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 16, 2020

ABSTRACT

Your abstract goes here

ACKNOWLEDGMENTS

I acknowledge the people who helped me.

Contents

List of Figures

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 object comprising a set of independent parameters, $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^{N_\theta}$ which makes a set of predictions, $\boldsymbol{\mu}_y = \{\mu_{y,j}\}_{j=1}^{N_y}$ where $\boldsymbol{\mu}_y = \boldsymbol{f}(\boldsymbol{\theta})$. Using Bayes' theorem, we may write the *posterior* probability density function (PDF) of the model given a set of observations $\boldsymbol{y}$ as,

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{y})}, \tag{1.1}$$

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

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

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

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

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 evidence that the age of the universe is $\sim 14\,\mathrm{Gyr}$ [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.

The evidence is the PDF of the observational data. We write this as the normalisation of the numerator of Equation 1.1,

$$p(\boldsymbol{y}) = \int_{-\infty}^{+\infty} p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}. \tag{1.4}$$

There are many ways to determine the posterior PDF, either analytically or numerically using e.g. Markov chain Monte Carlo (MCMC) through algorithms such as Metropolis-Hastings and Hamiltonian Monte-Carlo (HMC) [CITE]. 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,

$$p(\theta_1|\boldsymbol{y}) = \int_{-\infty}^{+\infty} p(\boldsymbol{\theta}|\boldsymbol{y})\, \mathrm{d}\theta_2 \ldots \mathrm{d}\theta_{N_\theta}. \tag{1.5}$$

Therefore, we end up with a distribution which describes the probability of θ_1 given $\boldsymbol{y}$ which takes into account the distribution (or uncertainty) of all other parameters in the model.

The model described above can be applied to a single object such as a star. Let us now consider modelling a population of N_{obj} similar objects. We could combine the posteriors for each object to get a posterior for the population of objects,

$$p(\boldsymbol{\Theta}|\boldsymbol{Y}) = \prod_{k=1}^{N_\mathrm{obj}} p(\boldsymbol{\theta}_k|\boldsymbol{y}_k), \tag{1.6}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^{N_\mathrm{obj}}$ and $\boldsymbol{Y} = \{\boldsymbol{y}_k\}_{k=1}^{N_\mathrm{obj}}$ 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 $\boldsymbol{\theta}_i$ in the population? For example, if all the objects are stars in an open cluster which formed at roughly the same time, such as Messier 67 [CITE], 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 population mean and standard deviation in age. It has been shown that this method typically over-predicts the standard deviation because it propagates the object-level 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 object-level parameters are drawn from some common distribution, and the latter is the special case where all object-level parameters share the same value in the population.

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

$$p(\boldsymbol{\phi}, \boldsymbol{\Theta}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\Theta})\, p(\boldsymbol{\Theta}|\boldsymbol{\phi})\, p(\boldsymbol{\phi})}{p(\boldsymbol{Y})} \tag{1.7}$$

where the probability of $\boldsymbol{\Theta}$ given $\boldsymbol{\phi}$ is,

$$p(\boldsymbol{\Theta}|\boldsymbol{\phi}) = \prod_{k=1}^{N_\mathrm{obj}} d(\boldsymbol{\theta}_k|\boldsymbol{\phi}), \tag{1.8}$$

and $d(\boldsymbol{\theta}_k|\boldsymbol{\phi})$ is some chosen distribution from which the parameters for a given object are drawn from the population.

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

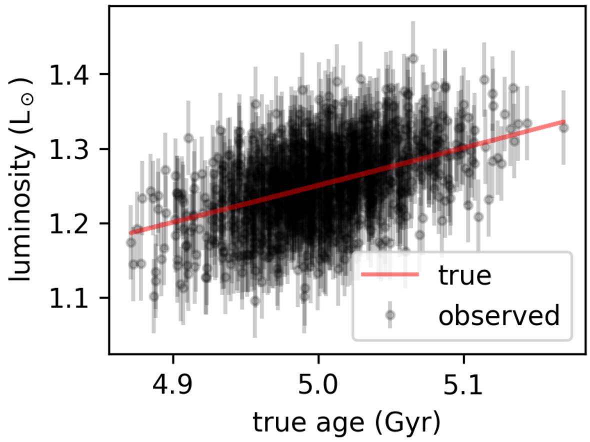


Figure 1.1: Luminosity against true ages of a fake stellar cluster. The true luminosities lie on the red line and the observed luminosities (black) have been artificially scattered by 0.05 $L_\odot$.

we get the posterior,

$$p(\boldsymbol{\tau}|\boldsymbol{L}) \propto \prod_{k=1}^{1000} p(L_k|\tau_k)\, p(\tau_k). \tag{1.9}$$

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,

$$p(\mu_\tau, \sigma_\tau, \boldsymbol{\tau}|\boldsymbol{L}) \propto p(\boldsymbol{L}|\boldsymbol{\tau})\, p(\boldsymbol{\tau}|\mu_\tau, \sigma_\tau)\, p(\mu_\tau, \sigma_\tau), \tag{1.10}$$

where,

$$p(\boldsymbol{\tau}|\mu_\tau, \sigma_\tau) = \prod_{k=1}^{1000} \mathcal{N}(\tau_k|\mu_\tau, \sigma_\tau). \tag{1.11}$$

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 1.9 and 1.10 and the Bayesian package `pymc3` [CITE]. The observed and true luminosities are plot 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 naïvely calculate a standard deviation from the resulting stellar ages. However, this overestimates the true standard deviation, getting 0.109 Gyr rather than 0.05 Gyr, because it includes the uncertainty in the individual ages. When we model the population mean and spread in the hierarchical model we get $\mu_\tau = 5.002 \pm 0.003$ Gyr and $\sigma_\tau = 0.042 \pm 0.007$ Gyr which are within $< 2\sigma$ of the truths. Therefore, the hierarachical model is a better way of determining population-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 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

How do we model stellar observables? A bit of history of the topic including Eddington (1926). Then Chandrasekhar 1939 and Schwartzschild 1958.

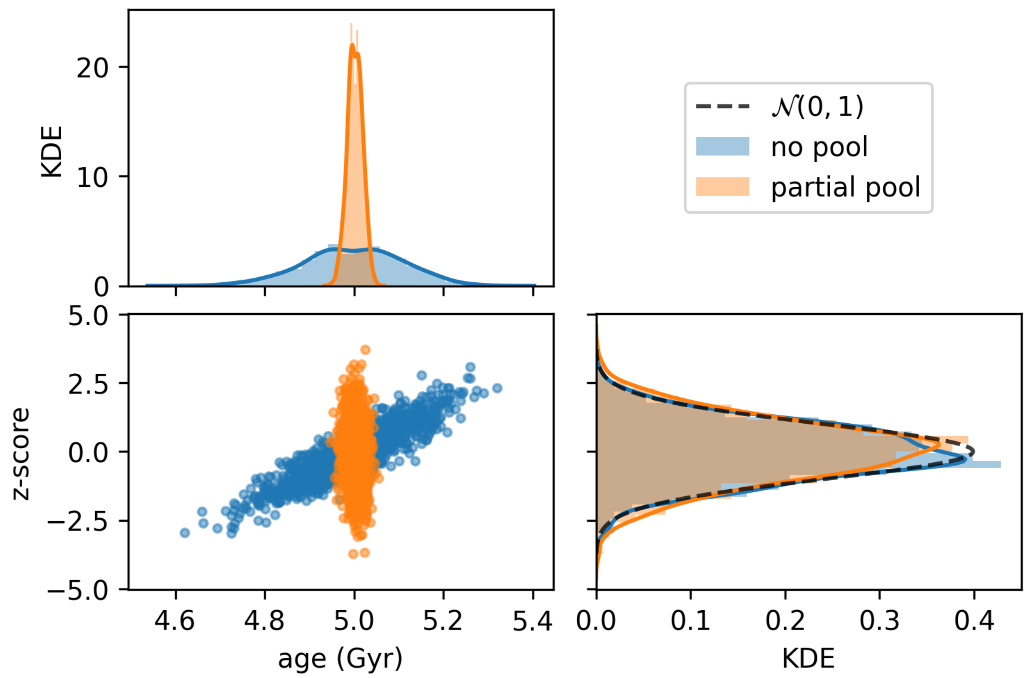


Figure 1.2: The z-score, $(\bar{\tau} - \tau_{\mathrm{true}})/s_{\tau}$, where $\bar{\tau}$ and s_{τ} are the respective sample mean and standard deviation of the posterior ages from each of the no- and partially-pooled models.

Introduction of stellar computational codes in the 1960s e.g. Iben and Ehrman 1962 and Kippenhahn et al. 1967 to solve the complicated differential equations

Today, many codes exist from one-dimensional [CITE] to three-dimensional and their outputs are often compared [CITE aarhus red giants challenge and Magic papers].

Why did we chose MESA?

What are the basic need-to-knows of stellar evolution in order to understand this work?

Basic scalings of observables with fundamentals, e.g. what is luminosity, effective temperature.

What are the evolutionary phases, a stellar track may be useful here?

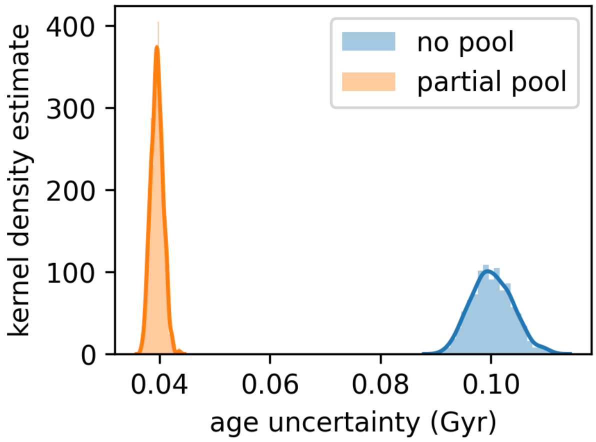


Figure 1.3: Standard deviations, s_τ of the age posteriors from both the no- and partially-pooled models.

What is X Y and Z and w

What is the mixing-length theory of convection? A diagram of the sun may be helpful.

What is element diffusion and why is it better to include it?

Finally, what is asteroseismology and why is it useful in stellar evolution?

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 using Hertzsprung-Russell (HR) diagrams. 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, Davies, et al., 2015, see, e.g.) to masses and radii from scaling relations () and fitting stellar models().

Solar-like oscillators are stars which typical 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 in the case of rotating bodies, azimuthal order, m. To simplify this discussion, I will assume the case where the star is non-rotating.

To first order in $\Delta\nu$, I may express the eigenfrequency as follows [CITE],

$$\nu_{nl} \simeq \Delta\nu \left(n + \frac{l}{2} + \epsilon \right) \tag{1.12}$$

where,

$$\Delta\nu = \left(2\int_0^R \frac{\mathrm{d}r}{c(r)}\right)^{-1} \tag{1.13}$$

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 1.13, it has been shown 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],

$$\langle\Delta\nu_{nl}\rangle \propto \langle\rho\rangle^{1/2}. \tag{1.14}$$

The diagram in Figure **??** shows the path of the asteroseismic wave fronts through a cross-section of a stellar interior. One can see how modes of different angular degree penetrate the star at different depths.

How do we observe ν?

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

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, Davies, 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

How do we observe stars? E.g. how do we determine luminosity from parallax and magnitude.

How do we determine effective temperature from spectroscopy?

How do we determine metallicity?

1.5.1 Detecting Asteroseismic Oscillation Modes

Why do we care?

Name some missions which were able to detect asteroseismic oscillations and review their limitations.

Give an example from PBjam of detecting modes of oscillation.

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

Ahumada, Romina et al. (2019). "The Sixteenth Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra". In: *arXiv e-prints* 1912, arXiv:1912.02905.

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.

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.

Asplund, Martin et al. (2009). "The Chemical Composition of the Sun". In: *ARA&A* 47, pp. 481–522. DOI: 10.1146/annurev.astro.46.060407.145222.

Astropy Collaboration, A. M. Price-Whelan, et al. (2018). "The Astropy Project: Building an Open-Science Project and Status of the v2.0 Core Package". In: *AJ* 156, p. 123. DOI: 10.3847/1538-3881/aabc4f.

Astropy Collaboration, Thomas P. Robitaille, et al. (2013). "Astropy: A Community Python Package for Astronomy". In: *A&A* 558, A33. DOI: 10.1051/0004-6361/201322068.

Bahcall, John N. and Roger K. Ulrich (1988). "Solar Models, Neutrino Experiments, and Helio-seismology". In: *Reviews of Modern Physics* 60, pp. 297–372. DOI: 10.1103/RevModPhys.60.297.

Balser, Dana S. (2006). “The Chemical Evolution of Helium”. In: *AJ* 132, pp. 2326–2332. DOI: 10.1086/508515.

Barber, C. Bradford, David P. Dobkin, and Hannu Huhdanpaa (1996). “The Quickhull Algorithm for Convex Hulls”. In: *ACM Trans. Math. Softw.* 22.4, pp. 469–483. DOI: 10.1145/235815.235821.

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.

Berger, Travis A., Daniel Huber, Eric Gaidos, et al. (2018). “Revised Radii of *Kepler* Stars and Planets Using *Gaia* Data Release 2”. en. In: *ApJ* 866.2, p. 99. DOI: 10.3847/1538-4357/aada83.

Berger, Travis A., Daniel Huber, Jennifer L. van Saders, et al. (2020). “The $Gaia$-$Kepler$ Stellar Properties Catalog I: Homogeneous Fundamental Properties for 186,000 $Kepler$ Stars”. In: *arXiv e-prints* 2001, arXiv:2001.07737.

Bossini, D. et al. (2019). “Age Determination for 269 *Gaia* DR2 Open Clusters”. en. In: *A&A* 623, A108. DOI: 10.1051/0004-6361/201834693.

Bovy, Jo et al. (2016). “On Galactic Density Modeling in the Presence of Dust Extinction”. In: *ApJ* 818, p. 130. DOI: 10.3847/0004-637X/818/2/130.

Brogaard, K. et al. (2012). “Age and Helium Content of the Open Cluster NGC 6791 from Multiple Eclipsing Binary Members. II. Age Dependencies and New Insights”. In: *Astronomy and Astrophysics* 543, A106. DOI: 10.1051/0004-6361/201219196.

Broomhall, A.-M. et al. (2011). “Solar-Cycle Variations of Large Frequency Separations of Acoustic Modes: Implications for Asteroseismology”. In: *MNRAS* 413, pp. 2978–2986. DOI: 10.1111/j.1365-2966.2011.18375.x.

Campante, T. L. et al. (2016). “Spin-Orbit Alignment of Exoplanet Systems: Ensemble Analysis Using Asteroseismology”. In: *ApJ* 819, p. 85. DOI: 10.3847/0004-637X/819/1/85.

Casagrande, Luca et al. (2007). “The Helium Abundance and $\Delta Y/\Delta Z$ in Lower Main-Sequence Stars”. In: *MNRAS* 382, pp. 1516–1540. DOI: 10.1111/j.1365-2966.2007.12512.x.

Chan, Victor C. and Jo Bovy (2020). “The Gaia DR2 Parallax Zero-Point: Hierarchical Modelling of Red Clump Stars”. In: *MNRAS* 493, pp. 4367–4381. DOI: 10.1093/mnras/staa571.

Chaplin, William J. and Andrea Miglio (2013). “Asteroseismology of Solar-Type and Red-Giant Stars”. In: *Annual Review of Astronomy and Astrophysics* 51, pp. 353–392. DOI: 10.1146/annurev-astro-082812-140938.

Chiappini, C. et al. (2015). “Young [α /Fe]-Enhanced Stars Discovered by CoRoT and APOGEE: What Is Their Origin?” en. In: *A&A* 576, p. L12. DOI: 10.1051/0004-6361/201525865.

Chiosi, C. and F. M. Matteucci (1982). “The Helium to Heavy Element Enrichment Ratio, Delta Y/Delta Z”. In: *A&A* 105, pp. 140–148.

Christensen-Dalsgaard, J. (1982). “Seismological Studies of the Sun and Other Stars”. In: *Advances in Space Research* 2, pp. 11–19. DOI: 10.1016/0273-1177(82)90250-2.

— (1984). “What Will Asteroseismology Teach Us”. In: p. 11.

Christensen-Dalsgaard, J. and D. O. Gough (1980). “Is the Sun Helium-Deficient”. In: *Nature* 288, pp. 544–547. DOI: 10.1038/288544a0.

Clevert, Djork-Arné, Thomas Unterthiner, and Sepp Hochreiter (2015). “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”. In: *arXiv e-prints* 1511, arXiv:1511.07289.

Corsaro, E., J. De Ridder, and R. A. García (2015). “High-Precision Acoustic Helium Signatures in 18 Low-Mass Low-Luminosity Red Giants: Analysis from More than Four Years of *Kepler* Observations★”. en. In: *A&A* 578, A76. DOI: 10.1051/0004-6361/201525922.

Das, Payel and Jason L. Sanders (2019). “MADE: A Spectroscopic Mass, Age, and Distance Estimator for Red Giant Stars with Bayesian Machine Learning”. en. In: *Mon. Not. R. Astron. Soc.* 484, p. 294. DOI: 10.1093/mnras/sty2776.

Davies, G. R., V. Silva Aguirre, 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.

Davies, G. R., W. J. Chaplin, et al. (2015). “Asteroseismic Inference on Rotation, Gyrochronology and Planetary System Dynamics of 16 Cygni”. In: *Monthly Notices of the Royal Astronomical Society* 446, pp. 2959–2966. DOI: 10.1093/mnras/stu2331.

Davies, G. R. and A. Miglio (2016). “Asteroseismology of Red Giants: From Analysing Light Curves to Estimating Ages”. In: *Astronomische Nachrichten* 337, p. 774. DOI: 10.1002/asna.201612371.

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.

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

Ferguson, Jason W. et al. (2005). “Low-Temperature Opacities”. In: *The Astrophysical Journal* 623, pp. 585–596. DOI: 10.1086/428642.

Foreman-Mackey, Daniel et al. (2013). “Emcee: The MCMC Hammer”. In: *Publications of the Astronomical Society of the Pacific* 125, p. 306. DOI: 10.1086/670067.

Frankel, Neige et al. (2020). “Keeping It Cool: Much Orbit Migration, yet Little Heating, in the Galactic Disk”. In: *arXiv e-prints* 2002, arXiv:2002.04622.

Gaia Collaboration, A. G. A. Brown, et al. (2018). “Gaia Data Release 2. Summary of the Contents and Survey Properties”. In: *Astronomy and Astrophysics* 616, A1. DOI: 10.1051/0004-6361/201833051.

Gaia Collaboration, T. Prusti, et al. (2016). “The Gaia Mission”. In: *A&A* 595, A1. DOI: 10.1051/0004-6361/201629272.

Ginsburg, Adam et al. (2019). “Astroquery: An Astronomical Web-Querying Package in Python”. In: *AJ* 157, p. 98. DOI: 10.3847/1538-3881/aafc33.

Glorot, Xavier, Antoine Bordes, and Y. Bengio (2011). “Deep Sparse Rectifier Neural Networks”. In: *Proc. 14th Int. Conf. Artif. Intell. Statisitics AISTATS 2011* 15, pp. 315–323.

Goupil, MarieJo (2019). “Seismic Probe of Transport Processes in Red Giants”. In: *ArXiv191009361 Astro-Ph*. arXiv: 1910.09361 [astro-ph].

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.

Grevesse, N. and A. J. Sauval (1998). “Standard Solar Composition”. In: *Space Sci. Rev.* 85, pp. 161–174. DOI: 10.1023/A:1005161325181.

Hahnloser, Richard H. R. et al. (2000). “Digital Selection and Analogue Amplification Coexist in a Cortex-Inspired Silicon Circuit”. In: *Nature* 405, pp. 947–951. DOI: 10.1038/35016072.

Hall, Oliver J. et al. (2019). “Testing Asteroseismology with Gaia DR2: Hierarchical Models of the Red Clump”. In: *MNRAS* 486, pp. 3569–3585. DOI: 10.1093/mnras/stz1092.

Hawkins, Keith et al. (2017). “Red Clump Stars and Gaia: Calibration of the Standard Candle Using a Hierarchical Probabilistic Model”. In: *MNRAS* 471, pp. 722–729. DOI: 10.1093/mnras/stx1655.

Hekker, S. et al. (2012). “Solar-like Oscillations in Red Giants Observed with Kepler: Influence of Increased Timespan on Global Oscillation Parameters”. In: *Astronomy and Astrophysics* 544, A90. DOI: 10.1051/0004-6361/201219328.

Hendriks, L. and C. Aerts (2019). “Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent Oscillation Modes”. In: *PASP* 131, p. 108001. DOI: 10.1088/1538-3873/aaeeec.

Hogg, David W. (2012). “Data Analysis Recipes: Probability Calculus for Inference”. In: *arXiv e-prints* 1205, arXiv:1205.4446.

Hogg, David W., Jo Bovy, and Dustin Lang (2010). “Data Analysis Recipes: Fitting a Model to Data”. In: *arXiv e-prints* 1008, arXiv:1008.4686.

Hogg, David W. and Daniel Foreman-Mackey (2018). “Data Analysis Recipes: Using Markov Chain Monte Carlo”. In: *ApJS* 236, p. 11. DOI: 10.3847/1538-4365/aab76e.

Hogg, David W., Adam D. Myers, and Jo Bovy (2010). "Inferring the Eccentricity Distribution". In: *ApJ* 725, pp. 2166–2175. DOI: 10.1088/0004-637X/725/2/2166.

Hon, Marc, Dennis Stello, and Jie Yu (2017). "Deep Learning Classification in Asteroseismology". en. In: *Mon. Not. R. Astron. Soc.* 469.4, pp. 4578–4583. DOI: 10.1093/mnras/stx1174.

Hon, Marc, Dennis Stello, and Joel C. Zinn (2018). "Detecting Solar-like Oscillations in Red Giants with Deep Learning". en. In: *ApJ* 859.1, p. 64. DOI: 10.3847/1538-4357/aabfdb.

Howe, Rachel et al. (2020). "Solar Cycle Variation of *N*max in Helioseismic Data and Its Implications for Asteroseismology". In: *MNRAS* 493, pp. L49–L53. DOI: 10.1093/mnrasl/slaa006.

Huber, D., T. R. Bedding, et al. (2011). "Testing Scaling Relations for Solar-like Oscillations from the Main Sequence to Red Giants Using Kepler Data". In: *ApJ* 743, p. 143. DOI: 10.1088/0004-637X/743/2/143.

Huber, D., D. Stello, et al. (2009). "Automated Extraction of Oscillation Parameters for Kepler Observations of Solar-Type Stars". In: *Communications in Asteroseismology* 160, p. 74.

Kahn, F. D. (1961). "Sound Waves Trapped in the Solar Atmosphere." In: *The Astrophysical Journal* 134, p. 343. DOI: 10.1086/147164.

Kinemuchi, K. et al. (2012). "Demystifying Kepler Data: A Primer for Systematic Artifact Mitigation". In: *Publications of the Astronomical Society of the Pacific* 124, p. 963. DOI: 10.1086/667603.

Kingma, Diederik P. and Jimmy Ba (2014). "Adam: A Method for Stochastic Optimization". In: *arXiv e-prints* 1412, arXiv:1412.6980.

Kjeldsen, H. and T. R. Bedding (1995). "Amplitudes of Stellar Oscillations: The Implications for Asteroseismology." In: *Astronomy and Astrophysics* 293, pp. 87–106.

Kuszlewicz, James S. et al. (2019). "Bayesian Hierarchical Inference of Asteroseismic Inclination Angles". In: *MNRAS* 488, pp. 572–589. DOI: 10.1093/mnras/stz1689.

Lebreton, Y., M.J. Goupil, and J. Montalbán (2014). "How Accurate Are Stellar Ages Based on Stellar Models?: I. The Impact of Stellar Models Uncertainties". en. In: *EAS Publications*

Series 65. Ed. by Y. Lebreton, D. Valls-Gabaud, and C. Charbonnel, pp. 99–176. DOI: 10.1051/eas/1465004.

Leistedt, Boris and David W. Hogg (2017). "Hierarchical Probabilistic Inference of the Color-Magnitude Diagram and Shrinkage of Stellar Distance Uncertainties". In: *AJ* 154, p. 222. DOI: 10.3847/1538-3881/aa91d5.

Licquia, Timothy C. and Jeffrey A. Newman (2015). "Improved Estimates of the Milky Way's Stellar Mass and Star Formation Rate from Hierarchical Bayesian Meta-Analysis". In: *ApJ* 806, p. 96. DOI: 10.1088/0004-637X/806/1/96.

Lightkurve Collaboration et al. (2018). "Lightkurve: Kepler and TESS Time Series Analysis in Python". In: *Astrophysics Source Code Library*, ascl:1812.013.

Lindegren, L. et al. (2018). "Gaia Data Release 2. The Astrometric Solution". In: *A&A* 616, A2. DOI: 10.1051/0004-6361/201832727.

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.

Magic, Z., A. Weiss, and M. Asplund (2015). "The Stagger-Grid: A Grid of 3D Stellar Atmosphere Models. III. The Relation to Mixing Length Convection Theory". In: *A&A* 573, A89. DOI: 10.1051/0004-6361/201423760.

Martinez, Gregory D. (2015). "A Robust Determination of Milky Way Satellite Properties Using Hierarchical Mass Modelling". In: *MNRAS* 451, pp. 2524–2535. DOI: 10.1093/mnras/stv942.

Morton, Timothy D. (2015). "Isochrones: Stellar Model Grid Package". In: *Astrophys. Source Code Libr.* ascl:1503.010.

Morton, Timothy D. and Joshua N. Winn (2014). "Obliquities of Kepler Stars: Comparison of Single- and Multiple-Transit Systems". In: *ApJ* 796, p. 47. DOI: 10.1088/0004-637X/796/1/47.

Mosser, B. et al. (2015). "Period Spacings in Red Giants. I. Disentangling Rotation and Revealing Core Structure Discontinuities". In: *Astronomy and Astrophysics* 584, A50. DOI: 10.1051/0004-6361/201527075.

Nissen, P. E. et al. (2017). “High-Precision Abundances of Elements in Kepler LEGACY Stars. Verification of Trends with Stellar Age”. In: *Astronomy and Astrophysics* 608, A112. DOI: 10.1051/0004-6361/201731845.

Ohn, Ilsang and Yongdai Kim (2019). “Smooth Function Approximation by Deep Neural Networks with General Activation Functions”. In: *Entropy* 21.7, p. 627. DOI: 10.3390/e21070627. arXiv: 1906.06903.

Paxton, Bill, Lars Bildsten, et al. (2011). “Modules for Experiments in Stellar Astrophysics (MESA)”. In: *ApJS* 192, p. 3. DOI: 10.1088/0067-0049/192/1/3.

Paxton, Bill, Matteo Cantiello, et al. (2013). “Modules for Experiments in Stellar Astrophysics (MESA): Planets, Oscillations, Rotation, and Massive Stars”. In: *The Astrophysical Journal Supplement Series* 208, p. 4. DOI: 10.1088/0067-0049/208/1/4.

Paxton, Bill, Pablo Marchant, et al. (2015). “Modules for Experiments in Stellar Astrophysics (MESA): Binaries, Pulsations, and Explosions”. In: *The Astrophysical Journal Supplement Series* 220, p. 15. DOI: 10.1088/0067-0049/220/1/15.

Paxton, Bill, Josiah Schwab, et al. (2018). “Modules for Experiments in Stellar Astrophysics (MESA): Convective Boundaries, Element Diffusion, and Massive Star Explosions”. In: *The Astrophysical Journal Supplement Series* 234, p. 34. DOI: 10.3847/1538-4365/aaa5a8.

Paxton, Bill, R. Smolec, et al. (2019). “Modules for Experiments in Stellar Astrophysics (MESA): Pulsating Variable Stars, Rotation, Convective Boundaries, and Energy Conservation”. In: *The Astrophysical Journal Supplement Series* 243, p. 10. DOI: 10.3847/1538-4365/ab2241.

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.

Qian, Ning (1999). “On the Momentum Term in Gradient Descent Learning Algorithms”. en. In: *Neural Networks* 12.1, pp. 145–151. DOI: 10.1016/S0893-6080(98)00116-6.

Ramachandran, Prajit, Barret Zoph, and Quoc V. Le (2017). “Searching for Activation Functions”. In: *arXiv e-prints* 1710, arXiv:1710.05941.

Reese, D. R. et al. (2016). "SpaceInn Hare-and-Hounds Exercise: Estimation of Stellar Properties Using Space-Based Asteroseismic Data". In: *A&A* 592, A14. DOI: 10.1051/0004-6361/201527987.

Ribas, Ignasi et al. (2000). "Chemical Composition of Eclipsing Binaries: A New Approach to the Helium-to-Metal Enrichment Ratio". In: *MNRAS* 313, pp. 99–111. DOI: 10.1046/j.1365-8711.2000.03195.x.

Rogers, F. J. and A. Nayfonov (2002). "Updated and Expanded OPAL Equation-of-State Tables: Implications for Helioseismology". In: *The Astrophysical Journal* 576, pp. 1064–1074. DOI: 10.1086/341894.

Rogers, Leslie A. (2015). "Most 1.6 Earth-Radius Planets Are Not Rocky". In: *ApJ* 801, p. 41. DOI: 10.1088/0004-637X/801/1/41.

Ruder, Sebastian (2016). "An Overview of Gradient Descent Optimization Algorithms". In: *arXiv e-prints* 1609, arXiv:1609.04747.

Salvatier, John, Thomas V. Wiecki, and Christopher Fonnesbeck (2016). "Probabilistic Programming in Python Using PyMC3". en. In: *PeerJ Comput. Sci.* 2, e55. DOI: 10.7717/peerj-cs.55.

Sandquist, Eric L. et al. (2020). "Variability in the Massive Open Cluster NGC 1817 from K2: A Rich Population of Asteroseismic Red Clump, Eclipsing Binary, and Main Sequence Pulsating Stars". In: *ArXiv200101839 Astro-Ph*. arXiv: 2001.01839 [astro-ph].

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.

Serenelli, Aldo M. and Sarbani Basu (2010). "Determining the Initial Helium Abundance of the Sun". In: *ApJ* 719, pp. 865–872. DOI: 10.1088/0004-637X/719/1/865.

Serenelli, Aldo, Jennifer Johnson, et al. (2017). "The First APOKASC Catalog of Kepler Dwarf and Subgiant Stars". In: *ApJS* 233, p. 23. DOI: 10.3847/1538-4365/aa97df.

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.

Silva Aguirre, V., G. R. Davies, 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.

Silva Aguirre, Víctor, Mikkel N. Lund, et al. (2017). "Standing on the Shoulders of Dwarfs: The Kepler Asteroseismic LEGACY Sample. II.Radii, Masses, and Ages". In: *ApJ* 835, p. 173. DOI: 10.3847/1538-4357/835/2/173.

Smith, Jeffrey C. et al. (2012). "Kepler Presearch Data Conditioning II - A Bayesian Approach to Systematic Error Correction". In: *Publications of the Astronomical Society of the Pacific* 124, p. 1000. DOI: 10.1086/667697.

Soderblom, David R. (2010). "The Ages of Stars". en. In: *ARA&A* 48.1, pp. 581–629. DOI: 10.1146/annurev-astro-081309-130806.

Soni, Harsh et al. (2020). "Phases and Excitations of Active Rod-Bead Mixtures: Simulations and Experiments". In: *ArXiv200100173 Cond-Mat*. arXiv: 2001.00173 [cond-mat].

Stumpe, Martin C. et al. (2012). "Kepler Presearch Data Conditioning I—Architecture and Algorithms for Error Correction in Kepler Light Curves". In: *Publications of the Astronomical Society of the Pacific* 124, p. 985. DOI: 10.1086/667698.

Sutskever, Ilya et al. (2013). "On the Importance of Initialization and Momentum in Deep Learning". In: ed. by Sanjoy Dasgupta and David McAllester. Vol. 28. Proceedings of Machine Learning Research. Atlanta, Georgia, USA: PMLR, pp. 1139–1147.

Townsend, R. H. D. and S. A. Teitler (2013). "GYRE: An Open-Source Stellar Oscillation Code Based on a New Magnus Multiple Shooting Scheme". In: *Monthly Notices of the Royal Astronomical Society* 435, pp. 3406–3418. DOI: 10.1093/mnras/stt1533.

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.

Ulrich, R. K. (1986). "Determination of Stellar Ages from Asteroseismology". In: *The Astrophysical Journal Letters* 306, pp. L37–L40. DOI: 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.

VanderPlas, Jacob T. (2018). “Understanding the Lomb-Scargle Periodogram”. In: *The Astrophysical Journal Supplement Series* 236, p. 16. DOI: 10.3847/1538-4365/aab766.

Verma, Kuldeep, Shravan Hanasoge, et al. (2016). “Asteroseismic Determination of Fundamental Parameters of Sun-like Stars Using Multilayered Neural Networks”. In: *MNRAS* 461, pp. 4206–4214. DOI: 10.1093/mnras/stw1621.

Verma, Kuldeep, Keyuri Raodeo, et al. (2019). “Helium Abundance in a Sample of Cool Stars: Measurements from Asteroseismology”. In: *MNRAS* 483, pp. 4678–4694. DOI: 10.1093/mnras/sty3374.

Viani, Lucas S. et al. (2018). “Investigating the Metallicity-Mixing-Length Relation”. In: *The Astrophysical Journal* 858, p. 28. DOI: 10.3847/1538-4357/aab7eb.

von Hippel, Ted et al. (2006). “Inverting Color-Magnitude Diagrams to Access Precise Star Cluster Parameters: A Bayesian Approach”. en. In: *ApJ* 645.2, pp. 1436–1447. DOI: 10.1086/504369.

Vrard, M. et al. (2015). “Helium Signature in Red Giant Oscillation Patterns Observed by *Kepler*”. en. In: *A&A* 579, A84. DOI: 10.1051/0004-6361/201425064.

White, Timothy R. et al. (2011). “Calculating Asteroseismic Diagrams for Solar-like Oscillations”. In: *The Astrophysical Journal* 743, p. 161. DOI: 10.1088/0004-637X/743/2/161.

Zinn, Joel C., Marc H. Pinsonneault, Daniel Huber, and Dennis Stello (2019). “Confirmation of the Gaia DR2 Parallax Zero-Point Offset Using Asteroseismology and Spectroscopy in the Kepler Field”. In: *ApJ* 878, p. 136. DOI: 10.3847/1538-4357/ab1f66.

Zinn, Joel C., Marc H. Pinsonneault, Daniel Huber, Dennis Stello, et al. (2019). “Testing the Radius Scaling Relation with Gaia DR2 in the Kepler Field”. In: *ApJ* 885, p. 166. DOI: 10.3847/1538-4357/ab44a9.

Appendix A

Accompanying Paper

MNRAS 000, 1–11 (2020) Preprint 16 September 2020 Compiled using MNRAS LaTeX style file v3.0

TBC: Hierarchically modelling *Kepler* dwarfs using machine learning to uncover helium enrichment in the solar neighbourhood

Alexander J. Lyttle,[1,2]⋆ Tanda Li,[1,2] Guy R. Davies,[1,2] Lindsey M. Carboneau[1,2] and TBC
[1]School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, UK
[2]Stellar Astrophysics Centre (SAC), Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

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

1 INTRODUCTION

Motivation - precise and accurate stellar fundamentals. Useful for e.g. galactic archaeology and exoplanet research.

Audience - astrophysicist with some knowledge Introduce new method and reference Guy's paper:

- Summarise the typical way in which stellar fundamentals are estimated and their pitfalls (e.g. discrete sampling, and assuming solar calibrated mixing-length parameter and helium enrichment)
  - Problems with grid-based-modelling (e.g. proper sampling)
  - assuming fixed DYDZ and MLT bad; attempts to interpolate, slow and hard to scale
- Why hierarchical models are good with examples of HBMs in astrophysics
- Advantage of HBM is to incorporate population-level distributions
- Why HBMs are difficult with stellar models.
- Introduce the neural network as a way to overcome these issues and give examples of neural networks to approximate models in astrophysics
- Highlight the novel element of this paper - the first application of combining a neural network emulator with a hierarchical model to provide shrinkage of fundamentals uncertainties and simultaneously study a helium enrichment relation
- Use a helium enrichment law prior, and assume a distribution of mixing-length of the population-level, to inform object-level parameters

Why do we care about helium and mixing-length? These parameters have a large (be quantitative) affect on stellar ages. Good stellar ages allow us to better study galactic archaeology (with citations).

Given that we are assuming a helium enrichment prior, give a brief summary of research into the helium enrichment and typical values for $\Delta Y/\Delta Z$. Note that in reality there may not be a linear law, and more may be studied in future work (or using a GP like in Guy's paper?). Why do we care about an enrichment law? Why is it physically justified?

Given that we are assuming a mixing-length distribution, mention this is mainly a nuisance parameter which we will marginalize over, since this differs depending on model physics. However, later justify a normal spread by referring to work (e.g. Magic) which shows little variation in the area of the HRD we are studying.

Outline the structure of the work. We are demonstrating the method on an asteroseismic sample of dwarfs and subgiants from Serenelli 2017. We first introduce the data and why we choose to use spectroscopy and asteroseismology. We then introduce the method, from the grid of stellar models

Why asteroseismology and why this particular set of Kepler-field dwarfs? Acknowledge selection bias but explain that with TESS providing an all-sky sample of solar-like oscillators this method can be extended to a much larger sample size.

Note: here is an example of a paper which would benefit from a value of the intrinsic spread in helium enrichment: (Zinn et al. 2019) 'Until such a time as the intrinsic scatter in helium enrichment can be determined, which... hinders a comparison between the theoretical metallicity trend and the observed radius agreement... the asteroseismic scaling relation radius does not require a metallicity term...'. In other words, they assume a helium enrichment law but this hinders their ability to study the seismic scaling relation correction.

2 DATA

We began with the sample of 415 stars from the first APOKASC catalogue of dwarfs and subgiants (S17). It is, to date, the most comprehensive sample of asteroseismic dwarfs and subgiant stars observed by the *Kepler* mission. We adopted the global asteroseismic parameters – the large frequency separation, $\Delta\nu$, and the

⋆ E-mail: ajl573@student.bham.ac.uk

frequency of maximum power, $\nu_{\max}$ – determined by S17, and references therein. We then cross matched the sample with *Gaia* Data Release 2 (DR2) for high-precision parallaxes, and the Apache Point Observatory Galaxy Evolution Experiment (APOGEE) catalogue to obtain spectroscopic metallicities and effective temperatures. Using the cross matched catalogue, we computed luminosities for the full sample with Two-Micron All Sky Survey (2MASS) photometry and selected a subsample of stars which we determined to be within our grid of stellar models (see Section 3.1).

We cross-matched the *Kepler* input catalogue (KIC) for the sample with the *Gaia* DR2 catalogue taking the nearest neighbours within a 4" radius [CITE GAIA]. We then adopted the *Gaia* DR2 parallaxes, assuming a zero-point offset of 0.05 mas in the sense that the *Gaia* parallaxes are underestimated. We chose this value in line with recent studies on the *Gaia* zero-point parallax offset in the *Kepler* field [CITATIONS].

We adopted spectroscopic metallicities, [M/H], and effective temperatures, $T_{\rm eff}$determined by the APOGEE stellar parameters and chemical abundances pipeline (ASPCAP) from the second data release of the fourth phase of the Sloan Digital Sky Survey (SDSS) otherwise known as Data Release 14 (DR14). We cross matched the APOGEE DR14 catalogue with our *Kepler*-*Gaia* DR2 cross match yielding spectroscopic parameters for 413 stars in the S17 sample.

We derived stellar surface gravity, g using the asteroseismic scaling relation for the frequency at maximum power, $\nu_{\max}$ [CITE],

$$\log g \approx \log g_{\odot} + \log\left(\frac{\nu_{\max}}{\nu_{\max,\odot}}\right) - \frac{1}{2}\log\left(\frac{T_{\rm eff}}{T_{\rm eff,\odot}}\right), \qquad (1)$$

where we adopted solar reference values of $\nu_{\max,\odot} = 3090 \pm 30\,\mu$Hz (Huber et al. 2011) $\log g_{\odot} = 4.44$ dex and $T_{\rm eff,\odot} = 5777$ K.

We determined luminosities for the sample using the direct method of ISOCLASSIFY [CITE HUBER]. We calculated absolute K_S-band magnitudes using K_s-band photometry from the 2MASS, distances from the *Gaia* DR2 parallaxes, and extinctions determined from the 3D galactic reddening maps of Green et al. (2019) [CITE]. We then determined absolute bolometric magnitudes by interpolating the MIST bolometric correction tables using ASPCAP [M/H]and $T_{\rm eff}$, asteroseismic $\log g$ and absolute magnitude as inputs. An uncertainty of 0.02 mag was assumed for both the extinctions and bolometric corrections, in line with typical uncertainties from randomly sampling the input data within their errors.

We selected a subset from the above sample which we determined to lie within the bounds of the model grid described in Section 3.1 using mass estimates from S17. We determined such "on-grid" stars where their estimated mass (from S17) and observed metallicity were within one standard deviation of the grid boundary, from 0.8 to 1.2 $M_{\odot}$ in mass and from −0.5 to 0.5 dex in metallicity. We also cut targets in the sample with an asteroseimic $\log g$ less than 3.8 dex to remove more evolved stars. The cut in mass was motivated by our choice of model physics described in Section 3.1. Stars with $M \gtrsim 1.15\,M_{\odot}$ are understood to have a convective, hydrogen-burning core, with some dependence on the choice of stellar physics [CITE Appourchaux](). Modelling stars with a convective core requires the consideration of extra mixing due to the overshooting of convective cells at the core boundary [CITE OVERSHOOT PAPERS], which is beyond the scope of this work.

The final sample comprised 81 stars, after removing stars with null observables and is shown in Table A1. The Hertzsprung-Russell diagram in Figure 1 shows the sample plot above a selection of stellar evolutionary tracks from the grid described in Section 3.1.

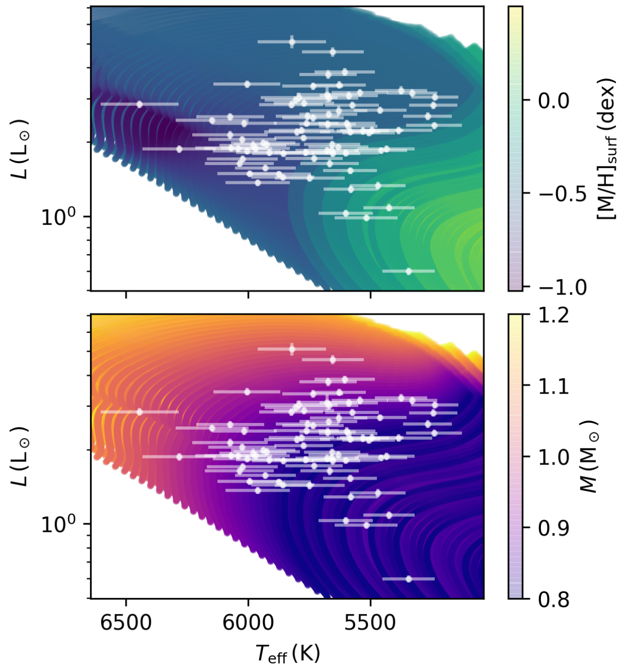


Figure 1. The luminosity, L against effeective temperature, $T_{\rm eff}$ of the sample of 81 *Kepler* dwarfs and subgiants plot against a subset of the grid of stellar models computed in Section 3.1. The top plot is coloured by stellar surface metallicity and the bottom plot is coloured by stellar mass.

Comparing the spectroscopic data in our sample to S17 we found that $T_{\rm eff}$ was offset by ∼ 170 K less than the preferred photometric temperature scale of S17, with a dispersion of ∼ 120 K. The median uncertainty in our sample of spectroscopic $T_{\rm eff}$ from DR14 is 125 K, which is greater than in previous ASPCAP data releases [CITE DR13].

3 METHODS

The principle goal of our new method was to reduce the statistical uncertainties on our inference of fundamental stellar parameters. To achieve this, we derived 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 use a stellar evolutionary codes to compute a grid of models (see 3.1). However, in order to sample our HBM using a modern Markov-chain Monte Carlo (MCMC) sampling method, we needed a way to continuously sample the grid and quickly evaluate the gradient. We approximated the grid of stellar models using an artificial neural network (ANN) in Section 3.2. Finally, we derive the statistical models compared in this work in Section 3.3, testing them on fake stars before running them on the sample of stars from Section 2.

3.1 Grid of stellar models

We built up a stellar model grid to train the NN model. The gird includes four independent model inputs: stellar mass (M), initial

Table 1. Stellar model grid parameters for training and test datasets.

Stellar model grid			
Input Parameter	Range	Increment	$N_{\rm track}$
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
$Y_{\rm init}$	0.24 – 0.32	0.02	5
$\alpha_{\rm mlt}$	1.5 – 2.5	0.2	6
Total			17,220

helium fraction ($Y_{\rm init}$), initial metallicity ([Fe/H]), and the mixing-length parameter ($\alpha_{\rm MLT}$). Ranges and grid steps of the four model inputs are summarised in Table 1. 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 4,000 evolutionary tracks with random input values in the parameter space for validating the results.

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:

$$\log(Z_{\rm init}/X_{\rm init}) = \log(Z/X)_\odot + [{\rm Fe/H}]. \tag{2}$$

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 $\alpha_{\rm MLT} = \ell_{\rm MLT}/H_p$ is the mixing-length parameter. We also applied the MESA predictive mixing scheme (Paxton et al. 2018, 2019) in the model computation.

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). And we computed radial modes (for $\ell = 0$) by solving the adiabatic stellar pulsation equations with the structural models generated by MESA. We computed 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.

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, $\boldsymbol{x}$ based on trainable weights, $\boldsymbol{w}$ 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,

$$x_{i,k} = f_k(\boldsymbol{w}_{i,k} \cdot \boldsymbol{x}_{k-1} + b_{i,k}) \tag{3}$$

where f_k is the *activation* function for the k-th layer, $\boldsymbol{w}_{i,k}$ 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,

$$\boldsymbol{x}_k = f_k(\boldsymbol{W}_k \cdot \boldsymbol{x}_{k-1} + \boldsymbol{b}_k), \tag{4}$$

where $\boldsymbol{W}_k$ 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,

$$\boldsymbol{Y} = \boldsymbol{W}_M \cdot f_{M-1}(\ldots f_1(\boldsymbol{W}_1 \cdot f_0(\boldsymbol{W}_0 \cdot \boldsymbol{X} + \boldsymbol{b}_0) + \boldsymbol{b}_1)) + \boldsymbol{b}_M \tag{5}$$

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, $\boldsymbol{D}_{\rm train} = \{(\boldsymbol{X}_1, \boldsymbol{Y}_1)\ldots(\boldsymbol{X}_{N_{\rm train}}, \boldsymbol{Y}_{N_{\rm train}})\}$ comprising $N_{\rm train}$ input-output pairs. We split the training data into pseudo-random batches, $\boldsymbol{D}_{\rm batch}$ 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, $E(\boldsymbol{D}_{\bf batch})$, 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 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, $\boldsymbol{D}$)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

To build the training dataset, we randomly sampled 7.736×10^6 points from the grid of stellar models, with the remaining $\sim 2 \times 10^6$ points given to the test dataset. Before we trained a given ANN architecture, we standardised the training dataset by subtracting the median, $\mu_{1/2}$ and dividing by the standard deviation, σ. We show the locations and scales of the standardisation for each parameter in Table B1.

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 1. We also constrained the validation dataset to $\tau < 15\,\mathrm{Gyr}$ because ages above $\sim 15\,\mathrm{yr}$ are unphysical and such points are sparse in the training data. This dataset was set aside and evaluated on the final ANN.

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.

$$f_\mathrm{evol} = \begin{cases} f_\mathrm{MS}, & f_\mathrm{MS} \leq 1 \\ 1 + \frac{\tau - \tau_\mathrm{MS}}{\tau_{\log g = 3.6} - \tau_\mathrm{MS}}, & f_\mathrm{MS} > 1 \end{cases} \tag{6}$$

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

$$f_\mathrm{MS} = \frac{\tau}{\tau_\mathrm{MS}}, \tag{7}$$

and τ_MS is the main sequence lifetime, defined in the models as the point where the central hydrogen fraction, $X_c < 0.XX$. In other words, a star with $f_\mathrm{evol} \in (0, 1]$ is in its main sequence phase, burning hydrogen in its core, and $f_\mathrm{evol} \in (1, 2]$ 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 the chosen loss function.

After preliminary tuning, we chose the ANN input and output parameters to be $\boldsymbol{X} = \{f_\mathrm{evol}, M, \alpha_\mathrm{mlt}, Y_\mathrm{init}, Z_\mathrm{init}\}$ and $\boldsymbol{Y} = \{\log(\tau), T_\mathrm{eff}, R, \Delta\nu, [\mathrm{M/H}]_\mathrm{surf}\}$ respectively. The inputs correspond to initial conditions in the stellar modelling code and the outputs correspond to surface conditions throughout the lifetime of the star, with the exception of age which is mapped from f_evol. A generalised form of our neural network is depicted in Figure 2. In the following subsection, we tune the number of hidden layers, M and neurons per layer, N along with other ANN parameters.

3.2.2 Tuning

We needed to train an ANN which would reproduce stellar observables according to our choice of physics with greater accuracy

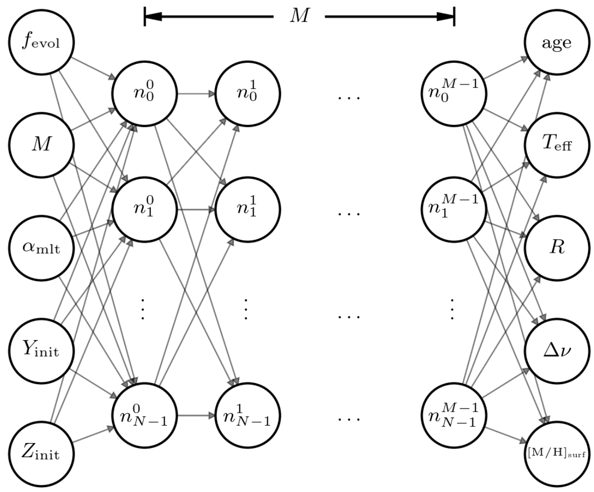


Figure 2. An artificial neural network comprising M hidden layers with N neurons per layer. Arrows connecting the nodes represent tunable weights.

than the typical observational precision and systematic uncertainties. 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.

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

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

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

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

Table 2. The median, $\mu_{1/2}$ and median absolute deviation estimator, $\sigma_{\mathrm{MAD}} = 1.4826 \cdot \mathrm{median}(|x_{\mathrm{true}} - x_{\mathrm{pred}}|)$, for each parameter, x between the validation dataset and the ANN predictions.

Error	$\mu_{1/2}$	σ_{MAD}
$\tau^{\mathrm{true}} - \tau^{\mathrm{pred}}$ (Gyr)	-0.0001	0.0066
$T_{\mathrm{eff}}^{\mathrm{true}} - T_{\mathrm{eff}}^{\mathrm{pred}}$ (K)	-0.0998	1.5949
$R^{\mathrm{true}} - R^{\mathrm{pred}}$ ($\mathrm{R}_{\odot}$)	0.0000	0.0009
$L^{\mathrm{true}} - L^{\mathrm{pred}}$ ($\mathrm{L}_{\odot}$)	0.0008	0.0021
$\Delta\nu^{\mathrm{true}} - \Delta\nu^{\mathrm{pred}}$ (μHz)	-0.0071	0.0836
$[\mathrm{M/H}]_{\mathrm{surf}}^{\mathrm{true}} - [\mathrm{M/H}]_{\mathrm{surf}}^{\mathrm{pred}}$ (dex)	0.0000	0.0007

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 robust to outliers. However, we tracked both metrics regardless of which was added to the loss function and found that MAE converged faster. Although it is less robust to outliers, we were able to achieve sufficient accuracy faster with MAE.

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-ouptut 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, $x_{\mathrm{true}} - x_{\mathrm{pred}}$.

We found good agreement between the validation dataset and ANN predictions, within observational uncertainties and model systematics. 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, such as the spread in error increasing at high temperatures. Such large errors mostly lie outside of the range of observables in our sample of *Kepler* dwarfs. 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.

We present the median, $\mu_{1/2}$ and MAD estimator, $\sigma_{\mathrm{MAD}} = 1.4826 \cdot \mathrm{median}(|x_{\mathrm{true}} - x_{\mathrm{pred}}|)$ in Table 2. The median is close to zero for all parameters, showing little systematic bias. The MAD is lower than, or of the order of observational uncertainties where relevant, and well below typical model systematics in age, effective temperature and radius [CITE e.g. RED GIANTS CHALLENGE]. Although the error in $\Delta\nu$ is $\sim 0.1\ \mu$Hz is comparable to observations with the best signal-to-noise, we note that the systematics due to not including the surface term in our calculation of $\Delta\nu$ from the stellar models would likely have a greater effect.

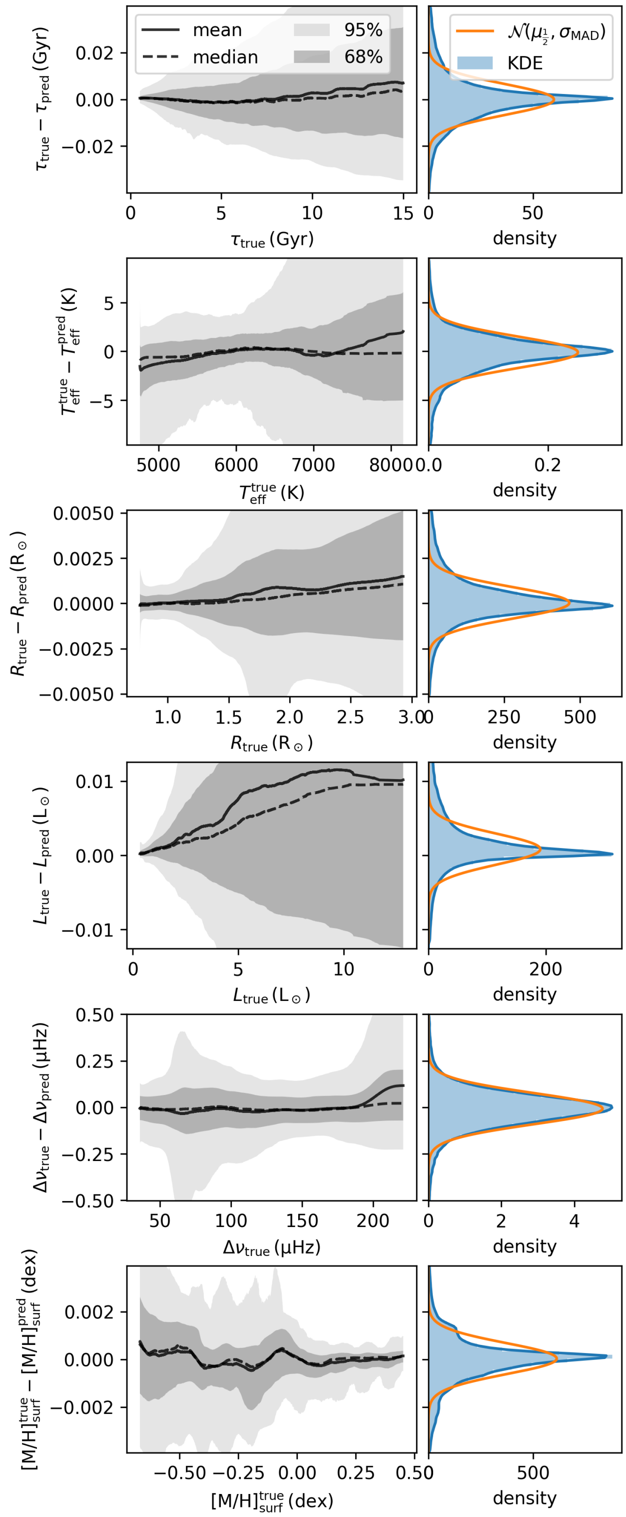


Figure 3. *Left*: the rolling error between the validation dataset (*true*) and the ANN predictions (*pred*) plotted against each parameter. *Right*: a kernel density estimate (KDE) of the validation error and a normal distribution centred on the median, $\mu_{1/2}$ with an estimator for the standard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

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 shrinkage 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 A1 (with and without Solar data for two of the models) and compared the results with that of S17.

Our first model was equivalent to modelling each star individually and featured no pooling; henceforth, we refer to it as the no-pooled (NP) model (see Section 3.3.1). We then derived two hierarchical Bayesian models (HBMs) which use population-level parameters to describe their distribution in the sample. Both of these models partially-pooled helium using a linear enrichment law. We drew the initial helium fraction for each star from a normal distribution with a mean described by the enrichment law and standard deviation representing its spread. Similarly, we partially-pooled the mixing-length theory parameter, 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.

For both the PP and MP models, we ran versions with and without data for the Sun included in the population. We adopted the solar data shown in Table **??** as observables with uncertainties limited to the accuracy of the neural network or typical stellar model systematics. We refer to the partial-pooled and max-pooled models including the Sun as PPS and MPS respectively.

We sampled the posterior for each model using the No U-Turn Sampler (NUTS) of PyMC4 (a new version of PyMC3 based on Tensorflow [CITE PYMC3 AND TENSORFLOW]). Initially, we modelled each star individually in order to identify stars outside the grid range and highlight other sampling problems. We flagged stars with output parameters near the gird boundaries. We also flagged stars with many model divergences, indicative of problems during sampling. Then, we modelled the remaining sample of 65 stars using each of the NP, PP, PPS, MP and MPS models. When we encountered problems with model convergence in the pooled models, we removed stars with large values of the Gelman-Rubin diagnostic (Gelman & Rubin 1992) and reran.

3.3.1 No-pooled model

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

$$p(\boldsymbol{\theta}_i|\boldsymbol{d}_i) \propto p(\boldsymbol{\theta}_i)\, p(\boldsymbol{d}_i|\boldsymbol{\theta}_i), \tag{8}$$

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

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

$$p(\boldsymbol{\theta}_i) = \prod_{k=1}^{N_\theta} \left[\theta_{k,\mathrm{min}} + (\theta_{k,\mathrm{max}} - \theta_{k,\mathrm{min}})\, \mathcal{B}(\theta_{k,i}|1.2, 1.2)\right], \tag{9}$$

where the beta distribution is defined as,

$$\mathcal{B}(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\mathrm{d}u}. \tag{10}$$

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

We made predictions for each star using the trained ANN, $\{\log(\tau)_i, T_{\mathrm{eff},i}, R_i, \Delta\nu_i, [\mathrm{M/H}]_{\mathrm{surf},i}\} = \boldsymbol{f}_{\mathrm{ANN}}(\boldsymbol{\theta}_i)$, 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 $\boldsymbol{\mu}_{d,i} = \boldsymbol{f}(\boldsymbol{\theta}_i)$. Thus, we write the likelihood we observe any $\boldsymbol{d}_i$ with known uncertainty, $\boldsymbol{\sigma}_{d,i}$ given the model as,

$$p(\boldsymbol{d}_i|\boldsymbol{\theta}_i) = \prod_{k=1}^{N_{\mathrm{obs}}} \frac{1}{\sigma_{d,k,i}\sqrt{2\pi}} \exp\left[-\frac{(d_{k,i} - \mu_{d,k,i})^2}{2\sigma_{d,k,i}^2}\right], \tag{11}$$

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

Using the above model, we sampled from the posterior for each individual star separately and then together as a population of N_{stars} stars,

$$p(\boldsymbol{\Theta}|\boldsymbol{D}) = \prod_{i=1}^{N_{\mathrm{stars}}} p(\boldsymbol{\theta}_i|\boldsymbol{d}_i). \tag{12}$$

Modelling stars separately allowed us to identify poorly sampled posteriors, whether the model indicated a fit outside the given input range, or other sampling issues. Once a refined sample was chosen, we modelled the sample all together as a natural application of the ANN through the use of batching. We modelled the ANN inputs as independent distributions, from which the random variables were batched together and passed through the ANN to produce predictions for each star. 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, $\boldsymbol{Y}_{\mathrm{init}}$ and mixing-length theory parameter, $\boldsymbol{\alpha}_{\mathrm{mlt}}$ are drawn from a common distribution characterised by the set of hyperparameters, $\boldsymbol{\phi} = \{\Delta Y/\Delta Z, Y_P, \sigma_Y, \mu_\alpha, \sigma_\alpha\}$. Thus, Bayes' theorem becomes,

$$p(\boldsymbol{\phi}, \boldsymbol{\Theta}|\boldsymbol{D}) \propto p(\boldsymbol{\phi})\, p(\boldsymbol{Y}_{\mathrm{init}}, \boldsymbol{\alpha}_{\mathrm{mlt}}|\boldsymbol{\phi})\, p(\boldsymbol{f}_{\mathrm{evol}}, \boldsymbol{M}, \boldsymbol{Z})\, p(\boldsymbol{D}|\boldsymbol{\Theta}), \tag{13}$$

where $\boldsymbol{\Theta}$ is the same as in the NP model, i.e. each object-level parameter, $\boldsymbol{\theta}_j = \{\theta_{j,i}\}_{i=1}^{N_{\mathrm{stars}}}$.

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,

$$p(\boldsymbol{Y}_{\text{init}}, \boldsymbol{\alpha}_{\text{mlt}}|\boldsymbol{\phi}) = p(\boldsymbol{Y}_{\text{init}}|\boldsymbol{\mu}_Y, \sigma_Y)\, p(\boldsymbol{\alpha}_{\text{mlt}}|\mu_\alpha, \sigma_\alpha). \tag{14}$$

Regarding the first term of this equation, the mean initial helium follows a linear enrichment law with respect to the initial fraction of heavy-elements for a given star,

$$\boldsymbol{\mu}_Y = Y_P + \frac{\Delta Y}{\Delta Z}\boldsymbol{Z}_{\text{init}}, \tag{15}$$

where Y_P is the primordial helium abundance fraction and $\Delta Y/\Delta Z$ is the so-called enrichment ratio. Therefore, we may write the prior PDF of initial helium given its population-level hyperparameters as,

$$p(\boldsymbol{Y}_{\text{init}}|\boldsymbol{Z}_{\text{init}}, \Delta Y/\Delta Z, Y_P, \sigma_Y) = \prod_{i=1}^{N_{\text{stars}}} \mathcal{N}(Y_{\text{init},i}|\mu_{Y,i}, \sigma_Y) \tag{16}$$

We justified this assumption based on theoretical and empirical evidence for a linear enrichment law [CITE], but taking into account an intrinsic spread, σ_Y about this law due to random variations in chemical abundance throughout the interstellar medium.

Similarly, for the second term of Equation 14, 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 $\boldsymbol{\alpha}_{\text{mlt}}$ as,

$$p(\boldsymbol{\alpha}_{\text{mlt}}|\mu_\alpha, \sigma_\alpha) = \prod_{i=1}^{N_{\text{stars}}} \mathcal{N}(\alpha_{\text{mlt},i}|\mu_\alpha, \sigma_\alpha) \tag{17}$$

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,

$$\begin{aligned}
\Delta Y/\Delta Z &\sim 4.0\,\mathcal{B}(1.2, 1.2),\\
Y_P &\sim \mathcal{N}(0.247, 0.1),\\
\sigma_Y &\sim \mathcal{LN}(0.01, 1.0),\\
\mu_\alpha &\sim 1.5 + \mathcal{B}(1.2, 1.2),\\
\sigma_\alpha &\sim \mathcal{LN}(0.1, 1.0),
\end{aligned}$$

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

$$\mathcal{LN}(x|m, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{\ln(x/m)^2}{2\sigma^2}\right]. \tag{18}$$

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, $\boldsymbol{\phi} = \{\Delta Y/\Delta Z, Y_P, \sigma_Y, \alpha_{\text{mlt}}\}$. The posterior distribution of the model takes the same form as in Equation 13 except that now,

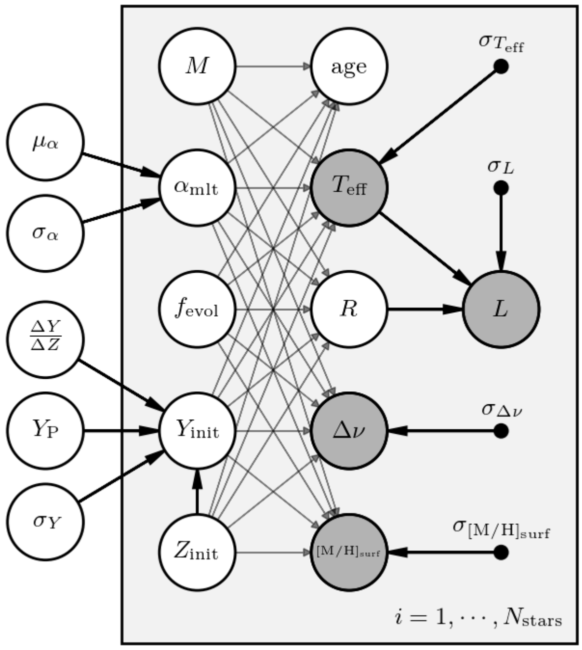


Figure 4. A probabilistic graphical model (PGM) of the partially-pooled (PP) hierarchical model. Nodes outside of the grey rectangle represent the hyperparameters in the model. Nodes inside the grey rectangle represent individual stellar parameters. Dark grey nodes represent observables which each have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

$$p(\boldsymbol{\alpha}_{\text{mlt}}|\alpha_{\text{mlt}}) = \prod_{i=1}^{N_{\text{stars}}} \delta(\alpha_{\text{mlt},i}|\alpha_{\text{mlt}}) \tag{19}$$

where $\delta(x|\alpha)$ is defined as,

$$\delta(x|\alpha) = \begin{cases} +\infty, & x = \alpha \\ 0, & x \neq \alpha \end{cases} \tag{20}$$

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

4 RESULTS

We obtained model stellar parameters for the set of test stars and the sample obtained in Section 2. For each of the NP, PP and MP models, we took the median and 68 percent credible region from the marginalised posterior samples. In this section, we present the results from the test dataset and compare them to the tue values. We then present the results for the *Kepler* dwarfs and compare the results to S17.

4.1 Test stars

We generated parameters for 100 stars randomly assuming a true population distribution of Y_{init} and α_{mlt}. We then modelled the whole population using each of our NP, PP and MP models. We also randomly selected 10 and 50 stars from the population and modelled them separately using the PP model to show the shrinkage in uncertainties with increasing sample size. In this section, we present the results for the set of test stars and show that our method is able to recover the true values well.

We found that each pooled model was able to recover the true population hyperparameters within uncertainty, as shown in Figure C1.

We found good agreement with the true values for the test stars. We found that the NP model underestimates the parameter residuals given their uncertainties. We expected this was due to the poorly constrained parameters, $\alpha_{\rm mlt}$ and $Y_{\rm init}$, for which their marginalised posteriors were not Gaussian (see Figure [TODO: MAKE CORNER PLOT FOR TEST STARS]). Boundary effects for these parameters bias their mean values towards the centre of the prior, which reduces the difference between the truth and sample mean over the uncertainty. We found that the effects of this where mitigated in the PP and MP models because pooling the parameters among the stars improved their uncertainties and reduced the boundary effects. This is apparent in the marginalised distributions of $\alpha_{\rm mlt}$ and $Y_{\rm init}$ for the PP model compared to the NP model.

We confirmed that the statistical uncertainties on the parameters described hierarchically, $\alpha_{\rm mlt}$ and $Y_{\rm init}$ reduce with the number of stars modelled, $N_{\rm stars}$ by a factor of $\sqrt{N_{\rm stars}}$, as shown in 5. The effects of the uncertainty shrinkage propagate to other fundamental parameters, particularly the age and mass which are influenced heavily by the precision in $\alpha_{\rm mlt}$ and $Y_{\rm init}$. Notably, the statistical uncertainties from the NP model are larger in all parameters, except $Z_{\rm init}$ which is dominated by the uncertainty in observed metallicity.

Fitting the helium enrichment law and spread in mixing-length to the NP model results recovered the true hyperparameters with higher precision than the PP and MP models. However, fitting this way limited the precision of the stellar parameters, wheras the hierarchical models reduced the uncertainties by roughly a factor of $\sqrt{N_{\rm stars}}$.

4.2 The sample

With confidence that the models were able to obtain accurate stellar parameters, in accordance with our choice of stellar model physics, we present results for the sample of 81 *Kepler* dwarfs for each of our statistical models. We ran each of our pooled models with and without the Sun and found it greatly affected the resulting hyperparameters. We then compared our results to S17 and found that our method was able to improve on their statistical uncertainties, despite the additional free parameters.

We obtained values for the hyperparameters for each of the models and present them in Table 3 along with their upper and lower 68 per cent confidence regions. We fit the same hyperparameters from the PP model to the NP model results for the purpose of comparison. However, as discovered in Section 4.1, the NP model results suffer from boundary effects which we expect to skew the hyperparameter fit. The joint-posterior distributions between the hyperparameters for each model are shown in Figure 6.

The helium hyperparameters were reasonably consistent between models excluding the Sun. We found the slope of the enrichment law to be $\Delta Y/\Delta Z \sim 1.6$ for MP and PP which is consistent with other values in the literature [CITE]. When we introduce the Sun, we see the slope reduces to $\Delta Y/\Delta Z \lesssim 1$. The addition of the Sun is more obvious in the MPS model, yet the similar σ_Y between all the models implies that the initial helium of the Sun is consistent with the spread in helium among the rest of the sample. In Figure 7, we show random samples from the hyperparameter posteriors for helium. Here, the smaller $\Delta Y/\Delta Z$ for the models including the Sun is evident, especially in the MPS model. We also note the anti-correlation between $\Delta Y/\Delta Z$ and μ_α or $\alpha_{\rm mlt}$ visible in the joint posterior distributions plot in Figure 6.

We found very little difference in results for μ_α and $\alpha_{\rm mlt}$ between the PP and MP models respectively. However, when we added the sun, both the PPS and MPS models yielded significantly different results. The MPS model obtained a global value of $\alpha_{\rm mlt} = 2.09 \pm 0.03$, whereas the PPS model found $\mu_\alpha = 1.90 \pm 0.09$ with $\sigma_\alpha = 0.13^{+0.06}_{-0.05}$. When we modelled the Sun separately, it yielded a value of $\alpha_{\rm mlt} = 2.11 \pm 0.01$ which is far from $\alpha_{\rm mlt} = 1.72^{+0.08}_{-0.07}$ obtained by the MP model. When we partially pooled $\alpha_{\rm mlt}$ without the Sun, allowing for a population mean and spread, we get $\mu_\alpha = 1.74^{+0.08}_{-0.07}$ and $\sigma_\alpha = 0.06^{+0.05}_{-0.03}$, which is consistent with the MP model. However, when we add the Sun, the PPS model copes with the difference between the Solar $\alpha_{\rm mlt}$ and the rest of the sample by doubling the spread in $\alpha_{\rm mlt}$ over the PP model. The larger spread in the PPS model is apparent in Figure 8 which shows hyperparameter posterior samples for the two PP models.

We present the results for the NP model in Table **??** and for the four pooled models in Tables E1 to E4. We compared our results for each model in this work with the ages, masses and radii from the photometric effective temperature scale results of S17.

Firstly, we compared the statistical uncertainties as shown in Figure 9. Despite the additional free parameters, we obtained median uncertainties of 20 per cent in age, 4.5 per cent in mass and 1.9 per cent in radius for the NP model, which were comparable to **?**. However, we improved on the precision of S17 by factor of ~ 1.5 with the inclusion of pooling. We expected this because we saw the a similar shrinkage in uncertainty in the test stars (see the bottom row of Figure 5). We saw the highest precision in the MP and MPS models of 9.2 per cent in age, 2.8 per cent in mass and 1.3 per cent in radius when including the Sun. We expected the MP models to be the most precise because it drops all uncertainty in $\alpha_{\rm mlt}$ which is assumed the same for all stars. Partially pooling $\alpha_{\rm mlt}$ in the PP and PPS resulted in uncertainties of 15.6 per cent in age, 3.0 per cent in mass and 1.4 per cent in radius.

We then compared the results from the PPS model with that of S17 in Figure 10. The plot shows the distribution of both results, and the normalised residuals between the two with an $\mathcal{N}(0,1)$ distribution. We found good agreement between the models, with our model favouring similar ages but slightly higher masses and radii by 1.2 per cent and 1.0 per cent respectively.

5 DISCUSSION

We have shown that, utilising hierarchical models, it is possible to reduce the statistical uncertainties on fundamental stellar parameters by sharing information between the stars. However, as such uncertainties reduce, systematics begin to dominate. We stress that the next challenge is to robustly model the systematic uncertainties between stellar modelling codes and within observables. Until then, we hesitate to publish one preferred set of parameters for our sample of stars. Instead, this paper is a demonstration of our method.

Note, possibility of a systematic shift in effective temperature which could affect our results. E.g. we see our radii are greater than S17 yet our temperatures are lower. Could this be the radii in the model compensating for a shift in temperature at the same luminosity.

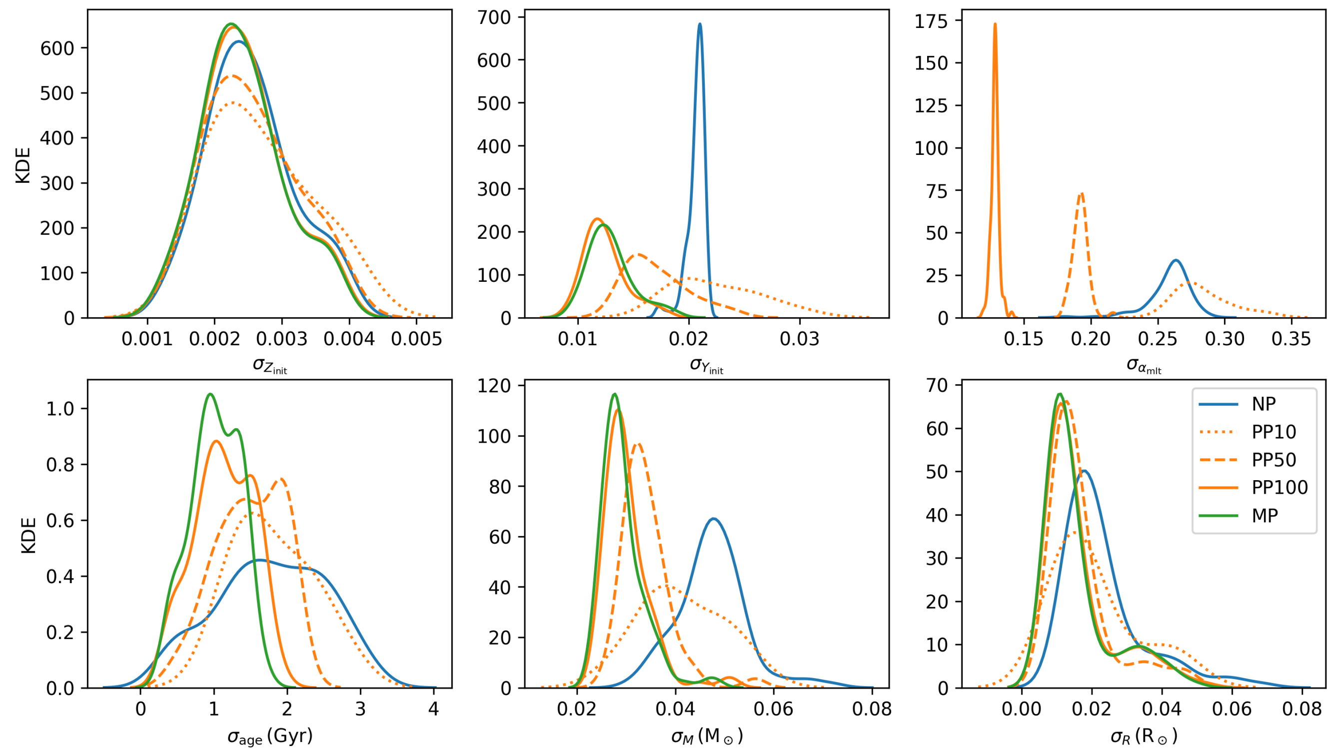


Figure 5. Kernal density estimates (KDEs) showing the shrinkage of statistical uncertainties between models of the sample of test stars. The PP model was run with 10, 50 and 100 stars and is denoted PP10, PP50, and PP100 respectively. The NP and MP models were both ran with the full set of 100 stars.

Table 3. Hyperparameter results for each model in descending order of the helium enrichment ratio, $\Delta Y/\Delta Z$.

Model	$\Delta Y/\Delta Z$	σ_Y	μ_α	σ_α	$\alpha_{\rm mlt}$
NP	$1.69^{+0.21}_{-0.21}$	$0.0074^{+0.0026}_{-0.0022}$	$1.954^{+0.040}_{-0.041}$	$0.065^{+0.030}_{-0.024}$	—
MP	$1.60^{+0.45}_{-0.42}$	$0.0051^{+0.0044}_{-0.0027}$	—	—	$1.728^{+0.077}_{-0.066}$
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	$1.05^{+0.28}_{-0.25}$	$0.0045^{+0.0038}_{-0.0023}$	$1.900^{+0.095}_{-0.088}$	$0.133^{+0.057}_{-0.047}$	—
MPS	$0.76^{+0.24}_{-0.27}$	$0.0049^{+0.0039}_{-0.0025}$	—	—	$2.088^{+0.031}_{-0.029}$

6 CONCLUSIONS

ACKNOWLEDGEMENTS

REFERENCES

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

Clevert D.-A., Unterthiner T., Hochreiter S., 2015, arXiv e-prints, 1511, arXiv:1511.07289

Ferguson J. W., Alexander D. R., Allard F., Barman T., Bodnarik J. G., Hauschildt P. H., Heffner-Wong A., Tamanai A., 2005, The Astrophysical Journal, 623, 585

Gelman A., Rubin D. B., 1992, Statistical Science, 7, 457

Glorot X., Bordes A., Bengio Y., 2011, Proc. 14th Int. Conf. Artif. Intell. Statisitics AISTATS 2011, 15, 315

Hahnloser R. H. R., Sarpeshkar R., Mahowald M. A., Douglas R. J., Seung H. S., 2000, Nature, 405, 947

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

Kingma D. P., Ba J., 2014, arXiv e-prints, 1412, arXiv:1412.6980

Magic Z., Weiss A., Asplund M., 2015, A&A, 573, A89

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

Qian N., 1999, Neural Networks, 12, 145

Rogers F. J., Nayfonov A., 2002, The Astrophysical Journal, 576, 1064

Ruder S., 2016, arXiv e-prints, 1609, arXiv:1609.04747

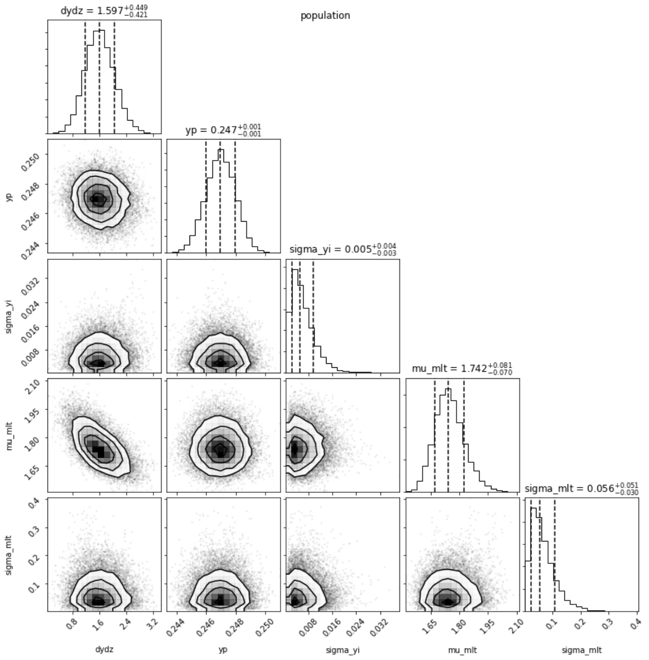
Serenelli A., et al., 2017, ApJS, 233, 23

Townsend R. H. D., Teitler S. A., 2013, Monthly Notices of the Royal Astronomical Society, 435, 3406

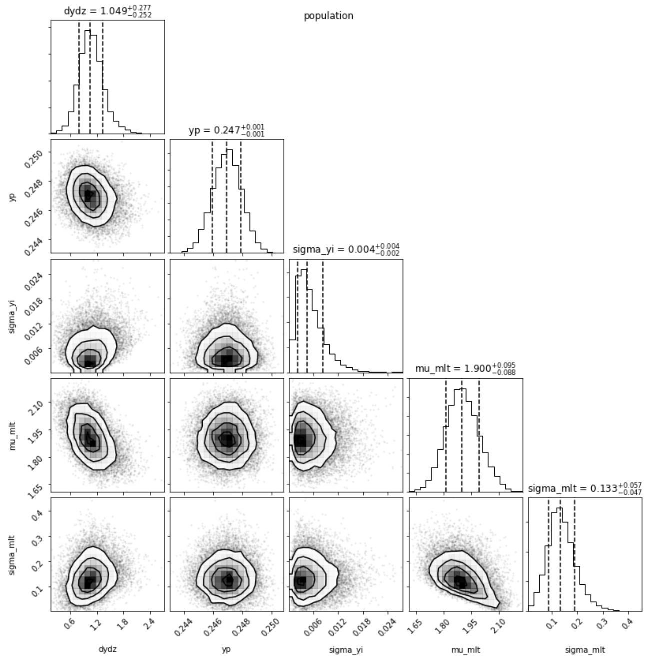
Viani L. S., Basu S., Ong J. M. J., Bonaca A., Chaplin W. J., 2018, The Astrophysical Journal, 858, 28

White T. R., Bedding T. R., Stello D., Christensen-Dalsgaard J., Huber D., Kjeldsen H., 2011, The Astrophysical Journal, 743, 161

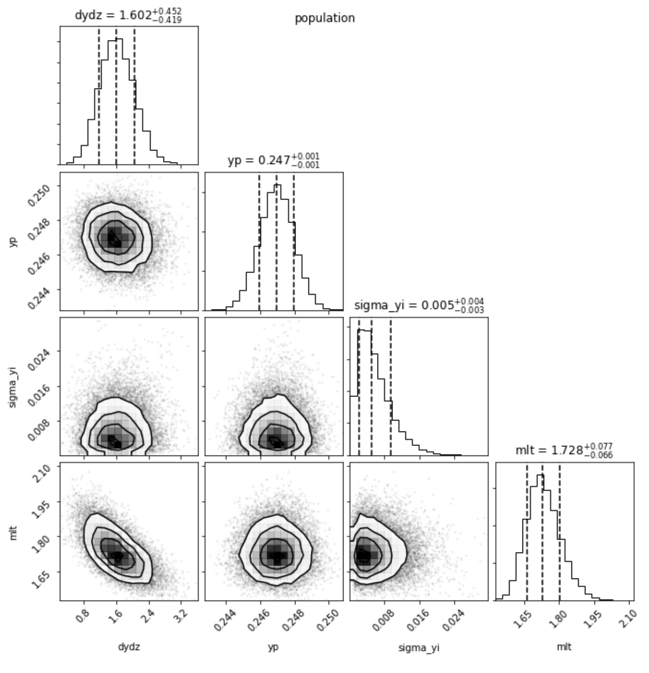
Zinn J. C., Pinsonneault M. H., Huber D., Stello D., Stassun K., Serenelli A., 2019, ApJ, 885, 166



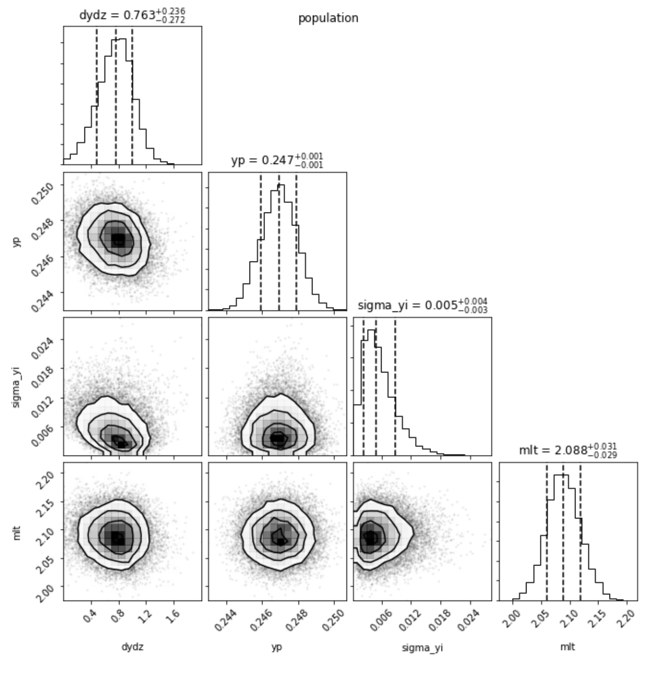
(a) PP model



(b) PPS model



(c) MP model



(d) MPS model

Figure 6. Corner plots showing the joint and marginalised sampled posterior distributions for the hyperparameters for both the PP and MP models, with and without the Sun. The vertical dashed lines give the 16th, 50th and 84th percentiles.

APPENDIX A: DATA

APPENDIX B: NEURAL NETWORK

APPENDIX C: TESTING THE METHOD

We tested the ability of the method to recover stellar fundamental properties in accordance with our choice of stellar evolution code and physics.

APPENDIX D: SOLAR RESULTS

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.

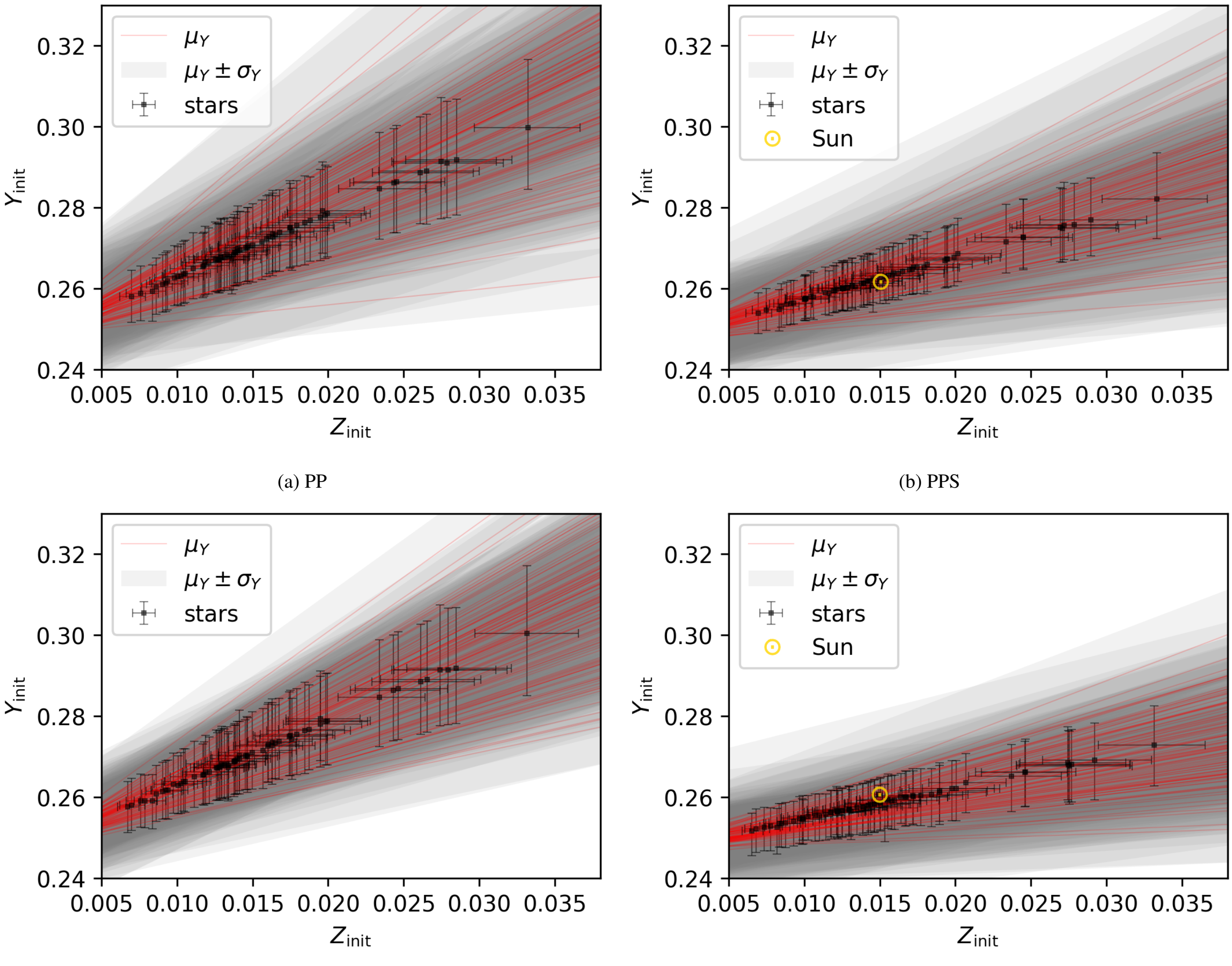


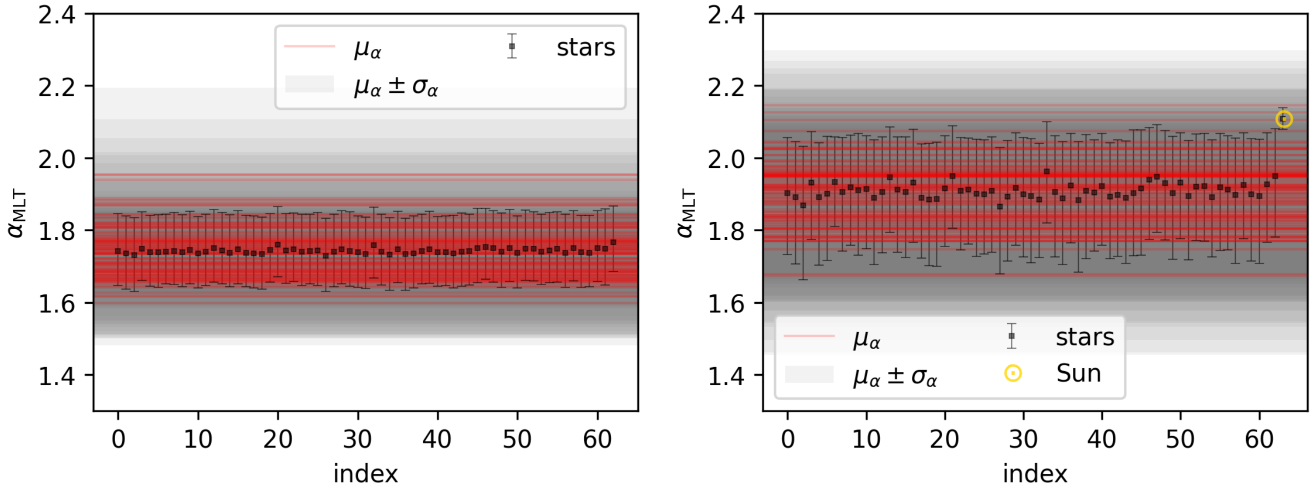
Figure 7. The results for initial helium against initial heavy-element fraction for each star. 100 random samples from the posterior for the population mean, $\mu_Y = Y_P + (\Delta Y/\Delta Z)Z_{\rm init}$ and spread, $\mu_Y \pm \sigma_Y$ are shown in red and grey respectively. The Sun is shown by the solar symbol, $\odot$ in yellow for the models which included the Sun.

Table A1. The observables and their respective uncertainties for the 10 stars in sample of 81 stars. The whole table is available online.

Name	$T_{\rm eff}$ (K)	$\sigma_{T_{\rm eff}}$ (K)	L (L$_\odot$)	σ_L (L$_\odot$)	$\Delta\nu$ (μHz)	$\sigma_{\Delta\nu}$ (μHz)	[M/H] (dex)	$\sigma_{\rm [M/H]}$ (dex)	log g (dex)	$\sigma_{\log g}$ (dex)
KIC10079226	5928.84	124.84	1.57	0.05	116.04	0.73	0.16	0.07	4.36	0.01
KIC10215584	5666.92	119.33	1.64	0.06	115.16	2.83	0.04	0.07	4.27	0.09
KIC10319352	5456.17	106.65	1.85	0.06	78.75	1.73	0.27	0.06	3.96	0.13
KIC10322381	6146.79	148.58	2.44	0.08	86.64	6.57	-0.32	0.08	4.19	0.04
KIC10417911	5628.26	109.99	3.41	0.12	56.14	2.10	0.34	0.07	3.94	0.02
KIC10732098	5669.65	119.28	3.02	0.12	62.18	1.92	0.05	0.07	3.96	0.02
KIC10794845	6035.12	140.46	1.64	0.06	116.35	6.70	-0.21	0.08	4.40	0.11
KIC10963065	6039.78	139.10	1.88	0.06	103.21	0.11	-0.16	0.08	4.30	0.01
KIC10971974	5748.00	142.40	1.43	0.05	106.63	3.31	-0.07	0.09	4.32	0.04
KIC11021413	5329.18	102.98	3.16	0.11	48.16	1.29	0.01	0.04	3.84	0.01

APPENDIX E: MODEL RESULTS

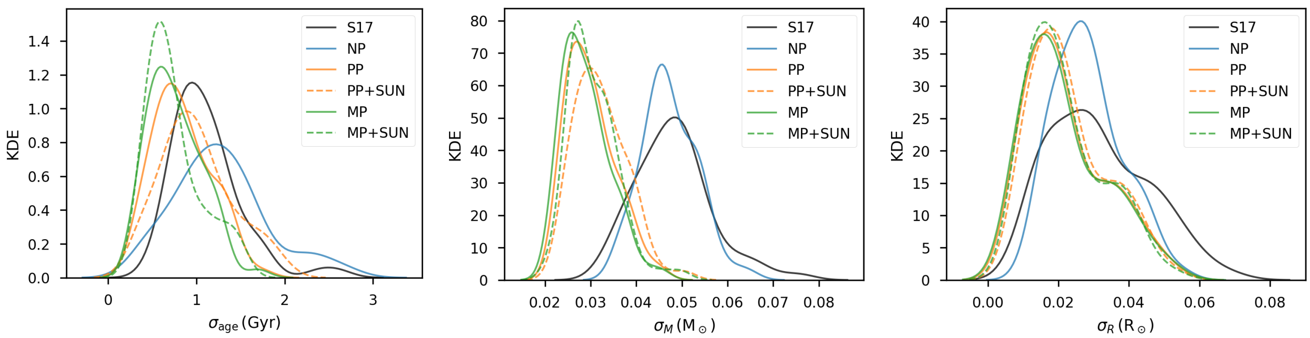
This paper has been typeset from a TEX/LATEX file prepared by the author.



(a) PP

(b) PPS

Figure 8. The results for the mixing-length theory parameter, $\alpha_{\rm mlt}$ for each star in the PP and PPS models. 100 random samples from the posterior for the population mean, μ_α and spread, $\mu_\alpha \pm \sigma_\alpha$ are shown in red and grey respectively. The Sun is shown by the solar symbol, $\odot$ in yellow for the PPS model.

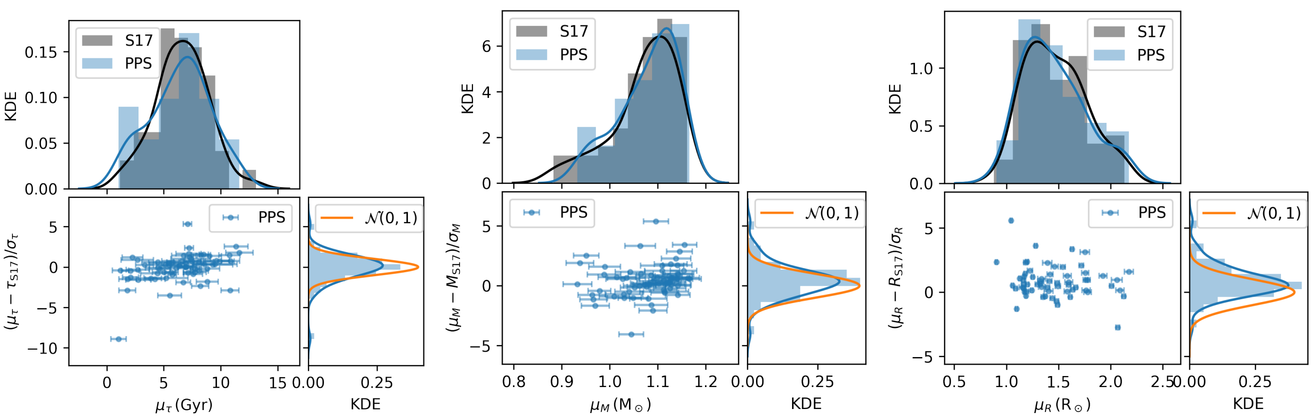


(a) Age

(b) Mass

(c) Radius

Figure 9. Kernel density estimates (KDEs) of the uncertainties in the results from each model compared with that of (S17).



(a) Age

(b) Mass

(c) Radius

Figure 10. 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.

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**				
	$f_{\rm evol}$	M ($\mathrm{M}_\odot$)	$\alpha_{\rm mlt}$	$Y_{\rm init}$	$Z_{\rm init}$	log(age/Gyr)	$T_{\rm eff}$ (K)	R ($\mathrm{R}_\odot$)	$\Delta\nu$ (μHz)	$[\mathrm{M/H}]_{\rm surf}$ (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

Table E1. The median and upper and lower 68 per cent confidence intervals for parameters output by the PP model. For the full table, see online.

Name	$f_{\rm evol}$	M ($\mathrm{M}_\odot$)	$Y_{\rm init}$	$Z_{\rm init}$	$[\mathrm{M/H}]_{\rm init}$ (dex)	age (Gyr)	$T_{\rm eff}$ (K)	R ($\mathrm{R}_\odot$)	$\Delta\nu$ (μHz)	$[\mathrm{M/H}]_{\rm surf}$ (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}	$1.18^{+0.02}_{-0.02}$	$112.5^{+2.6}_{-2.7}$	$0.07^{+0.07}_{-0.07}$
KIC10319352	$1.41^{+0.11}_{-0.27}$	$1.08^{+0.03}_{-0.03}$	$0.29^{+0.02}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.35^{+0.06}_{-0.07}$	$8.6^{+1.1}_{-1.0}$	5512^{+45}_{-46}	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.6}_{-1.6}$	$0.28^{+0.06}_{-0.07}$
KIC10322381	$0.78^{+0.23}_{-0.19}$	$1.14^{+0.03}_{-0.06}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.002}$	$-0.07^{+0.06}_{-0.07}$	$3.6^{+1.7}_{-1.1}$	6081^{+95}_{-92}	$1.41^{+0.05}_{-0.05}$	$86.2^{+4.8}_{-5.2}$	$-0.31^{+0.07}_{-0.07}$
KIC10732098	$1.50^{+0.13}_{-0.14}$	$1.14^{+0.03}_{-0.04}$	$0.28^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.15^{+0.06}_{-0.07}$	$6.4^{+0.6}_{-0.6}$	5701^{+59}_{-58}	$1.78^{+0.03}_{-0.03}$	$62.2^{+1.7}_{-1.7}$	$0.06^{+0.06}_{-0.06}$
KIC10794845	$0.39^{+0.19}_{-0.19}$	$1.09^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.001}$	$-0.09^{+0.07}_{-0.06}$	$2.0^{+1.4}_{-1.0}$	6060^{+75}_{-78}	$1.16^{+0.03}_{-0.03}$	$113.1^{+4.6}_{-4.9}$	$-0.19^{+0.07}_{-0.07}$
KIC10963065	$0.52^{+0.10}_{-0.10}$	$1.12^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.013^{+0.002}_{-0.002}$	$-0.01^{+0.06}_{-0.06}$	$2.6^{+0.6}_{-0.6}$	6060^{+40}_{-41}	$1.25^{+0.01}_{-0.01}$	$103.2^{+0.1}_{-0.1}$	$-0.15^{+0.07}_{-0.07}$
KIC10971974	$0.76^{+0.10}_{-0.12}$	$1.03^{+0.04}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.015^{+0.003}_{-0.002}$	$0.07^{+0.07}_{-0.08}$	$5.5^{+1.4}_{-1.3}$	5782^{+60}_{-61}	$1.19^{+0.02}_{-0.02}$	$106.5^{+3.1}_{-3.2}$	$-0.04^{+0.08}_{-0.08}$
KIC11021413	$1.86^{+0.02}_{-0.03}$	$1.11^{+0.03}_{-0.02}$	$0.27^{+0.01}_{-0.01}$	$0.015^{+0.001}_{-0.001}$	$0.05^{+0.04}_{-0.04}$	$7.2^{+0.4}_{-0.3}$	5334^{+55}_{-56}	$2.09^{+0.04}_{-0.04}$	$48.2^{+1.2}_{-1.2}$	$0.02^{+0.04}_{-0.04}$
KIC11027406	$0.97^{+0.07}_{-0.06}$	$1.04^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.012^{+0.002}_{-0.002}$	$-0.04^{+0.06}_{-0.06}$	$6.1^{+1.1}_{-1.0}$	5866^{+45}_{-43}	$1.36^{+0.02}_{-0.01}$	$88.3^{+1.0}_{-1.0}$	$-0.19^{+0.07}_{-0.07}$

Table E2. The median and upper and lower 68 per cent confidence intervals for parameters output by the PPS model. For the full table, see online.

Name	$f_{\rm evol}$	M ($\mathrm{M}_\odot$)	$Y_{\rm init}$	$Z_{\rm init}$	$[\mathrm{M/H}]_{\rm init}$ (dex)	age (Gyr)	$T_{\rm eff}$ (K)	R ($\mathrm{R}_\odot$)	$\Delta\nu$ (μHz)	$[\mathrm{M/H}]_{\rm surf}$ (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	$1.51^{+0.10}_{-0.22}$	$1.09^{+0.03}_{-0.03}$	$0.28^{+0.01}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.34^{+0.06}_{-0.06}$	$9.6^{+1.1}_{-1.2}$	5507^{+47}_{-48}	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.6}_{-1.6}$	$0.28^{+0.06}_{-0.06}$
KIC10322381	$0.89^{+0.21}_{-0.22}$	$1.12^{+0.05}_{-0.06}$	$0.26^{+0.01}_{-0.01}$	$0.010^{+0.002}_{-0.002}$	$-0.10^{+0.06}_{-0.07}$	$4.3^{+1.7}_{-1.2}$	6093^{+92}_{-89}	$1.41^{+0.04}_{-0.04}$	$86.1^{+5.0}_{-4.9}$	$-0.31^{+0.07}_{-0.08}$
KIC10732098	$1.60^{+0.11}_{-0.14}$	$1.14^{+0.03}_{-0.04}$	$0.27^{+0.01}_{-0.01}$	$0.017^{+0.002}_{-0.002}$	$0.13^{+0.06}_{-0.07}$	$6.9^{+0.6}_{-0.6}$	5704^{+62}_{-61}	$1.78^{+0.04}_{-0.03}$	$62.2^{+1.8}_{-1.7}$	$0.06^{+0.06}_{-0.06}$
KIC10794845	$0.46^{+0.20}_{-0.20}$	$1.09^{+0.03}_{-0.04}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.001}$	$-0.10^{+0.07}_{-0.06}$	$2.5^{+1.6}_{-1.2}$	6074^{+77}_{-80}	$1.15^{+0.03}_{-0.03}$	$113.9^{+4.8}_{-5.1}$	$-0.19^{+0.07}_{-0.07}$
KIC10963065	$0.64^{+0.09}_{-0.13}$	$1.11^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.013^{+0.002}_{-0.002}$	$-0.02^{+0.06}_{-0.06}$	$3.3^{+0.7}_{-0.8}$	6064^{+41}_{-42}	$1.24^{+0.01}_{-0.01}$	$103.2^{+0.1}_{-0.1}$	$-0.15^{+0.07}_{-0.07}$
KIC10971974	$0.83^{+0.10}_{-0.13}$	$1.03^{+0.04}_{-0.04}$	$0.26^{+0.01}_{-0.01}$	$0.015^{+0.003}_{-0.002}$	$0.06^{+0.07}_{-0.07}$	$6.5^{+1.7}_{-1.6}$	5786^{+61}_{-62}	$1.19^{+0.02}_{-0.02}$	$106.5^{+3.2}_{-3.2}$	$-0.04^{+0.08}_{-0.08}$
KIC11021413	$1.87^{+0.02}_{-0.03}$	$1.12^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.015^{+0.001}_{-0.001}$	$0.05^{+0.04}_{-0.04}$	$7.2^{+0.4}_{-0.4}$	5329^{+55}_{-55}	$2.09^{+0.04}_{-0.04}$	$48.2^{+1.2}_{-1.2}$	$0.02^{+0.04}_{-0.04}$
KIC11027406	$1.06^{+0.10}_{-0.09}$	$1.03^{+0.04}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.002}$	$-0.06^{+0.07}_{-0.07}$	$7.3^{+0.9}_{-1.1}$	5878^{+45}_{-46}	$1.35^{+0.02}_{-0.02}$	$88.3^{+1.0}_{-0.9}$	$-0.19^{+0.07}_{-0.07}$

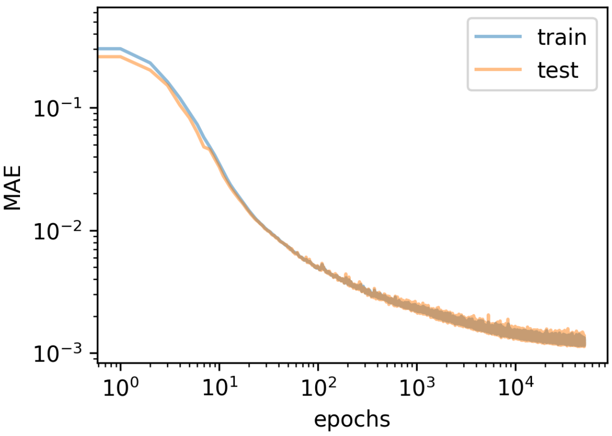
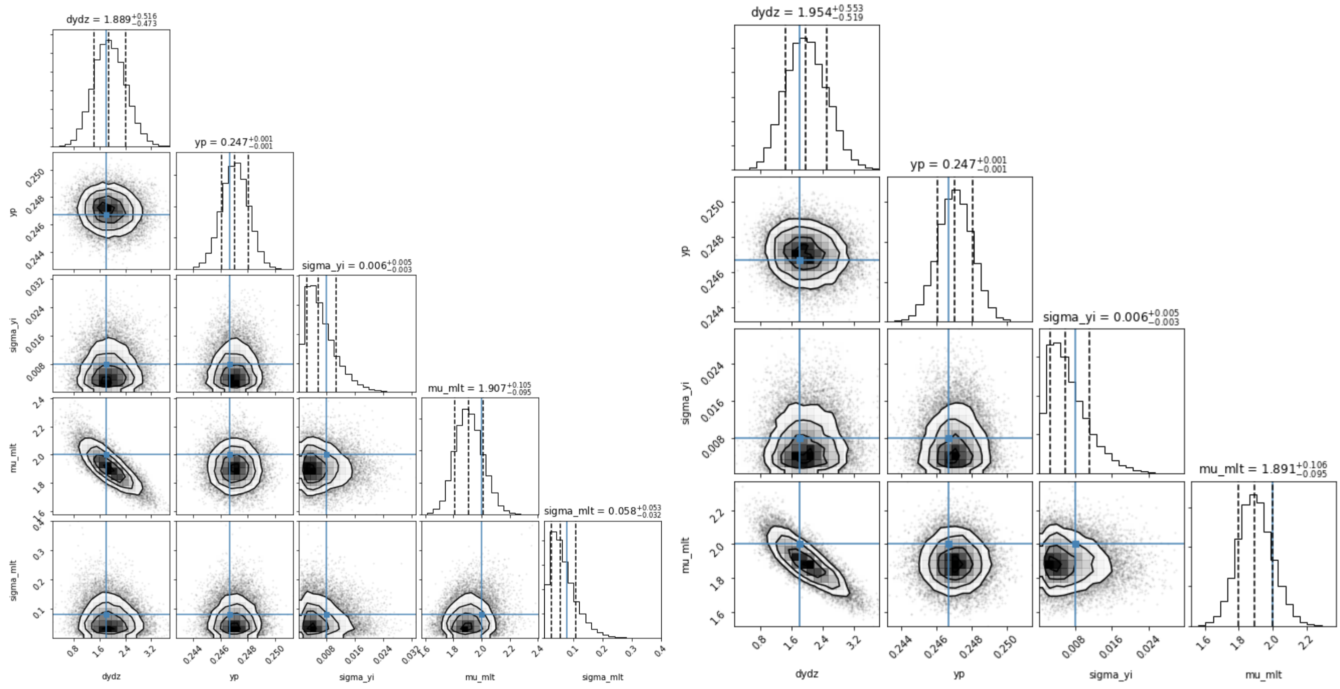


Figure B1. The MAE as a function of epochs for the train and test dataset.



(a) PP model (b) MP model

Figure C1. Corner plots showing the marginalised and joint distributions between the model hyperparameters for the test stars model. The true values are shown in blue.

Table E3. The median and upper and lower 68 per cent confidence intervals for parameters output by the MP model. For the full table, see online.

Name	f_{evol}	M ($M_\odot$)	Y_{init}	Z_{init}	[M/H]$_{init}$ (dex)	age (Gyr)	T_{eff} (K)	R ($R_\odot$)	$\Delta\nu$ (μHz)	[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}	$1.18^{+0.02}_{-0.02}$	$112.5^{+2.6}_{-2.7}$	$0.07^{+0.06}_{-0.07}$
KIC10319352	$1.41^{+0.10}_{-0.25}$	$1.08^{+0.03}_{-0.03}$	$0.29^{+0.02}_{-0.01}$	$0.028^{+0.004}_{-0.004}$	$0.36^{+0.06}_{-0.07}$	$8.6^{+1.0}_{-0.9}$	5512^{+44}_{-45}	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.7}_{-1.6}$	$0.28^{+0.06}_{-0.07}$
KIC10322381	$0.77^{+0.23}_{-0.19}$	$1.14^{+0.04}_{-0.06}$	$0.27^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.002}$	$-0.07^{+0.06}_{-0.07}$	$3.5^{+1.6}_{-1.0}$	6076^{+96}_{-91}	$1.41^{+0.05}_{-0.05}$	$86.1^{+4.7}_{-5.3}$	$-0.32^{+0.07}_{-0.07}$
KIC10732098	$1.50^{+0.13}_{-0.13}$	$1.14^{+0.03}_{-0.04}$	$0.28^{+0.01}_{-0.01}$	$0.018^{+0.002}_{-0.002}$	$0.15^{+0.06}_{-0.07}$	$6.4^{+0.6}_{-0.6}$	5702^{+56}_{-58}	$1.78^{+0.03}_{-0.03}$	$62.2^{+1.7}_{-1.7}$	$0.06^{+0.06}_{-0.06}$
KIC10794845	$0.38^{+0.18}_{-0.18}$	$1.09^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.001}$	$-0.10^{+0.07}_{-0.06}$	$1.9^{+1.3}_{-1.0}$	6059^{+74}_{-75}	$1.16^{+0.03}_{-0.03}$	$113.2^{+4.6}_{-4.9}$	$-0.19^{+0.07}_{-0.07}$
KIC10963065	$0.51^{+0.09}_{-0.09}$	$1.12^{+0.02}_{-0.02}$	$0.27^{+0.01}_{-0.01}$	$0.013^{+0.002}_{-0.002}$	$-0.01^{+0.06}_{-0.06}$	$2.5^{+0.5}_{-0.5}$	6058^{+40}_{-39}	$1.25^{+0.01}_{-0.01}$	$103.2^{+0.1}_{-0.1}$	$-0.15^{+0.07}_{-0.07}$
KIC10971974	$0.75^{+0.09}_{-0.10}$	$1.03^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.015^{+0.003}_{-0.002}$	$0.06^{+0.07}_{-0.07}$	$5.4^{+1.3}_{-1.2}$	5782^{+59}_{-58}	$1.19^{+0.02}_{-0.02}$	$106.6^{+3.1}_{-3.1}$	$-0.04^{+0.09}_{-0.08}$
KIC11021413	$1.86^{+0.02}_{-0.03}$	$1.11^{+0.03}_{-0.02}$	$0.27^{+0.01}_{-0.01}$	$0.015^{+0.001}_{-0.001}$	$0.05^{+0.04}_{-0.04}$	$7.2^{+0.3}_{-0.3}$	5336^{+55}_{-56}	$2.09^{+0.04}_{-0.04}$	$48.2^{+1.2}_{-1.2}$	$0.02^{+0.04}_{-0.04}$
KIC11027406	$0.96^{+0.05}_{-0.05}$	$1.05^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.012^{+0.002}_{-0.002}$	$-0.04^{+0.06}_{-0.06}$	$6.0^{+1.0}_{-0.8}$	5866^{+42}_{-44}	$1.36^{+0.01}_{-0.01}$	$88.3^{+1.0}_{-1.0}$	$-0.19^{+0.08}_{-0.07}$

Table E4. The median and upper and lower 68 per cent confidence intervals for parameters output by the MPS model. For the full table, see online.

Name	f_{evol}	M ($M_\odot$)	Y_{init}	Z_{init}	[M/H]$_{init}$ (dex)	age (Gyr)	T_{eff} (K)	R ($R_\odot$)	$\Delta\nu$ (μHz)	[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^{+55}_{-56}	$1.18^{+0.02}_{-0.02}$	$112.7^{+2.7}_{-2.7}$	$0.08^{+0.06}_{-0.07}$
KIC10319352	$1.61^{+0.04}_{-0.06}$	$1.08^{+0.03}_{-0.03}$	$0.27^{+0.01}_{-0.01}$	$0.028^{+0.004}_{-0.003}$	$0.33^{+0.06}_{-0.06}$	$10.8^{+0.7}_{-0.8}$	5516^{+46}_{-47}	$1.49^{+0.02}_{-0.02}$	$78.6^{+1.7}_{-1.6}$	$0.28^{+0.07}_{-0.06}$
KIC10322381	$0.98^{+0.19}_{-0.20}$	$1.10^{+0.06}_{-0.05}$	$0.26^{+0.01}_{-0.01}$	$0.010^{+0.002}_{-0.001}$	$-0.13^{+0.07}_{-0.07}$	$5.1^{+1.3}_{-1.5}$	6106^{+94}_{-80}	$1.40^{+0.04}_{-0.04}$	$85.8^{+5.6}_{-4.3}$	$-0.30^{+0.08}_{-0.08}$
KIC10732098	$1.69^{+0.06}_{-0.09}$	$1.14^{+0.03}_{-0.04}$	$0.26^{+0.01}_{-0.01}$	$0.017^{+0.002}_{-0.002}$	$0.12^{+0.06}_{-0.07}$	$7.4^{+0.5}_{-0.5}$	5715^{+61}_{-61}	$1.77^{+0.04}_{-0.03}$	$62.3^{+1.8}_{-1.8}$	$0.07^{+0.06}_{-0.07}$
KIC10794845	$0.55^{+0.17}_{-0.20}$	$1.08^{+0.03}_{-0.04}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.001}$	$-0.09^{+0.07}_{-0.06}$	$3.2^{+1.5}_{-1.4}$	6089^{+77}_{-79}	$1.15^{+0.03}_{-0.03}$	$114.3^{+5.2}_{-5.4}$	$-0.19^{+0.07}_{-0.07}$
KIC10963065	$0.73^{+0.04}_{-0.05}$	$1.10^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.012^{+0.002}_{-0.002}$	$-0.03^{+0.06}_{-0.06}$	$4.0^{+0.5}_{-0.4}$	6074^{+40}_{-40}	$1.24^{+0.01}_{-0.01}$	$103.2^{+0.1}_{-0.1}$	$-0.15^{+0.07}_{-0.07}$
KIC10971974	$0.90^{+0.08}_{-0.08}$	$1.02^{+0.04}_{-0.04}$	$0.26^{+0.01}_{-0.01}$	$0.015^{+0.003}_{-0.002}$	$0.06^{+0.08}_{-0.08}$	$7.7^{+1.7}_{-1.3}$	5794^{+60}_{-65}	$1.19^{+0.02}_{-0.02}$	$106.3^{+3.2}_{-3.6}$	$-0.04^{+0.08}_{-0.08}$
KIC11021413	$1.88^{+0.02}_{-0.02}$	$1.13^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.014^{+0.001}_{-0.001}$	$0.04^{+0.04}_{-0.04}$	$7.3^{+0.4}_{-0.4}$	5327^{+54}_{-52}	$2.09^{+0.04}_{-0.04}$	$48.2^{+1.2}_{-1.2}$	$0.02^{+0.04}_{-0.04}$
KIC11027406	$1.13^{+0.07}_{-0.07}$	$1.02^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$0.011^{+0.002}_{-0.001}$	$-0.08^{+0.06}_{-0.06}$	$7.9^{+0.8}_{-0.7}$	5895^{+40}_{-42}	$1.34^{+0.01}_{-0.01}$	$88.3^{+0.9}_{-0.9}$	$-0.19^{+0.07}_{-0.07}$

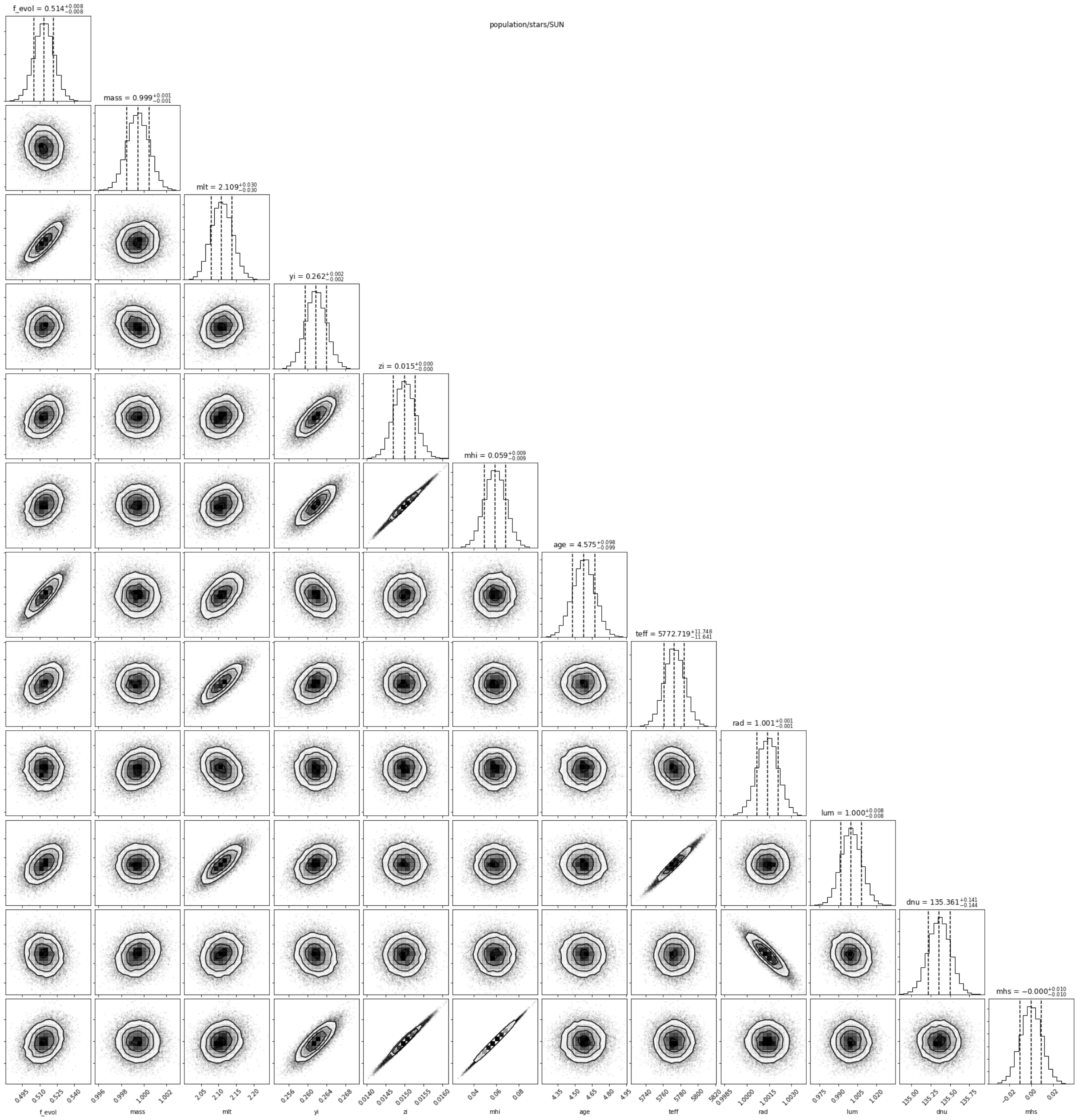


Figure D1. A corner plot showing the sampled marginal and joint posterior distributions for the Sun as a part of the PP model.