into account the number of results with a confidence level of 0. We have to make a trade-off to have the highest number of common stars with the lowest number of zero-confidence level stars.

We recap here how we select the best step. First, we remove the step where the number of stars with a confidence level of 0 is the highest. Then, we choose the step with the highest number of common stars and lowest number of stars with 0 confidence level. We select only the results where the confidence level is equal to 1 and 2, so we subtract the number of results with a zero-confidence level. Finally, we add up the stars found in the 2 best steps as all the common results might not be included in the common results of the other step.

By comparing the results to the input, a tolerance of 5% of the real value for $\langle \Delta \nu \rangle$ is accepted. If we are within this error, we say that the result is "correct".

For instance, for the case of Pancho (Table 1), we notice that for steps 1 and 2, the number of confidence level of 0 is quite close: 6 and 9. But the number of common results is higher for step 2, 259 compared to 231 for step 1. So we choose the common results of step 2. When we remove the results with a confidence level of 0, this gives 250 confident results. By applying it to all the stars, we select: 25% of the common results with Arthur, 34% for Boris, 29% for Pancho, and 40% for the Aarhus-asteroFLAG simulated stars. Among them, we know that there are respectively 9%, 2%, 11%, and 6% of incorrect results.

Finally, according to the last step of our methodology, if we add the stars that are in common in step 1 in the case of Pancho and that do not appear in the common stars in step 2, we obtain 293 stars for which we are confident. Thus we would have 29% of common results for Arthur, 38% for Boris, 34 % for Pancho, and 40% for the Aarhus-asteroFLAG simulated stars, with respectively 16.8%, 2.8%, 13.3%, and 6% of incorrect results.

Now, if we look at the results in terms of magnitude (Table 2, first half) we notice that before applying the last step, for a magnitude V=9, we manage to have 87% of common results, with only 0.5% of incorrect results. Then, for V=10, the number of common results decreases to 25.6% with 10.8% of bad determinations. Finally, with higher magnitudes, the results are not convincing as we have more than 50% of wrong determinations. So we can say that our pipeline works well for stars with V<11.

After applying the last step (Table 2, second half), the results are the following: for V=9, we have 91.5% of common results, with only 0.7% of incorrect results, for V=10, we have 33.3% of common results with 14.3% of bad determinations.

For the simulated stars of Aarhus/asteroFLAG, this strategy leads to 71 common results out of 176 (40%), among which 67 results are correct. Applying the last step does not change the result. Fig. 9 represents the distribution of correct results in terms of magnitude, T_{eff}, and log g. We notice that for magnitudes lower than 8, 75% of the stars are retrieved. For $8 \leq V < 9$, we have a good estimation of $\langle \Delta \nu \rangle$ for 63.1% of the stars, for $9 \leq V < 10$, 48.9% of the stars, and for $V \geq 10$, 27.1% of the stars. These results are quite close to what we have obtained with the cats.

We have applied the other packages, BF and BAL0, to these 67 stars and estimated the maximum amplitude per radial mode. Fig. 10 shows the difference between the maximum amplitude estimated (using a smoothing of $3 \times \Delta \nu$) and the input. We clearly notice that for most of the stars, we underestimate the amplitude per radial mode. This can be due to the smoothing of the spectrum and maybe to the background fitting. We notice that the Gaussian fit is most of the time below the p-mode envelope. There is not any correlation between the correct estimation of the amplitude and the magnitude of the star (Fig. 10, left panel), or with the position of the maximum amplitude in frequency (Fig. 10, right panel). For 73% of the stars, we manage to retrieve the maximum amplitude per radial mode within 3 σ. We have also calculated these amplitudes with different values for q. The bias varies between -10.8 ±9% and -14.6 ±8%. For instance, with a smoothing of $1 \times \Delta \nu$, we obtain that 81 % of the results are within 3 σ.

For a few tens of the survey stars (squares in Fig. 9), we have calculated the radius following the method described in Sect. 3. Fig. 9 shows the models that converged with our method (green squares) as well as the models that do not converge (red squares). We selected only a few tens of stars for time and computer lim-
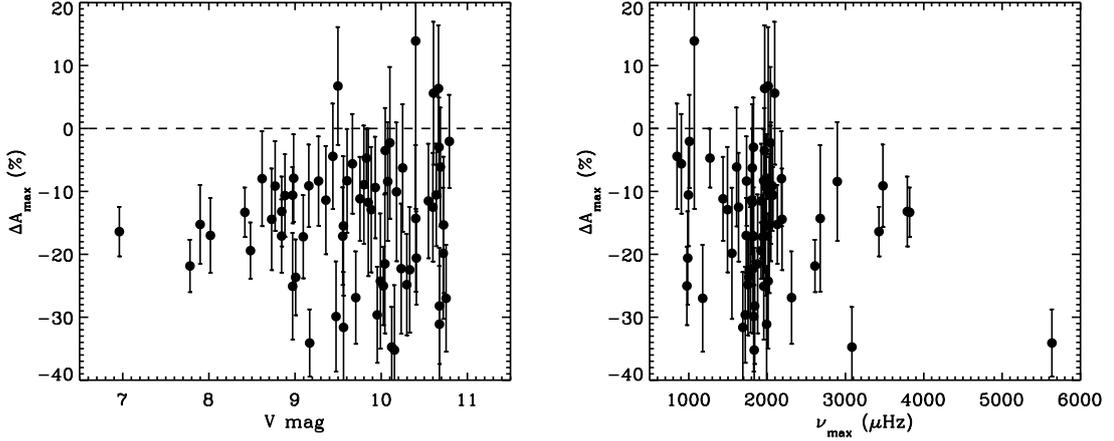
Fig. 10. Relative difference between the maximum amplitude calculated with the *A2Z* pipeline (using q=3) and the input values of the Aarhus-asteroFLAG simulated stars as a function of the magnitude (left panel) and the frequency of the maximum amplitude (right panel). We overplotted the error bars for each estimation.
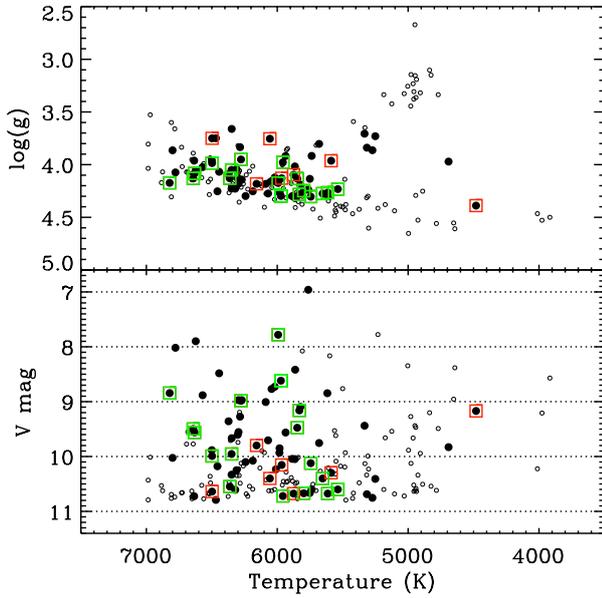

Fig. 9. Distribution of the correct results obtained with the simulated stars from the Aarhus-asteroFLAG simulator in term of V magnitude, $T_{\rm eff}$, and log *g*. The open circles represent the 176 simulated stars. The full black circles represent the stars for which the correct value of the large separation was obtained. The squares are the stars selected to determine the radius: the green squares are the converging models and the red squares are the diverging models.

itations. The selection was also done such as to have a panel of different stars, specially with a wide range for the mean large separation. The results are given in Table 3. We also give the difference between the true input value and the value obtained with the pipeline and among 18 stars, we obtain the radius within 3σ for 13 stars. Fig. 11 shows the input values of the radius for the simulated data versus the fitted values and their errors bars obtained with the pipeline. The continuous line is the x=y line.
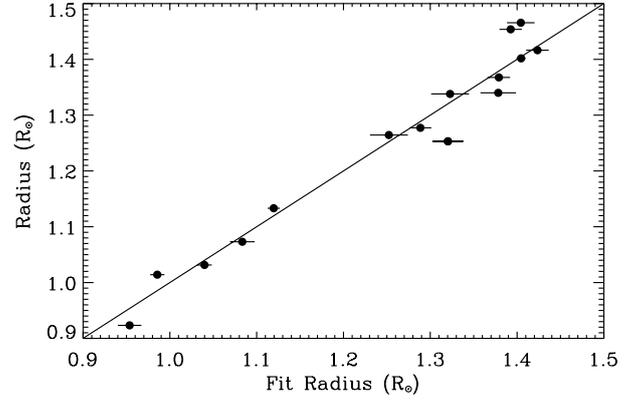

Fig. 11. The true radius versus the fitted input radius for a selection of the simulations. The continuous line is the x=y line.

The figure shows that the determination of the radius from the pipeline is a rather good determination of the true input values within a few%. We note that there are some stars that were not estimated, and the stellar models most likely did not converge due to unphysical sets of parameters as input. Some more work is needed in the code, so that the automatic pipeline will give results for > 95% of the stars. At the moment the success rate is ~70%.

4.2. Solar data

In this section, we have analysed solar data to test our pipeline on some real data. We have estimated the solar surface rotation with the wavelets. We could not use intensity measurements from the VIRGO instrument aboard SoHO as the data available are filtered above 1 and 10 days, which prevents us from measuring the rotation period. Therefore, we have applied the wavelets to velocity data obtained during the first year after the launch of the GOLF (Global Oscillation at Low Frequency) instrument (García et al. 2005). We have rebinned the data to have one point every 2 hours. We have also applied a filter to remove periodic-

ities above 50 days. First we did a simple fast Fourier transform to calculate the PSD (Fig. 12). We notice four high peaks below 1 μHz, at 0.41, 0.8, 1.2, and 1.6 μHz, which can be attributed to the rotation period and a high peak at 0.1 μHz. However, we cannot disentangle the fundamental period.
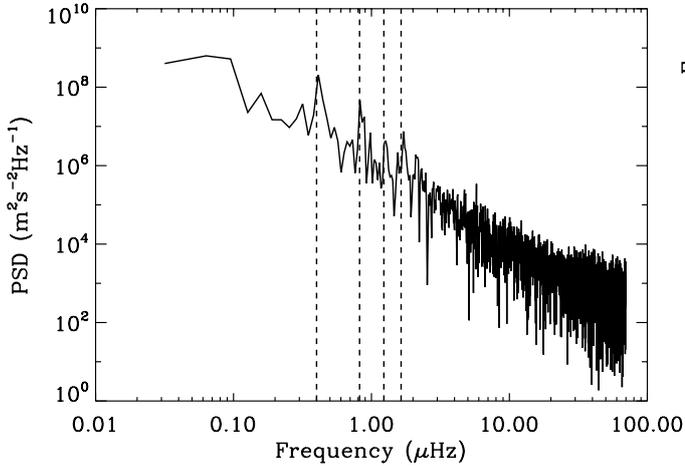

Fig. 12. Low-frequency part of the PSD obtained with one year of solar data from GOLF instrument as a function of the frequency. The dotted lines highlight the highest peaks below 1 μHz that represent the rotation period and the first harmonics.

We have then applied the wavelets to these data (Fig. 13). We notice that there is an accumulation of power along time around 0.4-0.5 μHz. The COI (green grid) shows that a frequency below $\approx$0.2 μHz is not reliable, rejecting thus the peak at 0.1 μHz. The Global Wavelet Power Spectrum, which is a collapsogram over time, shows a maximum at 27 days (0.42 μHz). The width of the peak is related to differential rotation visible at the surface of the Sun. We can also see a peak around 0.8 μHz, but in the WPS there is a power excess only for a short period of time, telling us that this is not the fundamental peak. Thus, this tool is very powerful as it enables us to recover without any ambiguity the real rotation period at the surface of the Sun. This was confirmed using several sets of simulated data as well as solar data obtained with GOLF instrument.

We have used the package PMRS to estimate the mean large separation for the Sun. We have applied it to the VIRGO data, by taking series of 30 days of the blue SpectroPhotoMultiplier, to simulate what we will have with the survey data of Kepler. Using Method 1, we have found $\langle \Delta\nu \rangle$ = 135.46 ±3.3 μHz. As the SNR is quite high, it is obtained with a window of 600 μHz. There is no double bump. This is also obtained with the highest peak in the PS2. Finally, the p-mode excess power is found between 2039 and 4919 μHz (Fig. 14). All these values have a confidence level of 95% and completely agree with the well-known oscillation parameters of the Sun.

Fig. 15 shows the background fit on the PSD. We have subtracted it from the original PSD and we have calculated the maximum amplitude per radial mode with the package BAL0. For the instrumental response of VIRGO, we have used the response function given in Michel et al. (2009). This gives the ampli-
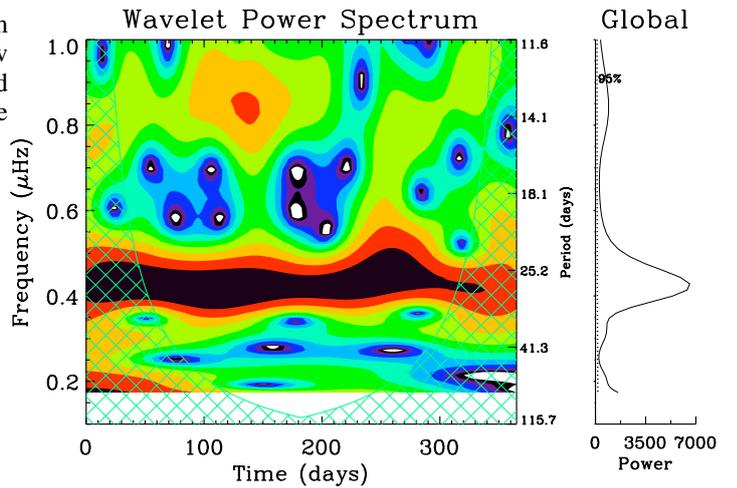

Fig. 13. Left panel: Wavelet power spectrum as a function of time and frequency of the Morlet wavelet. The period is shown on the right axis of the plot. The green grid represents the cone of influence delimiting the reliable periodicity. Black and red colours denote high-amplitude power whereas green and blue colours highlight low-amplitude power. Right panel: Global wavelet power spectrum as a function of the frequency of the wavelet. The dashed line represents the threshold for a 95% confidence level.
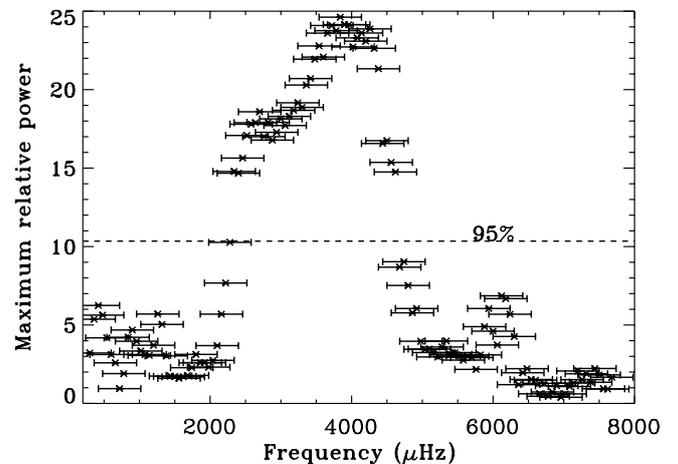

Fig. 14. Maximum relative power for one month of the VIRGO data of the Sun from package PMRS. Same legend as Fig. 2. The excess of power of the p modes is in the range 2039 to 4019 μHz.

tude envelope plotted on Fig. 16 (black curve). Thus, we obtain A_{max}=2.99 ±0.13 ppm and ν_{max}= 3082 ±1.08 μHz (Table 4).

Top panel of Fig. 17 shows the variation of the large separation with frequency. Our results are compatible with those of Frohlich et al. (1997).

The result for the radius of the Sun is given in Tables 5 and 6 and we see that it is fully retrieved within the error bar.
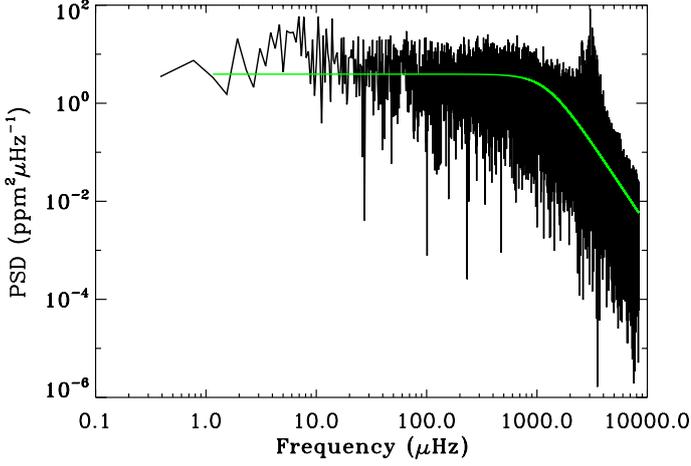
Fig. 15. Background fitting (green curve) of the solar PSD for one month of data from VIRGO.

4.3. Solar-like oscillating stars observed by CoRoT

We have applied our pipeline to three targets of the CoRoT mission: HD49933 (Appourchaux et al. 2008; Benomar et al. 2009), HD181906 (García et al. 2009), and HD181420 (Barban et al. 2009). For HD49933, two runs are available: (1) 60 days and (2) 137 days. For HD181906 and HD181420, the data sets are 156 days long.

The results of the pipeline are summarised in Tables 4, 5, and 6. The large separations and the p-mode range were obtained with the Method 1 of PMRS (Table 4). They all agree with results already obtained in published papers. Fig. 16 shows the bolometric amplitude in the frequency range where the p-mode bump was found for all these stars.
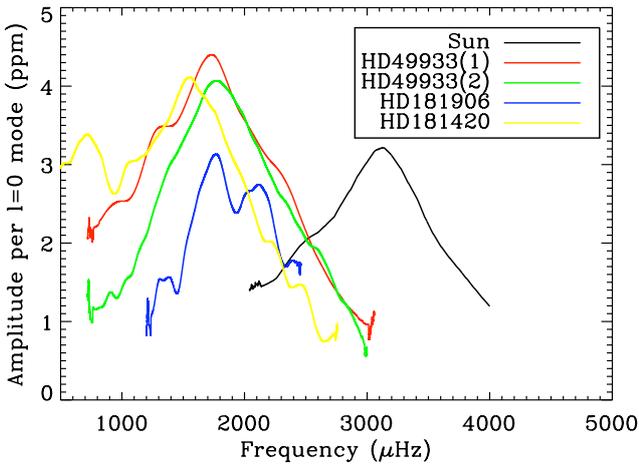

Fig. 16. Smoothed bolometric amplitude per radial mode (rms) for the Sun, HD49933 (run 1 and 2), HD181906, and HD181420.

The frequency dependence of the large separation extracted (Sect. 2.5) is illustrated in Fig. 17. The large separations thus measured were compared with published frequency tables (García et al. 2009; Barban et al. 2009; Benomar et al. 2009). They provide a qualitative good first estimates of their frequency dependence.
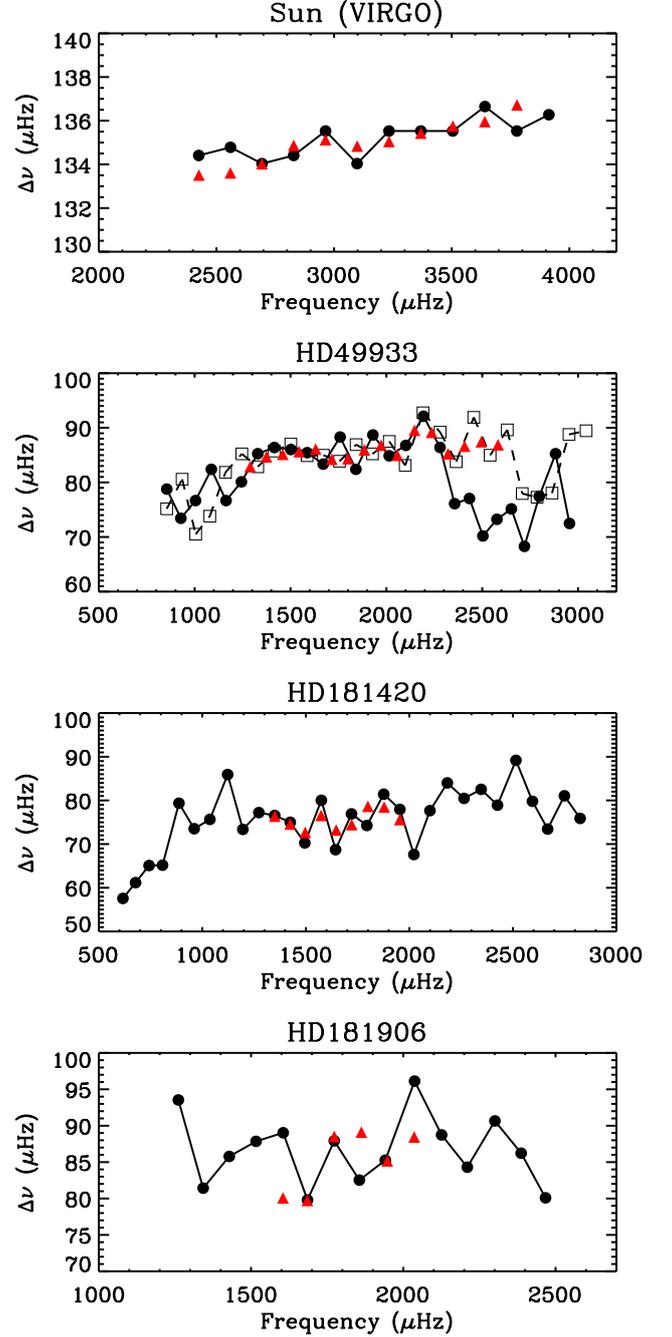

Fig. 17. Frequency dependence of the large separation obtained as in Sect. 2.5 in the case of the Sun (by using 30 days of VIRGO observations) and three COROT stars: HD49933, HD181420, HD181906 (circles). The results based on published frequencies are also shown (triangles). In the case of HD49933, the squares represent the results obtained from the second run of observations. Only the group of even modes ℓ=0-2 are represented.

With the package FDRM, we have estimated the mass and radius of the stars (Table 5). Within the error bars, they are compatible with the data we have with spectroscopic observations. So we can have a rough estimation of the mass and radius if we determine $\langle \Delta \nu \rangle$, and the frequency range of the p modes.

Using the package RadEx, we have estimated the radius of the CoRoT stars (Table 6). For HD49933, we find a radius of R=1.367 ±0.016 $R_\odot$, while from spectroscopic measurements, R=1.34 ±0.016 $R_\odot$. For HD181906, we obtain R=1.415 ±0.016 $R_\odot$, to be compared to R=1.392 ±0.054 $R_\odot$. Finally for HD181420, spectroscopic observations have given R=1.595 ±0.032 $R_\odot$ we obtain R=1.623 ±0.026 $R_\odot$. All of these values are in agreement.

5. Discussion and conclusions

We have described all the different packages of our *A2Z* pipeline to determine the global parameters of stellar oscillations: the mean large separation, the frequency range of the power excess due to the p modes, the background, the maximum bolometric amplitude per radial mode, the guess table of frequencies, and to obtain the radius and the mass of the star.

The first most important step of the pipeline is the estimation of the large separation. We have adopted a 3-step strategy by comparing Method 1 and Method 2 of the package PMRS and by interchanging the results of each of them. We have selected the common results of the step where we obtain the highest number of coincidences with the lowest number of zero confidence level. Applying this methodology to simulated stars has showed that we have obtained from 25 to 40% of common results with errors from 2 to 11%. An optional step consists in adding up the common results that are in the 2 best steps. This has led to 29 to 40% common results with errors ranging from 2.8 to 16.8%.

In term of magnitude of the stars, our pipeline seems to be more robust with stars of magnitude below 11.

It is not yet decided if the last step of our strategy, which adds up the common result from the two best steps, will be used in the future as the number of common results does not increase necessarily and the number of errors increases. It will mainly depend on the the the data that we will have with Kepler.

For the estimation of the mode amplitude, we have managed to estimate 73% within 3 σ of the amplitudes for the Aarhus-asteroFLAG simulations.

Concerning the estimation of the radius of the simulated stars using stellar models, we have succeeded in calculating 70% of the selected simulated stars.

Using the data of the Sun and solar-like targets of the CoRoT mission, we have retrieved the global parameters already published and we have also estimated the radius and the mass of the stars with very good accuracy.

By applying the *A2Z* pipeline to three CoRoT targets, we found for HD49933 $\langle \Delta \nu \rangle$=86.2 μHz, for HD181906, $\langle \Delta \nu \rangle$=86.53 μHz, and for 181420, $\langle \Delta \nu \rangle$=75.35 μHz.

We were able to estimate the maximum amplitude per radial mode for these stars: from 3.9 to 4.08 ppm for HD49933 (depending on the run), 2.85 ppm for HD181906, and 3.89 ppm for HD191420.

We estimated roughly the mass and the radius of these CoRoT targets with the FDRM package, which are, within the large error bars, compatible with the published values from spectroscopic observations. The radii retrieved from stellar models are in agreement with published results: for HD49933, R=1.367 ±0.016 $R_\odot$, for HD181906, R=1.415 ±0.016 $R_\odot$, and for HD181420, R=1.623 ±0.026 $R_\odot$.

We have to keep in mind that our codes have been modified and tuned to give the best results with the simulated data that we analysed. However, we might have to tune them again with real data from Kepler. So they are not unchangeable and they might evolve in the next months.

We have already applied this pipeline to one month of Kepler data and obtained some promising results with solar-type stars and red giants (Bedding et al. 2010; Chaplin et al. 2010; Stello et al. 2010).

References

Appourchaux, T., Michel, E., Auvergne, M., et al. 2008, A&A, 488, 705
Arentoft, T., Kjeldsen, H., Bedding, T. R., et al. 2008, ApJ, 687, 1180
Asplund, M., Grevesse, N., & Sauval, A. J. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 336, Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis, ed. T. G. Barnes, III & F. N. Bash, 25
Baglin, A., Auvergne, M., Boisnard, L., et al. 2006, in COSPAR, Plenary Meeting, Vol. 36, 36th COSPAR Scientific Assembly, 3749
Ballot, J., Appourchaux, T., Toutain, T., & Guittet, M. 2008, A&A, 486, 867
Ballot, J., García, R. A., & Lambert, P. 2006, MNRAS, 369, 1281
Ballot, J., Turck-Chièze, S., & García, R. A. 2004, A&A, 423, 1051
Barban, C., Deheuvels, S., Baudin, F., et al. 2009, A&A, 506, 51
Basu, S. & Antia, H. M. 1995, MNRAS, 276, 1402
Basu, S., Christensen-Dalsgaard, J., Chaplin, W. J., et al. 1997, MNRAS, 292, 243
Bedding, T. R., Huber, D., Stello, D., et al. 2010, ApJ, accepted
Bedding, T. R. & Kjeldsen, H. 2003, Publications of the Astronomical Society of Australia, 20, 203
Benomar, O. 2008, Communications in Asteroseismology, 157, 98
Benomar, O., Baudin, F., Campante, T. L., et al. 2009, A&A, 507, L13
Borucki, W., Koch, D., Batalha, N., et al. 2009, in IAU Symposium, Vol. 253, IAU Symposium, 289–299
Bruntt, H., Kjeldsen, H., Buzasi, D. L., & Bedding, T. R. 2005, ApJ, 633, 440
Chaplin, W. J., Appourchaux, T., Arentoft, T., et al. 2008a, Astronomische Nachrichten, 329, 549
Chaplin, W. J., Appourchaux, T., Elsworth, Y. P., et al. 2010, ApJ, accepted
Chaplin, W. J., Elsworth, Y., Isaak, G. R., et al. 2002, MNRAS, 336, 979
Chaplin, W. J., Houdek, G., Appourchaux, T., et al. 2008b, A&A, 485, 813
Christensen-Dalsgaard, J. 2008a, Ap&SS, 316, 113
Christensen-Dalsgaard, J. 2008b, Ap&SS, 316, 13
Christensen-Dalsgaard, J. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 431–442
Christensen-Dalsgaard, J., Arentoft, T., Brown, T. M., et al. 2008, Journal of Physics Conference Series, 118, 012039
Christensen-Dalsgaard, J., Gough, D., & Toomre, J. 1985, Science, 229, 923
de Ridder, J., Barban, C., Baudin, F., et al. 2009, Nature, 459, 398
Frohlich, C., Andersen, B. N., Appourchaux, T., et al. 1997, Sol. Phys., 170, 1
Gabriel, A. H., Baudin, F., Boumier, P., et al. 2002, A&A, 390, 1119
García, R. A., Mathur, S., Ballot, J., et al. 2008, Sol. Phys., 251, 119
García, R. A., Régulo, C., Samadi, R., et al. 2009, A&A, 506, 41
García, R. A., Turck-Chièze, S., Boumier, P., et al. 2005, A&A, 442, 385
Gizon, L. & Solanki, S. K. 2003, ApJ, 589, 1009
Grevesse, N., Noels, A., & Sauval, A. J. 1993, A&A, 271, 587
Harvey, J. 1985, in ESA Special Publication, Vol. 235, Future Missions in Solar, Heliospheric & Space Plasma Physics, ed. E. Rolfe & B. Battrick, 199–+
Hekker, S., Broomhall, A., Chaplin, W. J., et al. 2009a, ArXiv e-prints 0911.2612
Hekker, S., Kallinger, T., Baudin, F., et al. 2009b, ArXiv e-prints 0906.5002
Huber, D., Stello, D., Bedding, T. R., et al. 2009, ArXiv e-prints 0910.2764
Karoff, C., Metcalfe, T. S., Chaplin, W. J., et al. 2009, ArXiv e-prints
Kjeldsen, H. & Bedding, T. R. 1995, A&A, 293, 87
Kjeldsen, H., Bedding, T. R., Arentoft, T., et al. 2008, ApJ, 682, 1370

Latham, D. W., Brown, T. M., Monet, D. G., et al. 2005, in Bulletin of the American Astronomical Society, Vol. 37, Bulletin of the American Astronomical Society, 1340

Mathur, S., Eff-Darwich, A., García, R. A., & Turck-Chièze, S. 2008a, A&A, 484, 517

Mathur, S., Jimenez-Reyes, S. J., & Garcia, R. A. 2008b, in Proceedings of GONG 2008/SOHO XXI, Solar-stellar dynamos as revealed by helio- and asteroseismology

Mathur, S., Turck-Chièze, S., Couvidat, S., & García, R. A. 2007, ApJ, 668, 594

Metcalfe, T. S., Creevey, O. L., & Christensen-Dalsgaard, J. 2009, ApJ, 699, 373

Metcalfe, T. S., Dziembowski, W. A., Judge, P. G., & Snow, M. 2007, MNRAS, 379, L16

Michel, E., Baglin, A., Auvergne, M., et al. 2008, Science, 322, 558

Michel, E., Samadi, R., Baudin, F., et al. 2009, A&A, 495, 979

Mosser, B., Michel, E., Appourchaux, T., et al. 2009, A&A, 506, 33

Régulo, C. & Roca Cortés, T. 2002, A&A, 396, 745

Stello, D., Basu, S., Bruntt, H., et al. 2010, ApJ, accepted

Stello, D., Chaplin, W. J., Basu, S., Elsworth, Y., & Bedding, T. R. 2009a, MNRAS, Submitted

Stello, D., Chaplin, W. J., Bruntt, H., et al. 2009b, ApJ, 700, 1589

Stello, D., Kjeldsen, H., Bedding, T. R., et al. 2004, Sol. Phys., 220, 207

Thompson, M. J., Toomre, J., Anderson, E., et al. 1996, Science, 272, 1300

Torrence, C. & Compo, G. P. 1998, Bulletin of the American Meteorological Society, 79, 61

Turck-Chièze, S., Couvidat, S., Kosovichev, A. G., et al. 2001, ApJ, 555, L69

Vorontsov, S. V., Baturin, V. A., & Pamiatnykh, A. A. 1991, Nature, 349, 49

Walker, G., Matthews, J., Kuschnig, R., et al. 2003, PASP, 115, 1023

Table 1. Number of common results for the asteroFLAG and Aarhus/asteroFLAG stars, for the 3 steps of our strategy and the corresponding number of results with a 0 confidence level.

Star	Step 1	Level 0	Step 2	Level 0	Step 3	Level 0
Arthur	182	11	223	10	294	62
Boris	300	5	305	11	531	135
Pancho	231	6	259	9	310	273
Aarhus/asteroFLAG	33	3	75	4	78	36

Table 2. Number of common results for all the steps of our strategy with the errors (incorrect results) as a function of the magnitude. The second half of the table shows the results when we add the common results of the two best steps. In the parentheses, we put the percentage of incorrect results.

Star / V	9	Error	10	Error	11	Error	12	Error
Arthur	176	2 (1%)	25	5 (20%)	5	5 (100%)	7	7 (100%)
Boris	209	0 (0%)	89	2 (2%)	3	1 (33%)	2	2 (100%)
Pancho	179	1 (0.5%)	52	11 (21%)	10	8 (80%)	9	8 (89%)
Arthur	182	2 (1%)	39	14 (36%)	14	12 (85%)	14	14 (100%)
Boris	214	0 (0%)	107	3 (3%)	7	2 (28%)	4	4 (100%)
Pancho	197	2 (1%)	70	14 (20%)	13	11 (84%)	13	11 (84%)

Table 3. Radius (R_F) from stellar modelling for some Aarhus/asteroFLAG simulated stars compared to the simulation input (R_M).

KIC ID	$R_{\rm F}$ ($R_\odot$)	$\sigma(R)$	$R_{\rm M}$ ($R_\odot$)	$|R_{\rm F} - R_{\rm M}|$
3648936	1.454	0.013	1.393	4.692σ
4041406	1.032	0.008	1.040	1.000σ
4757931	1.416	0.013	1.424	0.615σ
4861991	1.340	0.021	1.378	1.810σ
5790807	1.540	0.020	1.559	0.950σ
6129877	1.014	0.008	0.986	3.500σ
6307459	1.368	0.013	1.379	0.846σ
8349612	0.695	0.000	1.268	—
8745924	1.264	0.022	1.252	0.545σ
9204313	1.133	0.007	1.120	1.857σ
10644253	0.923	0.014	0.954	2.214σ
10454113	1.253	0.018	1.320	3.722σ
3221671	1.277	0.013	1.289	0.923σ
5295670	1.466	0.016	1.404	3.875σ
6285677	1.338	0.022	1.323	0.682σ
6306896	1.402	0.005	1.405	0.600σ
8561664	1.073	0.014	1.083	0.714σ
10454113	1.253	0.018	1.320	3.722σ

F=fitted; M=model

Table 4. Pipeline results for real data: rotation period, mean large separation, frequency-range of the p modes, position of maximum amplitude, and bolometric maximum amplitude per radial mode.

Star	Period (days)	$\langle\Delta\nu\rangle$ (μHz)	ϵ (μHz)	$f_{\rm min}$ (μHz)	$f_{\rm max}$ (μHz)	$\nu_{\rm max}$ (μHz)	ϵ (μHz)	$A_{\rm max}$ (ppm)	ϵ (ppm)
Sun	27.8	135.46	3.26	2040	4920	3074.7	1.02	2.99	0.13
HD49933 (1)	3.2	86.14	1.58	720	3060	1699.88	1.05	4.08	0.15
HD49933 (2)	3.3	86.33	1.73	720	3000	1804.34	0.28	3.90	0.08
HD181906	2.8	86.53	1.69	1200	2460	1896.9	1.2	2.85	0.13
HD181420	2.6	75.35	1.53	480	2760	1536.26	0.42	3.89	0.1

Table 5. Estimation of the radius (R') and the mass (M') of real data with the package FDRM compared to spectroscopic observations (R(obs) and M(obs)).

Star	R' ($R_\odot$)	ϵ ($R_\odot$)	R (obs) ($R_\odot$)	M' ($M_\odot$)	ϵ ($M_\odot$)	M (obs) ($M_\odot$)
Sun	1.003	0.1	1	1.0167	0.2	1
HD49933 (1)	1.483	0.15	1.34±0.016	1.3279	0.26	1.2
HD49933 (2)	1.567	0.15	1.34±0.016	1.5741	0.35	1.2
HD181906	1.581	0.16	1,392 ±0.054	1.6231	0.32	1.144 ±0.119
HD181420	1.7255	0.17	1.595 ±0.032	1.600	0.32	1.43 ±0.05

Table 6. Radius and fitted parameters for real data from stellar modelling (package RadEx).

Star	R_F ($R_\odot$)	$\sigma(R)$ ($R_\odot$)	$T_{\rm eff,F}$ (K)	$\langle \Delta\nu \rangle_F$ (μHz)	$\langle \Delta\nu \rangle_O$ (μHz)	$\epsilon_{\langle \Delta\nu \rangle_{\rm Obs}}$ (μHz)
Sun	1.016	0.023	5879	135.46	134.67	3.26
HD49933	1.398	0.040	6835	86.14	84.60	1.58
HD181420	1.613	0.032	6596	75.35	76.43	1.53
HD181906	1.420	0.018	6294	86.53	86.12	1.69

F= fitted; O = Observed