



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 21, 2020

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

ABSTRACT

Your abstract goes here

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

	Page
1 Introduction	1
1.1 Hierarchical Bayesian Models	1
1.2 Modelling a Star	6
1.3 Asteroseismology of Solar-Like Oscillators	10
1.4 Sampling Stellar Models	13
1.5 Observing Stars	14
1.5.1 Photometry & Astrometry	14
1.5.2 Spectroscopy	16
1.5.3 Detecting Asteroseismic Oscillation Modes	17
2 Hierarchically Modelling Many Stars	18
3 Future Work	19
3.1 Including the Helium II Glitch	19
3.2 Increasing the Sample Size	19
3.3 To Higher Mass Stars and Beyond	19
References	20
Appendix A Accompanying Paper	25

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 $0.05 L_{\odot}$.	4
1.2	The z -score, $(\bar{\tau} - \tau_{\text{true}})/s_{\tau}$, where $\bar{\tau}$ and s_{τ} are the respective sample mean and standard deviation of the posterior ages from each of the no- and partially-pooled models.	5
1.3	Standard deviations, s_{τ} of the age posteriors from both the no- and partially-pooled models.	6
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 $\log g = 3.6$.	9
1.5	Spherical harmonic modes of oscillation for various combinations of angular degree (l) and azimuthal order (m).	12
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.	13

List of Tables

Chapter 1

Introduction

I will show how we did hierarchical modelling of stars.

1.1 Hierarchical Bayesian Models

NEEDS INTRODUCING AND REMEMBER THE AUDIENCE. TURN PACKED SENTANCES INTO PARAGRAPHS TO GUIDE THE READER THROUGH.

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, σ_y , we may write the likelihood function as a normal distribution,

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

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

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

The prior PDF of the model, assuming the parameters are independent, is $p(\boldsymbol{\theta}) = \prod_i p(\theta_i)$. Encoding our prior understanding of the model this way is useful for improving our inference. For example, we have independent prior understand that the age of the universe is ~ 14 Gyr [CITE]. Hence, we may choose to give the age parameter for a stellar model a uniform prior PDF from 0 to 14 Gyr such that our posterior PDF is not 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 θ_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 θ_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_{obj} similar stars. We could combine the posteriors for each star to get a posterior for the population of stars like so,

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

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

However, what if we have a model which describes the distribution of a particular parameter θ_i in the population? For example, if we have prior knowledge that stars in an stellar cluster have to

have formed at roughly the same time, we might want to encode such information into the model. One method would be to independently model the stars in the cluster and then find their mean and standard deviation in age. This method typically over-predicts the standard deviation because it fails to account the individual stellar uncertainties [CITE]. Alternatively, we can incorporate the assumption that stars in a cluster formed at the same time using 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 same value in the population.

We refer to models which pool parameters in this way as hierarachical models[CITE]. For a hierarchical model, we describe the distribution of Θ in the population by a set of *hyper-parameters*, $\phi = \{\phi_l\}_{l=1}^{N_\phi}$. Bayes' equation now becomes,

$$p(\phi, \Theta | Y) \propto p(Y|\Theta) p(\Theta|\phi) p(\phi) \quad (1.6)$$

where the probability of Θ given ϕ is,

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

and $d(\theta_k|\phi)$ 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, L from the ages, τ of $N_{\text{obj}} = 1000$ stars in a cluster formed at roughly the same time. Modelling the population independently, we get the posterior,

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

Now, let us consider a partially-pooled model where the stellar ages are drawn from a normal distribution centred on a mean, μ_τ and standard deviation, σ_τ . The posterior now becomes,

$$p(\mu_\tau, \sigma_\tau, \tau | L) \propto p(L|\tau) p(\tau|\mu_\tau, \sigma_\tau) p(\mu_\tau, \sigma_\tau), \quad (1.9)$$

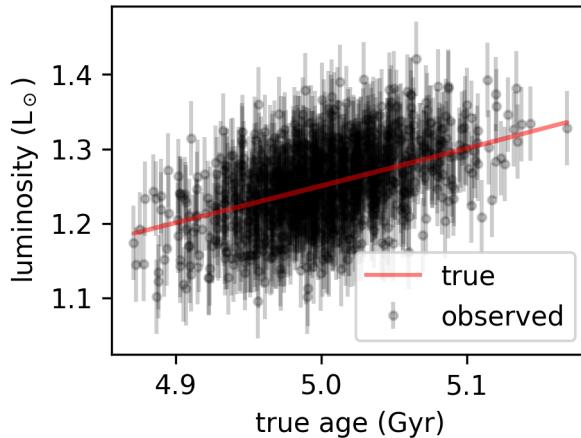


Figure 1.1: Luminosity against true ages of a fake stellar cluster. The true luminosities lie on the red line and the observed luminosities (black) have been artificially scattered by $0.05 L_\odot$.

where,

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

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 with a standard deviation of $0.05 L_\odot$ and proceeded to model the stellar ages using Equations 1.8 and 1.9 and the Bayesian package `pymc3` [CITE]. The observed and true luminosities are plotted against the true ages in Figure 1.1 to show

If we wished to determine spread of stellar ages in the cluster using the no-pooled model, we might naively 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. Therefore, the hierarchical model is a better way of determining population-

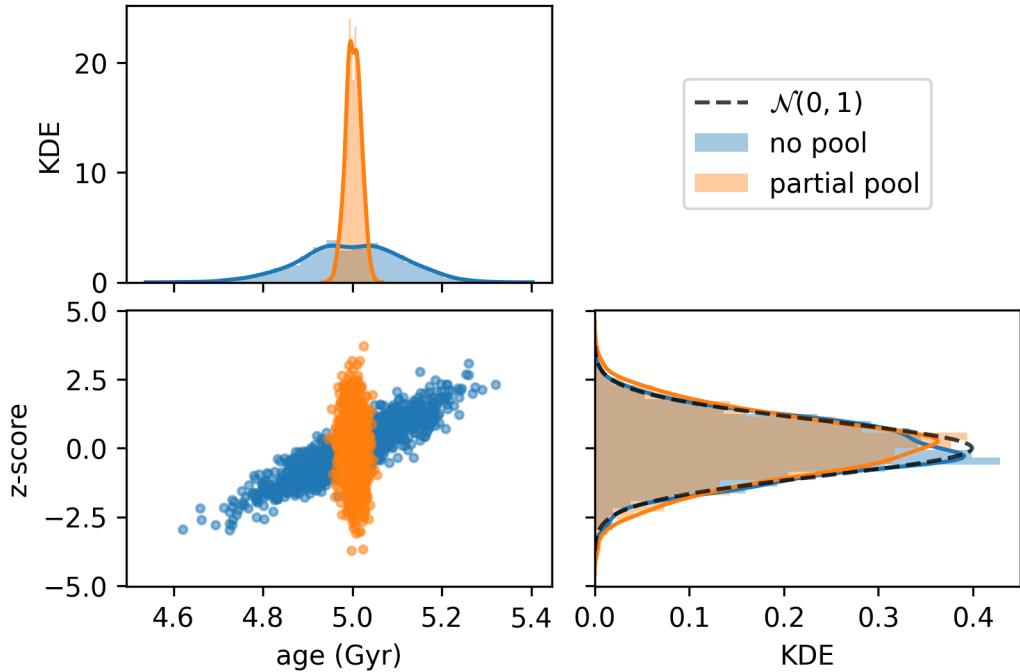


Figure 1.2: The z -score, $(\bar{\tau} - \tau_{\text{true}})/s_{\tau}$, where $\bar{\tau}$ and s_{τ} are the respective sample mean and standard deviation of the posterior ages from each of the no- and partially-pooled models.

level statistics than the traditional 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. However, 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 hyper-parameters. The reduced scatter on stellar ages is also reflected in the top-left plot of Figure 1.2.

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 analytical as in the example above. Before we can statistically model a population of stars, we must

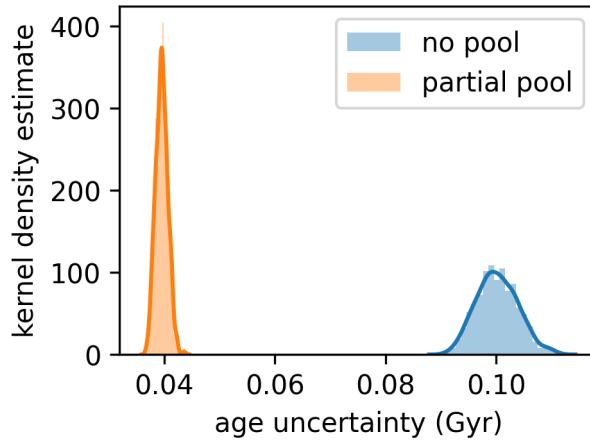


Figure 1.3: Standard deviations, s_τ of the age posteriors from both the no- and partially-pooled models.

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 its age and initial bulk properties. Following the development of quantum theories of ionization and radiation, the theory of the stellar interior 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 have 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 Ω is the gravitational potential energy of the star, γ is the adiabatic exponent which related the pressure of the star to the internal energy density, u , $P = (\gamma - 1)u$ assuming an ideal gas, and U is the total internal energy of the star. We show the virial theorem here to demonstrate that if a star contracts, its potential energy decreases which in tern 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 drawfs and neutron stars (Yakovlev and Urpin, 1980). Energy transportation is important when modelling stars. However, convective energy transfer is notoriously difficult to calculate. 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 α_{mlt} of the pressure scale height $H_P = -P(\text{d}r/\text{d}P)$, where α_{mlt} is of order unity. Since the mixing-length theory is an approximation of convection, its accuracy varies between stellar models and as such the value of α_{mlt} is often calibrated to the Sun. However, studies of 2D (Ludwig, Freytag, and Steffen, 1999) and 3D (Trampedach et al., 2014) hydrodynamical simulations calibrated α_{mlt} in the range 1.5 to 2.5 for different stars similar 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 a high enough 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 described by the dominant process of energy transport in the 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 increase 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_{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. 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_{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

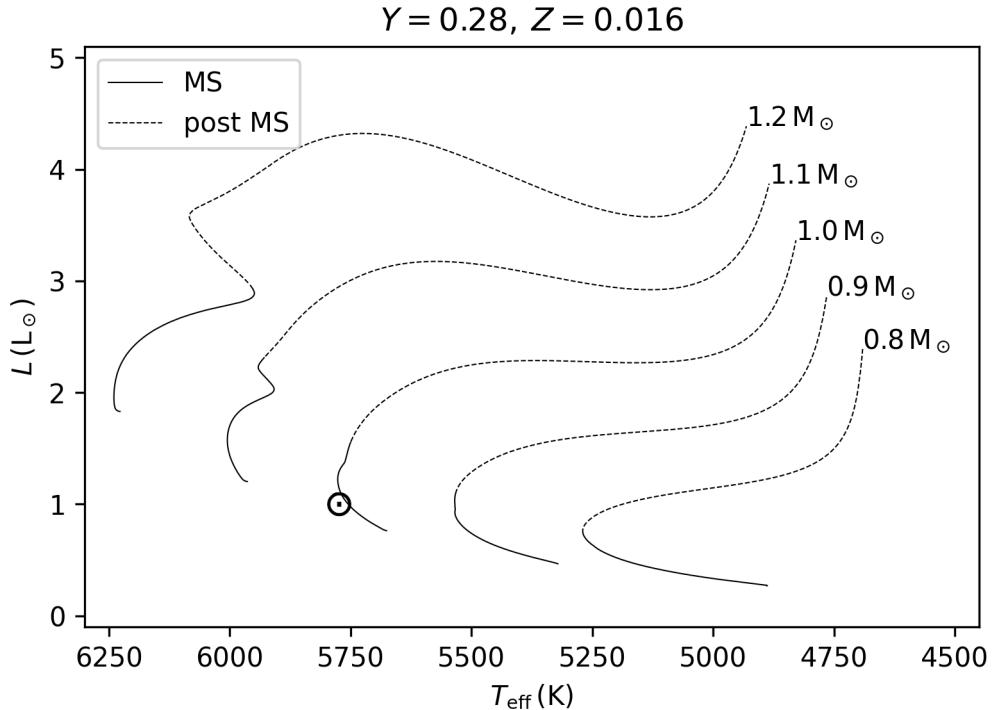


Figure 1.4: A series of stellar evolutionary tracks plot on a Hertzsprung-Russell diagram, starting at the ZAMS and evolved through the MS (black line) to the post MS (dashed black line) until the stellar surface gravity $\log g = 3.6$.

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) to model 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 called the 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 into a mesh of points at which it numerically integrates the stellar differential equations.

Example inputs to MESA include: the mass, M , the mixing-length theory parameter, α_{mlt} , and the fractional composition of hydrogen, helium and other heavy elements characterised by the 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 HR diagrams 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 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 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).

Give examples of the sorts of things asteroseismology can help us uncover, from ages (Ulrich, 1986; Soderblom, 2010; Silva Aguirre et al., 2015, see, e.g.) to masses and radii from scaling relations () and fitting stellar models().

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. This work focuses on main sequence stars for which p modes are only present in their spectra. Hence, in this section I will summarise the theory behind acoustic waves present in main sequence stars.

The theory which predicts the locations of the asteroseismic oscillation modes has its roots in the spherical harmonic oscillator. The eigenfrequencies, ν_{nlm} are categorised into modes of radial order, n , angular degree, l and azimuthal order, m . The radial order represents the number of standing wave nodes. 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 principle of hydrostatic equilibrium. 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 angular degree l . The waves curve due to the changing sound speed profile inside the star. We can see how waves with different l probe different depths of the stellar interior. In the case of the Sun, where modes of high l are able to be resolved, this has allowed us to uncover its density profile and through m , its internal rotation rates () .

To first order in $\Delta\nu$, I may express the eigenfrequency as follows (Mosser et al., 2011),

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

where,

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

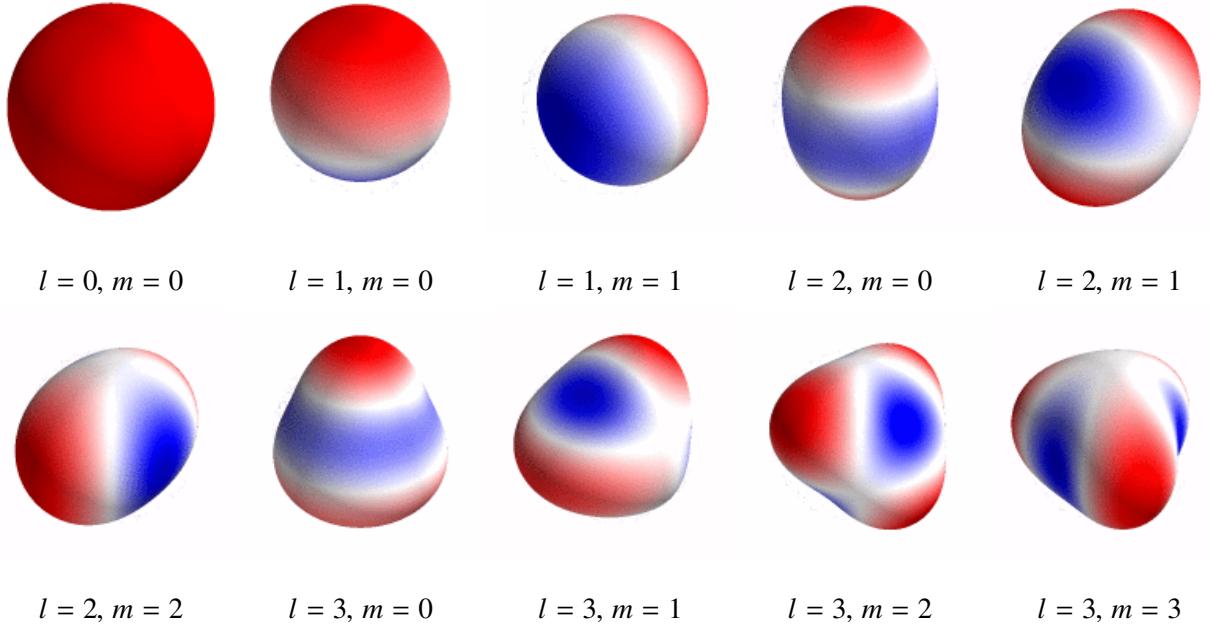


Figure 1.5: Spherical harmonic modes of oscillation for various combinations of angular degree (l) and azimuthal order (m).

is proportional to the inverse of the sound travel time over the stellar diameter, $2R$ where the speed of sound c is a function of stellar radii. The large frequency separation, $\Delta\nu$ is approximately the frequency difference between consecutive modes of the same l . From Equation 1.15, one can show by substitution of the speed of sound in a gas, that the average large frequency separation, $\langle\Delta\nu\rangle$ scales with the average stellar density, $\langle\rho\rangle$ [CITE Ulrich 1989],

$$\langle\Delta\nu_{nl}\rangle \propto \langle\rho\rangle^{1/2}. \quad (1.16)$$

How do we observe ν ?

Review some work on solar-like oscillators and fundamental parameters.

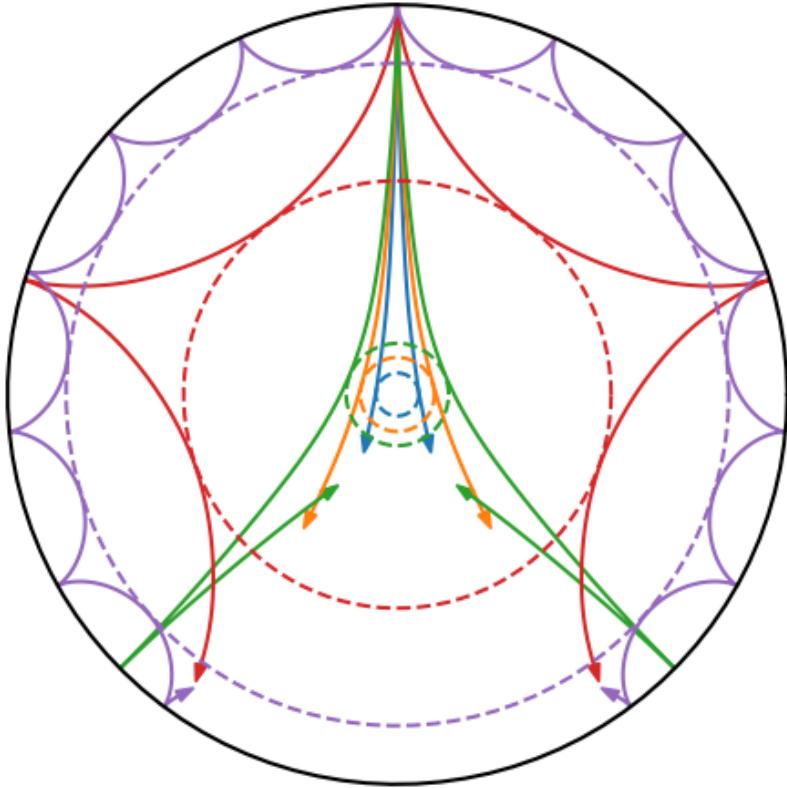


Figure 1.6: The asteroseismic wave fronts in a typical stellar interior. The blue, orange, green, red and purple lines represent the paths of oscillations with angular degree, $l = 1, 2, 3, 25, 75$ respectively. The dashed lines represent the depth probed by each mode.

1.4 Sampling Stellar Models

Typically, we start by producing a large grid of stellar models. Some are available online.

KEEP THIS SHORT DON'T SUBSECTION

GBM each star on its own. We can't do hierarchical models this way.

We intend to use a hierarachical model to model stars

What is GBM and give some examples e.g. BASTA Silva Aguirre et al. (2015).

What is wrong with GBM?

Why Interpolation is useful?

How we might interpolate, e.g. linear ND interpolator example.

A new alternative to GBM and interpolation is machine learning. Give examples of papers which have done this with stellar models.

Although ML stellar models is not new, it has not yet been applied to an HBM.

1.5 Observing Stars

In this section, we recall the fundamentals of observational astronomy. We first describe a method for determining luminosity from photometric and astrometric measurements. Then, we show how spectroscopy can determine the chemical composition and temperature at the surface of a star. Finally, we 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 ,

$$m_V = -2.5 \log\left(\frac{F_V}{F_{V,0}}\right), \quad (1.17)$$

where $F_{V,0}$ is the reference flux at a magnitude of zero, and its colour, $m_B - m_V$.

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 (ϖ) – the apparent difference in angular position of a star in the sky between observations separated by 1 astronomical unit (AU¹). We can then derive the distance, d to the star assuming the small-angle formula, $d \approx 1/\varpi$. We typically measure ϖ in arc seconds (" or 'as'), which yields a distance in parsecs (pc) – the distance from the Sun to an object which has moved 1 " over a baseline of 1 AU.

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,

$$\mathcal{M}_V = m_V - 5 \log(d) + 5 + A_V, \quad (1.18)$$

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 empirically determined or modelled by the surface properties of the star. The bolometric magnitude, \mathcal{M}_{bol} ,

$$\mathcal{M}_{\text{bol}} = \mathcal{M}_V - BC, \quad (1.19)$$

is related to the luminosity of the star,

$$\frac{L}{L_\odot} = 10^{\frac{2}{5} \cdot (\mathcal{M}_{\text{bol},\odot} - \mathcal{M}_{\text{bol}})}, \quad (1.20)$$

in terms of the solar luminosity, L_\odot .

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

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 surface to the effective temperature, T_{eff} .

The abundances of elements are typically reported as the ratio of the element to a readily abundant element. For example, the iron 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 write the metallicity as the relative abundance of all elements except for H and He,

$$[\text{M}/\text{H}] = \log(Z/X) - \log(Z/X)_{\odot}, \quad (1.21)$$

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_{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 the luminosity using Wein's law,

$$\lambda_{\text{max}} T_{\text{eff}} \approx 2.9 \times 10^{-3} \text{ m K}, \quad (1.22)$$

where λ_{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

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 ?? shows an example of an asteroseismic power spectrum.

There are several methods which exist to extract the radial and quadrupolar ($l = 0, 2$) oscillation modes from the asteroseismic power spectra (see e.g. Mosser et al., 2011; Appourchaux et al., 2012; Davies et al., 2016). In this work, we use the Python package `PBJam` (Nielsen et al. in preparation) to demonstrate one such method.

Chapter 2

Hierarchically Modelling Many Stars

See the accompanying paper (Appendix [A](#)).

Chapter 3

Future Work

3.1 Including the Helium II Glitch

3.2 Increasing the Sample Size

3.3 To Higher Mass Stars and Beyond

Our next step is to include intermediate-mass stars with masses from approx. 1.2 solar masses to 3.0 solar masses.

References

- Anderson, Lauren et al. (2018). “Improving Gaia Parallax Precision with a Data-Driven Model of Stars”. In: *AJ* 156, p. 145. doi: [10.3847/1538-3881/aad7bf](https://doi.org/10.3847/1538-3881/aad7bf).
- 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).

REFERENCES

- 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).
- Bellinger, E. P. et al. (2019). “Stellar Ages, Masses, and Radii from Asteroseismic Modeling Are Robust to Systematic Errors in Spectroscopy”. en. In: *A&A* 622, A130. doi: [10.1051/0004-6361/201834461](https://doi.org/10.1051/0004-6361/201834461).
- 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).
- 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).
- 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.

REFERENCES

- 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).
- Green, Gregory M. et al. (2019). “A 3D Dust Map Based on Gaia, Pan-STARRS 1, and 2MASS”. In: *ApJ* 887, p. 93. doi: [10.3847/1538-4357/ab5362](https://doi.org/10.3847/1538-4357/ab5362).
- 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.
- Kippenhahn, R., A. Weigert, and Emmi Hofmeister (1967). “Methods for Calculating Stellar Evolution”. In: *Methods Comput. Phys.* 7, pp. 129–190.
- 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.
- 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. 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).
- 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. (2012). “A Revised Effective Temperature Scale for the Kepler Input Catalog”. In: *ApJS* 199, p. 30. doi: [10.1088/0067-0049/199/2/30](https://doi.org/10.1088/0067-0049/199/2/30).

REFERENCES

- 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).
- 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).
- Si, Shijing et al. (2017). “A Hierarchical Model for the Ages of Galactic Halo White Dwarfs”. In: *MNRAS* 468, pp. 4374–4388. doi: [10.1093/mnras/stx765](https://doi.org/10.1093/mnras/stx765).
- 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).
- 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).
- 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).

REFERENCES

- 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,^{1,2*} Tanda Li,^{1,2} Guy R. Davies,^{1,2} Lindsey M. Carboneau,^{1,2} and To B. Confirmed

¹School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, UK

²Stellar 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

Context: Aims: To improve inference of stellar properties with a sample of asteroseismic dwarfs and subgiants. **Methods:** Using a hierarchical model to encode information about the population. Used machine learning to speed up inference from a grid of stellar models. Pooling initial helium using a freely varying enrichment law. Pooling mixing-length theory parameter. Tested on synthetic stars. Experimented with including solar data as a calibrator. **Results:** We reduced statistical uncertainties to 2.3, 1.1 and 12 per cent in mass (M), radius (R) and age (τ) respectively. **Conclusions:** HBMs can be used with our new method to improve precision of stellar fundamentals and uncover helium enrichment in the Milky Way.

Key words:

asteroseismology – methods: miscellaneous – methods: statistical – stars: fundamental parameters – stars: low-mass

1 INTRODUCTION

In recent years, the inference of stellar ages, masses and radii have been improved through the field of asteroseismology. 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 uncovered information about the stellar interior and improved estimates of stellar fundamental properties (see e.g. the review by Chaplin & Miglio (2013)). For example, stellar interior stuff like helium and convection. Also, fundamental properties for use in exoplanet and galactic stuff.

Typically, asteroseismic studies make use of either the direct approach via scaling relations, or by comparing data to asteroseismic stellar evolutionary models. Give examples for each of these. For each example, point out flaws and assumptions.

In this work, we apply a new method (Davies et al. in prep.) for determining stellar properties using a hierarchical Bayesian model to a sample of *Kepler* dwarfs and subgiants. Explain the crux of the model. In many stellar models, a solar-calibrated enrichment law and mixing-length theory parameter are assumed. We want to increase free parameters to include initial helium and mixing-length, then marginalise over their uncertainty. We can do this by assuming a population level prior distribution for both quantities.

Give examples of hierarchical models in astrophysics and some in asteroseismology.

Why are HBMs difficult with stellar models? As mentioned

before, stellar evolutionary models are computationally expensive. Evaluating the likelihood at points on a grid can give inaccurate posterior uncertainties depending on the grid resolution, but increasing the resolution is comp. expensive. We could speed up calls to the model by interpolating. However, interpolation is slow and does not scale well with dimensionality and grid size. In order to use sampling algorithms such as NUTS, we also require calculation of the likelihood gradient which is difficult with interpolation.

We suggest the use of machine learning in the form of an artificial neural network which is trained on a grid of stellar models.

There have been some examples of alternative methods to do seismology using machine learning (Bellinger et al. 2016). More specifically, neural networks (Verma et al. 2016; Hendriks & Aerts 2019). How these can speed up the process and to what level of success.

We explore different configurations of the HBM, with and without the inclusion of solar data as a calibrator.

2 DATA

We began with the sample of 415 stars from the first APOKASC catalogue of dwarfs and subgiants (S17). This sample provides an extensive set of dwarf and subgiant stars with asteroseismic detections observed by the *Kepler* mission. S17 used grid-based modelling to determine the ages, τ , masses, M , radii, R and surface gravity, $\log g$ of stars in the sample, using global asteroseismic

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

parameters, effective temperature T_{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, ν_{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_{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_{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 ~ 120 K. The median uncertainty in our adopted ASPCAP T_{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 for the *Kepler* field (Zinn et al. 2019; Hall et al. 2019). 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_{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 T_{eff} , $\log g$ and [M/H] (Dotter 2016; Choi et al. 2016). An uncertainty of 0.02 mag was assumed in ISOCLASSIFY 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_{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_{eff} for the sam-

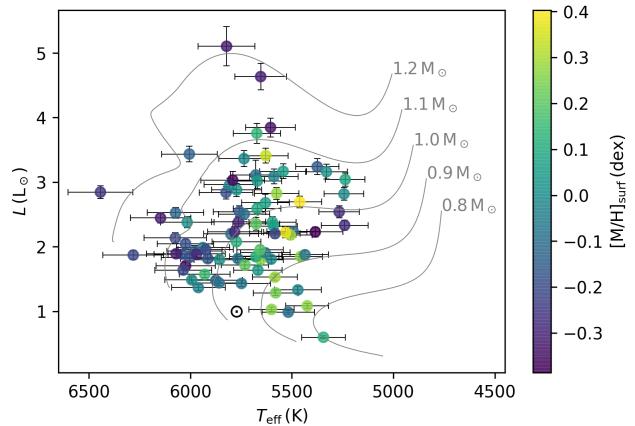


Figure 1. The luminosity, L against effective temperature, T_{eff} of the sample of 81 *Kepler* dwarfs and subgiants studied in this work. Each star is coloured according to metallicity. The grey lines depict evolutionary tracks with $[\text{M}/\text{H}]_{\text{init}} = 0.0$ dex, $Y_{\text{init}} = 0.28$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

ple plot on a Hertzsprung-Russell diagram in context with stellar evolutionary tracks at solar metallicity.

3 METHODS

Our principle goal was to improve 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, $[\text{M}/\text{H}]_{\text{init}}$, helium fraction, Y_{init} and mixing-length theory parameter, α_{mlt} our stellar models evolve the star outputting T_{eff} , L and chemical composition as a function of age, τ . 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

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_{eff} (K)	$\sigma_{T_{\text{eff}}}$ (K)	L (L_{\odot})	σ_L (L_{\odot})	$\Delta\nu$ (μHz)	$\sigma_{\Delta\nu}$ (μHz)	[M/H] _{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_{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_{init}	0.24 – 0.32	0.02	5
α_{mlt}	1.5 – 2.5	0.2	6
Total			17,220

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_{init}), initial metallicity ([M/H]_{init}), and the mixing-length parameter (α_{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 $\log g = 3.6$ 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:

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

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 (^1H , ^3He , ^4He , ^{12}C , ^{14}N , ^{16}O , ^{20}Ne , and ^{24}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 , ρ , 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 ~ 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). 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_{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 , α_{mlt} , Y_{init} and the initial heavy-elements fraction, Z_{init} . We also included the T_{eff} , $\log g$, $\Delta\nu$, stellar age (τ), 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 τ_{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 7.736×10^6 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 ~ 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_{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 τ_{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_{evol} gives the ANN information about the internal state of the star which affects the output observables. Otherwise, f_{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_{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 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_{evol} .

We standardised the training dataset by subtracting the median, $\mu_{1/2}$ and dividing by the standard deviation, σ . We found that the ANN performed better when the training data was scaled in this way. In Table B1, 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

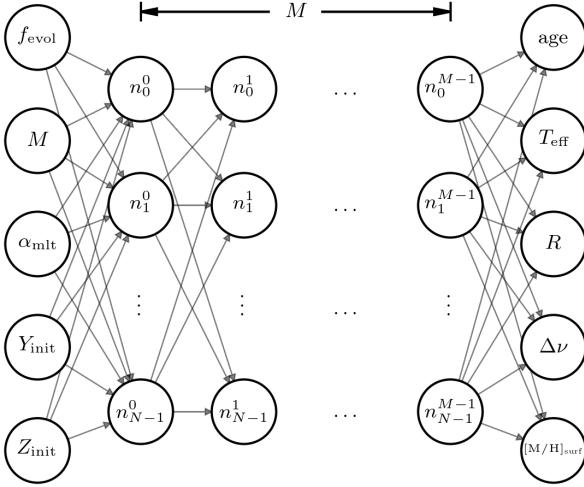


Figure 2. An artificial neural network comprising M hidden layers with N neurons per layer. Arrows connecting the nodes represent tunable weights.

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 λ_k and found that if 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.

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

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 1×10^{-4} and momentum of 0.999 with an MAE loss function. Training took ~ 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.

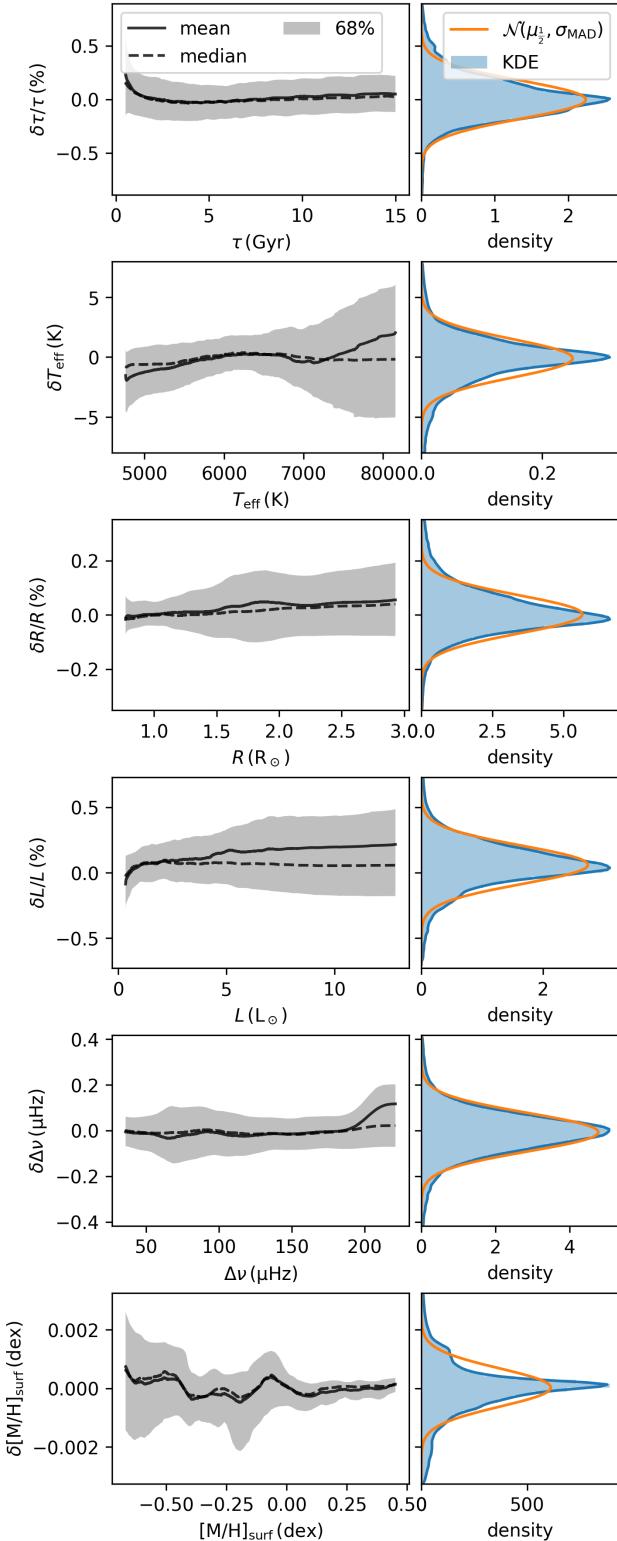


Figure 3. Left: the rolling error between the validation dataset (*true*) and the ANN predictions (*pred*) plotted against each parameter, where $\delta x = x_{\text{pred}} - x_{\text{true}}$ for a given output x . Right: a kernel density estimate (KDE) of the validation error and a normal distribution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

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, α_{mlt} in one model, whereas we maximally-pooled α_{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 $\theta_i = \{\text{evol}_i, M_i, \alpha_{\text{mlt},i}, Y_i, Z_i\}$ for a given star, i . Using Bayes' theorem, the *posterior* probability density function (PDF) of the model parameters given a set of observed data, d_i is,

$$p(\theta_i | d_i) \propto p(\theta_i) p(d_i | \theta_i), \quad (7)$$

where $p(\theta_i)$ is the *prior* PDF of the model parameters and $p(d_i | \theta_i)$ 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,

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

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 for θ_i as,

$$\begin{aligned} f_{\text{evol},i} &\sim 0.01 + 1.99 \cdot \mathcal{B}(1.2, 1.2), \\ M_i &\sim 0.8 + 0.4 \cdot \mathcal{B}(1.2, 1.2), \\ \alpha_{\text{mlt},i} &\sim 1.5 + \mathcal{B}(1.2, 1.2), \\ Y_{\text{init},i} &\sim 0.24 + 0.08 \cdot \mathcal{B}(1.2, 1.2), \\ Z_{\text{init},i} &\sim 0.005 + 0.035 \cdot \mathcal{B}(1.2, 1.2), \end{aligned}$$

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, $\{\log(\tau)_i, T_{\text{eff},i}, R_i, \Delta\nu_i, [\text{M}/\text{H}]_{\text{surf},i}\} = f_{\text{ANN}}(\theta_i)$, 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 \mathbf{d}_i with known uncertainty, σ_i given the model as,

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

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

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

$$p(\boldsymbol{\Theta}|\mathbf{D}) = \prod_{i=1}^{N_{\text{stars}}} p(\boldsymbol{\theta}_i|\mathbf{d}_i), \quad (11)$$

where $\boldsymbol{\Theta}$ is the matrix of model parameters and \mathbf{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_{init} to Y_{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 [CITE Gelman?], which builds upon the NP model by introducing population-level *hyperparameters*. Specifically, we chose to describe initial helium and α_{mlt} by partially-pooling them.

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

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

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

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_{init} and α_{mlt} given ϕ 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 μ_Y and is the mean initial helium fraction as described by the linear helium enrichment law [CITE WORKS WHICH USE THIS 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 α_{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_{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, σ_α . Therefore, we decided to describe the prior on α_{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 the mean [CITE PLANK] with a standard deviation representative of the range of values in the literature [CITE]. We assumed priors on the hyperparameters as follows,

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

$$Y_P \sim \mathcal{N}(0.247, 0.1),$$

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

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

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

where, for instance, $x \sim \mathcal{LN}(m, \sigma)$ 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 α_{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,

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

In other words, α_{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_{init} and α_{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 inputs.

For both the PP and MP models, we ran versions with and without data for the Sun included in the APOKASC population,

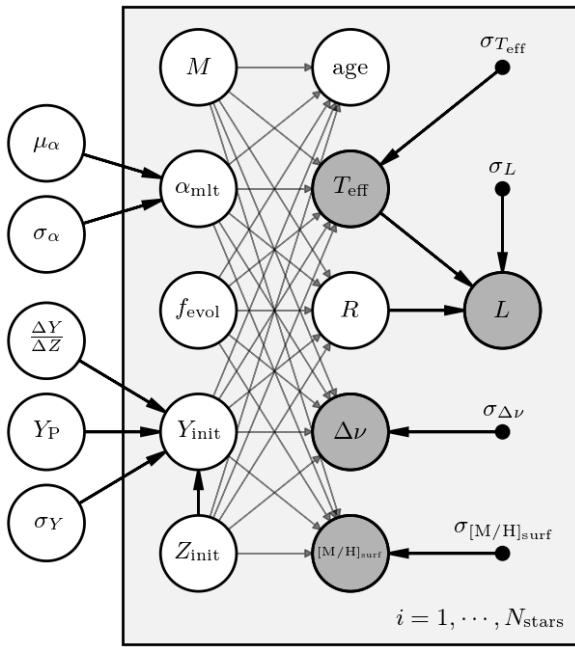


Figure 4. A probabilistic graphical model (PGM) of the partially-pooled (PP) hierarchical model. Nodes outside of the grey rectangle represent the hyperparameters in the model. Nodes inside the grey rectangle represent individual stellar parameters. Dark grey nodes represent observables which each have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

Table 4. Solar input data.

Input	μ	σ
$M (M_\odot)$	1.000	0.001
$\tau (\text{Gyr})$	4.6	0.1
$T_{\text{eff}} (\text{K})$	5777	20
$R (R_\odot)$	1.000	0.001
$L (L_\odot)$	1.00	0.01
$\Delta\nu (\mu\text{Hz})$	135.1	0.2
$[\text{M}/\text{H}]_{\text{surf}} (\text{dex})$	0.00	0.01

referred to as PPS and MPS respectively. We adopted the solar data shown in Table 4 as observables with uncertainties conservatively limited to the accuracy of the neural network. We adopt the central Solar $\Delta\nu$ from Huber et al. (2011), determined as if the Sun was observed by the *Kepler* telescope.

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 2000 samples across 10 MCMC chains and computed sample statistics for each parameter in the 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 rerun. We also dropped stars with problems during tuning where appropriate,

using the Gelman-Rubin diagnostic (\hat{r} ; Gelman & Rubin 1992). We were happy with 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 C.

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 effects on our inference with the addition of a solar calibrator. 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 *Kepler* 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

4.1.1 Stellar parameter results

After an initial run of the NP model on the APOKASC sample of 81 stars, 16 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 C). We saw the same boundary effects in the posteriors for Y_{init} and α_{mlt} here.

Out of the 65 stars sampled by the NP model, we obtained results for 63 with the PP model, after 2 stars were dropped due to convergence problems. In Tables 6 and 7 we present the results for 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 C1. We found that the statistical uncertainties in mass from the pooled models were reduced by a factor of ~ 2 over S17 with a median of ~ 2.5 per cent. We also obtained smaller uncertainties in radius and age of ~ 1.2 per cent 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 it 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 would make the resulting fit unreliable.

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_{evol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (Gyr)	T_{eff} (K)	$R (R_\odot)$	$\Delta\nu (\mu\text{Hz})$	[M/H] _{surf} (dex)
KIC5950854	$0.92^{+0.13}_{-0.15}$	$1.02^{+0.05}_{-0.05}$	$0.28^{+0.02}_{-0.02}$	$0.012^{+0.002}_{-0.002}$	$-0.02^{+0.06}_{-0.07}$	$5.8^{+1.7}_{-1.6}$	5955^{+55}_{-55}	$1.26^{+0.02}_{-0.02}$	$96.9^{+0.2}_{-0.2}$	$-0.15^{+0.07}_{-0.07}$
KIC3942719	$1.84^{+0.03}_{-0.05}$	$1.08^{+0.05}_{-0.05}$	$0.27^{+0.03}_{-0.02}$	$0.007^{+0.001}_{-0.001}$	$-0.24^{+0.06}_{-0.06}$	$5.6^{+0.4}_{-0.4}$	5761^{+73}_{-71}	$2.15^{+0.05}_{-0.05}$	$45.2^{+1.1}_{-1.1}$	$-0.29^{+0.06}_{-0.06}$
KIC4457351	$1.79^{+0.05}_{-0.06}$	$0.94^{+0.03}_{-0.04}$	$0.26^{+0.02}_{-0.01}$	$0.008^{+0.001}_{-0.001}$	$-0.20^{+0.05}_{-0.05}$	$10.7^{+0.8}_{-0.8}$	5476^{+67}_{-69}	$1.75^{+0.05}_{-0.04}$	$57.7^{+2.2}_{-2.2}$	$-0.26^{+0.05}_{-0.05}$
KIC3223000	$0.43^{+0.18}_{-0.19}$	$1.12^{+0.04}_{-0.04}$	$0.27^{+0.03}_{-0.02}$	$0.012^{+0.002}_{-0.002}$	$-0.03^{+0.06}_{-0.06}$	$2.0^{+1.1}_{-1.0}$	6187^{+67}_{-67}	$1.19^{+0.02}_{-0.02}$	$110.3^{+1.6}_{-1.6}$	$-0.13^{+0.07}_{-0.07}$
KIC7429287	$1.52^{+0.12}_{-0.12}$	$0.96^{+0.05}_{-0.05}$	$0.27^{+0.03}_{-0.02}$	$0.009^{+0.001}_{-0.001}$	$-0.17^{+0.06}_{-0.06}$	$8.5^{+1.3}_{-1.2}$	5808^{+58}_{-59}	$1.53^{+0.03}_{-0.03}$	$71.3^{+1.2}_{-1.2}$	$-0.28^{+0.07}_{-0.07}$

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

Name	f_{evol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (Gyr)	T_{eff} (K)	$R (R_\odot)$	$\Delta\nu (\mu\text{Hz})$	[M/H] _{surf} (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}	$1.18^{+0.02}_{-0.02}$	$112.5^{+2.7}_{-2.7}$	$0.07^{+0.07}_{-0.07}$
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_{evol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (Gyr)	T_{eff} (K)	$R (R_\odot)$	$\Delta\nu (\mu\text{Hz})$	[M/H] _{surf} (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	$1.51^{+0.10}_{-0.27}$	$1.09^{+0.03}_{-0.03}$	$0.28^{+0.01}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.34^{+0.06}_{-0.06}$	$9.6^{+1.1}_{-1.0}$	5507^{+47}_{-48}	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.6}_{-1.6}$	$0.28^{+0.06}_{-0.07}$
KIC10322381	$0.89^{+0.23}_{-0.22}$	$1.12^{+0.05}_{-0.06}$	$0.26^{+0.01}_{-0.01}$	$0.010^{+0.002}_{-0.002}$	$-0.10^{+0.06}_{-0.07}$	$4.3^{+1.7}_{-1.2}$	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}	$1.78^{+0.04}_{-0.03}$	$62.2^{+1.8}_{-1.7}$	$0.06^{+0.06}_{-0.06}$

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

Name	f_{evol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (Gyr)	T_{eff} (K)	$R (R_\odot)$	$\Delta\nu (\mu\text{Hz})$	[M/H] _{surf} (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}	$1.18^{+0.02}_{-0.02}$	$112.5^{+2.7}_{-2.7}$	$0.07^{+0.06}_{-0.07}$
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_{evol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (Gyr)	T_{eff} (K)	$R (R_\odot)$	$\Delta\nu (\mu\text{Hz})$	[M/H] _{surf} (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^{+35}_{-36}	$1.18^{+0.02}_{-0.02}$	$112.7^{+2.7}_{-2.7}$	$0.08^{+0.06}_{-0.07}$
KIC10319352	$1.61^{+0.04}_{-0.06}$	$1.08^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.33^{+0.06}_{-0.06}$	$10.8^{+0.7}_{-0.8}$	5516^{+46}_{-47}	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.7}_{-1.6}$	$0.28^{+0.06}_{-0.06}$
KIC10322381	$0.98^{+0.19}_{-0.20}$	$1.10^{+0.05}_{-0.05}$	$0.26^{+0.01}_{-0.01}$	$0.010^{+0.001}_{-0.001}$	$-0.13^{+0.07}_{-0.07}$	$5.1^{+1.2}_{-1.2}$	6106^{+94}_{-80}	$1.40^{+0.04}_{-0.04}$	$85.8^{+5.9}_{-4.3}$	$-0.30^{+0.08}_{-0.08}$
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}	$1.77^{+0.04}_{-0.03}$	$62.3^{+1.8}_{-1.8}$	$0.07^{+0.06}_{-0.07}$

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

Model	$\Delta Y/\Delta Z$	σ_Y	μ_α	σ_α	α_{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	$0.76^{+0.24}_{-0.27}$	$0.0049^{+0.0039}_{-0.0025}$	—	—	$2.088^{+0.031}_{-0.029}$

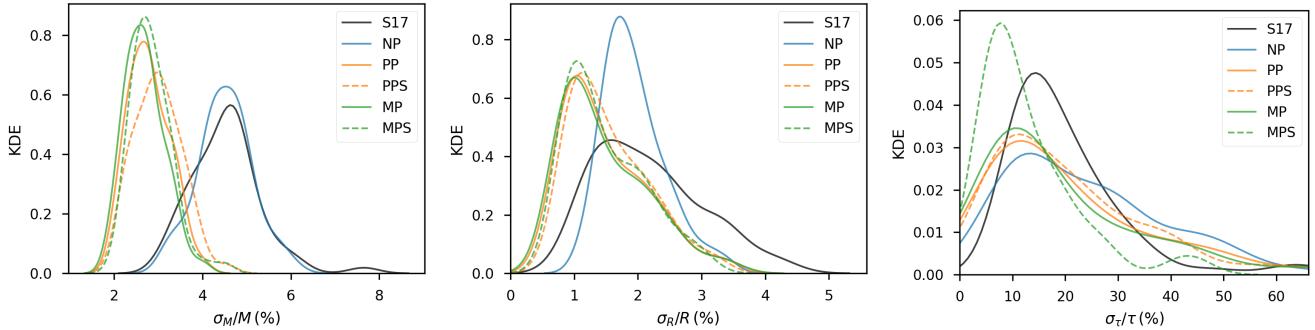


Figure 5. Kernel density estimates (KDEs) of the uncertainties in the results from each model compared with that of (S17) for the sample of APOKASC dwarfs and subgiants.

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 μ_α , 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 α_{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_{init} and Z_{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

5.1 Helium enrichment

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

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 $\Delta Y/\Delta Z$ of approximately 0.8 to 1.0 was consistent with the initial helium fraction expected from the solar model with our choice of Asplund et al. (2009) abundances (Serenelli & Basu 2010). However, such solar models do 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 our models without the Sun do not show the same effect as in the solar modelling problem. Therefore, the difference between the models with and without the Sun might indicate a problem with our sample.

We found little difference in the spread in helium, σ_Y between 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 parameters together which enables a better estimate of the population spread. Since there was little variation in σ_Y among hierarchical models, we can say that the spread of the initial helium fraction in our population of *Kepler* dwarfs and subgiants is ≈ 0.005 , assuming the linear helium enrichment law.

As shown in the sample of synthetic stars, 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 [CITE?]. Therefore, 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 limited by observational precision. However, hierarchical modelling as demonstrated in this work, allows us to get the most out of our data. Furthermore, minimising statistical uncertainties will improve our understanding and inference of systematic uncertainty.

Our results assume a linear helium enrichment law dependent only on the initial heavy element abundance, Z_{init} . However, helium in the galaxy could vary differently depending on the location of the star (galactic disk or bulge) [FIND PAPERS ON THIS]. 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, for example with the inclusion of α -element rich stars [PUT A BIT MORE INFO HERE].

What about our choice of Y_P and how e.g. Silva Aguirre et al. (2017) fit as a free parameter found sub-BBN values of Y_P fit to their results? We fixed this prior because, if a linear enrichment law existed, this is where it would start. It is more likely that a linear law is not realistic and there are other effects at play [CITE]

5.2 Mixing-length theory

Unlike chemical composition, the best-fitting α_{mlt} depends on the choice of model physics and to some degree, 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 α_{mlt} approximates convection varies across the HR diagram (Trampedach et al. 2014; Magic et al. 2015). [TODO] Something about Tayar17 and Viani18 which look at mlt with metallicity?

We found more of a difference in the α_{mlt} hyperparameter values between the models. The fit to the NP model results obtained a result for μ_α close to 2.0 – the midpoint of its prior distribution. This was expected because the tests on synthetic stars showed that α_{mlt} was biased towards the middle of its prior due to boundary effects from truncating the distribution between 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 solar results for the PPS model yielded a

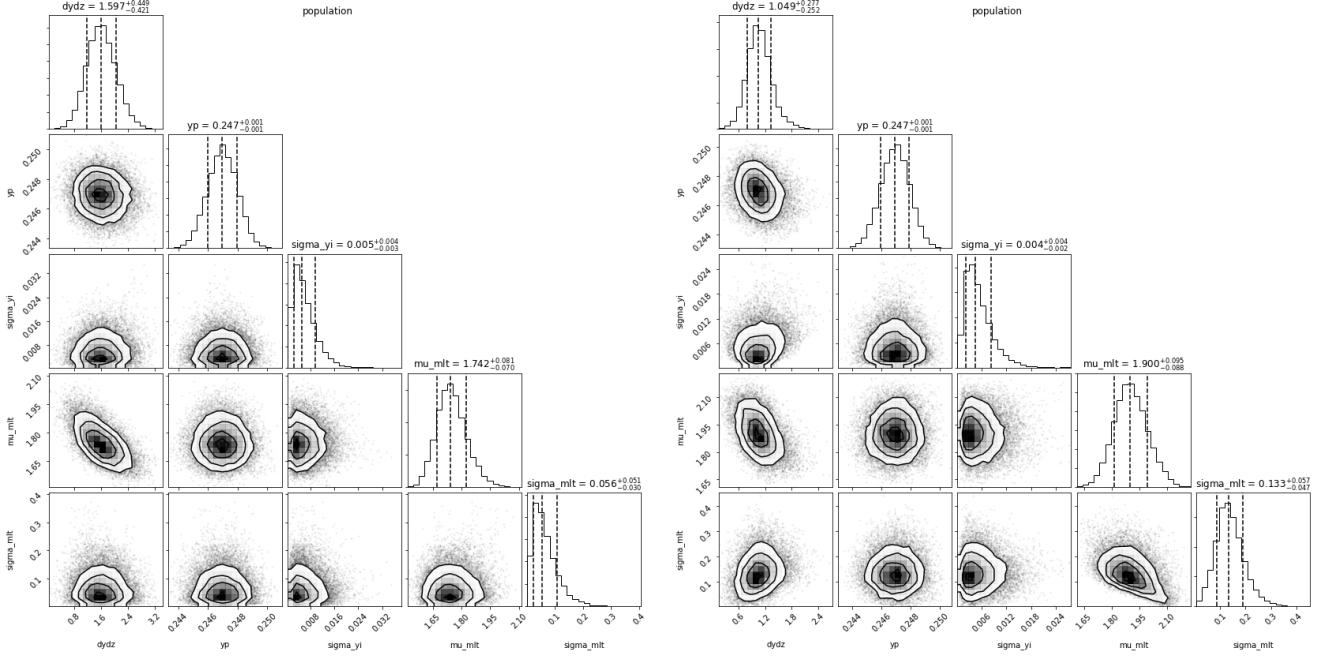


Figure 6. Corner plots showing the joint and marginalised sampled posterior distributions for the hyperparameters for the PP (left) and PPS (right) models. The vertical dashed lines give the 16th, 50th and 84th percentiles.

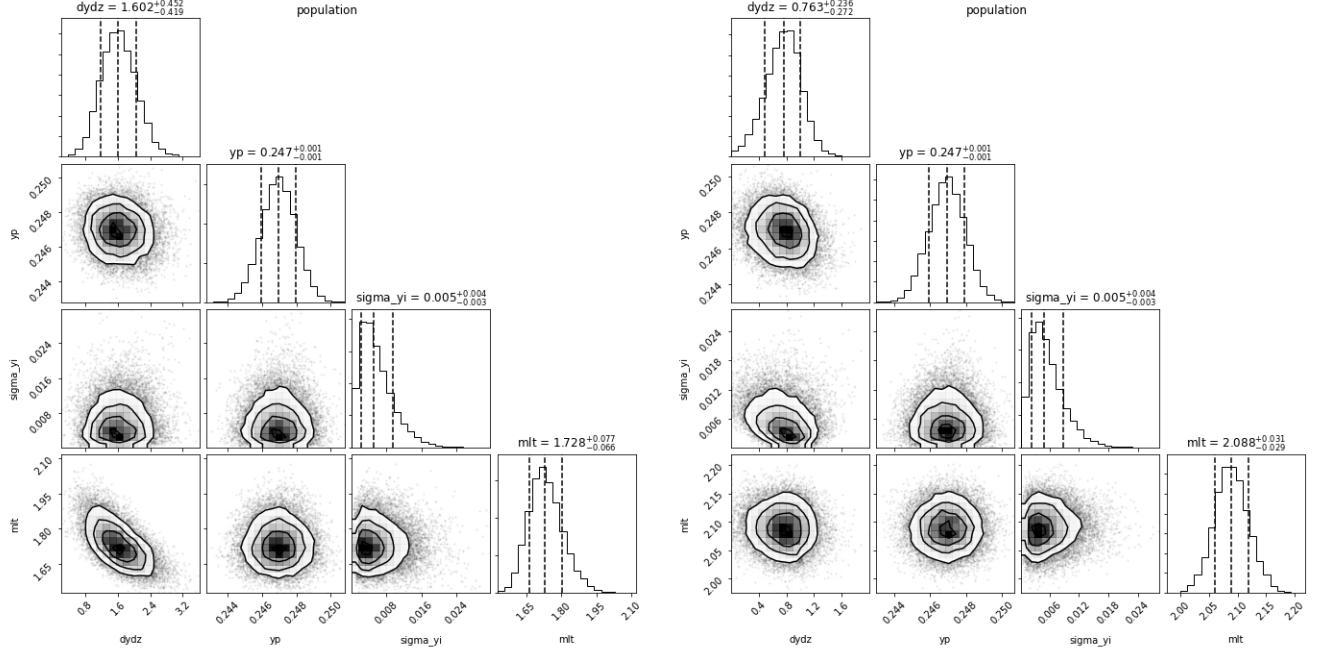


Figure 7. The same as Figure 6 but for the MP (left) and MPS (right) models.

value of $\alpha_{\text{mlt}\odot} = 2.11 \pm 0.03$ which was considerably higher than the α_{mlt} obtained for the other stars in the sample. The solar value also exceeds typical solar calibrated values of ~ 1.9 for the same stellar evolution code (). 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 α_{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 μ_α , 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 α_{mlt} could explain the difference we see. In other words, if we assume that the best-fitting α_{mlt} is normally distributed in our

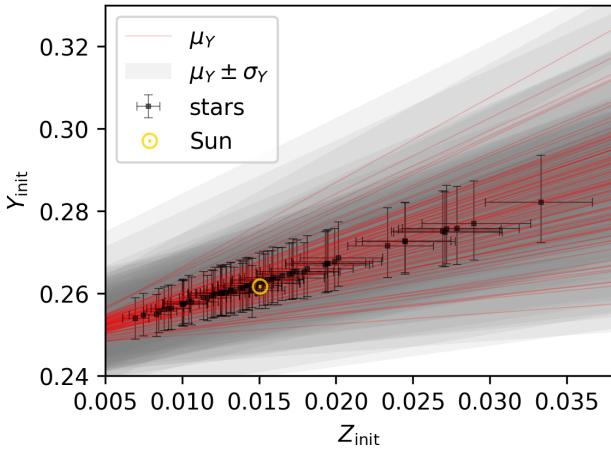


Figure 8. The results for initial helium fraction, Y_{init} against initial heavy-element fraction, Z_{init} for each star from the PPS model. 100 random samples from the posterior for the population mean, $\mu_Y = Y_P + (\Delta Y / \Delta Z)Z_{\text{init}}$ and spread, $\mu_Y \pm \sigma_Y$ are shown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

population, then the Sun lies within 2σ of the mean, among 95 per cent of all stars in the population.

There are no prior studies which look at the spread in α_{mlt} for a population of stars, but many examples which fit α_{mlt} as a function of $[\text{M}/\text{H}]$, T_{eff} and $\log g$. For example, results from Viani et al. (2018) for stellar models including diffusion, predict α_{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 α_{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 α_{mlt} in the MP and MPS models. The MP models yields a global α_{mlt} in line with μ_α 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 is close to assuming a solar calibrated value, because the model favours fitting to data with the best observational 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 α_{mlt} in both our NP and PP models. A possible explanation for the difference in α_{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_{eff} which typically underestimates T_{eff} compared to photometric scales. There is a positive correlation between α_{mlt} and T_{eff} , when holding all other variables constant. Therefore, the lower α_{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 helioseismic 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. Recent work by Nsamba et al. (2018) on a similar sample of stars, showed models without diffusion can lead to, on average, underestimated radii and masses and overestimated ages by 1, 3 and 16 per cent respectively, compared to those including diffusion.

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_{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_{eff} . The preferred results from S17 were determined using a photometric T_{eff} scale which we found to be ~ 170 K greater than our spectroscopic scale from DR14. In S17, they saw a similar offset between the DR13 T_{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_{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-most 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_{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 found that the results from all the pooled models showed the same difference in mass, with or without the Sun. Since including the Sun changed the resulting

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_{eff} scale and could also be explained by model physics.

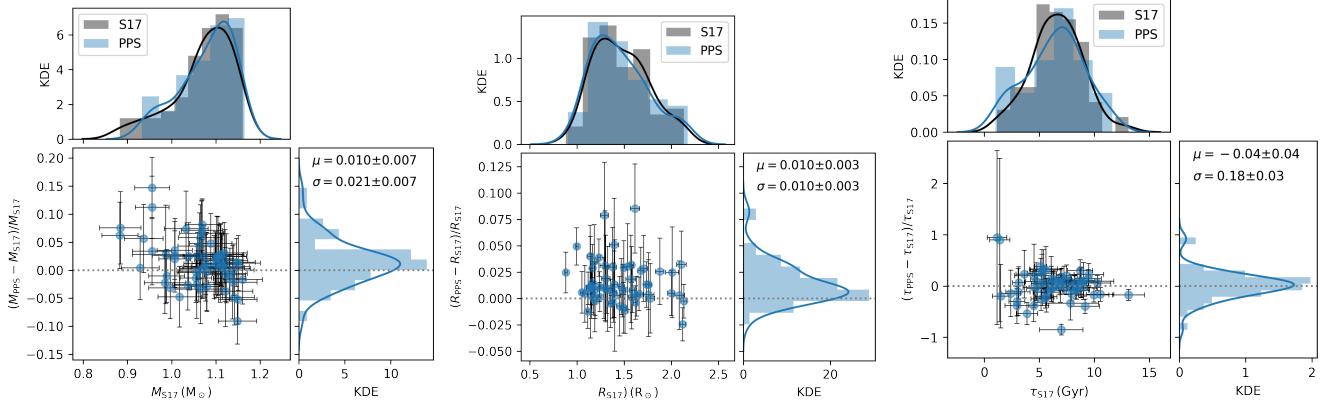


Figure 9. The mean and standard deviation in age, mass and radius results from the PPS model compared with the results (using the photometric temperature scale) from S17.

5.3.3 Age

Unlike mass and radius, our ages were largely consistent with S17. 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_{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.

5.4 Systematic uncertainties

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). Although our method can be adapted to different stellar evolutionary codes, a proper analysis of systematic uncertainties is left to future work.

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_{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- σ greater than its observed T_{eff} , and its modelled L was about 2- σ dimmer than its observed luminosity. Only ΔV 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_{eff} were consistent with a spectroscopic binary. We calculated a photometric T_{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, consistent with our modelled effective temperature and inconsistent with its spectroscopic T_{eff} . Therefore, we left KIC 9025370 in our results as an example of the robustness of our model.

5.6 The potential of pooling

Discuss the future of pooling and predictions of the kinds of uncertainties with more stars from TESS and PLATO.

6 CONCLUSIONS

ACKNOWLEDGEMENTS

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
- Appourchaux T., et al., 2015, *Astronomy and Astrophysics*, 582, A25
- Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481
- 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
- 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
- Casagrande L., Ramírez I., Meléndez J., Bessell M., Asplund M., 2010, *Astronomy and Astrophysics*, 512, A54
- Chaplin W. J., Miglio A., 2013, *Annual Review of Astronomy and Astrophysics*, 51, 353
- Choi J., Dotter A., Conroy C., Cantillo 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
- 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., Tamai A., 2005, *The Astrophysical Journal*, 623, 585
- 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
- Hendriks L., Aerts C., 2019, *PASP*, 131, 108001
- 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
- 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
- Lindgren L., et al., 2018, *A&A*, 616, A2
- Magic Z., Weiss A., Asplund M., 2015, *A&A*, 573, A89
- 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
- Pinsonneault M. H., An D., Molenda-Żakowicz J., Chaplin W. J., Metcalfe T. S., Bruntt H., 2012, *ApJS*, 199, 30
- Qian N., 1999, *Neural Networks*, 12, 145
- Ricker G. R., et al., 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 014003
- 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
- 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

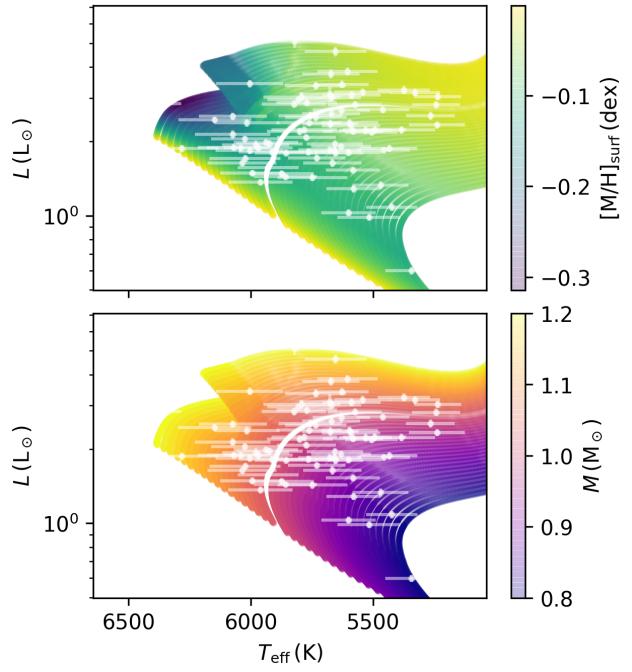


Figure A1. The luminosity, L against effective temperature, T_{eff} of the sample of 81 *Kepler* dwarfs and subgiants plot against a subset of the grid of stellar models computed in Section 3.1. The top plot is coloured by stellar surface metallicity and the bottom plot is coloured by stellar mass.

- 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
- Verma K., Hanasoge S., Bhattacharya J., Antia H. M., Krishnamurthi G., 2016, *MNRAS*, 461, 4206
- 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
- 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: GRID OF STELLAR MODELS

APPENDIX B: TRAINING THE NEURAL NETWORK

APPENDIX C: 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_{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 α_{init} 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

Table B1. The median, $\mu_{1/2}$ and standard deviation, σ for each parameter in the training data, used to standardise the dataset.

	Input					Output				
	f_{vol}	$M (\text{M}_\odot)$	α_{mlt}	Y_{init}	Z_{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
σ	0.651	0.118	0.338	0.028	0.011	0.467	601.172	0.503	42.582	0.361

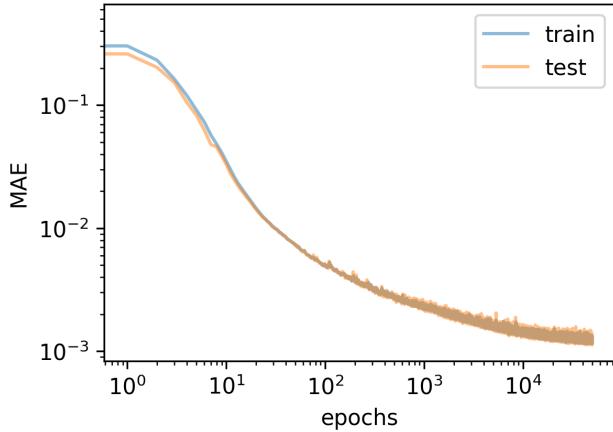


Figure B1. The MAE as a function of epochs for the train and test dataset.

MESA. We then took the output τ , T_{eff} , L , $\Delta\nu$ and $[\text{M}/\text{H}]_{\text{surf}}$ 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_{eff} , 3.5 per cent in L , 0.9 μHz in $\Delta\nu$ and 0.07 dex in $[\text{M}/\text{H}]_{\text{surf}}$ chosen to be representative of the APOKASC sample.

C1 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_{init} and α_{mlt} . As a result, their distributions were truncated at the bounds of their priors. These boundary effects skewed the marginalised posterior means for Y_{init} and α_{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_{init} and α_{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 C1, we show the uncertainties in the several parameters from the results of each of the models. For the two pooled parameters, Y_{init} and α_{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_{stars} is increased.

In Figure C1 we also see a similar reduction in uncertainty for τ , M and R , with all models improve upon the NP model. However, we do not see the same effect in Z_{init} for which the uncertainty appears dominated by observations of $[\text{M}/\text{H}]_{\text{surf}}$.

C2 Population parameters

In Figure C2, 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 C3, 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_{stars} .

Figure C4 shows the hyperparameter results for the MP model. Here, α_{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 α_{mlt} is within uncertainty of the true μ_α .

APPENDIX D: 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 Sun in the corner plot in Figure D1.

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

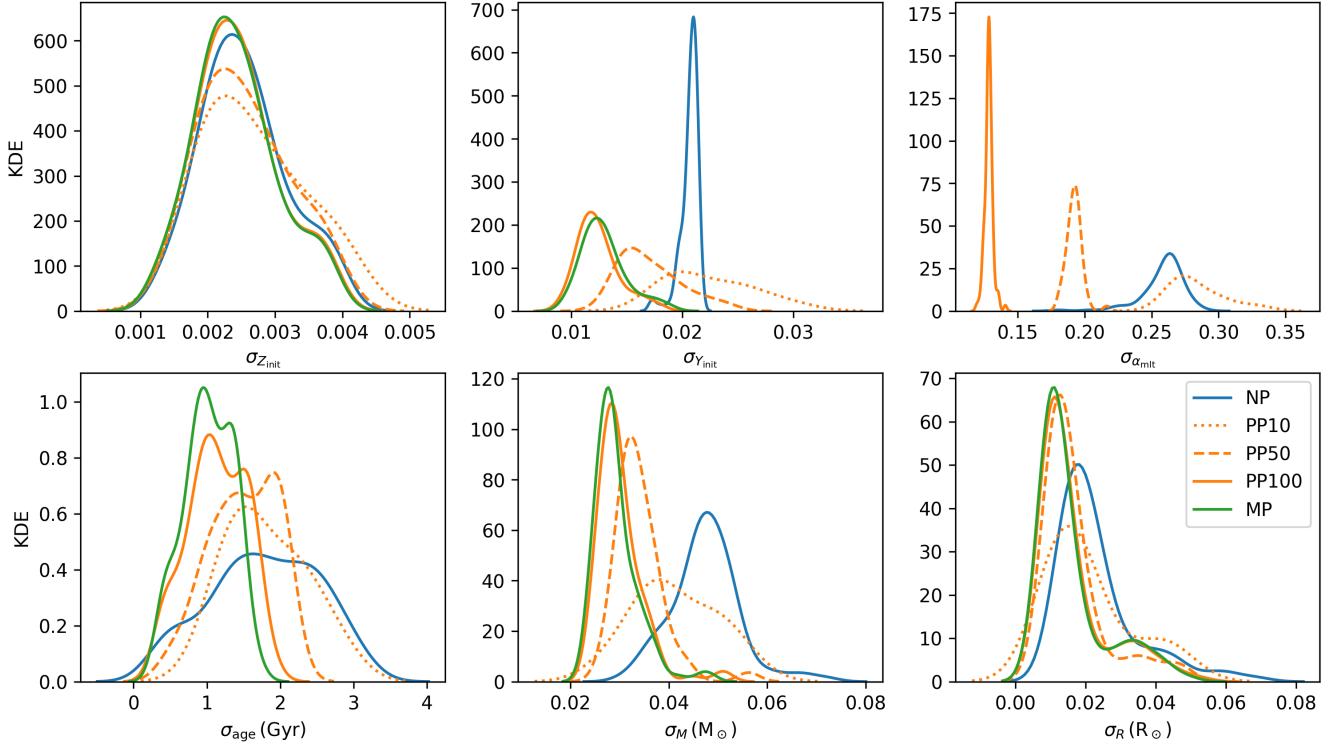


Figure C1. Kernel density estimates (KDEs) showing the reduction of statistical uncertainties between models of the sample of synthetic stars. The PP model was run with 10, 50 and 100 stars and is denoted PP10, PP50, and PP100 respectively. The NP and MP models were both run with the full set of 100 stars.

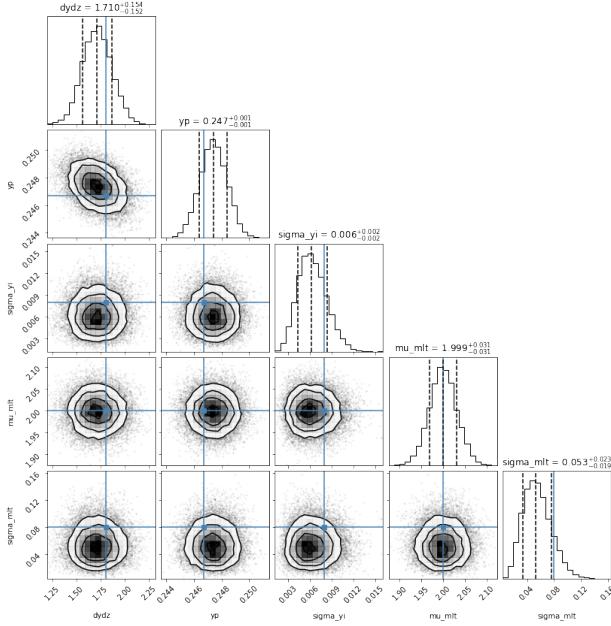


Figure C2. Corner plot showing the marginalised and joint posterior distributions between the NP model parameters for the synthetic stars. The true values are shown by the blue lines.

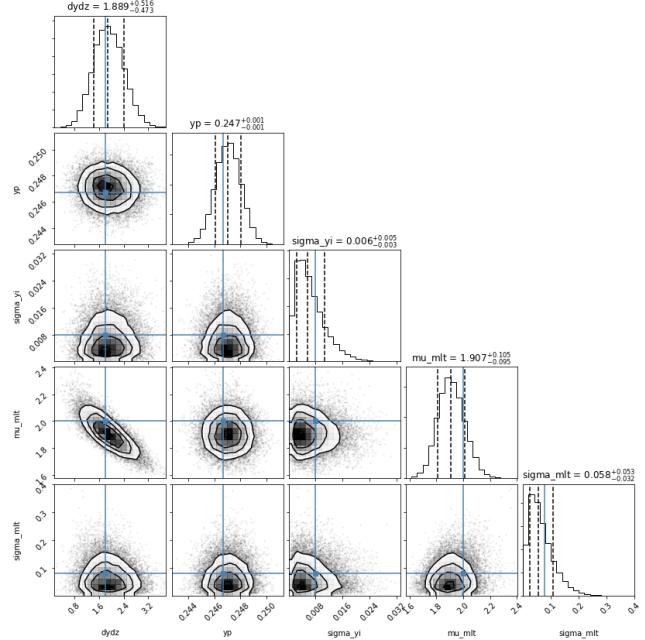


Figure C3. Corner plot showing the marginalised and joint posterior distributions between the NP model parameters for the synthetic stars. The true values are shown by the blue lines.

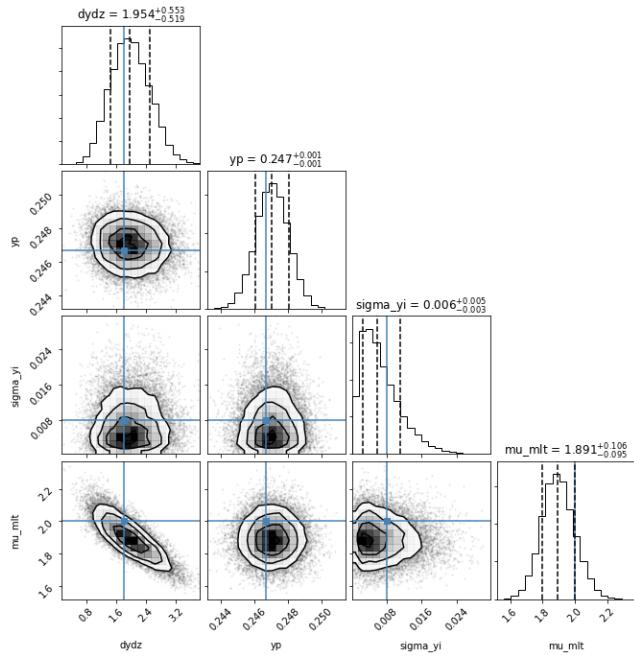


Figure C4. Corner plot showing the marginalised and joint posterior distributions between the MP model parameters for the synthetic stars. The true values are shown by the blue lines.

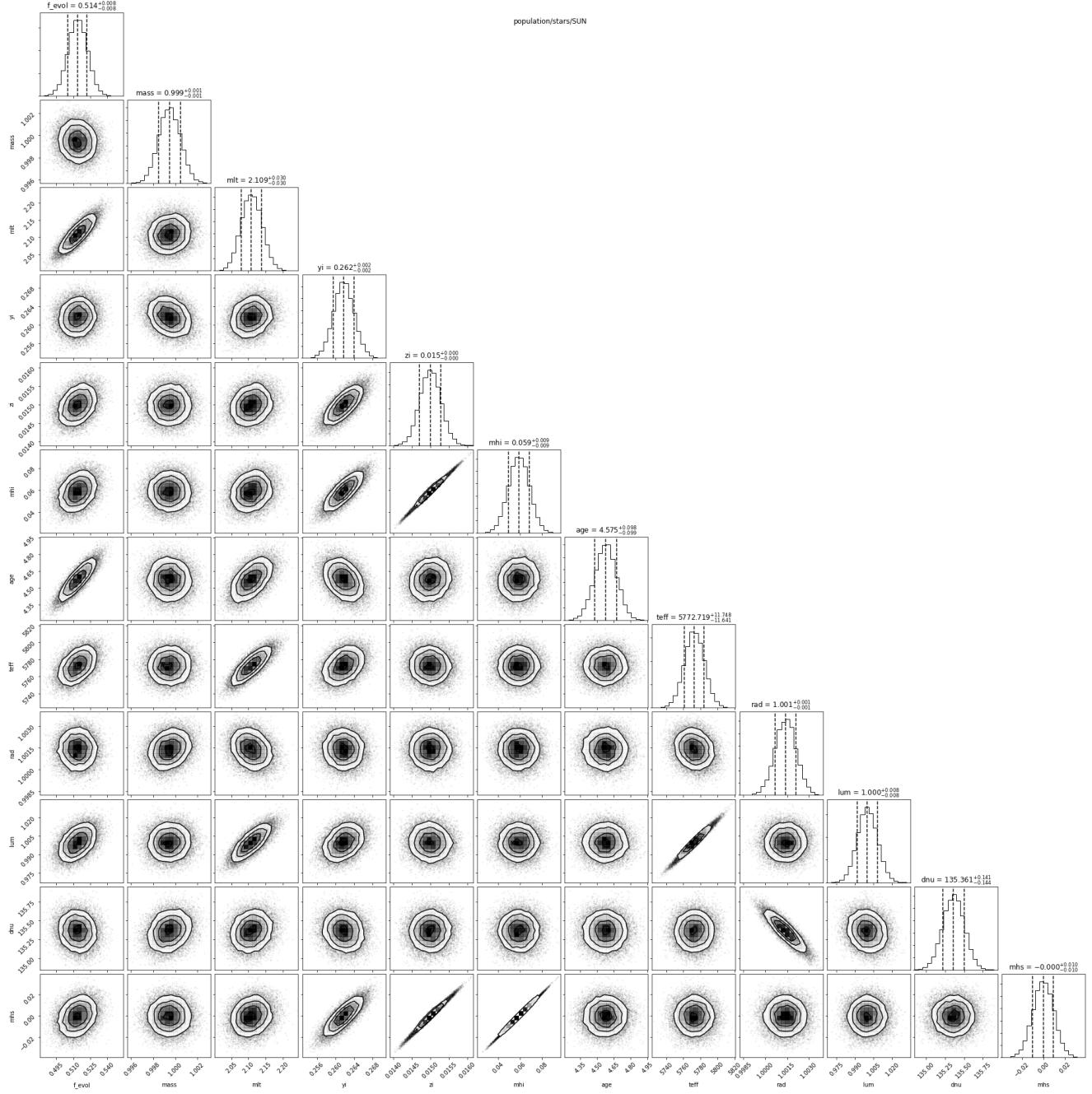


Figure D1. A corner plot showing the sampled marginal and joint posterior distributions for the Sun as a part of the PPS model.