BAYESIAN ASTEROSEISMOLOGY

by

Michael Gruberbauer

A thesis submitted to the faculty of
Saint Mary's University,
Halifax, Nova Scotia, Canada
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Astronomy

Approved: Dr. David B. Guenther, Advisor

Approved: Dr. Robert Thacker, Reader

Approved: Dr. Robert Deupree, Reader

Approved: Dr. Travis Metcalfe, External Examiner

Department of Astronomy and Physics
Saint Mary's University
June 7, 2013

ABSTRACT

BAYESIAN ASTEROSEISMOLOGY

by Michael Gruberbauer

June 7, 2013

This thesis presents a new probabilistic method for the asteroseismic analysis of stellar structure and evolution with the goal of providing a universal tool to improve our knowledge of stellar modelling. This new method implements the advantages of Bayesian analysis, such as the treatment of systematic errors and nuisance parameters, the modular structure of Bayesian analysis, and the correct normalization of all probabilities.

First, a general introduction to asteroseismology is provided, followed by an comprehensive guide to Bayesian analysis. The derivation of the new method then follows, and its subsequent application to current problems in asteroseismology is also presented. An in-depth analysis of the Sun is performed in order to investigate long standing problems with the solar chemical composition. This also reveals the presence of systematic problems in the modelling of the Sun, potentially requiring new developments in solar modelling. Finally, the new method is also applied to 23 stars that were observed with the Kepler satellite, in order to perform a comparative investigation with respect to published results from other teams, and to study systematic errors in the stellar models.

This thesis is dedicated to the memory of Dr. Peter "Piet" Reegen.

Contents

List of Tables

List of Figures

xiii

Chapter 1

Introduction I: Asteroseismology in the Age of Space Telescopes

Although hard data is lacking, it is probably safe to assume that a large percentage of PhD theses dedicated to the study of stellar structure and evolution start something like this: *For countless millennia human beings have looked at the sky and wondered about the nature of the Sun and the stars.* This is certainly a cliché, but it is so appealing because it is true, and it puts into perspective the enormous amount of understanding humans have gained about the stars in the last two centuries. It is illuminating to recall that the idea of gaining in-depth knowledge by probing the stellar interiors in some way (or even to realise that there is such a thing as "stellar interiors") must have seemed preposterous. This is nicely indicated by the well-known passage in Auguste Comte's "The Positive Philosophy" (1842)

> *To attain a true idea of the nature and composition of [astronomy], it is indispensable ... to mark the boundaries of the positive knowledge that we are able to gain of the stars. ... We can never by any means investigate their chemical composition.*

Today this notion is of course completely obsolete, since at least the basic (chemical) characteristics of stars have been accessible through spectroscopy of the stellar photospheres. Studying the deep interiors of stars observationally[1], on the other hand, has only recently become viable thanks to asteroseismology - studying the stars by understanding their oscillations.

---

[1]The first comprehensive work on stellar interiors (Eddington 1926) was, although reliant on the astronomical observations of the time, basically founded on theory.

The purpose of this chapter is to convey an understanding of the observational properties of stellar oscillations which can then be studied statistically for many stars or, alternatively, compared in more detail to theoretical models. Section 1.1 will elaborate on the fundamental characteristics and observables. Since this thesis is mostly concerned with the application of Bayesian methods to the analysis of solar-like oscillations, this section will not discuss other types of pulsations in great detail. Although the tools and methods developed during the course of my work are applicable to asteroseismic data gathered with many different methods, the current state-of-the-art data sets are almost exclusively obtained with space photometry. Consequently, section 1.2 will discuss how these observables are obtained from space photometry today. Finally, section 1.3 will briefly give recent examples for how these data are used to better understand stellar structure and evolution.

1.1   Fundamental aspects of stellar oscillations

1.1.1   Spherical harmonics and mode nomenclature

The first step to understanding stellar oscillation modes pertains to their description. The displacement and motion within the star produced by pulsation and, consequently, the brightness variations on the stellar surface can be expressed through spherical harmonics (see, e.g., the recent textbook by Aerts et al. 2010). These are defined by three quantum numbers: n, l, and m. n counts the number of nodes in the radial direction, l gives the number of nodes on the surface (the spherical degree), and $-l \leq m \leq l$ denotes the number of meridional nodes. Oscillations with $l = 0$ (and therefore $m = 0$) are called radial pulsations, where the mode with $n = 0$ is called the fundamental mode, followed by the first overtone with $n = 1$, and so on. It is these modes that have been known for the longest time and have been used in Cepheids and RR Lyrae stars to establish distance scales using the period-luminosity relation first discovered by Leavitt & Pickering (1912) and still studied and refined today (e.g. Bono et al. 2010). Modes with nodal lines on the surface are called nonradial modes.
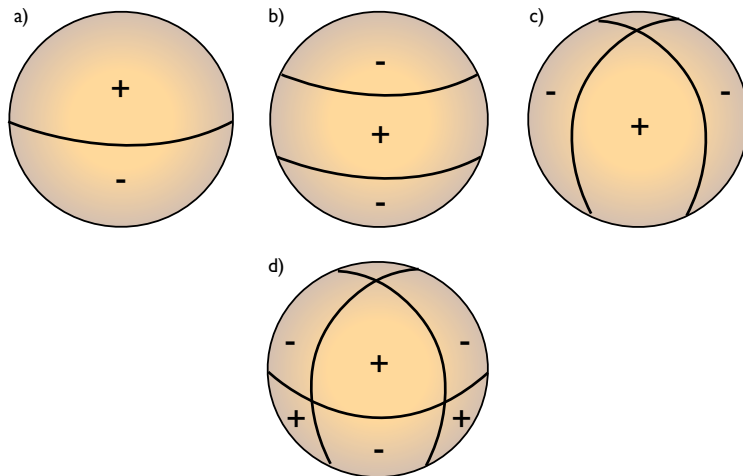
Figure 1.1: Cartoon representation of spherical harmonics. a) l=1, m=0; b) l=2, m=0; c)l=2, m=2; d) l=3, m=2. Signs represent the inward and outward motion at a snapshot in time.

For non-rotating stars, pulsations with $m \neq 0$ are not distinguishable from the $m = 0$ solution since the pulsation frequencies are identical. Once rotation is taken into account, however, *rotational splitting* can be observed. The splitting removes the degeneracy between the modes with different m and the corresponding frequencies can be described with

$$\nu_{n,l,m \neq 0} = \nu_{n,l,m=0} + mk\Omega, \tag{1.1}$$

where Ω is the rotation rate and $k \approx 1$. For faster rotators that can become substantially non-spherical, this approximation, as well as the description using single spherical harmonics, no longer holds. In such cases, for instance a linear combination of various spherical harmonics can be used to describe the pulsation. Here it starts to become problematic to uniquely identify each mode (Deupree & Beslin 2010). Similar expansions are necessary for pulsating stars with strong magnetic fields, resulting in comparable mode identification problems (Saio 2005; Cunha 2005).

Stellar pulsations can be understood as (damped and re-excited) harmonic oscillations. Depending on the restoring force, the modes that are commonly observed

in stars are either called p modes (pressure as restoring force), g modes (gravity modes - buoyancy in addition to pressure contributes to the restoring force), or r modes. For the latter, which are commonly called Rossby waves, the restoring force is the variation of the Coriolis force as a function of latitude, similar to the planetary waves on the Earth. The most common modes are the p and g modes. What distinguishes them is that the pressure modes are predominantly probing the bulk of the star, while the g modes are more sensitive to the core (see, e.g., Aerts et al. 2010, for more details). An effect termed "avoided crossing", where the frequency of a g mode in the core approaches the frequency of a p mode in the envelope due to evolutionary changes in the stellar structure, can lead to perturbations of the p-mode frequency. This results in so-called "mixed" or "bumped" modes that display theoretical characteristics of p modes in the outer layers and g modes in the interior. They are very sensitive to the evolutionary state or age of the star and can easily be distinguished from pure p modes (e.g., Bedding 2012).

1.1.2   Pulsation across the HR diagram and mode excitation

Different types of stars show different types of oscillations. For instance, Cepheids and RR Lyrae stars pulsate (predominantly) radially at very large amplitudes. They usually pulsate in only one or two modes with periods of several hours to days, although recent space missions have provided evidence for more radial and even nonradial pulsation at lower amplitudes (e.g., Gruberbauer et al. 2007; Guggenberger et al. 2012). Their long periods and large amplitudes are also the reason why they were among the first pulsating stars to be studied in great detail. δ Scuti stars, on the hand, pulsate in either radial or nonradial modes at slightly higher orders and at shorter periods down to several tens of minutes. Between a few and several dozen significant frequencies can usually be detected, even from ground-based observations (e.g. Breger et al. 2005). Recent space-based data have revealed even more frequencies although their nature is debated (see section 1.3). Since δ Scuti stars can show substantial rotation rates, which also appear to influence mode selection and mode

amplitudes, rotational splitting is observed, and other interesting features such as frequency multiplets and combination frequencies complicate the picture (Breger & Kolenberg 2006). The magnetic, rapidly oscillating Ap stars show nonradial pulsations at even higher radial orders with periods of minutes (Kurtz 1982).

The pulsations in these stars are thought to be excited by the so-called κ mechanism. As the name suggests, it is related to the radiative opacity of the material, and it is believed that this opacity, in particular in the ionization zones of hydrogen, helium, and some metals, provides the means to build up energy during contraction in a way as to perpetuate pulsation (see, e.g., Hansen et al. 2004). A different mechanism, called the ϵ mechanism for which the increased nuclear reaction rate during contraction is important, might also be responsible to drive pulsations in a subset of stars (see, for example Miller Bertolami et al. 2011). For γ Doradus stars, "convective blocking" (Guzik et al. 2000) has been found as a possible explanation to excite the g-mode pulsations found in these stars (and also in the many γ Dor-δ Scuti hybrid pulsators which have been found thanks to recent space missions).

In recent years, the discovery of solar-type oscillations in many different types of stars has been among the most exciting findings in stellar physics. These pulsations are named after the fact that they have first been observed in the Sun. They are thought to be p modes (and mixed modes) excited stochastically by the acoustic noise of convection and granulation which affects their observational properties, as discussed in section 1.2. Solar-type oscillations are now believed to be excited in all stars with convective envelopes, i.e., cool stars, from the main sequence up to the giant branch. As will be discussed in section 1.3, thanks to MOST, CoRoT and Kepler pulsation in thousands of red giants have now been analysed.

Aside from the types mentioned so far, many more named types of pulsating stars exist, so that the term "zoo" readily applies. For the purposes of this thesis, however, the types mentioned and the distinctions made so far are enough to proceed. More details on other types of pulsating stars can be found, e.g., in Aerts et al. (2010).

1.1.3 The asymptotic relation and scaling laws

The most important property of solar-type oscillations from the perspective of asteroseismology, is the fact that they excite modes at high radial orders, similar to the roAp stars, but without the complications of the magnetic field. Thanks to the approximate validity of the so-called asymptotic relation (Vandakurov 1967; Tassoul 1980)

$$\nu_{n,l} \approx \Delta\nu \left(n + \frac{l}{2} + \epsilon \right) - \delta\nu_{0l}, \tag{1.2}$$

at such high orders the p-mode pulsations show an almost uniform pattern. Modes of the same spherical degree l but adjacent radial order n are almost evenly spaced by $\Delta\nu$, which is called the characteristic frequency spacing. It is related to the run of the sound speed c_s through the star as in

$$\Delta\nu \approx \left(2 \int_0^R \frac{dr}{c_s} \right)^{-1}. \tag{1.3}$$

In practise, the characteristic frequency spacing is also sometimes identified with the large frequency spacing $\nu_{n+1,l} - \nu_{n,l}$ of a particular mode or its average over several orders for different modes, which is only approximately correct.

$\delta\nu_{0l} = \nu_{n,0} - \nu_{n+1,l}$ is called the small spacing or small separation. These patterns can be used to better visualize the structure in the pulsation spectrum by employing so-called échelle diagrams (Grec et al. 1983), in which the frequencies are plotted against their values modulo $\Delta\nu$. If the asymptotic relation were exact, all modes of the same spherical degree would line up on vertical ridges. Hence, the échelle diagram easily reveals such features as avoided crossings or other features (including instrumental artefacts) which do not line up with the expected pattern.

Figure 1.2 shows an exemplary échelle diagram for the Kepler target KIC 7976303, for which low-degree solar-type oscillations have been detected. It is obvious that the $l = 0$ and $l = 2$ modes correspond nicely to the asymptotic relation as they stack almost vertically. The $l = 1$ modes, on the other hand, are clearly affected by mode bumping, which produces deviations from the asymptotic relation.
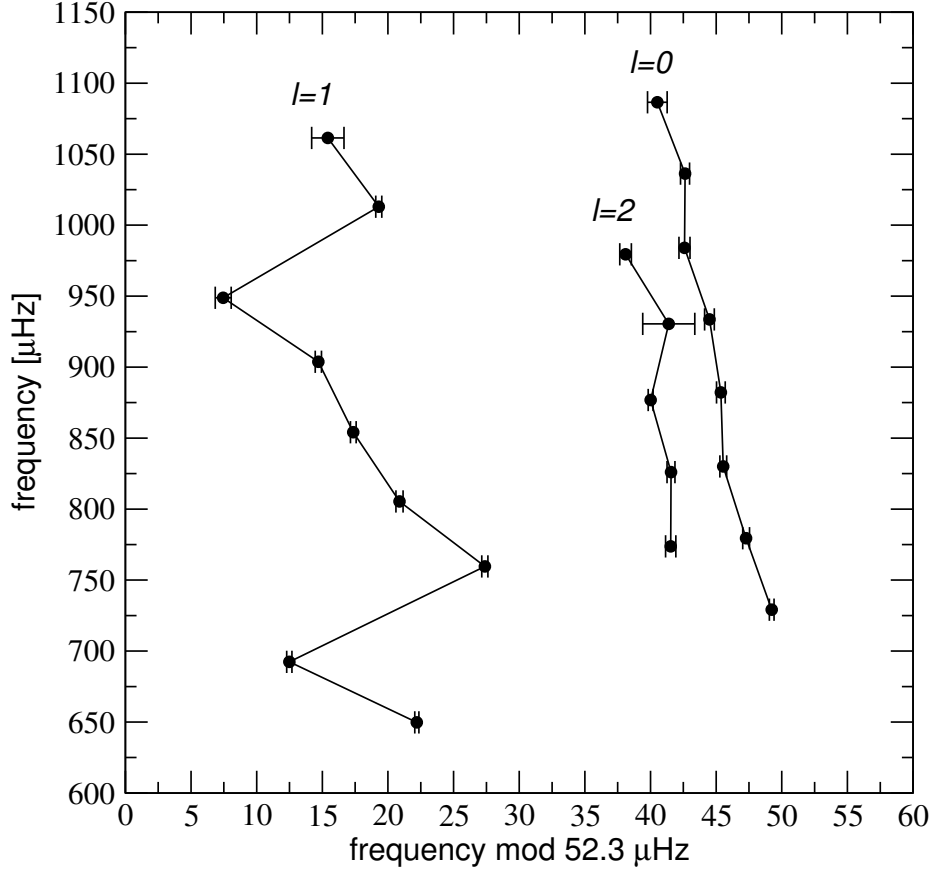
Figure 1.2: Échelle diagram of KIC 7976303 frequencies detected by the Kepler satellite (Mathur et al. 2012). Individual spherical degrees are indicated by labels and lines in between data points.

Compared to similar data obtained for the Sun (see, e.g., Figure 3.1 in Chapter 3), the frequency uncertainties are larger by about an order of magnitude and, overall, fewer modes can be detected, as will be explained in the next section.

Aside from employing $\Delta\nu$ to reveal the structure of the oscillation spectrum, it is also central to one of several important scaling relations. Using basic assumptions about stellar interiors, one can show that

$$\Delta\nu \propto \sqrt{M/R^3}. \tag{1.4}$$

Furthermore, for solar-type oscillations, the maximum pulsation amplitude occurs around the so-called *frequency of maximum oscillation power* $\nu_{\max}$ which is thought

7

to scale approximately as

$$\nu_{\max} \propto \frac{M}{R^2 \sqrt{T_{\mathrm{eff}}}}. \tag{1.5}$$

Therefore, knowing the approximate solar quantities and measuring the corresponding values in other stars (which is possible even in cases where individual modes are more difficult to extract) is helpful to further constrain the stellar properties. These relations and others for the amplitude of the pulsations as a function of stellar parameters, are described by Kjeldsen & Bedding (1995) in more detail. Recent results on the scaling relations will be discussed in the next section.

How the individual modes themselves are used to infer stellar properties is a major topic of this thesis and will be explained in detail in Chapter 3. First, however, it is helpful to understand how these observables are extracted from modern space-based data sets to better judge their reliability.

1.2 Obtaining pulsation characteristics from space photometry

1.2.1 The limits of detection and the need for satellites

Although for the Sun many thousands of modes of very high spherical degree have been detected via photometry as well as radial velocity variations, the situation is different for other stars. The fundamental constraints for the detection of stellar pulsations are provided by the magnitude of the pulsation amplitudes and the effect of *cancellation*. The latter is responsible for the fact that we cannot observe modes with very high l values in stars other than the Sun. Since the disks of most stars cannot be resolved, the disk-integrated observations result in the overall cancellation or averaging of brighter and darker areas on the surface, which strongly diminishes the effective amplitudes of the brightness variations. Thus, the solar-like oscillations in other stars have only been observed up to $l = 3$ and the claimed photometric detection of higher-l values in some δ Scuti stars is a matter of debate.

As mentioned above, the intrinsic amplitudes are also important and vary strongly among the different types of variable stars. Whereas Cepheids have light variations that can even be detected by eye, solar-type oscillations in Sun-like stars have amplitudes that are of the order of several parts per million (ppm). Ground-based observations, for instance for δ Scuti stars, can usually detect variability down to mmag-level precision. Therefore, due to the optical limitations of observing through the Earth's atmosphere, the sensitivity from ground is too small to, e.g., detect solar-type oscillations by approximately 3 orders of magnitude. Some of the first solar-type oscillations in other stars have been observed from ground, however, due to the more sensitive observations of radial velocity variations on the level of m s^{-1}. Even for radial velocity variations, however, a second major restriction applies for ground-based observations. In order to reliably extract pulsational information for many frequencies, long, uninterrupted, and rapidly sampled data sets have to be obtained. This is due to four properties of the Fourier domain that is often used to analyse such data sets (see, e.g., Reegen 2007):

- noise levels in the frequency domain decrease with the square root of the number of data points,

- the Rayleigh frequency resolution, which limits the frequency resolution in the discrete Fourier transform, is $1/\Delta T$, where ΔT is the length of the data set in time units,

- the Nyquist frequency, $1/(2\delta t)$, where for equidistant sampling δt is the time difference between adjacent data points, limits the highest frequency that can be reliably detected, and

- gaps in the sampling can produce artificial signal (aliasing) well above the noise level.

These limitations affect different types of pulsation data in different ways. Regular small-amplitude oscillations, e.g., as in δ Scuti stars, have mode lifetimes

9

that are much longer than the observed data sets. Therefore, even when multiple modes are excited, these sinusoidal signals can be easily resolved and decomposed into several significant frequency components. The uncertainties for the actual mode frequencies are then only limited by the frequency resolution and the signal-to-noise ratio of the actual pulsation signal.

Solar-type oscillations, on the other hand, are damped and re-excited on much shorter time scales. If the mode lifetimes are shorter than the data set time base, multiple realizations of the damped and re-excited signal are observed. This is known to lead to Lorentzian profiles in the power spectrum, where the width of the profile (the so-called *linewidth*) is inversely proportional to the mode lifetime (or mode coherence time) τ as in

$$\Gamma = \frac{1}{\pi\tau}. \tag{1.6}$$

For the Sun, for instance, the mode linewidths around ν_{max} are of the order of $1\,\mu\mathrm{Hz}$ (Chaplin et al. 1997). All this led to the realization that space-based missions, which are not negatively affected by gaps due to the day-night cycle, and which could observe targets for months at a time were the best possible solution to obtain asteroseismic data. The first dedicated asteroseismology mission was the Canadian MOST (Microvariability and Oscillation of STars) satellite (Walker et al. 2003) and launched in the summer of 2003 and is still in orbit obtaining data. The CoRoT mission (Michel et al. 2006) and Kepler mission (Borucki et al. 2010) followed a few years later and have since revolutionized the observational aspects of asteroseismology (see section 1.3).

1.2.2 The extraction of solar-like oscillations

The fact that individual pulsation modes are no longer represented by a sharp peak in the frequency domain but by Lorentzian profiles has severe consequences. In the case of coherent, long-lived oscillations, the individual signals can be determined by refining a multi-sine fit using iterative sine fitting and subtraction. This has traditionally been accomplished using software packages like Period04 (Lenz &

10

Breger 2005) or SigSpec (Reegen 2007), and more reliable probabilistic techniques for frequency extraction are still being developed (Zechmeister & Kürster 2009). In order to extract the frequencies associated with the Lorentzian profiles, however, it is necessary to model the power spectrum, including the granulation background (see, e.g., Kallinger & Matthews 2010), the (possibly skewed) shape of the Lorentzian profiles and the leakage from adjacent profiles. Furthermore, for sufficiently fast rotating stars and modes with $l > 0$, sidelobes to the Lorentzian profiles with asymmetric mode height also have to be considered. While this provides additional information on the stellar inclination (Gizon & Solanki 2003) it adds several new parameters per mode and complicates the analysis.

For solar data sets that span up to decades, modelling all this information is a computational challenge. Many different techniques have been developed to either model individual parts of the solar power spectrum or to implement a pseudo-global fitting strategy (e.g. Fletcher et al. 2009). For space-based data sets that are, at least for now, much shorter, the whole power spectrum can still be treated in a global approach where every significant mode is fit simultaneously. Recently, Bayesian methods (Gruberbauer et al. 2009; Benomar et al. 2009a; Handberg & Campante 2011) have proven to be the most promising approach to execute this task. They are able to propagate all the uncertainties and correlations of the power spectrum models with their sometimes more than 100 free parameters, into the eventual frequency uncertainties obtained from marginal distributions. They also help with the problem of mode identification, as will be discussed in the next chapter[2]. More details on Bayesian methods will be provided in Chapter 2.

Figure 1.3 summarizes the description of solar-type oscillation data by depicting a Bayesian fit to the CoRoT data of HD 49933 and a corresponding échelle diagram. Even without the échelle diagram, the regularity in the pulsation spectrum is obvious. Note that this plot only shows the region in the spectrum where significant

---

[2]A different but very promising Bayesian approach, which models the damped and randomly excited signal in the time domain, has been suggested by Brewer & Stello (2009). So far, however, it remains computationally intractable for the big data sets obtained from space.
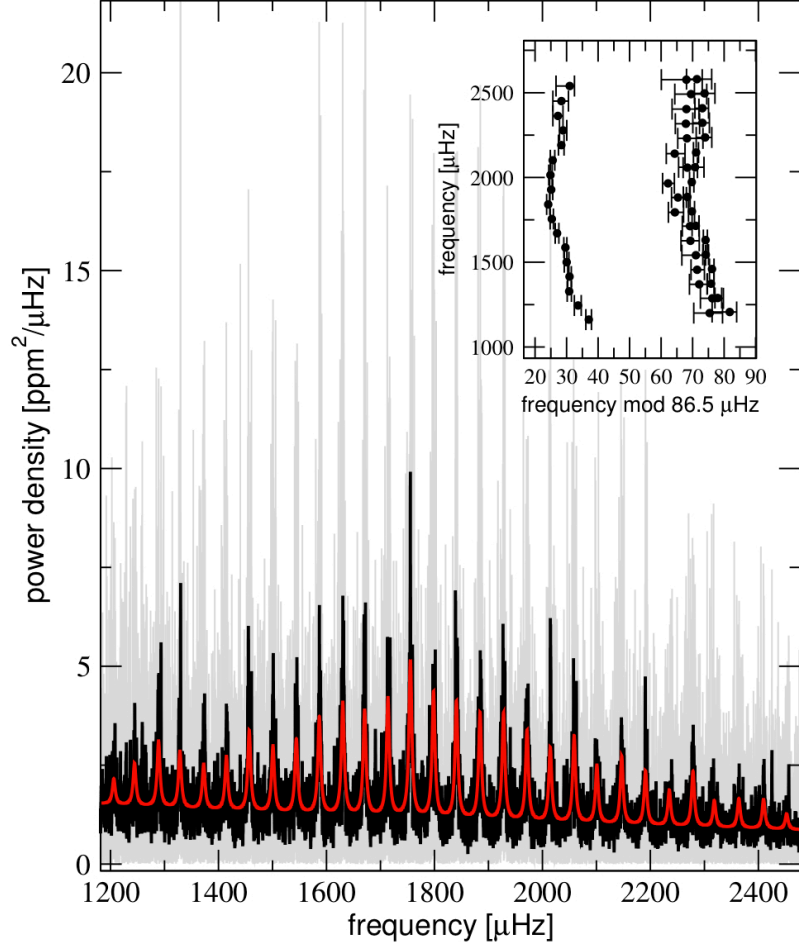
Figure 1.3: Power density spectrum of the solar-type oscillations detected in the CoRoT data for HD 49933 (grey) (Appourchaux et al. 2008; Gruberbauer et al. 2009; Benomar et al. 2009a). Also shown is a 10-pt running average (black) and a simple multi-component model fit consisting of a standard background model and several Lorentzian profiles (red). For simplicity, this model assumes equal linewidths for all modes and only considers $l = 0$ and $l = 1$ modes. The insert shows the échelle diagram produced using the frequencies from Benomar et al. (2009b).

pulsation signal can be detected. The power spectrum still contains more information on the granulation signal and possibly rotational modulation of the stellar brightness at much lower frequencies.

1.3 Recent advances in asteroseismology from space

1.3.1 Unveiling the hidden secrets of classical pulsators

Within the time span of a few years, the CoRoT and Kepler missions have lead to a revolution in the study of pulsating stars. It is self-evident that either stars with previously undetected pulsations or with many modes that are difficult to detect from ground are the biggest beneficiaries of the new instruments. However, even traditional pulsators like RR Lyrae stars have benefitted from the gapless sampling and high precision, for instance in the study of the so-called Blazhko cycles and a potentially connected new phenomenon discovered by the Kepler satellite called "period doubling" (Szabó et al. 2010). Moreover, as already suggested by data produced with previous space-based missions such as MOST (Gruberbauer et al. 2007), previously unknown additional modes were found in these stars as well (Guggenberger et al. 2012).

For δ Scuti stars, γ Dor stars, and hybrids, CoRoT and Kepler have revealed up to hundreds of additional and unexpected frequencies (e.g. Moya & Rodríguez-López 2010). This has produced quite a controversy about whether these frequencies could have been produced by pulsation (García Hernández et al. 2009) or are rather artefacts created by the noisy signals produced through very shallow surface convection and granulation (Kallinger & Matthews 2010). In any case, it seems clear that the incidence of hybrid pulsators that show both p and g modes appears to be much higher than previously thought (see, e.g Uytterhoeven et al. 2011). Many more unexpected discoveries, such as potential g modes in pre-main sequence stars (Zwintz et al. 2013a), as well as regularities (Breger et al. 2011; Zwintz et al. 2013b) and very high-frequencies in δ Scuti stars (Balona et al. 2012) have resulted from the CoRoT and Kepler space photometry as well. Naturally, roAp stars (Kurtz et al. 2011) and pulsating B stars (Balona et al. 2011) have also been observed, resulting in interesting new phenomena and detections.

1.3.2 The solar-like revolution

The reliable detection of nonradial solar-like oscillations in hundreds to thousands of red giants (e.g., Huber et al. 2010; Kallinger et al. 2010) is perhaps considered the most important asteroseismic discovery of these missions so far[3]. It has furthermore led to the detection of many g modes clustered around the $l = 1$ modes, as predicted by theory (Dupret et al. 2009). These g modes have been highly valuable since it was discovered that they can be employed to distinguish between hydrogen- and helium-burning giants (Bedding et al. 2011; Kallinger et al. 2012; Stello et al. 2013). Furthermore, they also allow to draw inferences about the rotation speed profile as a function of stellar radius (Beck et al. 2012).

Most relevant to the types of stars analyzed in this thesis, the data base for the reliable observation of solar-like oscillations in main sequence stars has increased by at least one order of magnitude. From one (the Sun) or perhaps a few objects, if one considers the ground-based observations of e.g., α Cen, η Boo, and a few other stars as reliable (see Aerts et al. 2008, for a good review of the state of asteroseismology before the impact of the two space-based missions), we have moved forward to single publications that contain frequencies for several dozens of solar-like pulsators (Appourchaux et al. 2012). Even just with a few quarters of Kepler data, as in the case of the 61 stars reported by Appourchaux et al., observational uncertainties of frequencies are approaching the theoretical random uncertainties typically reached by pulsation model frequencies. The observation of so many stars with solar-like oscillations also permit the explicit testing of the scaling relations previously mentioned in this chapter (Huber et al. 2011).

1.3.3 Realizing the promises of asteroseismology

A large asteroseismic community is currently at work, using previously unimaginable amounts of data, to investigate so many different phenomena that it is hopeless

---

[3]It should be noted that nonradial modes in giants were first detected by the MOST satellite Kallinger et al. (2008).

to mention them all in this introductory chapter. Instead, it is more important to connect these current advances with the aims of this thesis. To summarize, current state-of-the-art asteroseismology relying on space-based data is concerned with two types of studies. The first has been termed "ensemble asteroseismology" (e.g. Chaplin et al. 2011) and amounts to studying whole populations and types of stars using "comparative" and statistical asteroseismology. Due to their statistical nature, mistakes made in the modelling of individual stars, be it due to inadequacies of the scaling relations, slight systematics in the stellar modelling, or unavailable spectroscopic constraints are not as important.

The second important approach, however, requires the detailed modelling of individual stars in order to provide stronger constraints on the stellar parameters (e.g., Brandão et al. 2011; Metcalfe et al. 2010; Mathur et al. 2012; Metcalfe et al. 2012). This is, for instance, important to constrain properties of their transiting planets[4] (Huber et al. 2013). Of course, the ultimate goal for stellar astrophysicists is to better understand the stars themselves but it is vital to emphasise that for many types of stars there exist fundamental problems for the modelling approach. For many stars there exists the general problem of ambiguous mode identification (e.g., for δ Scuti and roAp stars) that prohibits a simple comparison of observed and calculated frequencies. Furthermore, the theoretical stellar models used to study stars are, at least for several classes of objects, still not good enough to perform asteroseismic modelling without significant systematic errors. For the aforementioned δ Scuti stars, for instance, systematic frequency errors arise due to rapid rotation.

One major problem for modelling of solar-like pulsation, which is of central importance for this thesis, lies in the inadequate modelling of the outer layers of solar-like stars which imprint the so-called "surface effect" onto the stellar frequency spectra[5]. This systematic error reveals itself through increasing deviations of the

---

[4]It is sometimes necessary to remind the asteroseismologist that both the CoRoT and Kepler mission are, fundamentally, missions for the detection and characterisation of exo-planets

[5]As of yet, it is uncertain in how far these "surface effects" also extend to other stars with convective envelopes such as red giants.

observed frequencies from the calculated values towards higher orders. Since this effect originates in the improper modelling of convection, different solutions have been proposed. Most widely used is the application of an empirical correction (Kjeldsen et al. 2008) which, however, has only been calibrated with a specific set of solar models and observations. In the next chapters, this thesis will introduce a more flexible solution for the consideration of such systematic errors. A different, more future-proof approach is to improve the convective modelling specifically targeted at reducing the surface effects (e.g., Grigahcène et al. 2012). However, most current studies are done using traditional stellar models and adiabatic pulsation codes that suffer from the surface effect and deficiencies due to the assumption of adiabaticity, and until recently there also was no framework for how to consistently compare the various modelling approaches. In general, even recent studies with many different modellers and codes, such as presented by Metcalfe et al. (2012), suffer from the lack of such a framework since many different incommensurable goodness-of-fit results (e.g., due to different parametrisation of the surface effect correction) have to be averaged to represent the results. This problem also persists outside the specific field of solar-like oscillations and is a problem for asteroseismic modelling in general. Providing this framework for a consistent comparison of model codes and results is the second major goal of this thesis.

Although the wealth of new data allows us to draw many new conclusions independent of the detailed modelling, to reach the fundamental goals of asteroseismology it is necessary to overcome these problems. I will argue in the chapters that follow that current methods for asteroseismic modelling can be improved upon by using probabilistic analysis. The goal is to develop a method that can be used to systematically work towards more realistic theoretical models and to critically evaluate these models at all stages. In order to make this a reality, a method has to be found that is able to reliably compare different modelling codes, different implementations of grid physics, and different formulations to correct against systematic errors in the

model. This method will be presented in the paper reproduced as Chapter 3. First, however, I will provide a brief introduction to Bayesian analysis.

Bibliography

Aerts, C., Christensen-Dalsgaard, J., Cunha, M., & Kurtz, D. W. 2008, Sol. Phys., 251, 3

Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, Asteroseismology (Springer Netherlands)

Appourchaux, T., et al. 2008, A&A, 488, 705

—. 2012, A&A, 543, A54

Balona, L. A., et al. 2011, MNRAS, 413, 2403

—. 2012, MNRAS, 424, 1187

Beck, P. G., et al. 2012, Nature, 481, 55

Bedding, T. R. 2012, in Astronomical Society of the Pacific Conference Series, Vol. 462, Progress in Solar/Stellar Physics with Helio- and Asteroseismology, ed. H. Shibahashi, M. Takata, & A. E. Lynas-Gray, 195

Bedding, T. R., et al. 2011, Nature, 471, 608

Benomar, O., Appourchaux, T., & Baudin, F. 2009a, A&A, 506, 15

Benomar, O., et al. 2009b, A&A, 507, L13

Bono, G., Caputo, F., Marconi, M., & Musella, I. 2010, ApJ, 715, 277

Borucki, W. J., et al. 2010, Science, 327, 977

Brandão, I. M., et al. 2011, A&A, 527, A37+

Breger, M., & Kolenberg, K. 2006, A&A, 460, 167

Breger, M., et al. 2005, A&A, 435, 955

—. 2011, MNRAS, 414, 1721

Brewer, B. J., & Stello, D. 2009, MNRAS, 395, 2226

Chaplin, W. J., Elsworth, Y., Isaak, G. R., McLeod, C. P., Miller, B. A., & New, R. 1997, MNRAS, 288, 623

Chaplin, W. J., et al. 2011, Science, 332, 213

Cunha, M. S. 2005, JApA, 26, 213

Deupree, R. G., & Beslin, W. 2010, ApJ, 721, 1900

Dupret, M.-A., et al. 2009, A&A, 506, 57

Eddington, A. S. 1926, The Internal Constitution of the Stars (Cambridge University Press)

Fletcher, S. T., Chaplin, W. J., Elsworth, Y., & New, R. 2009, ApJ, 694, 144

García Hernández, A., et al. 2009, A&A, 506, 79

Gizon, L., & Solanki, S. K. 2003, ApJ, 589, 1009

Grec, G., Fossat, E., & Pomerantz, M. A. 1983, Sol. Phys., 82, 55

Grigahcène, A., Dupret, M.-A., Sousa, S. G., Monteiro, M. J. P. F. G., Garrido, R., Scuflaire, R., & Gabriel, M. 2012, MNRAS, 422, L43

Gruberbauer, M., Kallinger, T., Weiss, W. W., & Guenther, D. B. 2009, A&A, 506, 1043

Gruberbauer, M., et al. 2007, MNRAS, 379, 1498

Guggenberger, E., et al. 2012, MNRAS, 424, 649

Guzik, J. A., Kaye, A. B., Bradley, P. A., Cox, A. N., & Neuforge, C. 2000, ApJ, 542, L57

Handberg, R., & Campante, T. L. 2011, A&A, 527, A56

Hansen, C. J., Kawaler, S. D., & Trimble, V. 2004, Stellar interiors : physical principles, structure, and evolution (Springer)

Huber, D., et al. 2010, ApJ, 723, 1607

—. 2011, ApJ, 743, 143

—. 2013, ArXiv e-prints

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

Kallinger, T., et al. 2008, A&A, 478, 497

—. 2010, A&A, 509, A77

—. 2012, A&A, 541, A51

Kjeldsen, H., & Bedding, T. R. 1995, A&A, 293, 87

Kjeldsen, H., Bedding, T. R., & Christensen-Dalsgaard, J. 2008, ApJ, 683, L175

Kurtz, D. W. 1982, MNRAS, 200, 807

Kurtz, D. W., et al. 2011, MNRAS, 414, 2550

Leavitt, H. S., & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1

Lenz, P., & Breger, M. 2005, Communications in Asteroseismology, 146, 53

Mathur, S., et al. 2012, ApJ, 749, 152

Metcalfe, T. S., et al. 2010, ApJ, 723, 1583

19

—. 2012, ApJ, 748, L10

Michel, E., et al. 2006, in ESA Special Publication, Vol. 1306, ESA Special Publication, ed. M. Fridlund, A. Baglin, J. Lochard, & L. Conroy, 39

Miller Bertolami, M. M., Córsico, A. H., & Althaus, L. G. 2011, ApJ, 741, L3

Moya, A., & Rodríguez-López, C. 2010, ApJ, 710, L7

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

Saio, H. 2005, MNRAS, 360, 1022

Stello, D., et al. 2013, ApJ, 765, L41

Szabó, R., et al. 2010, MNRAS, 409, 1244

Tassoul, M. 1980, ApJS, 43, 469

Uytterhoeven, K., et al. 2011, A&A, 534, A125

Vandakurov, Y. V. 1967, Astronomicheskii Zhurnal, 44, 786 (English translation: Soviet Astronomy AJ, 11, 630)

Walker, G., et al. 2003, PASP, 115, 1023

Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577

Zwintz, K., Fossati, L., Ryabchikova, T., Kaiser, A., Gruberbauer, M., Barnes, T. G., Baglin, A., & Chaintreuil, S. 2013a, A&A, 550, A121

Zwintz, K., et al. 2013b, ArXiv e-prints

Chapter 2

Introduction II: Probabilistic Inference Using Bayesian Analysis

In order to understand the modus operandi and the purpose of the asteroseismic methods developed in this thesis, some basic insights into Bayesian analysis are required. This chapter, which will strongly rely on the two textbooks by Gregory (2005) and Jaynes & Bretthorst (2003), will provide all the required preliminaries to do so. First, section 2.1 will provide a motivational discussion about scientific inference and the argument for probabilistic inference. Section 2.2 then sets out to explain the fundamental laws of probability theory and how they can be used to evaluate different hypotheses. This also includes a toy problem example related to stellar astrophysics. In section 2.3, I will then discuss some additional points that are helpful to understand, such as the nature of *priors*, the difference between a prior and an *observable*, the basic purpose of the *evidence*, and the modular nature of Bayesian analysis. Lastly, section 2.4 will present some recent results from stellar modelling and asteroseismology that were produced by the application of Bayesian analysis.

2.1 The scientific method and the argument for probabilistic inference

The scientific method is usually associated with the following narrative: science is a process that starts with the observation of nature, then leads to the formation of hypotheses, and then turns to experiments to verify or falsify the predictions that follow from the hypotheses. From this results a feedback loop that eventually leads to scientific theories as a collection of laws and relations between concepts and observed phenomena which have prevailed against all observational tests. Further observations

of nature are made to assess the applicability of the specific theories in previously unaccessible regimes until discrepancies are found.

Such an idealization of the scientific process, however, completely overlooks the actual sociological structure of science as a profession and the history of scientific development. This was impressively unveiled by Thomas Kuhn in the seminal work "The Structure of Scientific Revolutions" (Kuhn 1970). A simplification, which is often encountered when scientists describe the scientific method, is the notion that science works through falsification such that a single experiment can disprove a hypothesis or a theory. Kuhn, however, argues that scientists are much more adherent to the current paradigms in their field than the idea of falsification suggests. To paraphrase: once a theory is established, scientists use the tools that follow from it to further investigate nature as seen through its lens. Discrepancies between theory and experiment are not necessarily considered to be evidence against the paradigm's overall validity. Instead, slight modifications are made (e.g., *arbitrary fitting parameters*) to make the paradigm work also in such cases. This is fundamentally a good thing because otherwise science would be a very chaotic endeavour, constantly being overturned and questioned by the latest public claims of cold fusion or superluminal neutrinos. It is only when more and more evidence mounts in the peer-reviewed literature that makes it harder and harder to adhere to the current paradigm, that a *crisis* develops which can only be resolved by eventually changing to a different paradigm.

Therefore, falsification is either not as effective or as prevalent as the traditional narrative might suggest. One reason for this is certainly given by the difficulty to assess in practice whether a hypothesis, or one of its predictions, has actually been falsified. Complicated physical models usually have many (free) parameters and uncertainties or systematic errors in some of the modelled processes. Moreover, observations always have uncertainties, as well as known and unknown systematic errors. This naturally raises the question: is a model or a hypothesis automatically wrong if

the outcome of an experiment disagrees with what was predicted by more than the quoted uncertainties? Is a more or less conservative criterion necessary[1]?

This picture of the scientific process is clearly different from the classical narrative but strangely enough the idea of falsification still prevails. One could argue that this is partially due to long-held methodological practices. Science, in particular physics and also astronomy, was dominated for decades by the idea of frequentist hypothesis testing and the ideas of Fisher (see Jaynes & Bretthorst 2003)[2]. Given the increased complexity of todays problems, it is important to ask whether this school of scientific inference is reaching its limits. After all, these so-called "frequentist" methods, even though they haven been described as fundamentally different in their philosophy, are really approximations to general probabilistic inference. They certainly yield impressive and accurate results for many problems and were an obvious choice in particular at a time where computational power was limited. However, with the advent of more capable computers, the more general approach to scientific inference has re-emerged from the dustbins of history - Bayesian analysis.[3]

As will be described in the following sections, Bayesian analysis works in fundamentally different ways than traditional statistical tests (or frequentist methods), because it does away with the idea of simple hypothesis rejection as a sufficient tool for evaluating scientific models. The fundamental property of this probabilistic framework is that it is able to give every tested hypothesis an objective and comparative value. While the set of tested hypotheses, and the collection of prior knowledge on which their evaluation is based, is still subjective (or subject to economic or temporal constraints), a set of clearly stated assumptions always leads to the same outcome.

---

[1]This discussion leaves out the big problem of theories that have become very complicated and also more and more difficult to test. Such theories can reach a state where they are so flexible and powerful that they are utterly untestable. This is discussed for the case of string theory in the controversial book "The Trouble with Physics" (Smolin 2006)).

[2]For instance, as is discussed in Gregory (2005), the commonly cited p value is often used to reject (or accept) models based on the likelihood of the data sample under the null hypothesis. Such a test is only applicable if a commonly accepted threshold is employed which then assumes the responsibility for deciding whether the null hypothesis is considered "falsified".

[3]More details on the origins of Bayesian analysis and its connection to frequentist methods can be found, e.g., in Jaynes & Bretthorst (2003) and also McGrayne (2011).

Put in the terms of Jaynes & Bretthorst, Bayesian analysis is like a robot that evaluates new evidence using its prior knowledge and the rules of logic, and each robot with the same set of prior knowledge and the same set of new evidence must, by definition, come to the same conclusions. Except for the simplest of problems, this means that the process of scientific inference is never complete, because one can always envision new hypotheses, and one might always gather new evidence that can potentially have a huge impact on our prior knowledge for the next iteration of inference. However, such is the nature of the scientific method. It is certainly anchored in the well known principle of *Occam's Razor*, which says that the simplest available explanation that can explain all the available data is usually the most plausible. However, a new observational datum, if convincing enough, could suggest that what was deemed almost certain yesterday is now no longer so. Therefore, our set of prior assumptions should never have the upper hand in any test of our scientific hypotheses and theories. On the other hand, the famous motto of science, often publicized by Carl Sagan, that "extraordinary claims require extraordinary evidence", is certainly important as well. It is therefore assuring, that Bayesian analysis incorporates all these desiderata at the fundamental level.

Probabilistic inference has proven itself in countless practical applications even outside of science (see, e.g., McGrayne 2011), and it is gaining momentum in many previously "frequentist"-dominated scientific fields as well. One of these is asteroseismology, as will be discussed in section 2.4. It will be interesting to see whether the increasing prevalence of the more comparative Bayesian approach, which formally adheres to the fundamental principles of the scientific method, can change the narrative of how science is done in the future. Will Kuhn have to revise his conclusions if scientist will ever learn to use their Bayesian robots to evaluate competing paradigms?

2.2   Fundamentals of Bayesian inference

2.2.1   The laws of probability and Bayes' theorem

At the most fundamental level, probabilistic inference or Bayesian analysis is based on two "laws of probability" (Jaynes & Bretthorst 2003). The first is the *product rule*

$$P\left(A, B|C\right) = P\left(A|C\right) P\left(B|A, C\right) = P\left(B|C\right) P\left(A|B, C\right) \tag{2.1}$$

and the second is the *sum rule*

$$P\left(A + B|C\right) = P\left(A|C\right) + P\left(B|C\right) - P\left(A, B|C\right). \tag{2.2}$$

In these equations, P simply stands for probability or probability density, and the letters A, B, and C represent propositions that can be evaluated. Both rules follow from theorems developed by Cox (1961) as described at length by Jaynes & Bretthorst (2003). In short, these two laws are necessary for any inference that regards (1) the plausibility of a proposition as a numerical value, (2) requires consistency in the derived results, and (3) follows the requirements of logic. The notation used in these two equations associate the "," operator with a logical "AND", while the "+" operator describes the logical "OR". Finally, the "|" denotes conditionality in that the proposition on the left are evaluated conditional on the truth of the proposition on the right. Therefore the left side of the product rule, $P\left(A, B|C\right)$, reads as "the probability that A AND B are true, given C". The left side of the sum rule $P\left(A + B|C\right)$ reads as "the probability that A OR B OR both are true, given C".

By simply rearranging the terms in Equ. 2.1, one obtains *Bayes' theorem*

$$P\left(B|A, C\right) = \frac{P\left(B|C\right) P\left(A|B, C\right)}{P\left(A|C\right)} \tag{2.3}$$

which gives Bayesian analysis its name. It is important to remember, however, that

25

- Bayesian analysis is not merely the application of one mathematical theorem,

- the product and sum rules can be further applied to the terms in this and other probabilistic equations, and

- in some cases probabilities are of interest that do not require the specific use of Bayes theorem, but that would still be regarded as following a "Bayesian approach".

Still, although "probabilistic inference" is a better name to describe all these features, "Bayesian analysis" has established itself as the common term with which such analyses are described.

2.2.2 The Bayesian approach to scientific inference

In order to perform scientific inference, the terms in Equ. 2.3 need to be identified with propositions that are relevant for this task. This is done, as shown in Gregory (2005), by using for example the assignment

$$P\left(M_1|D, H, I\right) = \frac{P\left(M_1|H, I\right) P\left(D|M_1, H, I\right)}{P\left(D|H, I\right)} \tag{2.4}$$

In this assignment, M_1 represents the proposition that the model M_1 describes what is observed in the data D. Furthermore, all these terms are conditional on the hypothesis H with which the model M_1 was created. For example, H could stand for a specific combination of fundamental assumptions in stellar modelling (such as: treatment of convection, chemical composition, nuclear reaction rates). Different models M_i could then be created, for instance with different masses and metallicities or at different stages in stellar evolution, from the same common "hypothesis" H. Finally, I stands for the *prior information* or the set of prior assumptions made.

If a certain subset of mutually exclusive models M_i (e.g., a star cannot have two different masses at the same time) are to be considered as a combined proposition,

one can use the sum rule and the product rule and write for example

$$P\left(M_1 + M_2 + M_3|D, H, I\right) = \frac{\sum_{i=1}^{3} P\left(M_i|H, I\right) P\left(D|M_i, H, I\right)}{P\left(D|H, I\right)}. \qquad (2.5)$$

When this sum is extended to all models that we have available under the hypothesis H, the term on the left side becomes unity, since the posterior probability over all available hypotheses must be equal to one[4]. In this case the denominator of Bayes' theorem finally takes shape as

$$P\left(D|H, I\right) = \sum_i P\left(M_i|H, I\right) P\left(D|M_i, H, I\right). \qquad (2.6)$$

Rewriting the "scientific inference" version of Bayes' theorem, we obtain for the posterior of model M_1

$$P\left(M_1|D, H, I\right) = \frac{P\left(M_1|H, I\right) P\left(D|M_1, H, I\right)}{\sum_i P\left(M_i|H, I\right) P\left(D|M_i, H, I\right)}. \qquad (2.7)$$

This is another common and insightful way to write Bayes' theorem when used for scientific inference.

The individual terms in Equ. 2.4 or Equ 2.7 have very specific meanings:

1. $P\left(M_1|D, H, I\right)$ is the *posterior probability* of M_1 given the data D, the hypothesis H and the prior information I. It represents the actual probability of the model M_1 once all models M_i have been evaluated using the new data D.

2. $P\left(M_1|H, I\right)$ is the *prior probability* M_1 given the hypothesis H and the prior information I. This simply corresponds to the probability that model M_1 represents reality before the new data was obtained.

3. $P\left(D|M_1, H, I\right)$ is the *likelihood* of the obtaining the data D under the assumption that the model M_1 represents reality and given the prior information I.

---

[4]Note that this does not mean that one of the models must be "true" in any way. This will be discussed in more detail in section 2.3

4. $P(D|H, I)$, finally, is called the *evidence*[5] It is the likelihood of obtaining the data D under the assumption that the hypothesis H represents reality. As derived above, it is also equivalent to the normalization constant obtained when summing all the different numerators in Bayes' theorem.

All terms mentioned above warrant some more discussion that will be provided in section 2.3.

2.2.3 Marginalization

An important feature of Bayesian analysis that stems from the underlying rules of probability theory is *marginalization*. Marginalization has already been used above to derive the different expressions for the evidence. It is a process that employs the sum rule to remove a "dimension" in the problem that previously had to be considered. This is more easily understood if the models to be analyzed are analytical. For example, in simple sine fitting, a function

$$y(t) = c + A\sin(2\pi f t + \phi) \tag{2.8}$$

is fitted to the data (e.g., brightness variations), where c is the constant offset, A is the amplitude, f is the temporal frequency, ϕ is the phase shift, and t is the independent variable (e.g, time). All of these terms are required to describe a proper sine wave. However, when fitting such a function to the data, we are often not interested in all parameters. In other cases, strong correlations might arise that make it difficult to summarize the result as, e.g., a function of frequency alone. In a Bayesian analysis, while the analysis would proceed by calculating the posterior probability for all combinations of parameters, there is the option of marginalizing over the uninteresting parameters. For the sine wave example, the posterior probability is a

---

[5]Jaynes & Bretthorst uses the term *evidence* while Gregory uses the term *global likelihood*. This thesis predominantly uses the former term since it is more concise and prevents confusion with the regular likelihood.

function of all the parameters

$$P\left(c, A, f, \phi | D, I\right) \propto P\left(c, A, f, \phi | I\right) P\left(D | c, A, f, \phi, I\right). \tag{2.9}$$

If, say, the constant offset is uninteresting[6], it can be marginalized by calculating

$$P\left(A, f, \phi | D, I\right) \propto \sum_{c\,\text{values}} P\left(c, A, f, \phi | I\right) P\left(D | c, A, f, \phi, I\right), \tag{2.10}$$

or in the continuous case

$$P\left(A, f, \phi | D, I\right) \propto \int_{c_{min}}^{c_{max}} P\left(c, A, f, \phi | I\right) P\left(D | c, A, f, \phi, I\right) \mathrm{d}c. \tag{2.11}$$

Note that c is now missing in the list of arguments in the posterior probability. It is still considered in the evaluation of the *marginal posterior* probabilities of A, f, and ϕ, but these values are now independent of any specific choice for the value of c. Naturally, the marginalization will not change the value of the evidence, since the evidence is the marginal likelihood over the whole hypothesis space.

In order to properly calculate the marginal posterior probability, all values of the marginalized parameter have to be taken into account. This is of course impossible, unless the integral over the parameter can be performed analytically. In such cases, however, the computational savings can be tremendous, since the sampling of a complete parameter can be avoided by simply changing the function that is evaluated by integration. Nonetheless, even if the marginalization cannot be performed analytically, it is an important concept for the analysis of more complex problems via Markov Chain Monte Carlo (MCMC) methods.

2.2.4 A toy problem

Before proceeding with more details on the different terms in Bayes' theorem, it is helpful to present some numerical examples to better explain how these equations

---

[6]Gregory likes to use the term *nuisance parameter* for this purpose.

work and how they can be used to answer specific questions. Therefore, this section will show an intuitive but sufficiently complex example using a stellar toy problem.

Imagine that a new kind of peculiar star has been observed to have a certain effective temperature $T_{\text{eff,obs}} = 8000 \pm 100, \text{K}$. No other information about the object is available, except for the fact that T_{eff} is a very important probe of the particular and exotic physical properties of this star. Let us then assume that two sets of models, both of which propose to adequately describe this peculiar object, are available. One set uses fairly standard physics $\mathcal{S}$, while the other set incorporates non-standard physics $\mathcal{N}$ (e.g., a new description of the stratification of peculiar chemical elements). Let only three models be available per model family with specific T_{eff} predictions that are summarized in Table 2.1.

As shown by Gregory (2005), by assuming Gaussian uncertainties e we can use the ansatz

$$T_{\text{eff,obs}} = T_{\text{eff,M}_i} + e. \tag{2.12}$$

If one of the models was actually correct, we would observe the predicted effective temperature were it not for the random uncertainties. Thus, given the Gaussian uncertainties, we can write the likelihood of observing the data D ($= T_{\text{eff,obs}}$) as

$$P(T_{\text{eff,obs}}|M_i, \mathcal{H}, I) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(T_{\text{eff,obs}} - T_{\text{eff,M}_i})^2}{2\sigma^2}\right), \tag{2.13}$$

where $\mathcal{H}$ ("hypothesis") stands for either $\mathcal{S}$ ("$\mathcal{S}$ is true") or $\mathcal{N}$ ("$\mathcal{N}$ is true"), or the logical proposition $\mathcal{S} + \mathcal{N}$ ("$\mathcal{S}$ or $\mathcal{N}$ are true"), and $\sigma = 100\,\text{K}$. With these results, the posterior probabilities can then be calculated after the prior probabilities have been specified. If there is no specific information about which models are more realistic *a priori* (for more on this see the next section), an equal uniform probability can be assigned. The only condition to obtain a so-called *proper prior* is to ensure that they are properly normalized[7]. Therefore, if $\mathcal{N}$ and $\mathcal{S}$ are evaluated separately,

---

[7]Improper priors, which are not normalized to unity, do not produce evidence values that can be used as likelihood for further Bayesian analysis. This follows from the fact that the evidence is the prior-weighted mean likelihood.

Table 2.1: T_{eff}, likelihood, and posterior probability for all competing models in the stellar probability example.

observed	M_1	M_2	M_3
$T_{\mathrm{eff}}(\mathcal{S})$ [K]	7932	8305	12103
$T_{\mathrm{eff}}(\mathcal{N})$ [K]	7908	8102	9007
results	M_1	M_2	M_3
$P(T_{\mathrm{eff,obs}}|M,\mathcal{S},I)$	$3.166 \cdot 10^{-3}$	$3.8 \cdot 10^{-5}$	≈ 0.0
$P(T_{\mathrm{eff,obs}}|M,\mathcal{N},I)$	$2.613 \cdot 10^{-3}$	$2.37 \cdot 10^{-3}$	≈ 0.0
$P(M|T_{\mathrm{eff,obs}},\mathcal{S},I)$	0.988	0.012	≈ 0.0
$P(M|T_{\mathrm{eff,obs}},\mathcal{N},I)$	0.524	0.476	≈ 0.0
$P(T_{\mathrm{eff,obs}}|M,\mathcal{S}+\mathcal{N},I)$	$3.166 \cdot 10^{-3}$	$3.8 \cdot 10^{-5}$	≈ 0.0
$P(T_{\mathrm{eff,obs}}|M,\mathcal{S}+\mathcal{N},I)$	$2.613 \cdot 10^{-3}$	$2.37 \cdot 10^{-3}$	≈ 0.0
$P(M|T_{\mathrm{eff,obs}},\mathcal{S}+\mathcal{N},I)$	0.387	0.005	≈ 0.0
$P(M|T_{\mathrm{eff,obs}},\mathcal{S}+\mathcal{N},I)$	0.319	0.289	≈ 0.0

$P(M_i|\mathcal{H},I) = 1/3$. However, we can also evaluate all the individual models together to see which of the individual models is the most probable. This is done by combining the two model families so that $P(M_i|\mathcal{S}+\mathcal{N},I) = 1/6$. The results for the likelihoods and the posterior probabilities are also shown in Table 2.1.

It is clear that the likelihood values are the same whether they are evaluated conditional on one or both model families. The posterior probabilities, however, change dramatically. M_1 with the standard $\mathcal{S}$ physics, has a very high probability (0.988) compared to the other models in the same model family. When all models of both families are taken into account, it still has the highest overall posterior probability (0.387), but the second highest probability (0.319) now belongs to a model of the non-standard physics family, and the contrast between these two models is negligible. The reason for this difference is self-evident: within the $\mathcal{S}$ familiy, M_1 has no strong competitor, but in the $\mathcal{N}$ family, two models predict a very similar effective temperature[8].

---

[8]This is an example for how Bayesian analysis correctly takes into account the specific propositions that are investigated. Standard hypothesis tests, on the other hand, are usually only based on the likelihood itself. Therefore, based on the fundamental formalism, they cannot answer the questions that are related to the propositions tested here without further modifications.

Table 2.2: Posterior probabilities for the $\mathcal{S}$ and $\mathcal{N}$ families as a function of the evidence and several priors.

equal prior		$\mathcal{S}$	$\mathcal{N}$
$P\left(\mathcal{H}\vert I\right)$ (= prior)		0.5	0.5
$P\left(T_{\mathrm{eff,obs}}\vert\mathcal{H},I\right)$ (= evidence)		0.001068	0.001661
$P\left(\mathcal{H}\vert T_{\mathrm{eff,obs}},I\right)$ (= posterior)		0.391	0.609
unequal prior 1		$\mathcal{S}$	$\mathcal{N}$
$P\left(\mathcal{H}\vert I\right)$		0.9	0.1
$P\left(T_{\mathrm{eff,obs}}\vert\mathcal{H},I\right)$		0.001068	0.001661
$P\left(\mathcal{H}\vert T_{\mathrm{eff,obs}},I\right)$		0.853	0.147
unequal prior 2		$\mathcal{S}$	$\mathcal{N}$
$P\left(\mathcal{H}\vert I\right)$		0.99	0.01
$P\left(T_{\mathrm{eff,obs}}\vert\mathcal{H},I\right)$		0.001068	0.001661
$P\left(\mathcal{H}\vert T_{\mathrm{eff,obs}},I\right)$		0.985	0.015

The most interesting exercise is to evaluate the evidence for the two model families. The evidence can be understood as the prior-weighted mean likelihood, but it is also equivalent to the likelihood for the data under the hypothesis expressed by the summation. In the case of our example, therefore, the evidence for $\mathcal{S}$ (or $\mathcal{N}$) is equivalent to the likelihood for the data given $\mathcal{S}$ (or $\mathcal{N}$) as a whole. Hence, we can employ Bayes' theorem again to calculate the posterior probability for, e.g., the model family $\mathcal{S}$ with

$$P\left(\mathcal{S}\vert T_{\mathrm{eff,obs}},I\right) = \frac{P\left(\mathcal{S}\vert I\right)P\left(T_{\mathrm{eff,obs}}\vert\mathcal{S},I\right)}{P\left(\mathcal{S}\vert I\right)P\left(T_{\mathrm{eff,obs}}\vert\mathcal{S},I\right) + P\left(\mathcal{N}\vert I\right)P\left(T_{\mathrm{eff,obs}}\vert\mathcal{N},I\right)} \tag{2.14}$$

The equation now contains prior probabilities for the model families as a whole, $P\left(\mathcal{S}\vert I\right)$ and $P\left(\mathcal{N}\vert I\right)$, which again should be properly normalized to unity to be considered proper priors. The results for various prior values are presented in Table 2.2. As shown, if both model families are equally probable before the new observations are evaluated, we obtain the interesting situation where the most probable model belongs to model family $\mathcal{S}$, but where model family $\mathcal{N}$ as a whole is overall more probable (0.609 compared to 0.391). This is due to the fact that the prior-weighted mean

likelihood (and therefore the evidence) is larger for model $\mathcal{N}$ than for model $\mathcal{S}$. This interesting example also signifies that it is important to realize what is calculated at each step in the Bayesian analysis in contrast to the standard statistical tests. Are we interested in determining the most probable model, or are we interested in evaluating the physics behind the model? Answers to these questions usually belong to the evaluation of quite different propositions that nonetheless follow from the same exact formalism.

Finally, more details are in order for the priors $P\left(\mathcal{S}|I\right)$ and $P\left(\mathcal{N}|I\right)$. Since our example considers the comparison of standard to non-standard physics, it is important to assess how improbable the non-standard physics are, simply because they are non-standard. Thankfully, the rules of probability consider this implicitly due to the presence of the priors in Equ. 2.14. If it is simply a matter of convention, then the uniform prior is an adequate choice. However, if the non-standard physics of $\mathcal{N}$ are known to be otherwise much less plausible than those of $\mathcal{S}$, different priors are in order which can have a big impact on the posterior probabilities. Additional priors in Table 2.2 show how priors of various orders of magnitude in probability contrast affect the results. It can be difficult to assess which priors to use for things such as different input physics, but that does not render the analysis more subjective as is often claimed. More details on the choice of prior values will be discussed in the next section.

2.3 Advanced details of Bayesian analysis

Much of the skepticism that prevented the quick adaption of Bayesian analysis in many fields lies in its complexity and nuances, but also in the simple naming conventions of the different terms in, e.g., Equ 2.7. This section will provide a brief discussion of these terms, work out some differences to standard frequentist inference, and conclude with a detailed look at the modular nature of Bayesian analysis. Finally, at the end of this section, the Bayesian treatment of systematic errors will be introduced.

2.3.1 Prior probabilities

The prior probabilities and how they are used could be argued to leak "subjectivity" into statistical analyses. Several arguments can be given to better understand why priors are actually required that the claim of subjectivity is not necessarily correct. Firstly, *prior* might be a fitting name, but it misrepresents the fact that the term simply describes conditional probabilities that are *required by the product rule of probability theory.* It is therefore not the convenience of affecting the posterior probabilities that drives Bayesian analysts to incorporate their prior information. Doing so is a mathematical necessity.

Thus, follows the second argument, since prior probabilities *have to be specified* in order to produce the numerically correct results, not using prior probabilities is equivalent to always using equal priors for all parameters or hypotheses. As is argued by Jaynes & Bretthorst (2003), however, the uniform prior

$$P(x|I) = \frac{1}{x_{\max} - x_{\min}} \tag{2.15}$$

is not always the prior that correctly encodes a state of "missing information" for some parameter x. In the case of a scale parameter that varies over many orders of magnitude, for instance, the Jeffreys prior

$$P(x|I) = \frac{1}{x \log\left(x_{\max}/x_{\min}\right)} \tag{2.16}$$

is the correct choice. Jaynes tried to answer the general question which priors would be most non-committal to missing information (i.e., we are not putting in information that we do not have). The solution that he provides is to determine the prior which maximizes the entropy of the prior probability distribution, which leads to very particular prior choices for the simplest problems (i.e., uniform prior for a location parameter, Jeffreys prior for a scale parameter, Gaussian prior for a distribution with known mean and variance).

However, when the maximum entropy approach is not used, specifying the prior probability is still a requirement to (re-)producing the results. As mentioned above, Jaynes' "inference robot" will always come to the same conclusions if supplied with the same bits of prior information and the same observational data. Therefore, there is no more subjectivity in the results than in any other statistical test. What Bayes' theorem (or the robot) does, however, is to elevate the prior assumptions from some hidden area in statistical testing to being an integral part of the analysis to be put under scrutiny. As in the toy example above (or in the analysis of the Sun in Chapter 4), it is sometimes not easy to be absolutely clear on the choice of priors. In such cases, however, many different priors can be tested. The importance of Bayes' theorem is then that the choice of prior becomes a potentially revealing part of the analysis.

2.3.2 The posterior probability and the likelihood

The posterior probability is the probability of propositions (logical statements, models, parameters) after the evaluation of new information, but with consideration of the previous state of information. The likelihood, on the other hand, merely describes how likely it is to obtain the new information, given that the assumptions on which it is conditional are correct. From the formal equations alone, it is obvious that the two terms cannot mean the same thing, unless both the prior probability and the evidence either cancel or are equal to unity. In this case, however, the likelihood and the posterior themselves have to be equal to unity, and therefore no inference has occurred.

Seen from yet another angle, the likelihood alone cannot describe the probability of some hypothesis. Assume that there is some experiment, and that there are two hypotheses, both of which are equally probable *a priori*. Furthermore, both experiments predict the same outcomes with the same frequency although they make completely different but equally probable assumptions. Evaluated with some specific outcome observed from this experiment, they would both have the same prior

probability and they would lead to the same likelihood. Hence, if the two were the only hypotheses considered, they would obtain the same posterior probability: 0.5. Unless some experiment could be found where the predictions of both hypotheses differ, this result would always be obtained. The likelihood term, however, depends on the specific data set that was obtained. Sometimes the outcome of the experiment would be more, sometimes less likely. If the likelihood alone was responsible for the actual probability of the hypotheses, the probabilities for both models would constantly change. In the case of extremely unlikely outcomes, the experiment would even suggest that both hypotheses are wrong. Nonetheless, extremely unlikely events happen all the time. For instance, in a game of cards, the likelihood of obtaining a specific set of playing cards is astronomically small. Nonetheless, every player ends up with a hand of cards once they have been dealt.

Yet, the traditional χ^2 test, one of the test statistics most often encountered in traditional frequentist inference (see, e.g., Gregory 2005), uses this exact *likelihood principle* alone. It assesses the likelihood of obtaining the test statistic (the χ^2 value) given the assumptions (n degrees of freedom; observed variance due to independent, normally-distributed random variables). This is then turned into a "probability" (the p value) by using the cumulative distribution function of the χ^2 distribution for n degrees of freedom. However, this probability does not consider any alternate hypotheses, since it is only based on the likelihood of one model - the *null* hypothesis. Finally, the p value only encapsulates the frequency of the occurrence a χ^2 value like the one that was observed, if the experiment (or whatever led to the new information) was infinitely repeated. To summarize, it is merely a normalized likelihood.

The likelihood is only one term in the equations, and it only determines the likelihood of the data given that the conditional assumptions are correct. The probability of the tested hypotheses is quite different. This is also relevant for the distinction between prior information and new data. One of Cox's desiderata, that of consistency, requires that a calculus of inference must come to the same conclusions, no matter in which sequence information is evaluated. For instance, assume that

36

some information is obtained and evaluated at time $T1$, which leads to a certain state of prior information at later time $T2$. Then, another piece of information is obtained and evaluated at time $T2$. Alternatively, assume that the information at time $T1$ was put into an envelope instead of being evaluated. This envelope is then opened at time $T2$, at the same time as the second piece of information arrives, and both pieces are evaluated simultaneously. Naturally, the state of information after time $T2$ is the same in both cases. Similarly, Bayes' theorem will also produce the exact same posterior probability after time $T2$. However, even though the posterior probability at $T2$ is the same, the value of the likelihood at $T2$ differs depending on whether one or two pieces of information are evaluated at the same time. This is another indication that the likelihood alone is insufficient to evaluate hypotheses. It always depends on what is considered prior information and what is considered new data.

To conclude, the posterior probability, and the whole Bayesian approach, differs from the likelihood principle because the posterior probability and the likelihood of obtaining the data are two different parts of the problem of inference. The likelihood always has to be conditional on the hypothesis it is evaluating, and therefore it is insufficient for probabilistic inference.

2.3.3 The evidence and the modular nature of Bayesian analysis

One of the most useful features of Bayesian analysis is it's modular nature, provided by the correct normalization of probabilities through the evidence value. This was already encountered in the toy example above. In this section, however, a more complex conceptual example will be given to reflect the power of this approach. First, however, it is important to once more clarify that the evidence value is rigorously defined as the prior-weighted average likelihood of all propositions that are evaluated. In order to function in the way described below, proper priors have to be used so that the sum over all weights in the calculation of the evidence add up to unity.

Fig. 2.1 shows a fictional situation in which four different fundamental hypotheses $\mathcal{H}_{\mathrm{I}}$ to $\mathcal{H}_{\mathrm{IV}}$ are being evaluated. Each hypothesis has a distinct set of models M_i, and every model has several different parameters $\theta = \{\theta_1, \theta_2, ..., \theta_n\}$ that allow for up to several different parameter value combinations $C_{\theta,1}$, $C_{\theta,2}$, and so on. Probabilistic inference allows us to test this collection of models at every layer. The reason for this is that the likelihood terms originate at the smallest scale, i.e., at the parameter values in every model. For instance, after obtaining some data D relevant to these fundamental hypotheses, evaluating the parameter values $C_{\theta,1}$ for M_1 of $\mathcal{H}_{\mathrm{I}}$ leads to

$$P\left(C_{\theta,1}|D, M_1, \mathcal{H}_{\mathrm{I}}, I\right) = \frac{P\left(C_{\theta,1}|M_1, \mathcal{H}_{\mathrm{I}}, I\right) P\left(D|C_{\theta,1}, M_1, \mathcal{H}_{\mathrm{I}}, I\right)}{P\left(D|M_1, \mathcal{H}_{\mathrm{I}}, I\right)}. \qquad (2.17)$$
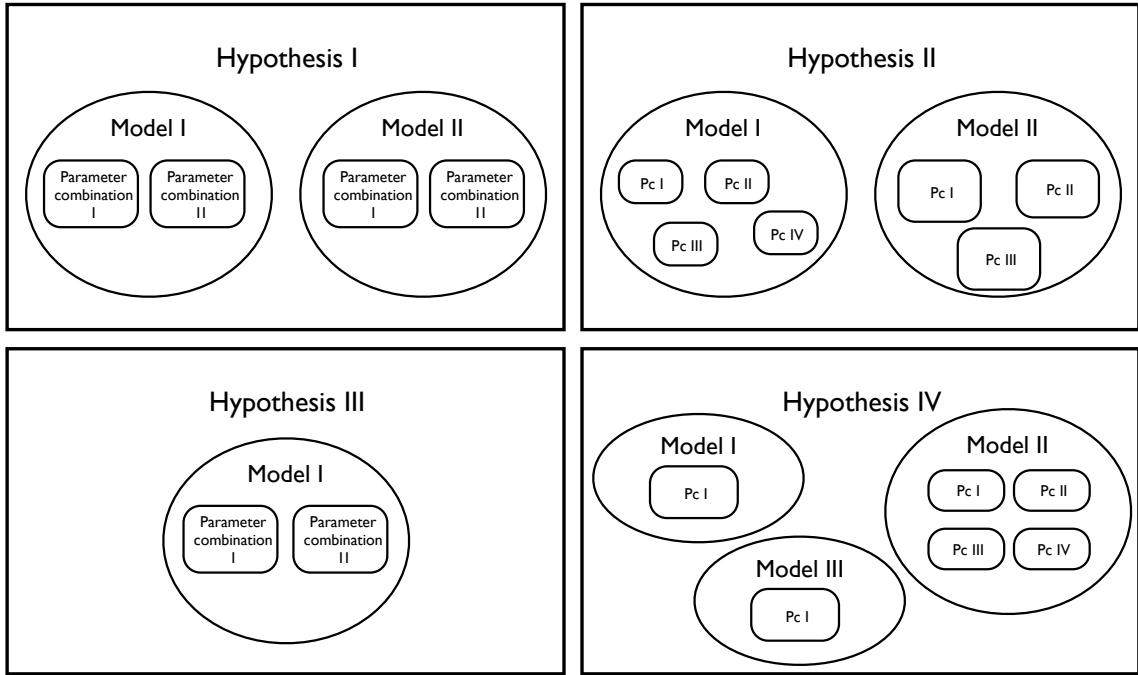
Figure 2.1: Cartoon representation of the modular nature of Bayesian analysis. Four different hypotheses, each with different model structure, down to the level of individual combination of parameter values (Pc), can be compared thanks to the correct normalization of the posterior probabilities and the evidence.

Here, the evidence $P\left(D|M_1, \mathcal{H}_{\mathrm{I}}, I\right)$ is equivalent to the sum of the numerator over all parameter combinations for $\mathcal{H}_{\mathrm{I}}$[9]. Note, however, that this evidence is formally equivalent to the likelihood of data given the model M_1 and the fundamental hypothesis $\mathcal{H}_{\mathrm{I}}$. Therefore, it is easy to evaluate the posterior probability for M_1 by calculating

$$P\left(M_1|D, \mathcal{H}_{\mathrm{I}}, I\right) = \frac{P\left(M_1|\mathcal{H}_{\mathrm{I}}, I\right) P\left(D|M_1, \mathcal{H}_{\mathrm{I}}, I\right)}{P\left(D|\mathcal{H}_{\mathrm{I}}, I\right)} \tag{2.18}$$

using the evidence values obtained at the more granular parameter value level. At this higher level of evaluating the M_i, however, the new evidence is obtained by summing over all the different models within $\mathcal{H}_{\mathrm{I}}$, rather than the parameter values. Consequently, it is equivalent to the likelihood of the data given the hypothesis $\mathcal{H}_{\mathrm{I}}$ as a whole[10]. It is therefore possible to again go up a layer and to directly compare the four different hypotheses $\mathcal{H}_{\mathrm{I}}$ to $\mathcal{H}_{\mathrm{IV}}$. The posterior probability for $\mathcal{H}_{\mathrm{I}}$, for instance, amounts to

$$P\left(\mathcal{H}_{\mathrm{I}}|D, I\right) = \frac{P\left(\mathcal{H}_{\mathrm{I}}|I\right) P\left(D|\mathcal{H}_{\mathrm{I}}, I\right)}{P\left(D|I\right)}. \tag{2.19}$$

This time, the evidence is equivalent to the sum over all four fundamental hypotheses.

The structure of the argument is also summarized in Fig. 2.2. Through the likelihood the observed data first affects the posterior probability of the parameter values. The evidence from this stage of the analysis then enters as the likelihood at the model level. This, in return, affects the posterior probability of the models and the evidence at the model layer. Lastly, the latter becomes the likelihood at the hypothesis layer, finally producing the posterior probabilities for the four fundamental hypotheses[11]. For the remaining chapters in this thesis, the evidence and the modular nature of the Bayesian approach are of fundamental importance.

---

[9]Note that for M_1 and M_3 of $\mathcal{H}_{\mathrm{IV}}$ there is only one set of parameter values. Hence for those models, the priors $P\left(C_{\theta,1}|M_1, \mathcal{H}_{\mathrm{IV}}, I\right) = 1$ and $P\left(C_{\theta,1}|M_3, \mathcal{H}_{\mathrm{IV}}, I\right) = 1$, and the likelihood is equivalent to the evidence.

[10]Just like $\mathcal{H}_{\mathrm{IV}}$ for the parameter values, $\mathcal{H}_{\mathrm{III}}$ is a special case at the level of individual models. It only has one model M_1. Therefore, $P\left(M_1|\mathcal{H}_{\mathrm{III}}, I\right) = 1$ and the evidence is equivalent to the likelihood of the model M_1.

[11]As the equations and the figure shows, each level also requires its own set of proper priors.
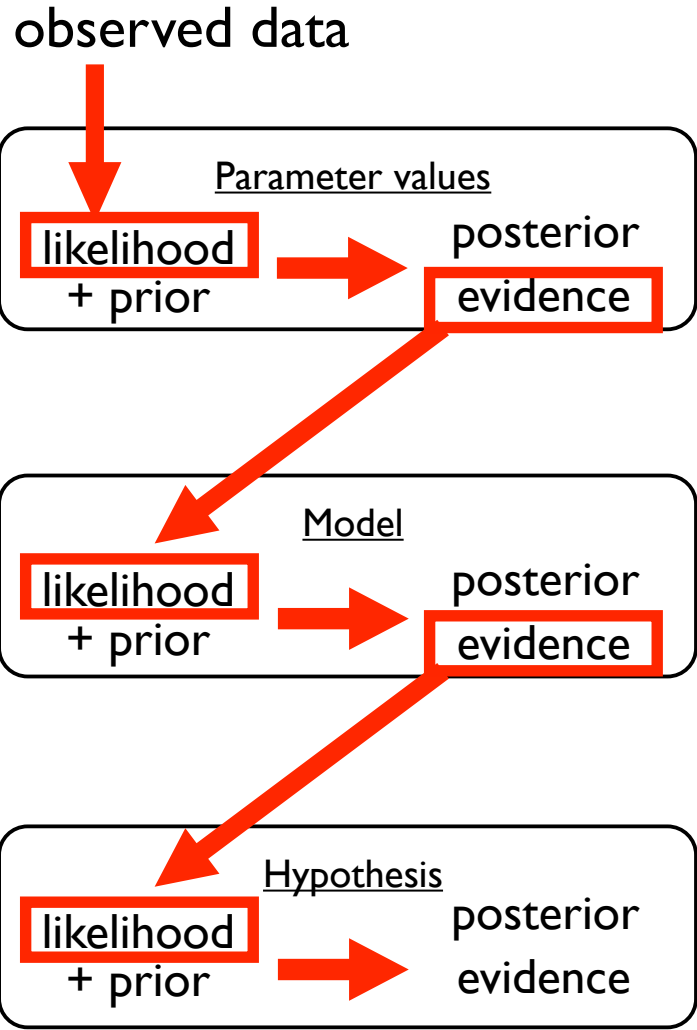
Figure 2.2: Flow of information from the observed data to the posterior probabilities of the different levels from Fig. 2.1.

2.3.4 Bayesian treatment of systematic errors

As will be shown in the following chapters, one of the central problems of asteroseismology is the assumption that the observables can in principle match the predictions from the model. However, in many modelling problems, and also in asteroseismic modelling, this not always the case. Therefore, a consistent treatment of systematic errors is important to properly assess the information provided by the

observations. In his book, Gregory provides a very straightforward way to solve this problem. Referring back to the toy example given in the previous section, a systematic error appears once the difference between model and observations cannot be reduced to random uncertainties. Therefore, instead of assuming $T_{\text{eff,obs}} = T_{\text{eff,calc}} + e$, the correct ansatz must be something like

$$T_{\text{eff,obs}} = T_{\text{eff,calc}} + e + \Delta, \tag{2.20}$$

where Δ is now some term describing the systematic error. In the simplest case, Δ could be a constant offset. In other cases, it could be a complicated function of $T_{\text{eff,calc}}$. Whatever the case might be, Δ can be included in the Bayesian analysis as a new dimension of the problem. Hence, again referring to our toy problem, the posterior probability is now not only a function of $\mathcal{S}$ or $\mathcal{N}$ but also of Δ. Since Δ is now part of the analysis, prior probabilities for different values of Δ have to be supplied that reflect the state of information about the systematic error. *Marginalization can then be used to integrate out the systematic errors, which takes their impact on the result fully into account without having to choose particular values of Δ.* On the other hand, Bayesian analysis can also be used to infer the most probable values of Δ. Thanks to the correct normalization of all probabilities to the parameter space of the investigation, adding systematic error parameters like Δ comes at a penalty (Gregory calls this a built-in "Occam's Razor"). All these properties for the analysis of systematic errors will be crucial in the analysis of the "surface effect" observed in the Sun and other Sun-like stars, as described in the later chapters.

2.4 Examples for Bayesian inference in asteroseismology

The most common application of Bayesian analysis in asteroseismology is the extraction of mode parameters for solar-type pulsators, which was already discussed in Chapter 1. The reason for this is two-fold and both are connected to the CoRoT data of the star HD 49933. This object was observed for 60 days, in the "initial run"

of CoRoT, and one of the first stars for which clear detection of solar-like oscillations was possible. Soon after this initial run, Appourchaux et al. (2008) presented a first analysis of the data, using a maximum likelihood-based approach to extract the mode parameters from the power spectrum. However, it turned out that the line widths (mode lifetimes) in HD 49933 were incredibly broad (short), so that there was no clear distinction between the $l = 0$ and $l = 1$ modes. With narrower line widths, such a distinction would have appeared either through rotational splitting in the $l = 1$ track, or through the extistence of an $l = 2$ track next to the $l = 0$ frequencies. Based on the maximum likelihood approach, Appourchaux et al. (2008) produced a favoured mode identification scenario and claimed high precision due to small frequency uncertainties.

However, soon thereafter, Bayesian analyses appeared (Gruberbauer et al. 2009; Benomar et al. 2009a; Kallinger et al. 2010) that showed strong evidence against the scenario proposed by Appourchaux et al. Gruberbauer et al. (2009) used Bayesian Markov-Chain Monte Carlo (MCMC) to extract the mode parameters which led to much larger uncertainties, and later on Kallinger et al. (2010) used these results to give a pulsation model-based argument against Appourchaux et al. Benomar et al., on the other hand, who also used an MCMC approach, calculated the evidence for the data, given both possible mode identification scenarios. This also turned out to contradict Appourchaux et al., as did an analysis based on even more data (Benomar et al. 2009b). In summary, these results showed that the single-value based maximum likelihood approach was inferior to a full evaluation of the parameter space. Bayesian analysis was already proposed even before the interpretation of the CoRoT observations were put into question (Brewer et al. 2007). Nonetheless, it obviously required a specific problematic object such as HD 49933, for which the probabilistic inference was clearly advantageous, to raise the community's awareness for these tools.

The same effect as for HD 49933 has since been identified in many F stars (thus it came to be known as the "bloody F-star" problem), and calculating the Bayesian evidence for both mode scenarios is now the standard approach to analyse these

stars (Handberg & Campante 2011; Appourchaux et al. 2012a). A different Bayesian approach for arriving at the correct mode identification has since been developed by White et al. (2012). Their method does not extract the individual mode properties but uses certain features in the power spectrum as well as spectroscopic constraints to calculate the evidence for both scenarios.

Further applications of Bayesian analysis in asteroseismology have been pursued for the analysis of the interplay between stellar granulation and pulsation (Kallinger & Matthews 2010), mode linewidths (Appourchaux et al. 2012b), amplitude scaling laws (Corsaro et al. 2013), basic stellar modelling (Bazot et al. 2008), and even more detailed modelling based on asteroseismic observables (Quirion et al. 2010). However, no asteroseismic modelling technique has yet reaped the full benefits of the Bayesian formalism. In the next chapters of this thesis, I therefore present several papers that rely on a completely new probabilistic approach to grid-based asteroseismic modelling. It makes use of all the key features of Bayesian analysis, from marginalization to the modular use of the evidence, and applies them to shed new light on some of the problems and paradigms of asteroseismology.

Bibliography

Appourchaux, T., et al. 2008, A&A, 488, 705

—. 2012a, A&A, 543, A54

—. 2012b, A&A, 537, A134

Bazot, M., Bourguignon, S., & Christensen-Dalsgaard, J. 2008, Mem. Soc. Astron. Italiana, 79, 660

Benomar, O., Appourchaux, T., & Baudin, F. 2009a, A&A, 506, 15

Benomar, O., et al. 2009b, A&A, 507, L13

Brewer, B. J., Bedding, T. R., Kjeldsen, H., & Stello, D. 2007, ApJ, 654, 551

Corsaro, E., Fröhlich, H.-E., Bonanno, A., Huber, D., Bedding, T. R., Benomar, O., De Ridder, J., & Stello, D. 2013, MNRAS

Cox, R. T. 1961, The Algebra of Probable Inference (Johns Hopkins)

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support, ed. Gregory, P. C. (Cambridge University Press)

Gruberbauer, M., Kallinger, T., Weiss, W. W., & Guenther, D. B. 2009, A&A, 506, 1043

Handberg, R., & Campante, T. L. 2011, A&A, 527, A56

Jaynes, E. T., & Bretthorst, G. L. 2003, Probability Theory, ed. Jaynes, E. T. & Bretthorst, G. L. (Cambridge University Press)

Kallinger, T., Gruberbauer, M., Guenther, D. B., Fossati, L., & Weiss, W. W. 2010, A&A, 510, A106

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

Kuhn, T. S. 1970, The structure of scientific revolutions (University of Chicago Press)

McGrayne, S. B. 2011, The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy (Yale University Press)

Quirion, P.-O., Christensen-Dalsgaard, J., & Arentoft, T. 2010, ApJ, 725, 2176

Smolin, L. 2006, The trouble with physics : the rise of string theory, the fall of a science, and what comes next (Houghton Mifflin Harcourt)

White, T. R., et al. 2012, ApJ, 751, L36

Chapter 3

Paper I: Toward a New Kind of Asteroseismic Grid Fitting

M. Gruberbauer[1], **D. B. Guenther**[1]

[1]Institute for Computational Astrophysics, Department of Astronomy and Physics, Saint Mary's University, B3H 3C3 Halifax, Canada

T. Kallinger[2,3]

[2]Instituut voor Sterrenkunde, K.U. Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium

[3]Institute for Astronomy, University of Vienna, Türkenschanzstrasse 17, 1180 Vienna, Austria

Co-author contributions

- DG: construction of solar model grids, valuable discussion.

- TK: valuable discussion.

Abstract

Recent developments in instrumentation (e.g., in particular the *Kepler* and *CoRoT* satellites) provide a new opportunity to improve the models of stellar pulsations. Surface layers, rotation, and magnetic fields imprint erratic frequency shifts,

trends, and other non-random behavior in the frequency spectra. As our observational uncertainties become smaller, these are increasingly important and difficult to deal with using standard fitting techniques. To improve the models, new ways to compare their predictions with observations need to be conceived. In this paper we present a completely probabilistic (Bayesian) approach to asteroseismic model fitting. It allows for varying degrees of prior mode identification, corrections for the discrete nature of the grid, and most importantly implements a treatment of systematic errors, such as the "surface effects." It removes the need to apply semi-empirical corrections to the observations prior to fitting them to the models and results in a consistent set of probabilities with which the model physics can be probed and compared. As an example, we show a detailed asteroseismic analysis of the Sun. We find a most probable solar age, including a 35 ± 5 million year pre-main sequence phase, of 4.591 billion years, and initial element mass fractions of $X_0 = 0.72$, $Y_0 = 0.264$, $Z_0 = 0.016$, consistent with recent asteroseismic and non-asteroseismic studies.

3.1   Introduction

The success of recent space missions *CoRoT* and *Kepler*, designed for the discovery of exoplanets and the analysis of stellar pulsation, have produced a large number of high-quality light curves (Chaplin et al. 2010). With these data sets, obtained over long time bases of several months, we are able to detect variability with semi-amplitudes down to a few parts per million. These observations have now firmly established the existence of solar-type pulsation in a large number of solar-like and red-giant stars. Moreover, observations of an unprecedented number of δ Scuti stars and other types of pulsators have also revealed rich mode spectra.

These data are now causing a paradigm shift for many topics in stellar astrophysics. In particular, the determination of fundamental stellar parameters, and any inferences regarding the physics of stellar interiors, have for a long time been restricted to testing theoretical models using classic observables such as photometric

indices or spectroscopic data. Even though these methods have become more advanced, for instance by applying complex Bayesian methods to determine stellar ages (Pont & Eyer 2004; Jørgensen & Lindegren 2005) and to evaluate competing models (Takeda et al. 2007; Bazot et al. 2008), the value of additional information provided by pulsation modes is tremendous, as they directly probe the whole star. Already, the asteroseismic community is successful in extracting general characteristics of the mode spectra for many different types of stars (e.g., Mathur et al. 2010; Kallinger et al. 2010c) and also in devising promising tools for a comparative interpretation of the observations (e.g., Bedding & Kjeldsen 2010). Average mode parameters, such as the large and small frequency separations, and the frequency of maximum power, have been shown to successfully constrain stellar parameters although certain correlations remain as a source for uncertainty (see, e.g., Kallinger et al. 2010b; Huber et al. 2011; Gai et al. 2011). These have been incorporated into the current advanced probabilistic pipelines to investigate stellar model grids (Quirion et al. 2010) and already been applied to recent observations (Metcalfe et al. 2010). The next step to improving our knowledge about stellar interiors is to analyze individual pulsation modes in an equally rigorous way, to see where our models agree or disagree.

In the past, χ^2-minimization techniques (Guenther & Brown 2004), or equivalent Bayesian analyses (e.g., Kallinger et al. 2010a), have been introduced to find the pulsation model that most closely reproduce the observed frequencies within a large and dense grid of models. The Bayesian analysis, in this context, only provides an additional framework for constraining solutions to models that match our prior knowledge about the stars' fundamental parameters. Due to the rich information provided by the pulsation frequencies, these approaches should be successful in many cases, which is why they are being applied also to the most recent *Kepler* data sets. For instance, Metcalfe et al. (2010) test various approaches from different modelers with different methods that actually use the individual frequencies. However, there are currently (at least) three major problems when applying these techniques.

Stellar rotation, at all but the slowest rotation speeds, has been shown to produce rotational splittings which are incompatible with the traditional linear approximations. It even perturbs the values of the axisymmetric ($m = 0$) frequencies (e.g., see Deupree & Beslin 2010, and references therein). In order to correctly take this into account, the rotation speed as a function of stellar depth needs to be known, and extensive computations would be necessary to do these effects justice. Given the large variety of possible rotation profile characteristics, this would greatly expand the dimensionality and size of the pulsation model grid. This implies currently insurmountable computational expenses for the types and sizes of grids that are necessary for a comprehensive asteroseismic analysis of many stars.

For stars with a convective envelope, model frequencies at high radial orders differ from observations due to problems in modeling the outer layers (see Figure 3.1). These so-called *surface effects* can be compensated by looking at ratios of frequency differences (Roxburgh 2005), or by "correcting" the observed frequencies through calibration of the surface effects seen for the Sun as proposed by Kjeldsen et al. (2008). It is likely that the surface correction as calibrated for the Sun is not universally applicable, and evidence for this has been mounting (e.g., Bedding et al. 2010). Moreover, neglecting (or correcting for) the surface effects in the observed frequencies is only reasonable when studying properties of the star for which the outer layers are unimportant. However, if we want the theoretical models to more closely reflect reality, we need to include more and better physics to bring the computed frequencies closer to the (un-corrected) observed ones.

Furthermore, the fact that *static asteroseismic grids can only have a finite resolution* in parameter space is often neglected. If the error bars of the observed frequencies are small compared to the differences between calculated frequencies in adjacent grid points, the likelihood of having a model in the grid that corresponds to the best model one's code could deliver decreases rapidly. The problem of finding the "true" model and the actual uncertainties with respect to the grid becomes apparent. Even grids with adaptive resolution have the same problem in principle, as the decision
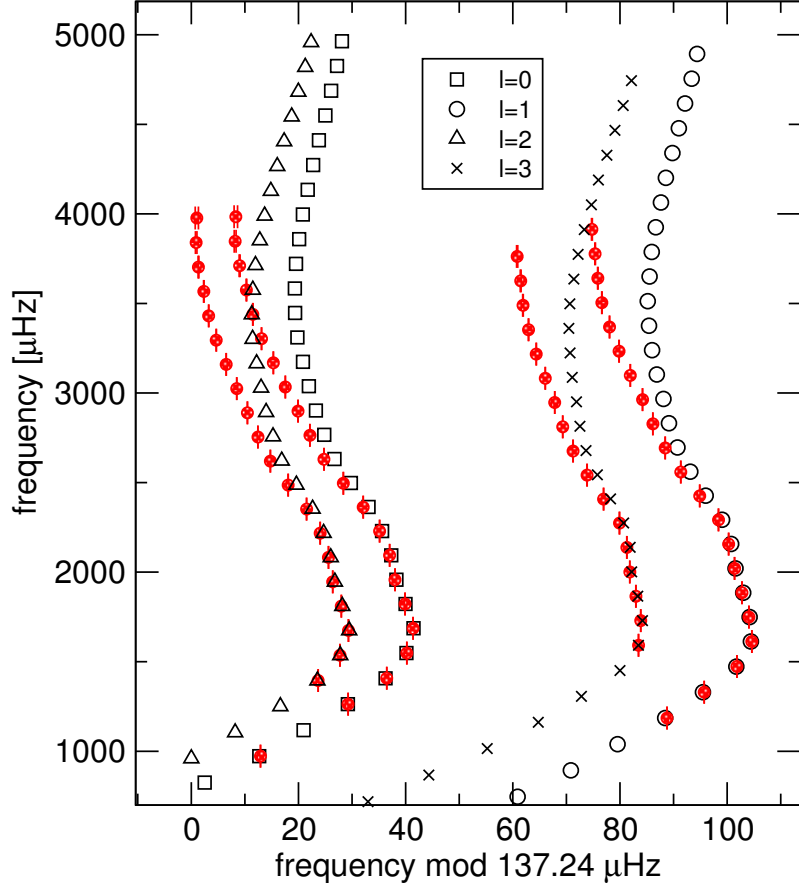
Figure 3.1: Echelle diagram of solar p modes taken from Broomhall et al. (2009) (filled circles) and an appropriate solar model constructed using YREC. The higher order model frequencies are increasingly deviating from the observations due to deficiencies in modeling the upper stellar layers. The systematic errors of the models are much bigger than the random observational uncertainties.

for further refining the resolution of a particular region in parameter space must always depend on a number of discrete grid points. This problem is much more severe if our aim is to calculate probabilities (or some summary statistics) to compare different model grids.

In this paper we present a new approach to asteroseismic model grid fitting. Our goal is to find a new way of putting our model physics to the test that can handle all of the aforementioned difficulties. Even restricted to models that are unable to produce all the details of the observations, we want to know which models are most "correct" (i.e., consistent with appropriate fundamental parameters and physics),

and how well the solution is constrained. We show how to quantitatively assess our model grids as a function of the observational uncertainties, the uncertainties of the calculated frequencies, and our general prior knowledge about the star and possible shortcomings of our models.

3.2 Bayesian treatment of systematic errors

3.2.1 Basics of Bayesian inference

Bayes' theorem, applied to the problem of inference, states that the probability of a particular hypothesis after obtaining new data (i.e., the posterior) is proportional to the probability of the hypothesis prior to obtaining the new data (i.e., the prior) times the likelihood of obtaining the new data, under the assumption that the hypothesis is true (i.e., the likelihood function). This approach to inference is derived from the product and sum rules of probability theory that have shown to be necessary and sufficient for consistent, quantitative logical reasoning[1](see Jaynes & Bretthorst 2003).

In this paper, we stay as close as possible to the general notation used in Jaynes & Bretthorst (2003) or Gregory (2005). We start with Bayes' theorem applied to the problem of comparing observations with the predictions of a model M. If the predictions of a model M are governed by a set of n parameters $\boldsymbol{\theta} = \{\theta_1, ..., \theta_n\}$, and we define the observations to be represented by the symbol D (for data), it is commonly formulated by expressing the posterior probability

$$P(\boldsymbol{\theta}|M, D, I) = \frac{P(\boldsymbol{\theta}|M, I)\, P(D|\boldsymbol{\theta}, M, I)}{P(D|M, I)}.$$
(3.1)

The symbol I is equivalent to the prior information about the problem that is investigated. The first term in the numerator of Equation (3.1) is the prior probability of

[1]Strictly speaking, Bayes' theorem is only one result that derives from these rules. Consistent use of Bayes' theorem, in particular the assignment of the various terms in Equation (3.1), also requires knowledge of its origin and consistent application of the product and sum rule. However, for the sake of brevity we will simply call our approach in this manuscript to be "Bayesian" rather than "based on probability theory as extended logic".

a particular set of parameter values $\boldsymbol{\theta}$, given the model M and our prior information I about the problem. It is independent of any new data which are supposed to be analyzed. The second term in the numerator is called the likelihood. It gives the likelihood of obtaining the observed data under the assumption that the predictions of model M are correct, given the particular choice of its parameter values $\boldsymbol{\theta}$. The denominator in Equation (3.1) is called the global likelihood, or *evidence*, and is the sum (or integral) of the numerator over the whole parameter space of model M. It therefore acts as a normalization constant. Most importantly, if the prior probabilities are adequately normalized, it also represents the likelihood of obtaining the data given the *whole* model M, independent of the particular choice of $\boldsymbol{\theta}$. Thus, it can be used as a likelihood for comparisons among different alternative models.

More details on the application of Bayes' theorem, in particular with respect to data analysis in astronomy, can be found in Gregory (2005).

3.2.2   Systematic errors in the Bayesian framework

One of the strengths of the Bayesian framework is that a parameter θ_n, known to be necessary to describe a model M, can be marginalized by applying the sum rule. In the case of continuous parameters, the sum turns into an integral, and by integrating the full posterior over the parameter range of θ_n, one obtains the marginal posterior

$$P(\theta_1, ..., \theta_{n-1}|M, D, I) = \int P(\theta_1, ..., \theta_{n-1}, \theta_n|M, D, I) \, d\theta_n. \tag{3.2}$$

The marginal posterior retains the overall effects of including parameter θ_n in the model, but is independent of any particular choice of its value. In other words, θ_n is "removed" from the detailed analysis. This is similar to what is done for calculating the evidence in the denominator in Equation (3.1). The only difference is that the evidence is the marginal likelihood over all parameters of the model, weighted by the prior.

The reason this is useful is that if the data and the model are known to show systematic differences, like shifts or trends, such "systematic errors" can simply be

encoded introducing additional parameters to the model M, so that M is able to model these effects as well. By subsequently marginalizing over these "fudge parameters", one is then able to perform a standard Bayesian analysis without any need for knowing the exact value of the systematic error(s). However, even though the exact value is unknown, the presence of the error is being considered in the evaluation of the posterior probabilities. Furthermore, an increasing number of "fudge parameters" comes at a cost, because it potentially decreases the evidence for the model due to the increase in prior volume. As mentioned in Section 3.2.1 the evidence is used as the value for the likelihood of obtaining the data in Bayesian model comparison. It is therefore possible to compare models with and without "fudge parameters". Improved models that do not need them, but are able to explain the observations just as well, will be favored.

3.3 Toward a Bayesian solution to asteroseismic model fitting

3.3.1 Review and problems of the standard approach

The general problem of asteroseismic model fitting is to match observed frequencies $f_{i,\mathrm{o}}$ to those calculated from models $f_{i,\mathrm{m}}$. If the n_{obs} observed frequencies have individual uncertainties $\sigma_{i,\mathrm{o}}$, and the model frequencies have random uncertainties $\sigma_{i,\mathrm{m}}$ then a χ^2-statistic can be calculated according to

$$\chi^2 = \frac{1}{n_{\mathrm{obs}}} \sum_{i=1}^{n_{\mathrm{obs}}} \frac{(f_{i,\mathrm{o}} - f_{i,\mathrm{m}})^2}{\sigma_{i,\mathrm{o}}^2 + \sigma_{i,\mathrm{m}}^2}. \tag{3.3}$$

Searching a large grid of N stellar models M_j with fundamental parameters close to those estimated for the observed star will produce a minimum in χ^2 ($=$ best-fit model). In addition, uncertainties can be estimated from the change in χ^2 as the distance in parameter space to the best fit increases. Calculated with adequate stellar evolution

and pulsation codes, it should be possible to infer details about the stellar interior and to obtain precise fundamental parameters.

In order to consistently encode prior information about the fundamental parameters and other model properties, and to make use of all the additional advantages that come with the Bayesian approach (all of which will become clear in the next section), it is much easier to perform the model fitting using probabilities. Assuming that the *random* uncertainties $\sigma_{i,\mathrm{o}}$ and $\sigma_{i,\mathrm{m}}$ are compatible with Normal distributions, one can define

$$\sigma_i^2 = \sigma_{i,\mathrm{o}}^2 + \sigma_{i,\mathrm{m}}^2. \tag{3.4}$$

This leads to the likelihood for observing the data ($=$ the specific values of $f_{i,\mathrm{o}}$), given a single observed and calculated frequency

$$P(f_{i,\mathrm{o}}|f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j, I) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(f_{i,\mathrm{o}} - f_{i,\mathrm{m}})^2}{2\sigma_i^2}\right]. \tag{3.5}$$

Here, $f_{i,\mathrm{o}\mapsto i,\mathrm{m}}$ stands for the proposition "*The observed mode $f_{i,\mathrm{o}}$ corresponds to the calculated mode $f_{i,\mathrm{m}}$.*"[2] Naturally, we want our models M_j to reproduce all observed frequencies. Assuming that each observed frequency is a statistically independent datapoint, this leads to a product for the likelihood of obtaining all observed frequency values given that the model is correct

$$P(D|M_j, I) = \prod_{i=1}^{n_{\mathrm{obs}}} P(f_{i,\mathrm{o}}|f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j, I). \tag{3.6}$$

Here, D stands for complete set of observed frequencies and their uncertainties. This can then be incorporated in the usual framework for Bayesian inference.

---

[2]Although the explicit notation seems clumsy at first glance, it is actually one of the major assets of the Bayesian approach. It visualizes exactly which propositions we are evaluating, and under which conditions the probabilities are calculated. Slightly different propositions or conditions can yield vastly different results. If the notation is explicit, there are no hidden variables or assumptions.

Alas, both of the mentioned, straightforward approaches above suffer from the following problems:

1. The most appropriate model is not necessarily the one that minimizes Equation (3.3) or maximizes Equation (3.6). There are many possible scenarios where this would be the case (e.g., due to surface effects, stellar activity, magnetic field effects, rotational effects). Straightforward application of the formalism above will then lead to wrong or nonsensical results in both best fit and derived uncertainties. Even worse, this would propagate into our assessment of the model physics that were used to produce the models.

2. In case of such systematic differences, we need to take into account that for each observed pulsation frequency, multiple model frequencies are possible candidates (not necessarily only the closest one). This is particularly problematic in cases were no prior mode identification is available.

3. As the observational uncertainties decrease, the contrast in χ^2 (and even more so the contrast in probabilities) between different models increases. If the model that minimizes/maximizes Equation (3.3)/Equation (3.6) is not the correct model due to missing physics, this increase in fitting contrast is misleading and unwarranted.

4. In static grids the finite grid resolution increases the risk of missing the most adequate model that the code could produce. If there are systematic differences between even the most adequate mode and the observations, the "contrast enhancement effect" will be magnified. For the same reason, adaptive grids run into the same problem and will miss the correct parameter space region to finer resolve in the first place.

As a consequence of all these shortcomings, it is clear that a method is needed that considers the *possibility of systematic differences*. It is also mandatory to consider the finite resolution of our model grids. Solutions to these problems are presented in the following sections.

3.3.2   The argument for probabilities

There are obvious benefits to quantifying the best fit and the uncertainties in terms of probabilities. With probabilities for each specific model, we automatically obtain probability distributions for each parameter of the model itself. We can furthermore consistently compare different grids and see which set of input physics is more probable, given all our current information and the data.

However, there are much stronger arguments for a probabilistic approach. Marginalization allows us to consistently treat nuisance parameters, while the sum and product rules allow us to clearly formulate the question we are asking. This question is "Given the observed frequencies, our knowledge about the star and model physics, which model(s) best represent the star in terms of its fundamental parameters and general physical properties as probed by the pulsation modes?" In reality, this general question has to be further refined as we encounter more complicated situations like: "We have model frequencies that could potentially show negative or positive systematic offsets, or no such offset at all, when compared to our observations. They could be influenced by rotation or actually be rotationally split frequencies themselves. They could be bumped $l = 1$ modes or $l = 0$ modes. Given all of these possibilities, which model is the most adequate one, and how well is the solution constrained?". From the viewpoint of probability theory the only way to treat such a set of possibilities and get meaningful answers is to use the sum rule and product rule, as we will show in the next section.

3.3.3   Ambiguous mode identification

As a first improvement to the general approach of asteroseismic model fitting, we can involve the sum rule to consistently consider uncertainties (or even ignorance) in our mode identification. In essence, if there is no unique proposition $f_{i,\mathrm{o}\mapsto i,\mathrm{m}}$,

Equation (3.6) changes to

$$P(D|M_j, I) = \prod_{i=1}^{n_{\mathrm{obs}}} \left\{ \sum_{k=1}^{n_{\mathrm{match}}} P(f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j, I) \right\} \tag{3.7}$$

with

$$\begin{aligned} P(f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j, I) = \\ P(f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j, I)P(f_{i,\mathrm{o}}|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j, I). \end{aligned} \tag{3.8}$$

Here the sum over the index k means that all possible and mutually exclusive assignments n_{match} of one observed mode to a number of calculated frequencies $f_{k,\mathrm{m}}$ have to be taken into account as an "or" proposition[3]. Note that due to the product rule of probability, the terms in each sum now include the conditional prior probabilities $P(f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j, I)$. These have to be normalized so that $\sum_{k=1}^{n_{\mathrm{match}}} P(f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j, I) = 1$. The most conservative assignment is to assign equal probabilities $P(f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j, I) = 1/n_{\mathrm{match}}$ to each possible scenario. However, if more information is available (e.g., a mode could be identified to be either $l = 0$ or $l = 2$ with specific probabilities for both cases as found by some peak bagging program), this can easily be encoded at this stage.

The end result is a product of weighted sums of probabilities, where the weights are given by the respective prior probabilities.[4] This product is the correctly normalized likelihood for obtaining the data, given the proposition that any one of the proposed scenarios is correct. Note that if there is an unambiguous assignment $f_{i,\mathrm{o}\mapsto i,\mathrm{m}}$ for every observed frequency, each prior probability $P(f_{i,\mathrm{o}\mapsto i,\mathrm{m}}|M_j, I) = 1$ and Equation (3.7) simplifies to Equation (3.6). Now that we have included our uncertainties concerning the assignment of model frequencies and observed frequencies, we will

---

[3]Hereafter, a possibly ambiguous frequency assignment will always be denoted as $f_{i,\mathrm{o}\mapsto k,\mathrm{m}}$.

[4]A common misconception is that these "priors" are only there to allow us to incorporate prior information. In reality, they are formally required by the product rule and ensure that the result of Equation (3.7) is always properly normalized.

deal with uncertainties in the validity of the model frequencies themselves in the next section.

3.3.4   Treatment of systematic errors

As a next step, we now show how to treat the problem of imperfect models. As mentioned before, applying standard techniques that rely on minimizing the quadratic differences between the observations and the models will give incorrect results if systematic differences exist. The alternative of correcting for such imperfections prior to modeling is also undesirable if the correction is not known to be universally applicable.

To treat any systematic deviation from the model frequencies due to unmodeled physical effects, we simply expand the models M_j by considering an additional systematic error parameter for each tested frequency. The aim is to construct new values $f_{i,\Delta}$ to compare with the observations according to

$$f_{i,\Delta} = f_{i,\mathrm{m}} + \gamma \Delta_i \tag{3.9}$$

Here, Δ_i is the absolute value of the systematic error. $\gamma = 1$ or $\gamma = -1$ and determines whether the model frequency is expected to be systematically higher or lower than the observed frequency. To keep our notation from occupying too much space, we will implicitly assume the value of γ to be constant throughout the following derivations, and attribute this to our prior information I. Δ_i is an unknown parameter but as long as its lower and upper boundaries can be roughly estimated, it can be treated fully consistently in the probabilistic framework.

In the following, we will again work out an example of only one observed and calculated frequency. Therefore, for the derivation the assignment $f_{i,\mathrm{o}\mapsto i,\mathrm{m}}$ is unique. We will then provide the extension to multiple frequencies and ambiguous mode identifications.

Using the new parameters, the equivalent to Equation (3.5) is

$$
\begin{aligned}
&P(f_{i,\mathrm{o}}, \Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I) = \\
&P(\Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I) P(f_{i,\mathrm{o}} | \Delta_i, f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I) = \\
&P(\Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I) \times \\
&\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(f_{i,\mathrm{o}} - f_{i,\mathrm{m}} - \gamma\Delta_i)^2}{2\sigma_i^2} \right].
\end{aligned}
\tag{3.10}
$$

Here the symbol M_j^{Δ} simply denotes the model M_j augmented by the new parameter Δ_i. Self-evidently, the product rule again requires that we introduce a prior probability $P(\Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I)$. This can either encode prior information about the expected behavior of the error, or be simply assigned by considerations of symmetry. Again it is required that the integral over the prior $\int P(\Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I) \, d\Delta_i = 1$.

It would now be possible to try to find the Δ_i that maximizes $P(f_{i,\mathrm{o}}, \Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I)$ in Equation (3.10). However, this is completely irrelevant for our needs. In case of multiple observed frequencies it would also quickly lead to a highly dimensional parameter space that we are not interested in navigating. Instead, we are interested in finding the probabilities of the models M_j^{Δ}. To do this it is necessary to integrate out Δ_i which we have just introduced. We obtain the marginal likelihood

$$
\begin{aligned}
&P(f_{i,\mathrm{o}} | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I) = \\
&\int_{\Delta_{i,\mathrm{min}}}^{\Delta_{i,\mathrm{max}}} P(f_{i,\mathrm{o}}, \Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, M_j^{\Delta}, I) \, d\Delta_i.
\end{aligned}
\tag{3.11}
$$

This integral naturally depends on the shape of the prior probability distribution for Δ_i, and can easily be evaluated numerically[5]. It represents the likelihood of obtaining the value of the observed frequency $f_{i,\mathrm{o}}$ given that M_j predicts a frequency $f_{i,\mathrm{m}}$ but that there is a possibility of a systematic difference Δ_i, between $\Delta_{i,\mathrm{min}}$ and $\Delta_{i,\mathrm{max}}$. Furthermore, it is fundamentally constrained and properly weighted by the prior we

---

[5]For several simple shapes, such as the *beta* prior introduced in the next section, there also exist analytical solutions.

assigned. This result is now easily extended to multiple modes and ambiguous mode identifications. Equation (3.7) becomes

$$P(D|M_j^\Delta, I) = \prod_{i=1}^{n_{\mathrm{obs}}} \left\{ \sum_{k=1}^{n_{\mathrm{match}}} P(f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j^\Delta, I) \right\} \tag{3.12}$$

and

$$
\begin{aligned}
&P(f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j^\Delta, I) = \\
&P(f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j^\Delta, I)P(f_{i,\mathrm{o}}|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j^\Delta, I).
\end{aligned} \tag{3.13}
$$

In summary, we have to calculate a product of weighted sums of integrals in the form of Equation (3.11), where the summation is performed over every possible assignment $f_{i,\mathrm{o}\mapsto k,\mathrm{m}}$.

Note that even when we choose to consider systematic deviations, we usually do not expect them to be significant for all frequencies. For good models some frequencies should already match well "right out of the box". In particular, this is true for all frequencies in the idealized case where we have (finally) found a way to correctly model all the effects that previously caused systematic deviations.

One might think that this is taken care of by setting $\Delta_{i,\mathrm{min}} = 0$. However, unless the prior $P(\Delta_i|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j^\Delta, I)$ is a δ function at $\Delta_i = 0$, it is much more likely that $\Delta_i > 0$. This means that a priori a model will be preferred which shows at least a small deviation from the observations, depending on the observational uncertainties and the steepness of the prior. The limiting case however, the δ function, corresponds to a whole different model which is simply the standard model without systematic deviations, M_j. Thanks to the sum rule, there is an elegant solution for taking this alternative into account.

For the mutually exclusive logical propositions[6] M_j^Δ and M_j we can calculate

$$
\begin{aligned}
&P(f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j^\Delta + M_j, I) = \\
&\frac{P(M_j^\Delta, f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|I) + P(M_j, f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|I)}{P(M_j^\Delta|I) + P(M_j|I)} = \\
&\frac{P(M_j^\Delta|I)}{P(M_j^\Delta|I) + P(M_j|I)} P(f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j^\Delta, I) + \\
&\frac{P(M_j|I)}{P(M_j^\Delta|I) + P(M_j|I)} P(f_{i,\mathrm{o}}, f_{i,\mathrm{o}\mapsto k,\mathrm{m}}|M_j, I).
\end{aligned}
\tag{3.14}
$$

Note that here $M_j^\Delta + M_j$ means "M_j^Δ or M_j is true". This is the likelihood of observing the frequency value $f_{i,\mathrm{o}}$, given that a systematic deviation either does or does not exist. The principle of indifference as the most conservative approach for the prior probabilities obviously demands $P(M_j^\Delta|I) = P(M_j|I) = 0.5$, but if more information is available, it can be encoded here. This result is also easily generalized to the case of multiple frequencies and ambiguous mode identification.

3.3.5 The choice of the prior for Δ_i

A very important detail to consider when extending the models with systematic error parameters is their prior probabilities $P(\Delta_i|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j^\Delta, I)$. There is a basic choice between two possibilities. The first is to use uninformative (or ignorance) priors, or alternatively, maximum entropy priors. Uninformative priors can be derived from arguments of invariance to specific transformations, while maximum entropy priors should satisfy the maximum entropy criterion for a given set of constraints. The other possibility is to use priors derived from heuristic or physical arguments.

The specific form of the prior probabilities of Δ_i are part of the model that is evaluated, as indicated by the notation.[7] They are not necessarily "prior" as in a sense of "before obtaining observations", but conditional probabilities required for the correct normalization, as demanded by the product rule of probabilities. They

---

[6]Note that from a logical standpoint even if $\Delta = 0$, M_j^Δ is still a different model than M_j because the prior is not a δ function. Therefore, they are always mutually exclusive.

[7]The prior is described by $P(\Delta_i|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j^\Delta, I)$ rather than, e.g., $P(\Delta_i|I)$

encode specific ways in which we expect Δ_i to behave, given our grid of frequencies and our information (which of course can be influenced by previous observations). For instance, if we expect our best model to minimize the systematic deviations, the prior should assign larger probability densities to smaller Δ_i, so that models with smaller deviations will be more probable. On the other hand, if we expect our best model to show more erratic deviations, a flat uninformative prior is a better choice. After a complete evaluation of the probabilities and likelihoods, the Bayesian evidence will indicate whether the state of information encoded by the priors is supported by the data or not.

As a first important example of an uninformative prior, consider a uniform prior

$$P(\Delta_i|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j^{\Delta}, I) = \frac{1}{\Delta_{i,\mathrm{max}} - \Delta_{i,\mathrm{min}}} = \text{const.} \qquad (3.15)$$

This means that all values of Δ_i are equally likely. With such a prior, every model that predicts frequencies at any value between $f_{i,\mathrm{o}} + \Delta_{i,\mathrm{min}}$ and $f_{i,\mathrm{o}} + \Delta_{i,\mathrm{max}}$ has the same maximum likelihood (i.e., the same maximum value for Equation (3.10)).

On the other hand, a Jeffreys prior

$$P(\Delta_i|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j^{\Delta}, I) = \frac{1}{\Delta_i \ln\left(\Delta_{i,\mathrm{max}}/\Delta_{i,\mathrm{min}}\right)}, \qquad (3.16)$$

assigns equal probability per decade and, in terms of the probability density, favors smaller values of Δ_i. This prior is obviously not defined for $\Delta_i = 0$, i.e., it requires $\Delta_{i,\mathrm{min}} > 0$. This is problematic for, e.g., surface effects that approach zero at low orders. However, when $\Delta_{i,\mathrm{min}} = 0$, one can use a modified version of this prior given by

$$P(\Delta_i|f_{i,\mathrm{o}\mapsto k,\mathrm{m}}, M_j^{\Delta}, I) = \frac{1}{(\Delta_i + c) \ln\left[(\Delta_{i,\mathrm{max}} + c)/c\right]}, \qquad (3.17)$$

where c is a small constant. For values smaller than c, this prior acts more or less like a uniform prior, while for higher values it behaves like the usual Jeffreys prior. This prior is nowadays often used in "peak-bagging" algorithms (e.g., Gruberbauer et al. 2009; Benomar et al. 2009; Handberg & Campante 2011). However, there is

no objective criterion for how to set c, and various tests we conducted with our grid fitting code have shown that the choice of c can have a large impact on the evidence values.

Consequently, we argue that any priors used for a systematic error parameter $\Delta_i = [0, \Delta_{i,\text{max}}]$ should be functions that are clearly defined by the parameter limits, without additional parameters that have large effects on the evidence. The uniform prior[8] is such a prior, as are priors derived from the beta distribution (given in units of our problem)

$$P(\Delta_i | f_{i,\text{o}\mapsto k,\text{m}}, M_j^\Delta, I) \propto \left(\frac{\Delta_i}{\Delta_{i,\text{max}}} \right)^{\alpha-1} \left(1 - \frac{\Delta_i}{\Delta_{i,\text{max}}} \right)^{\beta-1}. \tag{3.18}$$

With $\alpha = 1$ and $\beta = 2$ this simply leads to a linearly decreasing probability density

$$P(\Delta_i | f_{i,\text{o}\mapsto k,\text{m}}, M_j^\Delta, I) = \left(\frac{\sqrt{2}}{\Delta_{i,\text{max}}} \right)^2 (\Delta_{i,\text{max}} - \Delta_i). \tag{3.19}$$

It is the only prior that allows for a linearly decreasing probability density, is properly normalized, and reaches zero at $\Delta_i = \Delta_{i,\text{max}}$. It also leads to an analytical solution for the integral in Equation (3.11). Thus, it satisfies all our requirements for a prior with which to minimize systematic errors.

We have compared the results obtained from Equation (3.19) (hereafter: *beta* prior) with several other plausible choices, such as an exponential distribution with expectation value $\Delta_{i,\text{max}}/2$ and a modified Jeffreys prior with $c = \sigma_{i,\text{m}}$, and we find them to yield comparable results and evidence values. Due to the clarity of its definition and lack of additional parameters, we therefore argue that the *beta* prior is an appropriate choice for a non-flat prior. We will show how to use it in Section 3.4 in a worked example.

Lastly, priors based on heuristic or physical arguments obviously vary strongly with the specific problem to which the fitting method is applied. As an example, when

---

[8]In fact, the uniform prior is consistent with a beta distribution with $\alpha = \beta = 1$.

modeling surface effects on p-mode frequencies, the prior could be Gaussian, following the heuristic frequency correction proposed by Kjeldsen et al. (2008). It would be a function of frequency, expecting greater deviations toward higher-order modes. The width of the Gaussian, however, would be again a rather arbitrary choice, leading to potentially different evidence values. Such priors clearly need to have a strong basis either in theory or prior observations.

3.3.6 Bumped modes and finite grid resolution

Equation (3.12) represents the final likelihood for obtaining the observed frequencies given our (extended) model M_j^Δ. This model still represents only a single point in a discrete grid.[9] However, the probability is small that a single model in the grid corresponds to the "true best model" our code can produce. The problem becomes worse as the grid resolution is lowered, or as mode frequencies are changing quickly *or unpredictably* from one model to the next (e.g., avoided crossings, magnetic shifts). The probabilities (or χ^2-values) we obtain will not be a fair assessment of the model physics, even at higher grid resolutions. Even worse, the overall evidence for the grid will be finely tuned to the positions of all models in the grid. This makes it difficult to compare different grids with different physics. We will now show how to improve on this.

In a sequence of models along a single evolutionary track, except for the first and last models, each model M_j has two neighboring models M_{j-1} and M_{j+1}. In most cases these adjacent models will contain the same modes, and their changing values can be traced from M_{j-1} to M_j and M_{j+1}. Now we declare the difference between observed and calculated frequency as a new free parameter

$$\delta f_i = f_{i,\mathrm{o}} - f_{i,\mathrm{m}}. \tag{3.20}$$

[9]Note that this is also the case for approaches using an adaptive grid, since each iteration of an adaptive scheme is based on a discrete grid.

This value is fixed if only a single grid point is considered. However, we can split the evolutionary tracks into segments in between grid points, and define

$$\delta f_{i,j-} = f_{i,\text{o}} - \frac{f_{i,M_{j-1}} + f_{i,M_j}}{2} \tag{3.21}$$

and equivalently

$$\delta f_{i,j+} = f_{i,\text{o}} - \frac{f_{i,M_j} + f_{i,M_{j+1}}}{2}. \tag{3.22}$$

Adding δf_i as a new parameter to the equations derived in the earlier sections, we change our focus to evaluate probabilities of model track segments T_j^{Δ} centered around the models M_j^{Δ}. To do this, we again use marginalization to integrate out both Δ_i and δf_i to obtain the marginal likelihood. We obtain

$$\begin{aligned}
P(f_{i,\text{o}}|f_{i,\text{o}\mapsto i,\text{m}}, T_j^{\Delta}, I) = \\
\int_{\Delta_{i,\text{min}}}^{\Delta_{i,\text{max}}} \int_{\delta f_{i,j-}}^{\delta f_{i,j+}} P(f_{i,\text{o}}, \Delta_i, \delta f_i | f_{i,\text{o}\mapsto i,\text{m}}, T_j^{\Delta}, I) \, d\Delta_i \, d\delta f_i.
\end{aligned} \tag{3.23}$$

If the priors for Δ_i do not vary greatly from one model to the next, Δ_i and δf_i can be considered to be independent parameters. It is therefore possible to use the product rule to separate the conditional probabilities

$$\begin{aligned}
P(f_{i,\text{o}}, &\Delta_i, \delta f_i | f_{i,\text{o}\mapsto i,\text{m}}, T_j^{\Delta}, I) = \\
&P(\Delta_i | f_{i,\text{o}\mapsto i,\text{m}}, T_j^{\Delta}, I) \times \\
&P(\delta f_i | f_{i,\text{o}\mapsto i,\text{m}}, T_j^{\Delta}, I) \times \\
&P(f_{i,\text{o}} | \Delta_i, \delta f_i, f_{i,\text{o}\mapsto i,\text{m}}, T_j^{\Delta}, I).
\end{aligned} \tag{3.24}$$

Furthermore, since we evaluate the complete evolutionary track segment we can assume a uniform prior probability $P(\delta f_i | f_{i,\text{o}\mapsto i,\text{m}}, T_j^{\Delta}, I) = 1/(\delta f_{i,j+} - \delta f_{i,j-})$. With these definitions, the integral over δf_i can easily be calculated analytically. The equivalent to Equation (3.10) becomes

$$P(f_{i,\mathrm{o}}, \Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, T_j^\Delta, I) =$$

$$P(\Delta_i | f_{i,\mathrm{o}\mapsto i,\mathrm{m}}, T_j^\Delta, I) \times$$

$$\frac{1}{2\left(\delta f_{i,j+} - \delta f_{i,j-}\right)} \times \tag{3.25}$$

$$\left[\mathrm{erf}\left(\frac{\delta f_{i,j+} - \gamma\Delta_i}{\sqrt{2}\sigma_i}\right) - \mathrm{erf}\left(\frac{\delta f_{i,j-} - \gamma\Delta_i}{\sqrt{2}\sigma_i}\right)\right].$$

where erf is the error function. The remaining integral over Δ_i again has to be carried out numerically. Figure 3.2 shows an example for the definitions introduced above, given three models in a solar evolutionary track.

We have now used the free parameter δf_i to "trace" each mode through segments of the evolutionary track, and compare it to the observed frequencies, retaining the possibility of systematic differences. Note that our only assumption here is that the mode frequencies change smoothly between the frequencies given by the constraining models. In principle, this approach can be carried out in multiple dimensions (e.g., not only along the evolutionary track in stellar age but also between tracks in mass). As before, an extension to multiple frequencies and ambiguous mode identifications is straightforward.

We stress that this approach only locates the region of highest probability given the current grid, and given unspecified behavior of frequencies in between grid points. It is thus best used for frequencies whose behavior is difficult to capture, e.g., due to mode bumping, or for a first general assessment of a very coarse grid. Given a dense enough grid, regular frequencies that are expected to change approximately linearly from one grid point to the next need to be treated using interpolation, since the integration over the model gaps for individual modes, independently of all other modes, would allow for highly unphysical models.

Therefore, in order to obtain a final best model and uncertainties for the model parameters, the regions of substantial probability should be further refined after the track probabilities have been calculated. Eventually, the grid is resolved
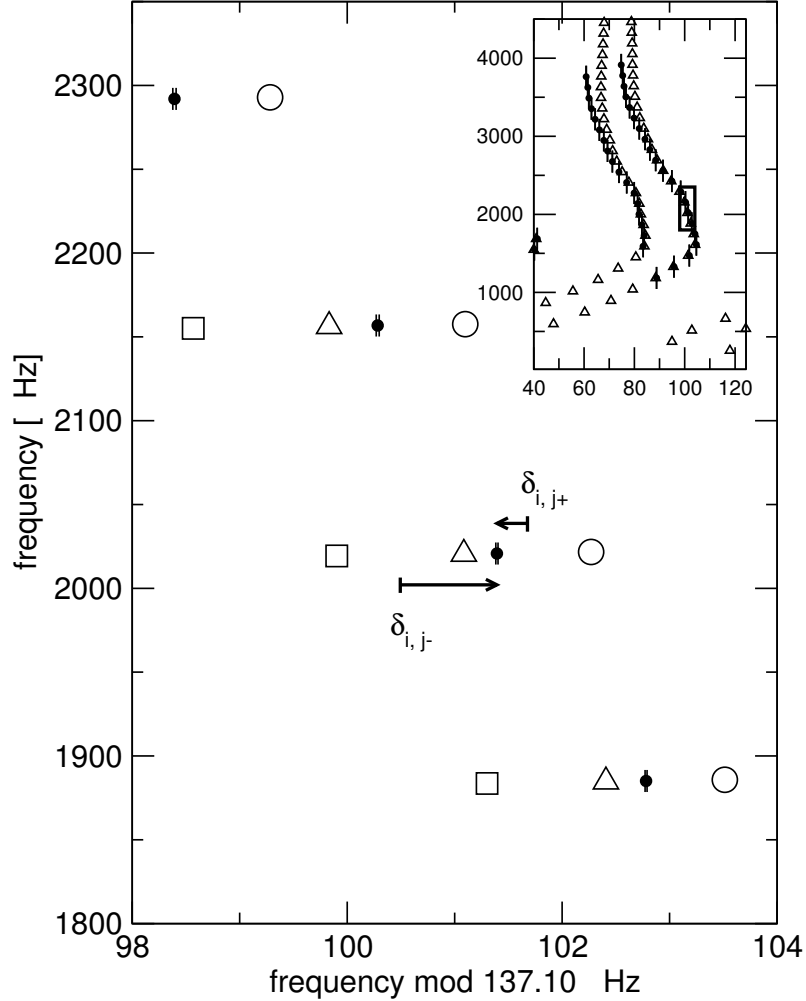
Figure 3.2: Example for the definition of $\delta_{i,j-}$ and $\delta_{i,j+}$ (see the text). Four radial orders of $l = 1$ modes from three adjacent models in a high-resolution grid of solar models are shown. The triangles represent the central model. The frequencies for the adjacent models along the evolutionary track sequence are depicted as squares and white circles. Black circles (and error bars) indicate observed frequencies published in Broomhall et al. (2009). $\delta_{i,j-}$ and $\delta_{i,j+}$ for a single mode are represented using arrows. The insert shows an unzoomed version of the $l = 1$ and $l = 3$ ridge.

enough so that well-defined distributions arise. In dense enough grids, this can easily be accomplished by interpolation of the frequencies in between grid points without violating the condition of hydrostatic equilibrium. This can also be done during run-time with arbitrary precision using probabilities, by interpolating between grid points and using the sum rule to calculate a probability representative of the original grid resolution using the interpolated models. Naturally, modes that change erratically,

66

should be excluded from such an interpolation routine and treated as shown above instead.

3.3.7 Model probabilities

So far we have only shown how to calculate the *likelihood* for standard pulsation models, models that contain systematic differences, and also evolutionary track segments. In order to obtain the *probabilities* for individual models (or track segments), we want to use Bayes' theorem, assign model priors, and calculate the total evidence for each model grid. The simplest method to assign model priors in the absence of any other prior information, is to use the principle of equipartition and assign a uniform prior

$$P(M_j|I) = 1.0/N_{\mathrm{M}}, \tag{3.26}$$

where N_{M} is the number of models (or, equivalently, N_{T} would be the number of track segments) that are analyzed.

Although each model or track only predicts a number of frequencies, it implicitly represents values or ranges for fundamental parameters like T_{eff} or L, which can be compared to (and constrained by) different and non-seismic observations. For instance, assuming our prior photometric and spectroscopic observations of a pulsating star indicate $T_{\mathrm{eff}} = T_{\mathrm{spec}} \pm \sigma_{\mathrm{spec}}$ then the prior probability density for the model temperature is

$$P(T_{\mathrm{eff},j}|I) = k \exp\left[-\frac{(T_{\mathrm{spec}} - T_{\mathrm{eff},j})^2}{2\sigma_{\mathrm{spec}}^2}\right]. \tag{3.27}$$

This example assumes that the uncertainty in T_{eff} follows a Gaussian distribution. k is a normalization constant depending on the absolute lower and upper plausible limits of T_{eff}.

All the different implicit parameters for which prior observations or other fundamental constraints are available, and hence prior knowledge exists, then can be used for prior probabilities which combine into an overall prior for model M_j. As an

example

$$P(M_j|I) = P(T_{\text{eff},j}|I)P(L_j|I)P([\text{Fe/H}]_j\,|I)... \tag{3.28}$$

if we assume separable priors for simplicity. If probabilities of track segments are calculated, such a prior could be approximated by a product of separable integrals, which are easily evaluated analytically. Again, in order to obtain a proper prior and therefore proper values for the evidence, the integral of the prior probability over all possible models/tracks in a grid should be 1.

By calculating, e.g.,

$$P(M_j^\Delta|D, I) = \frac{P(M_j^\Delta|I)P(D|M_j^\Delta, I)}{\sum_{k=1}^{N_\text{M}} P(M_k^\Delta|I)P(D|M_k^\Delta, I)} \tag{3.29}$$

or, e.g., in the case of rapidly changing modes with or without systematic errors,

$$P(T_j^\Delta + T_j|D, I) = \frac{P(T_j^\Delta + T_j|I)P(D|T_j^\Delta + T_j, I)}{\sum_{k=1}^{N_\text{T}} P(T_k^\Delta + T_j|I)P(D|T_k^\Delta + T_j, I)} \tag{3.30}$$

we obtain the probability of M_j^Δ or, respectively, $T_j^\Delta + T_j$ given our prior knowledge (or lack thereof), our grid, and the set of observed frequencies. Note that the denominators of these equations represent the evidence or likelihood for the grid as a whole. We can therefore use these as *likelihoods* when we want to compare different grids with different input physics.

3.4 Application to surface effects

As mentioned in the introduction, shortcomings in modeling the outer stellar layers produce systematic deviations in comparison to the observations. These deviations seem to be such that model frequencies tend to be higher than the observed frequencies, and therefore $\gamma = -1$ (see Equation (3.9)). Kjeldsen et al. (2008) have proposed to calibrate a power-law description of the deviations by measuring the surface effects in the Sun, and then fitting this relation to frequencies of other stars.

Their correction expressed in terms of our definitions has the general form of

$$\gamma \Delta_i \approx a \left(\frac{f_{i,\mathrm{m}}}{f_{\mathrm{ref}}} \right)^b, \tag{3.31}$$

where f_{ref} is some reference frequency, typically the frequency of maximum power ν_{max}, and a and b are parameters to be fitted. From their fits, Kjeldsen et al. determined $b \approx 4.90$ in the Sun, which has subsequently been used for other stars by a number of authors. A comprehensive implementation of this formalism into a χ^2-fitting algorithm was presented in a recent study by Brandão et al. (2011). However, even in this more advanced approach, there is still a choice of a and b required. Moreover, complications for modes of different spherical degree and also bumped modes arise because they do not necessarily conform to this relation. The authors propose to alleviate these problems by introducing additional model-dependent parameters that approximately correct for some of these deviations. While this approach is a great improvement over applying a fixed surface-effect correction (or no correction at all), our approach is much more powerful. It allows for a much greater flexibility and leads to clearly defined probabilistic results.

3.4.1 Priors for surface effects

As we want to treat systematic errors of more or less unknown magnitude, the most general approach is to use the flat uniform prior (Equation (3.15)). Imposing only minor additional constraints, as we argued in Section 3.3.5, the *beta* prior (Equation (3.19)) can also be used to give more weight to models which minimize these unknown errors. We can use both priors and compare the Bayesian evidence to tell us which interpretation of the surface effect is better supported by the data, given our model and everything we know. Moreover, irrespective of which prior is chosen, we also always allow for the possibility of no surface effects at all, as discussed in Section 3.3.4.
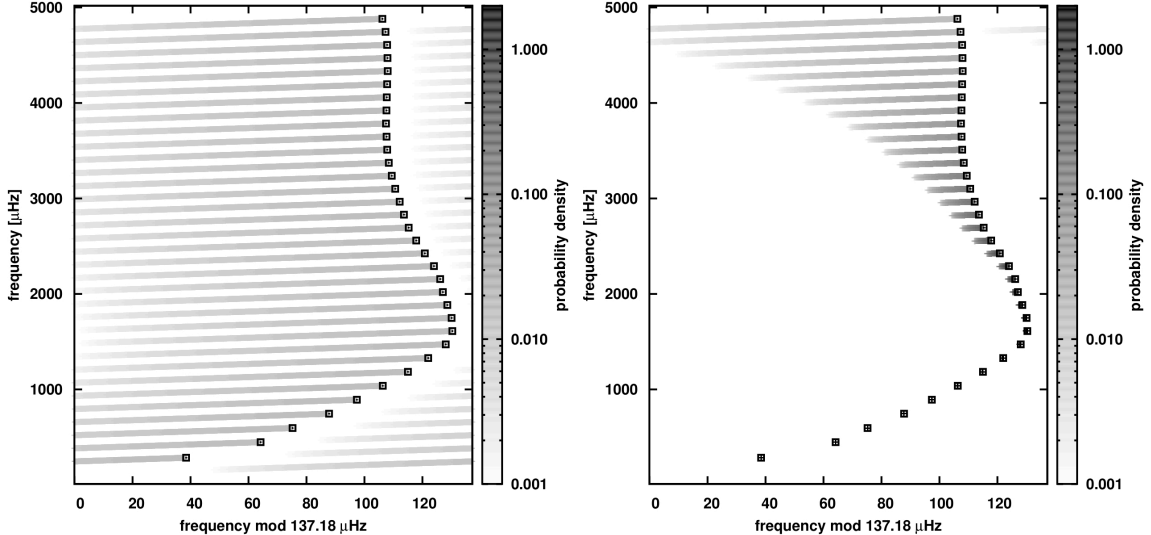
Figure 3.3: Behavior of the *beta* prior for systematic offsets in an echelle diagram. The squares represent model frequencies, while the shaded "trails" indicate the prior probability density for varying Δ_i. The left panel uses equal $\Delta_{\max}$, whereas the right panel uses a power-law $\Delta_{\max}$ with exponent $b = 4.9$ (see Section 3.4.1). Note that the uniform prior is not shown, since it simply assigns a constant probability density.

This now gives us enough flexibility to consider a possibly frequency-dependent behavior of the surface effects. However, instead of "predicting" the behavior of Δ_i as is done by modeling the surface deviations through a power law, we will prescribe the behavior of its *upper limit* $\Delta_{i,\max}$. In contrast, the lower limit should always remain 0, since our ultimate goal is to find models that correctly describe the surface layers (and therefore approach $\Delta_i \to 0$).

The choice of the largest allowed $\Delta_{\max}$ is not unique, but it should be sensible and used consistently throughout the analysis. A reasonable strategy is to use $\sup(\Delta_{\max}) = \Delta\nu$, the large frequency separation of each specific model, as a sensible upper limit. If the systematic differences between observations and models are larger than the average distance between modes of adjacent radial order, we no longer recognize this as a valid frequency assignment[10]. With this upper limit defined, we now want to model different types of surface effects. If we have no preference for any

---

[10]This condition may be relaxed at the highest radial orders.

frequency-dependent trend (i.e., all we know is that observed frequencies are lower than model frequencies) we require that all frequencies have equal Δ_{max}.

On the other hand we can also use a more specific model, such as Equation (3.31), but retain the same flexibility. The surface effect as shown in Equation (3.31) depends on two parameters. The power-law exponent b determines how quickly the surface effect increases as we move to higher frequencies, whereas a is simply a scaling factor. We are not interested in the scaling parameter, since the scaling (i.e., the magnitude of the offset) is governed by our condition that for each model $\sup(\Delta_{\mathrm{max}}) = \Delta\nu$. It is taken care of by the fact that we are marginalizing over Δ_i anyway. Since the largest surface effects are expected at the highest frequency $f_{\mathrm{max,m}}$ in the model, it follows that for a specific b

$$\Delta_{\mathrm{max},i} = \Delta\nu \left(\frac{f_{i,\mathrm{m}}}{f_{\mathrm{max,m}}} \right)^b. \tag{3.32}$$

Figure 3.3 shows how these definitions affect the prior probability density $P(\Delta_i | f_{i,\mathrm{o} \mapsto i,\mathrm{m}}, M_j^{\Delta}, I)$ as we increase the value of Δ_i for both the constant and the power-law approach. With all the $\Delta_{\mathrm{max},i}$ set, we can then use the priors as discussed above for all our calculations. Note that we can also easily evaluate new composite propositions at this stage and compute the probability for a hypothesis that allows for, e.g., a range $b = \{4.4, 4.65, 4.9, ...\}$. This is done in the same way as was explained earlier (see Equation (3.14)).

3.4.2 Detailed analysis of the Sun

As an example for how to implement the surface-effect treatment, we will consider the solar $l = 0$, 1, 2, and 3 p modes obtained by using BiSON data (Broomhall et al. 2009). For our models, we used a large and dense solar grid obtained with YREC (Demarque et al. 2008). The model grid spans: masses from 0.95 $M_\odot$ to 1.05 $M_\odot$ in steps of 0.005 $M_\odot$, initial hydrogen mass fractions from 0.68 to 0.74 in steps of 0.01, initial metal mass fractions from 0.016 to 0.022 in steps of 0.001, and mixing length

Table 3.1: Parameter Ranges for the Solar Grid.

Parameter	Range	Step Size
Mass	0.95–1.05	0.005
X_0	0.68–0.74	0.01
Z_0	0.016–0.022	0.001
α_{ml}	1.8–2.4	0.1

[0]**Notes.** Masses are given in units of solar masses; α_{ml} is the mixing length parameter.

parameters from 1.8 to 2.4 in steps of 0.1. These parameters are also summarized in Table 3.1.

Our model tracks begin as completely convective Lane–Emden spheres (Lane 1869; Chandrasekhar 1957) and are evolved from the Hayashi track (Hayashi 1961) through the zero-age main sequence (ZAMS) to 6 Gyr with each track consisting of approximately 2500 models. Only models between 4.0 and 6.0 Gyr are included in the model grid. Constitutive physics include the OPAL98 (Iglesias & Rogers 1996) and Alexander & Ferguson (1994) opacity tables using the GS98 mixture (Grevesse & Sauval 1998), and the Lawrence Livermore 2005 equation of state tables (Rogers 1986; Rogers et al. 1996). Convective energy transport was modeled using the Böhm-Vitense mixing-length theory (Böhm-Vitense 1958). The atmosphere model follows the (T–τ) relation by Krishna Swamy (1966). Nuclear reaction cross-sections are from (Bahcall et al. 2001). The effects of helium and heavy element diffusion (Bahcall et al. 1995) were included. Note that our atmosphere models and diffusion effects have been shown to require a larger value of mixing length parameter ($\alpha_{\mathrm{ml}} \approx 2.0 - -2.2$) than standard Eddington atmospheres ($\alpha_{\mathrm{ml}} \approx 1.7 - -1.8$) (Guenther et al. 1993).

The pulsation spectra were computed using the stellar pulsation code of Guenther (1994), which solves the linearized, non-radial, non-adiabatic pulsation equations using the Henyey relaxation method. The non-adiabatic solutions include radiative energy gains and losses but do not include the effects of convection. We estimate the random 1σ uncertainties of our model frequencies to be of the order of $0.1\,\mu$Hz.

We analyzed our grid using adiabatic and non-adiabatic frequencies, and employed three different surface-effect models:

- **M1**: frequency-independent surface effects with $\Delta_{i,\mathrm{max}} = \Delta\nu$

- **M2**: frequency-dependent, "canonical" surface effects with $\Delta_{i,\mathrm{max}}$ following Equation (5.3) with $b = 4.90$

- **M3**: same as **M2**, but with b as a free parameter marginalized from $b = 3.0$ to $b = 6.0$

For each frequency evaluated throughout our model grid, irrespective of the surface-effect model, we also considered the possibility of no surface effect, i.e., we consistently calculated $P(M_j^\Delta + M_j | D, I)$. To take into account the discrete nature of the grid, we interpolated along the evolutionary tracks during run-time by a factor of 20, thereby increasing the effective "frequency resolution" of the grid to below the random uncertainties of the model frequencies. All models were evaluated with

- **(a)** a uniform prior for all track segments

- **(b)** a prior using normal distributions for the constraints $M = 1.0000 \pm 0.0002\, M_\odot$, $\log T_{\mathrm{eff}} = 3.7617 \pm 0.01$, and $\log(L/L_\odot) = 0.00 \pm 0.01$, where YREC uses the following adopted values for $M_\odot = 1.9891 \pm 0.0004 \cdot 10^{33}$g (Cohen & Taylor 1986) and $L_\odot = 3.8515 \pm 0.009 \cdot 10^{33}$erg s^{-1} (the average of the ERB-Nimbus and SMM/ARCRIM measurements; Hickey & Alton (1983))

- **(c)** same as **(b)** but with an additional Gaussian constraint on the age of 4.603 ± 0.0075 Gyr, derived from the estimated age of the solar system found by Bouvier & Wadhwa (2010) and an average pre-main sequence phase of our models of 35 ± 5 Myr.

For the Δ_i we consistently used *beta* priors, as discussed in the previous section. Our calculations yield the most probable models and uncertainties for all these approaches, and they also give the Bayesian evidence for each approach. The results

Table 3.2: Evidence for the Solar Grid Using the BiSON Data Set

Surface Model	HRD Prior	$\log_{10}(evidence)$ (adiabatic)	$\log_{10}(evidence)$ (non-adiabatic)
M1	**a**	-233.4	-229.9
M2	**a**	-189.8	**-186.7**
M3	**a**	-189.8	-187.2
M1	**b**	-235.5	-231.6
M2	**b**	-190.8	**-187.1**
M3	**b**	-191.4	-187.9
M1	**c**	-236.7	-235.1
M2	**c**	-193.5	**-190.7**
M3	**c**	-192.6	-189.4

[0]**Notes.** See the text for the definition of models and prior a, b, and c. Results for models **M2a**, **M2b**, and **M2c**, which are analyzed in more detail, are indicated in bold face. Note that small numbers are expected.

are summarized in Table 3.2. We also computed the probabilities using uniform priors, but found similar results with lower evidence (several orders of magnitude) values than for the corresponding *beta* prior analysis.

The non-adiabatic frequencies consistently produce larger evidence values than for the respective adiabatic case. This is no surprise, as the non-adiabatic frequencies are in general better at reproducing the higher frequencies. Overall, model **M2a** shows the largest evidence, followed by **M2b** and **M3a**. Note that **M1a**, **M1b**, and **M1c**, which use frequency-independent priors for the surface effects, and therefore are extremely flexible, fail compared to the other models. Also, **M3a** and **M3b** cannot beat their **M2** counterparts. These are examples of how *marginalization and the consistent normalization of probabilities work together to penalize more flexible models if they cannot generate considerably better results.* Model M3c has a greater evidence than M2c, but the most probable stellar models are the same in both cases, suggesting that these models fit well, but do not necessarily adhere in detail to the standard surface correction.

Table 3.3: Most Probable Models for the Complete BiSON Data Set Model Fitting.

Model	Mass	Age	X_0	Z_0	Z_s	α_{ml}	Probability
M2a	1.015	4.885 ± 0.006	0.73	0.017	0.0153	2.2	0.54
	1.005	4.713 ± 0.006	0.72	0.017	0.0153	2.2	0.21
M2b	1.000	5.017 ± 0.006	0.72	0.018	0.0161	2.1	0.68
	1.000	4.983 ± 0.006	0.71	0.019	0.0160	2.2	0.17
M2c	1.000	4.591 ± 0.005	0.72	0.016	0.0144	2.2	0.95
	1.000	4.562 ± 0.005	0.71	0.017	0.0153	2.3	0.05

[0]**Notes.** Age is given in billion years and is computed from the pre-main-sequence birthline. The age from the ZAMS is 35 ± 5 million years less. X_0 and Z_0 are initial hydrogen and metal mass fractions, Z_s is the metal mass fraction in the envelope. Probabilities are given with respect to the specific surface-effect model and prior combination.

At first glance it might be unsettling that **M2a** has a slightly greater evidence than **M2b** (and significantly greater evidence than **M2c**). This indicates that there are models in our grid which reproduce the pulsation spectrum very well but do not match the solar fundamental parameters. A correctly calibrated grid would produce higher evidences with a prior restricted to the true solution. However, regardless of whether or not we include the fundamental parameter constraints, we are still finding models that match the oscillations constraints reasonably well. Furthermore, recall that the evidence is only the likelihood of obtaining the data, given that the approach is correct.[11] We know that the **a** prior is misrepresenting our state of information. The solar prior approach **b** more correctly encodes what we know about the Sun, and the age prior **c** puts even tighter constraints on the pulsation models. Ignoring this information (using prior **a** and setting equal conditional probabilities) is an interesting and necessary exercise to study the consistency of the results, and how the different models, approaches, and priors work. For an actual detailed study of the solar model physics, however, it is not appropriate. We can nonetheless compare the results,

[11]In order to obtain correctly normalized probabilities for the different approaches themselves, we have to introduce conditional probabilities like, e.g., $P(\mathbf{M2a}|I)$ or $P(\mathbf{M2b}|I)$ and use Bayes' theorem. Only comparing the evidence amounts to setting these conditional probabilities to be equal for all tested hypotheses (e.g., $P(\mathbf{M2a}|I) = P(\mathbf{M2b}|I)$).

restricting ourselves to the non-adiabatic frequencies and the on average best model for each prior, **M2**. The resulting parameters are displayed in Table 3.3.

The results obtained without using our prior knowledge of the Sun for model **M2a** are spread over several models in the parameter space that can fit the observations quite well. However, for the most probable models, the mass and age are inconsistent with our prior knowledge. These models seem to produce smaller surface effects, and are therefore preferred. For model **M2b** the situation is similar. Although the mass is now fixed to the true solar value, we do not obtain models that are consistent with the solar age.

For **M3c** a single combination of physical parameters dominates the results and manages to fit well all the constraints we impose (mass, luminosity, T_{eff}, pulsation frequencies, and age). Loosening the conditions on T_{eff} and the luminosity does not significantly change the result. We have also tested slight variations of up to 20 million years in the age prior and do not find the result to be affected. In all cases, we recover a tightly constrained most probable model with $Z_0 = 0.016$ and $Z_{\mathrm{s}} = 0.0144$, and an age of $4.591 \pm 0.005\,\mathrm{Gyr}$. We therefore find a result similar to Houdek & Gough (2011). Given that our models take $35 \pm 5\,\mathrm{Myr}$ to reach the main sequence, our result is also consistent with meteoritic age determinations of the solar system to within several million years (see, e.g. Bouvier & Wadhwa 2010). However, we also recover $X_0 = 0.72$, which leads to an initial helium mass fraction of $Y_0 = 0.264(1)$. This is different compared to the value of $Y_0 = 0.250(1)$ that was found by Houdek & Gough, but more consistent with Asplund et al. (2009).

Fitting the observations to the adiabatic frequencies, including the age prior, we also recover the exact same model. We also tested how sensitive the grid is to the prior constraints in order to estimate the actual impact of the pulsation frequencies on the probabilities. If we only evaluate the combined priors, ignoring the frequencies but including the prior on the age, we obtain $X_0 = 0.71 \pm 0.01$, $Z_0 = 0.019 \pm 0.002$, $Z_{\mathrm{s}} = 0.017 \pm 0.002$, age $= 4.603 \pm 0.008\,\mathrm{Gyr}$, and $\alpha_{\mathrm{ml}} = 2.1 \pm 0.2$. This leads us

to conclude that *the frequencies have a decisive impact and actually select the low-metallicity models, no matter whether adiabatic or non-adiabatic model frequencies are used.*

However, it has to be stressed again that the evidence drops by almost two orders of magnitude when we introduce the age prior. This can be understood by the fact that the solution is so well constrained and at the edge of our current parameter space in Z_0, and that many other models can also produce similar pulsation spectra. It could also suggest that we might not have covered the true best model parameters yet in our current grid. Therefore, our next goal will be to extend the grid to lower metallicities, and also include different abundance mixtures, but this is beyond the scope of this paper.

Figure 3.4 compares the BiSON observations with our most probable model at the correct solar age. Even with non-adiabatic frequencies, significant surface effects can still be found. The measured surface effects themselves are shown in Figure 3.5, together with least-squares fits following the relation proposed by Kjeldsen et al. (2008). The magnitude of the surface deviations depends on whether the non-adiabatic or the adiabatic frequencies are used for the fit. Nonetheless, our method manages to identify the same exact model to be the most probable, even using the same surface-effect model, thanks to the power of marginalization. However, the non-adiabatic models are vastly preferred in terms of the Bayesian evidence. This is an example for how the presented approach can be used to iterate toward improved stellar model physics, while still recovering meaningful stellar parameters from current asteroseismic investigations.

We also determined surface-correction power-law exponents for every spherical degree via least-squares fits. For both the non-adiabatic and adiabatic frequencies the best fitting exponents are markedly different from $b = 4.9$ which was both advocated by Kjeldsen et al. (2008) and also used as the basis for our probabilistic surface model **M2**. This is also the reason why the **M3c** models have a greater evidence than their **M2c** counterparts. The fitted values range from $b = 4.23$ for non-adiabatic ($l = 0$)
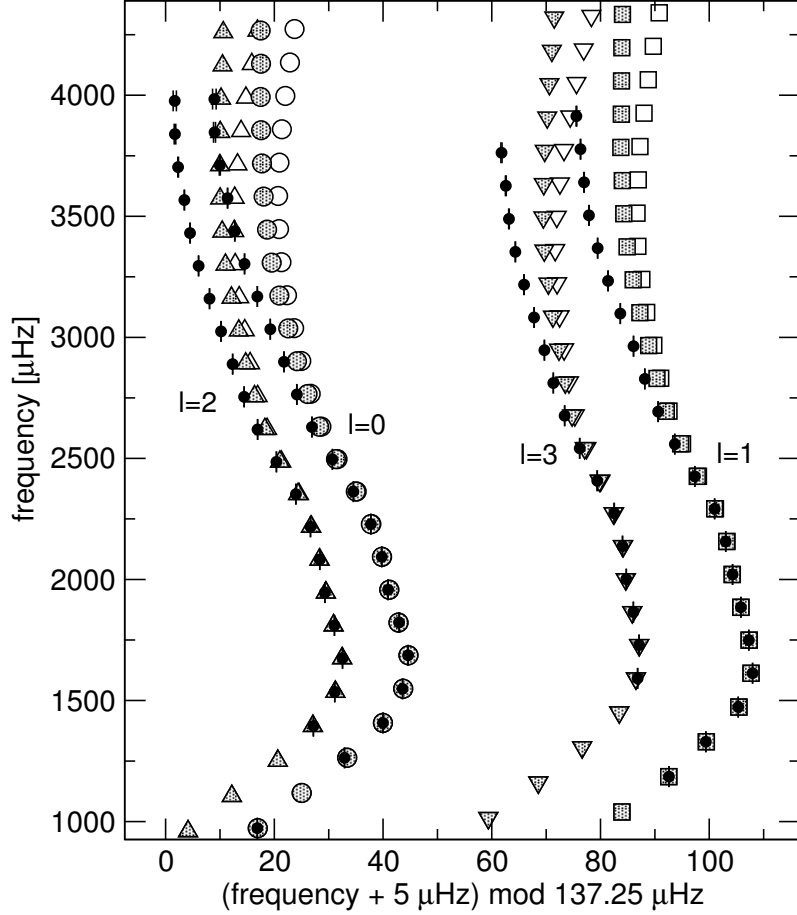
Figure 3.4: Non-adiabatic (shaded symbols) and adiabatic (open symbols) frequencies of the most probable solar model from evaluating the BiSON frequencies (black circles + error bars) using approach **M2c**. Note that frequencies have been shifted upward by $5\,\mu$Hz, before calculating the x-axis values in order to prevent the $l = 2$ modes from wrapping around.

frequencies to $b = 5.13$ for adiabatic ($l = 3$) frequencies. Moreover, the power-law fits do not match the deviations very well at intermediate radial orders near $2400\,\mu$Hz. From our point of view, fixing the exponent to $b = 4.9$ for a least-squares fit, as for instance done by Brandão et al. (2011), is therefore a potential problem since it does not even match the Sun very well, in particular when improved (e.g., non-adiabatic) physics are implemented. The probabilistic procedure has no problem with these deviations, even though it formally assumes an exponent of $b = 4.9$, since the magnitude of the surface effects is marginalized for every frequency.
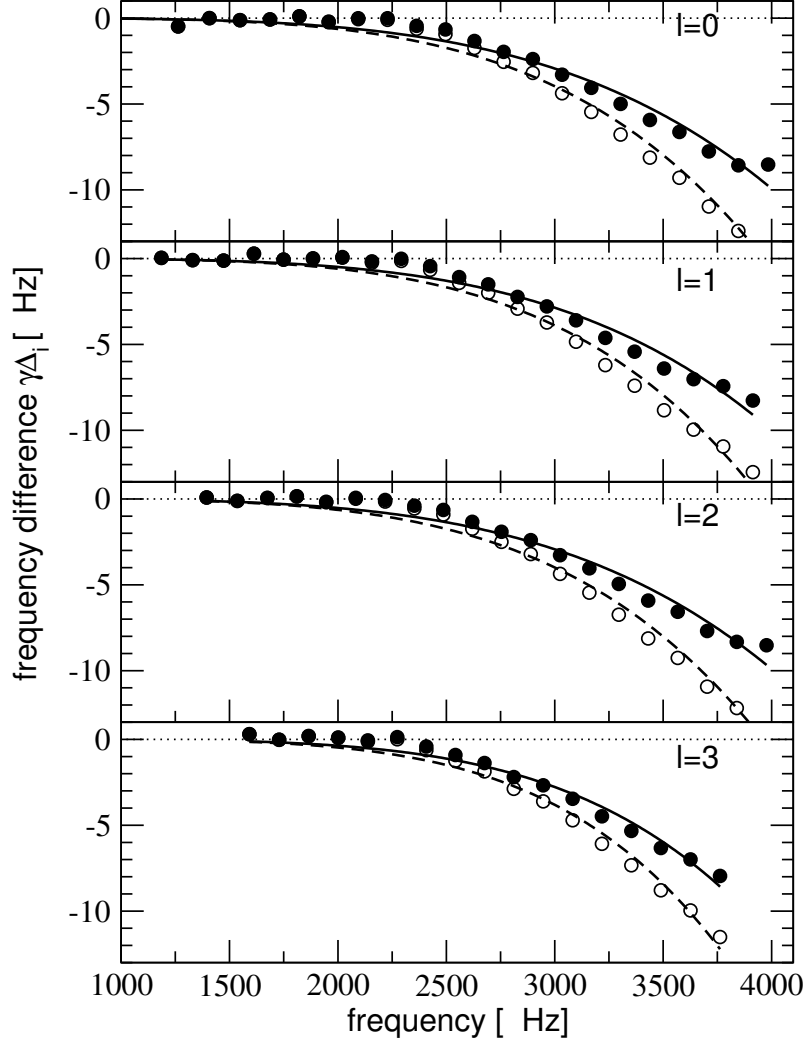
Figure 3.5: Measured surface effects for non-adiabatic (filled circles) and adiabatic (open circles) frequencies of the most probable solar model from evaluating the BiSON frequencies using approach **M2c**. The uncertainties of the differences are smaller than the symbols. Least-squares power-law fits (see Equation (3.31)) to the surface effects for the adiabatic (solid line) and non-adiabatic (dashed line) frequencies are also shown.

3.4.3 Asteroseismic analysis of a Sun-like star

To investigate the applicability of our method to current asteroseismic inves-
tigations, we also performed an "asteroseismic" analysis of a Sun-like star, simulated
by artificially "degrading" the set of observed BiSON frequencies to a precision and
accuracy expected from current space-based missions for average Sun-like solar-type
oscillators. We first multiplied the uncertainties of the BiSON observations by a

factor of 20, and then added corresponding random errors to the frequency values. Furthermore, we did not assume to have detailed prior information on the fundamental parameters. Instead, we fitted the "degraded" data set with a completely flat prior to the same grid as before, again using our surface effect model **M2**.

Although a different most probable model is identified, the overall results are comparable to our findings for **M2a**. They show a slightly larger spread of the model probabilities across the grid. Summarizing the uncertainties for the main parameters by calculating the first and second central moments of the probability distribution in our grid we approximately obtain $M = 1.015 \pm 0.007 \, M_\odot$, age $= 4.76 \pm 0.10 \, \text{Gyr}$, $X_0 = 0.72 \pm 0.01$, $Z_0 = 0.017 \pm 0.001$, $Z_\text{s} = 0.0148 \pm 0.0005$, and $\alpha_\text{ml} = 2.3 \pm 0.1$.

However, these results become worse if we systematically remove lower order modes which are crucial to "anchor" the surface effect relation. To illustrate, we further degraded our data set by only keeping 13 $l = 0$ modes from 1950 to 3580 μHz, 12 $l = 1$ modes from 2020 to 3505 μHz, 10 $l = 2$ modes between 2080 and 3300 μHz, and 8 $l = 3$ modes from 2270 to 3220 μHz. Similar data sets from *Kepler* and *CoRoT* with comparable uncertainties and numbers of modes have recently been analyzed in the literature. The results for the model parameters become $M = 1.046 \pm 0.007 \, M_\odot$, age $= 4.80 \pm 0.43 \, \text{Gyr}$, $X_0 = 0.72 \pm 0.01$, $Z_0 = 0.021 \pm 0.01$, $Z_\text{s} = 0.019 \pm 0.001$, and $\alpha_\text{ml} = 2.3 \pm 0.1$. Although the values are still within $\sim 5\%$ we are almost at the border of our parameter space, and higher-mass models systematically outperform lower-mass models.

We know from investigating the BiSON data using our grid that we require $\alpha_\text{ml} = 2.2$ to fit all solar observables. Therefore, in an analysis of a Sun-like star, we can constrain the fit to all models with this value or use a prior based on the marginal posterior probability for α_ml as determined from the fit to the Sun. In this case we obtain $M = 1.04 \pm 0.01 \, M_\odot$, age $= 4.41 \pm 0.29 \, \text{Gyr}$, $X_0 = 0.72 \pm 0.01$, $Z_0 = 0.020 \pm 0.002$, $Z_\text{s} = 0.018 \pm 0.001$. This is an improvement, but still not comparable to the results obtained when using the full data set.

Thanks to the probabilistic method, however, we can also easily add new observables as further constraints, such as the frequency of maximum power, which can also be inferred from a power spectrum analysis and which approximately scales for Sun-like stars as

$$\nu_{\text{max}} \approx \frac{M/M_\odot \, (T_{\text{eff}}/T_{\text{eff},\odot})^{3.5}}{L/L_\odot} \nu_{\text{max},\odot},$$

(3.33)

with $\nu_{\text{max},\odot} = 3120 \pm 5 \, \mu\text{Hz}$ (Kallinger et al. 2010b). Assuming an observed value of $\nu_{\text{max,obs}}$ and calculating $\nu_{\text{max,mod}}$ for each model according to Equation (5.1) we can then multiply the probability for each model with

$$P(\nu_{\text{max,obs}}|M_j^\Delta, I) = \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left[-\frac{(\nu_{\text{max,obs}} - \nu_{\text{max,mod}})^2}{2\sigma_\nu^2}\right],$$

(3.34)

where $\sigma_\nu = \sqrt{\sigma^2(\nu_{\text{max,obs}}) + \sigma^2(\nu_{\text{max,mod}})}$.

With $\nu_{\text{max,obs}} = 3120 \pm 20 \, \mu\text{Hz}$ for our simulated Sun-like star, we then obtain $M = 1.02 \pm 0.01 \, M_\odot$, age $= 4.39 \pm 0.28 \, \text{Gyr}$, $X_0 = 0.72 \pm 0.01$, $Z_0 = 0.019 \pm 0.002$, $Z_{\text{s}} = 0.017 \pm 0.002$. Finally, if we were able to determine $\nu_{\text{max,obs}}$ to about solar precision, the results would be $M = 1.008 \pm 0.006 \, M_\odot$, age $= 4.39 \pm 0.30 \, \text{Gyr}$, $X_0 = 0.71 \pm 0.01$, $Z_0 = 0.019 \pm 0.002$, $Z_{\text{s}} = 0.017 \pm 0.002$. Therefore, if our observations provide precise additional information such as ν_{max}, it can easily be implemented with our method. It then seems possible to obtain reasonably accurate results for Sun-like stars, even in the absence of low-order modes and without a fixed surface effect correction.

3.5 Conclusions

In this paper, we have derived a new, completely probabilistic framework for asteroseismic grid fitting. We explicitly used marginalization and the formulation of combined propositions to allow for the quantitative evaluation of the model grid physics. While computationally more intensive than the standard χ^2 evaluation, this approach has several benefits in that it

1. allows for the treatment *and analysis* of systematic errors such as *the surface effects*, therefore removing the need to apply corrections prior to fitting,

2. easily implements uncertainties in the mode identification,

3. takes into account the fact that grids are discrete representations of a continuous parameter space, which is especially important for rapidly varying bumped modes,

4. provides a consistent framework to use prior knowledge about stellar fundamental parameters or to evaluate additional observables such as ν_{max}, and

5. produces correctly normalized probabilities and likelihoods, respectively evidences, which can be used to assess the model grid physics and the calibration of the grids.

While the above was explicitly derived using the example of a static grid, the probabilistic approach would also be suited for an adaptive grid approach. The Bayesian evidence could be used as a formidable criterion to decide whether an adaptive grid needs to be further refined or not.

We also showed how to apply our method to study the Sun. The analysis based on our current grid and our prior information matches well the findings of Houdek & Gough (2011), and in general fits the up-to-date picture of the Sun. The age of our best model (measured from the pre-main-sequence birth line) is consistent with the meteoritic solar age. The solar model arrives on the ZAMS approximately $35 \pm 5\,\mathrm{Myr}$ after appearing on the birth line. *We found the same best model whether non-adiabatic or adiabatic frequencies were used.* This shows that our method can adequately deal with different shapes of surface effects, even when using the same (flexible) surface-effect model. One requirement, however, is that there exist enough lower-order modes to "anchor" the fit.

To our knowledge, this work is also the first completely grid-based asteroseismic analysis of the Sun, using *all the information provided by the frequencies* and

prior knowledge about the solar fundamental parameters, that results in the need for initial hydrogen, helium and metal mass fractions more consistent with Asplund et al. (2009) than the traditional higher-metallicity models. At least for our current grid, these values are required to produce a model that "looks" like the Sun, pulsates like the Sun, and has the correct solar age. We stress that a formal χ^2 fit to the Sun's oscillation frequencies (Guenther & Brown 2004) or even targeted nonlinear inversion of the oscillation frequencies (Marchenkov et al. 2000) will not necessarily yield the same model as our approach. With χ^2 fits it is difficult to provide an unbiased correction for surface effects that at the same time does not overly weight the deeper penetrating modes. Some of the deeper penetrating modes are sensitive to the base of the convection zone where the effects of convective overshoot and turbulence, introduced by rotation shears, are not included in the standard models. Inversion methods, where a standard base model is perturbed to fit the oscillations, are also distinct because even though the perturbed model obtained from inversions reveal regions of the standard model that are inadequate, e.g., the base of the convection zone, the inversion model is not an actual standard model in the sense that it is constrained and generated by the model physics.

We know our best-fit model is inaccurate at the surface and we suspect it is inaccurate at the base of the convection zone (the latter suspicion based on the inadequate model physics for this region). Regardless, the model is probabilistically the best model from the current model grid that matches all the known constraints. We speculate that preferring fits that match the oscillation frequencies at the expense of the other physical constraints may be the reason that helioseismologists have been unable reconcile the observed solar p-mode frequencies with frequencies derived from standard models based on the Asplund mixture and metal abundance (Serenelli et al. 2009; Guzik & Mussack 2010). We will pursue these matters in a future study where we include model grids based on the Asplund mixture.

While the purpose of our analysis of the Sun is to test the details of our model physics, our method can also be used in general asteroseismic investigations. When

applying our technique to stars other than the Sun, e.g., recent asteroseismic targets from the *Kepler* mission, tight prior constraints as in the solar case are generally not available. However, the probability formalism can simply assign uninformative (e.g., uniform) priors for the unknown parameters and still retain all the remaining benefits like treatment of missing mode identification and of finite grid resolution.

For current asteroseismology, however, the most important feature is the flexible treatment of the surface effects that differs from the usual approach of employing the empirical correction by Kjeldsen et al. (2008) to the frequencies. Instead of measuring the empirical correction for the Sun with the help of a reference model, we use a flexible probabilistic model that allows us to measure surface effects in any star given our current asteroseismic grids. We do not rely on the validity of the solar surface-effect correction and can test new surface-effect models that deviate from the solar power-law approach. Correctly treating the impact of the surface effects on the model probabilities, this also yields correctly propagated uncertainties, and therefore a less biased (but model-dependent) assessment of the stellar fundamental parameters.

The results presented in the previous section indicate that the accuracy of such current asteroseismic analyses is still an open question and heavily dependent on the number of unaffected, lower-order modes. If there are not enough lower-order modes the surface effect will lead to systematic errors in the fundamental parameter determination. However, even in such a case, by looking at how the evidence changes as better physics are included in the models, our method can be used to iterate toward improved models, hopefully solving the surface-effect problem eventually.

3.6   Acknowledgements

Bibliography

Alexander, D. R., & Ferguson, J. W. 1994, ApJ, 437, 879

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481

Bahcall, J. N., Pinsonneault, M. H., & Basu, S. 2001, ApJ, 555, 990

Bahcall, J. N., Pinsonneault, M. H., & Wasserburg, G. J. 1995, Reviews of Modern Physics, 67, 781

Bazot, M., Bourguignon, S., & Christensen-Dalsgaard, J. 2008, Mem. Soc. Astron. Italiana, 79, 660

Bedding, T. R., & Kjeldsen, H. 2010, Communications in Asteroseismology, 161, 3

Bedding, T. R., et al. 2010, ApJ, 713, 935

Benomar, O., Appourchaux, T., & Baudin, F. 2009, A&A, 506, 15

Böhm-Vitense, E. 1958, ZAp, 46, 108

Bouvier, A., & Wadhwa, M. 2010, Nature Geoscience, 3, 637

Brandão, I. M., et al. 2011, A&A, 527, A37+

Broomhall, A.-M., Chaplin, W. J., Davies, G. R., Elsworth, Y., Fletcher, S. T., Hale, S. J., Miller, B., & New, R. 2009, MNRAS, 396, L100

Chandrasekhar, S. 1957, An introduction to the study of stellar structure., ed. Chandrasekhar, S. (Dover Publications, New York)

Chaplin, W. J., et al. 2010, ApJ, 713, L169

Cohen, E. R., & Taylor, B. N. 1986, Codata Bulletin No. 63, (New York: Pergamon Press)

Demarque, P., Guenther, D. B., Li, L. H., Mazumdar, A., & Straka, C. W. 2008, Ap&SS, 316, 31

Deupree, R. G., & Beslin, W. 2010, ApJ, 721, 1900

Gai, N., Basu, S., Chaplin, W. J., & Elsworth, Y. 2011, ApJ, 730, 63

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support, ed. Gregory, P. C. (Cambridge University Press)

Grevesse, N., & Sauval, A. J. 1998, Space Sci. Rev., 85, 161

Gruberbauer, M., Kallinger, T., Weiss, W. W., & Guenther, D. B. 2009, A&A, 506, 1043

Guenther, D. B. 1994, ApJ, 422, 400

Guenther, D. B., & Brown, K. I. T. 2004, ApJ, 600, 419

Guenther, D. B., Pinsonneault, M. H., & Bahcall, J. N. 1993, ApJ, 418, 469

Guzik, J. A., & Mussack, K. 2010, ApJ, 713, 1108

Handberg, R., & Campante, T. L. 2011, A&A, 527, A56+

Hayashi, C. 1961, PASJ, 13, 450

Hickey, J. R., & Alton, B. M. 1983, in Solar Irradiance Variations of Active Region Time Scales, NASA Conference Publication 2310, ed. B. .J. LaBonte, G. A. Chapman, H. S. Hudson, & R. C. Wilson, 43

Houdek, G., & Gough, D. O. 2011, MNRAS, 1623

Huber, D., et al. 2011, ApJ, 743, 143

Iglesias, C. A., & Rogers, F. J. 1996, ApJ, 464, 943

Jaynes, E. T., & Bretthorst, G. L. 2003, Probability Theory, ed. Jaynes, E. T. & Bretthorst, G. L. (Cambridge University Press)

Jørgensen, B. R., & Lindegren, L. 2005, A&A, 436, 127

Kallinger, T., Gruberbauer, M., Guenther, D. B., Fossati, L., & Weiss, W. W. 2010a, A&A, 510, A106+

Kallinger, T., et al. 2010b, A&A, 522, A1

—. 2010c, A&A, 509, A77+

Kjeldsen, H., Bedding, T. R., & Christensen-Dalsgaard, J. 2008, ApJ, 683, L175

Krishna Swamy, K. S. 1966, ApJ, 145, 174

Lane, J. H. 1869, Amer. J. Sci., 2nd ser., 50, 57

Marchenkov, K., Roxburgh, I., & Vorontsov, S. 2000, MNRAS, 312, 39

Mathur, S., et al. 2010, A&A, 511, A46+

Metcalfe, T. S., et al. 2010, ApJ, 723, 1583

Pont, F., & Eyer, L. 2004, MNRAS, 351, 487

Quirion, P.-O., Christensen-Dalsgaard, J., & Arentoft, T. 2010, ApJ, 725, 2176

Rogers, F. J. 1986, ApJ, 310, 723

Rogers, F. J., Swenson, F. J., & Iglesias, C. A. 1996, ApJ, 456, 902

Roxburgh, I. W. 2005, A&A, 434, 665

Serenelli, A. M., Basu, S., Ferguson, J. W., & Asplund, M. 2009, ApJ, 705, L123

Takeda, G., Ford, E. B., Sills, A., Rasio, F. A., Fischer, D. A., & Valenti, J. A. 2007, ApJS, 168, 297

Chapter 4

Paper II: Bayesian Seismology of the Sun

M. Gruberbauer[1], **D.B. Guenther**[1]

[1]Institute for Computational Astrophysics, Department of Astronomy and Physics, Saint Mary's University, B3H 3C3 Halifax, Canada

Co-author contributions

DG: construction of solar model grids, valuable discussion.

Abstract

We perform a Bayesian grid-based analysis of the solar l=0,1,2 and 3 p modes obtained via BiSON in order to deliver the first Bayesian asteroseismic analysis of the solar composition problem. We do not find decisive evidence to prefer either of the contending chemical compositions, although the revised solar abundances (AGSS09) are more probable in general. We do find indications for systematic problems in standard stellar evolution models, unrelated to the consequences of inadequate modelling of the outer layers on the higher-order modes. The seismic observables are best fit by solar models that are several hundred million years older than the meteoritic age of the Sun. Similarly, meteoritic age calibrated models do not adequately reproduce

the observed seismic observables. Our results suggest that these problems will affect any asteroseismic inference that relies on a calibration to the Sun.

4.1 Introduction

The study of solar-type pulsation with its reliance on scaling relations (e.g., Huber et al. 2011) and calibrations of fundamental free parameters in stellar models (i.e., mixing length parameter and helium abundance) is ultimately anchored by what we know about the Sun and by how well seismology performs at identifying the Sun's key properties. Recent asteroseismic investigations of sun-like pulsators (e.g., Metcalfe et al. 2010; Mathur et al. 2012) are able to give precise model-dependent constraints but it is difficult to assess their accuracy. Inferences from certain asteroseismic observables are not necessarily model dependent as can be verified using spectroscopy or interferometry (e.g., Huber et al. 2012). However, full asteroseismic analyses that determine stellar ages and compositions, or decide among different implementations of how to model important physical processes (e.g., different approaches to convection) rely on a thorough calibration of the properties and parameters of the model.

Several incompatibilities remain between solar modelling and the results inferred from helioseismology that can potentially affect our calibrations (for a recent comprehensive review see Christensen-Dalsgaard 2009). For example, many investigators find that models that use the previous generation (Grevesse & Sauval 1998) abundances fit helioseismic observables better than the current revised solar abundances (Asplund et al. 2005, 2009). Consequently, the helium abundance and the resulting value for the ratio of the metal mass fraction to hydrogen mass fraction at the surface, $(Z_s/X_s)_\odot$, is uncertain. We do know that inadequate modelling of the outer layers leads to the so-called "surface effects" (see, e.g., Kjeldsen et al. 2008; Gruberbauer et al. 2012) that worsens the model fit to higher order frequencies. Uncertainties in opacities, equations of state, nuclear reaction rates, and other global parameters also influence the properties of the solar model and, as a consequence, its seismic calibration. Recently, for instance, an increase in the opacities (Serenelli et al.

2009) and different accretion scenarios (Serenelli et al. 2011) have been identified as possible remedies for the disagreement between the results of helioseismic inversion and models based on the previous and current generation of chemical compositions.

Previous studies testing different model configurations, for example, different chemical mixtures, have often relied on the direct comparison of non-seismic observables and general properties inferred from helioseismology to stellar models calibrated to the non-seismic observables: age, radius, mass, luminosity, and in some cases also surface abundances. More recent approaches (Basu et al. 2007; Chaplin et al. 2007; Serenelli et al. 2009) compared low-degree p modes, or rather various spacings derived from them, to models with solar characteristics. The result again suggests that they cannot be reconciled with the revised solar abundances. Houdek & Gough (2011) also developed an approach that uses quantities derived from the observed modes to infer solar model properties via iterative calibration procedures.

What is missing, though, is a test of the solar model with a tool that takes into account all the information given by the low-degree solar p modes and other constraints and which then results in a quantitative comparison of how much certain model properties are actually preferred on a global, probabilistic level. In our previous paper (Gruberbauer et al. 2012, hereafter Paper I) we introduced a new Bayesian method that uses prior information and properly treats known systematic effects (i.e., "surface effects"). We performed a state-of-the-art, albeit, abbreviated grid-based asteroseismic analysis of the solar model. In this paper we build on and extend our solar modelling by testing various chemical compositions and nuclear reaction rates. Our goal is to answer the following questions:

1. Which models fit the solar modes and other observables the best?

2. Is there a clear preference for any of the chemical compositions and reaction rates?

3. How do the surface effects affect the fit?

4. How do our results affect the calibrations for asteroseismology?

90

We approach our analysis, leaning more toward utilising the techniques applicable to asteroseismology than helioseismology. Specifically we will only utilise the lower l-valued p modes and we will allow all parameters except for the mass to remain unconstrained. We are, therefore, setting out to model the global properties of the Sun as a star, hence, to perform asteroseismology of the Sun.

4.2 Grid-based fitting approach

4.2.1 Observations

As in Paper I, we fit our models to the activity-corrected solar $l = 0$, 1, 2, and 3 p modes obtained by using BiSON data (Broomhall et al. 2009). For our prior probabilities on other solar observables, we take an investigative approach by using both broad and narrow priors for the most important solar quantities: $T_{\rm eff}$, L, and age. This will help us to study the systematic dependencies of our results on the imposed constraints. For the general properties of the Sun, we use both a broad prior with $\log T_{\rm eff} = 3.7617 \pm 0.01$, and $\log (L/L_\odot) = 0.00 \pm 0.01$, or alternatively a more realistic but still conservative prior with $\log T_{\rm eff} = 3.7617 \pm 0.002$, and $\log (L/L_\odot) = 0.00 \pm 0.002$. Here $L_\odot = 3.8515 \pm 0.009 \cdot 10^{33} {\rm erg\, s^{-1}}$ (the average of the ERB-Nimbus and SMM/ARCRIM measurements; Hickey & Alton (1983)). For the solar mass, we use $M_\odot = 1.9891 \pm 0.0004 \cdot 10^{33} {\rm g}$ (Cohen & Taylor 1986). As a reference for the solar age we take the result from Bouvier & Wadhwa (2010) who determined a meteoritic age of the solar system of $\tau \approx 4.5682\, {\rm Gyrs}$. As will be discussed in Section 4.3.3, we construct various uniform priors to allow a range of ages centred on this value. Finally, we also use the helioseismically inferred value of the radius of the base of the convection zone, $R_{\rm BCZ} = 0.713 \pm 0.001\, R_\odot$ (Christensen-Dalsgaard et al. 1991; Basu & Antia 1997). All priors are assumed to be normal distributions.

4.2.2 Model physics

Just as in Paper I, our aim was to employ YREC (Demarque et al. 2008) and produce a set of dense grids covering a wide range in initial hydrogen mass fractions

X_0, initial metal mass fractions Z_0, and mixing length parameters α_{ml}. For this study we kept all model masses constrained to $1\,M_\odot$, but we additionally varied the chemical composition and the nuclear reaction rates.

Our model tracks begin as completely convective Lane-Emden spheres (Lane 1869; Chandrasekhar 1957) and are evolved from the Hayashi track (Hayashi 1961) through the zero-age-main-sequence (ZAMS) to 6 Gyrs with each track consisting of approximately 2500 models. Only models between 4.0 and 6.0 Gyrs are included in the model grid. Constitutive physics include the OPAL98 (Iglesias & Rogers 1996) and Alexander & Ferguson (1994) opacity tables, as well as the Lawrence Livermore 2005 equation of state tables (Rogers 1986; Rogers et al. 1996). Convective energy transport was modelled using the Böhm-Vitense mixing-length theory (Böhm-Vitense 1958). The atmosphere model follows the (T-τ) relation by Krishna Swamy (1966). For each grid, we varied the chemical composition and tested two different nuclear reaction rates. We considered the GS98 mixture (Grevesse & Sauval 1998), the AGS05 mixture (Asplund et al. 2005), and the AGSS09 mixture (Asplund et al. 2009). Nuclear reaction cross-sections were taken from Bahcall et al. (2001) and the nuclear reaction rates from Table 21 in Bahcall & Ulrich (1988). In addition, we also calculated grids using the NACRE rates (Angulo et al. 1999). The effects of helium and heavy element diffusion (Bahcall et al. 1995) were included. Note that our atmosphere models and diffusion effects have been shown to require a larger value of mixing length parameter ($\alpha_{\mathrm{ml}} \approx 2.0 - 2.2$) than standard Eddington atmospheres ($\alpha_{\mathrm{ml}} \approx 1.7 - 1.8$) (Guenther et al. 1993). The model grid spans: X_0 from 0.68 to 0.74 in steps of 0.01, Z_0 from 0.014 to 0.026 in steps of 0.001, and α_{ml} from 1.3 to 2.5 in steps of 0.1.

The pulsation spectra were computed using the stellar pulsation code of Guenther (1994), which solves the linearized, non-radial, non-adiabatic pulsation equations using the Henyey relaxation method. The non-adiabatic solutions include radiative energy gains and losses but do not include the effects of convection. We estimate

the random 1σ uncertainties of our model frequencies to be of the order of $0.1\,\mu$Hz. These uncertainties are properly propagated into all further calculations.

4.2.3 Fitting method

Our Bayesian fitting method is explained in detail in Paper I. To briefly summarize, we compare theoretical and observed frequencies by calculating the likelihood that the two values agree were it not for the presence of random and systematic errors. These likelihoods are then combined using the sum rule and product rules of probability theory, and weighted by priors to arrive at correctly normalised probabilities. The random errors are assumed to be independent and Gaussian. Although frequency uncertainties are likely to be somewhat correlated depending on the data set quality and extraction technique, independence is a fundamental necessity to allow the independent treatment of surface effects. In the solar case the observational uncertainties are rather small, and so random errors in the model frequencies due to the model shell resolution ($\sim 0.1\,\mu$Hz), the influence of priors, and the surface effect treatment will outweigh the influence of correlations[1]. The systematic errors in the case of solar-like stars are assumed to be similar to "surface effects". At higher orders, observed frequencies are systematically lower than model frequencies, and the absolute frequency differences increase with frequency. This is modelled by introducing a systematic difference parameter, Δ, between observed and calculated frequency so that

$$f_{\mathrm{obs,i}} = f_{\mathrm{calc,i}} + \gamma\Delta_i. \tag{4.1}$$

In the case of surface effects, $\gamma = -1$. These Δ_i are then allowed to become larger at higher frequencies. The upper limit at each frequency is determined by the large frequency separation and a power law similar to the standard correction introduced by Kjeldsen et al. (2008). The Δ parameter is incorporated in a completely

---

[1]Moreover, if frequency errors are derived from their marginal distribution as in Bayesian peakbagging (e.g., Gruberbauer et al. 2009; Handberg & Campante 2011), they can be treated as independent.

Bayesian fashion, using a β prior to prefer smaller values over larger ones (see Paper I for more details). In addition, we always allow for the possibility that a mode is not significantly affected by any kind of systematic error. Altogether, this allows us to fully propagate uncertainties originating from the surface effects into all our results, and at the same time gives us more flexibility than the standard surface-effect correction.

We obtain probabilities for every evolutionary track in our grids, and within the tracks also for every model. We also obtain the correctly propagated distributions for systematic errors so that the model-dependent surface effect can be measured. In order to fully resolve the changes in stellar parameters and details in the stellar-model mode spectra, we oversample the evolutionary tracks via linear interpolation until the (normalised) probabilities no longer change significantly. Eventually, we obtain so-called evidence values, equivalent to the prior-weighted average likelihood, for every grid as a whole. These evidence values are also identical to the likelihood of the data (i.e., the solar frequency values) given the particular grids as conditional hypotheses. Just as the likelihoods for individual stellar models or, one step further, for evolutionary tracks can be used to compare their probabilities and evaluate the stellar parameters, the evidence values as likelihoods for whole grids can therefore be employed to perform a quantitative comparison between different input physics used in the grids[2]. This exemplifies the hierarchical structure of Bayesian analysis, which is discussed in more detail in Paper I and also in the more general literature (e.g., Gregory 2005).

4.2.4 Analysis procedure

The advantage of the Bayesian analysis method from Paper I is that many different approaches to fitting the same data set can be compared using the evidence values. Our goal is to see if there is a strong preference for either the GS98, AGS05, or AGSS09 mixture. We also want to test whether or not the NACRE nuclear reaction

---

[2]Other hypothesis modifications (e.g, different shapes of systematic errors) can in principle also be compared.

rates are an improvement. The corresponding grids will be designated as GS98N, AGS05N, and AGSS09N[3]. We use priors for the HRD position and age, as well as R_{BCZ}, to see which of these grids are more consistent with well-known solar properties other than the frequencies. Also, by turning off the priors we can tell which solar-mass models best reproduce the frequencies irrespective of their fundamental parameters.

We will start our analysis without any priors and successively increase the prior information we use, to answer the questions outlined in Section 5.1. For example, if we were to find that the best solar models are much too old and luminous, or if the evidence values decrease when the priors are turned on, we will then have evidence that the model physics cannot reproduce an accurately calibrated solar model.

It should be noted that all results presented in the following sections are highly dependent on the models used (i.e., what was described in Section 4.2.2). We therefore cannot claim that our results represent the real Sun, as indeed we perform our analysis to investigate the similarities and systematic differences between models and observations. However, as explained in Paper I, our approach is able to compare different grids produced from different codes and thus draw probabilistic inferences about systematic differences between these codes as well.

4.3   Results

4.3.1   No priors

For the first test we did not use any of the luminosity, temperature, age, or R_{BCZ} constraints as formal priors. In this case the only effective prior is provided by the selection of the model grid parameters and the restriction to solar-mass models, hence, every model in the grids was given equal prior weight.

Fig. 4.1 compares the grids in terms of the logarithm of the evidence. Note that differences between these values are equivalent to the logarithm of the posterior

[3]Statements that are valid for both reaction rates will refer to both grids at once using the notation GS98(N), AGS05(N), or AGSS09(N)
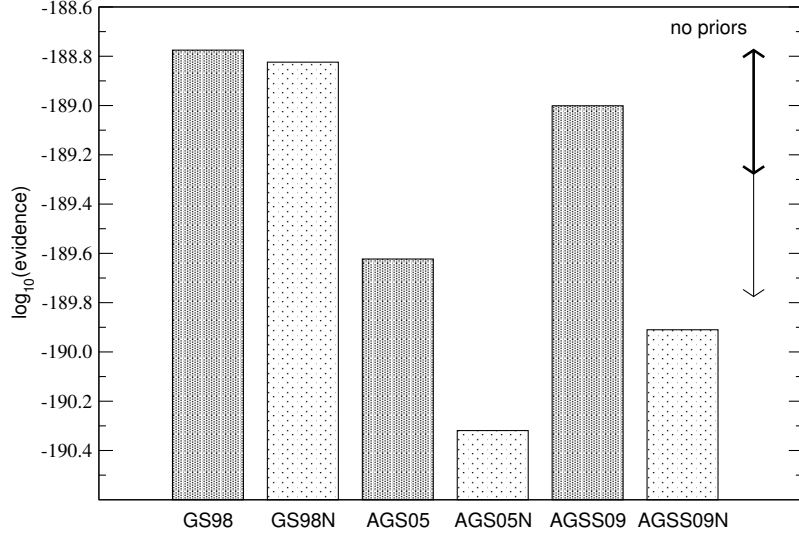
Figure 4.1: Model grid performance without HRD, age or R_{BCZ} priors. The thick double-sided and thin arrows indicate strength of evidence that is "barely worth mentioning" and "substantial" respectively. Differences larger than the thin arrow can be considered "strong" evidence (see text).

probability ratios for the grids as a whole under the condition that they have equal prior probabilities.

Following the guidelines provided by Jeffreys (1961), differences of up to 0.5 (or likelihood ratios up to 3) are considered "barely worth mentioning". Differences between 0.5 and 1.0 indicate "substantial" strength of evidence. Only when the differences rise above 1.0 (i.e., likelihood ratios > 10) should the strength of evidence be considered strong. Accordingly, the GS98-mixture models are not significantly better than AGSS09-mixture models. However, there is substantial evidence that the AGS05, AGS05N, and AGSS09N models do not reproduce the solar frequencies adequately, i.e., the GS98, GS98N, and AGSS09 are significantly better than the AGS05, AGS05N, and AGSS09N models. This indicates that there are problems with the AGS05 mixture and it also suggests that the NACRE rates have a detrimental effect on the model frequencies. Inspection of the frequencies for AGSS09N and AGSS05N reveals that, compared to the corresponding models in the AGSS09 and AGSS05 grids, the lower order modes do not fit as well and the surface effect also increases[4].

---

[4]As will be shown in Section 4.4.3, the former is usually more important than the latter.
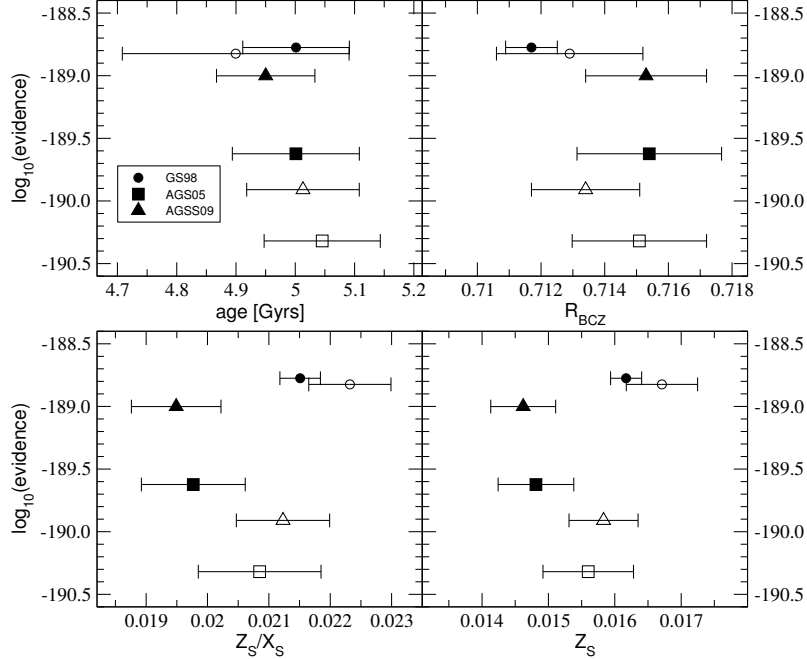
Figure 4.2: Grid evidence versus mean values and uncertainties of some model properties when fitting the observed frequencies without any priors. Open symbols denote the corresponding NACRE grids.

For instance, when just considering the best evolutionary track for AGSS09, adopting the NACRE rates for the same track leads to decrease in probability by a factor of ~ 125. The NACRE models are also older by ~ 16 Myrs and there is a significant increase in R_{BCZ} from 0.7164 to 0.7182.

In Fig. 4.2 we show the mean values and uncertainties of some model properties, corresponding to the grids in Fig. 4.1. Note that these uncertainties are caused by spreading the probabilities over a few different evolutionary tracks with models that fit the frequencies best. If we were to restrict the parameter space by using priors as described in the next sections, then the probabilities will be mostly concentrated on only one or two evolutionary tracks and, consequently, the formal uncertainties will be reduced.

Table 4.1 contains more details for the most probable model parameters of the best and second-best evolutionary tracks in all grids. Considering the metallicities and the locations of the base of the convection zone, the results are similar to the

Table 4.1: Most probable parameters without priors. The quoted probabilities refer to the probability of the evolutionary track within each grid. X_0, Z_0: initial hydrogen and metal mass fractions; Z_s: metal mass fraction in the envelope; R_BCZ: fractional radius of the base of the convection zone; α_ml: mixing length parameter.

grid	$T_\mathrm{eff}[K]$	$L/L_\odot$	$R/R_\odot$	Age	X_0	Z_0	Z_s	Z_s/X_s	R_BCZ	α_ml	Probability
GS98	5718	0.958	1.0001	5.022	0.72	0.018	0.0161	0.0214	0.7116	2.1	0.89
	5802	1.016	0.9998	4.656	0.71	0.018	0.0162	0.0218	0.7139	2.2	0.05
GS98N	5660	0.920	1.0000	5.046	0.72	0.019	0.0170	0.0226	0.7114	2.0	0.52
	5816	1.025	0.9997	4.637	0.71	0.018	0.0161	0.0217	0.7160	2.2	0.34
AGS05	5711	0.953	0.9997	4.967	0.72	0.016	0.0143	0.0190	0.7173	2.1	0.51
	5754	0.983	1.0000	4.975	0.71	0.017	0.0152	0.0204	0.7139	2.2	0.38
AGS05N	5694	0.942	1.0000	5.041	0.71	0.018	0.0161	0.0216	0.7139	2.1	0.50
	5647	0.911	0.9997	5.029	0.72	0.017	0.0152	0.0202	0.7165	2.0	0.26
AGSS09	5718	0.958	0.9998	4.932	0.72	0.016	0.0143	0.0190	0.7164	2.1	0.70
	5761	0.988	1.0000	4.941	0.71	0.017	0.0152	0.0205	0.7132	2.2	0.26
AGSS09N	5701	0.947	1.0001	5.006	0.71	0.018	0.0161	0.0216	0.7128	2.1	0.67
	5654	0.916	0.9998	4.993	0.72	0.017	0.0152	0.0202	0.7155	2.0	0.11

general picture that has emerged in the literature. The GS98 and GS98N models requires higher metallicities and a deeper base of the convection zone. Concerning the latter, the uncertainties are such that both AGSS09 and GS98(N) are in general agreement with R_BCZ. None the less, the GS98(N) models fit this value a little bit better. Using the R_BCZ prior in the next sections will put a formal constraint on this as well.

It is disturbing, however, to see that all of the best models greatly overestimate the age of the Sun by several hundred million years. Furthermore, most of the models do not match the solar T_eff and luminosity very well. Therefore our next step is to "switch on" either the broad or the more realistic priors constraining the Sun's position in the HR diagram.

4.3.2 T_eff and L priors

As in Paper I we now use normal distributions as priors for $\log T_\mathrm{eff}$ and $\log (L/L_\odot)$ (hereafter: HRD prior). More weight is put on models that match the solar position in the HRD. Note that this does not mean that the best models will match the solar values. In this paper we employ slightly different HRD priors, using either a broad prior or a more realistic narrow prior based on current observational
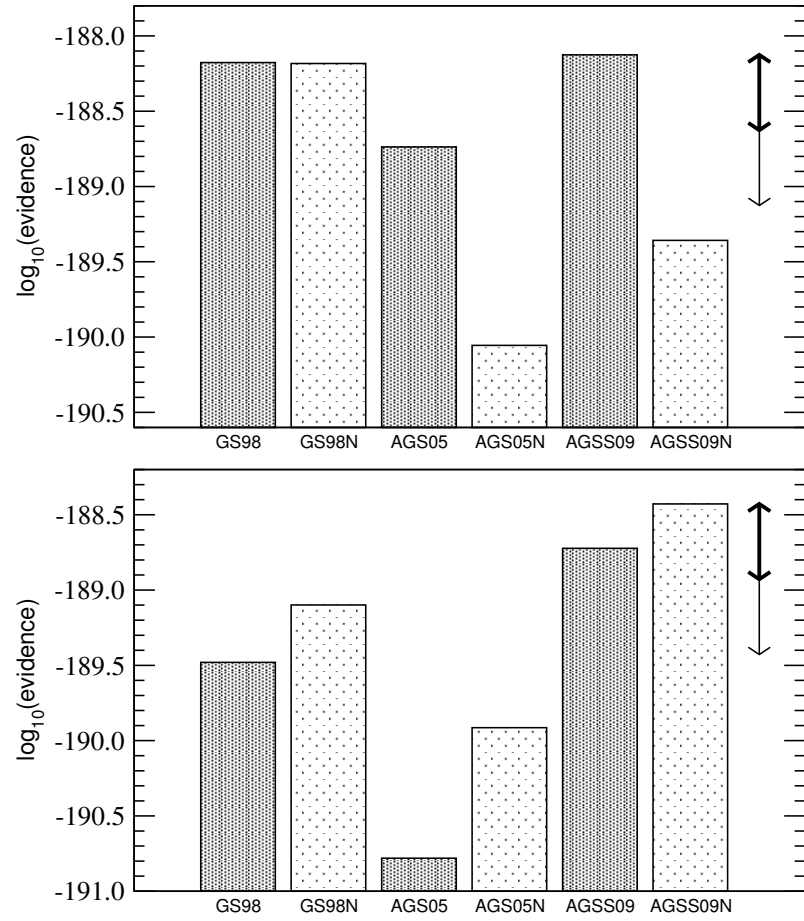
Figure 4.3: Model grid performance with the broad (top panel) and the realistic (bottom panel) HRD prior.

uncertainties. As we show below, the differences between the more realistic prior and the broad prior enable us to distinguish the chemical compositions. The resulting grid evidences are shown in the two panels of Fig. 4.3.

For the broad HRD prior, an increase in evidence for all grids can be seen. This indicates that the models that are somewhat consistent with the solar values do include the majority of the best fit models. Since the evidence is a weighted average of the likelihood, however, most of the increase in evidence is caused by putting less weight on the many models that are clearly outside the solar values and do not match the solar frequencies at all. The relative likelihood ratios remain comparable to the "no prior" case, but now AGSS09 is actually slightly more probable than the GS98(N) models. As before, the evidences of the three best grids are not different enough to clearly prefer one grid over the other. Table 4.2 again gives information on the best fitting evolutionary tracks within each grid for the broad HRD prior. About half of the best or second-best models from the "no prior" analysis remain among the most probable models but only GS98 shows the same models and ranking as before. It is interesting that the best-fitting model from the AGSS09 grid, which also is the overall best fit using the broad HRD prior, now matches the observed base of the convection zone closest from all models considered. Except for GS98N, the NACRE grids again perform worse than their counterparts. Note, however, that with the broad HRD prior the most probable basic model parameters are the same whether or not NACRE rates are used.

For the realistic HRD prior, on the other hand, the GS98(N) grids receive an evidence penalty. Here, the preference for AGSS09(N) is more pronounced, and the previous decrease in evidence for AGSS09N is now compensated by its much closer match to the solar HRD position. As is shown in Table 4.3, the most probable models for AGS05(N) and AGSS09(N) remain the same. However, the best AGS05 model underestimates luminosity and effective temperature and therefore its evidence decreases compared to the broad HRD prior.

100

Table 4.2: Same as Table 4.1 but with the broad HRD priors.

grid	$T_{\text{eff}}[K]$	$L/L_\odot$	$R/R_\odot$	Age	X_0	Z_0	Z_{s}	Z_s/X_s	R_{BCZ}	α_{ml}	Probability
GS98	5718	0.958	1.0001	5.022	0.72	0.018	0.0161	0.0214	0.7116	2.1	0.77
	5802	1.016	0.9998	4.656	0.71	0.018	0.0162	0.0218	0.7139	2.2	0.20
GS98N	5816	1.025	0.9997	4.637	0.71	0.018	0.0161	0.0217	0.7160	2.2	0.85
	5732	0.967	1.0000	5.002	0.72	0.018	0.0161	0.0214	0.7127	2.1	0.10
AGS05	5754	0.983	1.0000	4.975	0.71	0.017	0.0152	0.0204	0.7139	2.2	0.84
	5711	0.953	0.9997	4.967	0.72	0.016	0.0143	0.0190	0.7173	2.1	0.15
AGS05N	5768	0.992	0.9999	4.957	0.71	0.017	0.0152	0.0204	0.7161	2.2	0.46
	5725	0.962	0.9996	4.951	0.72	0.016	0.0143	0.0189	0.7189	2.1	0.36
AGSS09	5761	0.988	1.0000	4.941	0.71	0.017	0.0152	0.0205	0.7132	2.2	0.66
	5718	0.958	0.9998	4.932	0.72	0.016	0.0143	0.0190	0.7164	2.1	0.33
AGSS09N	5775	0.997	1.0000	4.923	0.71	0.017	0.0152	0.0204	0.7149	2.2	0.69
	5701	0.947	1.0001	5.006	0.71	0.018	0.0161	0.0216	0.7128	2.1	0.23

Table 4.3: Same as Table 4.1 but with the realistic HRD priors.

grid	$T_{\text{eff}}[K]$	$L/L_\odot$	$R/R_\odot$	Age	X_0	Z_0	Z_{s}	Z_s/X_s	R_{BCZ}	α_{ml}	Probability
GS98	5767	0.992	1.0002	4.980	0.71	0.019	0.0170	0.0229	0.7096	2.2	0.83
	5802	1.016	0.9998	4.656	0.71	0.018	0.0162	0.0218	0.7139	2.2	0.15
GS98N	5780	1.001	1.0002	4.959	0.71	0.019	0.0170	0.0228	0.7109	2.2	0.997
	5769	0.992	0.9995	4.660	0.72	0.017	0.0152	0.0203	0.7184	2.1	2.6e-3
AGS05	5754	0.983	1.0000	4.975	0.71	0.017	0.0152	0.0204	0.7139	2.2	0.90
	5789	1.006	0.9995	4.848	0.72	0.015	0.0134	0.0178	0.7205	2.2	0.07
AGS05N	5768	0.992	0.9999	4.957	0.71	0.017	0.0152	0.0204	0.7161	2.2	0.99996
	5779	1.000	0.9997	4.680	0.70	0.018	0.0161	0.0220	0.7177	2.2	3.7e-5
AGSS09	5761	0.988	1.0000	4.941	0.71	0.017	0.0152	0.0205	0.7132	2.2	0.9996
	5796	1.011	0.9996	4.814	0.72	0.015	0.0134	0.0178	0.7191	2.2	3.0e-4
AGSS09N	5775	0.997	1.0000	4.923	0.71	0.017	0.0152	0.0204	0.7149	2.2	0.999998
	5787	1.005	0.9998	4.646	0.70	0.018	0.0161	0.0220	0.7158	2.2	1.0e-6

Table 4.4: Same as Table 4.1 but with the R_{BCZ} and realistic HRD priors.

grid	$T_{\mathrm{eff}}[K]$	$L/L_\odot$	$R/R_\odot$	Age	X_0	Z_0	Z_s	Z_s/X_s	R_{BCZ}	α_{ml}	Probability
GS98	5802	1.016	0.9998	4.656	0.71	0.018	0.0162	0.0218	0.7139	2.2	0.83
	5789	1.007	1.0000	4.941	0.72	0.017	0.0152	0.0202	0.7130	2.2	0.14
GS98N	5780	1.001	1.0002	4.959	0.71	0.019	0.0170	0.0228	0.7109	2.2	0.99998
	5746	0.977	0.9998	4.694	0.71	0.019	0.0171	0.0230	0.7142	2.1	6.2e-6
AGS05	5754	0.983	1.0000	4.975	0.71	0.017	0.0152	0.0204	0.7139	2.2	0.99996
	5798	1.014	1.0001	4.947	0.70	0.018	0.0161	0.0219	0.7119	2.3	1.8e-5
AGS05N	5768	0.992	0.9999	4.957	0.71	0.017	0.0152	0.0204	0.7161	2.2	0.9999997
	5779	1.000	0.9997	4.680	0.70	0.018	0.0161	0.0220	0.7177	2.2	1.2e-7
AGSS09	5761	0.988	1.0000	4.941	0.71	0.017	0.0152	0.0205	0.7132	2.2	0.9999997
	5773	0.996	0.9998	4.664	0.70	0.018	0.0162	0.0221	0.7139	2.2	2.6e-7
AGSS09N	5775	0.997	1.0000	4.923	0.71	0.017	0.0152	0.0204	0.7149	2.2	0.9999998
	5787	1.005	0.9998	4.646	0.70	0.018	0.0161	0.0220	0.7158	2.2	1.6e-7

All the conclusions drawn from the "no prior" approach still apply, i.e., the model fits give us no clear indication for, e.g., preferring GS98(N) over AGSS09, but they do show significant evidence against AGS05 and for the detrimental effect of the NACRE rates.

Lastly, we turn on the R_{BCZ} prior in tandem with the HRD priors, which puts stronger constraints on a proper fit to the interior. The results are shown in Fig. 4.4 and the corresponding model parameters for the realistic HRD prior are summarized in Table 4.4. Interestingly, for both HRD priors, AGSS09 manages to increase the probability contrast to the other models. The evidence rises once more, which signifies that the models that fit the pulsation frequencies also are among those that fit best to R_{BCZ}. This is also confirmed by Table 4.4 which shows that the most probable models for AGSS09(N) and AGS05(N) have not changed. *For these mixtures the models that are best at reproducing the pulsation and broad HRD constraints also fit the realistic HRD constraints and the base of the convection zone.* This is also responsible for producing the enormous concentration of probability on the best evolutionary tracks. The bottom panel in Fig. 4.4 also indicates that, with the realistic HRD prior and the R_{BCZ} constraint, there is formally strong evidence for the AGSS09 mixture to provide the overall most realistic solar model.

Nonetheless, all ages are still too high compared to the well-established meteoritic age estimate. We cannot consider these models to be properly calibrated to
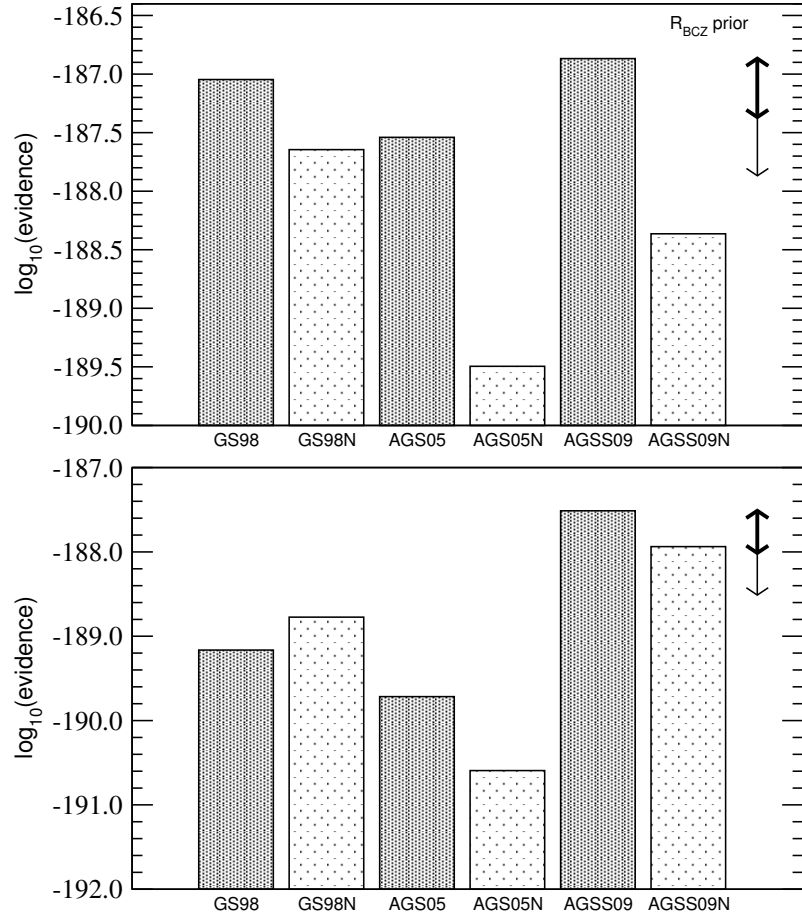
Figure 4.4: Model grid performance with the R_{BCZ} prior, as well as the the broad (top panel) and the realistic (bottom panel) HRD prior.

the Sun, even though the frequencies clearly prefer these solutions. We therefore now turn to age priors to avoid the solutions that are clearly too old (or too young).

4.3.3 HRD and age priors

In Paper I, we used a similar approach to rule out older models and employed a Gaussian prior centred on the meteoritic solar age but allowed for a few tens of millions of years of PMS evolution. In this paper, however, we chose to take a more careful approach.

Different authors often use different definitions for the age of their solar model (e.g., age from the birthline or age from the ZAMS). Therefore, Fig. 4.5 presents the age-related details in our solar model evolution. The meteoritic age is measured from the time when the initial abundance of the isotopes used to date the meteorites are no longer kept in equilibrium. This probably occurs at some point on the Hayashi track. We take the zero age of our models to coincide with the birthline as defined in Palla & Stahler (1999). This introduces an uncertainty of $\sim 7\,$Myrs between the meteoritic age and the birthline age, which is still smaller than the systematic errors in our model ages, which we estimate are of the order of a few tens of Myrs. Note, for example, that switching to the NACRE rates leads to a change in age of about $20\,$Myrs.

In order to avoid putting too much weight on slight differences in the age, and to allow for systematic errors in the meteoritic age determination of perhaps a few Myrs, we will only use uniform age priors centred on the meteoritic age. The purpose of the age prior is therefore only to provide a cut-off for model ages above or below certain limits. We chose two different age priors, one more restrictive than the other, and we continue to use the HRD and R_{BCZ} priors.

4.3.3.1 Broad age prior

The broad age prior is a uniform prior that rules out very old or young models. We designed it to allow for an age range of 4.4 – 4.7 Gyrs. This removes most of
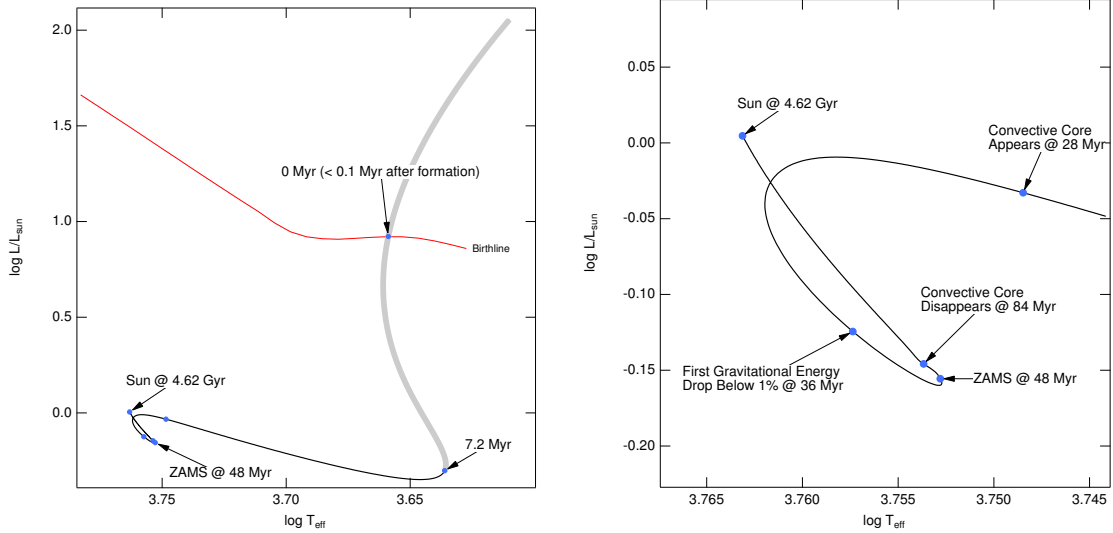
Figure 4.5: Evolution of a one solar mass model. Evolution is started above the birthline. The model crosses the birthline after < 0.1 Myrs, at which point, the age of the model is reset to zero. The thick grey line indicates the period in which the primordial meteoritic material will cool and lock in the initial isotopic abundances used to date meteorites.

our previous best fits, but retains the good GS98(N) models which have ≈ 4.65 Gyrs. Fig. 4.6 shows the results in terms of evidence. Clearly, the AGS05 and AGSS09 mixture have suffered a severe penalty for their older models are now outside the range allowed by the prior. The GS98 and GS98N models, on the other hand, show an increase in evidence compared to Fig. 4.3 and therefore the evidence contrast has increased markedly. The realistic HRD prior does affect and slightly decrease this contrast, but since the AGS05(N) and AGSS09(N) grids have lost their previous best models to the age prior, the effect is not as pronounced as in Fig. 4.3. In terms of the strength of evidence, this result would amount to decisive evidence for the GS98(N) grids. Since the best models are the same for the broad and the realistic HRD prior, we only list the results for the latter in Table 4.5.

For GS98, the probability is now concentrated in the best model from Table 4.4. Note that both the best GS98 and second best GS98N models have the same fundamental parameters, differing only in their nuclear reaction rates. The AGS05(N)
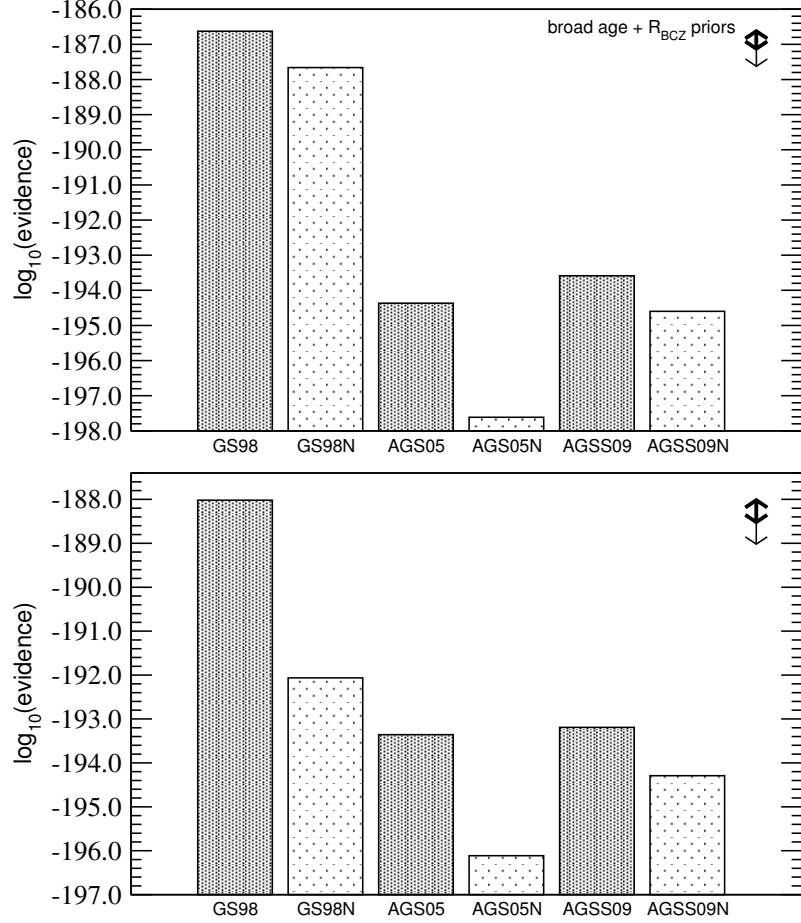
Figure 4.6: Model grid performance with the R_{BCZ} and broad age prior, as well as the broad (top panel) and realistic (bottom panel) HRD prior.

and AGSS09(N) grids all find the same basic model (except for the different mixture and reaction rates) with intermediate $Z_0 = 0.018$ proving to be the most probable.

Without the R_{BCZ} prior (not shown) the best GS98 and GS98N models are the same, and the overall evidence distribution is very similar as in Fig. 4.6. However, the AGS05(N) and AGSS09(N) grids would prefer models with low metallicity ($Z_0 = 0.016$) which produce values of R_{BCZ} well outside the range supported by the inversion results[5]. For all grids, the ages of the best models are still too high by up to 150 Myrs.

---

[5]These models will nonetheless turn out to be the most probable when we make the age constraint even stronger in the next section.

Table 4.5: Same as Table 4.1 but with R_{BCZ}, realistic HRD and broad age priors.

| grid	$T_{\mathrm{eff}}[K]$	$L/L_\odot$	$R/R_\odot$	Age	X_0	Z_0	Z_{s}	Z_s/X_s	R_{BCZ}	α_{ml}	Probability
GS98	5802	1.016	0.9998	4.656	0.71	0.018	0.0162	0.0218	0.7139	2.2	0.9998
	5755	0.983	0.9996	4.678	0.72	0.017	0.0153	0.0203	0.7167	2.1	1.0e-4
GS98N	5746	0.977	0.9998	4.694	0.71	0.019	0.0171	0.0230	0.7142	2.1	0.43
	5816	1.025	0.9997	4.637	0.71	0.018	0.0161	0.0217	0.7160	2.2	0.30
AGS05	5766	0.990	0.9997	4.697	0.70	0.018	0.0162	0.0220	0.7149	2.2	0.9999997
	5795	1.010	0.9994	4.613	0.71	0.016	0.0143	0.0193	0.7204	2.2	2.9e-7
AGS05N	5779	1.000	0.9997	4.680	0.70	0.018	0.0161	0.0220	0.7177	2.2	1.00
	5733	0.968	0.9995	4.687	0.71	0.017	0.0152	0.0205	0.7194	2.1	1.8e-13
AGSS09	5773	0.996	0.9998	4.664	0.70	0.018	0.0162	0.0221	0.7139	2.2	0.999999
	5802	1.015	0.9995	4.580	0.71	0.016	0.0144	0.0193	0.7192	2.2	9.9e-7
AGSS09N	5787	1.005	0.9998	4.646	0.70	0.018	0.0161	0.0220	0.7158	2.2	0.999998
	5767	0.991	0.9999	4.671	0.69	0.020	0.0180	0.0248	0.7121	2.2	1.0e-6

4.3.3.2 Narrow age prior

In order to see how fully age-constrained solar models in the GS98(N) grids compare to the AGS04 and AGSS09 models, we restricted the age even further by employing a narrow uniform age prior that only allows ages of 4.52 – 4.62 Gyrs. As shown in Fig. 4.7, the narrow age prior has a big effect on the analysis. Since it is interesting to see whether models at the correct age can fit the base of the convection zone, we first perform the analysis without the R_{BCZ} prior.

Compared to Fig. 4.6 and for the broad HRD prior, the narrow age constraint strongly decreases the evidence for the GS98(N) models, while increasing the evidence for the other models. GS98 still comes out to be the most probable by an order of magnitude. The remaining grids show more or less comparable evidences, but AGS05N and AGSS09N are worse than their non-NACRE counterparts. All solutions for AGS05 and AGSS09 favour the same basic model parameters. Table 4.6 lists the corresponding most probable models. Ultimately, the narrow age prior has led to models which are very close to the meteoritic solar age without constraining them too strongly (as would be the case for a non-uniform, e.g., Gaussian, age prior) so that we do not rule out completely the possibility of systematic errors in the stellar model age. All of the best models, irrespective of mixture or reaction rates, have $X_0 = 0.71$,
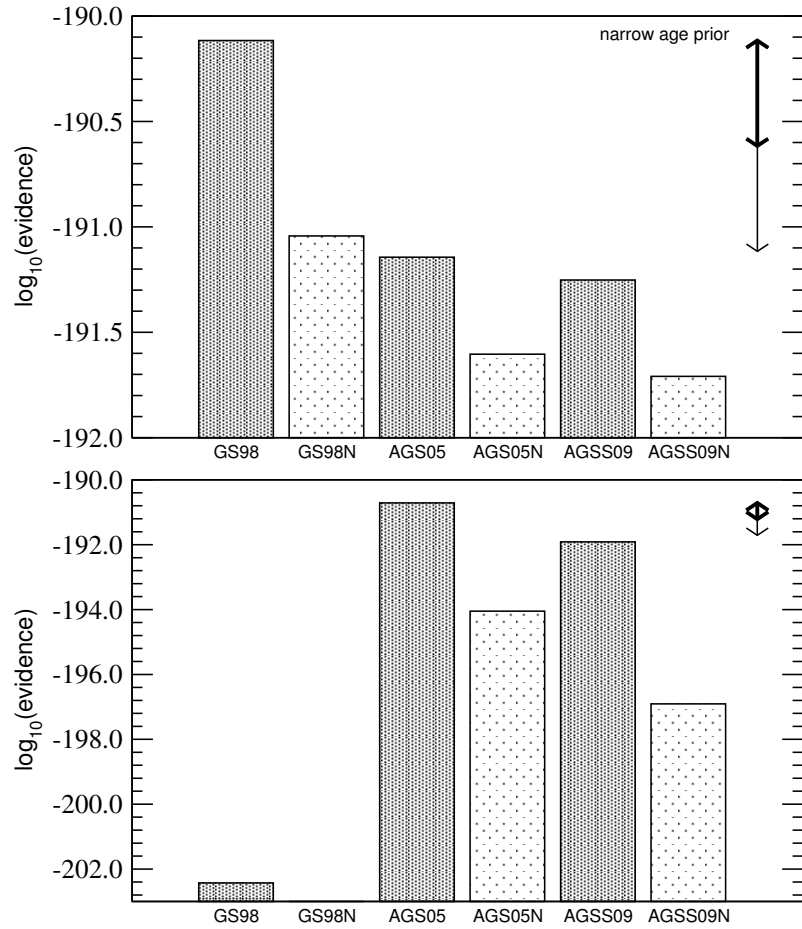
Figure 4.7: Model grid performance with the narrow age prior, as well as the broad (top panel) and realistic (bottom panel) HRD prior.

$Z_0 < 0.018$, and $R_{\mathrm{BCZ}} > 0.716$. Compared to the revised mixtures, GS98(N) has slightly higher metallicities and requires a larger mixing length parameter.

For the realistic HRD prior, however, the situation is completely different. All GS98(N) models drop out of the discussion due to a big decrease in evidence. The effective temperature and luminosity values (see, e.g., the results for GS98N in Table 4.6) are so far outside the prior range that the prior probability terms become close to our numerical threshold values. The only GS98(N) models that still retain what is left of the evidence have $Z_0 < 0.017$. Even though the evidence picture has changed drastically, the actual best models for the AGS05 and AGSS09 grids are still the same as with the broad HRD prior.

Finally, we again turn on the R_{BCZ} prior. The evidence results are depicted in Fig. 4.8. For the broad HRD prior, the evidence present a similar picture as before, but the contrast between GS98(N) and the AGS05(N) and AGSS09(N) models has intensified. Furthermore, the most probable models for GS98N, AGS05(N) and AGSS09(N) now have higher metal mass fraction as before. The overall most probable model of the GS98 grid, which far outweighs the others in terms of evidence, still is the same as in Table 4.6.

For the realistic HRD prior, which most closely reflects all of our prior knowledge of the Sun, the verdict is clear as well. However, for this prior, it is the AGS05 and the AGSS09 models which are preferred. *The evidence contrast between these two grids and the others is the highest contrast measured in all analyses performed in this paper.* The influence of the realistic HRD prior is again substantial and even produces a null result for the GS98N grid because of our numerical thresholds. The parameters for the most probable models are given in Table 4.7. All best models now have $X_0 < 0.73$, $Z_0 = 0.016$, $\alpha_{\mathrm{ml}} = 2.2$, and $R_{\mathrm{BCZ}} > 0.719$, and the models with revised composition agree on $X_0 = 0.71$ as well.

4.3.3.3 Summary

Our detailed analysis using various priors has shown:
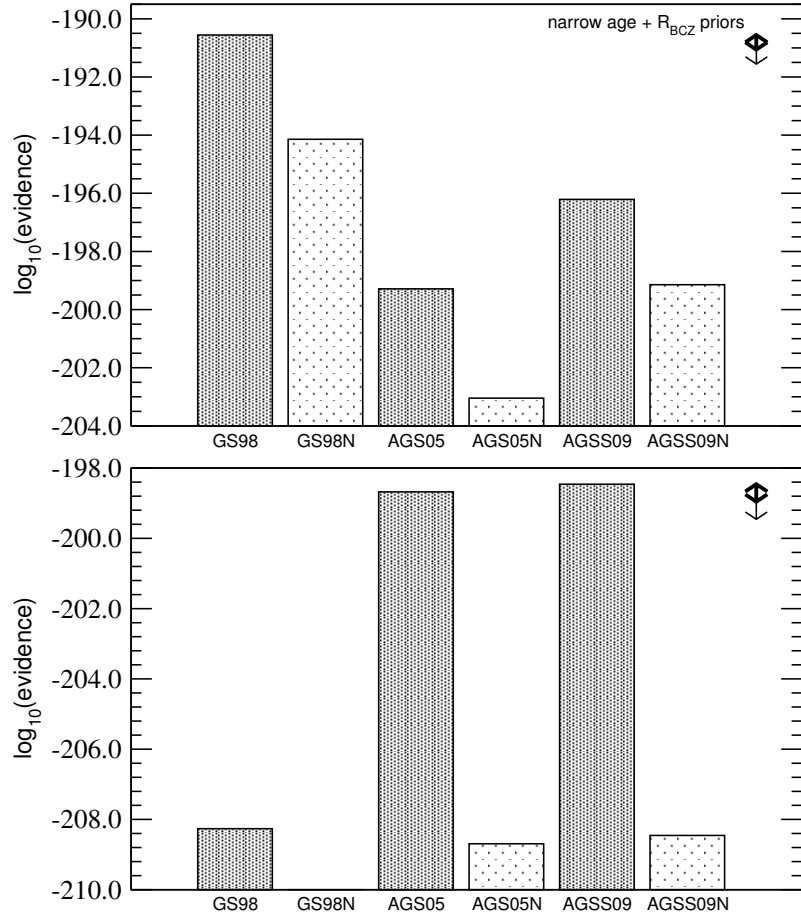
Figure 4.8: Model grid performance with the R_{BCZ} and narrow age priors, as well as the broad (top panel) and realistic (bottom panel) HRD prior.

Table 4.6: Same as Table 4.1 but with broad HRD and narrow age priors.

grid	$T_{\mathrm{eff}}[K]$	$L/L_\odot$	$R/R_\odot$	Age	X_0	Z_0	Z_{s}	Z_s/X_s	R_{BCZ}	α_{ml}	Probability
GS98	5872	1.065	0.9997	4.566	0.71	0.017	0.0153	0.0205	0.7162	2.3	0.98
	5829	1.034	0.9995	4.591	0.72	0.016	0.0144	0.0191	0.7192	2.2	0.02
GS98N	5885	1.075	0.9996	4.547	0.71	0.017	0.0152	0.0205	0.7183	2.3	0.999
	5843	1.043	0.9994	4.573	0.72	0.016	0.0143	0.0190	0.7212	2.2	4.3e-4
AGS05	5795	1.010	0.9994	4.613	0.71	0.016	0.0143	0.0193	0.7204	2.2	0.9999
	5837	1.040	0.9997	4.605	0.70	0.017	0.0153	0.0208	0.7180	2.3	1.6e-5
AGS05N	5809	1.019	0.9994	4.596	0.71	0.016	0.0143	0.0192	0.7220	2.2	0.9999
	5851	1.050	0.9996	4.588	0.70	0.017	0.0152	0.0207	0.7194	2.3	2.5e-5
AGSS09	5802	1.015	0.9995	4.580	0.71	0.016	0.0144	0.0193	0.7192	2.2	0.9999
	5845	1.045	0.9997	4.572	0.70	0.017	0.0153	0.0208	0.7162	2.3	2.5e-5
AGSS09N	5816	1.024	0.9994	4.564	0.71	0.016	0.0143	0.0193	0.7209	2.2	0.9999
	5858	1.055	0.9997	4.555	0.70	0.017	0.0152	0.0207	0.7180	2.3	4.9e-5

Table 4.7: Same as Table 4.1 but with R_{BCZ}, realistic HRD and narrow age priors.

grid	$T_{\mathrm{eff}}[K]$	$L/L_\odot$	$R/R_\odot$	Age	X_0	Z_0	Z_{s}	Z_s/X_s	R_{BCZ}	α_{ml}	Probability
GS98	5829	1.034	0.9995	4.591	0.72	0.016	0.0144	0.0191	0.7192	2.2	1.0
	5788	1.004	0.9991	4.580	0.73	0.015	0.0135	0.0177	0.7224	2.1	1.0e-18
AGS05	5795	1.010	0.9994	4.613	0.71	0.016	0.0143	0.0193	0.7204	2.2	1.0
	5741	0.974	0.9998	4.538	0.68	0.022	0.0199	0.0278	0.7101	2.2	5.1e-70
AGS05N	5809	1.019	0.9994	4.596	0.71	0.016	0.0143	0.0192	0.7220	2.2	1.0
AGSS09	5802	1.015	0.9995	4.580	0.71	0.016	0.0144	0.0193	0.7192	2.2	1.0
	5762	0.986	0.9991	4.547	0.72	0.015	0.0135	0.0179	0.7227	2.1	6.8e-26
AGSS09N	5816	1.024	0.9994	4.564	0.71	0.016	0.0143	0.0193	0.7209	2.2	0.999997
	5834	1.039	0.9999	4.607	0.69	0.019	0.0171	0.0235	0.7134	2.3	2.9e-6

- Without priors, the frequencies fit best to models with significantly underestimated luminosities and ages of about 5 Gyrs. There is no clear preference for any specific composition.

- Models that are constrained by the solar L, T_{eff} and R_{BCZ} prefer the revised composition but are still too old. Except for the age, they can reproduce all known parameters, as well as the frequencies, quite well.

- Models that are tightly constrained by our information on the solar age suffer a strong degradation in their quality of fit. Depending on whether L and T_{eff} are included as tight constraints, there is a either a clear preference for the old or the revised composition. In any case, for solar-age models T_{eff} is overestimated while the stellar radius is slightly underestimated, producing a significantly overestimated luminosity. The model values for R_{BCZ} are too high and well outside the observational uncertainties.

4.4 Discussion

In the following section, we will consider the questions formulated at the beginning of the paper.

4.4.1 The "best fit"

Our first question was "Which models fit the solar modes and other observables the best?". This can be answered by looking at the evidence values for all the grids we tested in the previous section.

Considering only the p-mode frequencies, i.e., no priors, Fig. 4.1 shows that GS98, GS98N, and AGSS09 all perform comparably well, as they are all able to reproduce the observed solar frequencies. Taking into account some of our prior knowledge by using the broad HRD prior, we find a similar result in Fig. 4.3. This also increases the evidence of all models. For a tighter, more realistic HRD prior the evidence for AGSS09(N) is significantly higher than for the other models.

However, we have to reject these results, as the models are clearly too old. By removing the older models via uniform age priors, we first see a large drop in the evidence values for the AGS05(N) and AGSS09(N) grids. Indeed, when we employed the narrow age prior, which is still comparatively broad (100 Myrs) to allow for systematic errors in the model evolution, the GS98(N) grid evidences suffer the same effect. We are forced to conclude that the model frequencies are getting worse as we approach the (presumably) correct solar age. A similar conclusion was reached in Paper I, but here we have shown that this is not affected by the contested different chemical compositions or the two different nuclear reaction rates.The different compositions only produce clearly different results when using additional constraints, as discussed in the next section.

In Fig. 4.9 we have plotted the relative difference between the solar sound speed profile as measured from inversion (Basu et al. 2009) and as determined from some of our best fit models. The Model S from Christensen-Dalsgaard et al. (1996) is plotted as well. This reflects our summary from above, concluding that models constrained to the solar observables are worse at reproducing the observed frequencies and therefore the solar sound speed profile. The figure also shows, contrary to what is commonly reported in the literature, that when using all our prior information the best GS98 model performs worse than the the models with the revised composition. It would
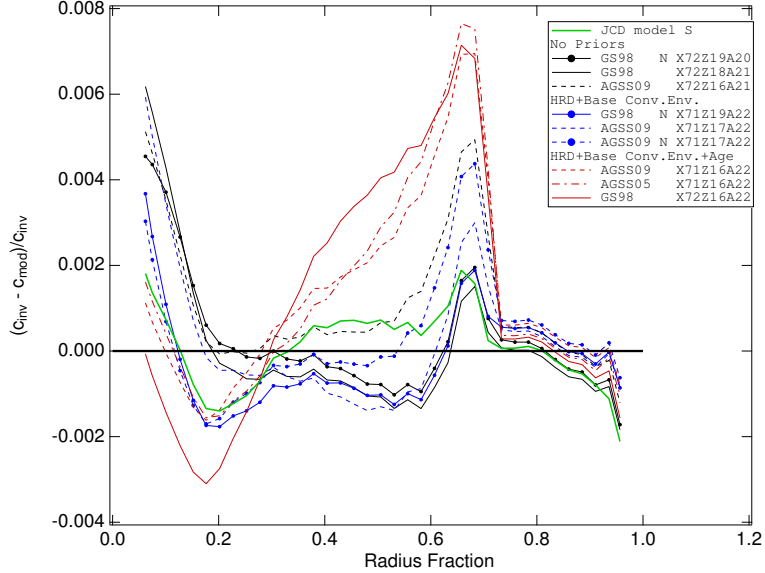
Figure 4.9: Relative difference between the solar sound speed as measured from inversion and determined from our various models. The legend indicates which models and priors were used. Only the realistic HRD and narrow age prior results are plotted. N denotes the NACRE reaction rates. (A colour version of this figure is available in the electronic version of the paper)

be interesting to include the sound speed profile information in the fitting procedure as well, but the systematic differences between observations and calculations are substantial and the analysis would be non-trivial. It is therefore beyond the scope of this paper and should be targeted for future work.

To summarize, the argument for or against the grids presented in this paper cannot be made by simply claiming that one grid produces better frequencies in one particular setup of priors and observables. As we have shown, the grids are able to deliver similar fits in various conditions, and all grids actually have problems to fit both seismic and solar parameters. Therefore, we cannot identify a clear "best fitting model".

4.4.2 Composition

Contrary to most studies in the literature, our results lead us to reject any clear preference for any of the contested chemical compositions over the others. Looking at the frequencies alone, no composition is clearly preferred, but there is very strong

evidence against AGS05. Ignoring the model ages but using our other priors leads to a significant preference for AGSS09(N). Employing the R_{BCZ}, narrow age, and broad HRD prior leads to decisive evidence for GS98. On the other hand, using all priors leads to a clear preference for AGS05 and AGSS09. Hence, which composition better represents the Sun depends on the consideration of tight constraints on R_{BCZ}, T_{eff}, L, and age. This suggests that our models are not calibrated well enough to the Sun, so that prior information is playing an important role compared to the observed frequencies. The latter do have an effect, however, in selecting the models that are compatible with our prior information, and thus the results cannot simply be dismissed.

We have to conclude that without solving the general problem of how to produce solar-age models that look like the Sun and produce adequate frequencies, any discussion of the contending compositions has to remain unresolved. Therefore, we also have to refute the claim that the AGSS09 composition is incompatible with helioseismic results.

We want to exemplify this, and contrast it to arguments used in the past, by looking at some of the solar parameters obtained from the fits. As shown in Table 4.7, the best GS98 model when subject to all our prior knowledge constraints, has $X_0 = 0.72$ and $Z_0 = 0.016$, and therefore $Y_0 = 0.264$. For the helium mass fraction in the envelope we obtain $Y_{\mathrm{s}} = 0.234$, which clearly does not agree with helioseismic inferences ($Y_{\mathrm{s}} = 0.2485 \pm 0.0035$, Basu & Antia (2004)). Furthermore, this GS98 model over-estimates the luminosity and effective temperature. Also the location of the base of the convection zone remains a problem. For the corresponding AGSS09 model, our best model when using all prior constraints, we obtain $X_0 = 0.71$ and $Z_0 = 0.016$, which leads to $Y_0 = 0.274$. In the envelope, this then amounts to $Y_{\mathrm{s}} = 0.243$, which lies almost within the 1σ uncertainties. Judging from the goodness of fit to the frequencies, as well as from the agreement to these helioseismically determined values, we would have to conclude that AGSS09 outperforms GS98. Naturally, this is only true if we ignore the overestimated luminosities, and R_{BCZ} values. Also, as

shown in Fig. 4.9 both models produce some of the strongest deviations from the solar sound speed profile.

In a similar case, when ignoring the age and using just the R_{BCZ} and realistic HRD prior we find $Y_{\mathrm{s}} = 0.2413$ and $R_{\mathrm{BCZ}} = 0.7096$ (GS98), or alternatively $Y_{\mathrm{s}} = 0.2413$ and $R_{\mathrm{BCZ}} = 0.7132$ (AGSS09). Both models now fit the helioseismic inferences quite well, but again AGSS09 is better at reproducing the overall solar parameters and also the base of the convection zone. In this case, however, we have to consider the big problem that the ages are wrong by almost ten percent.

Consequently, we reiterate that these arguments, as well as potentially serendipitous matches of specific model properties are insufficient to solve the composition problem.

4.4.3 Surface effects

As mentioned in the introduction, previous studies mostly looked at frequency differences and spacings, due to the surface effect problem. We have employed the Bayesian formalism to take the surface effects into account while still using the full information provided by each frequency. We now want to analyse how they affect the analysis and to what extent they influence our conclusions.[6] This is possible because, as discussed in Paper I, our method provides the most probable systematic deviations between observed and theoretical frequencies, as well as their uncertainties, for every observed mode.

One explanation for the higher evidences at older ages could be that the older models show smaller surface effects. Such a situation would pose a difficult problem, because we cannot assume that our models are already good enough in the outer layers. The determination of the basic solar parameters would again depend on the surface layers which is not what we want. Fig. 4.10 shows that, fortunately, the opposite is the case. It compares the surface effects as determined from our GS98 models with the "no prior" approach to respective values from the narrow

---

[6]Note that the surface effects are always measured with respect to specific models and using, e.g., adiabatic rather than non-adiabatic frequencies will result in different surface effects.

age prior approach. The mean most probable deviation for the latter amounts to $\langle \gamma\Delta_i \rangle = -2.495$, while the narrow age prior yields $\langle \gamma\Delta_i \rangle = -2.301$. Obviously, the surface effects are larger for the "no prior" mode, yet this model has a much higher evidence. This shows that in the case of large surface effects, the probabilities are, as required, more sensitive to the lower order modes. Fig. 4.11 shows these modes in more detail. In addition, we have plotted linear fits to these deviations to underline that for $l=0$ modes in particular, the "no prior" approach is more consistent with the $\gamma\Delta = 0$ baseline.

In a similar comparison, it is also interesting to probe the differences between the surface effects for the different compositions. Fig. 4.12 and 4.13 show a comparison of the AGSS09 and GS98 results obtained with the R_{BCZ}, realistic HRD, and narrow age prior. The AGSS09 model exhibits a larger surface effect at the lowest orders and therefore gets penalized in the probability terms for these modes. However, it also fits better on average at the lowest $l = 0$, $l = 1$, and $l = 3$ modes and, at the highest orders, has slightly smaller surface effects.

Comparing the systematic differences plotted in Fig. 4.11 and Fig. 4.13 reveals that the most important component of the frequency fit is indeed the overall goodness of fit at the lower order modes. While the "no prior" frequencies do have significantly larger surface effects above $\sim 3000\,\mu$Hz, the systematic differences are significantly smaller between 2000 and 3000 μHz.

In conclusion, our surface effect treatment performs favourably by allowing low-order modes to dominate the fitting process, while still being flexible enough to allow us to properly measure the most probable surface effect at higher orders for every frequency.

4.4.4 Implications for asteroseismology

General properties of Sun-like stars can now be inferred via scaling laws using high-quality frequencies from space missions and other asteroseismic observables (see, e.g., Huber et al. 2012). While the uncertainties of the current solar frequency sets are
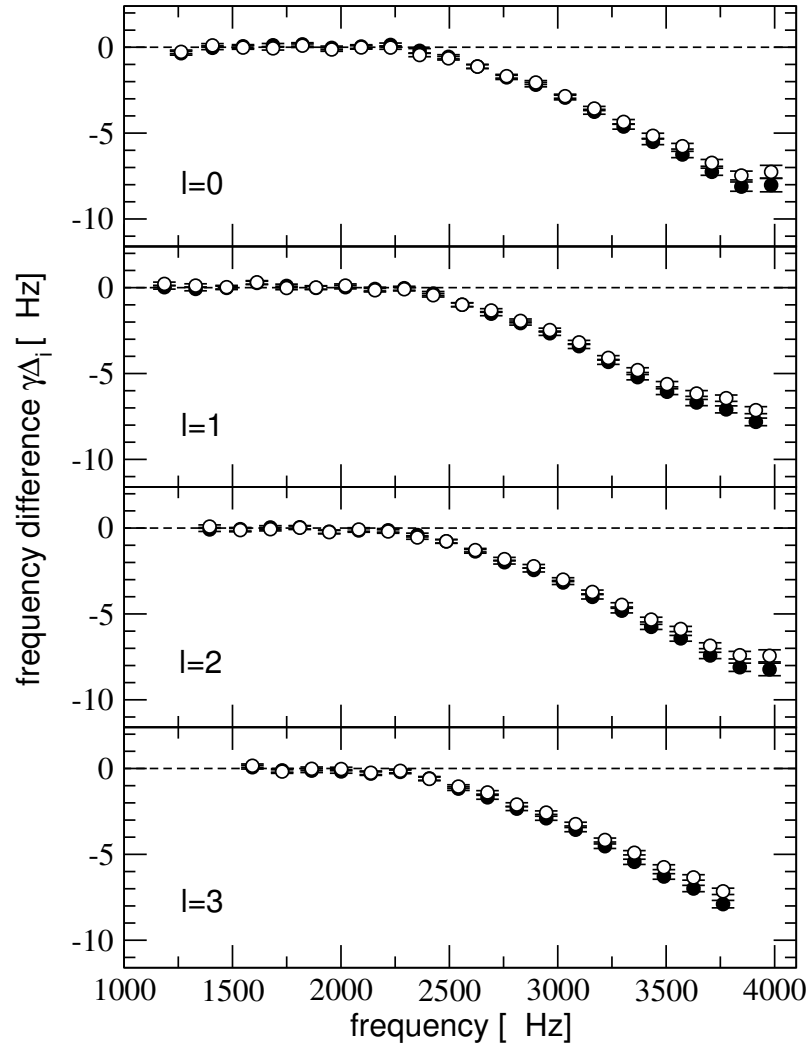
Figure 4.10: Most probable systematic deviations and uncertainties as measured using the GS98 grid. The black circles represent the "no prior" approach, while open circles derive from the broad HRD + narrow age prior. Note that the uncertainties are dominated by the theoretical frequency uncertainties (0.1 μHz) except at the highest orders. The "no prior" approach results in larger surface effects. The dashed guide line shows a frequency difference of zero.

117
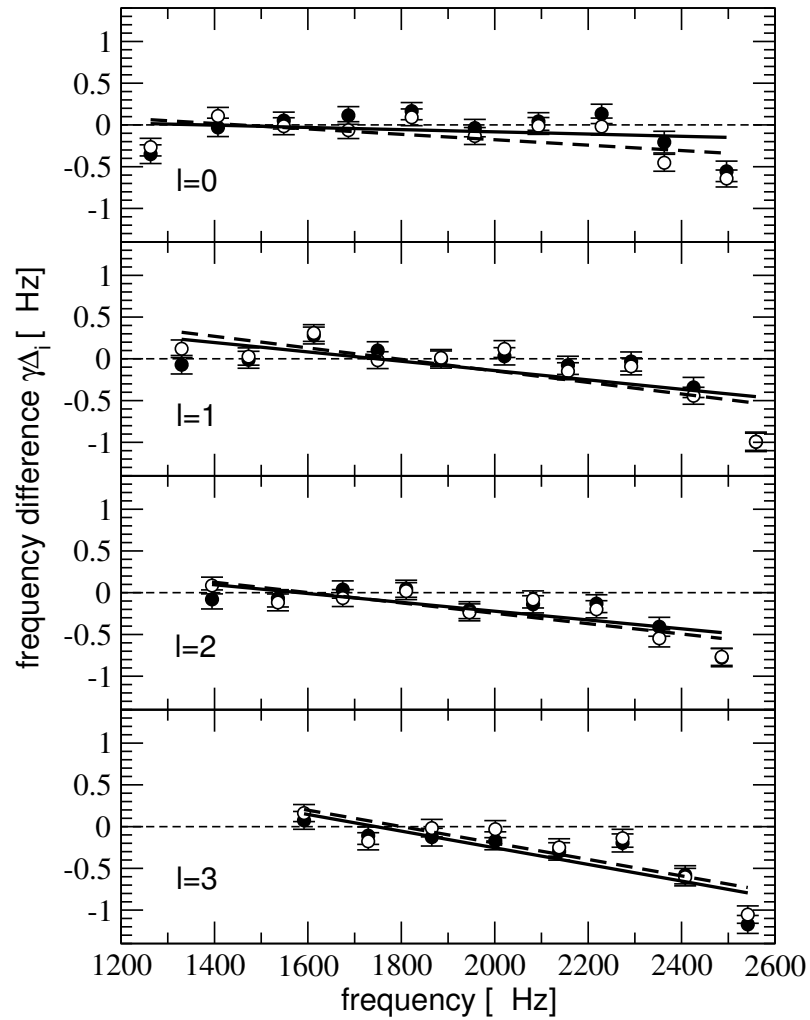
Figure 4.11: Same as Fig. 4.10 but zoomed in on the lower-order modes. In addition, long dashed and solid lines represent linear fits to the open and black circles. Note that the increase in slope towards higher spherical degree is an artifact due to missing lower-order modes.

118
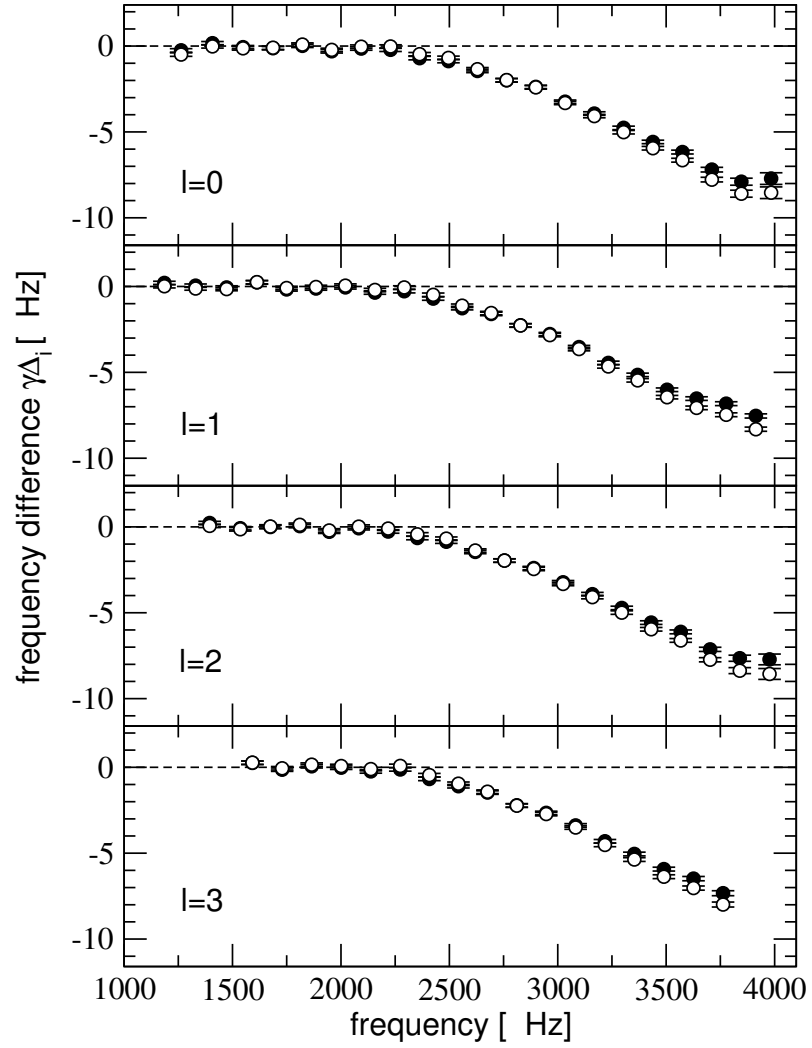
Figure 4.12: Same as Fig. 4.10 but for GS98 (open circles) and AGSS09 (black circles). Both results are based on the realistic HRD + narrow age + R_{BCz} prior analysis.
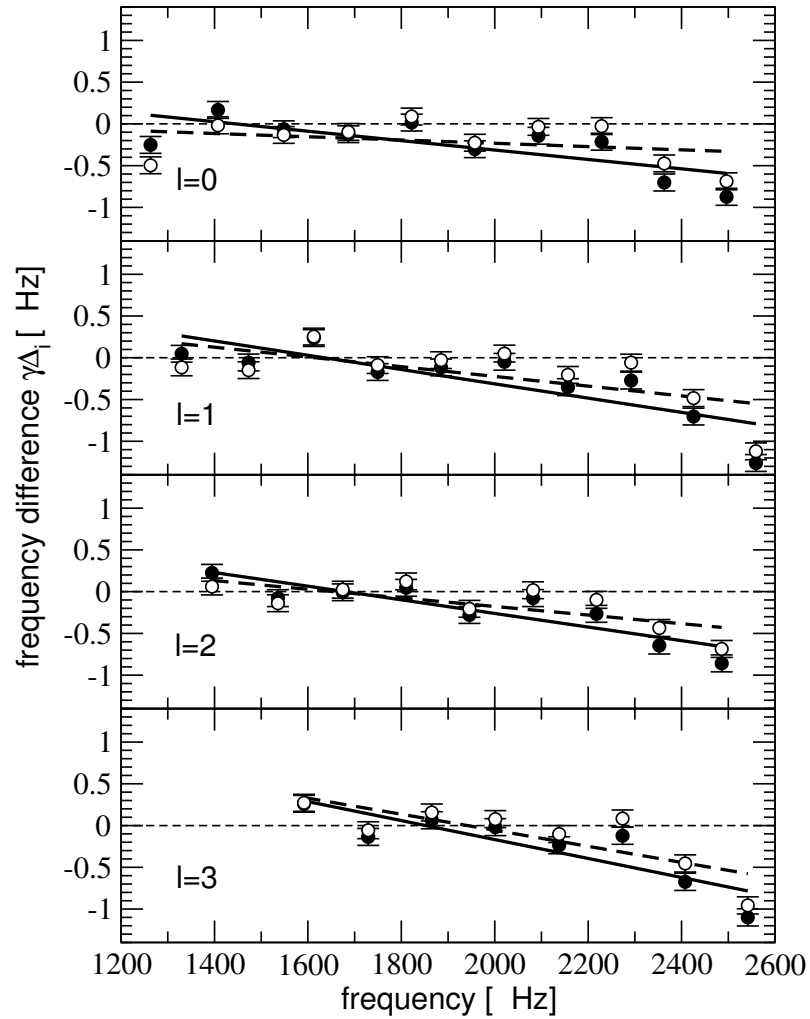
Figure 4.13: Same as Fig. 4.12 but zoomed in on the lower-order modes. In addition, long dashed and solid lines represent linear fits to the open and black circles.

still smaller than those of the best Kepler targets, the asteroseismology community expects to be able to go beyond scaling laws and probe details of the stellar physics (e.g., determining ages and chemical compositions). Our results suggest that in order to obtain accurate results, more work is needed to first understand the properties of the Sun. As we know from meteoritic data, we obtain solar ages that are wrong by hundreds of millions of years unless we restrict the model space. Furthermore, when we perform a full grid-based analysis, we cannot yet properly distinguish between the competing chemical compositions which have an effect on all the involved quantities.

The impact of our analysis also extends beyond the purely asteroseismic applications. For instance, for our best models presented in this paper we obtain values for $(Z_s/X_s)_\odot$ ranging from 0.0190 to 0.0230. If we constrain ourselves to our models at the approximate solar age, we require $(Z_s/X_s)_\odot < 0.0205$, which is quite different from the standard value that is often used to transform between [Fe/H] and Z_s. In addition, uncertainties and systematic errors in the metallicity and helium abundance will naturally propagate into the results of other fields (e.g., the study of Galactic abundances) that rely on the solar calibration.

4.5 Conclusions

In this paper we have reported on our extensive grid-based "asteroseismic" investigation of the Sun using the Bayesian formalism developed in Paper I. We extended our previous study by using different grids with competing chemical compositions (GS98, AGS05 and AGSS09) and nuclear reaction rates. We found that we cannot accurately reproduce the solar properties by fitting the frequencies alone without using prior information. On the other hand, when using prior information, we observe a strong degradation in the goodness-of-fit for the frequencies. This leads us to conclude that we cannot yet give preferential weight to either of the competing chemical compositions (or nuclear reaction rates for that matter) since the evidence values contradict our prior information. In other words, the grids are not properly calibrated and some parts of the fundamental model physics are inappropriate. Our

work does not suggest that the revised compositions are any more incompatible with helioseismology in some systematic way than the traditional GS98 abundances. We have also established that it is not the outer layers which cause the problem, as our Bayesian treatment of surface effects all but removes their impact.

The meteoritic age of the solar system of about 4.568 Gyrs is very well established (even if we allow for a systematic error of perhaps a few Myrs) and its relation to the solar model age, although not precisely known, cannot be expected to introduce a larger uncertainty than the dynamical time scales associated with evolution down the Hayashi track. Yet, if we do not constrain the solar age, we obtain values around 4.9 to 5 Gyrs, which is an error of about 10 percent. Systematic errors in the models are well below 100 Myrs, and therefore below the discrepancy between the asteroseismic solar age and the meteoritic age[7]. So although ultimately this may not be the best way to untangle our results and characterise, in a simple way, what is wrong with the models, we have come to see the problem as one related to, or at least indicated by, the age. Unfortunately, nearly every model assumption, e.g., opacities (especially of the metals), primordial abundances (especially of helium, neon, carbon, oxygen, and nitrogen), convective transport theory and the modelling of the surface convective envelope, diffusion of helium and heavier elements, winds, mass loss, magnetic fields, rotational shear at the base of the convection zone, can affect the model age.

Our conviction is that the problems reported in this paper are not caused by inadequate frequencies or the general inability to use asteroseismology in the way we have presented. Rather, we think that all the tools and the data are now at an adequate level to show us the limitations of our models. Indeed, we would like to emphasise that evidence-based Bayesian studies are an excellent way to accurately assess future developments in solar modelling. They provide a fully consistent framework

---

[7]The age discrepancy predates the present work. The standard, often used, reference solar model Model S by Christensen-Dalsgaard et al. (1996) uses an age of 4.6 Gyrs measured from the ZAMS or approximately 4.64 Gyrs measured from the birthline.

to test observables, treat systematic errors (e.g., surface effects) and use prior information, in order to iterate towards more accurate model physics. Such is necessary to both better understand our Sun and to reap the full benefits of asteroseismology.

4.6 Acknowledgments

Bibliography

Alexander, D. R., & Ferguson, J. W. 1994, ApJ, 437, 879

Angulo, C., et al. 1999, Nuclear Physics A, 656, 3

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

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481

Bahcall, J. N., Pinsonneault, M. H., & Basu, S. 2001, ApJ, 555, 990

Bahcall, J. N., Pinsonneault, M. H., & Wasserburg, G. J. 1995, Reviews of Modern Physics, 67, 781

Bahcall, J. N., & Ulrich, R. K. 1988, Reviews of Modern Physics, 60, 297

Basu, S., & Antia, H. M. 1997, MNRAS, 287, 189

—. 2004, ApJ, 606, L85

Basu, S., Chaplin, W. J., Elsworth, Y., New, R., & Serenelli, A. M. 2009, ApJ, 699, 1403

Basu, S., Chaplin, W. J., Elsworth, Y., New, R., Serenelli, A. M., & Verner, G. A. 2007, ApJ, 655, 660

Böhm-Vitense, E. 1958, ZAp, 46, 108

Bouvier, A., & Wadhwa, M. 2010, Nature Geoscience, 3, 637

Broomhall, A.-M., Chaplin, W. J., Davies, G. R., Elsworth, Y., Fletcher, S. T., Hale, S. J., Miller, B., & New, R. 2009, MNRAS, 396, L100

Chandrasekhar, S. 1957, An introduction to the study of stellar structure., ed. Chandrasekhar, S. (Dover Publications, New York)

Chaplin, W. J., Serenelli, A. M., Basu, S., Elsworth, Y., New, R., & Verner, G. A. 2007, ApJ, 670, 872

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., Gough, D. O., & Thompson, M. J. 1991, ApJ, 378, 413

Christensen-Dalsgaard, J., et al. 1996, Science, 272, 1286

Cohen, E. R., & Taylor, B. N. 1986, Codata Bulletin No. 63, (New York: Pergamon Press)

Demarque, P., Guenther, D. B., Li, L. H., Mazumdar, A., & Straka, C. W. 2008, Ap&SS, 316, 31

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support, ed. Gregory, P. C. (Cambridge University Press)

Grevesse, N., & Sauval, A. J. 1998, Space Sci. Rev., 85, 161

Gruberbauer, M., Guenther, D. B., & Kallinger, T. 2012, ApJ, 749, 109

Gruberbauer, M., Kallinger, T., Weiss, W. W., & Guenther, D. B. 2009, A&A, 506, 1043

Guenther, D. B. 1994, ApJ, 422, 400

Guenther, D. B., Pinsonneault, M. H., & Bahcall, J. N. 1993, ApJ, 418, 469

Handberg, R., & Campante, T. L. 2011, A&A, 527, A56+

Hayashi, C. 1961, PASJ, 13, 450

Hickey, J. R., & Alton, B. M. 1983, in Solar Irradiance Variations of Active Region Time Scales, NASA Conference Publication 2310, ed. B. .J. LaBonte, G. A. Chapman, H. S. Hudson, & R. C. Wilson, 43

Houdek, G., & Gough, D. O. 2011, MNRAS, 418, 1217

Huber, D., et al. 2011, ApJ, 743, 143

—. 2012, ApJ, 760, 32

Iglesias, C. A., & Rogers, F. J. 1996, ApJ, 464, 943

Jeffreys, H. 1961, Theory of Probability, 3rd edn. (Oxford, England: Oxford University Press)

Kjeldsen, H., Bedding, T. R., & Christensen-Dalsgaard, J. 2008, ApJ, 683, L175

Krishna Swamy, K. S. 1966, ApJ, 145, 174

Lane, J. H. 1869, Amer. J. Sci., 2nd ser., 50, 57

Mathur, S., et al. 2012, ApJ, 749, 152

Metcalfe, T. S., et al. 2010, ApJ, 723, 1583

Palla, F., & Stahler, S. W. 1999, ApJ, 525, 772

125

Rogers, F. J. 1986, ApJ, 310, 723

Rogers, F. J., Swenson, F. J., & Iglesias, C. A. 1996, ApJ, 456, 902

Serenelli, A. M., Basu, S., Ferguson, J. W., & Asplund, M. 2009, ApJ, 705, L123

Serenelli, A. M., Haxton, W. C., & Peña-Garay, C. 2011, ApJ, 743, 24

Chapter 5

Paper III: Bayesian Asteroseismology of 23 Solar-Like Kepler Targets

M. Gruberbauer[1], D.B. Guenther[1], K. MacLeod[1]

[1]Institute for Computational Astrophysics, Department of Astronomy and Physics, Saint Mary's University, B3H 3C3 Halifax, Canada

T. Kallinger[2]

[2]Institute for Astrophysics, University of Vienna, Türkenschanzstrasse 17, A-1180 Vienna, Austria

Co-author contributions

- DG: construction of stellar model grids, valuable discussion.

- KM: analysis of Kepler-36, valuable discussion.

- TK: $\nu_{\max}$ for 16 Cyg A & 16 Cyg B, valuable discussion.

Abstract

We study 23 previously published Kepler targets to perform a consistent grid-based Bayesian asteroseismic analysis and compare our results to those obtained via the Asteroseismic Modelling Portal (AMP). We find differences in the derived stellar parameters of many targets and their uncertainties. While some of these differences can be attributed to systematic effects between stellar evolutionary models,

we show that the different methodologies deliver incompatible uncertainties for some parameters. Using non-adiabatic models and our capability to measure surface effects, we also investigate the dependency of these surface effects on the stellar parameters. Our results suggest a dependence of the magnitude of the surface effect on the mixing length parameter which also, but only minimally, affects the determination of stellar parameters. While some stars in our sample show no surface effect at all, the most significant surface effects are found for stars that are close to the Sun's position in the HR diagram.

5.1 Introduction

Ultra-precise long-term photometric time series from space have revolutionized the study of stellar variability in recent years. The CoRoT (Michel et al. 2008) and the *Kepler* (Borucki et al. 2010) space telescopes in particular have produced high-quality data sets for thousands of stars in order to detect planets down to Earth size and below. Particularly interesting for the study of stellar interiors and stellar evolution is their ability to detect solar-type oscillations from giants to subdwarfs. The pulsational characteristics of these stars adhere, at least to a very good first approximation, to scaling relations (e.g., Huber et al. 2011) permitting the study of large populations of stars with "ensemble asteroseismology" (Chaplin et al. 2011) and even Galactic archeology (Miglio et al. 2013).

The same information can also be exploited to infer the parameters of individual stars, e.g., to better constrain their planets' properties. For stellar astrophysics, however, the ultimate goal is to use asteroseismology to study stellar interiors. Instead of direct inversion of the pulsation information, asteroseismology usually employs a comparison between observed and modelled pulsation frequencies (e.g., Guenther & Brown 2004). Various new tools have been developed to facilitate a state-of-the-art version of such a comparison using different approaches, such as the Asteroseismic Modelling Portal (AMP) (Metcalfe et al. 2009) and Bayesian grid-based analysis (Gruberbauer et al. 2012, hereafter Paper I). The major differences between these methods

128

lie in their different statistical basis and their different applications of what is known as the surface effect correction (see Paper I for an in-depth discussion). Already, the AMP has been used to analyse some Kepler targets in detail and to compare the results with those from other modellers (Metcalfe et al. 2010, 2012). Such a comparison is advantageous, because asteroseismic modelling often relies on a specific set of stellar models with a specific set of input physics. Slight systematic differences among these models are therefore not only plausible but unavoidable, resulting in underestimated uncertainties. A different approach is to study a larger sample of stars self-consistently with one particular method and model base to facilitate a pool of results to be compared with other researcher's results (Mathur et al. 2012).

In this paper we reexamine some of the previously published studies based on Kepler data with a strong emphasis on AMP results, employing our own set of models and our Bayesian method described in Paper I. We will discuss how the results differ and whether the methodologies themselves introduce systematic deviations. We will also perform the first detailed study on surface effects for a sample of stars with our flexible method.

5.2 Methods, models and observations

5.2.1 Target selection and observations

In order to investigate the impact of the stellar models and methodologies in the most general sense, we analyse stars for which the p-mode frequency sets and detailed prior information used in previous asteroseismic fitting procedures are available in the literature. We furthermore constrain ourselves to stars that do not show strong signatures of deviations from the asymptotic relation, i.e., avoided crossings such as in KIC 11026764 (Metcalfe et al. 2010). While those signatures are very valuable for asteroseismic inferences and can be easily taken into account with our method as mentioned in Paper I, they would constitute special cases in the comparison between methods. We therefore postpone such an analysis to a future paper and restricted ourselves to 20 of the 22 stars analysed by Mathur et al. (2012) (hereafter M20), the

129

solar analogues 16 Cyg A & B (Metcalfe et al. 2012), and the planet-host Kepler-36 (Carter et al. 2012). Where available in the previously cited papers, we use prior constraints on $\log T_{\text{eff}}$, $\log L/L_\odot$, Z/X (adopting $[\text{Fe}/\text{H}]_\odot = 0.0245$) and $\log g$. Following our description in Paper I, these prior constraints are modelled as separate Gaussian probability distributions.

As is common in recent asteroseismic analyses, we treat the frequency of maximum power ν_{max} as an additional and independent observable by using the scaling relation

$$\nu_{\text{max,mod}} \approx \frac{M/M_\odot \, (T_{\text{eff}}/T_{\text{eff},\odot})^{3.5}}{L/L_\odot} \nu_{\text{max},\odot}. \tag{5.1}$$

where we employ the solar value $\nu_{\text{max},\odot} = 3120.0 \pm 5 \, \mu\text{Hz}$ given by Kallinger et al. (2010) based on VIRGO data. For M20, the observed values and uncertainties of ν_{max} have been taken from Mathur et al. (2012), and for Kepler-36 we have used the value published in Carter et al. (2012). For 16 Cyg A & B, we have determined the values ourselves by performing a Bayesian multi-component model fit, consisting of a flat background, three super-Lorentzian profiles and a Gaussian power hump, to the power density spectra of both data sets[1]. In this case, the central frequency of the Gaussian power hump and the corresponding uncertainties are interpreted as a good proxy for ν_{max}. The method employs the nested sampling algorithm MultiNest (Feroz et al. 2009) and is described in more detail in Kallinger et al. (2010). We find $\nu_{\text{max}} = 2215.6 \pm 6.5 \, \mu\text{Hz}$ for 16 Cyg A, and $\nu_{\text{max}} = 2571.9 \pm 12.6 \, \mu\text{Hz}$ for 16 Cyg B.

5.2.2 Models

A wide parameter range has to be spanned in order to perform a meaningful grid-based analysis. We therefore employed YREC (Demarque et al. 2008) to produce a set of dense grids covering a wide range in initial masses, and several values for the

---

[1]Note that only Q7 data obtained between September 2010 to December 2010 have been used in Metcalfe et al. (2012). We therefore restrict ourselves to this data set as well.

helium mass fraction Y_0, initial metal mass fraction Z_0, and mixing length parameter α_{ml}.

Our model tracks begin as completely convective Lane-Emden spheres (Lane 1869; Chandrasekhar 1957) with the stellar age reset to zero when the star crosses the birthline ($10^{-5}\,\mathrm{M_\odot/yr}$, Palla & Stahler 1999). They are evolved from the Hayashi track (Hayashi 1961) through the zero-age-main-sequence (ZAMS) to the base of the red giant branch. Constitutive physics include the OPAL98 (Iglesias & Rogers 1996) and Alexander & Ferguson (1994) opacity tables, as well as the Lawrence Livermore 2005 equation of state tables (Rogers 1986; Rogers et al. 1996). Convective energy transport was modelled using the Böhm-Vitense mixing-length theory (Böhm-Vitense 1958, MLT,). The atmosphere is implemented using Eddington gray atmosphere. Nuclear reaction cross-sections were taken from Bahcall et al. (2001) and the nuclear reaction rates from Table 21 in Bahcall & Ulrich (1988). The effects of helium and heavy element diffusion (Bahcall et al. 1995) were included. The model grid contains models with $M/M_\odot$ from 0.8 to 1.3 in steps of 0.01 and Y_0 from 0.210 to 0.315 in steps of 0.005. Furthermore, Z_0 is varied from 0.005 to 0.04 in steps of 0.005, but with the overall constraint that $X_0 \geq 0.68$. Lastly, we also vary α_{ml} from 1.8 to 2.4 in steps of 0.1.

The pulsation spectra were computed using the stellar pulsation code of Guenther (1994), which solves the linearized, non-radial, non-adiabatic pulsation equations using the Henyey relaxation method. The non-adiabatic solutions include radiative energy gains and losses but do not include the effects of convection. We estimate the random 1σ uncertainties of our model frequencies to be of the order of $0.1\,\mu\mathrm{Hz}$. These uncertainties are properly propagated into all further calculations.

5.2.3 Bayesian asteroseismic grid fitting

Our Bayesian fitting method is explained in detail in Paper I, and it has been previously applied to analyse the Sun (Gruberbauer & Guenther 2013). We compare theoretical (f_m) and observed (f_o) frequencies by calculating the likelihood that the

two values agree were it not for the presence of random and systematic errors, i.e.,

$$f_{\rm o} - f_{\rm m} = \gamma\Delta + e. \tag{5.2}$$

Here, the random errors e are assumed to be independent and Gaussian. The systematic errors $\gamma\Delta$ in the case of solar-like stars are assumed to be similar to "surface effects". At higher orders, observed frequencies are systematically lower than model frequencies, and the absolute frequency differences increase with frequency. This is modelled by introducing Δ as free parameters for each observed mode and by setting $\gamma = -1$.

These Δ terms are then allowed to become larger at higher radial orders. The upper limit $\Delta_{\rm max}$ for each model frequency $f_{\rm m}$ is determined by the large frequency separation and a power law similar to the standard correction introduced by Kjeldsen et al. (2008) so that

$$\Delta_{\rm max} = \Delta\nu \left(\frac{f_{\rm m}}{f_{\rm max,m}} \right)^b, \tag{5.3}$$

where $b = 4.9$, $\Delta\nu$ is the large frequency separation of the corresponding model, and $f_{\rm max,m}$ is the frequency of the highest order in the model[2].

The Δ parameter is incorporated in a completely Bayesian fashion, using a β prior to prefer smaller values over larger ones (see Paper I for more details). In addition, we always allow for the possibility that a mode is not significantly affected by any kind of systematic error by explicitly including the null hypothesis, that is by combining the probabilities of two hypotheses: one with and one without the Δ parameter. Altogether, this allows us to fully propagate uncertainties originating from the surface effects, or other potential systematic differences, into all our results. At the same time it gives us more flexibility than the standard surface-effect correction. Whereas the latter prescribes a fixed power-law behaviour for the actual surface

---

[2]This means that for the highest order in the model $\Delta_{\rm max} = \Delta\nu$ and guarantees that we do not introduce ambiguities in the radial orders by implementing the Δ terms.

effects, our method only prescribes such a behaviour for the *upper limits* of the surface effects for the individual radial model frequencies.

For each model in our grid, all the likelihood terms from the different frequencies are combined to yield an overall weighted likelihood for the model, where the weights are provided either via prior information or using ignorance priors (i.e., information that simply encodes the dimension of the grid). These weights provide correctly normalized probabilities that allow us to derive distributions for all model properties (e.g., mass, age, fractional radius of the base of the convection zone R_{BCZ}, mixing-length parameter α_{ML}, and so on).

In summary, we obtain probabilities for every evolutionary track in our grid, and within the tracks also for every model. We also obtain the correctly propagated distributions for systematic errors so that the model-dependent surface effect can be measured. In order to fully resolve the changes in stellar parameters and details in the stellar-model mode spectra, we oversample the evolutionary tracks via linear interpolation until the (normalised) probabilities no longer change significantly. Eventually, we obtain so-called evidence values, equivalent to the prior-weighted average likelihood, for the grid as a whole. These could, in principle, be used to perform a quantitative evaluation of different input physics (see Gruberbauer & Guenther 2013) or even different stellar evolution and pulsation codes. We will use them in this study to analyse the significance of the measured surface effects.

In order to facilitate this, we propose two alternative systematic error models in addition to the standard surface effect (SSE) model described above. First, we employ a less restrictive systematic error model where

$$\Delta_{\mathrm{max}} = \Delta\nu/2 \tag{5.4}$$

for every frequency of each particular model. Furthermore, the observed frequencies are allowed to deviate in either direction (first $\gamma = 1$ is evaluated, then $\gamma = -1$ follows,

and then both results are combined using the sum rule). We call this model the "arbitrary systematic error" (ASE) model since it allows, in principle, very large differences between observed and calculated frequencies without prescribing any systematic behaviour or preferred sign. Note that this is not equivalent to simply increasing the Gaussian uncertainties of the observed frequencies to $\Delta\nu/2$.

Finally, we will also employ a third error model which only consists of the probabilities obtained without any Δ parameters. This model therefore assumes that no systematic errors are present so that $f_{\mathrm{o}} - f_{\mathrm{m}} = e$. We will call this the "no systematic error" (NSE) model. Together, the three systematic error models will allow us to estimate the significance of surface effects or other systematic differences between observed and calculated frequencies.

5.3 Dependence of surface effects on non-adiabaticity and mixing length

The advantage of our method to include systematic frequency errors over the standard surface correction is its universality. The standard surface effect exponent of $b = 4.9$ as obtained by Kjeldsen et al. (2008) has been derived for adiabatic pulsation frequencies and for a solar-calibrated model with a calibrated parametrization of the mixing length[3]. More advanced pulsation models and different stellar models (see, e.g., Grigahcène et al. 2012) are not necessarily consistent with such a relation. This is also the case for our non-adiabatic pulsation frequencies. Since the only way to improve our modelling of outer layers is to compare more advanced models to observations, it is necessary to relax the constraint of a definite empirical surface correction relation dependent on adiabatic pulsation codes. Aside from the surface effect, our method also allows various other kinds of parameterisations for systematic errors, such as our ASE model.

---

[3]As explained in the previous section, technically we also use $b = 4.9$ for our surface effect modelling, but the exponent is only employed to derive an upper limit for the surface effect for each mode. This does not enforce the usual power-law like behaviour of the surface effect.

The drawback of our method, as discussed in Paper I, is that in the absence of strong prior information about the stellar parameters, a lack of lower-order modes will potentially result in an underestimated magnitude of the surface effects. This follows from the fact that we always obtain the most probable result given our state of information including the new data set; if we cannot constrain the stellar model parameters using our prior knowledge, the pulsation frequencies are our only reference. When conditions are encountered under which the empirical correction law of Kjeldsen et al. (2008) does not apply, or if one rejects such a correction on other grounds, we have to evaluate the models acknowledging the presence of less well-specified systematic errors.

As described in Paper I, neglecting lower order modes leads to overestimated α_{ml}, mass, and metallicity for the Sun, simply because such models can fit the higher order modes better. The same models cannot fit the lower order modes as well, but when they are not included in the list of fitted modes no penalty ensues. For stars other than the Sun, we usually do not have a complete list of lower-order modes, nor do we have as accurate non-seismic constraints (e.g., mass, luminosity, age). Even in such cases, however, stellar metallicity, T_{eff} and L can be estimated from spectroscopic and photometric observations. Furthermore, equation 5.1 reveals that ν_{max} provides valuable if approximate constraints for the fundamental parameters, including the stellar mass, in particular when spectroscopic constraints are available.

Two adjustable parameters of the stellar model, the helium abundance and α_{ml}, affect the structure of the surface layers. The mixing length parameter is normally tuned to produce a model of the Sun at the observed composition and (meteoritic) age that matches the limb-darkening-corrected radius of the Sun. The helium abundance is either derived also from a tuned model of the Sun, matching its luminosity, or extrapolated from the observed rate of Galactic nucleosynthesis. Both the helium abundance and the α_{ml} affect the depth of the convection zone (i.e., the fitted adiabat) and the temperature gradient in the superadiabatic layer (SAL)[4] via the

---

[4]Below the SAL, the temperature gradient is adiabatic.
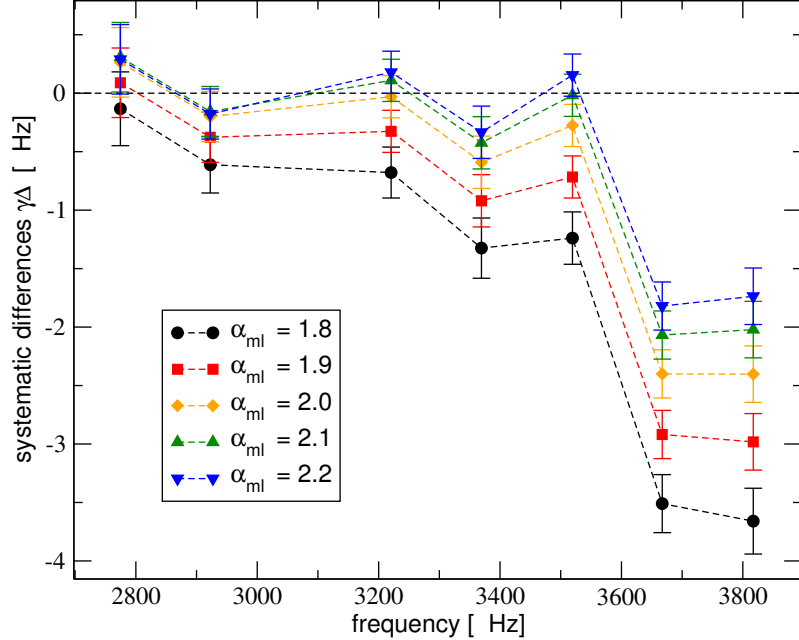
Figure 5.1: Systematic differences between observed and computed $l = 0$ modes for KIC 8006161 when fitted to models with varying mixing length but otherwise fixed initial parameters.

mixing length theory. We stress that the mixing length parameter of the MLT is used primarily to control the efficiency of convection and its adjustment is primarily used to fix the radius of the star. As is well known for the case of the Sun the MLT does not correctly predict the temperature gradients in the SAL so even though it may be providing an accurate radius for the star it may, at the same time, be providing a poor model of the SAL (e.g., Robinson et al. 2003). The surface effect is sensitive to α_{ml} since the p-mode frequencies are sensitive to the SAL. But at the same time the large frequency spacings are also sensitive to the α_{ml} via its effect on the star's radius. The interplay of the two effects of the mixing length parameter on the frequencies makes it difficult to isolate the surface effect completely from α_{ml}.

Figure 5.1 shows the effect of fitting one of the stars in our sample, KIC 8006161, to a specific evolutionary track with $M = 1.11\,M_\odot$, $Y_0 = 0.22$ and $Z_0 = 0.04$, but varying values of α_{ml} (note that these models are not equivalent to our most probable models as determined in the next section). At the highest frequencies, the larger α_{ml} values clearly reduce the measured surface effect by almost 50%, and the effect
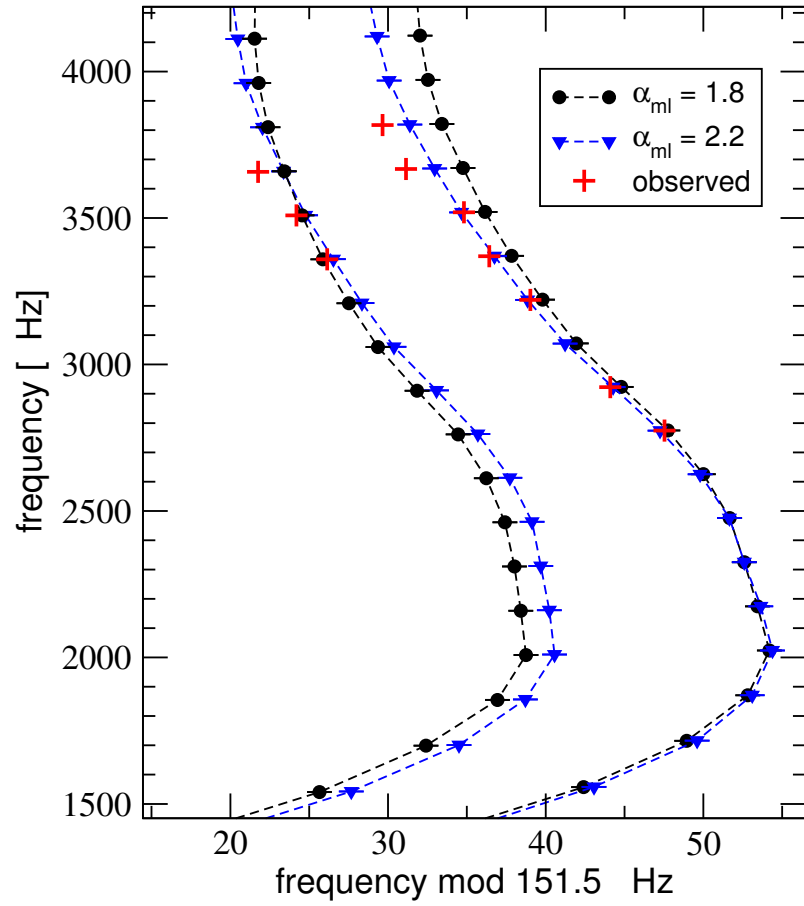
Figure 5.2: Echelle diagrammes for the l=0 (right sequence) and 2 (left sequence) modes of KIC 8006161 and two models with different α_{ml}. The uncertainties of the observed frequencies are of the order of the symbol size.

is even more pronounced at the lower orders.[5] On the other hand, the plot suggests that at the lowest observed radial order, the frequencies for the higher-α_{ml} models are somewhat too low. Figure 5.2 presents the echelle diagramme for the l=0,2 modes of the $\alpha_{\mathrm{ml}} = 1.8$ and $\alpha_{\mathrm{ml}} = 2.2$ models. We observe that if the set of l=2 modes extended below $\sim 3000\,\mu$Hz, we would be able to clearly distinguish between these two models. At the l=0 orders below $\sim 2000\,\mu$Hz the two models show small but systematic differences as well. With the current set of observed modes, however, we cannot clearly determine whether a lower or higher mixing length parameter value is more probable. Yet, we want to find the model with the smallest surface effects that still fits all other constraints. Therefore, in our example the higher α_{ml} values become more probable automatically. As long as we have limited knowledge on the magnitude of the surface effects across the HR diagram, however, this increase in probability might not be warranted. In the given example, it does seem as if the $\alpha_{\mathrm{ml}} = 2.2$ model is more consistent with the observed small spacing, but we know that the solar-calibrated value is closer to $\alpha_{\mathrm{ml}} \sim 1.8$, so deviations from this value should not be taken lightly[6]. Nonetheless, studying the possible variation of the mixing length parameter across the HRD and its interplay with the surface effects is important, so setting a fixed (calibrated) value is also not desirable.

We therefore propose the following solution: we perform our analysis using three different approaches to constraining the mixing length. The first approach is to not use any prior on α_{ml}. The second approach is to employ a Gaussian prior with $\alpha_{\mathrm{ml}} = 1.8 \pm 0.075$, based on the solar calibrated value. The standard deviation of the prior (0.075) is somewhat arbitrary, but we choose it to permit deviations from the calibrated value in the presence of strong evidence. As the maximum value of α_{ml} in our grid is 2.4, such a model would represent an *a priori* 8σ outlier. For such a model to still be more probable, it would require differences in likelihood of about 14 orders

---

[5]It is necessary to point out, however, that for adiabatic frequencies, the relative impact of the mixing length is not as big as for the non-adiabatic frequencies.

[6]Note that such deviations are also a non-negligible problem when applying the standard surface correction since it relies on the solar-calibrated values at the solar mixing length parameter.
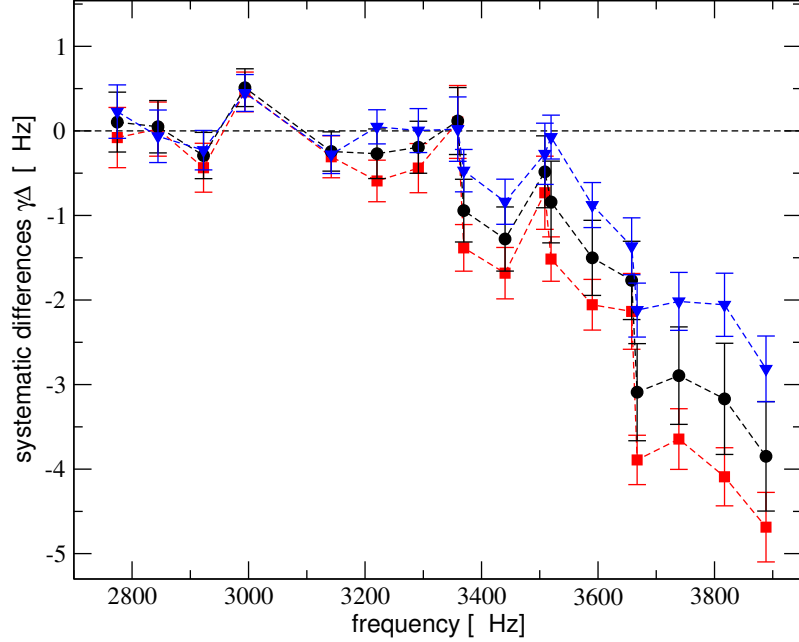
Figure 5.3: Systematic differences between all observed and computed modes of KIC 8006161 for the whole grid, calculated with (black circles) and without (blue triangles) α_{ml} prior. Results for only the $\alpha_{ml} = 1.8$ models (red squares) are also shown.

of magnitude, and therefore a large amount of evidence from the frequencies and the fit to the other stellar parameters. The prior should therefore only lead to $\alpha_{ml} > 2.1$ for stars that can be matched very well both in terms of their frequencies and in terms of their fundamental parameters. Lastly, the third approach is to constrain ourselves to $\alpha_{ml} = 1.8$ in reference to the Sun-calibrated value for Eddington atmospheres. This set of different constraints on α_{ml} will allow us to show its impact on the stellar parameters and the surface effects. By comparing the Bayesian evidence for the result obtained with different priors, we can also quantify the formal preference of one prior over the others.

As an example, we present the results for the surface effect analysis of KIC 8006161, based on the complete grid rather than just one evolutionary track, in Figure 5.3. While for this star the prior does not have a big effect at the lower order modes, we obtain significantly larger surface effects beyond $3300\,\mu$Hz with the α_{ml}

priors. Even with the Gaussian α_{ml} prior, as will be shown below, the most probable posterior value for α_{ml} lies above 1.8.
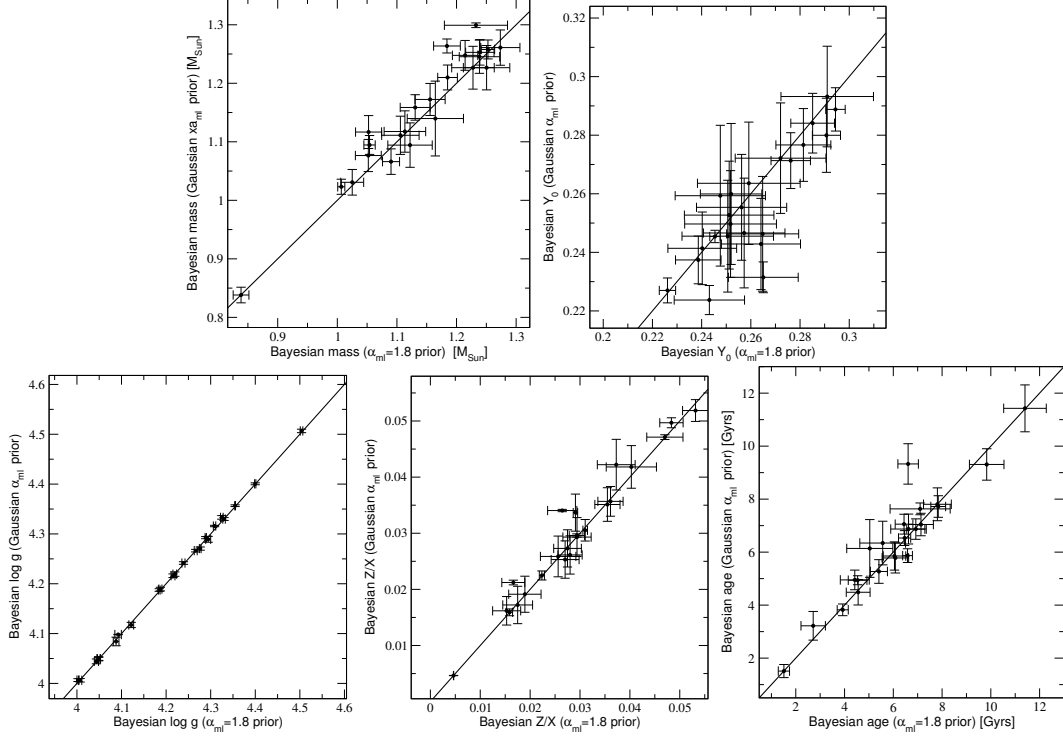
5.4 Results

As described in the previous sections, we have analysed all 23 stars in our sample with the same grid, using priors on their fundamental parameters if available and three different models relating to the treatment of systematic errors. Moreover, we perform this analysis three times, first setting $\alpha_{\mathrm{ml}} = 1.8$, then with a Gaussian prior, and lastly without a prior on α_{ml}. The results are given in Table 5.1, Table 5.2, and Table 5.3, and the most probable α_{ml} priors and surface effect models are also indicated.

5.4.1 The influence of the α_{ml} priors

Before we move on to a comparison to the literature, we first study the effect of the α_{ml} priors on our results. Figure 5.4 shows the posterior mean values and uncertainties of α_{ml}, M, Y_0, $\log g$, Z/X, and age for all stars and compares the results with and without the Gaussian α_{ml} prior. The Gaussian α_{ml} prior leads to slightly lower values of α_{ml} as was expected from the discussion in Section 5.3. Furthermore, the stellar masses are also slightly lower with an average difference $\langle \Delta M \rangle = -0.021\,\mathrm{M}_\odot$, and, although there is a larger scatter in Y_0, slightly larger values in the initial helium mass fraction are also preferred with an average difference of $\langle \Delta Y_0 \rangle = 0.008$. On the other hand, $\log g$ remains basically unaffected as expected, since the radius of the stars are well constrained by the large spacings (as we will see below, this also extends to a comparison with the other α_{ml} prior and the literature). Z/X and age also do not show strong systematic effects. Nonetheless, the latter does exhibit a strong outlier with 16 Cyg B, for which the age changes from $9.279 \pm 0.473\,\mathrm{Gyrs}$ to $6.532 \pm 0.281\,\mathrm{Gyrs}$. Note that even though the Bayesian evidence is clearly in favour of the older model, the younger value is much more reasonable,
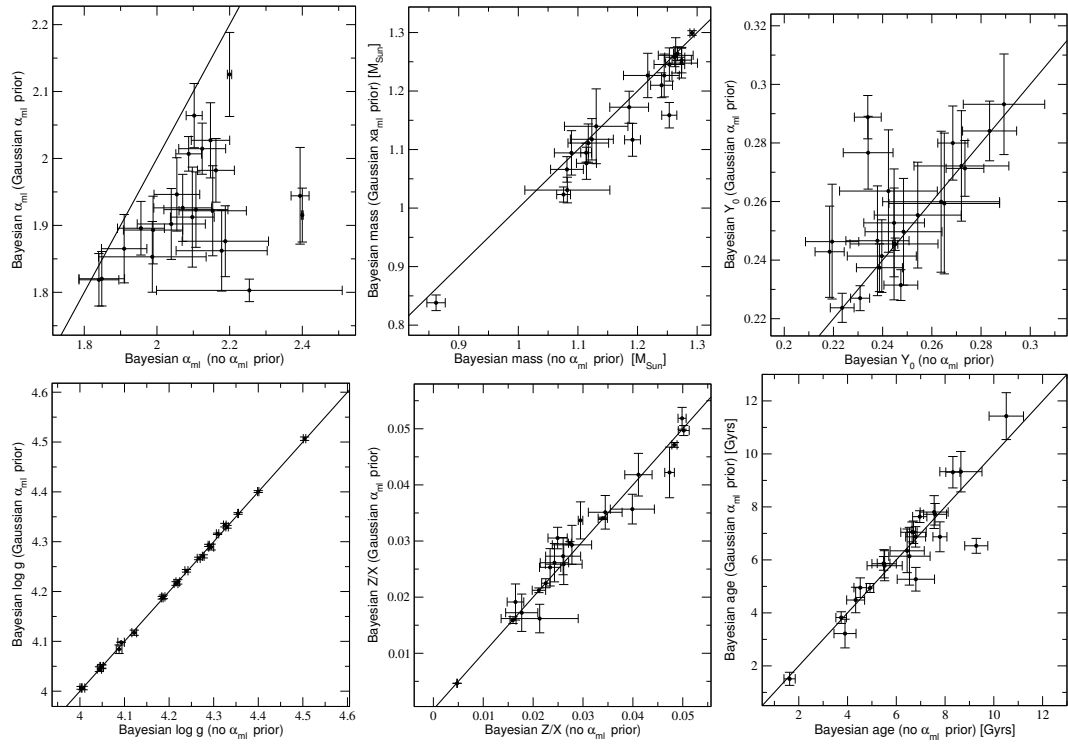
Figure 5.4: The effect of the Gaussian α_{ml} prior on the posterior value of various model parameters. Results are plotted for the most probable systematic error model as given in, e.g., Table 5.1. The black line indicates a ratio of unity.

141

Figure 5.5: Same as Figure 5.4 but for the $\alpha_{ml} = 1.8$ prior and the Gaussian α_{ml} prior. Comparison of α_{ml} is not shown.

given the results from Metcalfe et al. (2012) and also given the value of the age for 16 Cyg A. This is a good test case for the impact of the α_{ml} prior.

A comparison of the results from the Gaussian prior and the fixed $\alpha_{ml} = 1.8$ prior is presented in Figure 5.5. In general, the results fall in line with our expectations: the mass is now slightly larger for the Gaussian prior with $\langle \Delta M \rangle = 0.014$, and Y_0 is slightly smaller with $\langle \Delta Y_0 \rangle = -0.005$. Z/X and age values are again quite comparable except for a few outliers. In general the systematic differences between the Gaussian prior and the $\alpha_{ml} = 1.8$ prior results are smaller than those obtained in a comparison without any priors on α_{ml}. Overall, our comparison reveals that stronger constraints on α_{ml} do not perturb the parameters outside the uncertainties and produce slightly lower stellar masses and higher Y_0.

In terms of the systematic errors, in particular the surface effect, the results also follow our conclusions from the previous section. Figure 5.6 shows the differences

in the systematic errors that arise by using the two α_{ml} priors for every mode of every star in our sample. Using the Gaussian α_{ml} prior leads to an increase in the magnitude of the surface effects (= more negative systematic differences between observed and calculated frequencies) in general. There are only a few stars in the sample for which the effect is very pronounced. It is interesting that for many modes the Gaussian α_{ml} prior does not produce large differences for the surface effects. This is due to the fact that we find a number of stars for which the surface effects are not significant unless we restrict the analysis to the $\alpha_{\mathrm{ml}} = 1.8$ models. Consequently, switching to the $\alpha_{\mathrm{ml}} = 1.8$ prior results in even bigger surface effects and to significant surface effects for more stars in the sample. This is also reflected in the strong preference for the "surface effect" systematic error model, as shown in the result tables. In both panels there are also a few outliers for which the priors lead to decreased surface effects (= more positive systematic differences between observed and calculated frequencies), but these modes belong to the stars for which the "arbitrary systematic error" model is either preferred or very similar in probability to the surface effect model.

As indicated in our result tables, using no α_{ml} prior often leads to the highest evidence. Larger evidence values require that the models are formally more consistent with all our available constraints while also minimizing the systematic errors, i.e., the surface effects. Therefore, the analysis which yields the highest evidence and thus the corresponding stellar parameters are usually interpreted as being most appropriate. As explained in Section 5.3, however, we stress that at this point it is necessary to present the results from all approaches, and not to put too much confidence into the formal preference over to α_{ml} priors. This follows simply because we do not possess enough low-order modes or additional information to anchor the surface effect relation. The only clear exception to this are KIC 8379927 and KIC 10516096, for which we do not detect significant systematic errors irrespective of the α_{ml} priors but still find higher α_{ml} to be most probable. Concerning the impact of α_{ml} on the other stellar parameters, however, only the stellar mass and Y_0 seem to be somewhat systematically affected by the choice of priors. Even for those parameters the deviations are usually
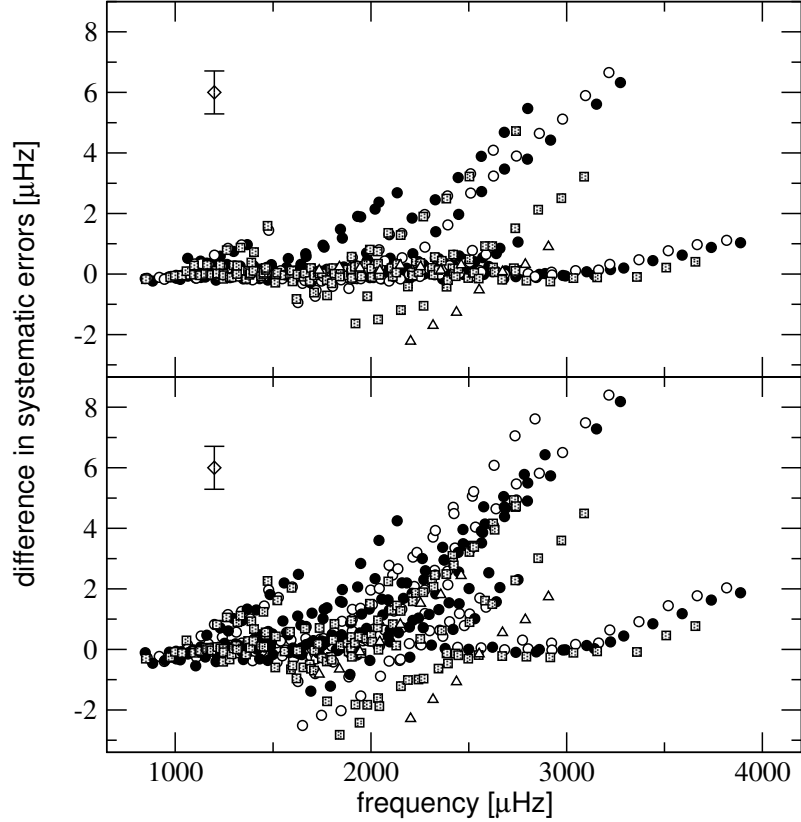
Figure 5.6: Differences in the measured systematic errors that arise from using the Gaussian α_{ml} prior (top panel) or the $\alpha_{\mathrm{ml}} = 1.8$ prior (bottom panel). All modes of all stars are shown: l=0 modes (open circles), l=1 modes (black circles), l=2 modes (shaded squares), and l=3 modes (open triangles). Positive (negative) values denote bigger (smaller) systematic errors in terms of surface effects when the α_{ml} priors are used. The average uncertainty of the differences is indicated by the diamond in the upper left. For each star, the plotted differences were obtained using the most probable systematic difference model for the respective α_{ml} prior.

within the quoted uncertainties. Thus, for our comparison with the values published in the literature, which also allow different values of α_{ml}, we constrain ourselves to the results obtained using the "intermediate approach", the Gaussian α_{ml} prior, and refer to our tabulated results for the differences arising from the different priors.

5.4.2 Surface effects and other systematic frequency differences

As previously alluded to, Figure 5.6 suggests that many stars do not show strong evidence for surface effects when our non-adiabatic models are used in tandem with the Gaussian prior. The situation changes, however, when the $\alpha_{\mathrm{ml}} = 1.8$ prior is

used. This implies that, depending on the prior, the convective contributions to the surface effects are either more or less significant. Since a proper normalization of the surface effect amplitudes is not trivial and the shape of the surface effects can vary from star to star, we instead quantify the significance of the surface effect in terms of probabilities. As discussed in Section 5.2.3, our calculations consider three different systematic error models: SSE, ASE, and NSE. Therefore, in order to quantify the surface effect significance for every star, we simply calculate the odds ratio

$$ODDS = \frac{ev(SSE)}{ev(ASE) + ev(NSE)}, \tag{5.5}$$

where $ev(SSE)$, $ev(ASE)$, and $ev(NSE)$ are the evidence values obtained for the analysis using each specific systematic error model[7]. This is the probability ratio between the hypotheses "standard surface effect" and "either arbitrary systematic errors or no systematic errors". Therefore, if surface effects are needed to explain the observations, we expect that $ODDS \gg 1$. According to the convention established by Jeffreys (1961), the evidence for or against one of the two hypotheses is considered "substantial" for a factor of 3 to 10, "strong" for a factor of 10 to 30, "very strong" for a factor of 30 to 100, and "decisive" for factors above 100. Hence, when the surface effects become more significant with respect to the other hypotheses, $ODDS$ will increase as well.

Our calculations show that for some stars the significance of the individual systematic error models depend on the specific prior for α_{ml}, in accordance with what was discussed in Section 5.3. However, there are four stars for which $ODDS < 1$ irrespective of mixing length parameter: KIC 6933899, KIC 8379927, KIC 10516096, and Kepler-36. The latter three objects do not require any systematic errors at all. Furthermore, for KIC 6106415, KIC 6603624, and KIC 11244118 the surface effect model is only significant for the $\alpha_{ml} = 1.8$ prior.

---

[7]This assumes that *a priori* all three surface effect models are equally probable.
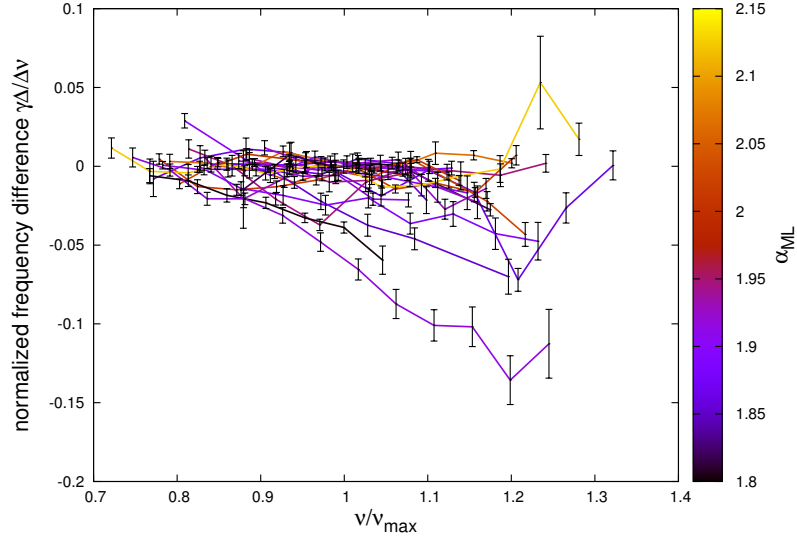
Figure 5.7: Normalized systematic frequency differences as a function of normalized frequency for $l = 0$ modes for the results obtained with the Gaussian α_{ml} prior. The colour represents the mean posterior α_{ml}. For each star, the plotted differences were obtained using the most probable systematic difference model.

Figure 5.7 shows the actual systematic error measurements obtained when using the Gaussian α_{ml} prior that have been rescaled and plotted as a function of their mean α_{ml}. For many stars the individual deviations do not seem to correspond to the clear power-law behaviour that can be identified for the Sun. Furthermore, there appears to be a very weak dependence on α_{ml}, where higher values are related to smaller normalised surface effects, as expected from the discussion in Section 5.3. Whether this dependence is physically meaningful depends on whether these stars actually have higher values of α_{ml}, or if it is simply the case that our α_{ml} prior is too weak. In any case, α_{ml} and surface effects are related.

Similar to Mathur et al. (2012), we do not find any simple correlations of the normalised surface effect with any of the other parameters in Table 5.1 to Table 5.3. However, studying the *significance* of the surface effect in terms of probabilities reveals some interesting results. Figure 5.8 shows the logarithm of the odds ratio for all stars in our sample as a function of their position in the HR diagram. The most significant detections appear to be situated at close-to-solar values of T_{eff} and the picture is

146
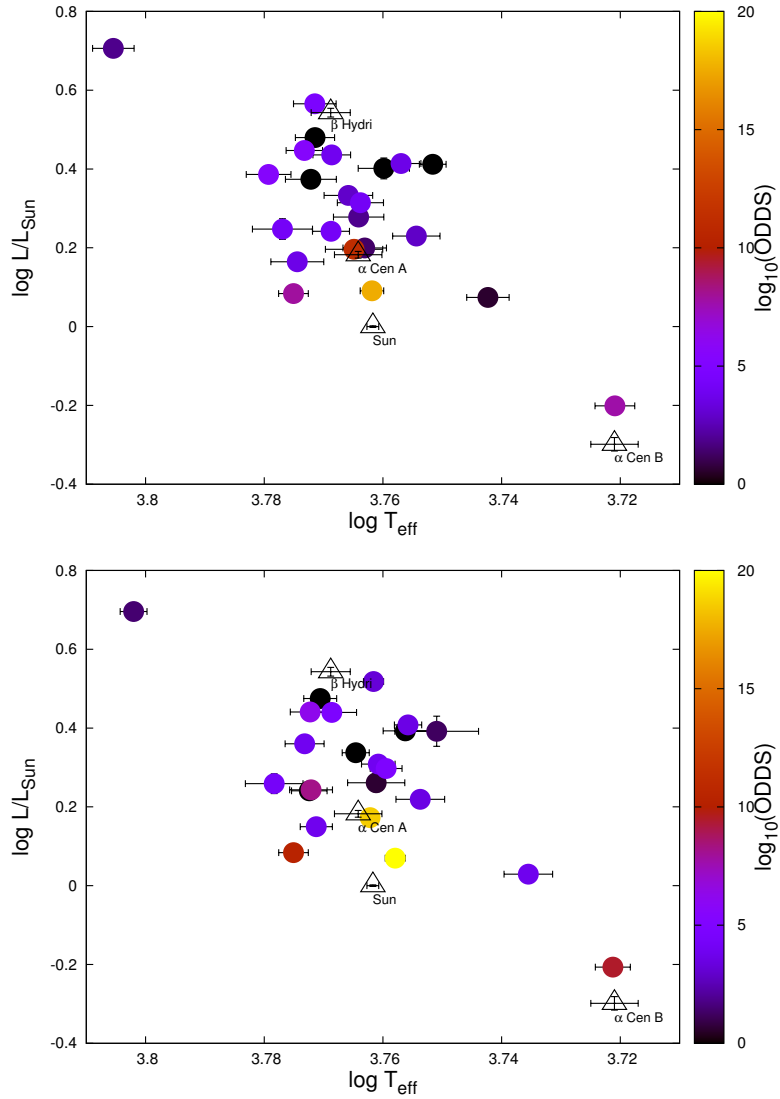
Figure 5.8: HR diagram of all stars in our sample (filled circles) with parameter taken from Table 5.1 to Table 5.3, using the results from the Gaussian (top panel) and $\alpha_{\mathrm{ml}} = 1.8$ (bottom panel) prior. The colour indicates the significance of the detected surface effect using $\log_{10}(ODDS)$. Four other well-known stars with surface effects are also shown as triangles. For each star, the plotted parameters were obtained using the most probable systematic difference model for the respective α_{ml} prior.

similar whether the Gaussian α_{ml} prior, the $\alpha_{\mathrm{ml}} = 1.8$ prior, or no α_{ml} prior is used. Furthermore, the coolest star in the sample, KIC 8006161, also displays highly significant surface effects but lies far off from the main bulk of the sample. We have also added symbols representing the Sun (Gruberbauer & Guenther 2013), β Hydri (Brandão et al. 2011), and α Cen A& B (Eggenberger et al. 2004), all of which were used by Kjeldsen et al. (2008) to define the surface effect correction. Except for β Hydri[8], the stars fit well into the pattern given by the Kepler stars. 16 Cyg A, 16 Cyg B,α Cen A, and of course the Sun, appear to lie on the "surface effect locus" in the HRD diagram of our sample. α Cen B, on the other hand, is situated very close to KIC 800616.

The stars for which no significant surface effects were detected do mix with stars that show less significant detections, which is why there does not seem to be a strong correlation of the surface effect with any particular parameter. On average, however, lower luminosities and higher effective temperatures correspond to more significant surface effects. Plotting $\log T_{\mathrm{eff}}$ against $\log g$ (not shown) necessarily yields a very similar picture which again clusters the most significant detections at the solar values. A correlation of the surface effect amplitude with $\log g$ was already noted by Mathur et al. (2012). Our comparison of the significance of the surface effect would be more in line with their investigation of the normalised surface effect for which they could not find a strong correlation. It will be intriguing to see whether a bigger sample and additional lower order modes could lead to a clearer detection of a "surface effect locus" in the HR diagram.

In any case, the non-detection of surface effects in some stars, as well as the concentration of very significant surface effects for stars with close to solar values should be a warning for unreflected usage of the standard surface correction for all solar-like stars.

---

[8]Note that the surface effects detected in β Hydri have only been measured using adiabatic frequencies which do not contain the correction for radiative gains and losses.

### 5.4.3   Comparison with non-Bayesian results

5.4.3.1   M20

In this section we investigate the presence of potential systematic differences between our results and those obtained using the AMP pipeline. Figure 5.9 shows that there are no strong systematic trends in either of the plotted parameters. As in the comparison between our three different Bayesian analyses (Figure 5.4, Figure 5.5), the determined $\log g$ values are very similar, but the Bayesian uncertainties are usually smaller. The results for α_{ml} show large scatter which is mostly compensated by the large uncertainties. It should be noted that our grid only extends from $\alpha_{\mathrm{ml}} = 1.8$ to 2.4, and therefore we do not cover the lower values that AMP returns for some of the stars.

The masses that were determined are quite similar for most stars, but the AMP delivers smaller uncertainties on average. For several stars, only the larger uncertainties reported by the Bayesian method can help to reconcile the results. There exists also a clear outlier with KIC 11244118 where the masses differ by about $0.3\,M_{\odot}$, more than 15 times our statistical uncertainty.[9] The initial helium mass fraction again displays large but seemingly unsystematic scatter, in particular when compared to some of the uncertainties reported by AMP. On the other hand, the results for Z/X are more similar, but our values appear to be slightly larger in a systematic way. In general, we have to stress that concerning the chemical composition, our grid is quite coarse compared to the capabilities of AMP's genetic algorithm.

Lastly, significant differences appear in the comparison of the determined ages. Irrespective of potential differences in the definition of zero-age models, the two methods yield different results with significant scatter. Moreover, the Bayesian age uncertainties appear to be bigger on average by a factor of 6, which is substantial, necessary, but insufficient to reconcile the results in many cases.

---

[9]This star is also problematic since it fits best to models near the border of our grid both in terms of mass and metallicity.
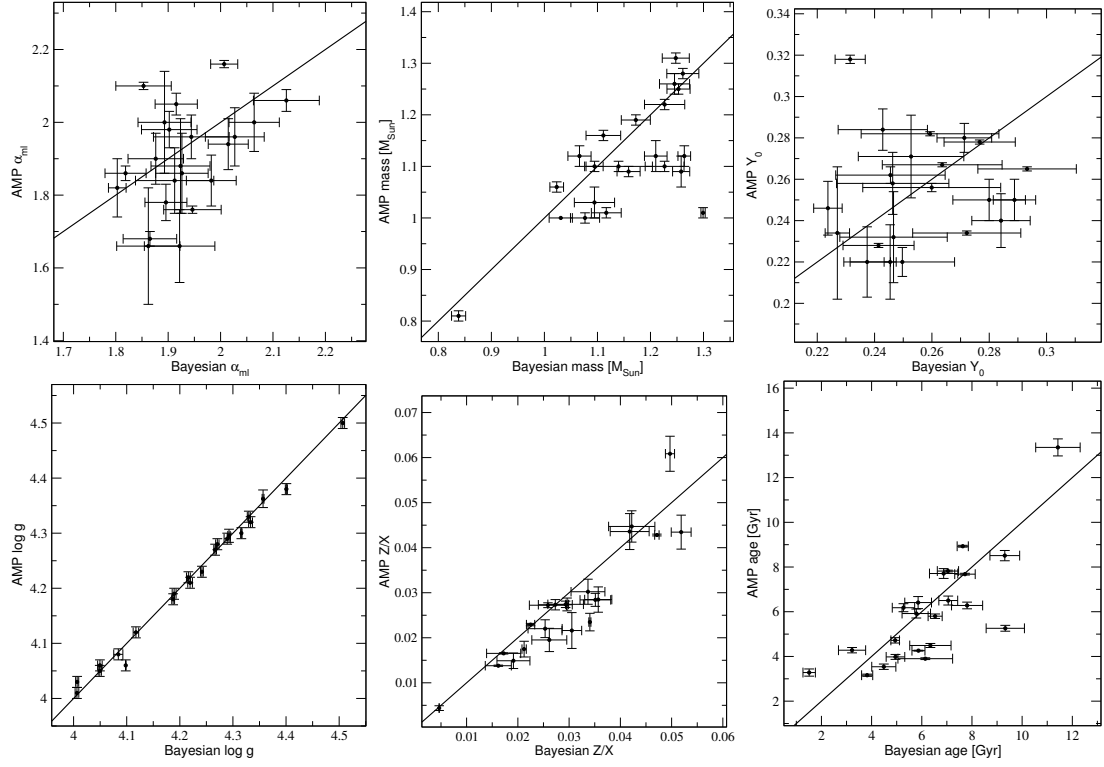
Figure 5.9: Same as Figure 5.4 but comparing the results from our Bayesian approach using the α_{ml} prior with the published results obtained via the AMP pipeline (Mathur et al. 2012; Metcalfe et al. 2012). Note that Kepler-36 is not included in these plots.

5.4.3.2 16 Cyg A & B

The modelling performed by Metcalfe et al. (2012) revealed that 16 Cyg A & B are of slightly different masses but have a similar age, as expected for a binary system. Several different grids and methods were used, including AMP, to arrive at an average ensemble solution. Our results compare favourably with this ensemble average, when it comes to the ages, the masses, Z_0, and α_{ml}. Except for the mass of 16 Cyg B, for which we obtain $1.023 \pm 0.013\, M_\odot$ compared to their result of $1.07 \pm 0.02\, M_\odot$, these parameters overlap within their respective 1σ uncertainties. It should be noted that we obtain a lower mass for 16 Cyg A as well, which might suggest a systematic difference between the methods and models used. As discussed in the previous section, however, we don't find that such a trend is true for our larger sample. The ages are fully consistent with a common origin, even though this constraint was not used in the analysis[10].

We find a slight discrepancy for the initial helium mass fraction. For 16 Cyg A we obtain $Y_0 = 0.282 \pm 0.01$ and for 16 Cyg B we find $Y_0 = 0.285 \pm 0.01$, while Metcalfe et al. report 0.25 ± 0.01. Overall, we observe that the differences between our results and the ensemble average in the literature are minor.

Comparing our results exclusively to the AMP values, we see a significant difference in the age and the value of α_{ml} for 16 Cyg B. It is interesting that this star is among the set of the most significant surface-effect detections in our sample. As the AMP results in a value of $\alpha_{\mathrm{ml}} = 2.05 \pm 0.03$, which is bigger than the ensemble average, it is perhaps the combination of the solar-calibrated surface effect correction and the use of a higher-than-solar α_{ml} which results in the discrepancy. For the age, we obtained $6.532 \pm 0.281\,\mathrm{Gyr}$ compared to 5.8 ± 0.1. Consistent with our findings in Section 5.4.3.1, we observe that our age uncertainties are significantly bigger.

In a recent paper, White et al. (2013) have combined interferometric diameters from CHARA observations with Hipparcos parallaxes, spectrophotometric bolometric

---

[10]The equal age is in even better agreement with our results for the $\alpha_{\mathrm{ml}} = 1.8$ prior, but for this approach we also obtain substantially smaller masses.

fluxes, and the asteroseismic large frequency separation, to obtain largely model-independent constraints for 16 Cyg A & B. In comparison to their results, for 16 Cyg A, our $\alpha_{ml} = 1.8$ prior produces a very close match in terms of mass and radius, but the model T_{eff} values are slightly too low and match better for the Gaussian α_{ml} prior. For 16 Cyg B, on the other hand, the higher α_{ml} values are more consistent with their results, predicting higher masses and larger radii but again T_{eff} values that are not quite high enough to match the mean observed values. These slight differences however are insignificant and, irrespective of the particular priors used, we find that our results match the masses, temperatures, and radii from White et al. (2013) reasonably well and in all cases to within the combined 1.5σ uncertainties. Therefore, the interferometric uncertainties are too large to give strong evidence for or against our particular solutions (i.e., in particular the different α_{ml} values). This can also be interpreted as additional justification for the various α_{ml} priors, since the range of results allows us to define a parameter space that is more in line with model-independent observations.

5.4.3.3   Kepler-36

With respect to Kepler-36, we find that we can match all parameters published in Carter et al. (2012) within the uncertainties. It is interesting, however, that we do not detect any surface effects for this star. Carter et al. report that the surface-effect correction was applied to the frequencies. Judging from our results, any surface effects necessary to be corrected for this star would have to originate from the radiative losses that are already taken into account in our non-adiabatic models.

5.5   Conclusions

In this paper we have reported on our asteroseismic analysis of 23 previously published stars that were observed with the Kepler satellite. We compared the results obtained with our Bayesian grid-based method to the results from the literature, most importantly those obtained with the AMP. Except for a weak trend towards larger

values of Z/X with our method, no obvious systematic differences in the basic stellar parameters can be found. In part, this is certainly due to spectroscopic constraints (T_{eff}, $\log g$, $[Fe/H]$, $L/L_\odot$) that were used in by all authors.

However, we observe that the uncertainties derived from the two methods differ substantially for some stellar parameters. Uncertainties in the stellar ages in particular are either significantly underestimated by AMP or significantly overestimated by the Bayesian method. We conclude that the flexible treatment of the surface effects in the Bayesian approach is probably responsible for this discrepancy. Different values of α_{ml} and the usage of non-adiabatic models require a more flexible treatment of the surface effect. Therefore, in our view the uncertainties derived with our method more adequately represent our actual state of knowledge about the surface effects and are therefore more realistic. On the other hand, the interplay between the surface effect and α_{ml} introduces another layer of complexity in the analysis which has to be taken into account in the determination of the stellar parameters. We propose that future studies with more stars should aim to reexamine this interdependence, especially as long as non-seismic constraints on α_{ml} are not available.

Concerning the surface effects themselves, we find that with a Gaussian prior on α_{ml}, only a few stars in our sample actually require larger corrections. 6 stars in our sample do not show strong evidence for any surface effect at all. Compared to the results in Mathur et al. (2012), this suggests that for many stars taking into account the radiative losses is already good enough. On the other hand, using only models with $\alpha_{\text{ml}} = 1.8$ leads to more significant detections. Irrespective of the prior on α_{ml}, we also discovered that the stars for which we do find a highly significant surface effect appear to be located very close to the Sun in the HR diagram (see Figure 5.8). A comparison with the stars that were used to derive the traditional surface-effect correction (Kjeldsen et al. 2008) shows that most of these calibrators - including the Sun - also fit the picture. As radiative losses are already taken into account in our models, the modelling of convection and its dependencies on element abundances,

opacities, and the equation of state remains a leading candidate to explain the cause of the surface effects.

To conclude, although systematic differences between stellar evolutionary codes are still affecting the individual stellar parameters, the systematic analysis of surface effects can already be pursued using more advanced methods than the standard surface correction, such as our Bayesian approach. No matter which surface correction is used, however, the constraints on α_{ml} will potentially affect the results in the absence of lower-order modes. The data sets on which this analysis is based have since been superseded by many more quarters of Kepler data. Also, many more stars have been observed, for which public frequencies are also available (Appourchaux et al. 2012). Strong spectroscopic constraints and access to lower-order modes will be necessary to improve our analysis, and to see whether the "surface effect locus" can be reproduced with a larger sample of stars and better data. Given the large number of subgiants and red giants observed with *Kepler* and CoRoT, a similar study for non-main sequence stars could be very illuminating as well.

Acknowledgments

Bibliography

Alexander, D. R., & Ferguson, J. W. 1994, ApJ, 437, 879

Appourchaux, T., et al. 2012, A&A, 543, A54

Bahcall, J. N., Pinsonneault, M. H., & Basu, S. 2001, ApJ, 555, 990

Bahcall, J. N., Pinsonneault, M. H., & Wasserburg, G. J. 1995, Reviews of Modern Physics, 67, 781

Bahcall, J. N., & Ulrich, R. K. 1988, Reviews of Modern Physics, 60, 297

Böhm-Vitense, E. 1958, ZAp, 46, 108

Borucki, W. J., et al. 2010, Science, 327, 977

Brandão, I. M., et al. 2011, A&A, 527, A37+

Carter, J. A., et al. 2012, Science, 337, 556

Chandrasekhar, S. 1957, An introduction to the study of stellar structure., ed. Chandrasekhar, S. (Dover Publications, New York)

Chaplin, W. J., et al. 2011, Science, 332, 213

Demarque, P., Guenther, D. B., Li, L. H., Mazumdar, A., & Straka, C. W. 2008, Ap&SS, 316, 31

Eggenberger, P., Charbonnel, C., Talon, S., Meynet, G., Maeder, A., Carrier, F., & Bourban, G. 2004, A&A, 417, 235

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Grigahcène, A., Dupret, M.-A., Sousa, S. G., Monteiro, M. J. P. F. G., Garrido, R., Scuflaire, R., & Gabriel, M. 2012, MNRAS, 422, L43

Gruberbauer, M., & Guenther, D. B. 2013, MNRAS

Gruberbauer, M., Guenther, D. B., & Kallinger, T. 2012, ApJ, 749, 109

Guenther, D. B. 1994, ApJ, 422, 400

Guenther, D. B., & Brown, K. I. T. 2004, ApJ, 600, 419

Hayashi, C. 1961, PASJ, 13, 450

Huber, D., et al. 2011, ApJ, 743, 143

Iglesias, C. A., & Rogers, F. J. 1996, ApJ, 464, 943

Jeffreys, H. 1961, Theory of Probability, 3rd edn. (Oxford, England: Oxford University Press)

Kallinger, T., et al. 2010, A&A, 522, A1

Kjeldsen, H., Bedding, T. R., & Christensen-Dalsgaard, J. 2008, ApJ, 683, L175

Lane, J. H. 1869, Amer. J. Sci., 2nd ser., 50, 57

Mathur, S., et al. 2012, ApJ, 749, 152

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

Metcalfe, T. S., et al. 2010, ApJ, 723, 1583

—. 2012, ApJ, 748, L10

Michel, E., et al. 2008, Science, 322, 558

Miglio, A., et al. 2013, MNRAS, 429, 423

Palla, F., & Stahler, S. W. 1999, ApJ, 525, 772

Robinson, F. J., Demarque, P., Li, L. H., Sofia, S., Kim, Y.-C., Chan, K. L., & Guenther, D. B. 2003, MNRAS, 340, 923

Rogers, F. J. 1986, ApJ, 310, 723

Rogers, F. J., Swenson, F. J., & Iglesias, C. A. 1996, ApJ, 456, 902

White, T. R., et al. 2013, MNRAS, in press, arXiv 1305.1934

Table 5.1: Mean parameters and uncertainties as a function of α_{ml} prior for KIC 3632418 to KIC 6603624. Bold font indicates the prior for which the highest evidence was obtained, as well as other priors for which the evidence was comparable (within a factor of 5). X_0, Z_0: initial hydrogen and metal mass fractions; Z_{s}: metal mass fraction in the envelope; R_{BCZ}: fractional radius of the base of the convection zone; α_{ml}: mixing length parameter; sys: the most probable systematic-error model is given (SSE = standard surface effect, ASE = arbitrary systematic errors, NSE = no systematic errors) and asterisks indicate a probability contrast of less than an order of magnitude with respect to any of the other systematic-error models.

Star	α prior	$M/M_\odot$	$\log T_{\mathrm{eff}}$	$\log L/L_\odot$	$\log R/R_\odot$	Age	X_0	Z_0	Z_{s}	Z_s/X_s	R_{BCZ}	α_{ml}	sys
3632418	$\boldsymbol{\alpha_{\mathrm{ml}}=1.8}$	1.273	3.802	0.696	0.268	3.926	0.735	0.0134	0.0130	0.0175	0.8397	1.80	SSE
		±0.033	±0.002	±0.007	±0.004	±0.227	±0.012	±0.0024	±0.0022	±0.0030	±0.0069		
	Gaussian	1.261	3.805	0.706	0.266	3.823	0.727	0.0130	0.0126	0.0172	0.8405	1.87	SSE
		±0.030	±0.003	±0.012	±0.004	±0.221	±0.024	±0.0025	±0.0022	±0.0033	±0.0088	±0.05	
	no α_{ml} prior	1.264	3.807	0.713	0.266	3.738	0.723	0.0133	0.0129	0.0177	0.8386	1.91	SSE
		±0.029	±0.003	±0.012	±0.004	±0.217	±0.024	±0.0024	±0.0021	±0.0032	±0.0093	±0.06	
3656476	$\alpha_{\mathrm{ml}}=1.8$	1.131	3.754	0.219	0.126	6.623	0.688	0.0310	0.0273	0.0373	0.6874	1.80	SSE
		±0.025	±0.004	±0.017	±0.003	±0.729	±0.011	±0.0028	±0.0026	±0.0038	±0.0090		
	Gaussian	1.159	3.754	0.230	0.130	6.871	0.689	0.0347	0.0308	0.0422	0.6732	1.94	SSE
		±0.022	±0.004	±0.017	±0.003	±0.564	±0.012	±0.0034	±0.0031	±0.0045	±0.0089	±0.07	
	no α_{ml} prior	1.253	3.766	0.301	0.143	7.789	0.726	0.0400	0.0359	0.0473	0.6591	2.39	NSE*
		±0.013	±0.002	±0.007	±0.002	±0.287	±0.010	±0.0005	±0.0005	±0.0010	±0.0015	±0.02	
4914923	$\alpha_{\mathrm{ml}}=1.8$	1.228	3.759	0.297	0.154	5.409	0.710	0.0306	0.0271	0.0361	0.7097	1.80	SSE
		±0.036	±0.003	±0.013	±0.004	±0.349	±0.021	±0.0017	±0.0016	±0.0025	±0.0078		
	Gaussian	1.227	3.764	0.314	0.153	5.269	0.707	0.0299	0.0266	0.0357	0.7075	1.88	SSE
		±0.037	±0.004	±0.018	±0.004	±0.446	±0.021	±0.0018	±0.0017	±0.0026	±0.0092	±0.05	
	no α_{ml} prior	1.245	3.769	0.343	0.157	6.802	0.724	0.0337	0.0302	0.0399	0.6812	2.19	SSE
		±0.025	±0.005	±0.021	±0.003	±0.766	±0.020	±0.0035	±0.0032	±0.0044	±0.0102	±0.12	
5184732	$\alpha_{\mathrm{ml}}=1.8$	1.239	3.761	0.261	0.132	4.421	0.684	0.0394	0.0350	0.0483	0.7258	1.80	SSE*
		±0.024	±0.005	±0.023	±0.003	±0.594	±0.008	±0.0016	±0.0015	±0.0023	±0.0143		
	Gaussian	1.253	3.764	0.278	0.135	4.951	0.689	0.0400	0.0360	0.0497	0.6995	2.03	SSE
		±0.022	±0.004	±0.020	±0.003	±0.370	±0.010	±0.0004	±0.0004	±0.0009	±0.0070	±0.06	
	no α_{ml} prior	1.274	3.771	0.312	0.137	4.521	0.687	0.0399	0.0362	0.0502	0.7022	2.15	SSE
		±0.016	±0.003	±0.016	±0.002	±0.257	±0.008	±0.0007	±0.0006	±0.0011	±0.0049	±0.05	
5512589	$\boldsymbol{\alpha_{\mathrm{ml}}=1.8}$	1.106	3.756	0.408	0.216	7.843	0.706	0.0222	0.0192	0.0256	0.6620	1.80	SSE
		±0.031	±0.002	±0.008	±0.004	±0.303	±0.018	±0.0026	±0.0024	±0.0036	±0.0051		
	Gaussian	1.111	3.757	0.414	0.217	7.722	0.706	0.0223	0.0194	0.0259	0.6629	1.82	SSE
		±0.033	±0.003	±0.015	±0.004	±0.408	±0.019	±0.0026	±0.0024	±0.0036	±0.0054	±0.04	
	no α_{ml} prior	1.117	3.758	0.421	0.218	7.588	0.706	0.0225	0.0196	0.0261	0.6640	1.84	SSE
		±0.034	±0.004	±0.019	±0.004	±0.472	±0.019	±0.0026	±0.0025	±0.0037	±0.0057	±0.05	
6106415	$\alpha_{\mathrm{ml}}=1.8$	1.184	3.772	0.243	0.101	4.536	0.733	0.0236	0.0204	0.0264	0.7446	1.80	SSE
		±0.022	±0.004	±0.014	±0.003	±0.383	±0.014	±0.0023	±0.0021	±0.0029	±0.0088		
	Gaussian	1.264	3.772	0.264	0.112	4.939	0.746	0.0300	0.0265	0.0341	0.7174	2.06	NSE
		±0.012	±0.002	±0.010	±0.001	±0.170	±0.005	$\pm < 0.0001$	±0.0001	±0.0002	±0.0038	±0.05	
	no α_{ml} prior	1.267	3.774	0.271	0.112	4.922	0.747	0.0299	0.0265	0.0340	0.7163	2.10	NSE*
		±0.007	±0.002	±0.008	±0.001	±0.150	±0.005	±0.0008	±0.0007	±0.0009	±0.0029	±0.02	
6116048	$\alpha_{\mathrm{ml}}=1.8$	1.090	3.772	0.241	0.099	6.608	0.745	0.0159	0.0132	0.0166	0.7290	1.80	ASE
		±0.014	±0.003	±0.011	±0.002	±0.420	±0.009	±0.0019	±0.0017	±0.0023	±0.0073		
	Gaussian	1.066	3.763	0.200	0.097	9.328	0.743	0.0200	0.0167	0.0212	0.6747	2.01	SSE
		±0.022	±0.004	±0.019	±0.003	±0.763	±0.008	±0.0003	±0.0003	±0.0004	±0.0082	±0.04	
	no α_{ml} prior	1.082	3.770	0.230	0.099	8.650	0.742	0.0197	0.0166	0.0212	0.6789	2.12	SSE
		±0.027	±0.005	±0.024	±0.004	±0.865	±0.009	±0.0012	±0.0010	±0.0013	±0.0100	±0.06	
6603624	$\alpha_{\mathrm{ml}}=1.8$	1.052	3.735	0.029	0.067	9.830	0.700	0.0356	0.0301	0.0403	0.6625	1.80	SSE
		±0.022	±0.004	±0.015	±0.003	±0.708	±0.016	±0.0040	±0.0034	±0.0050	±0.0057		
	Gaussian	1.117	3.742	0.074	0.076	9.309	0.720	0.0373	0.0319	0.0418	0.6627	1.98	SSE*
		±0.028	±0.004	±0.017	±0.004	±0.593	±0.015	±0.0029	±0.0026	±0.0038	±0.0043	±0.05	
	no α_{ml} prior	1.192	3.751	0.129	0.086	8.321	0.744	0.0371	0.0321	0.0411	0.6687	2.16	NSE*
		±0.013	±0.005	±0.018	±0.002	±0.290	±0.005	±0.0025	±0.0021	±0.0027	±0.0016	±0.05	

Table 5.2: Same as Table 5.1 but for KIC 6933899 to KIC 10963065

Star	α prior	$M/M_\odot$	$\log T_{\rm eff}$	$\log L/L_\odot$	$\log R/R_\odot$	Age	X_0	Z_0	$Z_{\rm s}$	Z_s/X_s	$R_{\rm BCZ}$	$\alpha_{\rm ml}$	sys
6933899	$\alpha_{\rm ml}=1.8$	1.164	3.756	0.393	0.208	7.808	0.728	0.0245	0.0212	0.0275	0.6779	1.80	ASE
		±0.047	±0.004	±0.022	±0.006	±0.571	±0.018	±0.0021	±0.0020	±0.0028	±0.0154		
	Gaussian	1.140	3.760	0.401	0.205	7.806	0.717	0.0237	0.0207	0.0273	0.6650	1.90	ASE
		±0.064	±0.004	±0.027	±0.008	±0.620	±0.024	±0.0026	±0.0024	±0.0033	±0.0189	±0.05	
	no $\alpha_{\rm ml}$ prior	1.131	3.766	0.426	0.204	7.553	0.713	0.0223	0.0196	0.0260	0.6589	2.04	ASE*
		±0.053	±0.005	±0.028	±0.007	±0.574	±0.022	±0.0027	±0.0025	±0.0035	±0.0146	±0.09	
7680114	$\alpha_{\rm ml}=1.8$	1.156	3.761	0.309	0.157	6.084	0.685	0.0294	0.0259	0.0355	0.7012	1.80	SSE
		±0.025	±0.003	±0.010	±0.003	±0.544	±0.009	±0.0018	±0.0017	±0.0025	±0.0068		
	Gaussian	1.172	3.766	0.333	0.159	5.780	0.687	0.0289	0.0256	0.0351	0.7004	1.89	SSE
		±0.027	±0.004	±0.017	±0.003	±0.567	±0.010	±0.0022	±0.0020	±0.0030	±0.0076	±0.05	
	no $\alpha_{\rm ml}$ prior	1.186	3.771	0.356	0.160	5.521	0.688	0.0281	0.0251	0.0345	0.6997	1.99	SSE
		±0.033	±0.005	±0.024	±0.004	±0.721	±0.011	±0.0025	±0.0023	±0.0034	±0.0101	±0.08	
8006161	$\alpha_{\rm ml}=1.8$	1.052	3.721	−0.207	−0.022	2.714	0.696	0.0395	0.0377	0.0532	0.6891	1.80	SSE
		±0.022	±0.003	±0.010	±0.003	±0.500	±0.015	±0.0015	±0.0015	±0.0026	±0.0026		
	Gaussian	1.077	3.721	−0.201	−0.019	3.220	0.714	0.0398	0.0378	0.0519	0.6847	1.91	SSE
		±0.027	±0.003	±0.011	±0.004	±0.541	±0.020	±0.0010	±0.0009	±0.0019	±0.0037	±0.07	
	no $\alpha_{\rm ml}$ prior	1.114	3.721	−0.188	−0.013	3.896	0.741	0.0400	0.0377	0.0499	0.6791	2.10	SSE
		±0.017	±0.003	±0.011	±0.002	±0.453	±0.011	±0.0003	±0.0003	±0.0008	±0.0024	±0.06	
8228742	$\alpha_{\rm ml}=1.8$	1.214	3.762	0.518	0.260	6.584	0.740	0.0200	0.0174	0.0224	0.6906	1.80	SSE
		±0.021	±0.002	±0.002	±0.003	±0.200	±0.014	±0.0003	±0.0003	±0.0005	±0.0025		
	Gaussian	1.248	3.771	0.565	0.264	5.868	0.739	0.0199	0.0175	0.0225	0.6996	1.95	SSE
		±0.025	±0.004	±0.017	±0.003	±0.259	±0.012	±0.0006	±0.0006	±0.0008	±0.0038	±0.05	
	no $\alpha_{\rm ml}$ prior	1.274	3.778	0.596	0.266	5.479	0.740	0.0199	0.0175	0.0225	0.7047	2.05	SSE
		±0.027	±0.004	±0.017	±0.003	±0.261	±0.014	±0.0006	±0.0006	±0.0008	±0.0039	±0.06	
8379927	$\alpha_{\rm ml}=1.8$	1.253	3.774	0.184	0.068	1.513	0.749	0.0250	0.0237	0.0310	0.7638	1.80	NSE*
		±0.011	±0.001	±0.007	±0.001	±0.231	±0.003	±0.0003	±0.0003	±0.0005	±0.0039		
	Gaussian	1.258	3.778	0.201	0.068	1.511	0.748	0.0246	0.0233	0.0305	0.7651	1.86	NSE*
		±0.016	±0.004	±0.018	±0.002	±0.248	±0.004	±0.0014	±0.0014	±0.0019	±0.0044	±0.06	
	no $\alpha_{\rm ml}$ prior	1.262	3.797	0.279	0.069	1.624	0.749	0.0204	0.0191	0.0249	0.7750	2.18	NSE*
		±0.017	±0.007	±0.029	±0.002	±0.231	±0.004	±0.0014	±0.0014	±0.0019	±0.0060	±0.13	
8760414	**$\alpha_{\rm ml}=1.8$**	0.839	3.775	0.084	0.016	11.400	0.750	0.0050	0.0038	0.0046	0.7212	1.80	SSE
		±0.013	±0.002	±0.014	±0.002	±0.873	±0.002	± < 0.0001	± < 0.0001	± < 0.0001	±0.0090		
	Gaussian	0.838	3.775	0.084	0.016	11.426	0.75	0.0050	0.0038	0.0046	0.7209	1.80	SSE
		±0.013	±0.002	±0.014	±0.002	±0.886	±0.002	± < 0.0001	± < 0.0001	± < 0.0001	±0.0092	±0.02	
	no $\alpha_{\rm ml}$ prior	0.862	3.789	0.147	0.020	10.511	0.750	0.0050	0.0039	0.0048	0.7181	2.25	SSE
		±0.015	±0.008	±0.037	±0.003	±0.706	±0.001	± < 0.0001	± < 0.0001	±0.0001	±0.0049	±0.26	
10516096	$\alpha_{\rm ml}=1.8$	1.185	3.765	0.338	0.163	6.049	0.718	0.0244	0.0213	0.0280	0.7128	1.80	NSE
		±0.017	±0.002	±0.009	±0.002	±0.461	±0.016	±0.0016	±0.0016	±0.0024	±0.0060		
	Gaussian	1.210	3.772	0.374	0.166	5.854	0.730	0.0229	0.0201	0.0261	0.7160	1.92	NSE
		±0.021	±0.004	±0.020	±0.003	±0.533	±0.019	±0.0025	±0.0023	±0.0034	±0.0079	±0.06	
	no $\alpha_{\rm ml}$ prior	1.240	3.781	0.419	0.170	5.500	0.741	0.0213	0.0189	0.0243	0.7188	2.11	NSE
		±0.018	±0.005	±0.022	±0.002	±0.500	±0.013	±0.0022	±0.0020	±0.0028	±0.0072	±0.09	
10963065	$\alpha_{\rm ml}=1.8$	1.122	3.778	0.259	0.097	5.035	0.731	0.0174	0.0147	0.0189	0.7511	1.80	SSE
		±0.037	±0.005	±0.025	±0.005	±0.945	±0.018	±0.0025	±0.0023	±0.0032	±0.0167		
	Gaussian	1.094	3.777	0.248	0.094	6.139	0.730	0.0176	0.0148	0.0191	0.7254	1.92	SSE
		±0.038	±0.005	±0.026	±0.005	±1.090	±0.18	±0.0025	±0.0022	±0.0032	±0.0198	±0.07	
	no $\alpha_{\rm ml}$ prior	1.089	3.785	0.278	0.093	6.538	0.740	0.0154	0.0129	0.0165	0.7162	2.15	SSE
		±0.029	±0.005	±0.026	±0.004	±0.846	±0.012	±0.0013	±0.0012	±0.0017	±0.0135	±0.09	

Table 5.3: Same as Table 5.1 but for KIC 11244118 to Kepler 36

Star	α prior	$M/M_\odot$	$\log T_{\mathrm{eff}}$	$\log L/L_\odot$	$\log R/R_\odot$	Age	X_0	Z_0	Z_{s}	Z_s/X_s	R_{BCZ}	α_{ml}	sys
11244118	$\alpha_{\mathrm{ml}}=1.8$	1.233	3.751	0.392	0.218	7.100	0.696	0.0388	0.0345	0.0470	0.6830	1.80	SSE
		±0.053	±0.007	±0.038	±0.006	±1.232	±0.014	±0.0026	±0.0025	±0.0036	±0.0267		
	Gaussian	1.299	3.752	0.412	0.227	7.633	0.729	0.0400	0.0359	0.0471	0.6557	2.01	NSE*
		±0.004	±0.002	±0.008	±0.000	±0.222	±0.005	±0.0001	±0.0001	±0.0004	±0.0017	±0.03	
	no α_{ml} prior	1.291	3.759	0.438	0.226	6.962	0.713	0.0400	0.0360	0.0483	0.6603	2.09	NSE*
		±0.004	±0.003	±0.011	±0.000	±0.292	±0.007	± < 0.0001	±0.0001	±0.0005	±0.0021	±0.04	
11713510	**$\alpha_{\mathrm{ml}}=1.8$**	1.025	3.772	0.441	0.200	7.135	0.695	0.0139	0.0115	0.0153	0.6992	1.80	SSE
		±0.019	±0.003	±0.013	±0.003	±0.508	±0.019	±0.0022	±0.0019	±0.0028	±0.0121		
	Gaussian	1.031	3.773	0.447	0.201	7.042	0.692	0.0145	0.0121	0.0162	0.6930	1.85	SSE
		±0.022	±0.003	±0.013	±0.003	±0.424	±0.017	±0.0019	±0.0017	±0.0025	±0.0121	±0.05	
	no α_{ml} prior	1.082	3.775	0.467	0.208	6.705	0.692	0.0182	0.0157	0.0213	0.6829	1.99	SSE*
		±0.071	±0.003	±0.025	±0.010	±0.522	±0.016	±0.0057	±0.0055	±0.0077	±0.0138	±0.15	
12009504	$\alpha_{\mathrm{ml}}=1.8$	1.238	3.773	0.360	0.157	4.558	0.724	0.0239	0.0207	0.0270	0.7451	1.80	SSE
		±0.034	±0.003	±0.017	±0.004	±0.488	±0.019	±0.0021	±0.0019	±0.0028	±0.0122		
	Gaussian	1.245	3.779	0.386	0.158	4.487	0.728	0.0223	0.0195	0.0253	0.7419	1.93	SSE
		±0.028	±0.004	±0.018	±0.003	±0.480	±0.019	±0.0025	±0.0022	±0.0033	±0.0112	±0.05	
	no α_{ml} prior	1.253	3.786	0.416	0.159	4.332	0.731	0.0205	0.0180	0.0234	0.7410	2.07	SSE
		±0.026	±0.005	±0.021	±0.003	±0.367	±0.016	±0.0016	±0.0014	±0.0021	±0.0103	±0.08	
12258514	$\alpha_{\mathrm{ml}}=1.8$	1.250	3.769	0.440	0.206	5.564	0.724	0.0256	0.0224	0.0294	0.7291	1.80	SSE
		±0.039	±0.004	±0.020	±0.004	±0.939	±0.018	±0.0021	±0.0019	±0.0028	±0.0137		
	Gaussian	1.227	3.769	0.436	0.204	6.342	0.729	0.0255	0.0225	0.0293	0.7086	1.90	SSE
		±0.038	±0.003	±0.017	±0.004	±0.823	±0.019	±0.0024	±0.0023	±0.0035	±0.0138	±0.04	
	no α_{ml} prior	1.217	3.771	0.445	0.203	6.445	0.731	0.0242	0.0213	0.0278	0.7046	1.96	SSE
		±0.041	±0.004	±0.020	±0.005	±0.701	±0.018	±0.0029	±0.0027	±0.0039	±0.0134	±0.06	
16CygA	$\alpha_{\mathrm{ml}}=1.8$	1.054	3.762	0.173	0.086	6.441	0.684	0.0250	0.0214	0.0291	0.7027	1.80	SSE
		±0.010	±0.001	±0.006	±0.001	±0.363	±0.006	± < 0.0001	±0.0002	±0.0004	±0.0036		
	Gaussian	1.095	3.765	0.196	0.092	7.055	0.692	0.0281	0.0247	0.0337	0.6730	2.13	SSE
		±0.016	±0.005	±0.023	±0.002	±0.375	±0.012	±0.0024	±0.0021	±0.0033	±0.0056	±0.06	
	no α_{ml} prior	1.114	3.771	0.225	0.095	6.647	0.706	0.0250	0.0220	0.0295	0.6795	2.20	SSE
		±0.009	±0.001	±0.004	±0.001	±0.206	±0.006	±0.0003	±0.0002	±0.0005	±0.0015	±0.01	
16CygB	$\alpha_{\mathrm{ml}}=1.8$	1.007	3.758	0.070	0.043	6.464	0.681	0.0247	0.0214	0.0294	0.6986	1.80	SSE
		±0.006	±0.002	±0.007	±0.001	±0.250	±0.004	±0.0012	±0.0010	±0.0015	±0.0035		
	Gaussian	1.023	3.762	0.091	0.045	6.532	0.686	0.0250	0.0217	0.0296	0.6942	1.92	SSE
		±0.013	±0.002	±0.010	±0.002	±0.281	±0.007	±0.0001	±0.0001	±0.0003	±0.0034	±0.04	
	no α_{ml} prior	1.076	3.764	0.116	0.054	9.279	0.741	0.0250	0.0214	0.0274	0.6621	2.40	SSE
		±0.012	±0.002	±0.009	±0.002	±0.473	±0.005	±0.0001	±0.0001	±0.0003	±0.0035	±0.00	
Kepler36	**$\alpha_{\mathrm{ml}}=1.8$**	1.113	3.771	0.475	0.220	6.923	0.729	0.0150	0.0124	0.0159	0.7059	1.80	NSE
		±0.035	±0.003	±0.015	±0.005	±0.372	±0.018	±0.0004	±0.0004	±0.0006	±0.0121		
	Gaussian	1.118	3.771	0.480	0.221	6.870	0.730	0.0150	0.0125	0.0159	0.7058	1.82	NSE
		±0.035	±0.003	±0.017	±0.005	±0.386	±0.018	±0.0004	±0.0004	±0.0006	±0.0122	±0.04	
	no α_{ml} prior	1.123	3.773	0.486	0.222	6.792	0.731	0.0150	0.0125	0.0160	0.7058	1.85	NSE
		±0.036	±0.004	±0.021	±0.005	±0.409	±0.018	±0.0004	±0.0004	±0.0006	±0.0122	±0.06	

Chapter 6

Concluding Remarks

Given that a wealth of new asteroseismic data and information has only become accessible very recently, it should not be surprising that we are discovering discrepancies between the data and our models. The prevalent way of performing asteroseismology is to employ a specific set of models (or rather a particular stellar evolution and pulsation code) and a very specific tool set in order to arrive at some conclusions about the stellar parameters, the stellar structure, and the general applicability of the theory of stellar evolution. In this respect, what was presented in this thesis might appear as just another adaptation of this approach. It is certainly correct that all the results given in this thesis are based on one stellar evolution code and one pulsation code alone.

The important breakthrough of the approach developed in this thesis, however, is the application of probabilistic inference to the comparison of observations and models. Chapter 2 described how it allows in principle to consistently compare many different stellar evolutionary models and evolutionary codes in one common hypothesis space. Furthermore, the Bayesian approach is uniquely modular, so that every layer of the stellar models can be studied in detail, using the same observed data and the same formalism. This is not only constrained to actual asteroseismic studies where, e.g., Kepler data is used to obtain results for actual stars. Instead, researchers could compare their codes in a new way. They could agree to use the exact same fundamental physics (equation of state, opacities, ...) in their otherwise different model toolset. The Bayesian formalism would then give them a clear answer as two whether they also obtain the same results at every layer of their model hierarchy when compared to some input data. Many different applications for theoretical modellers can be envisioned (e.g., hare-and-hound exercises), which might help to

study and better understand the systematic differences between the stellar evolution codes by using clearly defined probabilistic terms such as the evidence and posterior probability. The primary goal of the new method, however, is naturally to perform probabilistic inference and learn something about the real stars.

Here, the new method is uniquely able to make clear statements about all the stellar evolutionary models, systematic error models, or adapted prior assumption sets, that were used in the inference. It would certainly be a great advantage if, instead of merely comparing χ^2 values between different researchers, their different codes and assumptions could actually be evaluated in a common hypothesis space. A comparison that uses prior information about the stellar parameters from, e.g., spectroscopy, could immediately reveal which codes are best calibrated for the specific object that was studied.

The new method can also be used to answer very specific questions about specific objects. This follows from the fact that Bayesian analysis always evaluates specific propositions[1]. For example, the revised solar chemical compositions has been discredited for years based on specific fits of the helioseismic observables to the contending models. This is in conflict with the self-consistent and expansive probabilistic analysis of the problem described in Chapter 4. Thanks to the modular nature of Bayesian analysis, the new analysis method answers the question whether helioseismology actually has anything to say about the preference for either the old or the revised chemical composition. The answer, at least based on the larger grid of models that we have constructed, appears to be "no". There seem to be more fundamental problems with the solar model calibration that have to be solved first. Only then can the chemical composition really be tackled from the perspective of helioseismology. Aside from this example, many different variations of such investigations can be envisioned (e.g., testing different models for convection).

---

[1]Furthermore, the propositions can be constructed with the product and sum rules, which can be used to formulate specific compound propositions (such as: "this frequency is affected by the surface effect or not" or "this frequency is due to an $l = 0$ mode or due to an $l = 2$ mode") which correctly propagate into the posterior probabilities.

However, there also lies a danger in using the Bayesian approach. First off, both the prior information and the likelihood of the data has a big impact on the posterior probabilities calculated with Bayes' theorem. It is therefore important to make sure that what is used on their behalf makes sense. It is also the responsibility of the Bayesian analyst to clearly state which assumptions are made, and also in how far they are potentially aribtrary, so that the results can be reproduced and also properly assessed. More importantly, however, the answers obtained with this formalism are always related to a specific set of pre-defined propositions. In other words, if the same propositions (models, hypotheses) are always used to evaluate new data, there is no mechanism intrinsic to the posterior probability that tells us that "these models are wrong". In that respect, it is important to remember Thomas Kuhn's terminology and assessment once again. If we cling to the paradigms (and to our legacy codes) even in the presence of evidence against their validity, there is no mechanism in science that guarantees that other paradigms are getting their fair chance to overthrow the current one. At least, the Bayesian approach has the potential to reveal whether the paradigms we are clinging to are really applicable. For this to work, however, our analyses have to be as inclusive as possible. Then, if prior information and newly obtained information are in stark conflict, new explanations for what is observed can take the stage. On the other hand, clinging to current models that are no longer applicable is sometimes necessary as well. Fudging our paradigms to accommodate such problems is easily hidden in the traditional approach by performing "corrections" (e.g., such as a surface correction that yields reassuring χ^2 values). In the Bayesian approach, at least, these fudge parameters stand out in the list of propositions as part of the toolset that we apply to what we observe in nature.

To quote Kuhn in "The structure of scientific revolutions" directly:

Scientists work from models acquired through education and through subsequent exposure to the literature often without quite knowing or needing to know what characteristics have given these models the status of community paradigms. [..] The coherence displayed by the research tradition in

162

which they participate may not imply even the existence of an underlying body of rules and assumptions that additional historical or philosophical investigation might uncover. [..] Paradigms may be prior to, more binding, and more complete than any set of rules for research that could be unequivocally abstracted from them.

As physicists, our tools are quantitative in nature. Is Bayesian analysis a first step to perform a partial quantitative analysis of our paradigms, or is it perhaps just another paradigm itself that will be overthrown? Whether or not this is true, we should not forget that the paradigms are there, that they determine our results, and that at least in science they seem to never hold.

Acknowledgments

Bibliography

Kuhn, T. S. 1970, The structure of scientific revolutions (University of Chicago Press)

Appendix A

Curriculum Vitae

Michael Gruberbauer

Contact

Email: mgruberbauer@ap.smu.ca

Personal Data

Date of birth: November 24, 1983
Place of birth: St. Pölten, Austria
Citizenship: Austrian

Education

09/2009 – 07/2013   **Saint Mary's University**, Halifax, Canada. PhD in Astronomy; Thesis title (under supervision of Prof. David B. Guenther): *Bayesian Asteroseismology.*

10/2006 – 12/2008   **University of Vienna**, Vienna, Austria. MSc in Astronomy with honors; Thesis title (under supervision of Prof. Werner W. Weiss): *Asteroseismic modelling of the roAp star gamma Equulei based on observations obtained by the MOST satellite.*

10/2003 – 10/2006   **University of Vienna**, Vienna, Austria. BSc in Astronomy, minor in Meteorology; Thesis title (under supervision of Prof. Werner W. Weiss): *The MOST stray light behaviour.*

06/2002	**High School** Diploma with honors; BORG Krems, Krems, Lower Austria, Austria.

Employment

09/2012 – 12/2012	**Lecturer**, Saint Mary's University, Halifax, Canada.
09/2009 – 12/2009	**Teaching Assistant**, Saint Mary's University, Halifax, Canada.
03/2009 – 06/2009	**Project Scientist (part time)**, University of Vienna, Austria.
2007 – 2008	**Technical and Research Assistant (full time position)**, University of Vienna, Austria.
2005 – 2007	**Graduate Student Research Assistant**, University of Vienna, Austria
2002 – 2003	**Paramedic**, Civilian service at the Austrian Red Cross, Lower Austria, Austria.

Scholarships

2010 – 2013	**Vanier Canada Graduate Scholarship**
2005 – 2007	**Scholarship for Excellent Academic Achievements** (*"Leistungsstipendium"*), University of Vienna

Refereed Publications

Number of peer-reviewed publications as of 06/2013: **43** (**8** as first author)
For an up-to-date list, please query the NASA ADS System
(http://adsabs.harvard.edu) or go to tinyurl.com/8lg3d39

First-author Publications:

Gruberbauer M., Guenther D.B.: *Bayesian seismology of the Sun*, 2013, Monthly Notices of the Royal Astronomical Society, 432, 417

Gruberbauer M., Guenther D.B., Kallinger T.: *Toward a New Kind of Asteroseismic Grid Fitting*, 2012, The Astrophysical Journal, 749, 109

Gruberbauer M. et al.: *MOST observations of the roAp stars HD 9289, HD 99563, and HD 134214*, 2011, Astronomy & Astrophysics, 530, 135

Gruberbauer M. et al.: *On the detection of Lorentzian profiles in a power spectrum: a Bayesian approach using ignorance priors*, Astronomy & Astrophysics, 2009, Astronomy & Astrophysics, 506, 1043

Gruberbauer M. et al.: *MOST photometry and modeling of the rapidly oscillating (roAp) star γ Equ*, 2008, Astronomy & Astrophysics, 480, 223

Gruberbauer M., Weiss W.W..: *On the benefits of photometry for roAp theory: MOST observations of γ Equulei (HD 201601)*, 2008, Contributions of the Astronomical Observatory Skalnate Pleso, 38, 311

Gruberbauer M., Neuteufel R., Weiss W.W.: *Observing γ Doradus Stars with BRITE - an Outlook*, 2008, Communications in Asteroseismology, 152, 116

Gruberbauer M. et al.: *Photometry of the RRd Lyrae Variable AQ Leo: Two Radial Modes, 32 Combination Frequencies, and Beyond*, 2007, Monthly Notices of the Royal Astronomical Society, 379, 1498

Co-author Publications:

Karoff C. et al.: *Observations of Intensity Fluctuations Attributed to Granulation and Faculae on Sun-like Stars from the Kepler Mission*, 2013, The Astrophysical Journal, 767, 34

Zwintz K. et al.: γ *Doradus pulsation in two pre-main sequence stars discovered by CoRoT*, 2013, Astronomy & Astrophysics, 550, 121

Appourchaux T. et al.: *Oscillation mode frequencies of 61 main-sequence and subgiant stars observed by Kepler*, Astronomy & Astrophysics, 2012, Astronomy & Astrophysics, 543, 54

White T.R. et al.: *Solving the Mode Identification Problem in Asteroseismology of F Stars Observed with Kepler*, 2012, The Astrophysical Journal, 751L, 36

Kallinger T. et al.: *Evolutionary influences on the structure of red-giant acoustic oscillation spectra from 600d of Kepler observations*, Astronomy & Astrophysics, 2012, Astronomy & Astrophysics, 541, 51

Grupe D. et al.: *A Remarkable Long-term Light Curve and Deep, Low-state Spectroscopy: Swift and XMM-Newton Monitoring of the NLS1 Galaxy Mkn 335*, 2012, The Astrophysical Journal Supplement, 199, 28

Saio H. et al.: *Pulsation models for the roAp star HD 134214*, 2012, Monthly Notices of the Royal Astronomical Society, 420, 283

Beck P.G. et al.: *Fast core rotation in red-giant stars as revealed by gravity-dominated mixed modes*, 2012, Nature, 481, 55

Appourchaux T. et al.: *Oscillation mode linewidths of main-sequence and subgiant stars observed with Kepler*, Astronomy & Astrophysics, 2012, Astronomy & Astrophysics, 537, 134

Uytterhoeven K. et al.: *The Kepler characterization of the variability among A- and F-type stars. I. General overview*, 2011, Astronomy & Astrophysics, 534, 125

Kurtz D. et al.: *The first evidence for multiple pulsation axes: a new rapidly oscillating Ap star in the Kepler field, KIC 10195926*, 2011, Monthly Notices of the Royal Astronomical Society, 414, 2550

Balona L. et al.: *Rotation and oblique pulsation in Kepler observations of the roAp star KIC 10483436*, 2011, Monthly Notices of the Royal Astronomical Society, 413, 2651

Rucinski S. et al.: *"MOST" Satellite Photometry of Regulus*, 2011, Information Bulletin on Variable Stars, 5987, 1

Huber D. et al.: *Solar-like Oscillations and Activity in Procyon: A Comparison of the 2007 MOST and Ground-based Radial Velocity Campaigns*, 2011, The Astrophysical Journal, 731, 94

Basu S. et al.: *Sounding Open Clusters: Asteroseismic Constraints from Kepler on the Properties of NGC 6791 and NGC 6819*, 2011, The Astrophysical Journal, 729L, 10

Zwintz K. et al.: *Pulsational Analysis of V 588 Mon and V 589 Mon Observed with the MOST and CoRoT Satellites*, 2011, The Astrophysical Journal, 729, 20

Balona L. et al.: *Kepler observations of rapidly oscillating Ap, δ Scuti and γ Doradus pulsations in Ap stars*, 2011, Monthly Notices of the Royal Astronomical Society, 410, 517

Karoff C. et al.: *Asteroseismology of solar-type stars with Kepler I: Data analysis*, 2010, Astronomische Nachrichten, 331, 972

Kallinger T. et al.: *Asteroseismology of red giants from the first four months of Kepler data: Fundamental stellar parameters*, 2010, Astronomy & Astrophysics, 552, 1

Bedding T. et al.: *Solar-like Oscillations in Low-luminosity Red Giants: First Results from Kepler*, 2010, The Astrophysical Journal, 713L, 176

Chaplin W.J. et al.: *The Asteroseismic Potential of Kepler: First Results for Solar-Type Stars*, 2010, The Astrophysical Journal, 713L, 169

Kallinger T. et al.: *The nature of p-modes and granulation in HD 49933 observed by CoRoT*, 2010, Astronomy & Astrophysics, 510, 106

Kallinger T. et al.: *Oscillating red giants in the CoRoT exofield: asteroseismic mass and radius determination*, 2010, Astronomy & Astrophysics, 509, 77

Lüftinger T. et al.: *Surface structure of the CoRoT CP2 target star HD 50773*, 2010, Astronomy & Astrophysics, 509, 43

Huber D. et al.: *A search for p-modes and other variability in the binary system 85 Pegasi using MOST photometry*, 2009, Astronomy & Astrophysics, 505, 715

Gruber D. et al.: *SPB guide star photometry with MOST*, 2009, Communications in Asteroseismology, 158, 217

Zwintz K. et al.: *MOST photometry of the enigmatic PMS pulsator HD 142666*, 2009, Astronomy & Astrophysics, 494, 1031

Sachkov M. et al.: *On the excited mode stability in the roAp star γ Equ*, 2008, Communications in Asteroseismology, 157, 363

Guenther D.B. et al.: *Asteroseismology of sun-like stars with MOST*, 2008, Communications in Asteroseismology, 157, 64

Guenther D.B. et al.: *The Nature of p-modes and Granulation in Procyon: New MOST Photometry and New Yale Convection Models*, 2008, The Astrophysical Journal, 687, 1448

Reegen P. et al.: *CINDERELLA: Comparison of INDEpendent RELative least-squares Amplitudes. Times series data reduction in Fourier space*, 2008, Astronomy & Astrophysics, 484, 601

Huber D. et al.: *MOST photometry of the roAp star 10 Aquilae*, 2008, Astronomy & Astrophysics, 483, 239

Huber D. et al.: *MOST and 10 Aql*, 2008, Contributions of the Astronomical Observatory Skalnate Pleso, 38, 413

Aurière M. et al.: *Weak magnetic fields in Ap/Bp stars. Evidence for a dipole field lower limit and a tentative interpretation of the magnetic dichotomy*, 2007, Astronomy & Astrophysics, 475, 1053

Reegen P. et al.: *Reduction of time-resolved space-based CCD photometry developed for MOST Fabry Imaging data*, 2006, Monthly Notices of the Royal Astronomical Society, 367, 1417

Talks & Poster presentations

06/2012	**Presentations** at the MOST Science Team Meeting, University of Vienna, Austria.
06/2012	**Poster** at the 5th Kepler Asteroseismic Science Consortium (KASC) Workshop, Balatonalmadi, Hungary.
06/2012	**Presentations** at the MOST (Microvariability and Oscillations of STars) Science Team Meeting, University of Calgary, Canada.
05/2012	**Invited Talk** at the Department Colloquium of the High Altitude Observatory, Boulder, USA.
12/2011	**Presentations** at the MOST Science Team Meeting, University of Vienna, Austria.
05/2011	**Presentations** at the MOST Science Team Meeting, University of Western Ontario, London, Canada.
05/2011	**Contributed Talk** at the "1st Workshop on Pulsations in Pre-Main Sequence Stars and their Impact on Stellar Evolution", University of Vienna, Austria.
12/2010	**Presentations** at the MOST Science Team Meeting, Warzaw, Poland.
06/2010	**Contributed Talk** at the "3rd Kepler Asteroseismic Consortium (KASC) Workshop", Aarhus University, Denmark.

05/2010	**Presentations** at the MOST Science Team Meeting, Saint Mary's University, Halifax, Canada.
05/2010	**Poster** at the CASCA Meeting 2010, Saint Mary's University, Halifax, Canada.
10/2009	**Contributed Talk** at the "2nd Halifax Meeting on Computational Astrophysics", Saint Mary's University, Halifax, Canada.
07/2008	**Presentations** at the MOST Science Team Meeting, Institute of Astronomy, Vienna, Austria.
07/2008	**Contributed Talk** at the "2nd BRITE (BRight Target Explorer: http://www.brite-constellation.at) Workshop", University of Vienna, Austria.
09/2007	**Contributed Talk** at the "CP#AP Workshop", University of Vienna, Austria.
07/2007	**Contributed Talk & Poster** at the conference "Stellar Pulsation & Cycles of Discovery: New Tools, New Variables, New Astrophysics", University of British Columbia, Vancouver, Canada.
07/2007	**Presentations** at the MOST Science Team Meeting, University of British Columbia, Vancouver, Canada.
05/2007	**Contributed Talk** at the "BRITE Workshop", University of Vienna, Austria.
12/2005	**Presentations** at the MOST Science Team Meeting, University of Vienna, Austria.

Referee for

Astronomy & Astrophysics

Communications in Asteroseismology