



Mid-Course Assessment

# Hierarchically Modelling Stars Using Deep Learning and Asteroseismology

By

Alexander J. Lyttle

*Student ID* 1532473  
*Supervisor* Dr Guy R. Davies  
*Co-Supervisor* Dr Andrea Miglio

Solar and Stellar Physics Group  
School of Physics and Astronomy  
College of Engineering and Physical Sciences  
University of Birmingham

September 24, 2020

© Copyright by ALEXANDER J. LYTTLE, 2020  
All Rights Reserved

## ABSTRACT

Determining stellar parameters using high-precision asteroseismic constraints has improved our understanding of exoplanetary systems and galactic history. With such advances comes scrutiny of systematic uncertainties in our stellar models. In this report, I present a new method for determining fundamental stellar properties using a hierarchical Bayesian model. The hierarchical model enables the addition of model parameters, to address assumptions usually made when modelling stars, by encoding population-level information into a sample of stars. I present the first application of this method to a sample of low-mass dwarf stars ( $0.8$  to  $1.2 M_{\odot}$ ) and use it to infer the population distribution of helium ( $Y$ ) and mixing-length theory parameter ( $\alpha_{\text{mlt}}$ ). I find that the hierarchical model can obtain stellar ages, masses and radii to good precision despite the addition of uncertainty from  $Y$  and  $\alpha_{\text{mlt}}$ . Finally, I discuss a possible extension to the model which makes use of asteroseismic signatures of helium abundance to further constrain helium.

## **ACKNOWLEDGMENTS**

Thank you to my primary supervisor, Dr Guy Davies, for his support during the the first year of my PhD. Many thanks to my co-supervisor, Dr Andrea Miglio, for his additional comments and support. I would also like to thank Dr Warrick Ball for providing open-source code and Dr Tanda Li for computing stellar models, both used for Figures in this work.

# Contents

	<b>Page</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Hierarchical Bayesian Models . . . . .	3
1.2 Modelling a Star . . . . .	9
1.3 Asteroseismology of Solar-Like Oscillators . . . . .	13
1.4 Sampling Stellar Models . . . . .	16
1.5 Observing Stars . . . . .	18
1.5.1 Photometry & Astrometry . . . . .	18
1.5.2 Spectroscopy . . . . .	19
1.5.3 Detecting Asteroseismic Oscillation Modes . . . . .	20
<b>2 Hierarchically Modelling Many Stars</b>	<b>24</b>
<b>3 Future Work</b>	<b>25</b>
3.1 Including an Asteroseismic Signature of Helium . . . . .	25
<b>References</b>	<b>27</b>
<b>Appendix A Accompanying Paper</b>	<b>33</b>

# List of Figures

1.1 Luminosity against true ages of a fake stellar cluster. The true luminosities lie on the red line and the observed luminosities (black) have been artificially scattered by $0.05 L_{\odot}$ .	7
1.2 The $z$ -score, $(\bar{\tau} - \tau_{\text{true}})/s_{\tau}$ , where $\bar{\tau}$ and $s_{\tau}$ are the respective sample mean and standard deviation of the posterior ages from each of the no- and partially-pooled models.	8
1.3 Standard deviations, $s_{\tau}$ of the age posteriors from both the no- and partially-pooled models.	9
1.4 A series of stellar evolutionary tracks plot on a Hertzsprung-Russell diagram, starting at the ZAMS and evolved through the MS (black line) to the post MS (dashed black line) until the stellar surface gravity $\log g = 3.6$ .	12
1.5 Spherical harmonic modes of oscillation for various combinations of angular degree ( $l$ ) and azimuthal order ( $m$ ).	15
1.6 The asteroseismic wave fronts in a typical stellar interior. The blue, orange, green, red and purple lines represent the paths of oscillations with angular degree, $l = 1, 2, 3, 25, 75$ respectively. The dashed lines represent the depth probed by each mode.	16
1.7 The frequency-power spectrum for a red giant star.	21
1.8 The signal-to-noise (SNR) power spectrum for a red giant star with 100 random samples from the posterior locations of the $l = 0, 2$ oscillation mode pairs.	22
1.9 An echelle diagram for a red giant star with the locations of the radial ( $l = 0$ ) and quadrupole ( $l = 2$ ) oscillation modes.	23

# **List of Tables**

# Chapter 1

## Introduction

Historically, studying stars has been an integral part in our quest to uncover the secrets of the universe. Ever since our understanding of stellar structure and evolution progressed beyond that of a big ball of self-contained gas, following the pioneering works of Eddington (1926), we have studied stars to deepen our understanding in fields from particle physics to galactic evolution and cosmology.

Asteroseismolgy – the study of the oscillations of stars – has provided us with a new way to infer stellar properties and test fundamental physics. In a similar way to a seismologist studying earthquakes to probe the interior of the Earth, an asteroseismologist can infer stellar structure by measuring the frequencies of standing waves trapped within.

In recent years, asteroseismology has provided tight constraints on fundamental stellar properties such as age, mass and radius through direct scaling with the Sun or comparisons with theoretical stellar models. In the wake of missions such as *CoRoT*, *Kepler* and *TESS* (ongoing), asteroseismology has lead to large catalogues of precise fundamental properties for red giant (e.g. Pinsonneault et al., 2014; Serenelli et al., 2017) enabling better characterisation of exoplanetary systems and the chronology of the Milky Way. However, with the increased observational precision

from asteroseismology, the uncertainties in our theoretical stellar models are becoming more important.

The theory of stellar structure and evolution is not yet fully understood. We rely upon a series of assumptions often calibrated to the Sun or other well-studied stars. However, such assumptions carry with them a degree of uncertainty which propagates systematics on ages and masses determined by stellar modelling. For example, including the physics of heavy-element diffusion in low-mass stars ( $0.8 \lesssim M \lesssim 1.2 M_{\odot}$ ) can impact the efficiency of nuclear burning in its core and change the chemical composition at the surface with time. In heavier stars with a convective hydrogen-burning core ( $M \gtrsim 1.2 M_{\odot}$ ) the overshooting of material at the core boundary can also affect the predicted stellar age, yet this process is not fully understood.

Further uncertainty arises from our assumptions about the initial chemical composition of the star when it was formed, a requirement for an evolutionary stellar model. Helium ionisation lines are not measurable in the atmospheres of low-mass stars, making it difficult assume their initial chemical composition. Traditionally, a linear helium enrichment law is assumed which relates the helium abundance ( $Y$ ) to the measurable heavy-element abundance ( $Z$ ) through a ratio,  $\Delta Y / \Delta Z$  calibrated to the Sun and the a primordial abundance,  $Y_P$  (see e.g. Chiosi and Matteucci, 1982). However, overestimating the helium abundances can underestimate age determinations.

To make matters worse, inaccuracies in the mixing-length theory approximation of convection (parameterised by  $\alpha_{\text{mlt}}$ ) can cause further uncertainty. Individually modelling  $Y$  and  $\alpha_{\text{mlt}}$  is a computationally expensive process and our observational uncertainties are not good enough to individually constrain them in the majority of stars.

In this chapter, I will first introduce a new method for improving the inference of stellar parameters through the use of a hierarchical Bayesian model (HBM). This is a statistical model which allows us to encode prior information about the distribution of a stellar parameter in a population into the star-by-star analysis. In Section 1.1 I will describe the mathematics of an HBM

and give a simplified example. Then, I will review the basics of the evolution of low-mass stars in Section 1.2 to show how we can predict observables from models of stellar evolution.

In Section 1.3, I introduce the theory of asteroseismology and show how the detection of stellar oscillations can probe the structure of a star. Once we have a way of predicting stellar observables, I describe a few methods for comparing our models to observations in Section 1.4. Here, I introduce the need for using machine learning to speed-up the process and increase the scalability of our HBM. Finally, in Section 1.5, I briefly describe the conventional methods used to infer stellar properties by observation in order to understand the contents of the following chapters. This include an example of observational asteroseismology through the use of an open source code to which I have contributed.

In Chapter 2, I will introduce my paper which applies the methods in this chapter to a sample of low-mass, asteroseismic stars already studied by Serenelli et al. (2017). Here, I include  $Y$  and  $\alpha_{\text{mlt}}$  as additional parameters in our statistical model with population-level priors introduced by the HBM. I will show that HBMs can improve the precision of stellar fundamental parameters while introducing free parameters which account for uncertainties in our model physics.

In Chapter 3, I will propose an extension to the method which introduces a new observable to improve the inference of helium abundance in stars.

## 1.1 Hierarchical Bayesian Models

In this section, I begin by describing a Bayesian probabilistic model for a star. Then, I show how we can extend the model to an HBM which introduces parameters to describe a population of stars. I demonstrate, with a simple example that such a method can improve the inference of fundamental parameters.

Consider a model for a single star comprising a set of independent parameters,  $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^{N_\theta}$  which makes a set of predictions,  $\boldsymbol{\mu} = \mathbf{f}(\boldsymbol{\theta})$ , where  $\boldsymbol{\mu} = \{\mu_j\}_{j=1}^{N_y}$  and  $N_y$  is the number of observables. Using Bayes' theorem, we may write the *posterior* probability density function (PDF) of the model given a set of observations  $\mathbf{y}$  as,

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}), \quad (1.1)$$

where  $p(\mathbf{y}|\boldsymbol{\theta})$  is the *likelihood* of the data given the model and  $p(\boldsymbol{\theta})$  is the *a priori* PDF of the model parameters.

Assuming our observations of  $\mathbf{y}$  are uncorrelated and subjected to random, Gaussian noise with a known standard deviation,  $\sigma$ , we may write the likelihood function as a normal distribution,

$$p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{j=1}^{N_y} \frac{1}{\sigma_j \sqrt{2\pi}} \exp \left[ -\frac{(y_j - \mu_j)^2}{2\sigma_j^2} \right], \quad (1.2)$$

$$\equiv \prod_{j=1}^{N_y} \mathcal{N}(y_j | \mu_j, \sigma_j), \quad (1.3)$$

where  $\mathcal{N}(x|\mu, \sigma)$  is a normal distribution for  $x$  with a mean,  $\mu$  and standard deviation  $\sigma$ .

The prior PDF of the model is  $p(\boldsymbol{\theta}) = \prod_i p(\theta_i)$  where  $p(\theta_i)$  describes the prior distribution of the parameter  $\theta_i$ . The prior PDF could range from uninformative (flat) to informative (e.g. a normal distribution with a mean and standard deviation representative of our current knowledge of the parameter). Encoding our prior understanding of the model this way is useful for improving our inference. For example, we have independent prior knowledge that the age of the universe is  $\sim 14$  Gyr (Bennett et al., 2013; Planck Collaboration et al., 2016). Hence, we may choose to give the age parameter for a stellar model a weakly informative, uniform distribution prior from 0 to 14 Gyr. This would prevent our posterior from being influenced by unphysical ages.

Once we have the posterior, we can determine the marginalised posterior distribution of an individual parameter by integrating over all other parameters. For example, the marginalised

posterior for  $\theta_1$  is,

$$p(\theta_1 | \mathbf{y}) = \int_{-\infty}^{+\infty} p(\boldsymbol{\theta} | \mathbf{y}) d\theta_2 \dots d\theta_{N_\theta}. \quad (1.4)$$

Therefore, we end up with a distribution which describes the probability of  $\theta_1$  given  $\mathbf{y}$  which takes into account the distribution (or uncertainty) of all other parameters in the model.

Let us now consider modelling a population of  $N_{\text{stars}}$  similar stars. We could combine the posteriors for each star to get a posterior for the population of stars like so,

$$p(\boldsymbol{\Theta} | \mathbf{Y}) = \prod_{k=1}^{N_{\text{stars}}} p(\boldsymbol{\theta}_k | \mathbf{y}_k), \quad (1.5)$$

where  $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^{N_{\text{stars}}}$  and  $\mathbf{Y} = \{\mathbf{y}_k\}_{k=1}^{N_{\text{stars}}}$  are the matrices of model parameters and observations for the entire population. We refer to this as a *no-pooled* model because no information is shared between the objects.

However, what if we wish to model the distribution of a particular parameter  $\theta_i$  in the population? For example, if we have prior knowledge that stars in a stellar cluster formed at roughly the same time, but we don't know its age, we might want to encode such information into the model to influence individual stellar ages. One method would be to independently model the stars in the cluster and then find their mean and standard deviation in age. However, the individual ages will have been modelled under the assumption that they are uncorrelated with other stars in the cluster, but we have reason to believe otherwise. Alternatively, we can incorporate the assumption that stars in a cluster formed at the same time in one of two ways. The first is to *partially-pool* and the second is to *max-pool* the stellar ages respectively. The former assumes the stellar parameters are drawn from some common distribution, and the latter is the special case where all stellar parameters share the exactly same value in the population.

We refer to models which pool parameters in this way as hierarchical (or multilevel) model. For a hierarchical model, we describe the distribution of  $\boldsymbol{\Theta}$  in the population by a set of *hyper-*

parameters,  $\boldsymbol{\phi} = \{\phi_l\}_{l=1}^{N_\phi}$ . Bayes' equation now becomes,

$$p(\boldsymbol{\phi}, \boldsymbol{\Theta} | \mathbf{Y}) \propto p(\mathbf{Y} | \boldsymbol{\Theta}) p(\boldsymbol{\Theta} | \boldsymbol{\phi}) p(\boldsymbol{\phi}) \quad (1.6)$$

where  $p(\mathbf{Y} | \boldsymbol{\Theta})$  is the likelihood as before,  $p(\boldsymbol{\phi})$  the prior probability of the hyper-parameters. The probability of  $\boldsymbol{\Theta}$  given  $\boldsymbol{\phi}$  is,

$$p(\boldsymbol{\Theta} | \boldsymbol{\phi}) = \prod_{k=1}^{N_{\text{stars}}} d(\theta_k | \boldsymbol{\phi}), \quad (1.7)$$

where  $d(\theta_k | \boldsymbol{\phi})$  is some chosen distribution from which the parameters for a given star are drawn from the population.

Let us consider a simple model which predicts the luminosities,  $\mathbf{L}$  from the ages,  $\boldsymbol{\tau}$  of  $N_{\text{stars}} = 1000$  stars in a cluster formed at roughly the same time. Modelling the population independently, we get the posterior,

$$p(\boldsymbol{\tau} | \mathbf{L}) \propto \prod_{k=1}^{1000} p(L_k | \tau_k) p(\tau_k). \quad (1.8)$$

Now, let us consider a partially-pooled model where the stellar ages are drawn from a normal distribution centred on a mean,  $\mu_\tau$  and standard deviation,  $\sigma_\tau$ . This is appropriate for a stellar cluster because we expect the star formation to occur at some mean age over a characteristic timescale. The posterior now becomes,

$$p(\mu_\tau, \sigma_\tau, \boldsymbol{\tau} | \mathbf{L}) \propto p(\mathbf{L} | \boldsymbol{\tau}) p(\boldsymbol{\tau} | \mu_\tau, \sigma_\tau) p(\mu_\tau, \sigma_\tau), \quad (1.9)$$

where,

$$p(\boldsymbol{\tau} | \mu_\tau, \sigma_\tau) = \prod_{k=1}^{1000} \mathcal{N}(\tau_k | \mu_\tau, \sigma_\tau). \quad (1.10)$$

There is no known analytical or empirical relation between the age of a star and its luminosity, but for the purposes of this example let us say that we know  $L \propto \tau^2$ . I generated 1000 stellar ages from a normal distribution with a mean of 5 Gyr and a standard deviation of 0.05 Gyr, and computed their luminosities using this relation. Then, I added Gaussian noise to the luminosities

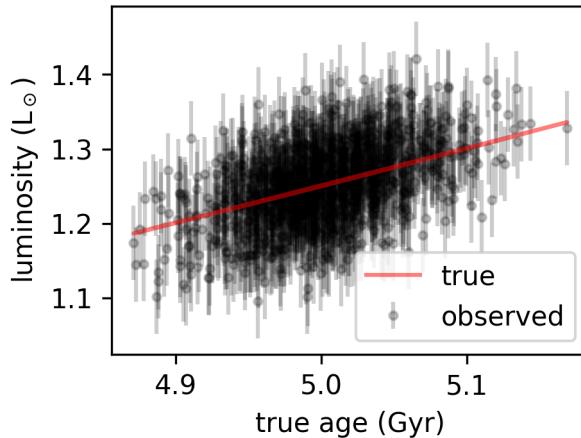


Figure 1.1: Luminosity against true ages of a fake stellar cluster. The true luminosities lie on the red line and the observed luminosities (black) have been artificially scattered by  $0.05 L_\odot$ .

with a standard deviation of  $0.05 L_\odot$  and proceeded to model the stellar ages using Equations 1.8 and 1.9 with the Bayesian package PyMC3 (Salvatier, Wiecki, and Fonnesbeck, 2016). The observed and true luminosities are plotted against the true ages in Figure 1.1.

If we wished to determine spread of stellar ages in the cluster using the no-pooled model, we might naïvely calculate a standard deviation from the resulting stellar ages. However, this overestimates the true standard deviation, getting 0.109 Gyr rather than 0.05 Gyr, because it includes the uncertainty in the individual ages. When we model the population mean and spread in the hierarchical model we get  $\mu_\tau = 5.002 \pm 0.003$  Gyr and  $\sigma_\tau = 0.042 \pm 0.007$  Gyr which are within  $< 2\sigma$  of the truths. In addition, the pooled model reduced individual uncertainties on ages over those of the no-pooled model.

Both models can accurately determine ages, but the hierarchical model returns more precise ages, assuming our prior assumptions are true. Figure 1.2 shows that the  $z$ -score for ages from both models match a normal distribution with a mean of 0 and standard deviation of 1, indicating the individual stellar ages and uncertainties are accurate. In the top left panel, we see so-called *shrinkage* of the ages. This is the tendency for the ages to be pulled towards the population mean in

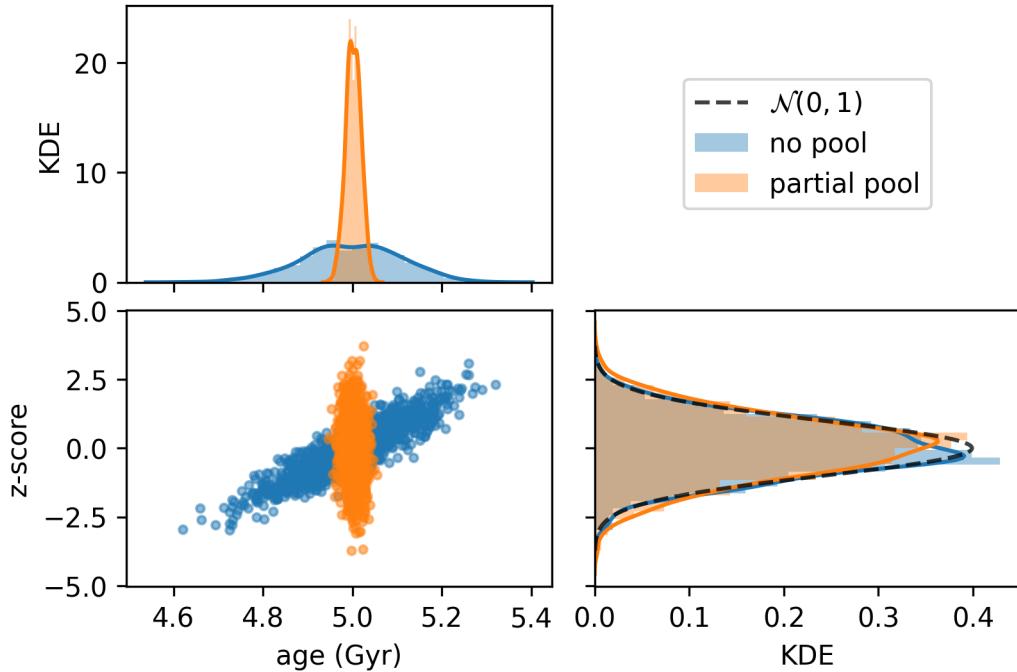


Figure 1.2: The  $z$ -score,  $(\bar{\tau} - \tau_{\text{true}})/s_{\tau}$ , where  $\bar{\tau}$  and  $s_{\tau}$  are the respective sample mean and standard deviation of the posterior ages from each of the no- and partially-pooled models.

a hierarchical model. In the bottom left panel we also see that the no-pooled model under-predicts low ages and over-predicts high ages. The hierarchical manages to separate population spread from individual uncertainties, all but eliminating this correlation.

Crucially, the partially pooled model produces more than doubly precise ages, as shown in Figure 1.3, because the model takes into account the population mean and spread as hyperparameters. The theoretical limit for this uncertainty reduction goes as  $1/\sqrt{N_{\text{stars}}}$ . With our sample of 1000 stars, the age uncertainties are about 0.04 Gyr. Therefore, a sample of 100,000 could reduce the uncertainties by up to 0.004 Gyr. With statistical uncertainties that small, we can start to confidently test the systematics of our models and confidence in our data.

If we wish to improve the precision of fundamental stellar parameters, using hierarchical models to encode our prior knowledge is essential. However, modelling stars is not as simple, nor

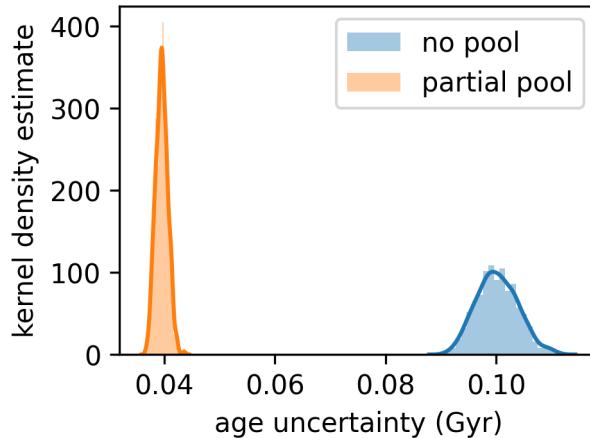


Figure 1.3: Standard deviations,  $s_\tau$  of the age posteriors from both the no- and partially-pooled models.

analytical as in the example above. Before we can statistically model a population of stars, we must have a way of generating stellar observables from fundamental parameters such as age and mass. In the next section, I give an overview of how we numerically model stellar observables and why traditional methods pose new problems when adapting the above model.

## 1.2 Modelling a Star

Modelling a star is a complicated process with no simple law which describes the evolution of measurable quantities as a function of age and initial bulk properties. Following the development of quantum theories of ionization and radiation, the theory of the stellar structure and evolution was able to progress from the assumption that stars were a ball of uniform, ideal gas Eddington (1926). Stellar structure could be simplified to the combined effects of a few differential equations.

The first is the equation of hydrostatic equilibrium which states that the outward pressure within the star acts to oppose the inward gravitational force. This may be written as,

$$\frac{\partial P(r)}{\partial r} = -\frac{G\rho(r)m(r)}{r^2}, \quad (1.11)$$

where  $P(r)$  is the pressure at a given radius  $r$  within the star,  $G$  is the gravitational constant,  $\rho(r)$  is the density at  $r$  and  $m(r)$  is the mass contained within  $r$ . Combined with equations for the conservation of mass, energy loss and internal energy transport, we can describe the structure of a star.

Integrating Equation 1.11 with respect to radius from the centre to the boundary  $R$  (where  $P(R) = 0$ ) leads to the virial theorem,

$$\Omega = -3(\gamma - 1)U, \quad (1.12)$$

where  $\Omega$  is the gravitational potential energy of the star,  $\gamma$  is the adiabatic exponent which relates the pressure of the star to the internal energy density,  $u$  ( $P = (\gamma - 1)u$ ) and  $U$  is the total internal energy of the star. We recall the virial theorem here to demonstrate that if a star contracts, its potential energy decreases which in turn increases its internal energy and flux at the surface.

Internal energy transport within a star is either radiative, convective, or conductive in the case of white dwarfs and neutron stars (Yakovlev and Urpin, 1980). Energy transportation is important when modelling stars. However, convective energy transfer is notoriously difficult to model. When solving for the evolution of stellar structure, we often use approximations of convection such as the mixing-length-theory (Böhm-Vitense, 1958; Gough, 1977). The mixing-length theory characterises the typical distance over which a blob of stellar material moves before dispersing as the fraction  $\alpha_{\text{mlt}}$  of the pressure scale height  $H_P = -P(\text{d}r/\text{d}P)$ , where  $\alpha_{\text{mlt}}$  is of order unity. Since the mixing-length theory is an approximation of convection, its accuracy varies between stellar models. Usually, the value of  $\alpha_{\text{mlt}}$  is calibrated to the Sun. However, studies of 2D (Ludwig, Freytag, and Steffen, 1999) and 3D (Trampedach et al., 2014) hydrodynamical simulations calibrate  $\alpha_{\text{mlt}}$  in the range  $\sim 1.5$  to  $2.5$  for different stars similar in mass to the Sun.

When a star of similar mass to the Sun ( $0.8$  to  $1.2 M_\odot$ ) begins its life, the conditions in its centre quickly become high enough in temperature and pressure to fuse hydrogen. We call this evolutionary phase the zero-age main sequence (ZAMS). Throughout its main sequence (MS)

lifetime, it burns hydrogen to produce helium via the proton-proton (p-p) chain reaction,



Stars in this mass range typically burn hydrogen in a radiative core surrounded by a convective envelope. Radiative and convective regions are characterised by the dominant process of energy transport in each region.

Figure 1.4 shows the evolutionary tracks near-solar-mass stars on a Hertzsprung-Russell diagram (HRD). As a star evolves through the MS, where it is commonly referred to as a *dwarf*, the mean molecular weight of its core increases as hydrogen is being converted to helium. Thus, the core contracts and heats up in accordance with the virial theorem, increasing the nuclear reaction rate and photon flux (or luminosity,  $L$ ) of the star. Towards the end of the MS lifetime, the convective envelope expands, cooling the temperature at the surface of the star (effective temperature,  $T_{\text{eff}}$ ). Once the star has extinguished available hydrogen in the core, its luminosity decreases and the core contracts because it is no longer supported by nuclear burning. In turn, the convective envelope continues to expand and the effective temperature of the star decreases, while the temperature near the contracting core increases due to the virial theorem. During this phase, the growing star is often referred to as a *subgiant*. Once conditions in a shell at the boundary of the core are sufficient, hydrogen begins to fuse once more, increasing the luminosity of the star as it evolves from a *subgiant* into a *red giant*.

Today, we are able to evolve stars to predict their surface quantities such as  $L$  and  $T_{\text{eff}}$  using stellar evolutionary codes. Early development of such codes began in the 1960s (see e.g. Kippenhahn, Weigert, and Hofmeister, 1967). Later, stellar models were being computed in large ranges of masses and chemical composition to fit isochrones (tracks at a constant age) to observations of Galactic open clusters (Vandenberg, 1985). Today, there are many codes available for scientists to evolve stellar models. In the case of 1D, non-rotating models, there are for example: ASTEC (Christensen-Dalsgaard, 2008), CESAM2k (Morel and Lebreton, 2008), GARSTEC (Weiss and

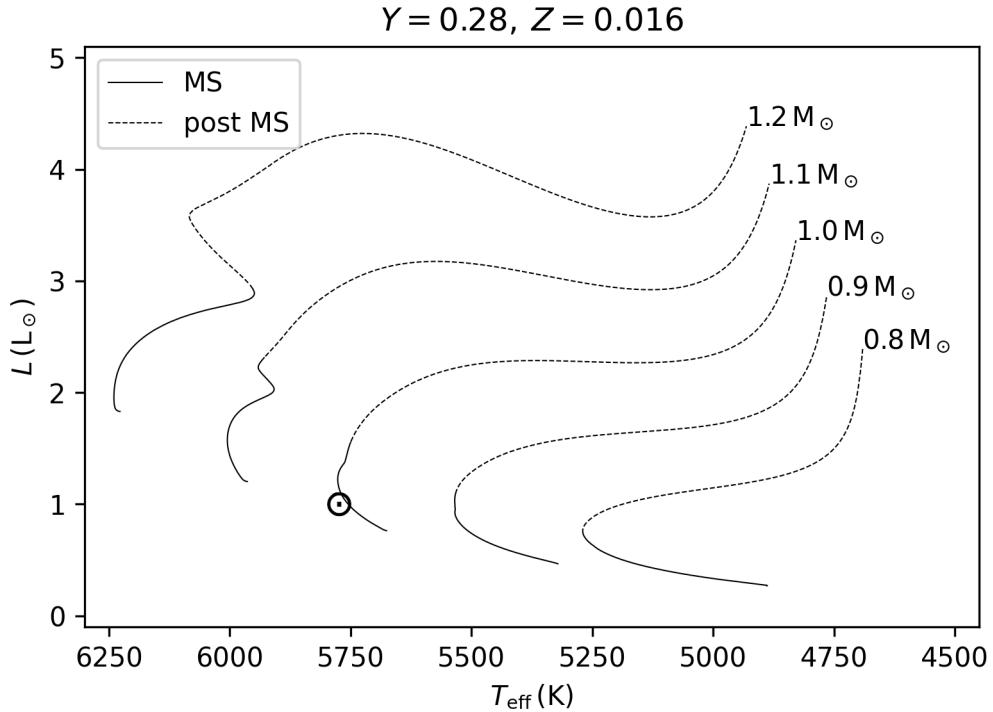


Figure 1.4: A series of stellar evolutionary tracks plot on a Hertzsprung-Russell diagram, starting at the ZAMS and evolved through the MS (black line) to the post MS (dashed black line) until the stellar surface gravity  $\log g = 3.6$ .

Schlattl, 2008) and MESA (Paxton et al., 2011). For 2D and 3D hydrodynamics, current codes include: MUSIC (Baraffe et al., 2017) for better modelling of convective mixing and 2DStars (Halabi et al., 2017) for rotating stars. Other 2D and 3D codes also exist for modelling short-timescale events such as rapidly rotation in stars (Roxburgh, 2004).

In this work, we use an open-source 1D stellar evolutionary code known as Modules for Experiments in Stellar Astrophysics (MESA; Paxton et al., 2011). The code evolves a star given a set of initial conditions over dynamically assigned time steps, producing models for the internal stellar structure and a summary of the state of the star at each step. It achieves this by dividing a 1D slice of the star (assuming spherical symmetry) into a mesh of points at which it numerically integrates the differential equations of state.

Example inputs to MESA include: the mass,  $M$ , the mixing-length theory parameter,  $\alpha_{\text{mlt}}$ , and the fractional composition of hydrogen, helium and other heavy elements given by  $X$ ,  $Y$  and  $Z$  respectively, where  $X + Y + Z = 1$ . If the macroscopic diffusion of heavy elements is considered, the chemical composition at the surface of the star will change with time.

## 1.3 Asteroseismology of Solar-Like Oscillators

For over a century, we have been able to map stars based on their photometric magnitude and spectroscopic colour with HRDs. Coupling such observational data with measurements of interstellar distances using parallax, we were able to determine stellar luminosities. The unique structure of early HRDs eluded to the idea that stars evolve over time. With the addition of nuclear physics, theories of stellar evolution could be put to the test. However, while we could only observe stellar surface properties, many modelling mysteries about the internal structure of stars would be left unsolved.

Until the last few decades, our understanding of stellar structure has been all but skin deep. In the 1960s, observations of 5-minute brightness fluctuations in the solar photosphere lead to the study of stochastically driven acoustic waves trapped beneath the surface of the Sun (Ulrich, 1970; Ando and Osaki, 1975). Later named helioseismology (Deubner and Gough, 1984), the study of oscillation modes allowed for further insights into the solar interior, such as rotation (Deubner, Ulrich, and Rhodes, 1979) and the solar neutrino production (Bahcall and Ulrich, 1988). In tandem with this research was the emergence of asteroseismology – the study of stars through their oscillation frequencies (Christensen-Dalsgaard, 1984). Asteroseismolgy has since been used to improve inference on the fundamental properties of stars (see, e.g. Ulrich, 1986; Soderblom, 2010; Silva Aguirre et al., 2015).

Solar-like oscillators are stars which typically exhibit two kinds of standing waves: acoustic

oscillation modes (or p modes) excited stochastically by convection in their outer layers and restored by pressure gradients, and internal gravity waves (or g modes) which are controlled by buoyancy in the stellar core. This work focuses on main sequence stars for which p modes are only present in their observed spectra. In this section I will summarise the theory behind acoustic waves present in dwarfs.

The theory which predicts the locations of the asteroseismic oscillation modes has its roots in the spherical harmonic oscillator. The eigenfrequencies,  $\nu_{nlm}$  are categorised into modes of radial order,  $n$ , angular degree,  $l$  and azimuthal order,  $m$ . The radial order represents the overtone number ( $\nu_n = n\nu_0$ ).

Figure 1.5 shows spherical oscillation modes for different  $l$  and  $m$ . Small changes in stellar radius manifest as small changes in luminosity, arising from the virial theorem. If we imagine the combination of such oscillations when viewing a star as a point source, we can imagine how higher  $l$  would be harder to detect due to cancellation across the stellar surface. Typically, we are able to resolve  $l \lesssim 3$  in the frequency spectra of stars for which we are unable to resolve the surface.

Figure 1.6 shows the paths of wave fronts travelling through the stellar interior at different  $l$ . The waves curve due to the changing sound speed profile inside the star. We can see how waves with difference  $l$  probe different depths of the stellar interior, given by the dashed lines. In the case of the Sun, where modes of high  $l$  are able to be resolved, this has allowed us to understand its density profile at shallower depths and, for example, measure its chemical composition (Basu and Antia, 2004).

We characterise the first three angular degrees as radial,  $l = 0$ , dipolar,  $l = 1$  and quadrupoloar,  $l = 2$  modes. Observable p modes in solar-like oscillators are typically at high  $n$  and low  $l$ , where asymptotic theory may be applied,  $l/n \rightarrow 0$  (Tassoul, 1980). Therefore, to first order in  $\Delta\nu$ , we can approximate the frequency for a given as follows (Christensen-Dalsgaard,

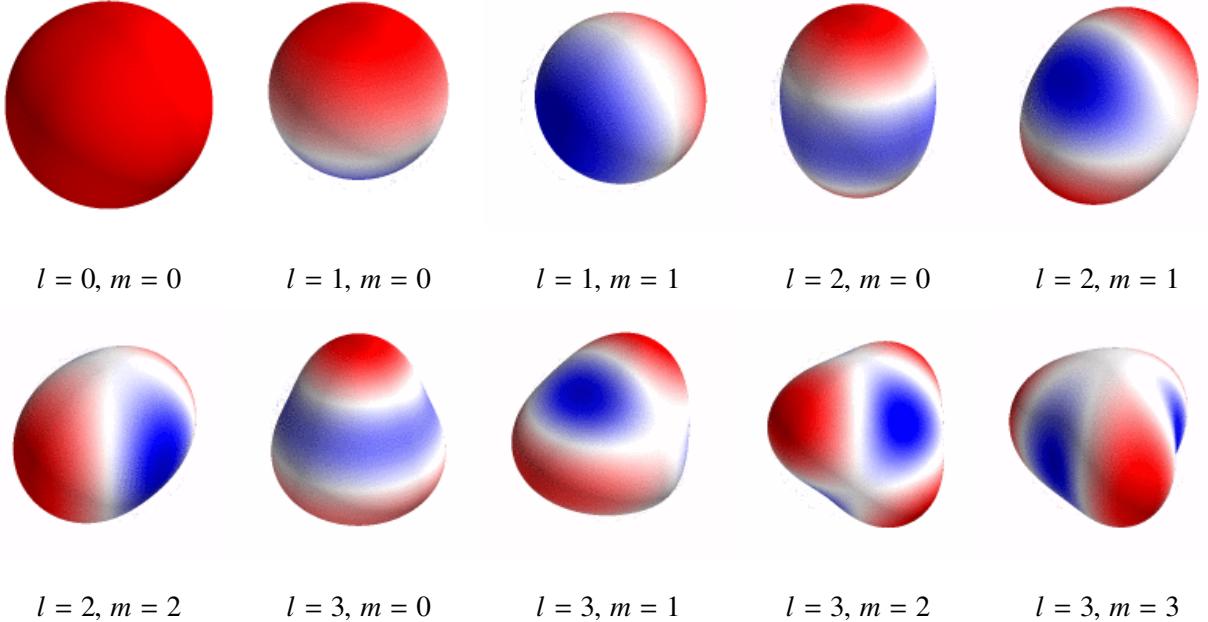


Figure 1.5: Spherical harmonic modes of oscillation for various combinations of angular degree ( $l$ ) and azimuthal order ( $m$ ).

1984),

$$\nu_{nl} \simeq \Delta\nu \left( n + \frac{l}{2} + \epsilon \right) \quad (1.14)$$

where  $\Delta\nu$  is proportional to the inverse of the sound travel time over the stellar diameter,

$$\Delta\nu = \left( 2 \int_0^R \frac{dr}{c(r)} \right)^{-1} \quad (1.15)$$

where the speed of sound  $c$  is a function of stellar radii,  $r$ . The large frequency separation,  $\Delta\nu$  is approximately the frequency difference between consecutive modes of the same  $l$ . Equation 1.14 also implies that consecutive modes of odd and even degree should be separated by  $\sim \Delta\nu/2$ . Second-order deviations from Equation 1.14 can be shown to describe the small frequency spacing,  $\delta\nu_{l,l+2}$ . In other words, we expect to find even modes clustered together, and likewise for odd modes. In practice, as shown in Section 1.5.3, we find pairs of radial and quadrupole modes separated by  $\sim \Delta\nu$  and dipole modes in between them.

We detect excited oscillation modes in an excess of power centred on the frequency at

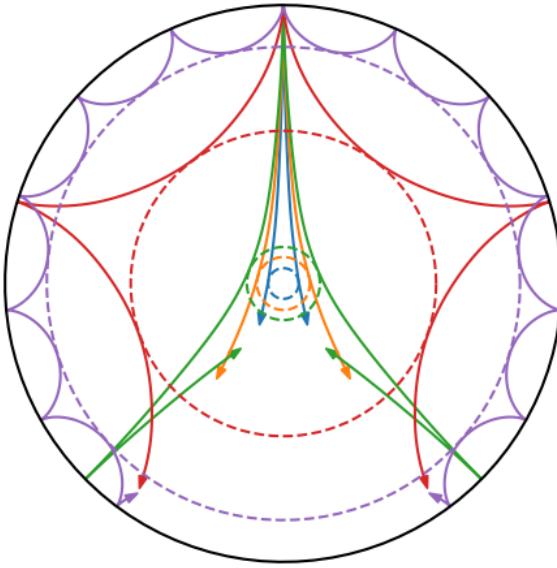


Figure 1.6: The asteroseismic wave fronts in a typical stellar interior. The blue, orange, green, red and purple lines represent the paths of oscillations with angular degree,  $l = 1, 2, 3, 25, 75$  respectively. The dashed lines represent the depth probed by each mode.

maximum power,  $\nu_{\max}$ . By characterising the behaviour of the waves close to the stellar surface, it can be shown that  $\nu_{\max} \propto g T_{\text{eff}}$  where  $g$  is the surface gravity of the star (Kjeldsen and Bedding, 1995).

We can determine theoretical oscillation modes by solving for the stellar profiles produced in the evolutionary models discussed in Section 1.2. One such code build to predict oscillation modes for a given stellar structure is GYRE (Townsend and Teitler, 2013).

## 1.4 Sampling Stellar Models

Stellar evolution and asteroseismic models can predict observables given the age and initial conditions (e.g. mass, chemical composition and other model physics) for a given star. To implement the

HBM described in Section 1.1 we need a way to generate predictions from our stellar models. There are generally two methods for achieving this: produce a discrete grid of models and evaluate the likelihood at each point, or interpolate a grid of models and continuously sample an approximation of the grid. It is difficult to implement an HBM using either of these methods, because we need it to scale well with input dimensions and number of stars in a population.

Grid-based modelling involves producing a large grid of stellar models with a range of input parameters. Then, the likelihood that a star is described by a point on the grid may be evaluated. One example of a method which does this is the BAyesian STellar Algorithm (BASTA; Silva Aguirre et al., 2015). However, using a discrete grid can lead to inaccurate model posteriors limited by the grid spacing. Computing finer grids is also computationally expensive.

Alternatively, we can interpolate the grid of models. This is common in the isochrone fitting method (see e.g. Berger et al., 2020). Here, a grid of models in 3 dimensions (mass, metallicity and age) may be easily interpolated to produce an approximation of the stellar models, continuously mapping inputs to observables. However, interpolation does not scale well with input dimensionality and size of the grid. In order to take full advantage of the HBM, we need the flexibility to expand the input dimensions to encompass extra model physics, such as the mixing-length theory parameter,  $\alpha_{\text{mlt}}$ . We need a faster way of approximating the grid of stellar models.

One alternative to interpolation is to use machine learning. We can train an artificial neural network on the grid of models. This is the method used in Chapter 2 and is explained in more detail in the attached paper (Appendix A). In summary, the neural network is made up of many layers of trainable weights which are optimised to minimise the error between predictions and the training data. Advantageously, it is easy to evaluate the gradient of the likelihood with a neural network, because the gradient is required to optimise the weights during training. Therefore, approximating the models this way allows us to sample the posterior of the HBM using modern algorithms such as Hamiltonian Monte Carlo (HMC) (see e.g. the No-U-Turn Sampler Homan and Gelman, 2014).

## 1.5 Observing Stars

In this section, I briefly recall the fundamentals of observational astronomy. Firstly, I describe a method for determining luminosity from photometric and astrometric measurements. Then, I show how spectroscopy can determine the chemical composition and temperature at the surface of a star. Finally, I show how we determine the asteroseismic oscillation modes from photometric time series measurements of a solar-like oscillator.

### 1.5.1 Photometry & Astrometry

Photometry is the measure of photon flux from an object. In stellar astronomy, we can use photometry to determine the luminosity,  $L$ , of a star and measure short-timescale variations in flux due to stellar activity and pulsations. We measure the flux from an object in a series of passbands. For example, if we measure the photometric flux,  $F$ , from a star in the blue ( $B$ ) and visual ( $V$ ) bands then we which we can determine its apparent visual magnitude,  $m_V$ ,

$$m_V = -2.5 \log \left( \frac{F_V}{F_{V,0}} \right), \quad (1.16)$$

where  $F_{V,0}$  is the reference flux at a magnitude of zero, and its colour,  $m_B - m_V$ . We can use colour to determine effective temperature,  $T_{\text{eff}}$ , but an explanation of this is beyond the scope of this work.

Astrometry is the measure of the position and movement of a star across the sky. We can measure the position of a star against a distant background throughout the year to determine its parallax ( $\varpi$ ) – the apparent difference in angular position of a star in the sky between observations separated by 1 astronomical unit (AU<sup>1</sup>). Thus, we can then derive the distance,  $d$  to the star assuming the small-angle formula,  $d \approx 1/\varpi$ . We typically measure  $\varpi$  in arc seconds (" or 'as'), which yields a distance in parsecs (pc) – the distance from the Sun to an object which has appeared to move 1" over a baseline of 1 AU.

---

<sup>1</sup>The average distance between the Earth and the Sun, 1 AU  $\approx 1.5 \times 10^8$  km.

If we wish to determine the luminosity of a star (photometric flux at the stellar surface), we need the distance to the star which we can determine using astrometry. We first calibrate the absolute magnitude,  $\mathcal{M}$  of the star to get its flux at a distance of 10 pc. For example, in the  $V$ -band,

$$\mathcal{M}_V = m_V - 5 \log(d) + 5 + A_V, \quad (1.17)$$

where  $A_V$  is the extinction – an estimate of the light absorbed by interstellar dust between the star and the observer. However, to get to luminosity, we need the total flux across the entire electromagnetic spectrum. We can estimate this by calculating a bolometric correction ( $BC$ ) which can be determined from models of the stellar atmosphere. The bolometric magnitude,  $\mathcal{M}_{\text{bol}}$ ,

$$\mathcal{M}_{\text{bol}} = \mathcal{M}_V - BC, \quad (1.18)$$

is related to the luminosity of the star,

$$\log\left(\frac{L}{L_\odot}\right) = \frac{2}{5} (\mathcal{M}_{\text{bol},\odot} - \mathcal{M}_{\text{bol}}), \quad (1.19)$$

in terms of the solar luminosity,  $L_\odot$ .

### 1.5.2 Spectroscopy

Stellar spectroscopy is the measure and analysis of the electromagnetic spectra of stars. We can use this to determine the abundance of elements in the stellar atmosphere and the effective temperature,  $T_{\text{eff}}$ .

The abundances of elements are typically reported as the ratio of the element to a readily abundant element. For example, the iron-to-hydrogen abundance is given as  $[\text{Fe}/\text{H}]$  and is the logarithmic difference between the fractional abundance in the stellar spectra with that of the Sun. We commonly refer to the metallicity as the relative abundance of all elements except for H and He,

$$[\text{M}/\text{H}] = \log(Z/X) - \log(Z/X)_\odot, \quad (1.20)$$

where  $(Z/X)_{\odot}$  is the relative abundance of heavy-elements to hydrogen in the Sun. The solar value has been revised downward over recent years, but is  $(Z/X)_{\odot} \approx 0.02$  (Grevesse and Sauval, 1998; Asplund, Grevesse, and Sauval, 2005; Asplund, Grevesse, Sauval, and Scott, 2009)

The stellar electromagnetic spectrum can also be used to determine its surface temperature,  $T_{\text{eff}}$  which is fundamentally related to the luminosity of a star of radius ( $R$ ),  $L \propto R^2 T_{\text{eff}}^4$ . We can determine  $T_{\text{eff}}$  assuming the star as an ideal black body using Wein's law,

$$\lambda_{\max} T_{\text{eff}} \approx 2.9 \times 10^{-3} \text{ m K}, \quad (1.21)$$

where  $\lambda_{\max}$  is the wavelength at maximum power.

We have shown the basic theory behind determining temperature and abundances from spectroscopy, but there are a whole lot more to the process which goes beyond the scope of this work. A continuously updated source of abundances and parameters is the APOGEE stellar parameters and chemical abundances pipeline (ASPCAP; García Pérez et al., 2016) of the Sloan Digital Sky Survey (SDSS; Blanton et al., 2017).

### 1.5.3 Detecting Asteroseismic Oscillation Modes

We can measure the asteroseismic frequencies to higher precision than traditional stellar parameters such as  $L$  and  $T_{\text{eff}}$ , allowing for better comparison with stellar models.

In order to characterise the asteroseismic oscillations of a star, we need measurements of its photometric time series. Several missions from *Kepler* (Borucki et al., 2010) to more recently, TESS (Ricker et al., 2015) have provided time series data. We can extract the frequency-power spectrum from the time series using the Lomb-Scargle method (Lomb, 1976; Scargle, 1982). Figure 1.7 shows an example of an asteroseismic power spectrum for TIC 38828538. There is a power excess at around 200  $\mu\text{Hz}$  which contains the excited oscillation modes. These are a series of peaks

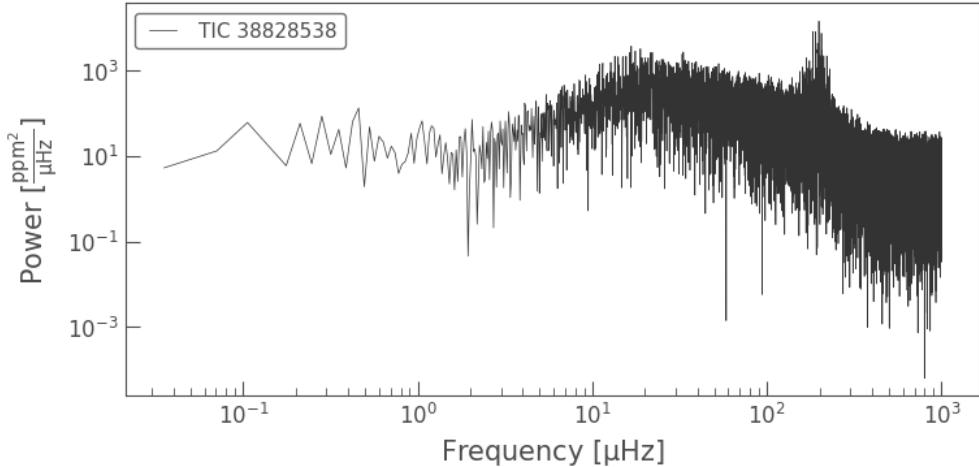


Figure 1.7: The frequency-power spectrum for a red giant star.

within the power element which correspond to frequencies,  $\nu_{nl}$  of different radial order ( $n$ ) and angular degree ( $l$ ).

Next, we divide the power by an estimate of the background noise to get the signal-to-noise (SNR) spectrum. Figure 1.8 spectrum for the star around the frequency of maximum power,  $\nu_{\max} \approx 200 \mu\text{Hz}$ . We can see the excited oscillation modes more closely here. Highlighted in red are the radial and quadrupole mode pairs, separated by a small frequency determined by the second order term omitted from Equation 1.14. The spacing between consecutive modes of the same angular degree is  $\approx \Delta\nu$ .

In between the radial and quadrupole mode pairs are the dipole modes. The dipole modes are harder to identify in more evolved stars like this red giant, because they can mix with the g modes present in the core of the star. Where the  $l = 0, 2$  modes are regularly spaced, the  $l = 1$  modes are more scattered throughout the spectrum.

In Figure 1.9 we see the SNR spectrum plot on an echelle diagram. An echelle diagram plots the frequency,  $\nu$  against  $\nu \text{mod} \Delta\nu$  to show ridges corresponding to each angular degree. We can see the regular spacing of the radial and quadrupole modes forms ridges to the left of the plot,

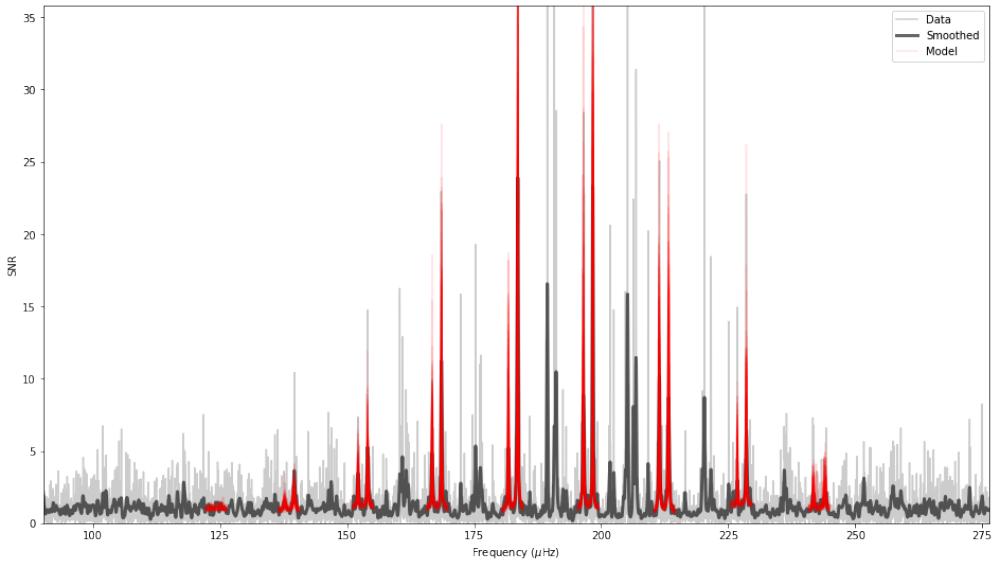


Figure 1.8: The signal-to-noise (SNR) power spectrum for a red giant star with 100 random samples from the posterior locations of the  $l = 0, 2$  oscillation mode pairs.

but the dipole modes are scattered to the right. Predicting the locations of dipole modes which are mixed in this way can be difficult, but can be used to probe the structure at the core of the star (see e.g. Mosser, Gehan, et al., 2018). However, the radial and quadrupole modes are still sufficient for determining average global stellar properties from  $\Delta\nu$  and for comparison with stellar models.

There are several methods which exist to extract the radial and quadrupole oscillation modes from the asteroseismic power spectra (see e.g. Mosser, Belkacem, et al., 2011; Appourchaux et al., 2012; Davies et al., 2016). This process is commonly referred to as *peakbagging*. Here, we use the Python package `PBjam` (Nielsen et al., in preparation) as an example. `PBjam` is an automated peakbagging pipeline which uses a population prior comprising thousands of already peakbagged stars to predict the locations of the radial modes. Once located, it fits a form of the asymptotic relation in Equation 1.14 to the SNR spectrum to get global properties such as  $\Delta\nu$  and  $\nu_{\max}$  to locate the mode locations. Finally, it drops all prior information and fits a Lorentzian to each mode individually to determine their central frequency.

The results of peakbagging are shown in both Figure 1.8 and 1.9. In the former, I show

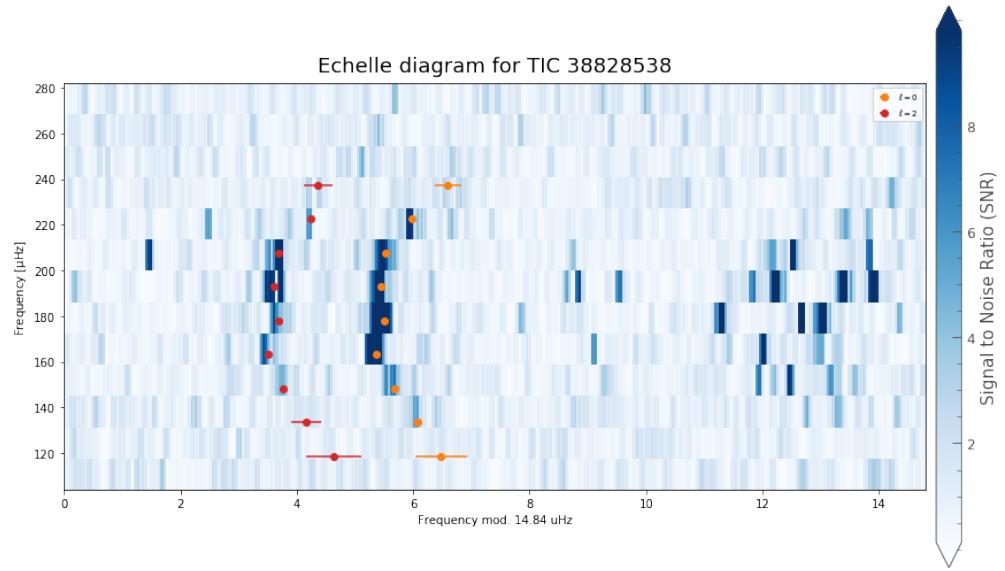


Figure 1.9: An echelle diagram for a red giant star with the locations of the radial ( $l = 0$ ) and quadrupole ( $l = 2$ ) oscillation modes.

random samples drawn from the posterior mode fits and in the latter, I show the mean and standard deviations of the posterior mode centres.

# Chapter 2

## Hierarchically Modelling Many Stars

I have demonstrated how an HBM can improve inference with a simple example in Section 1.1. In Section 1.2 I showed how we can use stellar evolutionary models to predict observables, and with the addition of asteroseismology (Section 1.3) improve constraints on the internal structure and average properties of a star. Now, I will show how these methods can be combined to hierarchically model many stars.

The paper accompanying this report (Lytle et al., in preparation) is attached in Appendix A. The paper in its current state is being prepared for submission, awaiting feedback and discussion from co-authors. In this paper, I report the first application of our new method for hierarchically modelling many stars to a sample of dwarfs and subgiants observed with *Kepler*.

# Chapter 3

## Future Work

I have demonstrated in Chapter 2 that our method works on a sample of previously studied stars, returning high-precision stellar parameters in-line with prior work (Serenelli et al., 2017). In this chapter, I will briefly discuss future extensions to the method.

### 3.1 Including an Asteroseismic Signature of Helium

We could extend our model to include an independent measure of helium abundance in the outer regions of the star. Since helium lines are not present in the spectra of low-mass stellar atmospheres, we must look to other indicators of helium abundance. One such approach utilises asteroseismology. Stellar oscillations can provide a probe of rapid variation in stellar structure via a periodic signature in the p mode separations (Broomhall et al., 2014). This has been used to study the base of the convective zone (Monteiro, Christensen-Dalsgaard, and Thompson, 2000) and crucially, the boundary of the second helium ( $\text{He II}$ ) ionization zone (Houdek and Gough, 2007). More recently, there have been attempts to measure helium abundance in cool stars (Mazumdar et al., 2014; Corsaro, De Ridder, and García, 2015; Verma, Raodeo, Antia, et al., 2017).

The boundary between the first and second helium ionization zones induces a peak in the adiabatic index (and sound-speed profile), particularly in the convective regions of sufficiently cool, low-mass stars. This causes an acoustic glitch in the oscillation modes,  $\nu$  – a small deviation from the normal pattern in a star without ionization. The amplitude of this glitch depends on the amount of helium present.

Verma, Raodeo, Basu, et al. (2019) characterises the deviation from the norm due to the He II ionization zone with the following equation based on the works of (Houdek and Gough, 2007),

$$\delta\nu_{\text{He}} = A_{\text{He}}\nu e^{-8\pi^2\Delta_{\text{He}}^2\nu^2} \sin(4\pi\tau_{\text{He}}\nu + \psi_{\text{He}}), \quad (3.1)$$

where  $A_{\text{He}}$  relates to the area beneath the peak in the adiabatic index and  $\Delta_{\text{He}}$  describes its width. The amplitude of the glitch depends on  $\nu$ , but it has been shown that an average amplitude across all observed frequencies,  $\langle A_\nu \rangle$ , can be a good indicator of helium abundance.

In a future extension to the HBM method detailed in this report, I will include  $\langle A_\nu \rangle$  as a nmodel output, calculated using our models of stellar structure from MESA. Then, either using published values of the glitch amplitude for the sample of *Kepler* dwarfs studied in this work (e.g. Verma, Raodeo, Antia, et al., 2017) or by performing my own analysis, I will introduce it as a new observable.

# References

- Ando, H. and Y. Osaki (1975). “Nonadiabatic Nonradial Oscillations - an Application to the Five-Minute Oscillation of the Sun”. In: *Publications of the Astronomical Society of Japan* 27, pp. 581–603.
- Appourchaux, T. et al. (2012). “Oscillation Mode Frequencies of 61 Main-Sequence and Subgiant Stars Observed by Kepler”. In: *Astronomy and Astrophysics* 543, A54. doi: [10.1051/0004-6361/201218948](https://doi.org/10.1051/0004-6361/201218948).
- Asplund, M., N. Grevesse, and A. J. Sauval (2005). “The Solar Chemical Composition”. In: *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis*. Ed. by III Barnes Thomas G. and Frank N. Bash. Vol. 336. Astronomical Society of the Pacific Conference Series, p. 25.
- Asplund, Martin, Nicolas Grevesse, A. Jacques Sauval, and Pat Scott (2009). “The Chemical Composition of the Sun”. In: *ARA&A* 47, pp. 481–522. doi: [10.1146/annurev.astro.46.060407.145222](https://doi.org/10.1146/annurev.astro.46.060407.145222).
- Bahcall, John N. and Roger K. Ulrich (1988). “Solar Models, Neutrino Experiments, and Helioseismology”. In: *Reviews of Modern Physics* 60, pp. 297–372. doi: [10.1103/RevModPhys.60.297](https://doi.org/10.1103/RevModPhys.60.297).
- Baraffe, I. et al. (2017). “Lithium Depletion in Solar-like Stars: Effect of Overshooting Based on Realistic Multi-Dimensional Simulations”. In: *The Astrophysical Journal Letters* 845, p. L6. doi: [10.3847/2041-8213/aa82ff](https://doi.org/10.3847/2041-8213/aa82ff).

---

## REFERENCES

- Basu, Sarbani and H. M. Antia (2004). “Constraining Solar Abundances Using Helioseismology”. In: *The Astrophysical Journal Letters* 606, pp. L85–L88. doi: [10.1086/421110](https://doi.org/10.1086/421110).
- Bennett, C. L. et al. (2013). “Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results”. In: *The Astrophysical Journal Supplement Series* 208, p. 20. doi: [10.1088/0067-0049/208/2/20](https://doi.org/10.1088/0067-0049/208/2/20).
- Berger, Travis A. et al. (2020). “The \$Gaia\$-\$Kepler\$ Stellar Properties Catalog I: Homogeneous Fundamental Properties for 186,000 \$Kepler\$ Stars”. In: *arXiv e-prints* 2001, arXiv:2001.07737.
- Blanton, Michael R. et al. (2017). “Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe”. In: *The Astronomical Journal* 154, p. 28. doi: [10.3847/1538-3881/aa7567](https://doi.org/10.3847/1538-3881/aa7567).
- Böhm-Vitense, E. (1958). “Über Die Wasserstoffkonvektionszone in Sternen Verschiedener Effektivtemperaturen Und Leuchtkräfte. Mit 5 Textabbildungen”. In: *Z. Für Astrophys.* 46, p. 108.
- Borucki, William J. et al. (2010). “Kepler Planet-Detection Mission: Introduction and First Results”. In: *Science* 327, p. 977. doi: [10.1126/science.1185402](https://doi.org/10.1126/science.1185402).
- Broomhall, A.-M. et al. (2014). “Prospects for Asteroseismic Inference on the Envelope Helium Abundance in Red Giant Stars”. In: *Monthly Notices of the Royal Astronomical Society* 440, pp. 1828–1843. doi: [10.1093/mnras/stu393](https://doi.org/10.1093/mnras/stu393).
- Chiosi, C. and F. M. Matteucci (1982). “The Helium to Heavy Element Enrichment Ratio, Delta Y/Delta Z”. In: *A&A* 105, pp. 140–148.
- Christensen-Dalsgaard, J. (1984). “What Will Asteroseismology Teach Us”. In: p. 11.
- Christensen-Dalsgaard, Jørgen (2008). “ASTEC—the Aarhus STellar Evolution Code”. In: *Astrophysics and Space Science* 316, pp. 13–24. doi: [10.1007/s10509-007-9675-5](https://doi.org/10.1007/s10509-007-9675-5).
- Corsaro, E., J. De Ridder, and R. A. García (2015). “High-Precision Acoustic Helium Signatures in 18 Low-Mass Low-Luminosity Red Giants: Analysis from More than Four Years of *Kepler* Observations★”. en. In: *A&A* 578, A76. doi: [10.1051/0004-6361/201525922](https://doi.org/10.1051/0004-6361/201525922).

---

## REFERENCES

- Davies, G. R. et al. (2016). “Oscillation Frequencies for 35 *Kepler* Solar-Type Planet-Hosting Stars Using Bayesian Techniques and Machine Learning”. en. In: *MNRAS* 456.2, pp. 2183–2195. doi: [10.1093/mnras/stv2593](https://doi.org/10.1093/mnras/stv2593).
- Deubner, F.-L., R. K. Ulrich, and E. J. Rhodes Jr. (1979). “Solar P-Mode Oscillations as a Tracer of Radial Differential Rotation”. In: *Astronomy and Astrophysics* 72, pp. 177–185.
- Deubner, Franz-Ludwig and Douglas Gough (1984). “Helioseismology: Oscillations as a Diagnostic of the Solar Interior”. In: *Annual Review of Astronomy and Astrophysics* 22, pp. 593–619. doi: [10.1146/annurev.aa.22.090184.003113](https://doi.org/10.1146/annurev.aa.22.090184.003113).
- Eddington, A. S. (1926). *The Internal Constitution of the Stars*. Cambridge: Cambridge University Press.
- García Pérez, Ana E. et al. (2016). “ASPCAP: The APOGEE Stellar Parameter and Chemical Abundances Pipeline”. In: *The Astronomical Journal* 151, p. 144. doi: [10.3847/0004-6256/151/6/144](https://doi.org/10.3847/0004-6256/151/6/144).
- Gough, D. O. (1977). “Mixing-Length Theory for Pulsating Stars”. In: *The Astrophysical Journal* 214, pp. 196–213. doi: [10.1086/155244](https://doi.org/10.1086/155244).
- Grevesse, N. and A. J. Sauval (1998). “Standard Solar Composition”. In: *Space Sci. Rev.* 85, pp. 161–174. doi: [10.1023/A:1005161325181](https://doi.org/10.1023/A:1005161325181).
- Halabi, G. M. et al. (2017). “2DStars: A Two-Dimensional Stellar Evolution Code.” In: *Memorie della Societa Astronomica Italiana* 88, p. 319.
- Homan, Matthew D. and Andrew Gelman (2014). “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo”. In: *J. Mach. Learn. Res.* 15.1, pp. 1593–1623.
- Houdek, G. and D. O. Gough (2007). “An Asteroseismic Signature of Helium Ionization”. In: *Monthly Notices of the Royal Astronomical Society* 375, pp. 861–880. doi: [10.1111/j.1365-2966.2006.11325.x](https://doi.org/10.1111/j.1365-2966.2006.11325.x).
- Kippenhahn, R., A. Weigert, and Emmi Hofmeister (1967). “Methods for Calculating Stellar Evolution”. In: *Methods Comput. Phys.* 7, pp. 129–190.

---

## REFERENCES

- Kjeldsen, H. and T. R. Bedding (1995). “Amplitudes of Stellar Oscillations: The Implications for Asteroseismology.” In: *Astronomy and Astrophysics* 293, pp. 87–106.
- Lomb, N. R. (1976). “Least-Squares Frequency Analysis of Unequally Spaced Data”. In: *Astrophysics and Space Science* 39, pp. 447–462. doi: [10.1007/BF00648343](https://doi.org/10.1007/BF00648343).
- Ludwig, Hans-Günter, Bernd Freytag, and Matthias Steffen (1999). “A Calibration of the Mixing-Length for Solar-Type Stars Based on Hydrodynamical Simulations. I. Methodical Aspects and Results for Solar Metallicity”. In: *Astronomy and Astrophysics* 346, pp. 111–124.
- Mazumdar, A. et al. (2014). “Measurement of Acoustic Glitches in Solar-Type Stars from Oscillation Frequencies Observed by Kepler”. In: *The Astrophysical Journal* 782, p. 18. doi: [10.1088/0004-637X/782/1/18](https://doi.org/10.1088/0004-637X/782/1/18).
- Monteiro, Mário J. P. F. G., Jørgen Christensen-Dalsgaard, and Michael J. Thompson (2000). “Seismic Study of Stellar Convective Regions: The Base of the Convective Envelope in Low-Mass Stars”. In: *Monthly Notices of the Royal Astronomical Society* 316, pp. 165–172. doi: [10.1046/j.1365-8711.2000.03471.x](https://doi.org/10.1046/j.1365-8711.2000.03471.x).
- Morel, P. and Y. Lebreton (2008). “CESAM: A Free Code for Stellar Evolution Calculations”. In: *Astrophysics and Space Science* 316, pp. 61–73. doi: [10.1007/s10509-007-9663-9](https://doi.org/10.1007/s10509-007-9663-9).
- Mosser, B., K. Belkacem, et al. (2011). “The Universal Red-Giant Oscillation Pattern. An Automated Determination with CoRoT Data”. In: *Astronomy and Astrophysics* 525, p. L9. doi: [10.1051/0004-6361/201015440](https://doi.org/10.1051/0004-6361/201015440).
- Mosser, B., C. Gehan, et al. (2018). “Period Spacings in Red Giants IV. Toward a Complete Description of the Mixed-Mode Pattern”. In: *Astronomy and Astrophysics* 618, A109. doi: [10.1051/0004-6361/201832777](https://doi.org/10.1051/0004-6361/201832777).
- Paxton, Bill et al. (2011). “Modules for Experiments in Stellar Astrophysics (MESA)”. In: *ApJS* 192, p. 3. doi: [10.1088/0067-0049/192/1/3](https://doi.org/10.1088/0067-0049/192/1/3).
- Pinsonneault, Marc H. et al. (2014). “The APOKASC Catalog: An Asteroseismic and Spectroscopic Joint Survey of Targets in the Kepler Fields”. In: *The Astrophysical Journal Supplement Series* 215, p. 19. doi: [10.1088/0067-0049/215/2/19](https://doi.org/10.1088/0067-0049/215/2/19).

---

## REFERENCES

- Planck Collaboration et al. (2016). “Planck 2015 Results. XIII. Cosmological Parameters”. In: *Astronomy and Astrophysics* 594, A13. doi: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830).
- Ricker, George R. et al. (2015). “Transiting Exoplanet Survey Satellite (TESS)”. In: *Journal of Astronomical Telescopes, Instruments, and Systems* 1, p. 014003. doi: [10.1117/1.JATIS.1.014003](https://doi.org/10.1117/1.JATIS.1.014003).
- Roxburgh, I. W. (2004). “2-Dimensional Models of Rapidly Rotating Stars I. Uniformly Rotating Zero Age Main Sequence Stars”. In: *Astronomy and Astrophysics* 428, pp. 171–179. doi: [10.1051/0004-6361:20041202](https://doi.org/10.1051/0004-6361:20041202).
- Salvatier, John, Thomas V. Wiecki, and Christopher Fonnesbeck (2016). “Probabilistic Programming in Python Using PyMC3”. en. In: *PeerJ Comput. Sci.* 2, e55. doi: [10.7717/peerj-cs.55](https://doi.org/10.7717/peerj-cs.55).
- Scargle, J. D. (1982). “Studies in Astronomical Time Series Analysis. II - Statistical Aspects of Spectral Analysis of Unevenly Spaced Data”. In: *The Astrophysical Journal* 263, pp. 835–853. doi: [10.1086/160554](https://doi.org/10.1086/160554).
- Serenelli, Aldo et al. (2017). “The First APOKASC Catalog of Kepler Dwarf and Subgiant Stars”. In: *ApJS* 233, p. 23. doi: [10.3847/1538-4365/aa97df](https://doi.org/10.3847/1538-4365/aa97df).
- Silva Aguirre, V. et al. (2015). “Ages and Fundamental Properties of Kepler Exoplanet Host Stars from Asteroseismology”. en. In: *MNRAS* 452.2, p. 2127. doi: [10.1093/mnras/stv1388](https://doi.org/10.1093/mnras/stv1388).
- Soderblom, David R. (2010). “The Ages of Stars”. en. In: *ARA&A* 48.1, pp. 581–629. doi: [10.1146/annurev-astro-081309-130806](https://doi.org/10.1146/annurev-astro-081309-130806).
- Tassoul, M. (1980). “Asymptotic Approximations for Stellar Nonradial Pulsations”. In: *The Astrophysical Journal Supplement Series* 43, pp. 469–490. doi: [10.1086/190678](https://doi.org/10.1086/190678).
- Townsend, R. H. D. and S. A. Teitler (2013). “GYRE: An Open-Source Stellar Oscillation Code Based on a New Magnus Multiple Shooting Scheme”. In: *Monthly Notices of the Royal Astronomical Society* 435, pp. 3406–3418. doi: [10.1093/mnras/stt1533](https://doi.org/10.1093/mnras/stt1533).
- Trampedach, Regner et al. (2014). “Improvements to Stellar Structure Models, Based on a Grid of 3D Convection Simulations - II. Calibrating the Mixing-Length Formulation”. In: *MNRAS* 445, pp. 4366–4384. doi: [10.1093/mnras/stu2084](https://doi.org/10.1093/mnras/stu2084).

---

## REFERENCES

- Ulrich, R. K. (1986). “Determination of Stellar Ages from Asteroseismology”. In: *The Astrophysical Journal Letters* 306, pp. L37–L40. doi: [10.1086/184700](https://doi.org/10.1086/184700).
- Ulrich, Roger K. (1970). “The Five-Minute Oscillations on the Solar Surface”. In: *The Astrophysical Journal* 162, p. 993. doi: [10.1086/150731](https://doi.org/10.1086/150731).
- Vandenberg, D. A. (1985). “Evolution of 0.7-3.0 Solar Masses Stars Having Fe/H between -1.0 and 0.0”. In: *The Astrophysical Journal Supplement Series* 58, pp. 711–769. doi: [10.1086/191055](https://doi.org/10.1086/191055).
- Verma, Kuldeep, Keyuri Raodeo, H. M. Antia, et al. (2017). “Seismic Measurement of the Locations of the Base of Convection Zone and Helium Ionization Zone for Stars in the Kepler Seismic LEGACY Sample”. In: *The Astrophysical Journal* 837, p. 47. doi: [10.3847/1538-4357/aa5da7](https://doi.org/10.3847/1538-4357/aa5da7).
- Verma, Kuldeep, Keyuri Raodeo, Sarbani Basu, et al. (2019). “Helium Abundance in a Sample of Cool Stars: Measurements from Asteroseismology”. In: *MNRAS* 483, pp. 4678–4694. doi: [10.1093/mnras/sty3374](https://doi.org/10.1093/mnras/sty3374).
- Weiss, Achim and Helmut Schlattl (2008). “GARSTEC—the Garching Stellar Evolution Code. The Direct Descendant of the Legendary Kippenhahn Code”. In: *Astrophysics and Space Science* 316, pp. 99–106. doi: [10.1007/s10509-007-9606-5](https://doi.org/10.1007/s10509-007-9606-5).
- Yakovlev, D. G. and V. A. Urpin (1980). “Thermal and Electrical Conductivity in White Dwarfs and Neutron Stars”. In: *Soviet Astronomy* 24, p. 303.

## **Appendix A**

### **Accompanying Paper**

# Hierarchically modelling *Kepler* dwarfs and subgiants to improve inference of stellar properties with asteroseismology

Alexander J. Lyttle,<sup>1,2\*</sup> Tanda Li,<sup>1,2</sup> Guy R. Davies,<sup>1,2</sup> Lindsey M. Carboneau,<sup>1,2</sup> and To B. Confirmed

<sup>1</sup>School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, UK

<sup>2</sup>Stellar Astrophysics Centre (SAC), Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

With recent advances in modelling stars using high-precision asteroseismology, the systematic effects associated with our assumptions of stellar helium abundance ( $Y$ ) and the mixing-length theory parameter ( $\alpha_{\text{mlt}}$ ) are becoming more important. In this paper, we apply a new method to improve the inference of stellar parameters for a sample of *Kepler* dwarfs and subgiants across a narrow mass range ( $0.8 < M < 1.2 M_{\odot}$ ). In this method, we include a statistical treatment of  $Y$  and the  $\alpha_{\text{mlt}}$ . We develop a hierarchical Bayesian model to encode information about the distribution of  $Y$  and  $\alpha_{\text{mlt}}$  in the population, fitting a linear helium enrichment law including an intrinsic spread around this relation and normal distribution in  $\alpha_{\text{mlt}}$ . We test various levels of pooling parameters, with and without solar data as a calibrator. We are able to constrain  $\alpha_{\text{mlt}}$  in the population with a mean of  $1.9 \pm 0.1$  with, and  $1.7 \pm 0.1$  without solar data. We find the gradient for the enrichment law,  $\Delta Y / \Delta Z$  also varied with and without the solar data ( $1.1 \pm 0.3$  and  $1.6 \pm 0.5$  respectively). While accounting for the uncertainty in  $Y$  and  $\alpha_{\text{mlt}}$ , we are still able to report statistical uncertainties of 2.5 per cent in mass, 1.2 per cent in radius, and 12 per cent in age, by pooling across our sample. Our method can also be applied to larger samples which will lead to improved constraints on both the population level inference and the star-by-star fundamental parameters. We show that more work is needed to study the validity of using the Sun as a calibrator for  $Y$  and  $\alpha_{\text{mlt}}$ , perhaps by including asteroseismic signatures of helium abundance.

## Key words:

asteroseismology – stars: fundamental parameters – stars: low-mass – stars: oscillations – stars: solar-type – stars: statistics

## 1 INTRODUCTION

In recent years, the inference of stellar ages, masses and radii have improved through the use of asteroseismology (e.g. see the review by Chaplin & Miglio 2013). Measuring the oscillation modes in stars using photometric time series data, from missions such as *CoRoT* (Baglin et al. 2006), *Kepler* (Borucki et al. 2010) and *TESS* (Ricker et al. 2015) has given us new insights into the structure and evolution of stars. Recent examples include a deeper understanding of stellar structure (Verma et al. 2017), chronology of a Milky Way merger (Chaplin et al. 2020) and classifying exoplanetary systems (Huber et al. 2019). Several studies have used grids of stellar models with constraints from asteroseismology to produce catalogues of precise stellar parameters (Pinsonneault et al. 2014; Silva Aguirre

et al. 2017). However, with increasing precision on fundamental parameters inferred from stellar models with asteroseismology, extra care should be taken to ensure that we are accounting for uncertainty in our choice of stellar physics.

Typically, stellar models are achieved on a star-by-star basis, with assumptions made about the bulk physical quantities based on empirical relations or solar calibrations. In many stellar models, a helium ( $Y$ ) to heavy-element ( $Z$ ) ratio,  $\Delta Y / \Delta Z$  and mixing-length theory parameter,  $\alpha_{\text{mlt}}$  are assumed. However, there has been little effort to account for the population distribution of such quantities. Assuming  $Y$  and  $\alpha_{\text{mlt}}$ , can result in inaccurate inference (Valle et al. 2015). Independently modelling  $Y$  and  $\alpha_{\text{mlt}}$  can also be computationally demanding and requires high-precision observations in order to return meaningful stellar properties.

In this work, we apply a new method (Davies et al. in prep.) to determine stellar properties for a sample of *Kepler* dwarfs and

\* E-mail: ajl573@student.bham.ac.uk

subgiants using a hierarchical Bayesian model (HBM). With an HBM, we introduce population-level distributions for  $Y$  and  $\alpha_{\text{mlt}}$  to encode prior information throughout the sample. We will show that when an HBM is used, increasing the number of free parameters over that of existing work can still lead to high-precision masses, radii and ages.

To describe the distribution of  $Y$ , we assume a linear helium enrichment law characterised by freely varied population parameters: the gradient ( $\Delta Y/\Delta Z$ ), the primordial helium abundance ( $Y_P$ , at  $Z = 0$ ) and an intrinsic spread in helium ( $\sigma_Y$ ). There have been many studies which fit a linear enrichment law, from modelling eclipsing binaries (Ribas et al. 2000) to measurements in galactic H-II regions (Balser 2006). In recent years, the value of  $\Delta Y/\Delta Z$  has been determined for samples of main sequence stars (Casagrande et al. 2007), open clusters (Brogaard et al. 2012) and more recently with asteroseismology (Silva Aguirre et al. 2017; Verma et al. 2019) ranging from around 1 to 3.

The widely used mixing-length theory of convection, parametrised by  $\alpha_{\text{mlt}}$ , has been tested throughout the Hertzsprung-Russell diagram with 3D hydrodynamical simulations (Trampedach et al. 2014; Magic et al. 2015) and asteroseismology (Tayar et al. 2017; Viani et al. 2018; Li et al. 2018). However, in many grids of stellar models, a constant value calibrated to reproduce the Sun is assumed. In this work, we experiment with two prior assumptions for  $\alpha_{\text{mlt}}$ , one assuming the best-fitting  $\alpha_{\text{mlt}}$  is normally distributed in our sample, and the other assuming it is constant throughout.

The use of HBMs has been demonstrated in other areas of astrophysics to reduce individual parameter uncertainties by encoding prior information about the distribution of said parameter in a population. For example, HBMs have been used with data from *Gaia* to improve distance measures (Leistedt & Hogg 2017; Anderson et al. 2018) and calibrate the red clump as a standard candle (Hawkins et al. 2017; Chan & Bovy 2020) using asteroseismology (Hall et al. 2019). In other instances, HBMs have been used to infer stellar eccentricity (Hogg et al. 2010) and obliquity (Morton & Winn 2014) of transit systems, and stellar inclination with asteroseismology (Campante et al. 2016; Kuszlewicz et al. 2019).

Our HBM requires a way to map from the stellar initial (or bulk) properties to predict observables. We can achieve this with a large grid of stellar evolutionary models. However, a discrete grid can produce inaccurate posterior distributions, limited to the grid resolution. Increasing the grid resolution is computationally demanding, especially when scaling to higher input dimensions. One method is to interpolate the stellar models, as is common in the isochrone-fitting method (see e.g. Berger et al. 2020). However, interpolation can become computationally expensive at high input dimensions and grid size, and evaluating the likelihood using modern Bayesian sampling techniques is slow.

We used machine learning to map stellar inputs to observables to provide a fast way to sample the HBM. Some examples of machine learning in asteroseismology include the use of unsupervised learning in oscillation mode detection (Davies et al. 2016) and a random forest to map observables to fundamental parameters (Bellinger et al. 2016). In this work, we train an artificial neural network (ANN) on a large grid of stellar models. There have also been similar applications of ANNs in asteroseismology (Verma et al. 2016; Hendriks & Aerts 2019) but not yet in the context of an HBM.

## 2 DATA

We began with the sample of 415 stars from the first APOKASC catalogue of dwarfs and subgiants (Serenelli et al. 2017, hereafter S17). This sample provides an extensive set of dwarfs and subgiant stars with asteroseismic detections observed by the *Kepler* mission. S17 used grid-based modelling to determine the ages,  $\tau$ , masses,  $M$ , radii,  $R$  and surface gravity,  $\log g$  of stars in the sample, using global asteroseismic parameters, effective temperature  $T_{\text{eff}}$ , and metallicity, [M/H] as inputs.

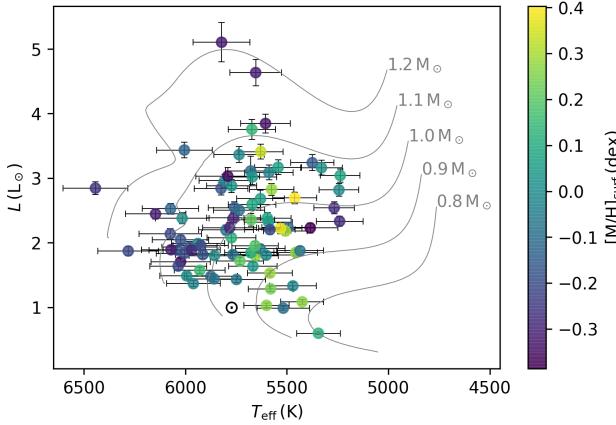
Using five independent pipelines, S17 determined values for global asteroseismic parameters – the large frequency separation  $\Delta\nu$  and the frequency at maximum power,  $\nu_{\text{max}}$  with median uncertainties of 1.7 per cent and 4 per cent respectively. We chose to adopt the  $\Delta\nu$  determined in their work as inputs for our method. They used [M/H] published in Data Release 13 (DR13) of the APOGEE stellar abundances pipeline (ASPCAP) with an additional uncertainty of 0.1 dex (García Pérez et al. 2016; Albareti et al. 2017). For their preferred set of results, they adopted a  $T_{\text{eff}}$  scale from the Sloan Digital Sky Survey (SDSS) *griz*-band photometry (Pinsonneault et al. 2012) with a median uncertainty of 70 K.

We made cuts to the APOKASC sample to remove more evolved stars, metal-poor stars and stars likely to be subjected to the effects of a convective, hydrogen-burning core. We removed more evolved stars by cutting those with  $\log g < 3.8$  dex. We then kept stars within  $1-\sigma$  of  $-0.5 < [\text{M}/\text{H}] < +0.5$  to remove metal-poor stars. Stars with  $M \gtrsim 1.2 M_{\odot}$  are understood to have a convective, hydrogen-burning core, with some dependence on the choice of stellar physics (Appourchaux et al. 2015). Stellar models with a convective core require the treatment of extra stellar physics such as overshooting, which is beyond the scope of this work. Therefore, we keep only stars with masses determined by S17 to within  $1-\sigma$  of 0.8 to  $1.2 M_{\odot}$ .

Following cuts to the sample, we adopted updated ASPCAP spectroscopic metallicities, [M/H], from Data Release 14 (DR14; García Pérez et al. 2016; Blanton et al. 2017) which had a median uncertainty of 0.07 dex. We also chose to adopt  $T_{\text{eff}}$  from the same catalogue to be internally consistent. We note that our chosen effective temperature scale is offset from the photometric temperature scale of S17 by approximately  $-170$  K with a dispersion of  $\sim 120$  K. The median uncertainty in our adopted ASPCAP  $T_{\text{eff}}$  was 125 K which is compatible with the dispersion observed.

To calculate luminosities for the sample, we used *Gaia* Data Release 2 (DR2) parallaxes (Gaia Collaboration et al. 2016, 2018). We cross-matched the remaining sample with the DR2 catalogue, taking the nearest neighbours within a  $4''$  radius. Although DR2 parallaxes have improved upon the DR1 values at the time of S17, there was still evidence for a zero-point offset (Lindegren et al. 2018). We adopted a global offset of 0.05 mas, in the sense that DR2 parallaxes were underestimated, representative of values obtained in the literature (see e.g. Riess et al. 2018; Zinn et al. 2019; Hall et al. 2019; Chan & Bovy 2020). We then cross-matched our sample with the Two-Micron All Sky Survey (2MASS) to obtain  $K_S$ -band ( $2.16 \mu\text{m}$ ) photometry (Skrutskie et al. 2006).

We determined luminosities,  $L$  for the sample using the direct method of ISOCLASSIFY with  $K_S$ -band photometry, *Gaia* DR2 parallaxes, ASPCAP [M/H] and  $T_{\text{eff}}$  and asteroseismic  $\log g$  as inputs (Huber et al. 2017; Berger et al. 2020). This involved computing absolute  $K_S$ -band magnitudes using the *Gaia* DR2 parallaxes and extinctions determined by the 3D galactic reddening maps of Green et al. (2018). We determined absolute bolometric magnitudes by interpolating the MIST bolometric correction tables as a function of



**Figure 1.** The luminosity,  $L$  against effective temperature,  $T_{\text{eff}}$  of the sample of 81 *Kepler* dwarfs and subgiants studied in this work. Each star is coloured according to metallicity. The grey lines depict evolutionary tracks with  $[M/H]_{\text{init}} = 0.0 \text{ dex}$ ,  $Y_{\text{init}} = 0.28$  and  $\alpha_{\text{mlt}} = 1.9$  for different stellar masses. The current position of the Sun is shown by the  $\odot$  symbol.

$T_{\text{eff}}$ ,  $\log g$  and  $[M/H]$  (Dotter 2016; Choi et al. 2016). An uncertainty of 0.02 mag was assumed in ISOCALIFY for both the extinctions and bolometric corrections, representative of typical systematics during interpolation (Huber et al. 2017). We obtained luminosities for the sample with a median uncertainty of 3.4 per cent.

The final sample comprised 81 stars for which we had data for  $T_{\text{eff}}$ ,  $[M/H]$ ,  $\Delta\nu$  and  $L$  to use as inputs for our stellar modelling method – see Table 1. Figure 1 shows the  $L$  and  $T_{\text{eff}}$  for the sample plot on a Hertzsprung-Russell diagram in context with stellar evolutionary tracks at solar metallicity.

### 3 METHODS

Our principle goal was to impove inference of fundamental stellar parameters for our set of stars. To achieve this, we constructed a hierarchical Bayesian model (HBM) which utilises a prior assumption of the distribution of stars in the population to share information between the stars. Based on the work of Davies et al. (in prep.), our HBM is a generative model which requires a function to map stellar initial conditions to their observables.

Firstly, we used a stellar evolutionary code to compute a grid of models to predict observable quantities (see 3.1). For a given stellar mass,  $M$ , metallicity,  $[M/H]_{\text{init}}$ , helium fraction,  $Y_{\text{init}}$  and mixing-length theory parameter,  $\alpha_{\text{mlt}}$  our stellar models evolve the star outputting  $T_{\text{eff}}$ ,  $L$  and chemical composition as a function of age,  $\tau$ . Calls to models of stellar evolution are slow and the grid produced is discrete. This makes it difficult to robustly evaluate an HBM using the grid alone. We could interpolate the grid of stellar models, for example with the isochrone fitting method [CITE]. However, interpolation does not scale well with the number of input dimensions and points on the grid, reducing the scalability of our method.

In Section 3.2, we describe a method to replace the grid of stellar models with a smooth function approximation using machine learning. In particular, we trained an artificial neural network (ANN) on the grid of stellar models to map stellar fundamentals to observables. Fast evaluation of the ANN gradient is required during training. Consequently, estimating the gradient of the model likelihood is possible with an ANN. With an ANN, we open up the

possibility of using the Hamiltonian Monte Carlo (HMC) algorithm which requires the gradient to sample the model posterior – for example, using the No-U-Turn Sampler (NUTS; Homan & Gelman 2014).

Finally, we constructed three Bayesian models in Section 3.3 which each used the trained ANN to estimate stellar fundamental parameters. We then tested the models on a set of synthetic stars generated by the stellar evolutionary code. Once we had tested the model accuracy using the synthetic stars, we evaluated each model on the subset of the APOKASC catalogue selected in Section 2.

#### 3.1 Grid of stellar models

We built a stellar model grid to use in training the ANN. The grid includes four independent model inputs: stellar mass ( $M$ ), initial helium fraction ( $Y_{\text{init}}$ ), initial metallicity ( $[M/H]_{\text{init}}$ ), and the mixing-length parameter ( $\alpha_{\text{mlt}}$ ). Ranges and grid steps of the four model inputs are summarised in Table 2. We computed each stellar evolutionary track from the Hayashi line and to the base of red-giant branch where  $\log g = 3.6 \text{ dex}$ . We also computed evolutionary tracks with input values at the midpoint between points on the grid for validating the ANN.

##### 3.1.1 Stellar models and input physics

We used Modules for Experiments in Stellar Astrophysics (MESA, version 12115) to establish a grid of stellar models. MESA is an open-source stellar evolution package which is undergoing active development. Descriptions of input physics and numerical methods can be found in Paxton et al. (2011, 2013, 2015). We adopted the solar chemical mixture,  $(Z/X)_\odot = 0.0181$ , provided by Asplund et al. (2009). The initial chemical composition was calculated by:

$$\log(Z_{\text{init}}/X_{\text{init}}) = \log(Z/X)_\odot + [M/H]_{\text{init}}. \quad (1)$$

We used the MESA  $\rho - T$  tables based on the 2005 update of OPAL EOS tables (Rogers & Nayfonov 2002) and OPAL opacity supplemented by low-temperature opacity (Ferguson et al. 2005). The MESA ‘simple’ photosphere were used as the set of boundary conditions for modelling the atmosphere. The mixing-length theory of convection was implemented, where  $\alpha_{\text{MLT}} = \ell_{\text{MLT}}/H_p$  is the mixing-length parameter. We also applied the MESA predictive mixing scheme (Paxton et al. 2018, 2019) in the model computation.

Atomic diffusion of helium and heavy elements was also taken into account. MESA calculates particle diffusion and gravitational settling by solving Burger’s equations using the method and diffusion coefficients of Thoul et al. (1994). We considered eight elements ( $^1\text{H}$ ,  $^3\text{He}$ ,  $^4\text{He}$ ,  $^{12}\text{C}$ ,  $^{14}\text{N}$ ,  $^{16}\text{O}$ ,  $^{20}\text{Ne}$ , and  $^{24}\text{Mg}$ ) for diffusion calculations, and had the charge calculated by the MESA ionization module, which estimates the typical ionic charge as a function of  $T$ ,  $\rho$ , and free electrons per nucleon from Paquette et al. (1986).

The evolution time step was mainly controlled by the set-up tolerances on changes in surface effective temperature and luminosity. We saved one structural model at every time step at main sequence and every two steps after central hydrogen exhaustion. For each evolutionary track, we obtained  $\sim 100$  at the main-sequence stage and 500 – 700 at evolved stages.

##### 3.1.2 Oscillation models and seismic $\Delta\nu$

Theoretical stellar oscillations were calculated with the GYRE code (version 5.1), which was developed by Townsend & Teitler (2013).

**Table 1.** The observables and their respective uncertainties for the 10 stars in sample of 81 stars. The whole table is available online.

Name	$T_{\text{eff}}$ (K)	$\sigma_{T_{\text{eff}}}$ (K)	$L$ ( $L_{\odot}$ )	$\sigma_L$ ( $L_{\odot}$ )	$\Delta\nu$ ( $\mu\text{Hz}$ )	$\sigma_{\Delta\nu}$ ( $\mu\text{Hz}$ )	$[\text{M}/\text{H}]_{\text{surf}}$ (dex)	$\sigma_{[\text{M}/\text{H}]}$ (dex)
KIC10079226	5928.84	124.84	1.57	0.05	116.04	0.73	0.16	0.07
KIC10215584	5666.92	119.33	1.64	0.06	115.16	2.83	0.04	0.07
KIC10319352	5456.17	106.65	1.85	0.06	78.75	1.73	0.27	0.06
KIC10322381	6146.79	148.58	2.44	0.08	86.64	6.57	-0.32	0.08
KIC10417911	5628.26	109.99	3.41	0.12	56.14	2.10	0.34	0.07

**Table 2.** Stellar model grid parameters for training and test datasets.

Stellar model grid			
Input Parameter	Range	Increment	$N_{\text{track}}$
$M$ ( $M_{\odot}$ )	0.80 – 1.20	0.01	41
[M/H] (dex)	-0.5 – 0.2/0.25 – 0.5	0.1/0.05	14
$Y_{\text{init}}$	0.24 – 0.32	0.02	5
$\alpha_{\text{mlt}}$	1.5 – 2.5	0.2	6
<b>Total</b>			17,220

We computed radial modes (for  $\ell = 0$ ) by solving the adiabatic stellar pulsation equations with the structural models generated by MESA. We determined a seismic large separation ( $\Delta\nu$ ) for each model with theoretical radial modes to avoid the systematic offset of the scaling relation. We derived  $\Delta\nu$  with the approach given by White et al. (2011), which is a weighted least-squares fit to the radial frequencies as a function of  $n$ .

We chose to ignore the well known, yet poorly characterised impact of modelled oscillation mode inaccuracies in the near-surface region of the star (Kjeldsen et al. 2008; Ball & Gizon 2014; Sonoi et al. 2015). This presents only a small effect when considering the average large frequency spacing,  $\Delta\nu$  and is beyond the scope of this paper.

### 3.2 Artificial neural network

Once we constructed our grid of models, we needed a way in which we could continuously sample the grid for use in our statistical model. We could interpolate the grid, as is common in the isochrone-fitting method [CITE], but this would be slow due to the high dimensionality of our inputs and the size of the dataset. Moreover, evaluating the gradient of an interpolated function is slow. In this work, we utilise deep learning (DL) to approximate the grid of stellar models via an artificial neural network (ANN). The ANN is advantageous over interpolation due to scaling well with dimensionality, fast training and evaluation, and easy gradient evaluation due to its roots in linear algebra [CITE].

We trained an ANN on the data generated by the grid of stellar models to map fundamentals to observables. Firstly, we split the grid into a *train* and *test* dataset for tuning the ANN, as described in Section 3.2.1. We then tested a multitude of ANN configurations and training data inputs, repeatedly evaluating them with the test dataset in Section 3.2.2. Finally, in Section 3.2.3, we reserved a set of off-grid stellar models as our final *validation* dataset to evaluate the approximation ability of the best-performing ANN. In this section, we briefly describe the theory and motivation behind the ANN.

An ANN is a network of artificial *neurons* which each transform some input vector,  $\mathbf{x}$  based on trainable weights,  $\mathbf{w}$  and a bias,  $b$  [CITATIONS]. The weights are represented by the connections

between neurons and the bias is a unique scalar associated with each neuron. Deep learning (DL) is the name given to the case where neurons are arranged into a series of layers such that any neuron in layer  $k - 1$  is connected to at least one of the neurons in layer  $k$ .

In this work, we considered a fully-connected ANN, where each neuron in layer  $k - 1$  is connected to every neuron in layer  $k$ . The output of a given neuron,  $i$  in layer  $k$  is,

$$x_{i,k} = f_k(\mathbf{w}_{i,k} \cdot \mathbf{x}_{k-1} + b_{i,k}) \quad (2)$$

where  $f_k$  is the *activation* function for the  $k$ -th layer,  $\mathbf{w}_{i,k}$  are the weights connecting all the neurons in layer  $k - 1$  to the current neuron, and  $b_{i,k}$  is the bias. This generalises such that the output of the  $k$ -th layer is,

$$\mathbf{x}_k = f_k(\mathbf{W}_k \cdot \mathbf{x}_{k-1} + \mathbf{b}_k), \quad (3)$$

where  $\mathbf{W}_k$  is the matrix of weights leading to all neurons in the  $k$ -th layer. For a regression neural network, we typically have a linear activation function applied to the output of the final layer. Therefore, the output of a network of  $M$  hidden layers with initial input  $X$  is,

$$\mathbf{Y} = \mathbf{W}_M \cdot f_{M-1}(\dots f_1(\mathbf{W}_1 \cdot f_0(\mathbf{W}_0 \cdot X + \mathbf{b}_0) + \mathbf{b}_1)) + \mathbf{b}_M \quad (4)$$

We also restricted our configuration to an ANN with the same number of neurons,  $N$  in each hidden layer. Hereafter, we refer to our choice of neurons per layer,  $N$  and hidden layers,  $M$  as the *architecture*.

To fit the ANN, we used a set of training data,  $\mathbf{D}_{\text{train}} = \{(X_1, Y_1) \dots (X_{N_{\text{train}}}, Y_{N_{\text{train}}})\}$  comprising  $N_{\text{train}}$  input-output pairs. We split the training data into pseudo-random batches,  $\mathbf{D}_{\text{batch}}$  because this has been shown to improve model convergence and computational efficiency [CITE]. The set of predictions made for each batch is evaluated with an error function,  $E(\mathbf{D}_{\text{batch}})$ , also known as the *loss* which quantifies the difference between the training data and predictions. We also considered an addition to the loss called *regularisation* which helps reduce over-fitting (CITE). During fitting, the weights are updated after each batch using an algorithm called the *optimizer*, back-propagating the error with the goal of minimising the loss.

We trained the ANN using TensorFlow (Abadi et al. 2016). We varied the architecture, number of batches, choice of loss function, optimizer and regularisation during the optimisation phase. For each set of ANN parameters, we initialised the ANN with a random set of weights and biases and minimized the loss over a given number of *epochs*. An epoch is defined as one iteration through the entire training dataset,  $\mathbf{D}_{\text{train}}$ . We tracked the loss for each ANN using an independent test dataset to determine the most effective choice of ANN parameters (see Section 3.2.2).

#### 3.2.1 Train, test and validation data

We built the train and test dataset from the outputs of the grid of stellar models in Section 3.1. This included the input parameters:  $M$ ,  $\alpha_{\text{mlt}}$ ,  $Y_{\text{init}}$  and the initial heavy-elements fraction,  $Z_{\text{init}}$ .

We also included the  $T_{\text{eff}}$ ,  $\log g$ ,  $\Delta\nu$ , stellar age ( $\tau$ ), radius ( $R$ ), surface metallicity ( $[\text{M}/\text{H}]_{\text{surf}}$ ) and other chemical composition information generated by the models. We determined the fractional main sequence (MS) lifetime,  $f_{\text{MS}} = \tau/\tau_{\text{MS}}$ , of each evolutionary track by taking  $\tau_{\text{MS}}$  as the age when the central helium fraction,  $X_c < 0.01$ . We then cut data where  $f_{\text{MS}} < 0.01$  to remove points on the grid prior to the MS.

Once we had refined the data from the grid of models, we randomly sampled  $7.736 \times 10^6$  points to use as the training dataset, with the remaining  $\sim 2 \times 10^6$  points given to the test dataset. We then varied our choice of ANN input and output parameters among those available in the training dataset during tuning (see Section 3.2.2).

We produced a validation dataset of  $\sim 2 \times 10^6$  stellar models evolved using MESA. Values for the initial mass, metallicity, helium and mixing-length-theory parameter were chosen at the midpoint of the grid parameters described in Table 2. We prepared this dataset in the same way as the training set, but also constrained it to  $\tau < 15$  Gyr because ages above  $\sim 15$  yr are unphysical and such points are sparse in the training data. This dataset was set aside and evaluated on the final ANN.

### 3.2.2 Tuning

We needed to train an ANN which would reproduce stellar observables according to our choice of physics with greater accuracy than typical observational precisions. We experimented with a variety of ANN parameter choices, such as the architecture, activation function, optimization algorithm and loss function. We tuned the ANN parameters by varying them in both a grid-based and heuristic approach, each time evaluating the accuracy using the test dataset.

During initial tuning, we found that having stellar age as an input was unstable, because it varied heavily with the other input parameters. We mitigated this by introducing an input to describe the fraction of time a star had spent in a given evolutionary phase,  $f_{\text{evol}}$ .

$$f_{\text{evol}} = \begin{cases} f_{\text{MS}}, & f_{\text{MS}} \leq 1 \\ 1 + \frac{\tau - \tau_{\text{MS}}}{\tau_{\log g=3.6} - \tau_{\text{MS}}}, & f_{\text{MS}} > 1 \end{cases} \quad (5)$$

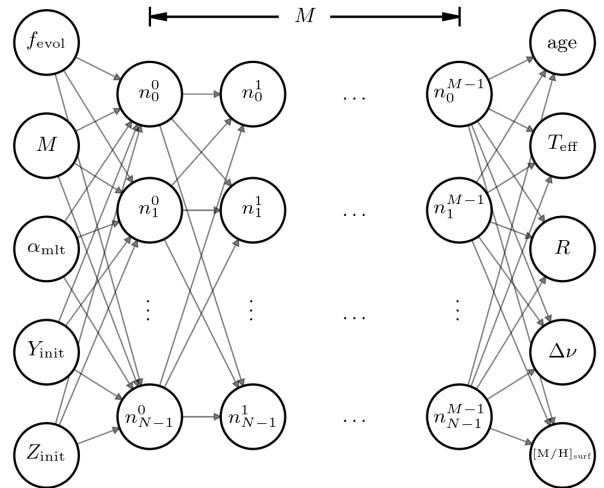
where  $\tau_{\log g=3.6}$  is the age of the star at the end of the track,

$$f_{\text{MS}} = \frac{\tau}{\tau_{\text{MS}}}, \quad (6)$$

and  $\tau_{\text{MS}}$  is the main sequence lifetime. In other words, a star with  $f_{\text{evol}} \in (0, 1]$  is in its main sequence phase, burning hydrogen in its core, and  $f_{\text{evol}} \in (1, 2]$  has left the main sequence and began burning hydrogen in a shell. Consequently,  $f_{\text{evol}}$  gives the ANN information about the internal state of the star which affects the output observables. Otherwise,  $f_{\text{evol}}$  is a meaningless parameter, although it could loosely be interpreted as a measure of the evolutionary phase of the star.

We also observed the ANN struggled to fit areas with a high rate of change in observables, partly because of poor grid coverage. To bias training to such areas, we calculated the gradient in  $T_{\text{eff}}$  and  $\log g$  between each point for each stellar evolutionary track and used them as optional weights to the loss during tuning. These weights multiplied the difference between the ANN prediction and the training data in our chosen loss function.

After preliminary tuning, we chose the ANN input and output parameters to be  $\mathbf{X} = \{f_{\text{evol}}, M, \alpha_{\text{mlt}}, Y_{\text{init}}, Z_{\text{init}}\}$  and  $\mathbf{Y} = \{\log(\tau), T_{\text{eff}}, R, \Delta\nu, [\text{M}/\text{H}]_{\text{surf}}\}$  respectively. A generalised form of our neural network is depicted in Figure 2. The inputs corresponded



**Figure 2.** An artificial neural network comprising  $M$  hidden layers with  $N$  neurons per layer. Arrows connecting the nodes represent tunable weights.

to initial conditions in the stellar modelling code and the outputs corresponded to surface conditions throughout the lifetime of the star, with the exception of age which is mapped from  $f_{\text{evol}}$ .

We standardised the training dataset by subtracting the median,  $\mu_{1/2}$  and dividing by the standard deviation,  $\sigma$ . We found that the ANN performed better when the training data was scaled in this way. In Table A1, we show the locations and scales of the standardisation for our chosen input and output parameters.

We found that the optimal choice of  $N$  and  $M$  varied depending on our choice of other ANN parameters. Therefore, each time we explored a new parameter, we trained an ANN with a grid of  $(N, M)$  ranging from  $(32, 2)$  to  $(512, 10)$ .

We evaluated the performance of three activation functions: the hyperbolic-tangent, the rectified linear unit (ReLU; Hahnloser et al. 2000; Glorot et al. 2011) and the exponential linear unit (ELU; Clevert et al. 2015). Although the ReLU activation function outperformed the other two in speed and accuracy, the ANN output was not smooth. The discontinuity in the ReLU function,  $f(x) = \max(0, x)$  caused the output to also be discontinuous. This made the ANN difficult to sample for our choice of statistical model (see Section 3.3). Out of the remaining activation functions, ELU performed the best, providing a smooth output which was well-suited to our probabilistic sampling methods.

We compared the performance of two optimisers: Adam (Kingma & Ba 2014) and stochastic gradient descent (SGD; see, e.g. Ruder 2016) with and without momentum (Qian 1999). Both optimizers required a choice of *learning rate* which determined the rate at which the weights were adjusted. We found that Adam performed well but the test loss was noisy as a function of epochs as it struggled to converge. The SGD optimizer was less noisy than Adam, but it was difficult to tune the learning rate. However, SGD with momentum allowed for more adaptive weight updates and outperformed the other configurations.

There are several ways to reduce over-fitting, from minimising the complexity of the architecture, to increasing the size and coverage of the training dataset. One alternative is to introduce weight regularisation. So-called L2 regularisation adds a term,  $\sim \lambda_k \sum_i w_{i,k}^2$  to the loss function for a given hidden layer,  $k$  which acts to keep the weights small. We varied the magnitude of  $\lambda_k$  and found that if

**Table 3.** The median error,  $\mu_{1/2}$  and median absolute deviation of the error,  $\sigma_{\text{MAD}} = 1.4826 \cdot \text{median}(|E(x) - \mu_{1/2}|)$ , for each output,  $x$ , where  $E(x)$  is the error given in the first column and  $\delta x = x_{\text{pred}} - x_{\text{true}}$ . All parameters are outputs of the ANN except for  $L$  which is derived.

Error	$\mu_{1/2}$	$\sigma_{\text{MAD}}$
$\delta\tau/\tau$ (%)	-0.003	0.178
$\delta T_{\text{eff}}$ (K)	-0.100	1.595
$\delta R/R$ (%)	0.002	0.071
$\delta L/L$ (%)	0.060	0.146
$\delta\Delta\nu$ ( $\mu\text{Hz}$ )	-0.007	0.084
$\delta[\text{M}/\text{H}]_{\text{surf}}$ (dex)	0.000	0.001

it was too large it would dominate the loss function, but if it was too small then performance on the test dataset was poorer.

We compared the choice of two error functions: mean squared error (MSE) and mean absolute error (MAE). The former is widely used among ANNs because it is more sensitive to large errors. However, we tracked both metrics regardless of which was added to the loss function and found that MAE converged faster. Although MAE is less effective at large errors, we found that these were typically at the edges of the grid and the accuracy was good enough everywhere else.

After extensive tuning, we opted for an ANN with  $N = 128$  neurons in each of  $M = 6$  hidden layers. Each of the hidden layers used an ELU activation function and L2 weight regularisation with  $\lambda = 1 \times 10^{-6}$ . We trained the ANN for 50,000 epochs with a 500 training data batches each containing 15,472 input-output pairs. To fit the ANN, we used an SGD optimiser with an initial learning rate of  $1 \times 10^{-4}$  and momentum of 0.999 with an MAE loss function. Training took  $\sim 48$  h on an NVidia Tesla V100 graphics processing unit (GPU).

### 3.2.3 Validation

The validation dataset contained  $\sim 2 \times 10^6$  models evolved in the same way as the training dataset but with initial conditions at the midpoint of those in the grid. We made predictions for the validation dataset, deriving luminosity from the output radius and effective temperature, using the final trained ANN as described in Section 3.2.2. We then evaluated the accuracy of the ANN by taking the difference between the validation truth and prediction,  $x_{\text{true}} - x_{\text{pred}}$ .

We found good agreement between the validation dataset and ANN predictions, within typical observational uncertainties. We found that the largest errors lay at the boundaries of the training data and in areas sparsely populated by the grid. This is apparent in Figure 3 where we plot the validation error against each parameter – for example, the spread in error increasing at high temperatures. Otherwise, the accuracy is very good within the observed range covered by our sample of 81 dwarfs and subgiants. Hence, we chose the median absolute deviation (MAD) as an estimator of the spread in error, because it is less sensitive to outliers than the standard deviation.

To represent the accuracy of the ANN, we present the median,  $\mu_{1/2}$  and MAD estimator,  $\sigma_{\text{MAD}} = 1.4826 \cdot \text{median}(|x_{\text{true}} - x_{\text{pred}}|)$  of the error in Table 3. The median is close to zero for all parameters, showing little systematic bias in the ANN. The MAD is also lower than observational uncertainties quoted in Section 2. Although the error in  $\Delta\nu$  is  $\sim 0.1 \mu\text{Hz}$  is comparable to observations with the best

signal-to-noise, this error is random throughout the validation data and should not produce any systematic bias.

### 3.3 Statistical models

We devised three Bayesian models, each with varying levels of parameter sharing (pooling) between stars in the population. Initially, we tested the models and demonstrated reduction of statistical uncertainties in the stellar fundamental parameters by analysing a random sample of 100 stars modelled using MESA. Then, we applied the models to the sample of stars in Table 1 (with and without Solar data for two of the models) and compared the results with that of S17.

Our first model was equivalent to modelling each star individually and featured no pooling; henceforth, we refer to it as the no-pooled (NP) model (see Section 3.3.1). We then derived two hierarchical Bayesian models (HBMs) which use population-level parameters to describe their distribution in the sample. Both of these models partially-pooled helium using a linear enrichment law. We drew the initial helium fraction for each star from a normal distribution with a mean described by the enrichment law and standard deviation representing its spread. Similarly, we partially-pooled the mixing-length theory parameter,  $\alpha_{\text{mlt}}$  in one model, whereas we maximally-pooled  $\alpha_{\text{mlt}}$  in the other, such that it assumes the same value for the entire sample. Hence, we refer to the former as the partial-pooled (PP) model and the latter as the max-pooled (MP) model, described in Sections 3.3.2 and 3.3.3 respectively.

#### 3.3.1 No-pooled model

Firstly, we constructed a model comprising independent parameters  $\boldsymbol{\theta}_i = \{f_{\text{evol},i}, M_i, \alpha_{\text{mlt},i}, Y_i, Z_i\}$  for a given star,  $i$ . Using Bayes' theorem, the *posterior* probability density function (PDF) of the model parameters given a set of observed data,  $\mathbf{d}_i$  is,

$$p(\boldsymbol{\theta}_i | \mathbf{d}_i) \propto p(\boldsymbol{\theta}_i) p(\mathbf{d}_i | \boldsymbol{\theta}_i), \quad (7)$$

where  $p(\boldsymbol{\theta}_i)$  is the *prior* PDF of the model parameters and  $p(\mathbf{d}_i | \boldsymbol{\theta}_i)$  is the *likelihood* of observing the data given the model.

We chose weakly-informative, bounded priors for the independent parameters, restricting them to their respective ranges in the ANN training data. Although the neural network is able to make predictions outside the training data range, these have not been tested and may be unreliable. Therefore, we used a beta distribution with  $\alpha = \beta = 1.2$  as the prior PDF on the independent parameters, transformed such that the probability is null outside the chosen range,

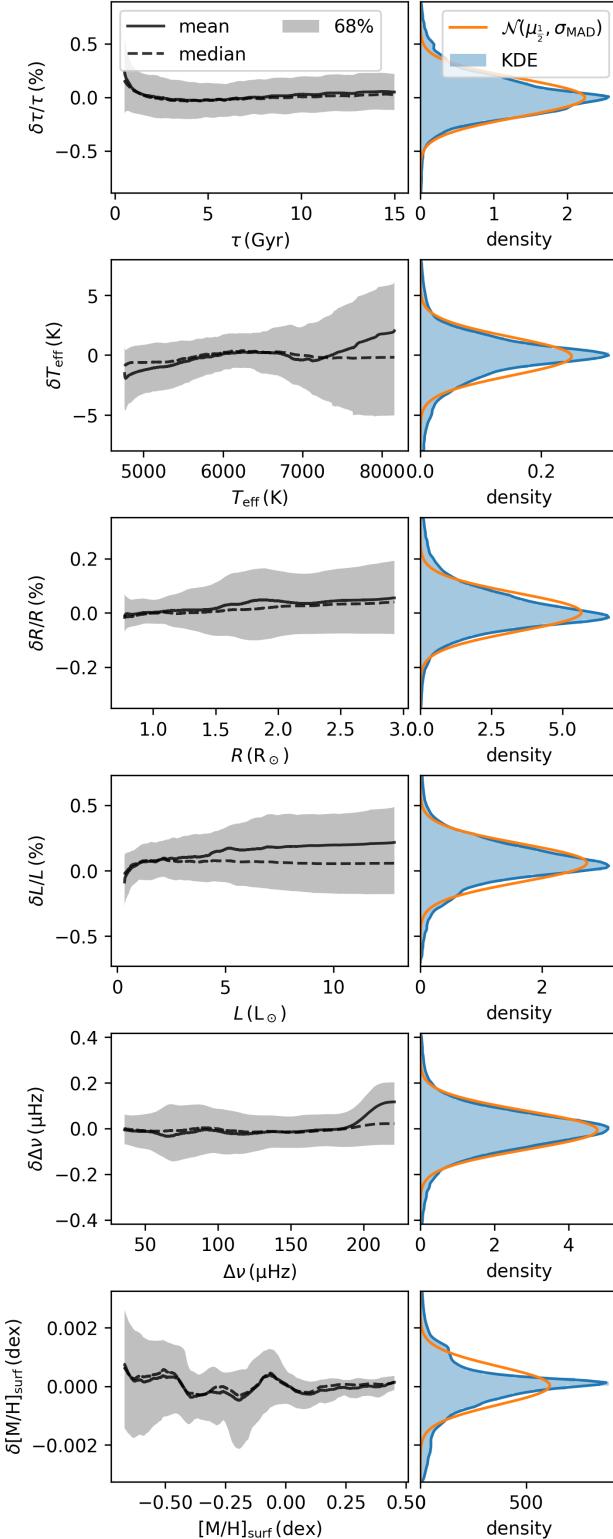
$$p(\boldsymbol{\theta}_i) = \prod_{k=1}^{N_\theta} [\theta_{k,\min} + (\theta_{k,\max} - \theta_{k,\min}) \mathcal{B}(\theta_{k,i} | 1.2, 1.2)], \quad (8)$$

where the beta distribution is defined as,

$$\mathcal{B}(x | \alpha, \beta) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du}. \quad (9)$$

The beta distribution was preferred over a bounded uniform distribution because our sampler evaluates the gradient of the posterior and hence sensitive to discontinuities.

Using notation which represents the distribution of some parameter  $x \sim g$  as equivalent to  $p(x) \propto g(x)$  where  $g(x)$  is some non-normalised probability density function, we write the priors



**Figure 3.** Left: the rolling error between the validation dataset (*true*) and the ANN predictions (*pred*) plotted against each parameter, where  $\delta x = x_{\text{pred}} - x_{\text{true}}$  for a given output  $x$ . Right: a kernel density estimate (KDE) of the validation error and a normal distribution centred on the median,  $\mu_{1/2}$  with an estimator for the standard deviation from the median absolute deviation,  $\sigma_{\text{MAD}}$ .

for  $\theta_i$  as,

$$f_{\text{evol},i} \sim 0.01 + 1.99 \cdot \mathcal{B}(1.2, 1.2),$$

$$M_i \sim 0.8 + 0.4 \cdot \mathcal{B}(1.2, 1.2),$$

$$\alpha_{\text{mlt},i} \sim 1.5 + \mathcal{B}(1.2, 1.2),$$

$$Y_{\text{init},i} \sim 0.24 + 0.08 \cdot \mathcal{B}(1.2, 1.2),$$

$$Z_{\text{init},i} \sim 0.005 + 0.035 \cdot \mathcal{B}(1.2, 1.2),$$

where each beta distribution is scaled to cover the boundaries of the grid of stellar models computed in Section 3.1.

We made predictions for each star using the trained ANN,  $\{\log(t_i), T_{\text{eff},i}, R_i, \Delta\nu_i, [\text{M}/\text{H}]_{\text{surf},i}\} = f_{\text{ANN}}(\theta_i)$ , from which we derived the luminosity,  $L_i$  using the Stefan-Boltzmann law. Any of the model parameters may be passed as an observable. Hereafter, we denote the set of model observables as  $\mu_i = f(\theta_i)$ . Thus, we write the likelihood we observe any  $d_i$  with known uncertainty,  $\sigma_i$  given the model as,

$$p(d_i|\theta_i) = \prod_{k=1}^{N_{\text{obs}}} \frac{1}{\sigma_{k,i} \sqrt{2\pi}} \exp \left[ -\frac{(d_{k,i} - \mu_{k,i})^2}{2\sigma_{k,i}^2} \right], \quad (10)$$

where  $N_{\text{obs}}$  is the number of observed variables. We chose to use observed  $T_{\text{eff}}$ ,  $L$ ,  $\Delta\nu$  and  $[\text{M}/\text{H}]$  collated for our sample as described in Section 2.

It follows that the posterior PDF for a population of  $N_{\text{stars}}$  stars for the NP model is,

$$p(\Theta|D) = \prod_{i=1}^{N_{\text{stars}}} p(\theta_i|d_i), \quad (11)$$

where  $\Theta$  is the matrix of model parameters and  $D$  is the matrix of observables. A graphical depiction of this model can be seen inside the grey box of Figure 4, without the arrow connecting  $Z_{\text{init}}$  to  $Y_{\text{init}}$ .

### 3.3.2 Partial-pooled model

Sharing, or pooling parameters between stars in a population can improve the uncertainties on stellar fundamentals by encoding our prior knowledge of their distribution in a population. We constructed a hierarchical model, which builds upon the NP model by introducing population-level *hyperparameters*. Specifically, we chose to describe initial helium and  $\alpha_{\text{mlt}}$  by partially-pooling them.

We constructed the PP model such that each of the initial helium,  $Y_{\text{init}}$  and mixing-length theory parameter,  $\alpha_{\text{mlt}}$  are drawn from a common distribution characterised by the set of hyperparameters,  $\phi$ . Thus, Bayes' theorem becomes,

$$p(\phi, \Theta|D) \propto p(\phi) p(Y_{\text{init}}, \alpha_{\text{mlt}}|\phi) p(f_{\text{evol}}, M, Z) p(D|\Theta), \quad (12)$$

where  $\Theta$  is the same as in the NP model, i.e. each object-level parameter,  $\theta_j = \{\theta_{j,i}\}_{i=1}^{N_{\text{stars}}}$ , and  $\phi = \{\Delta Y/\Delta Z, Y_P, \sigma_Y, \mu_\alpha, \sigma_\alpha\}$ . The hyperparameters for  $Y_{\text{init}}$  comprise the helium enrichment ratio,  $\Delta Y/\Delta Z$ , primordial helium abundance fraction,  $Y_P$  and the spread in helium,  $\sigma_Y$ . The remaining hyperparameters for  $\alpha_{\text{mlt}}$  comprise the mean,  $\mu_\alpha$  and spread,  $\sigma_\alpha$ .

We assumed the initial helium and the mixing-length parameter are each drawn from a normal distribution characterised by a population mean and standard deviation. The probability of  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$  given  $\phi$  is,

$$p(Y_{\text{init}}, \alpha_{\text{mlt}}|\phi) = p(Y_{\text{init}}|\mu_Y, \sigma_Y) p(\alpha_{\text{mlt}}|\mu_\alpha, \sigma_\alpha), \quad (13)$$

where  $\mu_Y$  and  $Z_{\text{init}}$  is the mean initial helium fraction as described by the linear helium enrichment law,

$$\mu_Y = Y_P + \frac{\Delta Y}{\Delta Z} Z_{\text{init}}. \quad (14)$$

Therefore, we may write the prior PDF of initial helium given its population-level hyperparameters as,

$$p(Y_{\text{init}} | Z_{\text{init}}, \Delta Y / \Delta Z, Y_P, \sigma_Y) = \prod_{i=1}^{N_{\text{stars}}} \mathcal{N}(Y_{\text{init},i} | \mu_Y, \sigma_Y). \quad (15)$$

Similarly, for the second term of Equation 13, we chose to partially-pool the mixing-length parameter. We assume that convection in stars of a similar mass, evolutionary stage and area of the HR diagram may be approximated using a similar value of  $\alpha_{\text{mlt}}$ , but the accuracy of the mixing-length theory may vary from star-to-star. There is theoretical evidence for such a variation with [M/H],  $T_{\text{eff}}$  and  $\log g$  in 3D hydrodynamical stellar models (Magic et al. 2015; Viani et al. 2018). However, investigating such dependencies are beyond this scope of this paper. Given the small range of our sample, any such variation will be absorbed by the spread parameter,  $\sigma_\alpha$ . Therefore, we decided to describe the prior on  $\alpha_{\text{mlt}}$  as,

$$p(\alpha_{\text{mlt}} | \mu_\alpha, \sigma_\alpha) = \prod_{i=1}^{N_{\text{stars}}} \mathcal{N}(\alpha_{\text{mlt},i} | \mu_\alpha, \sigma_\alpha) \quad (16)$$

We gave all of the hyperparameters weakly informative priors, with the exception of  $Y_P$  for which we adopt a recent measurement of the primordial helium abundance from big band nucleosynthesis (BBN) as the mean (Pitrou et al. 2018), with a standard deviation representative of the range of values in the literature (Aver et al. 2015; Peimbert et al. 2016; Cooke & Fumagalli 2018). Hence, we assumed priors on the hyperparameters as follows,

$$\Delta Y / \Delta Z \sim 4.0 \cdot \mathcal{B}(1.2, 1.2),$$

$$Y_P \sim \mathcal{N}(0.247, 0.001),$$

$$\sigma_Y \sim \mathcal{LN}(0.01, 1.0),$$

$$\mu_\alpha \sim 1.5 + \mathcal{B}(1.2, 1.2),$$

$$\sigma_\alpha \sim \mathcal{LN}(0.1, 1.0),$$

where,  $x \sim \mathcal{LN}(m, \sigma)$  represents a random variable drawn from the log-normal distribution,

$$\mathcal{LN}(x|m, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{\ln(x/m)^2}{2\sigma^2}\right]. \quad (17)$$

We produced a PGM for the model, depicted in Figure 4. The hyperparameters are shown outside of the grey box containing the individual stellar parameters to represent the hierarchical aspect of the model.

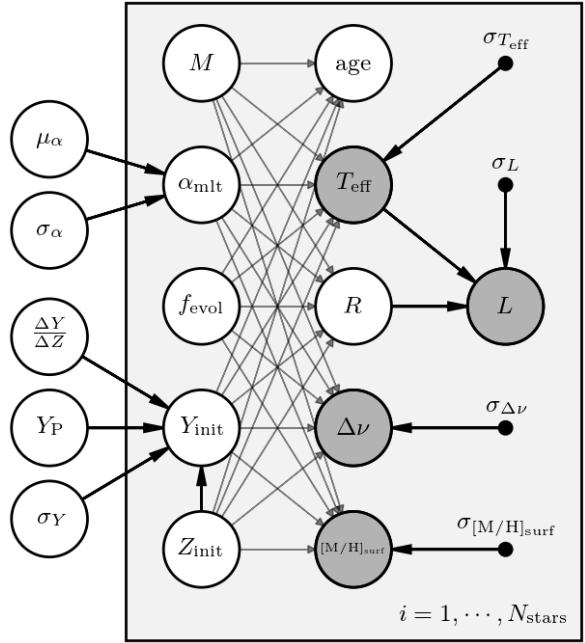
### 3.3.3 Max-pooled model

We built another hierarchical model similar to the PP model except that  $\alpha_{\text{mlt}}$  is max-pooled. In other words, we assumed that the mixing length must be the same value for every star in the sample, but still allowed it to freely vary. Thus the hyperparameters are now,  $\phi = \{\Delta Y / \Delta Z, Y_P, \sigma_Y, \alpha_{\text{mlt}}\}$ . The posterior distribution of the model takes the same form as in Equation 12 except that the mixing-length theory parameter for the  $i$ -th star is,

$$\alpha_{\text{mlt},i} = \alpha_{\text{mlt}}, \quad (18)$$

where,

$$\alpha_{\text{mlt}} \sim 1.5 + \mathcal{B}(1.2, 1.2). \quad (19)$$



**Figure 4.** A probabilistic graphical model (PGM) of the partially-pooled (PP) hierarchical model. Nodes outside of the grey rectangle represent the hyperparameters in the model. Nodes inside the grey rectangle represent individual stellar parameters. Dark grey nodes represent observables which each have their respective observational uncertainties given by the solid black nodes. The direction of the arrows represent the dependencies in the generative model.

In other words,  $\alpha_{\text{mlt}}$  is a free parameter in the model but is assumed to be the same in all stars.

### 3.4 Solar calibrator

Pooling parameters in an HBM allows us to use the Sun as a calibrator in a unique way. Rather than calibrating our model physics to the Sun and then assuming the calibrated parameters across our sample, we can include the Sun as a part of the same population. In other words, if we assume  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$  for the Sun are a part of the same prior distribution as for the sample, then we can add Solar data to our model.

For both the PP and MP models, we ran versions with and without data for the Sun included in the APOKASC population, referred to as PPS and MPS respectively. We adopted standard solar data as shown in Table 4 with uncertainties conservatively limited to the accuracy of the neural network. We also adopted  $\Delta\nu = 135.1 \pm 0.2 \mu\text{Hz}$  from Huber et al. (2011) with a standard deviation representative of variations in measurements of the solar  $\Delta\nu$  (Broomhall et al. 2011).

## 4 RESULTS

We obtained results for the models described above by sampling the posterior using the NUTS algorithm implemented in PyMC4 – a new version of PyMC3 (Salvatier et al. 2016) based on TensorFlow (Abadi et al. 2016). We took 20,000 samples split across 10 MCMC chains and computed summary statistics for each parameter in the

**Table 4.** Solar input data.

Input	$\mu$	$\sigma$	Reference
$M (M_\odot)$	1.000	0.001	—
$\tau (\text{Gyr})$	4.6	0.1	Connelly et al. (2012)
$T_{\text{eff}} (\text{K})$	5777	20	Scott et al. (2015)
$R (R_\odot)$	1.000	0.001	—
$L (L_\odot)$	1.00	0.01	—
$\Delta\nu (\mu\text{Hz})$	135.1	0.2	Huber et al. (2011)
[M/H] <sub>surf</sub> (dex)	0.00	0.01	Asplund et al. (2009)

model. Stars which were identified to be less than  $1-\sigma$  from the boundaries of the prior were dropped from the sample and the model was rerun. We also dropped stars with problems during tuning where appropriate, using the Gelman-Rubin diagnostic ( $\hat{r}$ ; Gelman & Rubin 1992). We settled from results from each run when  $\hat{r} < 1.04$  for all parameters.

Firstly, we created a random synthetic population of stars using MESA to test ability of the method to recover stellar properties according to our choice of model physics and population priors. We tested the NP, PP and MP models. Since our sample was fictitious, it would not have been appropriate to include solar data. We summarise the results for the synthetic stellar parameters and hyperparameters in Appendix B.

Then, we ran the models with the APOKASC sample collated in Section 2. Here, we included runs for the PPS and MPS models to test the effect of the addition of the Sun to the population. We summarise the results for the APOKASC sample in Section 4.1.

#### 4.1 The APOKASC sample

In this section, we present the results for each of the NP, PP and MP models ran with the sample of 81 APOKASC dwarfs and subgiants as inputs. We also present the results for the PPS and MPS models ran with Solar data as a calibrator. We show the reduction in age, mass and radius uncertainty with the addition of pooling in Section 4.1.1. We then show the results for model hyperparameters in Section 4.1.2 where we infer the initial helium abundance and mixing-length parameter distribution in the sample.

##### 4.1.1 Stellar parameter results

After an initial run of the NP model, 16 out of the 81 stars were dropped. The posteriors of 6 of the removed stars were skewed towards the upper mass limit of  $1.2 M_\odot$ , and the remaining 10 stars suffered poor convergence during sampling. In Table 5, we preset results for the remaining 65 stars using the NP model.

Running the NP model with synthetic stars resulted in unreliable uncertainties (see Appendix B). We saw the same boundary effects in the posteriors for  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$  here. Therefore, we present the NP results only for comparison purposes, but we exclude it from further discussion.

Out of the 65 stars sampled by the NP model, 2 stars were dropped with problems during convergence for the PP model. In Tables 6 and 7 we present the results for the remaining 63 stars from the PP and PPS model respectively. We did not see the same convergence problems with the MP and MPS models, hence we present results for the 65 stars in Tables 8 and 9 respectively.

Figure 5 compares the uncertainties in mass, radius and age for each of the models in this work, with those of the results for

the same stars in S17. We saw a similar improvement in uncertainty between the NP and pooled models as with the synthetic stars in Figure B1. We found that the statistical uncertainties in mass from the pooled models were reduced by a factor of  $\sim 2$  over S17 with a median of 2.5 per cent. We also obtained smaller uncertainties in radius and age of 1.2 and 12 per cent respectively when pooling the stellar parameters.

##### 4.1.2 Population parameter results

We obtained values for the hyperparameters for each of the models and present them in Table 10 along with their upper and lower 68 per cent credible regions. We omitted the results for  $Y_P$  because its posterior was the same as the prior,  $Y_P = 0.247 \pm 0.001$  for all the models.

We fit the same hyperparameters from the PP model to the NP model results for the purpose of comparison. However, the NP model results suffered from boundary effects which made the resulting fit unreliable, so it is left here only for comparison.

Figure 6 shows the joint and marginal distributions (corner plot) output by the PP and PPS model. We saw an anti-correlation between  $\Delta Y/\Delta Z$  and  $\mu_\alpha$ , expected due to the degeneracy between the two parameters in the stellar evolutionary models. In Figure 7, we also show the corner plot for the MP and MPS model output. Similarly, we see an anti-correlation between  $\Delta Y/\Delta Z$  and  $\alpha_{\text{mlt}}$ .

We present 100 random samples from the posterior for the helium enrichment relation from the PPS model in Figure 8. In this figure, we also plot the individual results for  $Y_{\text{init}}$  and  $Z_{\text{init}}$  for each of the stars in the sample. This is an example of *shrinkage* in the HBM; the estimates for individual stellar parameters move towards the mean of the population.

## 5 DISCUSSION

So far, we have shown that we can add parameters to stellar models without sacrificing statistical uncertainties through the application of an HBM. We freed the initial helium abundance and mixing-length theory parameter, using pooling to encode our prior knowledge of their distribution in the population. We also tested the impact of including the Sun in our population as a calibrator. We first discuss the impact of pooling and our choice of population priors for  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$  in Section 5.1 and 5.2.

To assess the accuracy of our model with respect to the literature, we compare our results to those of S17 in Section 5.3. We found good agreement between this work and their results, despite some differences in observables and stellar model physics which we discuss further.

Finally, we discuss sources of systematic uncertainties in Section 5.4. Although we have accounted for uncertainties in 5.1 and 5.2 in our model, there are still differences between stellar modelling codes and other model physics which should be considered. In Section 5.5, we highlight an outlier in our dataset. Then, we discuss the future scalability of this method in Section 5.6.

### 5.1 Helium enrichment

We found the value for the helium enrichment ratio,  $\Delta Y/\Delta Z$  to be the same in both the PP and MP models,  $\Delta Y/\Delta Z = 1.6^{+0.5}_{-0.4}$ . This is consistent with values of  $\sim 1.4$  in the literature for stellar models which include heavy element diffusion (Brogaard et al. 2012; Verma et al. 2019).

**Table 5.** The median of the marginalised posterior samples for each parameter output by the NP model, with their respective upper and lower 68 per cent confidence intervals. For the full table, see online.

Name	$f_{\text{evol}}$	$M$ ( $M_{\odot}$ )	$Y_{\text{init}}$	$Z_{\text{init}}$	[M/H] <sub>init</sub> (dex)	$\tau$ (Gyr)	$T_{\text{eff}}$ (K)	$R$ ( $R_{\odot}$ )	$\Delta\nu$ ( $\mu\text{Hz}$ )	[M/H] <sub>surf</sub> (dex)
KIC10079226	$0.44^{+0.16}_{-0.20}$	$1.14^{+0.04}_{-0.04}$	$0.28^{+0.02}_{-0.02}$	$0.021^{+0.003}_{-0.003}$	$0.21^{+0.07}_{-0.06}$	$2.5^{+1.2}_{-1.3}$	$5990^{+51}_{-52}$	$1.16^{+0.01}_{-0.02}$	$116.0^{+0.7}_{-0.7}$	$0.16^{+0.07}_{-0.07}$
KIC10215584	$0.50^{+0.21}_{-0.21}$	$1.13^{+0.04}_{-0.05}$	$0.27^{+0.02}_{-0.02}$	$0.018^{+0.002}_{-0.002}$	$0.15^{+0.06}_{-0.06}$	$2.9^{+1.6}_{-1.3}$	$5949^{+64}_{-65}$	$1.18^{+0.02}_{-0.02}$	$112.5^{+2.6}_{-2.6}$	$0.08^{+0.06}_{-0.06}$
KIC10319352	$1.51^{+0.15}_{-0.28}$	$1.09^{+0.05}_{-0.05}$	$0.28^{+0.03}_{-0.02}$	$0.028^{+0.004}_{-0.003}$	$0.34^{+0.06}_{-0.06}$	$9.6^{+1.7}_{-1.5}$	$5507^{+57}_{-56}$	$1.49^{+0.03}_{-0.03}$	$78.6^{+1.6}_{-1.6}$	$0.28^{+0.06}_{-0.06}$
KIC10322381	$0.98^{+0.23}_{-0.28}$	$1.07^{+0.07}_{-0.07}$	$0.28^{+0.02}_{-0.02}$	$0.010^{+0.002}_{-0.002}$	$-0.11^{+0.08}_{-0.08}$	$4.7^{+1.5}_{-1.7}$	$6132^{+94}_{-94}$	$1.39^{+0.05}_{-0.04}$	$85.6^{+4.9}_{-4.6}$	$-0.30^{+0.07}_{-0.08}$
KIC10732098	$1.60^{+0.14}_{-0.19}$	$1.12^{+0.04}_{-0.05}$	$0.28^{+0.02}_{-0.02}$	$0.017^{+0.002}_{-0.002}$	$0.13^{+0.06}_{-0.07}$	$6.7^{+0.8}_{-0.8}$	$5720^{+67}_{-66}$	$1.77^{+0.04}_{-0.04}$	$62.1^{+1.8}_{-1.7}$	$0.06^{+0.07}_{-0.07}$

**Table 6.** The same as Table 5, but for the PP model.

Name	$f_{\text{evol}}$	$M$ ( $M_{\odot}$ )	$Y_{\text{init}}$	$Z_{\text{init}}$	[M/H] <sub>init</sub> (dex)	$\tau$ (Gyr)	$T_{\text{eff}}$ (K)	$R$ ( $R_{\odot}$ )	$\Delta\nu$ ( $\mu\text{Hz}$ )	[M/H] <sub>surf</sub> (dex)
KIC10079226	$0.22^{+0.10}_{-0.09}$	$1.16^{+0.02}_{-0.03}$	$0.28^{+0.01}_{-0.01}$	$0.020^{+0.003}_{-0.002}$	$0.19^{+0.06}_{-0.06}$	$1.2^{+0.6}_{-0.5}$	$5962^{+44}_{-42}$	$1.17^{+0.01}_{-0.01}$	$115.9^{+0.7}_{-0.7}$	$0.15^{+0.07}_{-0.07}$
KIC10215584	$0.37^{+0.15}_{-0.13}$	$1.14^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.14^{+0.06}_{-0.06}$	$2.1^{+1.0}_{-0.8}$	$5941^{+57}_{-56}$	$1.18^{+0.02}_{-0.02}$	$112.5^{+2.6}_{-2.7}$	$0.07^{+0.07}_{-0.07}$
KIC10319352	$1.41^{+0.11}_{-0.27}$	$1.08^{+0.03}_{-0.03}$	$0.29^{+0.02}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.35^{+0.06}_{-0.07}$	$8.6^{+1.1}_{-1.0}$	$5512^{+45}_{-46}$	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.6}_{-1.6}$	$0.28^{+0.06}_{-0.07}$
KIC10322381	$0.78^{+0.23}_{-0.19}$	$1.14^{+0.03}_{-0.06}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.002}$	$-0.07^{+0.06}_{-0.07}$	$3.6^{+1.7}_{-1.1}$	$6081^{+95}_{-92}$	$1.41^{+0.05}_{-0.05}$	$86.2^{+4.8}_{-5.2}$	$-0.31^{+0.07}_{-0.07}$
KIC10732098	$1.50^{+0.13}_{-0.14}$	$1.14^{+0.03}_{-0.04}$	$0.28^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.15^{+0.06}_{-0.07}$	$6.4^{+0.6}_{-0.6}$	$5701^{+59}_{-58}$	$1.78^{+0.03}_{-0.03}$	$62.2^{+1.7}_{-1.7}$	$0.06^{+0.06}_{-0.06}$

**Table 7.** The same as Table 5, but for the PPS model.

Name	$f_{\text{evol}}$	$M$ ( $M_{\odot}$ )	$Y_{\text{init}}$	$Z_{\text{init}}$	[M/H] <sub>init</sub> (dex)	$\tau$ (Gyr)	$T_{\text{eff}}$ (K)	$R$ ( $R_{\odot}$ )	$\Delta\nu$ ( $\mu\text{Hz}$ )	[M/H] <sub>surf</sub> (dex)
KIC10079226	$0.35^{+0.11}_{-0.12}$	$1.17^{+0.02}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.020^{+0.003}_{-0.002}$	$0.20^{+0.06}_{-0.06}$	$2.1^{+0.8}_{-0.8}$	$5962^{+44}_{-43}$	$1.17^{+0.01}_{-0.01}$	$116.0^{+0.7}_{-0.7}$	$0.15^{+0.06}_{-0.07}$
KIC10215584	$0.47^{+0.16}_{-0.16}$	$1.14^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.14^{+0.06}_{-0.06}$	$2.7^{+1.2}_{-1.1}$	$5943^{+56}_{-58}$	$1.18^{+0.02}_{-0.02}$	$112.6^{+2.6}_{-2.6}$	$0.07^{+0.06}_{-0.07}$
KIC10319352	$1.51^{+0.10}_{-0.22}$	$1.09^{+0.03}_{-0.03}$	$0.28^{+0.01}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.34^{+0.06}_{-0.06}$	$9.6^{+1.1}_{-1.2}$	$5507^{+47}_{-48}$	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.6}_{-1.6}$	$0.28^{+0.06}_{-0.06}$
KIC10322381	$0.89^{+0.21}_{-0.22}$	$1.12^{+0.05}_{-0.06}$	$0.26^{+0.01}_{-0.01}$	$0.010^{+0.002}_{-0.002}$	$-0.10^{+0.06}_{-0.07}$	$4.3^{+1.7}_{-1.2}$	$6093^{+92}_{-89}$	$1.41^{+0.04}_{-0.04}$	$86.1^{+5.0}_{-4.9}$	$-0.31^{+0.07}_{-0.08}$
KIC10732098	$1.60^{+0.11}_{-0.14}$	$1.14^{+0.03}_{-0.04}$	$0.27^{+0.01}_{-0.01}$	$0.017^{+0.002}_{-0.002}$	$0.13^{+0.06}_{-0.07}$	$6.9^{+0.6}_{-0.6}$	$5704^{+62}_{-61}$	$1.78^{+0.04}_{-0.03}$	$62.2^{+1.8}_{-1.7}$	$0.06^{+0.06}_{-0.06}$

When we added the Sun to the pooled models, PPS and MPS, we obtained lower values of  $\Delta Y/\Delta Z$  with increased precision. In both models, the resulting  $\Delta Y/\Delta Z$  of approximately 0.8 to 1.0 was consistent with the initial helium fraction expected from solar models with our choice of Asplund et al. (2009) abundances (Serenelli & Basu 2010). However, such solar models have been shown to not recover helioseismic measurements of helium in the Sun (Basu & Antia 2004; Serenelli et al. 2009; Villante et al. 2014). Solar models with the older Grevesse & Sauval (1998) abundances typically yield higher helium fractions more in-line with helioseismology. Interestingly, the  $\Delta Y/\Delta Z$  from the PP and MP models are higher than those with the Sun, in closer agreement to those calibrated with the older abundances. We would need to extend our model to predict surface helium abundances and compare these with asteroseismic measurements indicators of helium to test whether this difference is significant.

We found little difference in the helium dispersion,  $\sigma_Y$  between all our pooled models. The main difference was the greater  $\sigma_Y \approx 0.007$  when fitting the enrichment law to the results of the NP model. This is an example of the hierarchical models pulling individual  $Y_{\text{init}}$  together, enabling a better estimate of the population spread. Since there was little variation in  $\sigma_Y$  among hierarchical models, we can conclude that the spread of  $Y_{\text{init}}$  about a linear helium enrichment law in our population of *Kepler* dwarfs and subgiants is  $\sigma_Y = 0.005^{+0.004}_{-0.003}$ .

Our results assume a linear helium enrichment law dependent only on the initial heavy element abundance,  $Z_{\text{init}}$ . However, helium in the galaxy could vary non-linearly and depending on other chemical abundances (West & Heger 2013) or the location of the

star in the Milky Way (Frebel 2010). Our model has the advantage of being adaptable to different population priors, stellar inputs and outputs. Future work will explore the helium enrichment relation further, with the inclusion of metal-poor stars and a dependence on different chemical abundances.

Our models assume a prior of  $Y_P = 0.247 \pm 0.001$  for the primordial helium fraction which dominated its posterior. This is a sensible assumption to make when assuming a linear enrichment law, because measurements of the primordial helium correspond to the fraction at the epoch of BBN according to current cosmological theory (Cyburt et al. 2016). However, if we used a less informative prior for  $Y_P$  we might yield more uncertain results, or a different value for  $Y_P$ . In prior work fitting a linear enrichment law, some results for  $Y_P$  suggested a value below the BBN value (Casagrande et al. 2007; Silva Aguirre et al. 2017). It is more probable that the assumption of a linear enrichment law is inaccurate than a sample of stars could contradict independent  $Y_P$  from cosmology. Therefore, we justify our prior on  $Y_P$  as in-line with the assumption of a linear enrichment law, but highlight the need to investigate other ways of describing helium in a population of stars.

## 5.2 Mixing-length theory

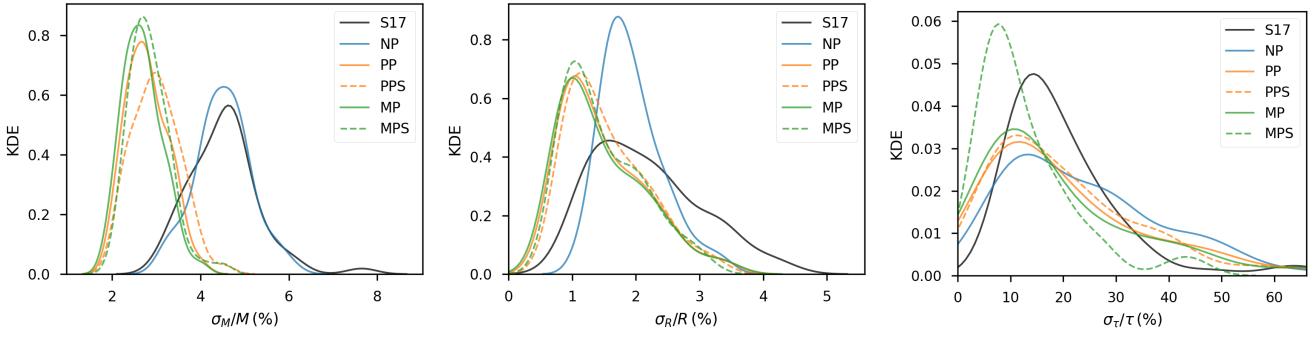
To a greater degree than chemical composition, the best-fitting  $\alpha_{\text{mlt}}$  depends on the choice of model physics and stellar modelling code. The mixing-length theory is an approximation of convection which is often calibrated to the Sun and then assumed for all stars in a model. However, studies of 3D hydrodynamical simulations suggest that the degree in which  $\alpha_{\text{mlt}}$  approximates convection varies across

**Table 8.** The same as Table 5, but for the MP model.

Name	$f_{\text{evol}}$	$M (M_{\odot})$	$Y_{\text{init}}$	$Z_{\text{init}}$	[M/H] <sub>init</sub> (dex)	$\tau$ (Gyr)	$T_{\text{eff}}$ (K)	$R (R_{\odot})$	$\Delta\nu (\mu\text{Hz})$	[M/H] <sub>surf</sub> (dex)
KIC10079226	$0.20^{+0.08}_{-0.08}$	$1.17^{+0.02}_{-0.03}$	$0.28^{+0.01}_{-0.01}$	$0.019^{+0.003}_{-0.002}$	$0.19^{+0.06}_{-0.06}$	$1.1^{+0.5}_{-0.4}$	$5961^{+42}_{-41}$	$1.17^{+0.01}_{-0.01}$	$115.9^{+0.7}_{-0.7}$	$0.15^{+0.06}_{-0.07}$
KIC10215584	$0.36^{+0.14}_{-0.13}$	$1.14^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.14^{+0.06}_{-0.06}$	$2.0^{+0.9}_{-0.8}$	$5941^{+57}_{-57}$	$1.18^{+0.02}_{-0.02}$	$112.5^{+2.6}_{-2.7}$	$0.07^{+0.06}_{-0.07}$
KIC10319352	$1.41^{+0.10}_{-0.25}$	$1.08^{+0.03}_{-0.03}$	$0.29^{+0.02}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.36^{+0.06}_{-0.07}$	$8.6^{+1.0}_{-0.9}$	$5512^{+44}_{-45}$	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.7}_{-1.6}$	$0.28^{+0.06}_{-0.07}$
KIC10322381	$0.77^{+0.23}_{-0.19}$	$1.14^{+0.04}_{-0.06}$	$0.27^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.002}$	$-0.07^{+0.06}_{-0.07}$	$3.5^{+1.6}_{-1.0}$	$6076^{+96}_{-91}$	$1.41^{+0.05}_{-0.05}$	$86.1^{+4.7}_{-5.3}$	$-0.32^{+0.07}_{-0.07}$
KIC10732098	$1.50^{+0.13}_{-0.13}$	$1.14^{+0.03}_{-0.04}$	$0.28^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.15^{+0.06}_{-0.07}$	$6.4^{+0.6}_{-0.6}$	$5702^{+56}_{-58}$	$1.78^{+0.03}_{-0.03}$	$62.2^{+1.7}_{-1.7}$	$0.06^{+0.06}_{-0.06}$

**Table 9.** The same as Table 5, but for the MPS model.

Name	$f_{\text{evol}}$	$M (M_{\odot})$	$Y_{\text{init}}$	$Z_{\text{init}}$	[M/H] <sub>init</sub> (dex)	$\tau$ (Gyr)	$T_{\text{eff}}$ (K)	$R (R_{\odot})$	$\Delta\nu (\mu\text{Hz})$	[M/H] <sub>surf</sub> (dex)
KIC10079226	$0.44^{+0.07}_{-0.06}$	$1.16^{+0.02}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.021^{+0.003}_{-0.002}$	$0.20^{+0.06}_{-0.06}$	$2.7^{+0.5}_{-0.4}$	$5965^{+40}_{-40}$	$1.17^{+0.01}_{-0.01}$	$116.0^{+0.7}_{-0.7}$	$0.15^{+0.06}_{-0.06}$
KIC10215584	$0.59^{+0.11}_{-0.13}$	$1.13^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.15^{+0.06}_{-0.06}$	$3.6^{+0.9}_{-0.9}$	$5952^{+55}_{-56}$	$1.18^{+0.02}_{-0.02}$	$112.7^{+2.7}_{-2.7}$	$0.08^{+0.06}_{-0.07}$
KIC10319352	$1.61^{+0.04}_{-0.06}$	$1.08^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.028^{+0.004}_{-0.003}$	$0.33^{+0.06}_{-0.06}$	$10.8^{+0.7}_{-0.8}$	$5516^{+46}_{-47}$	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.7}_{-1.6}$	$0.28^{+0.07}_{-0.06}$
KIC10322381	$0.98^{+0.19}_{-0.20}$	$1.10^{+0.06}_{-0.05}$	$0.26^{+0.01}_{-0.01}$	$0.010^{+0.002}_{-0.001}$	$-0.13^{+0.07}_{-0.07}$	$5.1^{+1.3}_{-1.5}$	$6106^{+94}_{-80}$	$1.40^{+0.04}_{-0.04}$	$85.8^{+5.6}_{-4.3}$	$-0.30^{+0.08}_{-0.08}$
KIC10732098	$1.69^{+0.06}_{-0.09}$	$1.14^{+0.03}_{-0.04}$	$0.26^{+0.01}_{-0.01}$	$0.017^{+0.002}_{-0.002}$	$0.12^{+0.06}_{-0.07}$	$7.4^{+0.5}_{-0.5}$	$5715^{+61}_{-61}$	$1.77^{+0.04}_{-0.03}$	$62.3^{+1.8}_{-1.8}$	$0.07^{+0.06}_{-0.07}$


**Figure 5.** Kernel density estimates (KDEs) of the uncertainties in the results from each model compared with that of (S17) for the sample of APOKASC dwarfs and subgiants.

the HR diagram (Magic et al. 2015) and this is confirmed when modelling stars with asteroseismology (Tayar et al. 2017).

We found more of a difference in the  $\alpha_{\text{mlt}}$  hyperparameter values between the models than with helium. The fit to the NP model results obtained a result for  $\mu_{\alpha}$  close to 2.0 – the midpoint of its prior distribution. This was expected because the tests on synthetic stars showed that  $\alpha_{\text{mlt}}$  was biased in this way due to boundary effects from truncating the distribution at 1.5 and 2.5.

The PP model (without the Sun) favoured a mean mixing-length parameter of  $\mu_{\alpha} \approx 1.7$ . Whereas, the PPS model yielded a higher value of  $\mu_{\alpha} \approx 1.9$ . We found this was attributed to the addition of the Sun. The individual solar results for the PPS model yielded a value of  $\alpha_{\text{mlt},\odot} = 2.11 \pm 0.03$  which was considerably higher than the  $\alpha_{\text{mlt}}$  obtained for the other stars in the sample (see Appendix C). The solar value also exceeds typical solar calibrated values of  $\sim 1.9$  for the same stellar evolution code (Stancliffe et al. 2016). The reason for this is unclear, except that the difference we see between between  $\alpha_{\text{mlt},\odot}$  and the rest of the sample is not unique. Previous work on the LEGACY sample of *Kepler* dwarfs found the best fitting  $\alpha_{\text{mlt}}$  for their sample approximately 90 per cent of the solar calibrated value for one of their pipelines with similar model physics to this work (Silva Aguirre et al. 2017).

Despite the difference in  $\mu_{\alpha}$ , the resulting spread in mixing-length for the PPS model  $\sigma_{\alpha} \approx 0.13$  is double that of the PP model

to cope with the high solar value. This implies that a large population spread in  $\alpha_{\text{mlt}}$  could explain the difference we see. In other words, if we assume that the best-fitting  $\alpha_{\text{mlt}}$  is normally distributed in our population, then the Sun lies within  $2\sigma$  of the mean, among 95 per cent of all stars in the population.

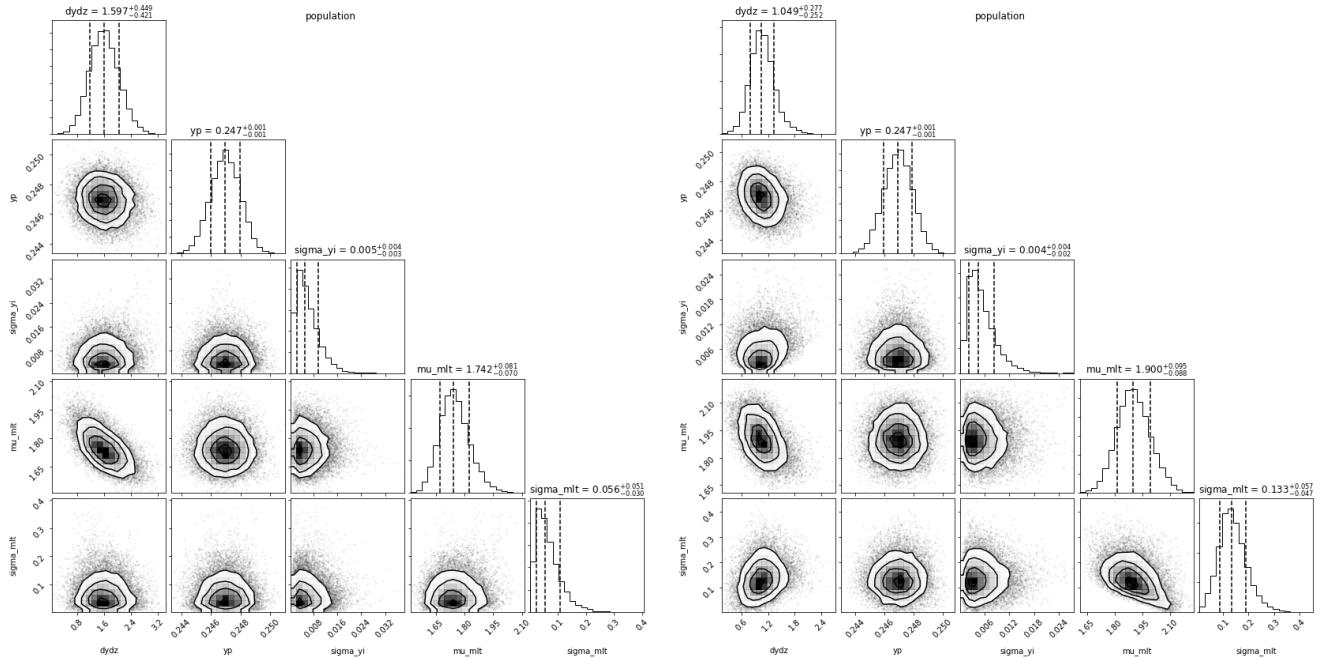
There are no prior studies which look at the spread in  $\alpha_{\text{mlt}}$  for a population of stars, but many examples which fit  $\alpha_{\text{mlt}}$  as a function of [M/H],  $T_{\text{eff}}$  and  $\log g$ . For example, results from Viani et al. (2018) for stellar models including diffusion, predict  $\alpha_{\text{mlt}}$  in the range 1.5 to 2.3 across our sample. This dispersion would be more compatible with the larger spread obtained by our PPS model. However, in future work we should further investigate how  $\alpha_{\text{mlt}}$  varies with stellar parameters, as our assumption of a normal distribution may not be accurate.

Additionally, we see the effect of adding the Sun increase when we max-pool  $\alpha_{\text{mlt}}$  in the MP and MPS models. The MP models yields a global  $\alpha_{\text{mlt}}$  in line with  $\mu_{\alpha}$  from the PP model. However, when we add the Sun, the model yields  $\alpha_{\text{mlt}} \approx 2.1$  which is in common with the solar results from the PPS model. This has a similar effect as assuming a solar calibrated value, because the model favours fitting to data with the best observed precision.

In all observables except for  $L$ , the Sun is near the centre of our distribution of stars. However, we found no relationship between  $L$  and  $\alpha_{\text{mlt}}$  in both our NP and PP models. A possible explanation

**Table 10.** Hyperparameter results for each model with the omission of  $Y_P$ .

Model	$\Delta Y/\Delta Z$	$\sigma_Y$	$\mu_\alpha$	$\sigma_\alpha$	$\alpha_{\text{mlt}}$
NP	$1.69^{+0.21}_{-0.21}$	$0.0074^{+0.0026}_{-0.0022}$	$1.954^{+0.040}_{-0.041}$	$0.065^{+0.030}_{-0.024}$	—
PP	$1.60^{+0.45}_{-0.42}$	$0.0051^{+0.0045}_{-0.0027}$	$1.742^{+0.081}_{-0.070}$	$0.056^{+0.051}_{-0.030}$	—
PPS	$1.05^{+0.28}_{-0.25}$	$0.0045^{+0.0038}_{-0.0023}$	$1.900^{+0.095}_{-0.088}$	$0.133^{+0.057}_{-0.047}$	—
MP	$1.60^{+0.45}_{-0.42}$	$0.0051^{+0.0044}_{-0.0027}$	—	—	$1.728^{+0.077}_{-0.066}$
MPS	$0.76^{+0.24}_{-0.27}$	$0.0049^{+0.0039}_{-0.0025}$	—	—	$2.088^{+0.031}_{-0.029}$

**Figure 6.** Corner plots showing the joint and marginalised sampled posterior distributions for the hyperparameters for the PP (left) and PPS (right) models. The vertical dashed lines give the 16th, 50th and 84th percentiles.

for the difference in  $\alpha_{\text{mlt}}$  with and without the Sun could be some systematic offset in our observational data for the sample. Here, we point to our choice of spectroscopic  $T_{\text{eff}}$  which typically underestimates  $T_{\text{eff}}$  compared to photometric scales. There is a positive correlation between  $\alpha_{\text{mlt}}$  and  $T_{\text{eff}}$ , when holding all other variables constant. Therefore, the lower  $\alpha_{\text{mlt}}$  obtained without the Sun as a calibrator could be caused by underestimated effective temperatures.

### 5.3 Comparison with APOKASC results

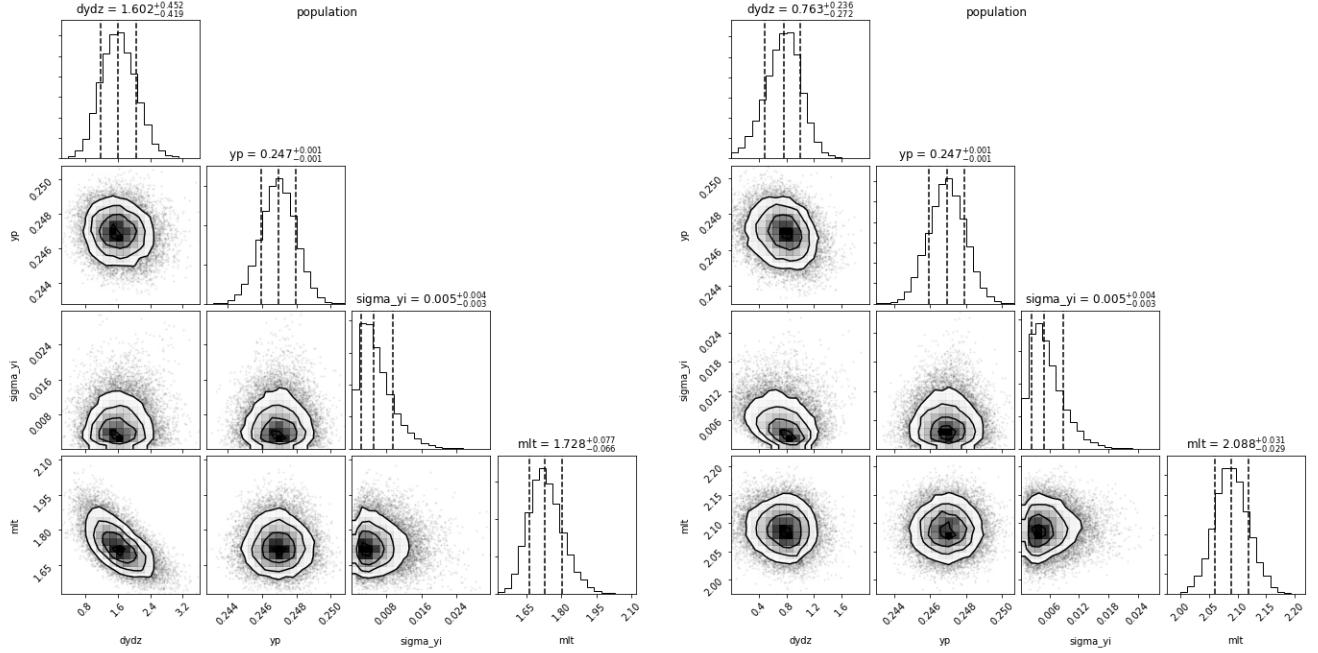
Before we compare our results to S17, we should highlight some key differences between our data and methodology. The results from S17 were determined using a grid-based-modelling technique, which estimates the likelihood across a dense grid of stellar models. They used results from several pipelines to estimate the systematic uncertainties. For the central values of their results, they used the Bayesian stellar algorithm (BASTA; Silva Aguirre et al. 2015) using a grid computed with GARSTEC (Weiss & Schlattl 2008). Their choice of stellar physics was similar to this work, except for two major differences.

Firstly, the results of S17 were determined using stellar models calculated without heavy-element diffusion. The inclusion of diffusion when modelling the Sun has been commonplace over the last few decades, with good agreement between models and helio-

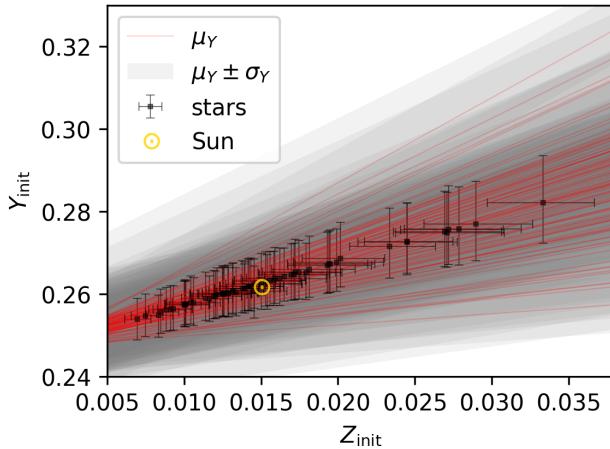
seismic observations (Christensen-Dalsgaard et al. 1993; Bahcall et al. 1995). More recent work explored the diffusion in cluster stars (Korn et al. 2007; Önehag et al. 2014) and another demonstrated the impact of including diffusion of stellar ages (Dotter et al. 2017). Our stellar models were computed with heavy-element diffusion. Recently, work by Nsamba et al. (2018) on a similar sample of stars showed models without diffusion compared to those including diffusion can lead to, on average, underestimated radii and masses and overestimated ages by 1, 3 and 16 per cent respectively.

Secondly, our choice of Asplund et al. (2009) solar chemical mixture differs from the Grevesse & Sauval (1998) mixtures adopted by S17. The former leads to a solar heavy-element to hydrogen ratio of  $(Z/X)_\odot = 0.0181$ , and the latter,  $(Z/X)_\odot = 0.0230$ . Typically, Grevesse & Sauval (1998) abundances are favoured in asteroseismic modelling because they are better able to reproduce measurements of helium in the Sun from helioseismology (Serenelli et al. 2009). An effect of using the Asplund et al. (2009) abundances, is that it favours lower  $Z_{\text{init}}$  for a given  $[\text{M}/\text{H}]_{\text{surf}}$ . As a result, models using Grevesse & Sauval (1998) abundances on average underestimate radii and mass compared to those without by about 1 and 0.5 per cent respectively (Nsamba et al. 2018).

Although updated, much of our observable data is comparable to that of S17, with the exception of  $T_{\text{eff}}$ . The preferred results from S17 were determined using a photometric  $T_{\text{eff}}$  scale which



**Figure 7.** The same as Figure 6 but for the MP (left) and MPS (right) models.



**Figure 8.** The results for initial helium fraction,  $Y_{\text{init}}$  against initial heavy-element fraction,  $Z_{\text{init}}$  for each star from the PPS model. 100 random samples from the posterior for the population mean,  $\mu_Y = Y_P + (\Delta Y / \Delta Z)Z_{\text{init}}$  and spread,  $\mu_Y \pm \sigma_Y$  are shown in red and grey respectively. The central location of the solar model is shown by the symbol,  $\odot$ .

we found to be on average  $\sim 170$  K greater than our spectroscopic scale from DR14. In S17, they saw a similar offset between the DR13  $T_{\text{eff}}$  available at the time. They found a median difference in mass, radius and age of approximately  $-6$ ,  $-2$  and  $+35$  per cent respectively with results from the photometric  $T_{\text{eff}}$  scale subtracted from the spectroscopic scale.

In the following subsections, we compare the results between our PPS model with that of S17 with reference to Figure 9.

### 5.3.1 Mass

In the left panel of Figure 9, we compare the masses obtained by the PPS model with S17 and found a dispersion of around 2 per cent. Our masses were on average 1 per cent above the results from S17. Although we might expect the lower  $T_{\text{eff}}$  scale in this work to underestimate the mass, we attribute this overall effect to our choice of stellar model physics. As previously discussed, the use of Asplund et al. (2009) solar abundances and heavy-element diffusion has the cumulative effect of overestimating stellar masses compared to the physics adopted by S17. We also found that the results from all the pooled models returned similar masses, with or without the Sun.

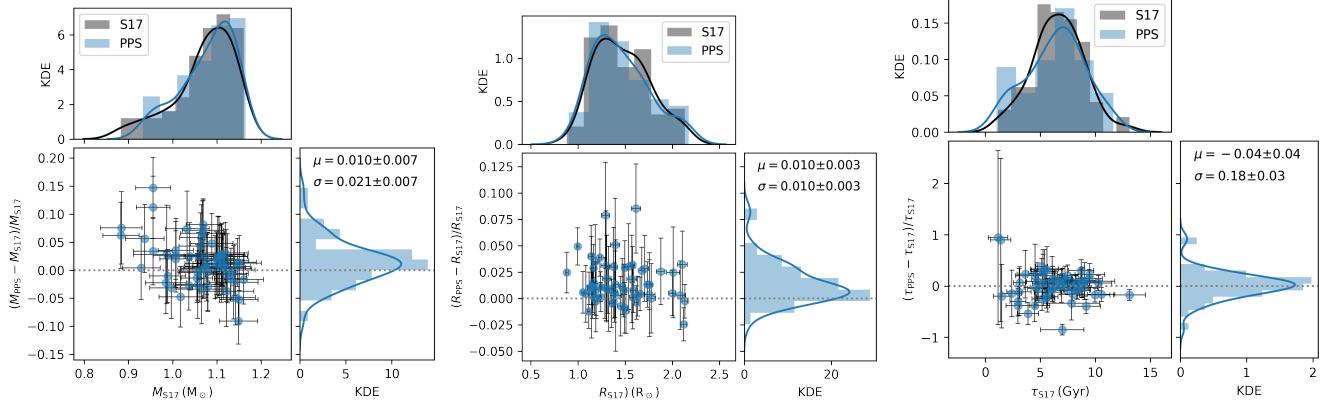
### 5.3.2 Radius

In the central panel of Figure 9, we show that our radii were similar to S17 with a spread of 1 per cent. We also found radii on average 1 per cent greater than the APOKASC results. Similarly to with mass, this contradicts what would be expected from a lower  $T_{\text{eff}}$  scale and could also be explained by model physics. Our radii also varied little between models with and without the Sun.

### 5.3.3 Age

Our ages were more consistent with S17 than our masses and radii. The right-most panel of Figure 9 shows the spread in the relative age differences to be about 18 per cent, slightly underestimated by 4 per cent. We would expect the lower  $T_{\text{eff}}$  scale to overestimate the ages as found in S17, but instead they are comparable. However, as discussed previously, including diffusion has been shown to reduce age estimates compared to those without.

Including the Sun in our pooled models affected the resulting ages more than mass and radius. Including the Sun typically overestimated the ages compared to models without the Sun. This



**Figure 9.** The mean and standard deviation in age, mass and radius results from the PPS model compared with the results (using the photometric temperature scale) from S17.

is expected given the higher  $\alpha_{\text{mlt}}$  for the models including solar data, because a larger mixing-length leads to more efficient nuclear burning and more time spent during the main sequence phase.

#### 5.4 Systematic uncertainties

We have already accounted for systematics due to the choice of helium enrichment and mixing-length parameter by marginalising over their uncertainties assuming their population distributions. However, there are other model physics which we have not freely varied, including diffusion and choice of solar mixture. Although our method can be adapted to different stellar evolutionary codes and choice of physics, an in-depth analysis of systematic uncertainties is left to future work.

In previous work studying stars in the APOKASC sample, several pipelines used a range of stellar evolutionary codes and model physics are employed to evaluate systematic uncertainties from the models (Serenelli et al. 2017; Silva Aguirre et al. 2017). Using a hierarchical model in this work enabled us to reduce median statistic uncertainties to 2.5 per cent in mass, 1.2 per cent in radius and 12 per cent in age. The systematic uncertainty analysis of S17 found median systematics of 3, 1 and 13 per cent in mass, radius and age respectively. Reducing statistical uncertainties highlights the importance of understanding systematics.

Other systematics could arise from observational data. For example, we chose the ASPCAP DR14  $T_{\text{eff}}$  scale which was systematically lower than the photometric scale of choice in S17. However, our method was still able to recover similar masses, radii and ages. This could be explained by our choice of stellar model physics, as discussed previously.

#### 5.5 Outliers

We identified KIC 9025370 as a possible outlier. Consistent across all our models, its output effective temperature,  $T_{\text{eff}} = 5934 \pm 50$  K was about 4- $\sigma$  greater than its observed  $T_{\text{eff}}$ , and its modelled  $L$  was about 2- $\sigma$  dimmer than its observed luminosity. Only  $\Delta\nu$  and  $[\text{M}/\text{H}]_{\text{surf}}$  were consistent between modelled and observed values. The difference was also apparent in our comparison of ages with S17 where we obtained an age of  $1.5^{+0.7}_{-0.6}$  Gyr compared to their value of  $7.0^{+2.0}_{-1.6}$  Gyr.

KIC 9025370 turned out to be a double-lined spectroscopic

binary (Nissen et al. 2017), discovered after S17 and hence included in the original sample. The brighter observed luminosity and possibly unreliable spectroscopic  $T_{\text{eff}}$  compared to our models were consistent with a spectroscopic binary. We calculated a photometric  $T_{\text{eff}}$  using the IRFM method (Casagrande et al. 2010) with the available 2MASS photometry for the target and obtained  $T_{\text{eff}} = 5983 \pm 120$  K, more consistent with our modelled effective temperature and inconsistent with its spectroscopic  $T_{\text{eff}}$ . Therefore, we left KIC 9025370 in our results as an example of the robustness of our model.

#### 5.6 The potential of pooling

Using HBMs has allowed us to introduce more free parameters without sacrificing statistical uncertainties. We used an ANN to approximate stellar models, a method which can be extended to higher input dimensions with little impact on training and evaluation time. Our model also scales well with the number of stars, making use of GPU parallel processing when sampling the posterior.

As shown in tests with synthetic stars (Appendix B) and apparent in Figure 5, increasing the number of stars decreases the statistical uncertainties when parameters are pooled. The theoretical limit to this improvement is  $\sqrt{N_1/N_2}$  for two populations of size  $N_1$  and  $N_2$ . For example, if we increase our sample to 300 stars, we would expect the uncertainties to reduce by up to a factor of 2. Naturally, the uncertainty is still limited by observational precision. However, hierarchical modelling as demonstrated in this work, allows us to get the most out of our data and paves the way for a data-driven analysis of model systematics.

Including all-sky data from *TESS* and in anticipation of *Plato* (Rauer et al. 2014) we can expect our sample size of asteroseismic dwarfs and subgiants only to increase. There is also scope to extend our grid of models to include red giants, for which there are vast catalogues of stars already studied with *Kepler* (Pinsonneault et al. 2018).

## 6 CONCLUSIONS

We have shown that modelling  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$  to improve inference of fundamental parameters can be done through the use of an HBM, whilst still improving statistical uncertainties. Our results were in

good agreement with S17 with small changes in mass and radii expected from our choice of model physics and updated observables. Taking our partially-pooled model including the Sun (PPS) as our preferred set of results, we obtained median statistical uncertainties on  $M$ ,  $R$  and  $\tau$  of 2.5, 1.2 and 12 per cent respectively. Furthermore, we demonstrated that the uncertainties reduced with increasing sample size in a population of synthetic stars, giving scope to further improve our inference on larger sample sizes from *TESS*.

We found that the gradient,  $\Delta Y/\Delta Z$  of the linear helium enrichment law ranged from 0.8 to 1.6 depending on the level of parameter pooling and the inclusion of the Sun in our sample, with  $\Delta Y/\Delta Z = 1.1^{+0.3}_{-0.3}$  from our preferred PPS model. Consistent across our models was the spread in initial helium about the enrichment law,  $\sigma_Y = 0.005^{+0.004}_{-0.003}$ . We also found the spread in  $\alpha_{\text{mlt}}$  doubled to  $\sigma_\alpha = 0.13^{+0.06}_{-0.05}$  to account for the addition of the Sun in our sample. We conclude that there are still discrepancies between the best-fitting  $\alpha_{\text{mlt}}$  in our population and that of the Sun which need to be investigated further. Perhaps, the addition of asteroseismic signatures of helium abundance (see e.g. Verma et al. 2017) would improve our constraints on  $Y_{\text{init}}$  and thus reduce star-by-star uncertainties in  $\alpha_{\text{mlt}}$ .

## ACKNOWLEDGEMENTS

This work is a part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (CartographY; grant agreement ID 804752).

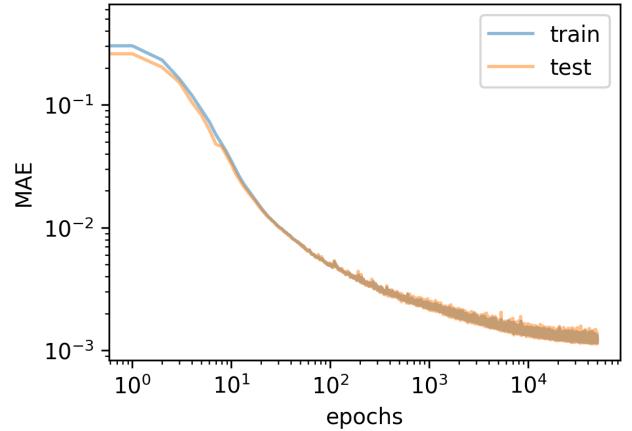
## REFERENCES

- Abadi M., et al., 2016, arXiv e-prints, 1605, arXiv:1605.08695  
 Albareti F. D., et al., 2017, *The Astrophysical Journal Supplement Series*, 233, 25  
 Anderson L., Hogg D. W., Leistedt B., Price-Whelan A. M., Bovy J., 2018, *AJ*, 156, 145  
 Appourchaux T., et al., 2015, *Astronomy and Astrophysics*, 582, A25  
 Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481  
 Aver E., Olive K. A., Skillman E. D., 2015, *Journal of Cosmology and Astroparticle Physics*, 07, 011  
 Baglin A., et al., 2006, in 36th COSPAR Scientific Assembly. p. 3749  
 Bahcall J. N., Pinsonneault M. H., Wasserburg G. J., 1995, *Reviews of Modern Physics*, 67, 781  
 Ball W. H., Gizon L., 2014, *Astronomy and Astrophysics*, 568, A123  
 Balser D. S., 2006, *AJ*, 132, 2326  
 Basu S., Antia H. M., 2004, *The Astrophysical Journal Letters*, 606, L85  
 Bellinger E. P., Angelou G. C., Hekker S., Basu S., Ball W. H., Guggenberger E., 2016, *The Astrophysical Journal*, 830, 31  
 Berger T. A., Huber D., van Saders J. L., Gaidos E., Tayar J., Kraus A. L., 2020, arXiv e-prints, 2001, arXiv:2001.07737  
 Blanton M. R., et al., 2017, *The Astronomical Journal*, 154, 28  
 Borucki W. J., et al., 2010, *Science*, 327, 977  
 Brogaard K., et al., 2012, *Astronomy and Astrophysics*, 543, A106  
 Broomhall A.-M., Chaplin W. J., Elsworth Y., New R., 2011, *MNRAS*, 413, 2978  
 Campante T. L., et al., 2016, *ApJ*, 819, 85  
 Casagrande L., Flynn C., Portinari L., Girardi L., Jimenez R., 2007, *MNRAS*, 382, 1516  
 Casagrande L., Ramírez I., Meléndez J., Bessell M., Asplund M., 2010, *Astronomy and Astrophysics*, 512, A54  
 Chan V. C., Bovy J., 2020, *MNRAS*, 493, 4367  
 Chaplin W. J., Miglio A., 2013, *Annual Review of Astronomy and Astrophysics*, 51, 353  
 Chaplin W. J., et al., 2020, *Nature Astronomy*, 4, 382  
 Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *The Astrophysical Journal*, 823, 102  
 Christensen-Dalsgaard J., Proffitt C. R., Thompson M. J., 1993, *The Astrophysical Journal Letters*, 403, L75  
 Clevert D.-A., Unterthiner T., Hochreiter S., 2015, arXiv e-prints, 1511, arXiv:1511.07289  
 Connally J. N., Bizzarro M., Krot A. N., Nordlund Å., Wielandt D., Ivanova M. A., 2012, *Science*, 338, 651  
 Cooke R. J., Fumagalli M., 2018, *Nature Astronomy*, 2, 957  
 Cyburt R. H., Fields B. D., Olive K. A., Yeh T.-H., 2016, *Reviews of Modern Physics*, 88, 015004  
 Davies G. R., et al., 2016, *MNRAS*, 456, 2183  
 Dotter A., 2016, *The Astrophysical Journal Supplement Series*, 222, 8  
 Dotter A., Conroy C., Cargile P., Asplund M., 2017, *The Astrophysical Journal*, 840, 99  
 Ferguson J. W., Alexander D. R., Allard F., Barman T., Bodnarik J. G., Hauschildt P. H., Heffner-Wong A., Tamanai A., 2005, *The Astrophysical Journal*, 623, 585  
 Frebel A., 2010, *Astronomische Nachrichten*, 331, 474  
 Gaia Collaboration et al., 2016, *A&A*, 595, A1  
 Gaia Collaboration et al., 2018, *Astronomy and Astrophysics*, 616, A1  
 García Pérez A. E., et al., 2016, *The Astronomical Journal*, 151, 144  
 Gelman A., Rubin D. B., 1992, *Statistical Science*, 7, 457  
 Glorot X., Bordes A., Bengio Y., 2011, Proc. 14th Int. Conf. Artif. Intell. Statistics AISTATS 2011, 15, 315  
 Green G. M., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, 478, 651  
 Grevesse N., Sauval A. J., 1998, *Space Sci. Rev.*, 85, 161  
 Hahnloser R. H. R., Sarpeshkar R., Mahowald M. A., Douglas R. J., Seung H. S., 2000, *Nature*, 405, 947  
 Hall O. J., et al., 2019, *MNRAS*, 486, 3569  
 Hawkins K., Leistedt B., Bovy J., Hogg D. W., 2017, *MNRAS*, 471, 722  
 Hendriks L., Aerts C., 2019, *PASP*, 131, 108001  
 Hogg D. W., Bovy J., Lang D., 2010, arXiv e-prints, 1008, arXiv:1008.4686  
 Homan M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, 15, 1593  
 Huber D., et al., 2011, *ApJ*, 743, 143  
 Huber D., et al., 2017, *The Astrophysical Journal*, 844, 102  
 Huber D., et al., 2019, *The Astronomical Journal*, 157, 245  
 Kingma D. P., Ba J., 2014, arXiv e-prints, 1412, arXiv:1412.6980  
 Kjeldsen H., Bedding T. R., Christensen-Dalsgaard J., 2008, *The Astrophysical Journal Letters*, 683, L175  
 Korn A. J., Grundahl F., Richard O., Mashonkina L., Barklem P. S., Collet R., Gustafsson B., Piskunov N., 2007, *The Astrophysical Journal*, 671, 402  
 Kuszlewicz J. S., Chaplin W. J., North T. S. H., Farr W. M., Bell K. J., Davies G. R., Campante T. L., Hekker S., 2019, *MNRAS*, 488, 572  
 Leistedt B., Hogg D. W., 2017, *AJ*, 154, 222  
 Li T., Bedding T. R., Huber D., Ball W. H., Stello D., Murphy S. J., Bland-Hawthorn J., 2018, *Monthly Notices of the Royal Astronomical Society*, 475, 981  
 Lindegren L., et al., 2018, *A&A*, 616, A2  
 Magic Z., Weiss A., Asplund M., 2015, *A&A*, 573, A89  
 Morton T. D., Winn J. N., 2014, *ApJ*, 796, 47  
 Nissen P. E., Silva Aguirre V., Christensen-Dalsgaard J., Collet R., Grundahl F., Slumstrup D., 2017, *Astronomy and Astrophysics*, 608, A112  
 Nsamba B., Campante T. L., Monteiro M. J. P. F. G., Cunha M. S., Rendle B. M., Reese D. R., Verma K., 2018, *Monthly Notices of the Royal Astronomical Society*, 477, 5052  
 Önehag A., Gustafsson B., Korn A., 2014, *Astronomy and Astrophysics*, 562, A102  
 Paquette C., Pelletier C., Fontaine G., Michaud G., 1986, *The Astrophysical Journal Supplement Series*, 61, 177  
 Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, *ApJS*, 192, 3  
 Paxton B., et al., 2013, *The Astrophysical Journal Supplement Series*, 208, 4

- Paxton B., et al., 2015, *The Astrophysical Journal Supplement Series*, 220, 15
- Paxton B., et al., 2018, *The Astrophysical Journal Supplement Series*, 234, 34
- Paxton B., et al., 2019, *The Astrophysical Journal Supplement Series*, 243, 10
- Peimbert A., Peimbert M., Luridiana V., 2016, *Revista Mexicana de Astronomia y Astrofisica*, 52, 419
- Pinsonneault M. H., An D., Molenda-Żakowicz J., Chaplin W. J., Metcalfe T. S., Bruntt H., 2012, *ApJS*, 199, 30
- Pinsonneault M. H., et al., 2014, *The Astrophysical Journal Supplement Series*, 215, 19
- Pinsonneault M. H., et al., 2018, *The Astrophysical Journal Supplement Series*, 239, 32
- Pitrou C., Coc A., Uzan J.-P., Vangioni E., 2018, *Physics Reports*, 754, 1
- Qian N., 1999, *Neural Networks*, 12, 145
- Rauer H., et al., 2014, *Experimental Astronomy*, 38, 249
- Ribas I., Jordi C., Torra J., Giménez Á., 2000, *MNRAS*, 313, 99
- Ricker G. R., et al., 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 014003
- Riess A. G., et al., 2018, *The Astrophysical Journal*, 861, 126
- Rogers F. J., Nayfonov A., 2002, *The Astrophysical Journal*, 576, 1064
- Ruder S., 2016, arXiv e-prints, 1609, arXiv:1609.04747
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, *PeerJ Comput. Sci.*, 2, e55
- Scott P., et al., 2015, *Astronomy and Astrophysics*, 573, A25
- Serenelli A. M., Basu S., 2010, *ApJ*, 719, 865
- Serenelli A. M., Basu S., Ferguson J. W., Asplund M., 2009, *The Astrophysical Journal Letters*, 705, L123
- Serenelli A., et al., 2017, *ApJS*, 233, 23
- Silva Aguirre V., et al., 2015, *MNRAS*, 452, 2127
- Silva Aguirre V., et al., 2017, *ApJ*, 835, 173
- Skrutskie M. F., et al., 2006, *The Astronomical Journal*, 131, 1163
- Sonoi T., Samadi R., Belkacem K., Ludwig H.-G., Caffau E., Mosser B., 2015, *Astronomy and Astrophysics*, 583, A112
- Stancliffe R. J., Fossati L., Passy J.-C., Schneider F. R. N., 2016, *Astronomy and Astrophysics*, 586, A119
- Taylor J., et al., 2017, *The Astrophysical Journal*, 840, 17
- Thoul A. A., Bahcall J. N., Loeb A., 1994, *The Astrophysical Journal*, 421, 828
- Townsend R. H. D., Teitler S. A., 2013, *Monthly Notices of the Royal Astronomical Society*, 435, 3406
- Trampedach R., Stein R. F., Christensen-Dalsgaard J., Nordlund Å., Asplund M., 2014, *MNRAS*, 445, 4366
- Valle G., Dell'Ommodarme M., Prada Moroni P. G., Degl'Innocenti S., 2015, *Astronomy and Astrophysics*, 575, A12
- Verma K., Hanasoge S., Bhattacharya J., Antia H. M., Krishnamurthi G., 2016, *MNRAS*, 461, 4206
- Verma K., Raodeo K., Antia H. M., Mazumdar A., Basu S., Lund M. N., Silva Aguirre V., 2017, *The Astrophysical Journal*, 837, 47
- Verma K., Raodeo K., Basu S., Silva Aguirre V., Mazumdar A., Mosumgaard J. R., Lund M. N., Ranadive P., 2019, *MNRAS*, 483, 4678
- Viani L. S., Basu S., Ong J. M. J., Bonaca A., Chaplin W. J., 2018, *The Astrophysical Journal*, 858, 28
- Villante F. L., Serenelli A. M., Delahaye F., Pinsonneault M. H., 2014, *The Astrophysical Journal*, 787, 13
- Weiss A., Schlattl H., 2008, *Astrophysics and Space Science*, 316, 99
- West C., Heger A., 2013, *The Astrophysical Journal*, 774, 75
- White T. R., Bedding T. R., Stello D., Christensen-Dalsgaard J., Huber D., Kjeldsen H., 2011, *The Astrophysical Journal*, 743, 161
- Zinn J. C., Pinsonneault M. H., Huber D., Stello D., 2019, *ApJ*, 878, 136

## APPENDIX A: TRAINING THE ANN

In Table A1 we give the scaling parameters used to standardise our training dataset. We determined the median,  $\mu_{1/2}$  and standard deviation,  $\sigma$  to 3 decimal places for each of the input and output



**Figure A1.** The MAE as a function of epochs for the train and test dataset.

dimensions. We then standardised the data by subtracting  $\mu_{1/2}$  and dividing by  $\sigma$ .

In Figure A1 we show the train and test MAE as a function of epochs for the final ANN configuration. The train and test loss were comparable throughout training.

## APPENDIX B: THE SYNTHETIC POPULATION

In this section, we present the results for the NP, PP and MP models run on a synthetic sample of 100 stars with the following initial conditions. We randomly generated initial  $M$  and  $[M/H]_{\text{init}}$  uniformly. We drew initial values for  $Y_{\text{init}}$  from a normal distribution centred on the helium enrichment law from Equation 14 with  $\Delta Y/\Delta Z = 1.8$  and  $Y_P = 0.247$ , and scaled by  $\sigma_Y = 0.008$ . We also generated initial values for  $\alpha_{\text{mlt}}$  from a normal distribution centred on  $\mu_\alpha = 2.0$  and scaled by  $\sigma_\alpha = 0.08$ .

We evolved the synthetic stars to randomly chosen ages using MESA. We then took the output  $\tau$ ,  $T_{\text{eff}}$ ,  $L$ ,  $\Delta\nu$  and  $[M/H]_{\text{surf}}$  from the models and used these as true values for each of the stars. We added random noise to the observed quantities centred on the true values with a standard deviation of 2.2 per cent in  $T_{\text{eff}}$ , 3.5 per cent in  $L$ , 0.9  $\mu\text{Hz}$  in  $\Delta\nu$  and 0.07 dex in  $[M/H]_{\text{surf}}$  chosen to be representative of the APOKASC sample.

### B1 Stellar parameters

We found that the NP model recovered the true values for the individual stellar parameters, but the uncertainties were unreliable. The observational quantities alone were not good enough to constrain  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$ . As a result, their distributions were truncated at the bounds of their priors. These boundary effects skewed the marginalised posterior means for  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$  towards the centre of the prior (0.28 and 2.0 respectively).

The PP model recovered true values for the synthetic stars with more reliable uncertainty than the NP model. The addition of pooling  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$  between the stars improved their uncertainty which reduced the effects of the prior as seen in the NP model. We

We found little difference between the results of the PP and MP models.

We reran the PP model with 10 and 50 stars chosen randomly from the sample of synthetic stars. In Figure B1, we show the uncertainties in the several parameters from the results of each

**Table A1.** The median,  $\mu_{1/2}$  and standard deviation,  $\sigma$  for each parameter in the training data, used to standardise the dataset.

	Input					Output				
	$f_{\text{vol}}$	$M (\text{M}_\odot)$	$\alpha_{\text{mlt}}$	$Y_{\text{init}}$	$Z_{\text{init}}$	$\log(\tau/\text{Gyr})$	$T_{\text{eff}} (\text{K})$	$R (\text{R}_\odot)$	$\Delta\nu (\mu\text{Hz})$	$[\text{M}/\text{H}]_{\text{surf}} (\text{dex})$
$\mu_{1/2}$	0.865	1.000	1.900	0.280	0.017	0.790	5566.772	1.224	100.720	0.081
$\sigma$	0.651	0.118	0.338	0.028	0.011	0.467	601.172	0.503	42.582	0.361

of the models. For the two pooled parameters,  $Y_{\text{init}}$  and  $\alpha_{\text{mlt}}$ , the uncertainty reduction due to pooling is most obvious. We see the PP model repeatedly improves on the uncertainties from the NP model when  $N_{\text{stars}}$  is increased.

In Figure B1 we also see a similar reduction in uncertainty for  $\tau$ ,  $M$  and  $R$ , with all models improve upon the NP model. However, we do not see the same effect in  $Z_{\text{init}}$  for which the uncertainty appears dominated by observations of  $[\text{M}/\text{H}]_{\text{surf}}$ .

## B2 Population parameters

In Figure B2, we show the joint posterior distributions for the hyperparameters of the PP model fit to the results of the NP model. We see that this method appears to recover the true values well. However, fitting the model this way does not benefit from the same uncertainty reduction on the stellar parameters as shown in the pooled models. Furthermore, the uncertainties on the individual stellar parameters were found to be unreliable due to boundary effects from the prior. This likely means that the uncertainties on the hyperparameter results for the NP model were underestimated.

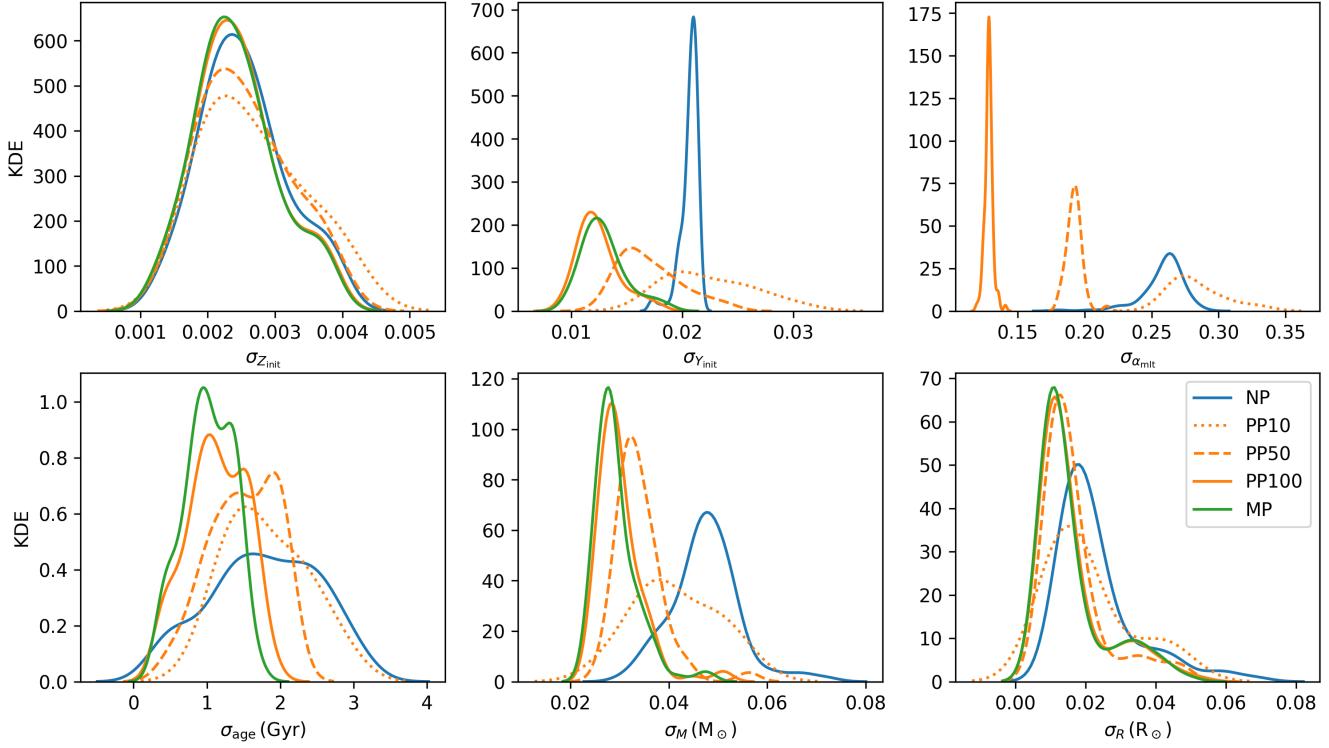
In Figure B3, we see that the PP model also recovers the hyperparameter truths well, with some noise due to random realisation error. Fitting the model this way has the added benefit over the NP model of improving the inference of the individual stellar parameters, as shown in the previous two sections. We also found that when we ran the PP model with 10 and 50 stars, the uncertainties on the hyperparameters also shrank with increasing  $N_{\text{stars}}$ .

Figure B4 shows the hyperparameter results for the MP model. Here,  $\alpha_{\text{mlt}}$  was assumed to be the same for all stars. This model also recovers the true hyperparameters for helium well, and the assumed value for  $\alpha_{\text{mlt}}$  is within uncertainty of the true  $\mu_\alpha$ .

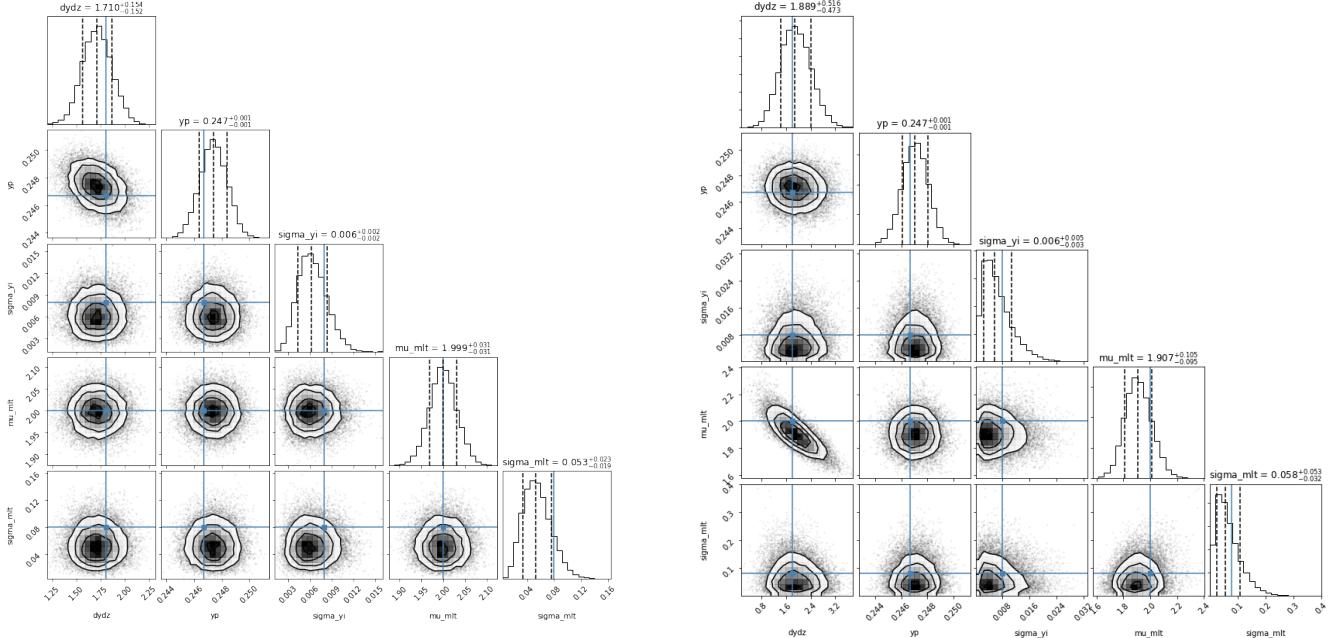
## APPENDIX C: THE SOLAR CALIBRATOR

We found that our model consistently recovers the Sun when modelled in each of the NP, PP and MP models. We show the marginal and joint posterior distributions for the solar parameters from the PPS model in the corner plot in Figure C1.

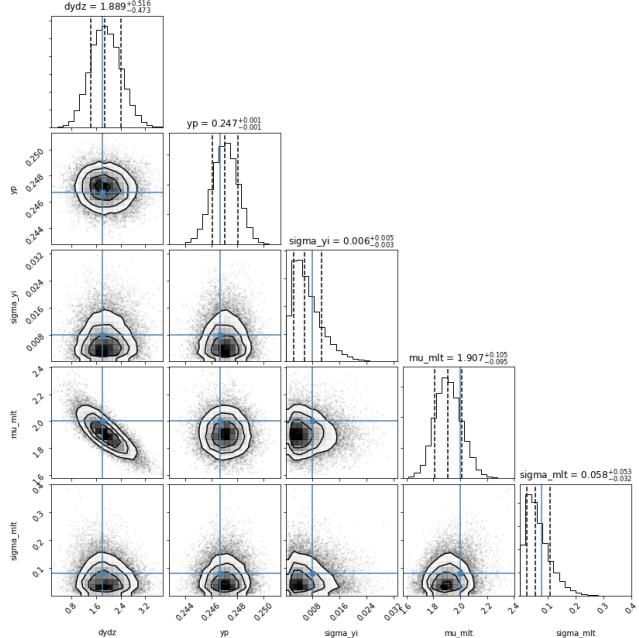
This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.



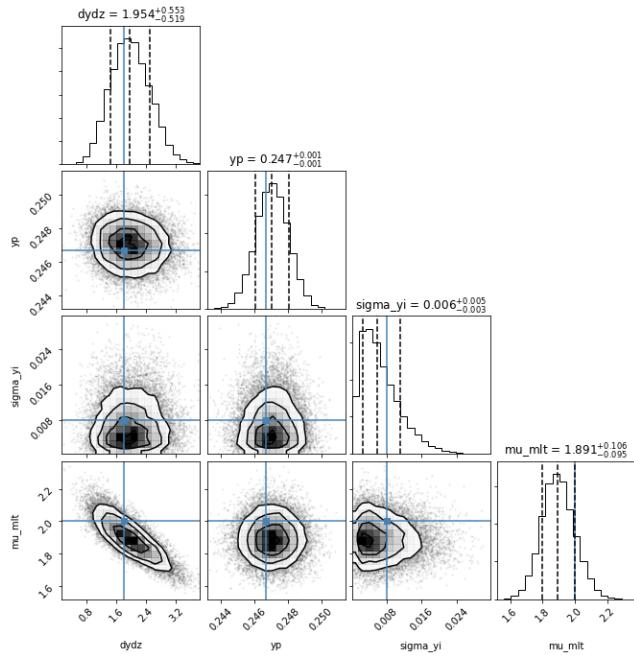
**Figure B1.** Kernel density estimates (KDEs) showing the reduction of statistical uncertainties between models of the sample of synthetic stars. The PP model was run with 10, 50 and 100 stars and is denoted PP10, PP50, and PP100 respectively. The NP and MP models were both run with the full set of 100 stars.



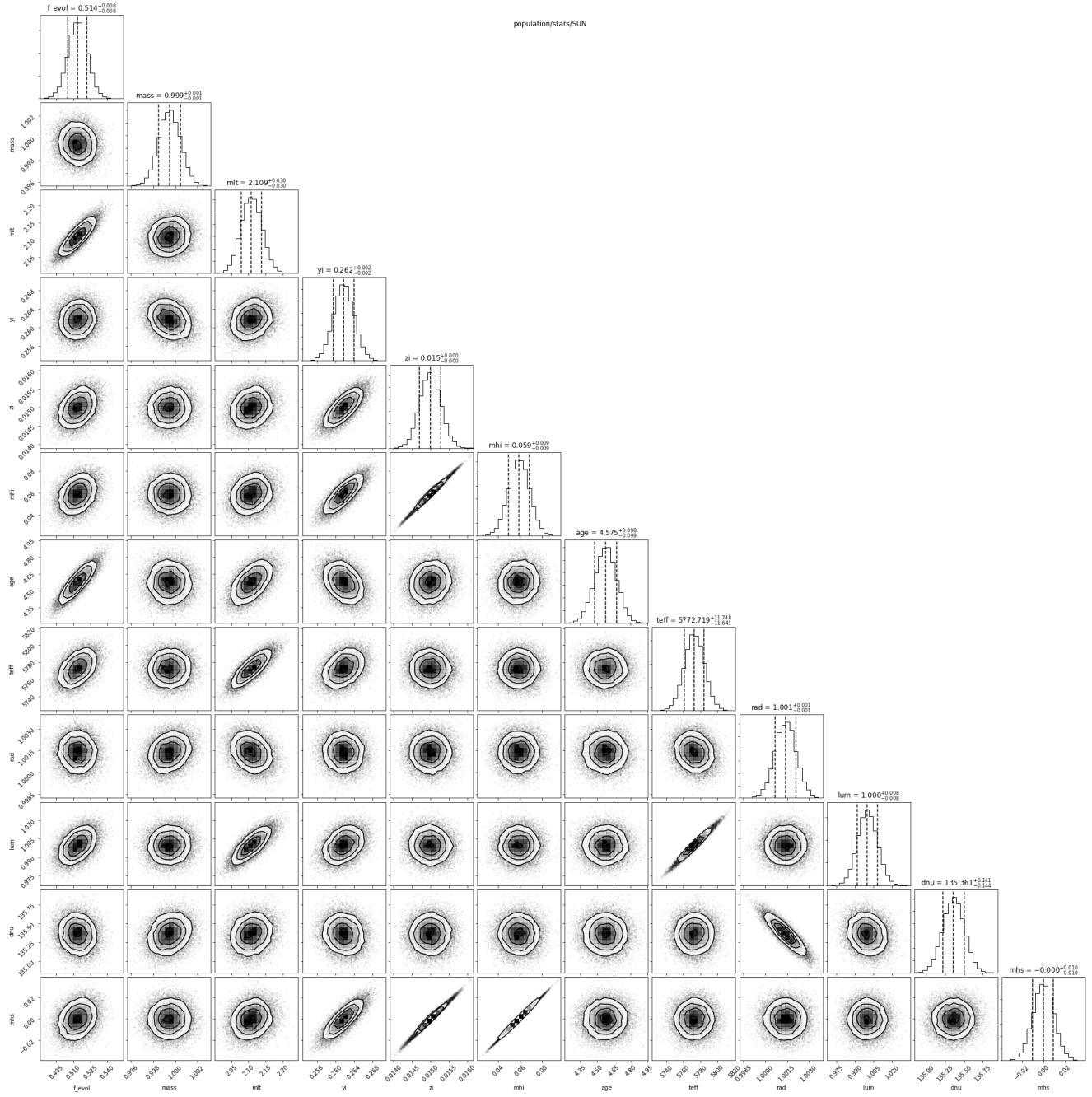
**Figure B2.** Corner plot showing the marginalised and joint posterior distributions between the NP model parameters for the synthetic stars. The true values are shown by the blue lines.



**Figure B3.** The same as Figure B2 but for the PP model.



**Figure B4.** The same as Figure B2 but for the MP model.



**Figure C1.** A corner plot showing the sampled marginal and joint posterior distributions for the Sun as a part of the PPS model.