



Graduate Program Astrophysics 2021-2022
Observatoire de Paris/PSL
M2 International Research Track

Master Class Data - Seismic Analysis

Solar-like oscillations have so small amplitudes that detecting them is challenging. The reliable detection of these oscillations motivates the use of efficient statistical tools. Here, we introduce basic elements of asteroseismology and data analysis to show how to implement a Bayesian fit of solar-like oscillations using Markov chain Monte Carlo.

1 Introduction

The chapter *Seismic Analysis* of the Master Class *Data Analysis* makes use of a series of papers to present the challenge in analyzing an asteroseismic signal. Different directions are studied: identifying small seismic signal at low signal-to-noise ratio; fitting seismic modes.

A key reference for *preparing* the MC is Appourchaux (2014)¹, especially their Sections 3 (Digital signal processing and spectral analysis) and 4 (Data analysis and statistics). The key reference for *performing the analysis* is Handberg & Campante (2011), who provide a comprehensive guide for Bayesian peak-bagging of solar-like oscillators using a Markov chain Monte Carlo (MCMC) method. In addition to these references, this document presents the expected work, organized as follows:

- Section 2 introduces basic seismic information, in order to describe the oscillation pattern.
- Section 3 presents the observations and the seismic data.
- Section 4 briefly recalls, if needed, important properties of the Fourier analysis.
- Different statistical tests used in the literature are very briefly presented in Section 5.
- [The work to be done in your project is described in Section 6.](#)

[Training exercises](#) are proposed all along the document, to illustrate the context, introduce concepts, and prepare the work. You are not obliged to present them in your report, but can do it you develop some point in detail. Don't spend too much time in the training exercises.

You may work alone, or in pair, or as a group. In the latter case, I need to identify precisely everyone's role. The larger the group, the higher the level of expectations!

2 Asteroseismology

2.1 Solar-like oscillation pattern

An introduction on asteroseismology can be found in Mosser (2015). Don't pay too much attention and too much time in reading the useful sections (Introduction and Section 5: Ensemble asteroseismology) of this article. Other papers dealing with seismology are referred to in this Section; however, their reading is not necessary.

2.1.1 Asymptotic pattern

Asteroseismology is used for deriving constraints on stellar structure modelling and stellar evolution modelling. Solar-like oscillations consist of pressure modes² that follow a regular pattern closely described by the asymptotic expansion (Tassoul 1980, ³). Oscillation frequencies of radial order n and degree ℓ , both with positive integer values, write

$$\nu_{n,\ell} \simeq \left(n + \frac{\ell}{2} + \varepsilon \right) \Delta\nu_{\text{as}} + \text{2nd order term}, \quad (1)$$

where $\Delta\nu_{\text{as}}$ is the asymptotic large separation. The oscillation pattern peaks at the frequency of maximum oscillation signal ν_{max} , which scales as the surface gravity. Observations are fitted in near-asymptotic conditions by

$$\nu_{n,\ell} \simeq \left(n + \frac{\ell}{2} + \varepsilon + d_{0\ell} + \frac{\alpha}{2}(n - n_{\text{max}})^2 \right) \Delta\nu_{\text{obs}}, \quad (2)$$

where $\Delta\nu_{\text{obs}}$, hereafter noted $\Delta\nu$ without subscript, is the observed large separation (close to $\Delta\nu_{\text{as}}$), the small terms $d_{0\ell}$ are used to locate non-radial modes with respect to radial modes, α represents the second-order correction, and $n_{\text{max}} = \nu_{\text{max}}/\Delta\nu - \varepsilon$ (Mosser et al. 2013b).

The asteroseismic global parameters $\Delta\nu$ and ν_{max} scale respectively as the stellar mean density and surface gravity,

$$\frac{\Delta\nu}{\Delta\nu_{\odot}} \simeq \left(\frac{M}{M_{\odot}} \right)^{1/2} \left(\frac{R}{R_{\odot}} \right)^{-3/2} \quad (3)$$

$$\frac{\nu_{\text{max}}}{\nu_{\text{max}\odot}} \simeq \left(\frac{M}{M_{\odot}} \right) \left(\frac{R}{R_{\odot}} \right)^{-2} \left(\frac{T_{\text{eff}}}{T_{\text{eff}\odot}} \right)^{-1/2}, \quad (4)$$

where the stellar parameters (mass M , radius R , effective temperature T_{eff}) are calibrated to solar parameters. Therefore, $\Delta\nu$ and ν_{max} can be used to infer proxies of the stellar mass and radius.

Training: Consider a star of mass M and radius R as a self-gravitating body and infer a characteristic frequency from dimensional analysis. Then, use the seismic scaling relations (Eqs. 3 and 4) and express the relations $R = R(\nu_{\text{max}}, \Delta\nu)$ and $M = M(\nu_{\text{max}}, \Delta\nu)$ that provides relevant and precise proxies of the stellar fundamental parameters.

2.1.2 Mixed modes

In evolved stars, dipole (with $\ell = 1$) mixed modes are observed, which derive from the coupling of pressure

²Pressure is the restoring force of the waves that interfere to construct the pressure modes.

³Do not try to read this reference!

¹<https://arxiv.org/pdf/1103.5352.pdf>

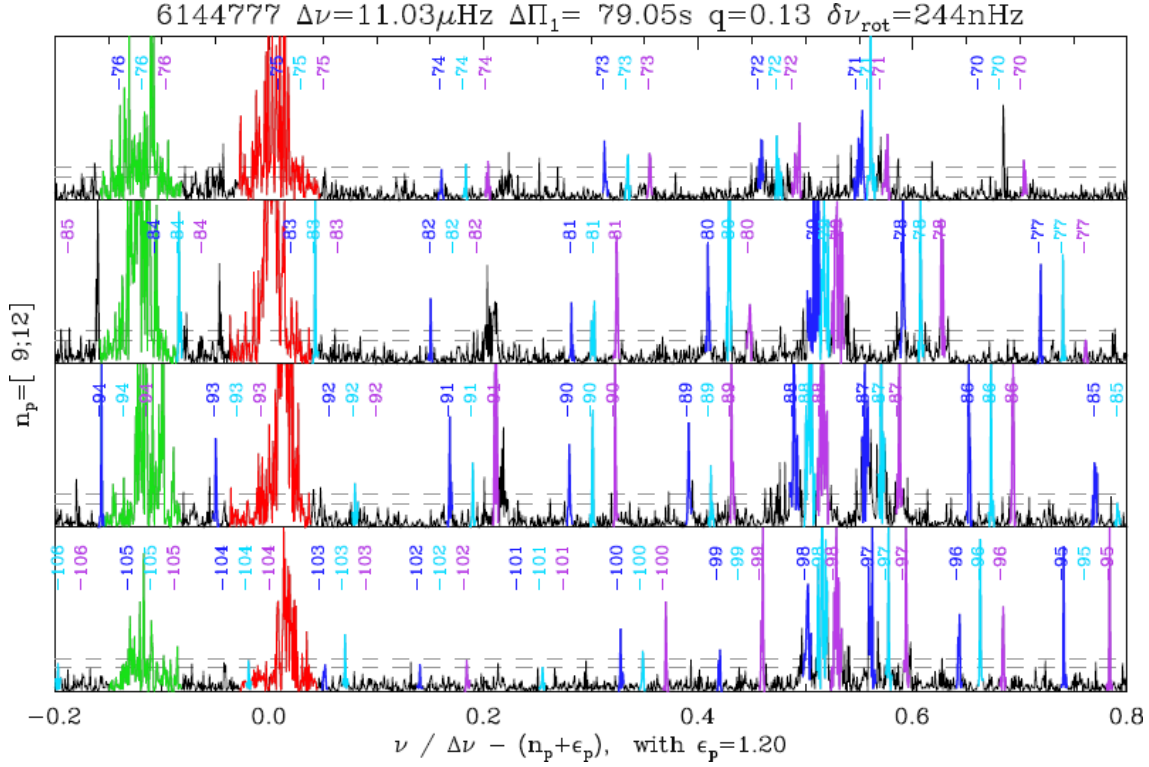


Figure 1: Fit of the oscillation pattern of the RGB star KIC 6144777, showing the pressure radial orders n_p (corresponding to n in Eqs. 1 and 2) from 9 to 12. The power spectrum density has been divided by the fit of the background. Radial and quadrupole modes are highlighted in red and green. The expected locations of dipole mixed modes are labelled with their mixed radial orders. When detected, mixed modes are highlighted in dark blue ($m = -1$), light blue ($m = 0$), or purple ($m = 1$). $\ell = 3$ modes, which are also mixed, are located near the abscissa 0.22; extra peaks in the range $[-0.2, -0.05]$ are mixed quadrupole modes. The gray dashed lines indicate the two thresholds used in this work, corresponding to height-to-background ratios of 7 and 10 (figure from Mosser et al. 2018).

waves propagating in the envelope and gravity waves propagating in the radiative core. As a consequences, many dipole mixed modes are present in a $\Delta\nu$ -wide frequency range (Fig. 1), most of them far away from the expected location of pure pressure modes (Eq. 2), according to the asymptotic expansion of mixed modes (Mosser et al. 2018, and references therein). Using the asymptotic expansion of mixed modes is beyond the scope of the work presented here.

2.1.3 Rotation

Oscillation spectra are complicated by rotation. In the framework of this MC, we limit the description of rotational multiplets to

$$\nu_{n,\ell,m} = \nu_{n,\ell} + m \delta\nu_{n,\ell}, \quad (5)$$

where the azimuthal order m is in the range $[-\ell, +\ell]$, and $\delta\nu_{n,\ell}$ is the rotational splitting.

For pure pressure modes, the rotational splittings write

$$\delta\nu_{n,\ell} = \delta\nu_{rot,p} \equiv \frac{1}{\langle T_{rot,env} \rangle}, \quad (6)$$

where $\langle T_{rot,env} \rangle$ is the mean rotation rate of the stellar envelope. For dipole ($\ell = 1$) mixed modes,

$$\delta\nu_{n,1} \simeq \zeta_{n,1} \delta\nu_{rot,g} \equiv \frac{\zeta_{n,1}}{\langle 2T_{rot,core} \rangle}, \quad (7)$$

where $\langle T_{rot,core} \rangle$ is the mean rotation rate of the stellar core.

Training: Identify dipole mixed modes in the oscillation pattern of the red giant branch (RGB) star KIC 6144777 observed by *Kepler* (Fig. 1).

2.2 Mode widths

Pressure-modes described by Eq. 2 are short lived, since they are stochastically excited, and damped. Their mean profile is then a Lorentzian. For a short-lived mode with radial order n , degree ℓ , and azimuthal order m , the profile is (Appourchaux 2014, and references therein)

$$M_{n,\ell,m}(\nu) = \frac{H_{n,\ell,m}}{1 + \frac{4}{\Gamma_{n,\ell,m}^2}(\nu - \nu_{n,\ell,m})^2}, \quad (8)$$

where H is the mode height and Γ is the mode width (half width at half-maximum). Mode visibilities, which depend on the mode degree, modulate the height; such point can be omitted here. The heights $H_{n,\ell,m}$ approximately follow a Gaussian distribution centered on ν_{\max} (Fig. 2).

The modes widths are important parameters of the model, among the many stellar and seismic parameters used for describing the oscillation modes. Observed modes have then, due to the stochastic nature of the oscillation, a Lorentzian profile multiplied by a random noise with a χ^2 probability distribution with two degrees of freedom (see Fig. 1 of Handberg & Campante 2011).

Training: Plot the mean and actual profiles of a short-lived modes. Provide an estimate of the mode width Γ to be used in the modeling you are interested in.

Long-lived modes, as the dipole mixed modes that lie away from the expected pure pressure dipole modes (close to the mid-point between two consecutive radial modes, as defined by Eq. 2) may have lifetimes longer than the total duration of the observation. As a consequence, they are unresolved, so that their profile is given by

$$U_{n,\ell,m}(\nu) = H_{n,\ell,m} \text{sinc} \pi \frac{(\nu - \nu_{n,\ell,m})}{\delta\nu_{res}}, \quad (9)$$

where $\delta\nu_{res}$ is the frequency resolution.

2.2.1 Rotational splittings

The different components of dipole rotational multiplets have different heights, depending on the stellar inclination i (Appourchaux 2014, and references therein):

$$H_{n,1,0} = H_{n,1} \cos^2 i \text{ and } H_{n,1,\pm 1} = H_{n,1} \frac{\sin^2 i}{2}. \quad (10)$$

2.3 Oscillation spectra

Oscillation spectra consist of the superimposition of three components: stellar background, noise, oscillation spectrum.

The mean background B is locally described around ν_{\max} by a scaling law

$$B(\nu) = B(\nu_{\max}) \left(\frac{\nu}{\nu_{\max}} \right)^\beta, \quad (11)$$

where $B(\nu_{\max})$ is the background at ν_{\max} and β is an exponent usually close to -2 . Such a power law is an approximation of the actual background that results from a superimposition of Harvey profiles (Eq. (11) of Handberg & Campante 2011). The observed background corresponds to the mean profile multiplied by a random noise with a χ^2 probability distribution with 2 degrees of freedom.

	KIC	Type	ν_{\max} (μHz)	$\Delta\nu$ (μHz)	SNR	References (for peak bagging)
1	11819446	Evolved red giant	0.84	0.21	poor	Mos2013
2	3645232	Evolved red giant	8.25	1.42		
3	11037292	Red giant	16.8	2.45		
4	6852836	Red clump	31.4	4.00		
5	1723700	Red clump	39.2	4.48	excellent	Mos2018
6	6144777	Red giant	128	11.03		
7	11925474	RGB	177	13.26		
8	7341231	Subgiant	383	28.92		Dev2012
9	5955122	Subgiant	857	49.2		App2012
10	6603624	Main sequence	2382	110.1		App2012

TABLE 1: Stellar time series or spectra available. App12 = Appourchaux et al. (2012); Dev12 = Deheuvels et al. (2012); Mos2013 = Mosser et al. (2013a); Mos2018 = Mosser et al. (2018)

Finally, the model of the power spectral density (PSD) is

$$P = \sum_n \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} M_{n,\ell,m} + B + W, \quad (12)$$

where the noise component N is supposed to be white noise; its observed value corresponds to a χ^2 probability distribution with 2 degrees of freedom, affected by the actual mean noise typically expressed in $\text{ppm}^2 \mu\text{Hz}^{-1}$ (as in Fig. 2). In Eq. 12, the values of the radial order n have to be considered around the value n_{\max} introduced in Eq. 2; only low degrees less than $\ell_{\max} = 3$ are present in oscillation spectra observed in photometry.

3 Oscillation spectra

3.1 Observations

Solar-like oscillations benefit from precise space-borne photometric observations. After some treatment, observations provide long, continuous, nearly uninterrupted times series of the relative photometric variations. As the time sampling is not completely regular, fast Fourier transform cannot apply. Instead, it is necessary to perform a Lomb-Scargle periodogram⁴ (see Section 4).

Kepler data were recorded with two cadences: short/long-cadence (SC/LC) data have sampling times close to 1/30 min, respectively. *Kepler* LC data may last as long as 4 years.

Training: Provide an estimate of the frequency resolution $\delta\nu_{\text{res}}$ and the Nyquist frequency ν_{Nyq} of 4-year LC data. Check the exact values in the seismic data presented in the next paragraph.

⁴<https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.lombscargle.html>

3.2 Data

Time series and spectra are provided for various stars listed in Table 1, as text table with a header and then two columns of data. Proxies of the global seismic parameters are given: $\Delta\nu$, ν_{\max} ... Stellar parameters are given too.

Training: Unsurprisingly, some of the data presented in Table 1 have to be read. The Python routine `numpy.loadtxt`⁵ can be used therefore.

4 Fourier analysis

4.1 Fourier transform, fft and Lomb Scargle periodogram

If you are not used to work in the Fourier space, you may refer to Butz T. (2006), Fourier Transformation for Pedestrians, Springer-Verlag Berlin Heidelberg⁶.

Asteroseismologists never use the Fourier transform, which is computationally too expensive. Instead, Lomb Scargle periodograms are used to cope with real data that are not recorded with a strictly regular sampling time (see Appourchaux 2014, and references therein). When data are regularly sampled, fft can be used, but requires that the number of points is, ideally, equal to a power of 2.

4.2 The Wiener Khintchine theorem

This theorem shows that the PSD of a stationary stochastic process is analogous to the Fourier transform of the corresponding autocorrelation function. The theorem, closely linked to the definition and use of PSD, is often used to define the properties of the

⁵<https://numpy.org/doc/stable/reference/generated/numpy.loadtxt.html>

⁶Available as an ebook

Fourier analysis. It is for instance used as the basis of the most efficient way for identifying the large separation $\Delta\nu$ in an oscillation spectrum, with the autocorrelation of the seismic time series computed as the Fourier transform of the oscillation spectrum (e.g. Mosser & Appourchaux 2009).

4.3 Parseval's identity

The calibration of the power spectrum must be properly performed. The Parseval's identity explains how the mean energy in the time series is linked to the mean energy density in the Fourier spectrum.

$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(\nu)|^2 d\nu \quad (13)$$

The total energy of a signal does not depend on the representation chosen: frequency or time.

Training: Transform a seismic time series into a PSD. Check the calibration. Gaps and accidents in the time series must be corrected; the mean value of the time series must be equal to zero. Why?

5 Statistical tools

The preparation work has introduced different elements that can be used to:

- check the properties of the observations;
- investigate their nature: signal, noise;
- provide a fit of seismic modes with uncertainties;
- and finally derive seismic and stellar parameters.

5.1. The null hypothesis

The null hypothesis, introduced by the geneticist and statistician Ronald Fischer in Fischer (1935)), consists in assuming that white noise can explain the observation. When applied to an oscillation spectrum, it consists in checking whether a given peak is signal or noise. If the height of the peak in the power spectrum is high enough, the H0 hypothesis is rejected, implying that a signal might have been detected (Appourchaux 2004). The H0 test is for instance used by Mosser & Appourchaux (2009) to determine the large frequency separation $\Delta\nu$ in an automated way.

5.2 Bayesian analysis

Bayes theorem introduces the prior probability, the posterior probability, the likelihood and the global likelihood. Refer to Appourchaux (2014) and Handberg & Campante (2011) for identifying the interest of the theorem and the way it can be used.

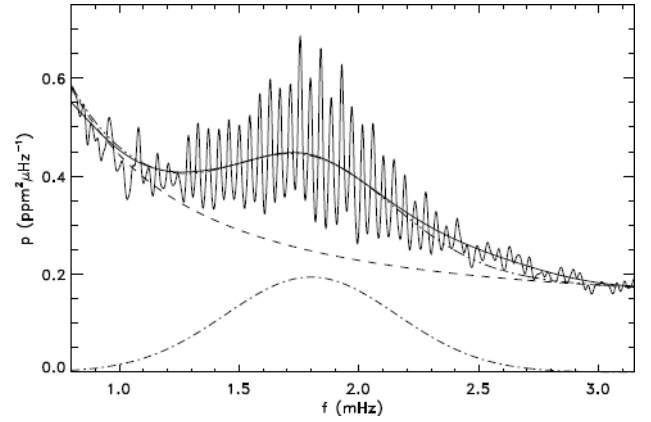


Figure 2: Smoothed power density distributions of the oscillation of HD 49933 observed by CoRoT (solid thin and thick lines). The dashed line represents the contributions of granulation and photon noise. The dash-dot lines account for the Gaussian modeling of the seismic envelope and the total contribution centered on ν_{\max} (figure from Mosser & Appourchaux 2009)

5.3 MCMC algorithm

Metropolis-Hastings and other Markov chain Monte Carlo (MCMC) algorithms are used for sampling multi-dimensional distributions, especially when the number of dimensions is high. A clear presentation of the Metropolis-Hastings algorithm is given by Fig. 3 of Handberg & Campante (2011). You can write your own code, use any routine already available⁷, or refer to Foreman-Mackey et al. (2013).

6 Your work

The aim of the project is to perform a Bayesian fit of an oscillation spectrum (or peak bagging, according to the jargon of asteroseismologists) with MCMC. In order to make this exercise as digestible as possible, we use the comprehensive guide written by Handberg & Campante (2011), hereafter noted HC11.

6.1 Fit

The set of parameters of the oscillation spectrum is large. For sake of simplification, you may reduce the fit to:

- the background and the oscillation power excess, as shown for instance in Fig. 2;
- or pure pressure modes described by Eq. 2, as shown in Fig. 5 of HC11, or as Fig. 3 of this presentation;

⁷For instance <https://emcee.readthedocs.io/en/stable/> for MCMC and <https://corner.readthedocs.io/en/latest/> for corner plots that show different projections of samples in high dimensional spaces.

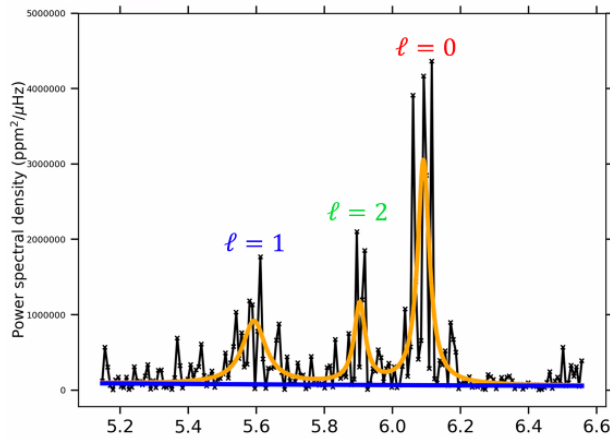


Figure 3: Observed (black line) and smoothed (orange line) oscillation spectrum of KIC 1432587 observed by *Kepler*. The blue line corresponds to the background (G. Dréau, private communication)

- or rotational multiplets; in this case you may consider the spectrum of KIC 6144777 (Fig. 1).

6.2 Bayesian terms

In order to compute the posterior probability, you may estimate first:

- the model, as given by Eq. (10) of HC11, but simplified according to the parameters you aim to fit. In each case, you have to define the useful parameters of the model; if you lack information about any seismic parameter and the role it plays in the modeled oscillation spectrum, just wake the prof up and ask him kindly. In a first stage, it is prudent to start the fit with a limited number of parameters, before increasing it.
- the prior probability comparing the observed spectrum and the model: Eqs. (1) or (2) of HC11. Here also, if you lack information, just ask.
- the priors for the useful parameters of the model, as shown in Section 3.3 of HC11 (and to be adapted). Then derive the posterior probability using Eq. (20) of HC11.

6.3 Report

6.3.1 Points to be evaluated

In your report, I am interested in:

- the fit you performed: explain your choice and the hypotheses that were used (for instance: terms that may vary slowly but are supposed to be fixed in a first approximation; terms that may vary with frequency, but are supposed to be constant⁸ over the frequency

⁸Constant with frequency, but considered as a variable for the fit.

range...)

- the definition of all terms you used (please provide a table with at least three information; the definition of the term, its physical symbol, its coding as a Python variable⁹; and avoid the confusion between physical symbol and code)
- the parameters used for modelling the spectrum;
- the priors of these parameters;
- the method: explain how you implement it, how you adapt the recipes given in the articles used as references...
- the algorithm you developed or the algorithms you used;
- the definition of any subroutine you developed;
- the results of the fits, with uncertainties, and explanation about the derivation of the uncertainties, expressed first in terms of frequencies, widths, heights... and possibly in terms of stellar parameters;
- figures illustrating the results;
- open questions, discussions...

The final mark will consider not only the results, but the way to reach them.

6.3.2 Deliverables

Documents to be placed on the Moodle, no later than two weeks after the second course:

- Your report, in pdf format, with an explicite name as `MCdata_seismo_Your Name(s)_2021.pdf`.
- Your code(s) and products, well documented and commented in the report, must be given too, independent of the report, in an archive with an explicite name as `MCdata_seismo_Your Name(s)_2021_code`.

References

- Appourchaux, T. 2004, *A&A*, 428, 1039
- Appourchaux, T. 2014, A crash course on data analysis in asteroseismology, ed. P. L. Pallé & C. Esteban, 123
- Appourchaux, T., Chaplin, W. J., García, R. A., et al. 2012, *A&A*, 543, A54
- Deheuvels, S., García, R. A., Chaplin, W. J., et al. 2012, *ApJ*, 756, 19
- Fischer, A. S. 1935, *The Design of Experiments* (Edinburgh: Oliver and Boyd), 18, 290
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Handberg, R. & Campante, T. L. 2011, *A&A*, 527, A56

⁹For instance: mode width, Γ , **Gamma**

Mosser, B. 2015, in EAS Publications Series, Vol. 73-74, EAS Publications Series, 3–110

Mosser, B. & Appourchaux, T. 2009, A&A, 508, 877

Mosser, B., Dziembowski, W. A., Belkacem, K., et al. 2013a, A&A, 559, A137

Mosser, B., Gehan, C., Belkacem, K., et al. 2018, A&A, 618, A109

Mosser, B., Michel, E., Belkacem, K., et al. 2013b, A&A, 550, A126

Tassoul, M. 1980, ApJS, 43, 469