

Bayesian peak-bagging of solar-like oscillators using MCMC: a comprehensive guide

R. Handberg¹ and T. L. Campante^{1,2}

¹ Danish AsteroSeismology Centre, Department of Physics and Astronomy, Aarhus University, 8000 Aarhus C, Denmark
e-mail: [rasmush; campante]@phys.au.dk

² Centro de Astrofísica, DFA-Faculdade de Ciências, Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal

Received 22 July 2010 / Accepted 29 December 2010

ABSTRACT

Context. Asteroseismology has entered a new era with the advent of the NASA *Kepler* mission. Long and continuous photometric observations of unprecedented quality are now available which have stimulated the development of a number of suites of innovative analysis tools.

Aims. The power spectra of solar-like oscillations are an inexhaustible source of information on stellar structure and evolution. Robust methods are hence needed in order to infer both individual oscillation mode parameters and parameters describing non-resonant features, thus making a seismic interpretation possible.

Methods. We present a comprehensive guide to the implementation of a Bayesian peak-bagging tool that employs a Markov chain Monte Carlo (MCMC). Besides making it possible to incorporate relevant prior information through Bayes' theorem, this tool also allows one to obtain the marginal probability density function for each of the fitted parameters. We apply this tool to a couple of recent asteroseismic data sets, namely, to CoRoT observations of HD 49933 and to ground-based observations made during a campaign devoted to Procyon.

Results. The developed method performs remarkably well at constraining not only in the traditional case of extracting oscillation frequencies, but also when pushing the limit where traditional methods have difficulties. Moreover it provides an rigorous way of comparing competing models, such as the ridge identifications, against the asteroseismic data.

Key words. methods: data analysis – methods: statistical – stars: late-type – stars: oscillations

1. Introduction

Seismology of solar-like stars is a powerful tool that can be used to increase our understanding of stellar structure and evolution. Solar-like oscillations in main-sequence stars and subgiants have been measured thanks to data collected from ground-based high-precision spectroscopy (for a review e.g., [Bedding & Kjeldsen 2008](#)) and, more recently, to photometric space-based missions such as CoRoT (e.g., [Michel et al. 2008](#)). Red giants also exhibit solar-like oscillations, although at lower frequencies, and hence require longer time series in order to resolve them (e.g., [De Ridder et al. 2009](#), and references therein). The launch of the NASA *Kepler* mission ([Koch et al. 2010](#)) definitely marked a milestone in the field of asteroseismology. *Kepler* will particularly lead to a revolution in the seismology of solar-like oscillators, since it will increase by more than two orders of magnitude the number of stars for which high-quality observations will be available, while allowing for long-term follow-ups of a selection of those targets. The large homogeneous sample of data made available by *Kepler* opens the possibility of conducting a seismic survey of the solar-like part of the colour-magnitude diagram, which researchers in the field already started naming as *ensemble asteroseismology*. As of the time of writing of this article, first results arising from the *Kepler* asteroseismic programme had already been made available ([Bedding et al. 2010a](#); [Chaplin et al. 2010](#); [Gilliland et al. 2010](#); [Hekker et al. 2010b](#); [Stello et al. 2010](#); [Christensen-Dalsgaard et al. 2010](#); [Metcalfe et al. 2010](#)).

The rich informational content of power spectra of solar-like oscillations allows fundamental stellar properties (e.g. mass, radius, and age) to be determined, and the internal structure to be constrained to unprecedented levels provided that individual oscillation mode parameters are measured (e.g., [Christensen-Dalsgaard 2004](#)). Furthermore, the measured stellar background signal provides us with valuable information on activity and convection. In the case of the highest signal-to-noise ratio (S/N) observations, for which it is possible to measure individual oscillation mode parameters, we expect asteroseismology to produce a major breakthrough on stellar structure and evolution, on topics as diverse as energy generation and transport, rotation and stellar cycles (e.g., [Karoff et al. 2009](#)).

For the past few years significant work has been invested in making preparations for the mode parameter analysis of *Kepler* data. This analysis involves the estimation of individual and average oscillation mode parameters, as well as estimation of parameters that describe non-resonant signatures of convection and activity. Examples include the work conducted in the framework of the AsteroFLAG consortium ([Chaplin et al. 2008a](#)) and the work undertaken by the CoRoT data analysis team ([Appourchaux et al. 2006](#)). This consequently paved the way for the development of suites of analysis tools for application to *Kepler* data ([Hekker et al. 2010a](#); [Huber et al. 2009](#); [Karoff et al. 2010](#); [Mathur et al. 2010](#); [Mosser & Appourchaux 2009](#); [Campante et al. 2010](#)).

In the present study we give continuity to this work by presenting a comprehensive guide to the implementation of a

Bayesian peak-bagging¹ tool that employs a MCMC. These techniques derive from the tools traditionally used in helioseismology and are in many ways an extension of the maximum likelihood estimation (MLE) methods. This peak-bagging tool is to be applied to the power spectra of solar-like oscillators and used as a means to infer both individual oscillation mode parameters and parameters describing non-resonant features. Besides making it possible to incorporate relevant prior information through Bayes' theorem, this tool also allows one to obtain the marginal probability density function (PDF) for each of the model parameters (frequencies, mode heights, mode lifetimes, rotational splitting, inclination angle etc.). This is one of the main advantages of these MCMC techniques, as it not only performs well in low signal-to-noise conditions, but also provides reliable error bars on the parameters. Parameter space is sampled using a *Metropolis-Hastings* algorithm featuring a built-in *statistical control system* that allows to automatically set an appropriate instrumental law during the burn-in stage. Also included is *parallel tempering*, which increases the mixing properties of the Markov chain.

The outline of the paper is as follows: we start in Sect. 2 by providing an overview of the theory behind the power spectrum of solar-like oscillations, introducing the assumptions and the set of parameters needed to model the spectrum to the level of detail required by modern asteroseismic data. In Sect. 3 we describe the subjacent Bayesian statistical framework by highlighting the topics of parameter estimation and model selection. Section 4 is devoted to the modus operandi of advanced Markov chain Monte Carlo methods and their implementation. In Sect. 5 we present a couple of examples where this tool has been applied to recent asteroseismic data sets, evidencing some of its capabilities and illustrating its functioning. A summary and discussion are presented in Sect. 6.

2. The power spectrum of solar-like oscillations

Solar-like oscillations or p modes (pressure playing the role of the restoring force) are global standing acoustic waves. They are characterized by being intrinsically damped while simultaneously stochastically excited by near-surface convection. Therefore, all stars cool enough to harbor an outer convective envelope – whose locus in a H–R diagram approximately extends from the cool edge of the Cepheid instability strip up to the red giant branch – may be expected to exhibit solar-like oscillations.

Modes of oscillation are characterized by three wave numbers: n , ℓ and m . The radial order n characterizes the behaviour of the mode in the radial direction. The degree ℓ and the azimuthal order m determine the spherical harmonic describing the properties of the mode as a function of colatitude and longitude. In the case of stellar observations, the associated whole-disk light integration and consequent lack of spatial resolution strongly suppress the signal from all but the modes of the lowest degree (with $\ell \leq 3$). For a spherically symmetric star mode frequencies depend only on n and ℓ .

2.1. Statistics and likelihood function of the spectrum

Stellar p modes can be modelled as stochastically excited and intrinsically damped harmonic oscillators (Kumar et al. 1988).

¹ The term “peak-bagging” has become the customary name for the examination of individual oscillation peaks in the field of asteroseismology. The origin of the name is explained in Appourchaux (2003b).

The frequency-power spectrum arising from such a system can in turn be modelled by a mean spectrum profile, $\mathcal{P}(v_j; \Theta)$, described by the set of parameters Θ which contain the desired physical information, multiplied by a random noise with a χ^2 probability distribution with 2 degrees of freedom (Woodard 1984; Duvall & Harvey 1986). This means that, at a fixed frequency bin j , the probability density, $f(P_j)$, that the observed power spectrum takes a particular value P_j , is related to the mean spectrum, $\mathcal{P}(v_j; \Theta)$, by:

$$f(P_j) = \frac{1}{\mathcal{P}(v_j; \Theta)} \exp \left[-\frac{P_j}{\mathcal{P}(v_j; \Theta)} \right]. \quad (1)$$

Very often when dealing with long time series, it is customary to divide the observational data set into several independent subsets, to compute their separate spectra and to average them. In doing so one aims at decreasing the variance in the power spectrum. The average power spectrum will then obey a χ^2 probability distribution with $2s$ degrees of freedom, χ_{2s}^2 , s being the number of combined spectra (Appourchaux 2003a):

$$f(P_j) = \frac{s^{s-1}}{(s-1)!} \frac{P_j^{s-1}}{\mathcal{P}(v_j; \Theta)^s} \exp \left[-\frac{s P_j}{\mathcal{P}(v_j; \Theta)} \right]. \quad (2)$$

Equation (2) also holds when binning the power spectrum over s bins (Appourchaux 2004).

We would now like to specify the likelihood function, i.e., the joint PDF for the data sample $\{P_j\}$. Assuming that the frequency bins are uncorrelated, the joint PDF is simply given by the product of $f(P_j; \Theta)$ over some frequency interval of interest spanned by j :

$$L(\Theta) = \prod_j f(P_j; \Theta). \quad (3)$$

Notice that we have written $f(P_j; \Theta)$ to make the dependence on the parameters Θ explicit. In spite of the fact that Eq. (3) is valid for an uninterrupted data set, the same is not true when gaps are present in the time series. In that event, Stahn & Gizon (2008) have derived an expression for the joint PDF of solar-like oscillations in complex Fourier space, in agreement with the earlier work of Gabriel (1994). The latter PDF explicitly takes into account frequency correlations introduced by the convolution with the spectral window.

The basic idea when employing a Maximum Likelihood Estimator (MLE) is to determine estimates $\tilde{\Theta}$ so as to maximize the likelihood function (e.g., Toutain & Appourchaux 1994). Due to improved numerical stability, however, it is more convenient, in practice, to work with logarithmic probabilities:

$$\begin{aligned} \mathcal{L}(\Theta) &\equiv \ln L(\Theta) \\ &= - \sum_j \left\{ \ln \mathcal{P}(v_j; \Theta) + \frac{P_j}{\mathcal{P}(v_j; \Theta)} \right\}. \end{aligned} \quad (4)$$

One therefore ends up maximizing the logarithm of the likelihood function instead:

$$\tilde{\Theta} = \arg \max_{\Theta} \{ \mathcal{L}(\Theta) \}. \quad (5)$$

2.2. Modelling the power spectrum²

The power spectrum of a single mode of oscillation is distributed around a mean profile with an exponential probability distribution according to Eq. (1). As already mentioned, this mean profile contains the information on the physics of the mode. In the limit of taking the ensemble average of an infinite number of realisations of the power spectrum, it can be shown (Anderson et al. 1990) that the limit spectrum thus obtained follows in fact a standard Lorentzian profile near the resonance, i.e., for $|\nu - \nu_0| \ll \nu_0$. A Lorentzian profile is defined as:

$$\mathcal{M}(\nu; S, \nu_0, \Gamma) = \frac{S}{1 + \frac{4}{\Gamma^2}(\nu - \nu_0)^2}, \quad (6)$$

where S is the mode height and Γ is the mode linewidth. Γ is related to the mode lifetime, τ , through $\Gamma = (\pi\tau)^{-1}$. In the case of solar-type stars and for low angular degree ℓ , we can assume that Γ is a function of frequency alone, which is supported both by observations of the Sun and by theoretical models (e.g. Aerts et al. 2010; Dupret et al. 2009).

A power spectrum of solar-like oscillations will, of course, contain a myriad of modes spanning a broad range in frequency, superimposed on a background signal of both stellar and instrumental origin. The overall limit spectrum is then given by the sum of the separate limit spectra arising from the different sources, since interference effects from beating between the modes average out in the limit. Notice that we are assuming that a mode is uncorrelated with any other modes or with the background signal. In doing so, we neglect any eventual asymmetries of the Lorentzian profiles (Duvall et al. 1993; Abrams & Kumar 1996). Nevertheless, when dealing with long time series, such asymmetries should be included in order to avoid biases in mode frequency determination. Furthermore, the presence of gaps and the finite length of the time series lead to a degradation of the observed power spectrum, which then results from the convolution of the true spectrum (i.e., the one that would be obtained were there no gaps) with the power spectrum of the window function (i.e., the spectral window). However, this problem is overcome by convolving the final limit spectrum with the spectral window.

Ignoring any departure from spherical symmetry, non-radial modes differing only on the azimuthal number m are degenerate and their profiles will be combined into a single profile, that of the (n, ℓ) multiplet. Stellar rotation removes the $(2\ell + 1)$ -fold degeneracy of the frequency of oscillation of non-radial modes, thus allowing for a direct measurement of the angular velocity of the star averaged over the region probed by these modes. When the angular velocity of the star, Ω , is small and in the case of rigid-body rotation, the frequency of a (n, ℓ, m) mode is given to first order by (Ledoux 1951):

$$\nu_{n\ell m} = \nu_{n\ell} + m \frac{\Omega}{2\pi} (1 - C_{n\ell}). \quad (7)$$

The kinematic splitting, $m\Omega/(2\pi)$, is corrected for the effect of the Coriolis force through the dimensionless quantity $C_{n\ell} > 0$. In the asymptotic regime, i.e., for high-order, low-degree p modes, rotational splitting is dominated by advection and the splitting between adjacent modes within a multiplet is $\nu_s \approx \Omega/(2\pi)$. Second-order rotational effects are related to the distortion of the

² To be precise, we will be modelling the power density spectrum and not the power spectrum. The former is independent of the window function and is obtained by multiplying the power spectrum by the effective length of the observational run, which can in turn be calculated as the reciprocal of the area integrated under the spectral window.

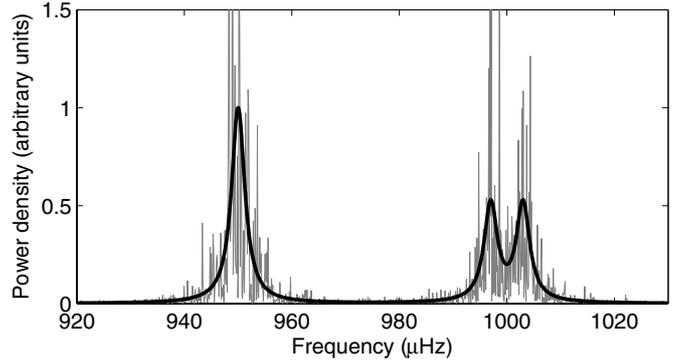


Fig. 1. Artificial power density spectrum of a $\ell = 0$ singlet and a $\ell = 1$ multiplet. One has $\nu_s = \Gamma = 3 \mu\text{Hz}$. Notice how the $\ell = 1$ multiplet splits into its m components. The power spectrum (grey) is distributed around a mean spectrum (black) with an exponential probability distribution.

equilibrium structure of the star caused by centrifugal forces. Although negligible in the Sun, these effects are significant for faster solar-type rotators where these effects can cause non-negligible biases on frequency determinations (e.g., Ballot 2010). Large-scale magnetic fields may also introduce further corrections to the oscillation frequencies.

Assuming energy equipartition between the components of a multiplet, we define the following symmetric profile for a (n, ℓ) multiplet:

$$\mathcal{M}_{n\ell}(\nu; S_{n\ell}, \nu_{n\ell}, \Gamma_{n\ell}, \nu_s, i) = \sum_{m=-\ell}^{\ell} \mathcal{E}_{\ell m}(i) \mathcal{M}(\nu; S_{n\ell}, \nu_{n\ell} + m\nu_s, \Gamma_{n\ell}), \quad (8)$$

where $\mathcal{E}_{\ell m}(i)$ represents mode visibility within a multiplet and i is the inclination angle between the direction of the stellar rotation axis and the line of sight. The overall profile of a multiplet thus consists of the sum of $2\ell + 1$ Lorentzian profiles regularly spaced in frequency and scaled in height according to the $\mathcal{E}_{\ell m}(i)$ factors (see Fig. 1), which in turn are given by (Gizon & Solanki 2003):

$$\mathcal{E}_{\ell m}(i) = \frac{(\ell - |m|)!}{(\ell + |m|)!} \left[P_{\ell}^{|m|}(\cos i) \right]^2, \quad (9)$$

where $P_{\ell}^m(x)$ are the associated Legendre functions. Notice that $\sum_m \mathcal{E}_{\ell m}(i) = 1$, meaning that $\mathcal{E}_{\ell m}(i)$ represents the relative power contained in a mode within a multiplet.

Since we are primarily interested in performing a so-called *global fit* (e.g., Appourchaux et al. 2008) to the observed power spectrum, whereby several radial orders are fitted simultaneously within a broad frequency range, we end up modelling the mean acoustic spectrum according to the following general relation:

$$\mathcal{P}(\nu; \Theta) = \sum_{n=n_0}^{n_{\max}} \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} \frac{\mathcal{E}_{\ell m}(i) S_{n\ell}}{1 + \frac{4}{\Gamma_{n\ell}^2}(\nu - \nu_{n\ell} - m\nu_s)^2} + N(\nu), \quad (10)$$

where we have also included a profile describing the background signal, $N(\nu)$. Granulation, faculae and active regions might contribute to the stellar background signal, which is commonly modelled as a sum of power laws describing these physical phenomena (Harvey 1985; Aigrain et al. 2004):

$$N(\nu) = \sum_{k=1}^{k_{\max}} \frac{4A_k^2 B_k}{1 + (2\pi B_k \nu)^{C_k}} + N, \quad (11)$$

Table 1. Relative spatial response functions, V_ℓ/V_0 , for a number of present and upcoming instruments/missions.

	Intensity			Velocity	
	VIRGO (862 nm)	CoRoT (660 nm)	<i>Kepler</i> (641 nm) ^a	HARPS (535 nm)	SONG (550 nm)
V_0/V_0	1.00	1.00	1.00	1.00	1.00
V_1/V_0	1.20	1.22	1.22	1.35	1.35
V_2/V_0	0.67	0.70	0.71	1.02	1.01
V_3/V_0	0.10	0.14	0.14	0.48	0.47
V_4/V_0	-0.10	-0.09	-0.08	0.09	0.09

Notes. Notice the increased sensitivity to $\ell = 3, 4$ modes in velocity. Negative values of V_ℓ mean that the oscillations will appear to have reversed phases. ^(a) Calculated as the weighted mean over the spectral response function.

$\{A_k\}$ and $\{B_k\}$ being, respectively, the corresponding amplitudes and characteristic time-scales, whereas the $\{C_k\}$ are the slopes of each of the individual power laws. A flat component, N , is needed in order to model the photon shot noise. Equation (11) might just well incorporate any instrumental background signal. We refer to $S_{n\ell}/N(v_{n\ell})$ as the signal-to-noise ratio (in power) of the multiplet (n, ℓ) .

Once again assuming energy equipartition between the different components of a multiplet, their heights can be expressed as:

$$S_{n\ell m} = \mathcal{E}_{\ell m}(i) S_{n\ell} = \mathcal{E}_{\ell m}(i) V_\ell^2 \alpha_{n\ell}. \quad (12)$$

The quantity V_ℓ^2 is an estimate of the geometrical visibility of the total power in a multiplet (n, ℓ) as a function of ℓ , whereas $\alpha_{n\ell}$ depends mainly on the frequency and excitation mechanism, i.e., $\alpha_{n\ell} \simeq \alpha(v_{n\ell})$. Christensen-Dalsgaard (2003) concisely treats this issue of spatial filtering. Equation (12), however, is only strictly valid under one assumption: when the stellar flux is integrated over the full apparent disc, one must assume that the weighting function, W , which gives the contribution of a surface element to the integral, is a function of the distance to the disc centre alone, i.e., $W = W(\theta')$, where θ' is defined in an inertial frame with polar axis pointing toward the observer. In this case, the apparent mode amplitude can effectively be separated into two factors: $\mathcal{E}_{\ell m}(i)$ and V_ℓ^2 . This assumption holds very well in the case of intensity measurements, since the weighting function is then mainly linked to the limb-darkening, whereas for velocity measurements departures might be observed due to asymmetries in the velocity field induced by rotation (see Ballot et al. 2006, 2008, and references therein). See Appendix A for how to compute $\mathcal{E}_{\ell m}(i)$ and V_ℓ .

The heights of non-radial modes are commonly defined based on the heights of radial modes according to Eq. (12), and taking into account the V_ℓ/V_0 ratios. Note that $\ell = 0$ modes constitute a sensible reference since they are not split by rotation. Table 1 displays the relative spatial response functions, V_ℓ/V_0 , computed according to Bedding et al. (1996), for a number of present and upcoming instruments/missions used when measuring solar-like oscillations. Those performing intensity measurements are the red channel of the VIRGO SPM instrument on board the SOHO spacecraft (Fröhlich et al. 1995), as well as the CoRoT and *Kepler* space missions. On the other hand, velocity measurements are performed by the HARPS spectrograph (Mayor et al. 2003) and are the purpose of the forthcoming SONG network (Grundahl et al. 2007).

Finally, a possible set of parameters going into the model is given by:

$$\Theta = \{S_{n\ell}, v_{n\ell}, \Gamma_{n\ell}, v_s, i, A_k, B_k, C_k, N\}. \quad (13)$$

We have described in detail how the modelling of a power spectrum of solar-like oscillations can be achieved. When actually fitting a model to an observed power spectrum, the set of parameters entering the model might differ from the one represented in Eq. (13). Moreover, it might be desirable to justifiably fix some of the parameters in order to reduce the dimension of parameter space.

3. Bayesian inference

Having set up the model of the power spectrum, we will now introduce the Bayesian statistical framework to be used for estimating the model parameters and for comparing competing models. Let us start by considering a set of competing hypotheses, $\{H_i\}$, not necessarily mutually exclusive. We should then be able to assign a probability, $p(H_i|D, I)$, to each hypothesis, taking into account the observed data, D , and available prior information, I , arising from theoretical considerations and/or previous observations. This is done through Bayes' theorem:

$$p(H_i|D, I) = \frac{p(H_i|I)p(D|H_i, I)}{p(D|I)}. \quad (14)$$

The probability of the hypothesis H_i in the absence of D is called the *prior probability*, $p(H_i|I)$, whereas the probability including D is called the *posterior probability*, $p(H_i|D, I)$. The quantity $p(D|H_i, I)$ is called the *likelihood* of H_i , $p(D|I)$ being the *global likelihood* for the entire class of hypotheses. Bayesian inference thus encodes our current state of knowledge into a posterior probability concerning each member of the hypothesis space of interest. Moreover, the sum of the posterior probabilities over the hypothesis space of interest is unity, and thus

$$p(D|I) = \sum_i p(H_i|I)p(D|H_i, I). \quad (15)$$

3.1. Parameter estimation

Very often a particular hypothesis, i.e., a model of the power spectrum, is assumed to be true and the hypothesis space of interest then relates to the values taken by the model parameters Θ . These parameters are continuous, which means that the quantity of interest is a PDF. The global likelihood of model M , assumed true, is now given by the continuous counterpart of Eq. (15):

$$p(D|I) = \int p(\Theta|I) p(D|\Theta, I) d\Theta. \quad (16)$$

Let us restate Bayes' theorem in order to account for this new formalism:

$$p(\Theta|D, I) = \frac{p(\Theta|I)p(D|\Theta, I)}{p(D|I)}, \quad (17)$$

where we have substituted the hypothesis, H_i , with the parameters of the model that is assumed true. The terms entering this equation have the same meaning as the corresponding terms entering Eq. (14). Use of Eq. (17) allows one to obtain the full joint posterior PDF, $p(\Theta|D, I)$, this being the Bayesian solution to the problem of parameter estimation in contrast to traditional point

estimation methods (e.g. MLE). The procedure of *marginalisation* makes it possible to derive the marginal posterior PDF for a subset of parameters Θ_A , by integrating out the remaining parameters Θ_B , called *nuisance parameters*:

$$p(\Theta_A|D, I) = \int p(\Theta_A, \Theta_B|D, I) d\Theta_B. \quad (18)$$

Furthermore, assuming that the prior on Θ_A is independent of the prior on the remaining parameters, then by applying the product rule we have:

$$p(\Theta_A, \Theta_B|I) = p(\Theta_A|I)p(\Theta_B|\Theta_A, I) = p(\Theta_A|I)p(\Theta_B|I). \quad (19)$$

We will be working, in practice, with logarithmic probabilities. The global likelihood of the model plays the role of a normalisation constant and we rewrite Eq. (17) as follows:

$$\ln p(\Theta|D, I) = \text{const.} + \ln p(\Theta|I) + \mathcal{L}(\Theta). \quad (20)$$

3.2. Model comparison

We might also be facing a situation wherein several parametrized models are available to describe the same physical phenomenon. We then expect Bayes' theorem to allow for a statistical comparison between these competing models. In fact, Bayesian model comparison has a built-in *Occam's razor*, a principle also known as *lex parsimoniae*, by which a complex model is automatically penalised, unless the available data justifies its additional complexity. Notice that these might be intrinsically different models or similar models with varying number of parameters, or even the same model with different priors for its parameters.

Given two or more competing models, $\{M_i\}$, and our prior information, I , being in the current context that *one and only one of the models is true*, we can assign individual probabilities similarly to what has been done in Eq. (14), after substituting H_i with M_i :

$$p(M_i|D, I) = \frac{p(M_i|I)p(D|M_i, I)}{p(D|I)}, \quad (21)$$

where $p(D|M_i, I)$, also called the *evidence* of model M_i , is given by Eq. (16). The problem of model comparison is therefore analogous to the problem of parameter estimation as can be seen by comparing Eqs. (17) and (21).

Of particular interest to us will be calculating the ratio of the probabilities of two competing models,

$$O_{ij} \equiv \frac{p(M_i|D, I)}{p(M_j|D, I)} = \frac{p(M_i|I)p(D|M_i, I)}{p(M_j|I)p(D|M_j, I)} = \frac{p(M_i|I)}{p(M_j|I)} B_{ij}, \quad (22)$$

where O_{ij} is the *odds ratio* in favour of model M_i over model M_j , B_{ij} is the so-called *Bayes' factor* and the remaining factor is the *prior odds ratio*. We will always assume that we have no prior information impelling us to prefer one model over the other, and hence $p(M_i|I)/p(M_j|I) = 1$. One is now naturally in need of a scale by which to judge the ratio of the evidences of two competing models. The usual scale employed is the Jeffreys' scale (Jeffreys 1961), which we display in Table 2 for convenience.

Furthermore, the Bayesian framework makes it possible to extract parameter constraints even in the presence of model uncertainty, i.e., when the implementation of model selection has not been successful. This is done by simply combining the probability distribution of the parameters within each individual model, weighted by the model probability. This procedure, called *Bayesian model averaging* (see Liddle 2009, and references therein), is an analogue of the superposition of eigenstates of an observable in quantum mechanics.

Table 2. Jeffreys' scale.

$\ln O_{ij}$	Strength of evidence
<1	Not worth more than a bare mention
1–2.5	Significant
2.5–5	Strong to very strong
>5	Decisive

3.3. Ignorance priors

The main advantage of the Bayesian framework when compared to a frequentist approach is the ability to incorporate relevant prior information through Bayes' theorem and evaluate its effect on our conclusions. Assuming that the prior on each parameter is independent of the prior on any other parameter, then according to Eq. (19) we have:

$$p(\Theta|I) = \prod_k f_k(\Theta_k), \quad (23)$$

where $f_k(\Theta_k)$ is the prior PDF associated with the k th parameter entering the model. As our state of knowledge of a particular physical phenomenon evolves through continued study and experimentation, the set of priors relevant for the analysis of a new data set will change. In the early stages of research, however, we look for a set of priors that encode our rather limited state of knowledge, i.e., a set of *ignorance priors* (e.g., Gregory 2005a, and references therein).

When dealing with *location parameters*, e.g. $\{v_{nl}\}$ in Eq. (13), our choice of prior would at first be the *uniform prior*:

$$f_k(\Theta_k) = \begin{cases} \frac{1}{\Theta_k^{\max} - \Theta_k^{\min}}, & \text{for } \Theta_k^{\min} \leq \Theta_k \leq \Theta_k^{\max}, \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

If we are ignorant about the limits Θ_k^{\min} and Θ_k^{\max} , then we refer to $f_k(\Theta_k)$ as an improper prior, meaning that it is not normalised. An improper prior is not suitable for model comparison problems. On the other hand, when dealing with *scale parameters*, e.g. $\{S_{nl}\}$ in Eq. (13), our choice of prior might be that of a *Jeffreys' prior*:

$$f_k(\Theta_k) = \begin{cases} \frac{1}{\Theta_k \ln[\Theta_k^{\max}/\Theta_k^{\min}]}, & \text{for } \Theta_k^{\min} \leq \Theta_k \leq \Theta_k^{\max}, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

By employing a Jeffreys' prior we are assigning equal probability per decade (scale invariance), mainly useful when the prior range spans several orders of magnitude. In case the prior lower limit includes zero, a *modified Jeffreys' prior* should be used instead to avoid the divergence at zero:

$$f_k(\Theta_k) = \begin{cases} \frac{1}{\Theta_k + \Theta_k^{\text{uni}} \ln[\Theta_k^{\text{uni}} + \Theta_k^{\max}/\Theta_k^{\text{uni}}]}, & \text{for } 0 \leq \Theta_k \leq \Theta_k^{\max}, \\ 0, & \text{otherwise.} \end{cases} \quad (26)$$

For $\Theta_k \gg \Theta_k^{\text{uni}}$, Eq. (26) behaves just like a Jeffreys' prior, whereas for $\Theta_k \ll \Theta_k^{\text{uni}}$ it behaves like a uniform prior, thus not diverging at zero. Θ_k^{uni} marks the transition between the two regimes.

4. Markov chain Monte Carlo

After inspection of Eq. (18), the need for a mathematical tool that is able to efficiently evaluate the multi-dimensional integrals required in the computation of the marginal posteriors becomes clear. This constitutes the rationale behind the method known as Markov chain Monte Carlo, first introduced in the early 1950s by statistical physicists and nowadays widely used in all areas of science and economics.

4.1. Metropolis-Hastings algorithm

The aim is to draw samples from the *target distribution*, $p(\Theta|D, I)$, by constructing a pseudo-random walk in model parameter space such that the number of samples drawn from a particular region is proportional to its posterior density. Such a pseudo-random walk is achieved by generating a Markov chain, whereby a new sample, X_{t+1} , depends on the previous sample³, X_t , in accordance with a time-independent quantity called the *transition kernel*, $p(X_{t+1}|X_t)$. After a burn-in phase, $p(X_{t+1}|X_t)$ is able to generate samples of Θ with a probability density converging on the target distribution. The Markov chain must fulfil three requirements in order to achieve this convergence: it must be *irreducible*, *aperiodic* and *positive recurrent* (Roberts 1996).

The algorithm that we employ in order to generate a Markov chain was initially proposed by Metropolis et al. (1953), and subsequently generalised by Hastings (1970), this latter version being commonly referred to as the Metropolis-Hastings algorithm. It works in the following way: Suppose the current sample, at some instant denoted by t , is represented by X_t . We would like to steer the Markov chain toward the next sampling state, X_{t+1} , by first proposing a new sample to be drawn, Y , from a *proposal distribution*, $q(Y|X_t)$, centred on X_t . Here we specifically treat $q(Y|X_t)$ as being a multivariate normal distribution with covariance matrix Σ . We employ independent Gaussian parameter proposal distributions and thus Σ is assumed diagonal. The proposed sample is then accepted with a probability given by:

$$\alpha(X_t, Y) = \min(1, r) = \min\left[1, \frac{p(Y|D, I) q(X_t|Y)}{p(X_t|D, I) q(Y|X_t)}\right], \quad (27)$$

where $\alpha(X_t, Y)$ is the *acceptance probability* and r is called the *Metropolis ratio*. In the present case of a symmetric proposal distribution, we have $q(X_t|Y) = q(Y|X_t)$. As a result, if the posterior density for the proposed sample is greater than or equal to that of the current sample, i.e., $p(Y|D, I) \geq p(X_t|D, I)$, then the proposal will always be accepted, otherwise it will be accepted with a probability given by the ratio of the posterior densities. If Y is not accepted, then the chain will keep the current sampling state, i.e., $X_{t+1} = X_t$. The procedure just described is repeated for a predefined number of iterations or, alternatively, for a number of iterations determined by a convergence test applied to the Markov chain (e.g., Gelman & Rubin 1992). The total number of iterations is denoted by n_{it} .

Once a Markov chain has been created, the problem of marginalization becomes trivial, as the way to extract information on the individual parameters is simply to generate a histogram for each parameter and thus obtain its PDF. An appropriate number of bins in the histograms can be selected using for example Scott's criterion (Scott 1979). Usually the information in the PDF will be condensed using some summary statistics, like for example finding the median of the distribution and the 68% credible region around it.

4.2. Parallel tempering

The Metropolis-Hastings algorithm outlined above might become stuck in a local maximum of the target distribution, thus failing to fully explore all regions in parameter space containing significant probability. A way of overcoming this is to employ parallel tempering (e.g., Earl & Deem 2005), whereby a discrete set of progressively flatter versions of the target distribution are

³ A remark on the notation: X_t may be thought of as a single vector in parameter space.

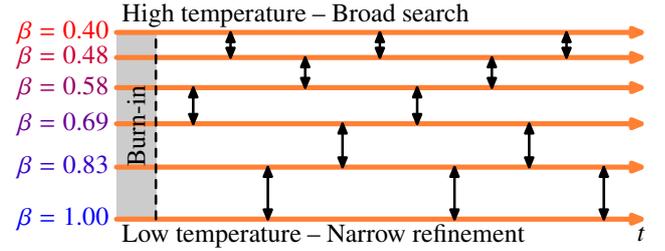


Fig. 2. Schematic representation of the functioning of the parallel tempering mechanism, whereby tempering chains are allowed to swap their parameter states (swaps are indicated by vertical arrows).

created by introducing a *temperature parameter*, \mathcal{T} . In practice, use is made of its reciprocal, $\beta = 1/\mathcal{T}$, referred to as the *tempering parameter*. By modifying Eq. (17), we generate the tempered distributions as follows:

$$p(\Theta|D, \beta, I) \propto p(\Theta|I) p(D|\Theta, I)^\beta, \quad 0 < \beta \leq 1. \quad (28)$$

For $\beta = 1$, we recover the target distribution, also called the *cold sampler*, whereas for $\beta < 1$ the hotter distributions are effectively flatter versions of the target distribution. Drawing samples from a hotter, i.e., flatter, version of the target distribution will allow, in principle, to visit regions of parameter space containing significant probability, otherwise not accessible to the basic algorithm. The problem of parameter estimation obviously continues to rely on samples drawn from the cold sampler. In Sect. 4.4 we describe how samples drawn from the remaining tempered distributions are useful in evaluating Bayes' factor.

Implementation of parallel tempering works in the following way: Several versions of the Metropolis-Hastings algorithm are launched in parallel (n_β in total), each being characterised by a different tempering parameter, β_i . At random intervals, comprehending a mean number (n_{swap}) of iterations, a pair of adjacent chains, labelled with β_i and β_{i+1} , is randomly chosen and a proposal is made to swap their parameter states. The proposed swap is then accepted with a probability given by:

$$\begin{aligned} \alpha_{\text{swap}} &= \min(1, r_{\text{swap}}) \\ &= \min\left[1, \frac{p(X_{t,i+1}|D, \beta_i, I) p(X_{t,i}|D, \beta_{i+1}, I)}{p(X_{t,i}|D, \beta_i, I) p(X_{t,i+1}|D, \beta_{i+1}, I)}\right], \end{aligned} \quad (29)$$

where, at instant t , chain β_i is in state $X_{t,i}$ and chain β_{i+1} is in state $X_{t,i+1}$. By running such a set of cooperative chains, we effectively enable the algorithm to sample the target distribution in a way that allows for both the investigation of its overall features (low- β chains) and the examination of the fine details of a local maximum (high- β chains). A schematic representation of the functioning of the parallel tempering mechanism is shown in Fig. 2. In Fig. 3, a version of the Metropolis-Hastings algorithm is shown, written in pseudocode, and with the inclusion of the parallel tempering mechanism.

Concerning the values taken by the tempering parameter, $\{\beta_i\}$, optimal values are chosen in order to achieve a *swap acceptance rate* between adjacent levels of $\sim 50\%$. Heuristically, we can assert that by employing a geometric progression (cf. Benomar et al. 2009a),

$$\beta_i = \lambda^{1-i}, \quad (30)$$

such a desideratum is reached by setting $\lambda \sim 1.2$. The number of chains, n_β , should be chosen such as to reach a desired balance between sampling efficiency and computational time. However,

```

1: procedure PARALLEL TEMPERING METROPOLIS–HASTINGS
2:    $X_{0,i} = X_0, 1 \leq i \leq n_\beta$ 
3:   for  $t = 0, 1, \dots, n_{it} - 1$  do
4:     for  $i = 1, 2, \dots, n_\beta$  do
5:       Propose a new sample to be drawn from a
         proposal distribution:  $Y \sim N(X_{t,i}; \Sigma_i)$ 
6:       Compute the Metropolis ratio:
          $\ln r = \ln p(Y|D, \beta_i, I) - \ln p(X_{t,i}|D, \beta_i, I)$ 
7:       Sample a uniform random variable:
          $U_1 \sim \text{Uniform}(0, 1)$ 
8:       if  $\ln U_1 \leq \ln r$  then
9:          $X_{t+1,i} = Y$ 
10:      else
11:         $X_{t+1,i} = X_{t,i}$ 
12:      end if
13:    end for
14:     $U_2 \sim \text{Uniform}(0, 1)$ 
15:    if  $U_2 \leq 1/n_{\text{swap}}$  then
16:      Select random chain:
         $i \sim \text{UniformInt}(1, n_\beta - 1)$ 
17:      Compute  $r_{\text{swap}}$ :
         $\ln r_{\text{swap}} = \ln p(X_{t,i+1}|D, \beta_i, I) + \ln p(X_{t,i}|D, \beta_{i+1}, I)$ 
         $- \ln p(X_{t,i}|D, \beta_i, I) - \ln p(X_{t,i+1}|D, \beta_{i+1}, I)$ 
18:       $U_3 \sim \text{Uniform}(0, 1)$ 
19:      if  $\ln U_3 \leq \ln r_{\text{swap}}$  then
20:        Swap parameter states of chains  $i$  and  $i + 1$ :
         $X_{t,i} \leftrightarrow X_{t,i+1}$ 
21:      end if
22:    end if
23:  end for
24:  return  $X_{t,i}, \forall t, i: \beta_i = 1$ 
25: end procedure

```

Fig. 3. Version of the Metropolis-Hastings algorithm written in pseudocode and with the inclusion of parallel tempering.

when using the parallel tempering mechanism in model comparison problems, as we will get back to in Sect. 4.4, a large number of tempering chains are needed (typically $n_\beta \gtrsim 10$). The value of n_{swap} should be chosen inversely proportional to n_β (typically a few dozens).

4.3. Automated MCMC

So far we have not mentioned the need to adequately choose the set $\{\sigma\}$ of diagonal elements of the Σ matrix, indicating the width of the Gaussian proposal distribution for each parameter. The set of individual σ values specifies the direction and step size in parameter space when proposing a new sample to be drawn. The optimal choice of $\{\sigma\}$ is closely related to the average rate at which proposed state changes are accepted, the so-called *acceptance rate*. Accordingly, small σ values will lead to a large acceptance rate, with successive samples being highly correlated and ultimately requiring a large number of iterations in order to yield equilibrium distributions of model parameters. On the other hand, large σ values will lead to a low acceptance rate, meaning that proposed state changes will seldom be accepted. This is illustrated in Fig. 4, where the same simplified target distribution is sampled by three chains, each being characterised by a set of σ values differing on the respective magnitudes. Roberts et al. (1997) recommend, based on empirical studies, calibrating the acceptance rate to $\sim 25\%$ when dealing with a high-dimensional model as it is the case when performing a global peak-bagging.

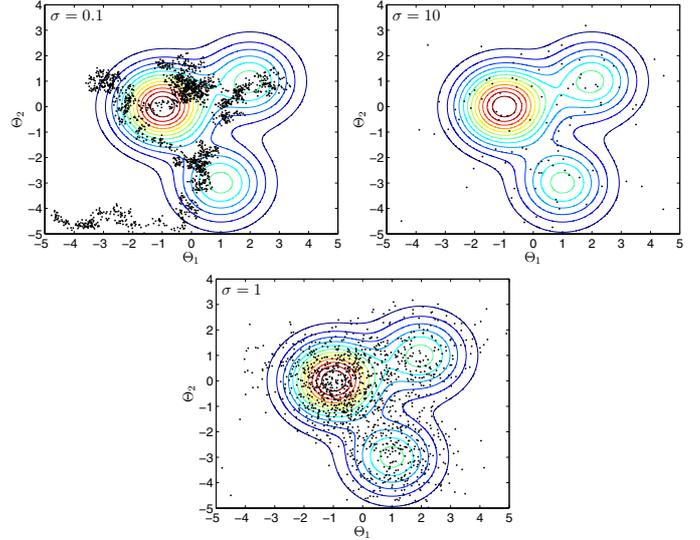


Fig. 4. The same target distribution sampled by three chains, each being characterised by a different set $\{\sigma\}$. The contours map the target distribution, which in turn depends only on the parameters Θ_1 and Θ_2 . The starting point in each of the chains is $(\Theta_1, \Theta_2) = (-4.5, -4.5)$ and all contain 2000 iterations. Both parameters share a common σ -value whose optimal setting is $\sigma = 1$. It is important to note that given a sufficiently large number of iterations, all the chains would eventually map out the target distribution, however an optimal choice of the proposal distribution will result in significantly faster convergence.

One could, of course, employ a trial-and-error approach and manually calibrate the σ values. However, since we are dealing with a large number of parameters that, in addition, correspond to several different physical quantities, this would quickly become very time-consuming and impractical. We instead employ an automated process of calibration of the proposal σ values, which is based on a statistical control system similar to the one described in Gregory (2005b). The control system makes use of an error signal to steer the selection of the σ values during the burn-in stage of a single parallel tempering MCMC run, acting independently on each of the tempered chains. The error signal is proportional to the difference between the current acceptance rate and the target acceptance rate. As soon as the error signal for each of the tempered chains is less than a measure of the Poisson fluctuation expected for a zero mean error (computed as the square root of the target acceptance rate times the number of iterations between changes in the σ values), the control system is turned off and the algorithm switches to the standard parallel tempering MCMC. In practice this effectively marks the end of the burn-in stage.

The control system as briefly described here is also used in Gruberbauer et al. (2009), whereas Benomar et al. (2009a) employ a self-learning process that appropriately adapts the covariance matrix, assumed non-diagonal.

4.4. Model comparison using parallel tempering MCMC

We are now interested in computing the odds ratio, O_{ij} , in favour of model M_i over model M_j according to Eq. (22). When analysing solar-like oscillations, a recurrent difficulty is to correctly tag the modes of oscillation by angular degree ℓ . There are two possible ways of tagging the modes or, equivalently, two competing models. Computation of O_{ij} is thus a means of assessing which of the two identification scenarios is statistically more

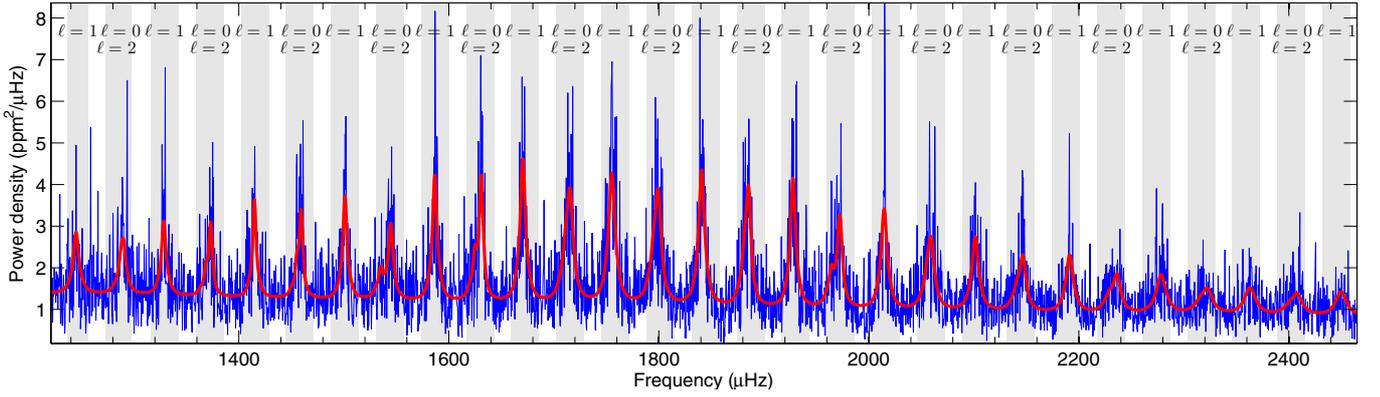


Fig. 5. Power density spectrum of HD 49933 with the best-fitting model (using prior set S2) overlaid. The shaded areas indicates the ranges of the uniform priors on the frequencies.

likely (although not necessarily physically more meaningful, as is often misinterpreted).

Samples drawn from the tempered distributions can, in principle, be used to compute the global likelihood, $p(D|M_i, I)$, of a given model M_i . Notice that Bayes' factor, B_{ij} , is defined as the ratio of the global likelihoods of two competing models:

$$B_{ij} \equiv \frac{p(D|M_i, I)}{p(D|M_j, I)} = \exp[\ln p(D|M_i, I) - \ln p(D|M_j, I)]. \quad (31)$$

It can be shown that the global likelihood of a model is given by (for a derivation see [Gregory 2005b](#)):

$$\ln p(D|M_i, I) = \int_0^1 \langle \ln p(D|M_i, X, I) \rangle_\beta d\beta, \quad (32)$$

where

$$\langle \ln p(D|M_i, X, I) \rangle_\beta = \frac{1}{n} \sum_t \ln p(D|M_i, X_{t,\beta}, I) \quad (33)$$

is the expectation value of the natural logarithm of the likelihood for a particular tempered chain characterised by β . The set $\{X_{t,\beta}\}$ represents the corresponding samples drawn after the burn-in stage, while n is the number of samples in each set. A sufficient number (≥ 10) of parallel tempered chains is required if we are to estimate the integral in Eq. (32) by interpolating values of $\langle \ln p(D|M_i, X, I) \rangle_\beta$.

5. Examples

In the following we will pick a couple of examples where we have applied the described Automated Parallel Tempering MCMC formalism to recent measurements of solar-like oscillators.

5.1. HD 49933: the importance of priors

We have performed an analysis of the star HD 49933, based on 180 days of photometry from the CoRoT satellite arising from two runs: The initial 60-day run, IRa01, and 120 days from the longer second run, LRa01. The time series was split up into segments of 30 days and the power spectra of the individual segments were averaged to construct a mean power spectrum ($s = 6$ in Eq. (2)). The acoustic spectrum of this F5 main-sequence star has proven to be very difficult to interpret mainly due to the relatively large linewidths (see Fig. 5). We assume the ridge identification denoted as ‘‘Scenario B’’ in [Benomar et al. \(2009b\)](#).

Table 3. Prior input for the HD 49933 analysis.

Parameter	Prior
Frequencies	Uniform
Heights	Modified Jeffreys
Linewidths	Uniform
Inclination	Uniform (0° – 90°)
Rotation	S1: Uniform on ν_s (0–10 μHz) S2: Gaussian on ν_s^* (1.65 ± 0.17 μHz)

The acoustic spectrum was fitted using the APT MCMC formalism, but using two different sets of priors (see Table 3). The first set (S1) was constructed using only ignorance priors, while the second set (S2) includes knowledge about the stellar rotation. From spectroscopic and asteroseismic studies of HD 49933, [Bruntt \(2009\)](#) was able to constrain the rotation of the star to $\nu \sin(i) = 10 \pm 1$ km s $^{-1}$ and the radius to $R/R_\odot = 1.385 \pm 0.031$, which can be combined to impose a constraint on the projected rotational splitting, $\nu_s^* = \nu_s \sin(i)$, of 1.65 ± 0.17 μHz . In set S2 this knowledge is added as a gaussian prior on the projected splitting of the star. In both cases the fits were done using the following configuration:

- 15 orders were fitted with $\ell = 0, 1, 2$ modes in a fitting window spanning from 1220 to 2465 μHz (see Fig. 5);
- one linewidth and one height per order assigned to the $\ell = 0$ mode, and then linearly interpolated by frequency and scaled to the higher degree modes;
- rotation and inclination angle fitted with the two free parameters, ν_s^* and i ;
- the background was parametrized as a sum of 3 Harvey-like models plus a white noise contribution;
- 800 000 samples were drawn from the target distribution, employing 10 parallel chains.

First of all, it is important to note that the results are consistent with the ones reported in [Benomar et al. \(2009b\)](#). For example the derived frequencies and linewidths are all well within the error bars. We will here focus on the results of the rotational splitting and inclination angle. The probability density functions for the fitted parameters when using ignorance priors (S1) are shown in Fig. 6a and, after applying the Gaussian prior on the projected splitting (S2), the results change to the ones shown in Fig. 6b.

What these results demonstrate is, first of all, that these techniques are extremely efficient at probing and constraining parameters which traditional methods would have considerable

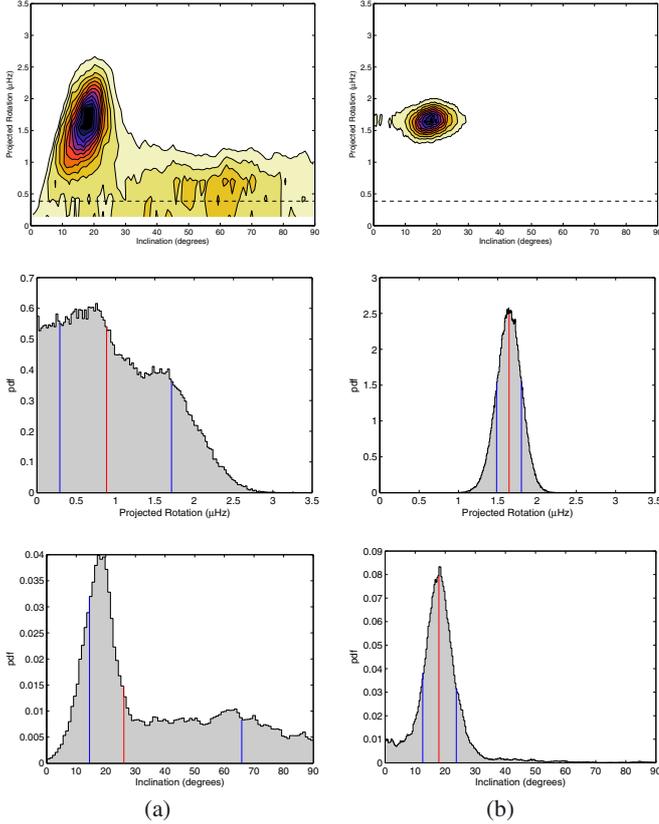


Fig. 6. Results on the rotation and inclination angle of HD 49933. In both cases the prior on the inclination angle is uniform in the interval 0° – 90° . **a)** We employ a uniform prior on rotation (S1). **b)** We employ a Gaussian prior on projected rotation (S2). The vertical lines in the histograms indicate the median and the boundaries of the 68% credible regions of the distributions. The dashed line in the top figures indicates the frequency resolution in the spectrum.

difficulty in constraining. In [Benomar et al. \(2009b\)](#), where similar techniques are employed, the derived inclination angle was $17^{+7}_{-9}^\circ$ and the rotational splitting placed in the range 3.5–6.0 μHz , with which our results are perfectly consistent with. Another point to be drawn from this example is the importance of the inclusion of prior knowledge. By incorporating our prior knowledge about the rotation of the star through the accurate measurements from a spectral analysis, we are able to yield a much cleaner constraint on the inclination angle, which in [Fig. 6a](#) has a considerable tail of probability towards large inclinations.

It is important to note that this prior on ν_s^* is strong in the sense that it dominates the fit. This simply comes from the fact that the data do not provide any further information on this parameter and so our prior knowledge of the model is still providing the best constraint. In such cases one of course has to be careful that such a strong prior is not wrongfully restrictive. If other effects (in this case, for instance, differential rotation or second-order rotational effects) were present, this could introduce biases to the fitted results. This could of course be tested using the methodology described in [Sects. 3.2 and 4.4](#), by constructing models that incorporate these effects and testing their significance.

5.2. Procyon: the problem of ridge identification

Here we address the issue of tagging the oscillation modes by angular degree in the case of the F5 star Procyon. We have thus reanalysed the data acquired during a multi-site campaign ([Arentoft et al. 2008](#); [Bedding et al. 2010b](#)) carried to observe oscillations in this star. The data consist of high-precision velocity observations obtained over more than three weeks with eleven telescopes, representing the most extensive campaign organised so far on a solar-type oscillator.

The problem of ridge identification in F stars dates back to when CoRoT observations of HD 49933 were first analysed ([Appourchaux et al. 2008](#)), a problem that would be recently solved for this star only after a new longer time series was made available ([Benomar et al. 2009b](#)). [Bedding et al. \(2010b\)](#) address this same problematic in the case of Procyon by employing three distinct methodologies: (i) a collapsed power spectrum along several radial orders; (ii) a scaled échelle diagram ([Bedding & Kjeldsen 2010](#)); and (iii) Bayesian model comparison (as described in [Sect. 4.4](#)). The last-mentioned methodology statistically favours their Scenario A over their Scenario B identification, whereas the first and second methodologies suggest the contrary although without quantifying their preference for Scenario B in a statistical sense.

We performed a peak-bagging of the power spectrum of Procyon considering both identification scenarios while simultaneously testing for the presence of $\ell = 3$ modes. This gives a total of four competing models, i.e., $\{M_A, M_A^{\ell=3}, M_B, M_B^{\ell=3}\}$, the notation chosen to be unambiguous. The details of the peak-bagging as implemented here slightly differ from those presented in [Bedding et al. \(2010b\)](#), and especially concern the limits of the fitting window and the way in which the background was parametrized. The details are as follows:

- The peak-bagging was performed on the sidelobe-optimised power density spectrum whose intrinsic frequency resolution is 0.77 μHz . Peaks were described by symmetric Lorentzians centred on the mode frequencies. Three frequencies were fitted per overtone, each with a different angular degree ($\ell = 0, 1, 2$). Type of prior imposed at first: independent and uniform, centred ($\pm 8 \mu\text{Hz}$) on the initial guesses. The mode frequencies were further constrained to lie close to the ridge centroids and to have only small jumps from one order to the next. Also, a Gaussian prior ($\mu = 4 \mu\text{Hz}$, $\sigma = 5 \mu\text{Hz}$) was imposed on the small frequency separation, $\delta\nu_{02}$, between adjacent modes with $\ell = 0$ and $\ell = 2$. The small separation was not itself a free parameter in the fit, but instead a derived quantity. Note that the last-mentioned constraint implies that the type of prior on the $\ell = 0, 2$ frequency parameters is ultimately not independent nor uniform. Optionally, modes with $\ell = 3$ could be included in the model with their frequencies fixed to

$$\nu_{n-1,3} = \nu_{n,1} - \frac{5}{3}(\nu_{n,0} - \nu_{n-1,2}), \quad (34)$$

according to the asymptotic relation ([Tassoul 1980](#)). A total of 14 overtones were considered and the fitting window runs from 500 to 1300 μHz . By employing this construction it is assumed that no mixed modes are present in the fitting window. The inclusion of $\ell = 3$ modes does not add any more free parameters, while adding however their features to the model spectrum which are derived from the fitted $\ell = 0, 1, 2$ frequencies.

- The linewidth was parametrized as a linear function of frequency, defined by two parameters Γ_{600} and Γ_{1200} , which are

Table 4. Model probabilities.

Model	$\ln p(D \text{Model}, I)$	Probability
M_A	2789.723	39.25%
$M_A^{\ell=3}$	2790.046	54.23%
M_B	2785.806	0.78%
$M_B^{\ell=3}$	2787.801	5.74%

the values at 600 and 1200 μHz , respectively. Both parameters were fitted. Type of prior imposed: uniform in the range 0–10 μHz .

- The heights of radial modes in units of power density were fixed according to [Chaplin et al. \(2008b\)](#):

$$S_{n0} = \frac{2A^2T}{\pi T \Gamma_{n0} + 2}, \quad (35)$$

where A^2 is the total power of the mode as determined from the power envelope for radial modes ([Kjeldsen et al. 2008](#)), and T is the effective length of the observational run. The heights of non-radial modes were then defined based on the heights of radial modes according to Eq. (12), and taking into account the appropriate V_ℓ/V_0 ratios given in Table 1 of [Kjeldsen et al. \(2008\)](#).

- The background was parametrized as a linear function of frequency since it had previously been suppressed at low frequencies (high-pass cut at 280 μHz) to effectively remove slow variations.
- The inclination angle between the direction of the stellar rotation axis and the line of sight was fixed at 31.1° , which is the inclination of the binary orbit and is consistent with the rotational modulation of the velocity curve. The rotational splitting was fixed at 0.7 μHz , which was chosen to match the observed value of $v \sin(i) = 3.16 \text{ km s}^{-1}$ ([Allende Prieto et al. 2002](#)), given the known radius of the star of $2.05 R_\odot$ ([Kervella et al. 2004](#)).
- We drew $\sim 800\,000$ samples from the target distribution after a burn-in phase. We employed 12 parallel tempered chains.
- We thus have a total of 46 free parameters, namely, 42 frequencies, 2 parameters for the linewidth and 2 parameters for the background.

Table 4 summarises the model selection calculations assuming equal prior probabilities for the models belonging to our discrete model space. Individual probabilities are assigned to models according to Eq. (21). Similarly to [Bedding et al. \(2010b\)](#), Bayesian model comparison again statistically favours Scenario A over Scenario B. Furthermore, the presence of residual power due to $\ell = 3$ modes is suggested. Computing Bayes' factor in favour of model $M_A^{\ell=3}$ over model $M_B^{\ell=3}$ gives a factor of approximately 9:1 or, equivalently, a logarithmic factor of 2.2, which classifies as “significant” on Jeffreys' scale. Figure 7 displays the power density spectrum of Procyon in échelle format with the fitted frequencies for model $M_A^{\ell=3}$ overlaid.

6. Summary and discussion

In this paper, we have presented the basic theory and methods behind the extraction of parameters from the power spectra of solar-like stars. In order to handle the ever rising quality and complexity of modern asteroseismic data, we have developed a tool (APT MCMC) that enables us to constrain parameters associated with the subtlest features in the spectra. The algorithm has been extensively tested and performs extremely well,

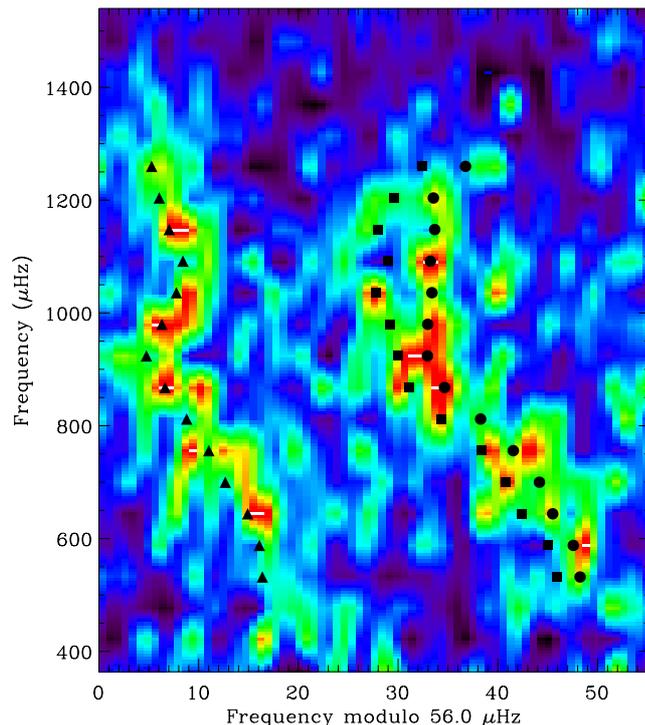


Fig. 7. Power density spectrum of Procyon (smoothed to 2 μHz) in échelle format. The fitted frequencies for model $M_A^{\ell=3}$ appear as overlaid filled symbols. Symbol shapes indicate mode degree: $\ell = 0$ (circles), $\ell = 1$ (triangles), $\ell = 2$ (squares).

not only in the traditional case of extracting oscillation frequencies, but also when pushing the limit where traditional methods have difficulties, such as constraining linewidths, rotational splittings and stellar inclination angles. In this work we have focused on data in the signal-to-noise regime of current asteroseismic measurements. In the case of very high signal-to-noise ratios, other features in the power spectrum becomes important, such as mode asymmetries and rotational splittings dependent on ℓ , arising from differential rotation with radius. In future work these effects will be incorporated into the program and tested on solar data.

One disadvantage of the method is that it can be quite computationally intensive, both to implement and run, when compared to traditional MLE fits. This is however balanced by the much added information outputted from the fits, specifically in the probability distributions of each parameter, making it easy to obtain accurate, reliable and realistic error bars on the results – a feature seriously missing from the traditional methods. The parameter estimation also benefits enormously from the possibilities the Bayesian formalism provides with inclusion of prior information. This not only allows control of the fit to, for example, not allow unphysical parameter combinations, but also include information into the fit that is better constrained by other measurements (as we saw in Sect. 5.1). Another powerful feature of the method lies in the parallel tempering, which not only keeps the fits from getting stuck in local maxima, but also provides an objective way of comparing different competing models, as it provides a way of calculating the global likelihood. This can for example be utilized in the familiar problem of ridge identification in solar-like stars (see Sect. 5.2).

A thing to keep in mind is also that the APT MCMC algorithm is completely general, in the sense that it could be applied to other problems without modification. MCMC methods are being used in various branches of astrophysics: cosmology (Liddle 2009), extra solar planets (Gregory 2005a) and stellar model fitting (Bazot et al. 2008), but in fact the methods would be applicable in any problem including parameter estimation. And as computational power continues to grow, the downsides are quickly becoming insignificant.

What could to some extent also be seen as a disadvantage of these methods is that they can never be fully automated, in the sense that they will not be able to handle a large number of stars without human interaction. The whole fundamental idea behind the Bayesian formalism is that it relies on “wise” human inputs on the priors and model setup that should not be done in an automated way. If nothing else, take this as a positive reassurance: You will, as an astrophysicist, never be obsolete to computers or monkeys with keyboards.

Acknowledgements. We would like to thank H. Kjeldsen, T. R. Bedding, C. Karoff, W. J. Chaplin, T. Appourchaux, R. A. García and O. Benomar for fruitful discussions and valuable comments. We also thank the anonymous referee for comments that helped improve the paper. T.L.C. is supported by grant with reference number SFRH/BD/36240/2007 from FCT/MCTES, Portugal. This work was supported by the project PTDC/CTE-AST/098754/2008 funded by FCT/MCTES, Portugal.

Appendix A: Computing $\mathcal{E}_{\ell m}(i)$ and V_ℓ

The $\mathcal{E}_{\ell m}(i)$ factors are given below for $\ell \in [0, 4]$, having used Eq. (9):

$$\begin{aligned} \mathcal{E}_{0,0}(i) &= 1, \\ \mathcal{E}_{1,0}(i) &= \cos^2 i, \\ \mathcal{E}_{1,\pm 1}(i) &= \frac{1}{2} \sin^2 i, \\ \mathcal{E}_{2,0}(i) &= \frac{1}{4} (3 \cos^2 i - 1)^2, \\ \mathcal{E}_{2,\pm 1}(i) &= \frac{3}{8} \sin^2(2i), \\ \mathcal{E}_{2,\pm 2}(i) &= \frac{3}{8} \sin^4 i, \\ \mathcal{E}_{3,0}(i) &= \frac{1}{64} (5 \cos(3i) + 3 \cos i)^2, \\ \mathcal{E}_{3,\pm 1}(i) &= \frac{3}{64} (5 \cos(2i) + 3)^2 \sin^2 i, \\ \mathcal{E}_{3,\pm 2}(i) &= \frac{15}{8} \cos^2 i \sin^4 i, \\ \mathcal{E}_{3,\pm 3}(i) &= \frac{5}{16} \sin^6 i, \\ \mathcal{E}_{4,0}(i) &= \frac{1}{64} (35 \cos^4 i - 30 \cos^2 i + 3)^2, \\ \mathcal{E}_{4,\pm 1}(i) &= \frac{5}{256} (\frac{7}{2} \sin(4i) + \sin(2i))^2, \\ \mathcal{E}_{4,\pm 2}(i) &= \frac{5}{128} (7 \cos(2i) + 5)^2 \sin^4 i, \\ \mathcal{E}_{4,\pm 3}(i) &= \frac{35}{16} \cos^2 i \sin^6 i, \\ \mathcal{E}_{4,\pm 4}(i) &= \frac{35}{128} \sin^8 i. \end{aligned} \tag{A.1}$$

Notice that when the rotation axis is aligned with the line of sight ($i = 0^\circ$), only the multiplet component with $m = 0$ is visible, thus making inviable an inference of rotation.

The spatial response function for each ℓ , V_ℓ , representing the ratio of the observed amplitude to the actual amplitude, is given

here for the five lowest degree modes (Bedding et al. 1996):

$$\begin{pmatrix} V_0 \\ V_1 \\ V_2 \\ V_3 \\ V_4 \end{pmatrix} = \begin{pmatrix} 1 & \frac{2}{3} & \frac{1}{2} & \frac{2}{5} \\ \frac{2}{\sqrt{3}} & \frac{\sqrt{3}}{2} & \frac{2\sqrt{3}}{5} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{5}}{4} & \frac{4}{3\sqrt{5}} & \frac{\sqrt{5}}{4} & \frac{8}{7\sqrt{5}} \\ 0 & \frac{\sqrt{7}}{12} & \frac{4}{5\sqrt{7}} & \frac{\sqrt{7}}{8} \\ -\frac{1}{8} & 0 & \frac{3}{32} & \frac{16}{105} \end{pmatrix} \times \begin{pmatrix} 1-c & c-1 & c-1 \\ c & 1-2c & -c \\ 0 & c & 1-c \\ 0 & 0 & c \end{pmatrix} \times \begin{pmatrix} 1 \\ u_2 \\ v_2 \end{pmatrix}, \tag{A.2}$$

where u_2 and v_2 are wavelength-dependent classical limb-darkening coefficients (Allen 1973) and c is a parameter defining the observational method. This matrix product can be used for velocity measurements by setting $c = 1$ and for intensity measurements by setting $c = 0$.

References

- Abrams, D., & Kumar, P. 1996, ApJ, 472, 882
Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, Asteroseismology, ed. C. Aerts, J. Christensen-Dalsgaard, & D. W. Kurtz (Springer)
Aigrain, S., Favata, F., & Gilmore, G. 2004, A&A, 414, 1139
Allen, C. W. 1973, Astrophysical quantities, 3rd edn., ed. C. W. Allen (Athlone Press)
Allende Prieto, C., Asplund, M., García López, R. J., & Lambert, D. L. 2002, ApJ, 567, 544
Anderson, E. R., Duvall, Jr., T. L., & Jefferies, S. M. 1990, ApJ, 364, 699
Appourchaux, T. 2003a, A&A, 412, 903
Appourchaux, T. 2003b, Ap&SS, 284, 109
Appourchaux, T. 2004, A&A, 428, 1039
Appourchaux, T., Berthomieu, G., Michel, E., et al. 2006, in ESA SP-10306, ed. M. Fridlund, A. Baglin, J. Lochard, & L. Conroy, 429
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
Ballot, J. 2010, Astron. Nachr., 331, 933
Ballot, J., García, R. A., & Lambert, P. 2006, MNRAS, 369, 1281
Ballot, J., Appourchaux, T., Toutain, T., & Guittet, M. 2008, A&A, 486, 867
Bazot, M., Bourguignon, S., & Christensen-Dalsgaard, J. 2008, Mem. Soc. Astron. Ital., 79, 660
Bedding, T. R., & Kjeldsen, H. 2008, in 14th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. G. van Belle, ASP Conf. Ser., 384, 21
Bedding, T. R., & Kjeldsen, H. 2010, Commun. Asteroseismol., 161, 3
Bedding, T. R., Kjeldsen, H., Reetz, J., & Barbuy, B. 1996, MNRAS, 280, 1155
Bedding, T. R., Huber, D., Stello, D., et al. 2010a, ApJ, 713, L176
Bedding, T. R., Kjeldsen, H., Campante, T. L., et al. 2010b, ApJ, 713, 935
Benomar, O., Appourchaux, T., & Baudin, F. 2009a, A&A, 506, 15
Benomar, O., Baudin, F., Campante, T. L., et al. 2009b, A&A, 507, L13
Bruntt, H. 2009, A&A, 506, 235
Campante, T. L., Karoff, C., Chaplin, W. J., et al. 2010, MNRAS, 408, 542
Chaplin, W. J., Appourchaux, T., Arentoft, T., et al. 2008a, Astron. Nachr., 329, 549
Chaplin, W. J., Houdek, G., Appourchaux, T., et al. 2008b, A&A, 485, 813
Chaplin, W. J., Appourchaux, T., Elsworth, Y., et al. 2010, ApJ, 713, L169
Christensen-Dalsgaard, J. 2003, Lecture Notes on Stellar Oscillations
Christensen-Dalsgaard, J. 2004, Sol. Phys., 220, 137
Christensen-Dalsgaard, J., Kjeldsen, H., Brown, T. M., et al. 2010, ApJ, 713, L164
De Ridder, J., Barban, C., Baudin, F., et al. 2009, Nature, 459, 398
Dupret, M., Belkacem, K., Samadi, R., et al. 2009, A&A, 506, 57
Duvall, Jr., T. L., & Harvey, J. W. 1986, in Seismology of the Sun and the Distant Stars, ed. D. O. Gough, NATO ASIC Proc. 169, 105
Duvall, Jr., T. L., Jefferies, S. M., Harvey, J. W., Osaki, Y., & Pomerantz, M. A. 1993, ApJ, 410, 829
Earl, D. J., & Deem, M. W. 2005, Phys. Chem. Chem. Phys. (Incorporating Faraday Transactions), 7, 3910
Fröhlich, C., Romero, J., Roth, H., et al. 1995, Sol. Phys., 162, 101
Gabriel, M. 1994, A&A, 287, 685
Gelman, A., & Rubin, D. 1992, Stat. Sci., 7, 457
Gilliland, R. L., Brown, T. M., Christensen-Dalsgaard, J., et al. 2010, PASP, 122, 131
Gizon, L., & Solanki, S. K. 2003, ApJ, 589, 1009
Gregory, P. C. 2005a, ApJ, 631, 1198

- Gregory, P. C. 2005b, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with “Mathematica” Support*, ed. P. C. Gregory (Cambridge University Press)
- Gruberbauer, M., Kallinger, T., Weiss, W. W., & Guenther, D. B. 2009, *A&A*, 506, 1043
- Grundahl, F., Kjeldsen, H., Christensen-Dalsgaard, J., Arentoft, T., & Frandsen, S. 2007, *Commun. Asteroseismol.*, 150, 300
- Harvey, J. 1985, *High-resolution helioseismology*, Tech. Rep.
- Hastings, W. K. 1970, *Biometrika*, 57, 97
- Hekker, S., Broomhall, A., Chaplin, W. J., et al. 2010a, *MNRAS*, 402, 2049
- Hekker, S., Debosscher, J., Huber, D., et al. 2010b, *ApJ*, 713, L187
- Huber, D., Stello, D., Bedding, T. R., et al. 2009, *Commun. Asteroseismol.*, 160, 74
- Jeffreys, H. 1961, *Theory of Probability*, 3rd edn. (Oxford University Press)
- Karoff, C., Metcalfe, T. S., Chaplin, W. J., et al. 2009, *MNRAS*, 399, 914
- Karoff, C., Campante, T. L., & Chaplin, W. J. 2010, *Astron. Nachr.*, in press [arXiv:1003.4167v1]
- Kervella, P., Thévenin, F., Morel, P., et al. 2004, *A&A*, 413, 251
- Kjeldsen, H., Bedding, T. R., Arentoft, T., et al. 2008, *ApJ*, 682, 1370
- Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, *ApJ*, 713, L79
- Kumar, P., Franklin, J., & Goldreich, P. 1988, *ApJ*, 328, 879
- Ledoux, P. 1951, *ApJ*, 114, 373
- Liddle, A. R. 2009, *Ann. Rev. Nucl. Part. Sci.*, 59, 95
- Mathur, S., García, R. A., Régulo, C., et al. 2010, *A&A*, 511, A46
- Mayor, M., Pepe, F., Queloz, D., et al. 2003, *The Messenger*, 114, 20
- Metcalfe, T. S., Monteiro, M. J. P. F. G., Thompson, M. J., et al. 2010, *ApJ*, 723, 1583
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *J. Chem. Phys.*, 21, 1087
- Michel, E., Baglin, A., Auvergne, M., et al. 2008, *Science*, 322, 558
- Mosser, B., & Appourchaux, T. 2009, *A&A*, 508, 877
- Roberts, G. O. 1996, in *Markov Chain Monte Carlo in Practice*, (London: Chapman and Hall), 45
- Roberts, G. O., Gelman, A., & Gilks, W. R. 1997, *Ann. Appl. Prob.*, 7, 110
- Scott, D. W. 1979, *Biometrika*, 66, 605
- Stahn, T., & Gizon, L. 2008, *Sol. Phys.*, 251, 31
- Stello, D., Basu, S., Bruntt, H., et al. 2010, *ApJ*, 713, L182
- Tassoul, M. 1980, *ApJS*, 43, 469
- Toutain, T., & Appourchaux, T. 1994, *A&A*, 289, 649
- Woodard, M. F. 1984, Ph.D. Thesis, AA (San Diego: University of California)