



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 18, 2020

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

ABSTRACT

Your abstract goes here

ACKNOWLEDGMENTS

I acknowledge the people who helped me.

Contents

	Page
1 Introduction	1
1.1 Hierarchical Bayesian Models	1
1.2 Modelling a Star	5
1.3 Asteroseismology of Solar-Like Oscillators	7
1.4 Sampling Stellar Models	9
1.5 Observing Stars	10
1.5.1 Detecting Asteroseismic Oscillation Modes	10
2 Hierarchically Modelling Many Stars	11
3 Future Work	12
3.1 Including the Helium II Glitch	12
3.2 Increasing the Sample Size	12
3.3 To Higher Mass Stars and Beyond	12
References	13
Appendix A Accompanying Paper	24

List of Figures

1.1	Luminosity against true ages of a fake stellar cluster. The true luminosities lie on the red line and the observed luminosities (black) have been artificially scattered by $0.05 L_{\odot}$.	4
1.2	The z -score, $(\bar{\tau} - \tau_{\text{true}})/s_{\tau}$, where $\bar{\tau}$ and s_{τ} are the respective sample mean and standard deviation of the posterior ages from each of the no- and partially-pooled models.	6
1.3	Standard deviations, s_{τ} of the age posteriors from both the no- and partially-pooled models.	7

List of Tables

LIST OF TABLES

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 $\boldsymbol{\mu}_y = \mathbf{f}(\boldsymbol{\theta})$. Using Bayes' theorem, we may write the *posterior* probability density function (PDF) of the model given a set of observations \mathbf{y} as,

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

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

Assuming our observations of \mathbf{y} are uncorrelated and subjected to random, Gaussian noise

with a known standard deviation, σ_y , we may write the likelihood function as a normal distribution,

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

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

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 ~ 14 Gyr [CITE]. Hence, we may choose to give the age parameter for a stellar model a uniform prior PDF from 0 to 14 Gyr such that our posterior PDF is not influenced by unphysical ages.

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

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

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,

$$p(\theta_1|\mathbf{y}) = \int_{-\infty}^{+\infty} p(\boldsymbol{\theta}|\mathbf{y}) d\theta_2 \dots d\theta_{N_\theta}. \quad (1.5)$$

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

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,

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

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^{N_{\text{obj}}}$ and $\mathbf{Y} = \{\mathbf{y}_k\}_{k=1}^{N_{\text{obj}}}$ are the matrices of model parameters and observations. We refer to this as a *no-pooled* model because no information is shared between the objects. However, what if we have a model which describes the distribution of a particular $\boldsymbol{\theta}_i$ 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 hierarchical 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,

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

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

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

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, \mathbf{L} from the ages, $\boldsymbol{\tau}$ of $N_{\text{obj}} = 1000$ stars in a cluster formed at roughly the same time. Modelling the population independently,

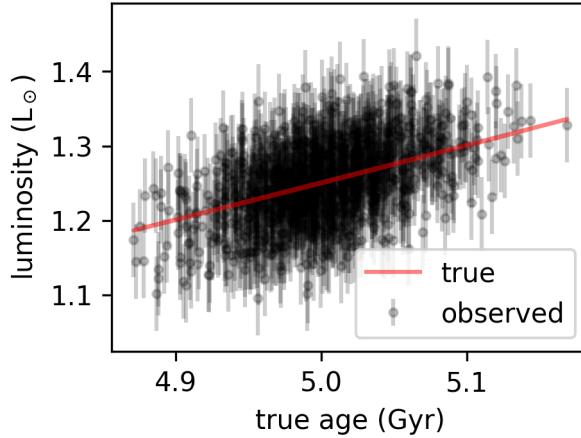


Figure 1.1: Luminosity against true ages of a fake stellar cluster. The true luminosities lie on the red line and the observed luminosities (black) have been artificially scattered by $0.05 L_\odot$.

we get the posterior,

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

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,

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

where,

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

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 plotted against the true ages in Figure 1.1 to show

If we wished to determine spread of stellar ages in the cluster using the no-pooled model, we might 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 $\mu_\tau = 5.002 \pm 0.003 \text{ Gyr}$ and $\sigma_\tau = 0.042 \pm 0.007 \text{ Gyr}$ which are within $< 2\sigma$ of the truths. Therefore, the hierarchical 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.

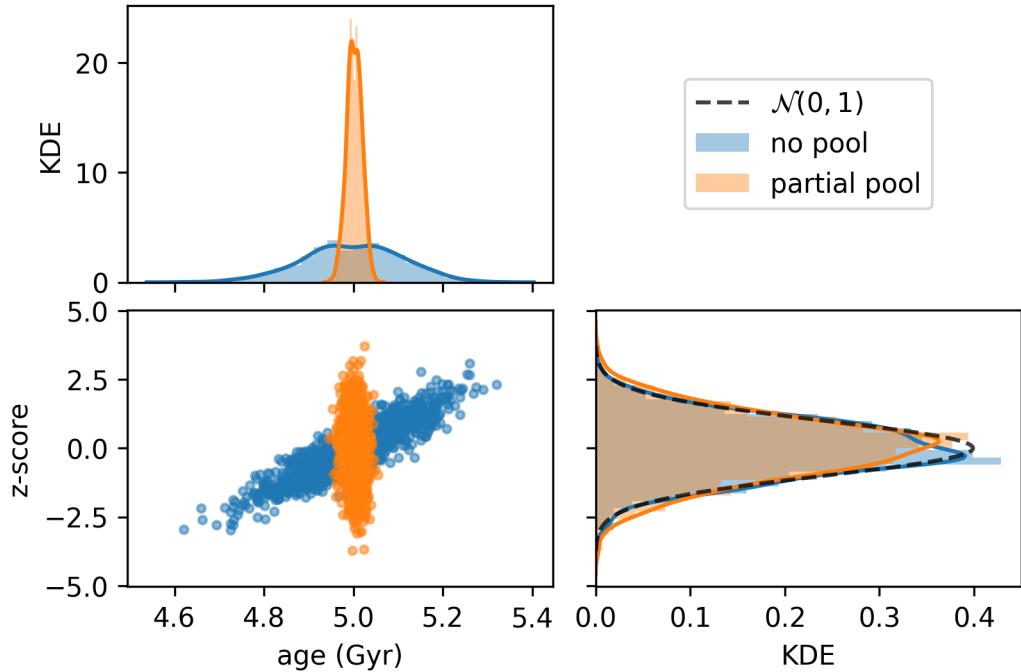


Figure 1.2: The z -score, $(\bar{\tau} - \tau_{\text{true}})/s_{\tau}$, where $\bar{\tau}$ and s_{τ} are the respective sample mean and standard deviation of the posterior ages from each of the no- and partially-pooled models.

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 aalborg 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?

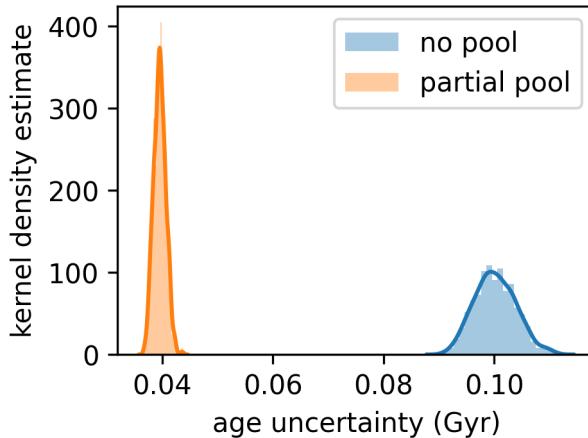


Figure 1.3: Standard deviations, s_τ of the age posteriors from both the no- and partially-pooled models.

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 typically exhibit two kinds of standing waves: acoustic oscillation modes (or p modes) excited stochastically by convection in their outer layers and restored by pressure gradients, and internal gravity waves (or g modes) which are controlled by buoyancy. This work focuses on main sequence stars for which p modes are only present in their spectra. Hence, in this section I will summarise the theory behind acoustic waves present in main sequence stars.

The theory which predicts the locations of the asteroseismic oscillation modes has its roots in the spherical harmonic oscillator. The eigenfrequencies, ν_{nlm} are categorised into modes of radial order, n , angular degree, l and 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],

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

where,

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

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],

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

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](https://doi.org/10.3847/1538-3881/aad7bf).
- Ando, H. and Y. Osaki (1975). “Nonadiabatic Nonradial Oscillations - an Application to the Five-Minute Oscillation of the Sun”. In: *Publications of the Astronomical Society of Japan* 27, pp. 581–603.
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1051/0004-6361/201322068).
- Bahcall, John N. and Roger K. Ulrich (1988). “Solar Models, Neutrino Experiments, and Helioseismology”. In: *Reviews of Modern Physics* 60, pp. 297–372. doi: [10.1103/RevModPhys.60.297](https://doi.org/10.1103/RevModPhys.60.297).

REFERENCES

- Balser, Dana S. (2006). “The Chemical Evolution of Helium”. In: *AJ* 132, pp. 2326–2332. doi: [10.1086/508515](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.3847/0004-637X/819/1/85).

REFERENCES

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1093/mnras/stv2593).

REFERENCES

- Davies, G. R., W. J. Chaplin, et al. (2015). “Astroseismic 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](https://doi.org/10.1093/mnras/stu2331).
- Davies, G. R. and A. Miglio (2016). “Astroseismology of Red Giants: From Analysing Light Curves to Estimating Ages”. In: *Astronomische Nachrichten* 337, p. 774. doi: [10.1002/asna.201612371](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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/aaeeeec](#).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.3847/1538-4357/aabfdb).
- Howe, Rachel et al. (2020). “Solar Cycle Variation of Nmax in Helioseismic Data and Its Implications for Asteroseismology”. In: *MNRAS* 493, pp. L49–L53. doi: [10.1093/mnrasl/slaa006](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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.
- Lindgren, L. et al. (2018). “Gaia Data Release 2. The Astrometric Solution”. In: *A&A* 616, A2. doi: [10.1051/0004-6361/201832727](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1051/0004-6361/201527075).

REFERENCES

- 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](https://doi.org/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](https://doi.org/10.3390/e21070627). arXiv: [1906.06903](https://arxiv.org/abs/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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.

REFERENCES

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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 \[astro-ph\]](https://arxiv.org/abs/2001.01839).
- Scargle, J. D. (1982). “Studies in Astronomical Time Series Analysis. II - Statistical Aspects of Spectral Analysis of Unevenly Spaced Data”. In: *The Astrophysical Journal* 263, pp. 835–853. doi: [10.1086/160554](https://doi.org/10.1086/160554).
- Serenelli, Aldo 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1093/mnras/stx765).

REFERENCES

- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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 \[cond-mat\]](https://arxiv.org/abs/2001.00173).
- 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](https://doi.org/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](https://doi.org/10.1093/mnras/stt1533).
- Trampedach, Regner et al. (2014). “Improvements to Stellar Structure Models, Based on a Grid of 3D Convection Simulations - II. Calibrating the Mixing-Length Formulation”. In: *MNRAS* 445, pp. 4366–4384. doi: [10.1093/mnras/stu2084](https://doi.org/10.1093/mnras/stu2084).
- Ulrich, R. K. (1986). “Determination of Stellar Ages from Asteroseismology”. In: *The Astrophysical Journal Letters* 306, pp. L37–L40. doi: [10.1086/184700](https://doi.org/10.1086/184700).

REFERENCES

- Ulrich, Roger K. (1970). “The Five-Minute Oscillations on the Solar Surface”. In: *The Astrophysical Journal* 162, p. 993. doi: [10.1086/150731](https://doi.org/10.1086/150731).
- VanderPlas, Jacob T. (2018). “Understanding the Lomb-Scargle Periodogram”. In: *The Astrophysical Journal Supplement Series* 236, p. 16. doi: [10.3847/1538-4365/aab766](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.3847/1538-4357/ab44a9).

Appendix A

Accompanying Paper

Hierarchically modelling asteroseismic dwarfs and subgiants to improve inference of stellar properties

Alexander J. Lyttle,^{1,2}★ Tanda Li,^{1,2} Guy R. Davies,^{1,2} Lindsey M. Carboneau^{1,2} and TBC

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

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

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

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. 2019b) ‘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). This sample provides an extensive set of dwarf and subgiant stars with asteroseismic detections observed by the *Kepler* mission. S17 used grid-based modelling to determine the ages, τ , masses, M , radii, R and sur-

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

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

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

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

To calculate luminosities for the sample, we used *Gaia* Data Release 2 (DR2) parallaxes (Gaia Collaboration et al. 2016, 2018). We cross-matched the remaining sample with the DR2 catalogue, taking the nearest neighbours within a $4''$ radius. Although DR2 parallaxes have improved upon the DR1 values at the time of S17, there was still evidence for a zero-point offset (Lindegren et al. 2018). We adopted a global offset of 0.05 mas, in the sense that DR2 parallaxes were underestimated, representative of values obtained in the literature for the *Kepler* field (Zinn et al. 2019a; Hall et al. 2019).

We then cross-matched our sample with the Two-Micron All Sky Survey (2MASS) to obtain K_S -band ($2.16 \mu\text{m}$) photometry (Skrutskie et al. 2006).

We determined luminosities, L for the sample using the direct method of ISOCLASSIFY with K_S -band photometry, *Gaia* DR2 parallaxes, ASPCAP $[\text{M}/\text{H}]$ and T_{eff} and asteroseismic $\log g$ as inputs (Huber et al. 2017; Berger et al. 2020). This involved computing absolute K_S -band magnitudes using the *Gaia* DR2 parallaxes and extinctions determined by the 3D galactic reddening maps of Green et al. (2018). We determined absolute bolometric magnitudes by interpolating the MIST bolometric correction tables with T_{eff} , $\log g$ and $[\text{M}/\text{H}]$ (Dotter 2016; Choi et al. 2016). An uncertainty of 0.02 mag was assumed in ISOCLASSIFY for both the extinctions and bolometric corrections (Huber et al. 2017). We obtained luminosities for the sample with a median uncertainty of 3.4 per cent.

The final sample comprised 81 stars for which we had data for T_{eff} , $[\text{M}/\text{H}]$, $\Delta\nu$ and L to use as inputs for our stellar modelling method – see Table A1. Figure 1 shows the L and T_{eff} for the sample plot on a Hertzsprung–Russell diagram in context with stellar evolutionary tracks at solar metallicity.

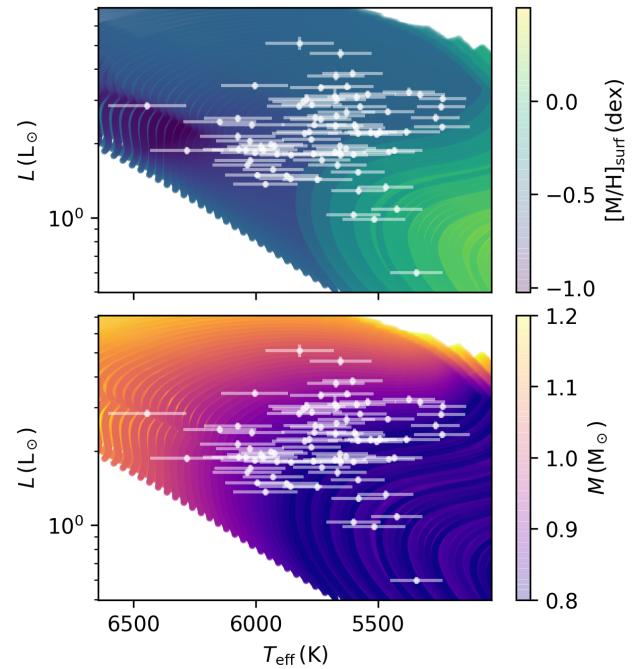


Figure 1. The luminosity, L against effective temperature, T_{eff} of the sample of 81 *Kepler* dwarfs and subgiants plot against a subset of the grid of stellar models computed in Section 3.1. The top plot is coloured by stellar surface metallicity and the bottom plot is coloured by stellar mass.

3 METHODS

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

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

In Section 3.2, we describe a method to replace the grid of stellar models with a smooth function approximation using machine learning. In particular, we trained an artificial neural network (ANN) on the grid of stellar models to map stellar fundamentals to observables. Fast evaluation of the ANN gradient is required during training. Consequently, estimating the gradient of the model likelihood is possible with an ANN. With an ANN, we open up the possibility of using the Hamiltonian Monte Carlo (HMC) algorithm which requires the gradient to sample the model posterior – for example, using the No-U-Turn Sampler (NUTS; Homan & Gelman 2014).

Finally, we constructed three Bayesian models in Section 3.3

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

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

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

3.1 Grid of stellar models

We built a stellar model grid to use in training the ANN. The grid includes four independent model inputs: stellar mass (M), initial helium fraction (Y_{init}), initial metallicity ([M/H]_{init}), and the mixing-length parameter (α_{mlt}). Ranges and grid steps of the four model inputs are summarised in Table 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 evolutionary tracks with input values at the midpoint between points on the grid for validating the ANN.

3.1.1 Stellar models and input physics

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

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

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

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

3.1.2 Oscillation models and seismic $\Delta\nu$

Theoretical stellar oscillations were calculated with the GYRE code (version 5.1), which was developed by Townsend & Teitler (2013). We computed radial modes (for $\ell = 0$) by solving the adiabatic stellar pulsation equations with the structural models generated by

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

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

3.2 Artificial neural network

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

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

An ANN is a network of artificial *neurons* which each transform some input vector, \mathbf{x} based on trainable weights, \mathbf{w} and a bias, b [CITATIONS]. The weights are represented by the connections between neurons and the bias is a unique scalar associated with each neuron. Deep learning (DL) is the name given to the case where neurons are arranged into a series of layers such that any neuron in layer $k - 1$ is connected to at least one of the neurons in layer k .

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

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

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

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

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

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

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

To fit the ANN, we used a set of training data, $\mathbf{D}_{\text{train}} =$

$\{(X_1, Y_1) \dots (X_{N_{\text{train}}}, Y_{N_{\text{train}}})\}$ comprising N_{train} input-output pairs. We split the training data into pseudo-random batches, D_{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, $E(D_{\text{batch}})$, also known as the *loss* which quantifies the difference between the training data and predictions. We also considered an addition to the loss called *regularisation* which helps reduce over-fitting (CITE). During fitting, the weights are updated after each batch using an algorithm called the *optimizer*, back-propagating the error with the goal of minimising the loss.

We 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, 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

We built the train and test dataset from the outputs of the grid of stellar models in Section 3.1. This included the input parameters: M , α_{mlt} , Y_{init} and the initial heavy-elements fraction, Z_{init} . We also included the T_{eff} , $\log g$, $\Delta\nu$, stellar age (τ), radius (R), surface metallicity ($[\text{M}/\text{H}]_{\text{surf}}$) and other chemical composition information generated by the models. We determined the fractional main sequence (MS) lifetime, $f_{\text{MS}} = \tau/\tau_{\text{MS}}$, of each evolutionary track by taking τ_{MS} as the age when the central helium fraction, $X_c < 0.001$. We then cut data where $f_{\text{MS}} < 0.01$ to remove points on the grid prior to the MS.

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

We produced a validation dataset of $\sim 2 \times 10^6$ stellar models evolved using MESA. Values for the initial mass, metallicity, helium and mixing-length-theory parameter were chosen at the midpoint of the grid parameters described in Table 1. We prepared this dataset in the same way as the training set, but also constrained it to $\tau < 15$ Gyr because ages above ~ 15 yr are unphysical and such points are sparse in the training data. This dataset was set aside and evaluated on the final ANN.

3.2.2 Tuning

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

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

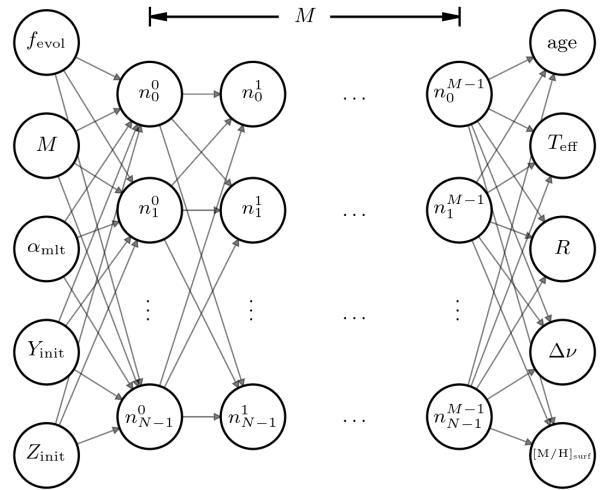


Figure 2. An artificial neural network comprising M hidden layers with N neurons per layer. Arrows connecting the nodes represent tunable weights.

f_{evol} .

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

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

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

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

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

After preliminary tuning, we chose the ANN input and output parameters to be $\mathbf{X} = \{f_{\text{evol}}, M, \alpha_{\text{mlt}}, Y_{\text{init}}, Z_{\text{init}}\}$ and $\mathbf{Y} = \{\log(\tau), T_{\text{eff}}, R, \Delta\nu, [\text{M}/\text{H}]_{\text{surf}}\}$ respectively. A generalised form of our neural network is depicted in Figure 2. The inputs corresponded to initial conditions in the stellar modelling code and the outputs corresponded to surface conditions throughout the lifetime of the star, with the exception of age which is mapped from f_{evol} .

We standardised the training dataset by subtracting the median, $\mu_{1/2}$ and dividing by the standard deviation, σ . We found that the ANN performed better when the training data was scaled in this way. In Table B1, we show the locations and scales of the standardisation for our chosen input and output parameters.

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

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

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

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

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

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

3.2.3 Validation

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

We found good agreement between the validation dataset and ANN predictions, within typical observational uncertainties. We found that the largest errors lay at the boundaries of the training data and in areas sparsely populated by the grid. This is apparent in Figure 3 where we plot the validation error against each parameter – for example, the spread in error increasing at high temperatures.

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

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

Otherwise, the accuracy is very good within the observed range covered by our sample of 81 dwarfs and subgiants. Hence, we chose the median absolute deviation (MAD) as an estimator of the spread in error, because it is less sensitive to outliers than the standard deviation.

To represent the accuracy of the ANN, we present the median, $\mu_{1/2}$ and MAD estimator, $\sigma_{\text{MAD}} = 1.4826 \cdot \text{median}(|x_{\text{true}} - x_{\text{pred}}|)$ of the error in Table 2. The median is close to zero for all parameters, showing little systematic bias in the ANN. The MAD is also lower than observational uncertainties quoted in Section 2. Although the error in $\Delta\nu$ is $\sim 0.1 \mu\text{Hz}$ is comparable to observations with the best signal-to-noise, this error is random throughout the validation data and should not produce any systematic bias.

3.3 Statistical models

We devised three Bayesian models, each with varying levels of parameter sharing (pooling) between stars in the population. Initially, we tested the models and demonstrated 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.

3.3.1 No-pooled model

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

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

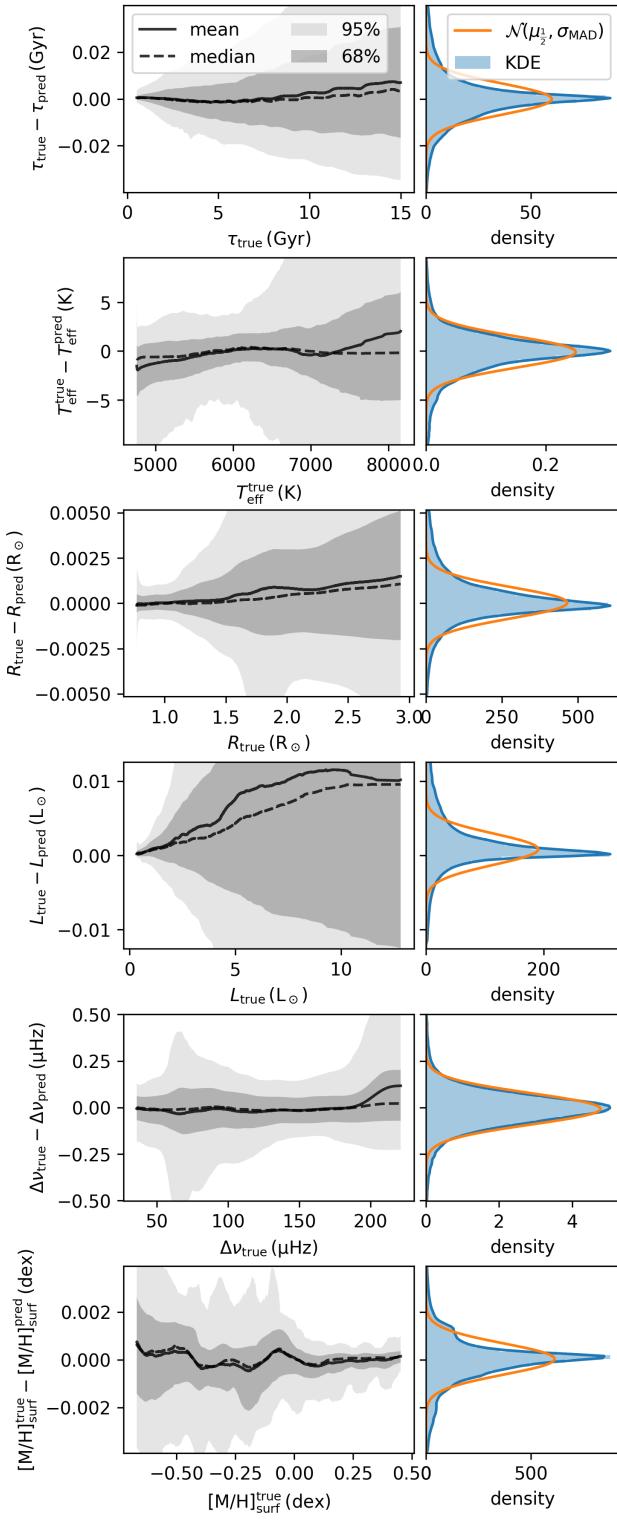


Figure 3. Left: the rolling error between the validation dataset (*true*) and the ANN predictions (*pred*) plotted against each parameter. Right: a kernel density estimate (KDE) of the validation error and a normal distribution centred on the median, $\mu_{1/2}$ with an estimator for the standard deviation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

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

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

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

where the beta distribution is defined as,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.3.2 Partial-pooled model

Sharing, or pooling parameters between stars in a population can improve the uncertainties on stellar fundamentals by encoding our prior

knowledge of their distribution in a population. We constructed a hierarchical model [CITE Gelman?], which builds upon the NP model by introducing population-level *hyperparameters*. Specifically, we chose to describe initial helium and α_{mlt} by partially-pooling them.

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

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

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

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

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

where μ_Y and σ_Y is the mean initial helium fraction as described by the linear helium enrichment law [CITE WORKS WHICH USE THIS LAW],

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

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

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

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

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

We gave all of the hyperparameters weakly informative priors, with the exception of Y_P for which we adopt a recent measurement of the primordial helium abundance the mean [CITE PLANK] with a standard deviation representative of the range of values in the literature [CITE]. We assumed priors on the hyperparameters as follows,

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

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

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

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

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

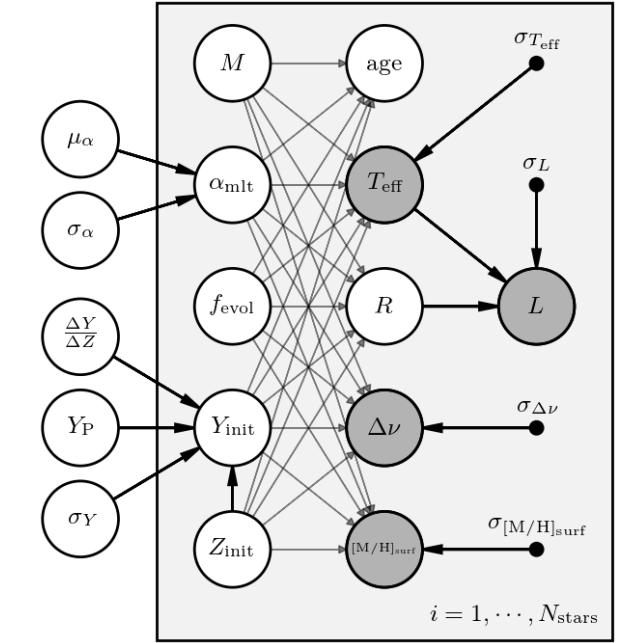


Figure 4. A probabilistic graphical model (PGM) of the partially-pooled (PP) hierarchical model. Nodes outside of the grey rectangle represent the hyperparameters in the model. Nodes inside the grey rectangle represent individual stellar parameters. Dark grey nodes represent observables which each have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

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

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

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

3.3.3 Max-pooled model

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

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

where,

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

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

Table 3. Solar input data.

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

3.4 Solar calibrator

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

For both the PP and MP models, we ran versions with and without data for the Sun included in the APOKASC population, referred to as PPS and MPS respectively. We adopted the solar data shown in Table 3 as observables with uncertainties conservatively limited to the accuracy of the neural network. We adopt the central Solar $\Delta\nu$ from Huber et al. (2011), determined as if the Sun was observed by the *Kepler* telescope.

4 RESULTS

TODO: Segment this into going through each table and figure with explanation, leave some for the discussion.

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 true values. We then present the results for the *Kepler* dwarfs and compare the results to S17.

4.1 The synthetic population

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

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

4.1.1 Stellar parameters

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

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

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

4.1.2 Shrinkage

We reran the PP model with 10 and 50 stars chosen randomly from the sample of synthetic stars. In Figure 5, we show the uncertainties in the several parameters from the results of each of the models. For the two pooled parameters, Y_{init} and α_{mlt} , the shrinkage due to pooling is most obvious. We see the PP model repeatedly improves on the uncertainties from the NP model when N_{stars} is increased.

In Figure 5 we also see a similar shrinkage in uncertainty for τ , M and R , with all models improve upon the NP model. However, we do not see the same effect in Z_{init} for which the uncertainty appears dominated by observations of $[\text{M}/\text{H}]_{\text{surf}}$.

4.1.3 Population parameters

In Figure C1, we show the joint posterior distributions for the hyperparameters of the PP model fit to the results of the NP model. We see that this method appears to recover the true values well. However, fitting the model this way does not benefit from the same shrinkage in uncertainty as shown in the pooled models. Furthermore, the uncertainties on the individual stellar parameters were found to be unreliable due to boundary effects from the prior. This likely means that the uncertainties on the hyperparameter results are underestimated.

In Figure C2, we see that the PP model also recovers the hyperparameter truths well, with some noise due to random realisation error. Fitting the model this way has the added benefit over the NP model of improving the inference of the individual stellar parameters, as shown in the previous two sections. We also found that when we ran the PP model with 10 and 50 stars, the uncertainties on the hyperparameters also shrank with increasing N_{stars} .

Figure C3 shows the hyperparameter results for the MP model. Here, α_{mlt} was assumed to be the same for all stars. This model also recovers the true hyperparameters for helium well, and the assumed value for α_{mlt} is within uncertainty of the true μ_α .

4.2 The APOKASC sample

In this section, we present the results for each of the NP, PP and MP models ran with the sample of 81 APOKASC *Kepler* dwarfs and subgiants as inputs. We also present the results for the PPS and MPS models ran with Solar data as a calibrator. We show the shrinkage in uncertainties in age, mass and radius with the addition of pooling in Section 4.2.1. We then show the results for model hyperparameters in Section 4.2.2.

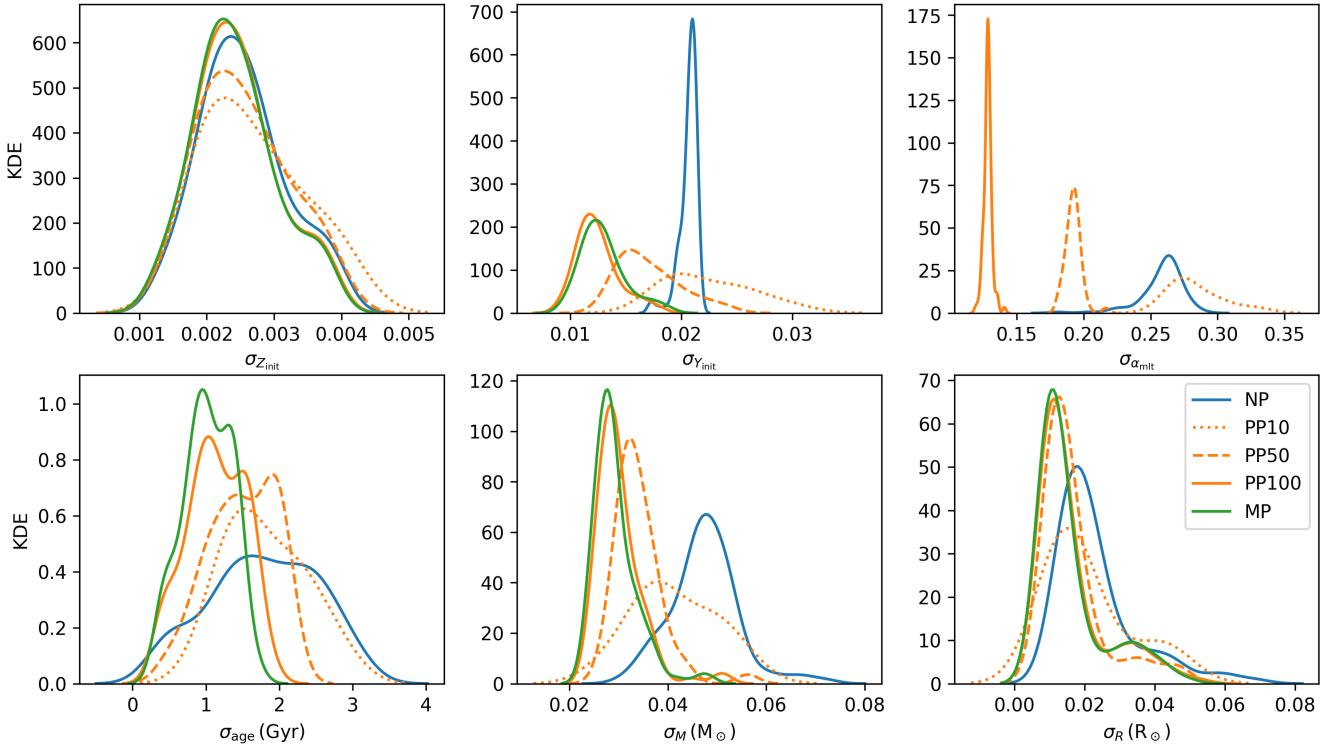


Figure 5. Kernel density estimates (KDEs) showing the shrinkage of statistical uncertainties between models of the sample of synthetic stars. The PP model was run with 10, 50 and 100 stars and is denoted PP10, PP50, and PP100 respectively. The NP and MP models were both run with the full set of 100 stars.

4.2.1 Stellar parameter results

In Table E1, we preset results for 65 of the 81 stars using the NP model. After an initial run, 16 stars were dropped. The posteriors of 6 of the dropped stars were skewed towards the upper mass limit of $1.2 M_{\odot}$, and the remaining 10 stars suffered poor convergence during sampling. Running the NP model with synthetic stars resulted in unreliable uncertainties. We saw the same boundary effects in the posteriors for Y_{init} and α_{mlt} here.

Out of the 65 stars sampled by the NP model, we obtained results for 63 with the PP model, because 2 stars were dropped due to convergence problems. In Tables E2 and E3 we present the results for the PP and PPS model respectively. We did not see the same convergence problems with the MP and MPS models, hence we present results for the 65 stars in Tables E4 and E5 respectively.

In Figure 6, we saw a similar improvement in uncertainty between the NP and pooled models as with the synthetic stars in Figure 5. These plots show the improvement in uncertainties for mass, radius and age over results for the same stars in S17 when pooling stellar parameters.

4.2.2 Population parameter results

We obtained values for the hyperparameters for each of the models and present them in Table 4 along with their upper and lower 68 per cent credible 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 would make the NP results unreliable.

5 DISCUSSION

5.1 Helium enrichment

Discuss the difference in helium with the NP and pooled models.

Discuss the difference in helium enrichment results with and without the Sun

Discuss why we made the choice of a linear enrichment law and other kinds of laws which could be explored as priors.

5.2 Mixing-length theory

Discuss difference in mlt in the MP and PP models and why this is an important distinction

Discuss the difference in mlt with and without the Sun

Discuss alternatives to normally distributed mixing length with reference to other possible laws.

5.3 Comparison with APOKASC results

We then compared the results from the PPS model with that of S17 in Figure 9. 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.

5.3.1 Mass

Discuss the difference in masses

Discuss the difference in a mass uncertainties

Table 4. 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	α_{mlt}
NP	$1.69^{+0.21}_{-0.21}$	$0.0074^{+0.0026}_{-0.0022}$	$1.954^{+0.040}_{-0.041}$	$0.065^{+0.030}_{-0.024}$	—
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	$0.76^{+0.24}_{-0.27}$	$0.0049^{+0.0039}_{-0.0025}$	—	—	$2.088^{+0.031}_{-0.029}$

5.3.2 Radius

Discuss the difference in radii
Discuss the difference in a radius uncertainties

5.3.3 Age

Discuss the difference in ages
Discuss the difference in a age uncertainties

5.4 Systematic uncertainties

What sorts of systematic uncertainties are there?
Model systematics
Discuss ways we may determine model systematics in the future.
Observable systematics (e.g. effective temperature)

5.5 The potential of pooling

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

6 CONCLUSIONS

ACKNOWLEDGEMENTS

REFERENCES

- Albareti F. D., et al., 2017, *The Astrophysical Journal Supplement Series*, 233, 25
 Appourchaux T., et al., 2015, *Astronomy and Astrophysics*, 582, A25
 Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481
 Ball W. H., Gizon L., 2014, *Astronomy and Astrophysics*, 568, A123
 Berger T. A., Huber D., van Saders J. L., Gaidos E., Tayar J., Kraus A. L., 2020, arXiv e-prints, 2001, arXiv:2001.07737
 Blanton M. R., et al., 2017, *The Astronomical Journal*, 154, 28
 Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *The Astrophysical Journal*, 823, 102
 Clevert D.-A., Unterthiner T., Hochreiter S., 2015, arXiv e-prints, 1511, arXiv:1511.07289
 Dotter A., 2016, *The Astrophysical Journal Supplement Series*, 222, 8
 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
 Gaia Collaboration et al., 2016, *A&A*, 595, A1
 Gaia Collaboration et al., 2018, *Astronomy and Astrophysics*, 616, A1
 García Pérez A. E., et al., 2016, *The Astronomical Journal*, 151, 144
 Glorot X., Bordes A., Bengio Y., 2011, Proc. 14th Int. Conf. Artif. Intell. Statistics AISTATS 2011, 15, 315
 Green G. M., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, 478, 651
 Hahnloser R. H. R., Sarpeshkar R., Mahowald M. A., Douglas R. J., Seung H. S., 2000, *Nature*, 405, 947
 Hall O. J., et al., 2019, *MNRAS*, 486, 3569

Homan M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, 15, 1593

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

Huber D., et al., 2017, *The Astrophysical Journal*, 844, 102

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

Lindgren L., et al., 2018, *A&A*, 616, A2

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

Pinsonneault M. H., An D., Molenda-Żakowicz J., Chaplin W. J., Metcalfe T. S., Bruntt H., 2012, *ApJS*, 199, 30

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

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

Skrutskie M. F., et al., 2006, *The Astronomical Journal*, 131, 1163

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., 2019a, *ApJ*, 878, 136

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

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.

APPENDIX E: MODEL RESULTS

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

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_{eff} (K)	$\sigma_{T_{\text{eff}}}$ (K)	L (L_{\odot})	σ_L (L_{\odot})	$\Delta\nu$ (μHz)	$\sigma_{\Delta\nu}$ (μHz)	[M/H]surf (dex)	$\sigma_{[\text{M}/\text{H}]}$ (dex)	$\log g$ (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

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

	Input					Output				
	f_{vol}	M (M_{\odot})	α_{mlt}	Y_{init}	Z_{init}	$\log(\tau/\text{Gyr})$	T_{eff} (K)	R (R_{\odot})	$\Delta\nu$ (μHz)	[M/H]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 NP model. For the full table, see online.

Name	f_{vol}	M (M_{\odot})	Y_{init}	Z_{init}	[M/H]init (dex)	τ (Gyr)	T_{eff} (K)	R (R_{\odot})	$\Delta\nu$ (μHz)	[M/H]surf (dex)
KIC5950854	0.92 ^{+0.13} _{-0.15}	1.02 ^{+0.05} _{-0.05}	0.28 ^{+0.02} _{-0.02}	0.012 ^{+0.002} _{-0.002}	-0.02 ^{+0.06} _{-0.07}	5.8 ^{+1.7} _{-1.6}	5955 ⁺⁵⁵ ₋₅₅	1.26 ^{+0.02} _{-0.02}	96.9 ^{+0.2} _{-0.2}	-0.15 ^{+0.07} _{-0.07}
KIC3942719	1.84 ^{+0.03} _{-0.05}	1.08 ^{+0.05} _{-0.05}	0.27 ^{+0.03} _{-0.02}	0.007 ^{+0.001} _{-0.001}	-0.24 ^{+0.06} _{-0.06}	5.6 ^{+0.4} _{-0.4}	5761 ⁺⁷³ ₋₇₁	2.15 ^{+0.05} _{-0.05}	45.2 ^{+1.1} _{-1.1}	-0.29 ^{+0.06} _{-0.06}
KIC4457351	1.79 ^{+0.05} _{-0.06}	0.94 ^{+0.03} _{-0.04}	0.26 ^{+0.02} _{-0.01}	0.008 ^{+0.001} _{-0.001}	-0.20 ^{+0.05} _{-0.05}	10.7 ^{+0.8} _{-0.8}	5476 ⁺⁶⁷ ₋₆₉	1.75 ^{+0.05} _{-0.04}	57.7 ^{+2.2} _{-2.2}	-0.26 ^{+0.05} _{-0.05}
KIC3223000	0.43 ^{+0.18} _{-0.19}	1.12 ^{+0.04} _{-0.04}	0.27 ^{+0.03} _{-0.02}	0.012 ^{+0.002} _{-0.002}	-0.03 ^{+0.06} _{-0.06}	2.0 ^{+1.1} _{-1.0}	6187 ⁺⁶⁷ ₋₆₇	1.19 ^{+0.02} _{-0.02}	110.3 ^{+1.6} _{-1.6}	-0.13 ^{+0.07} _{-0.07}
KIC7429287	1.52 ^{+0.12} _{-0.12}	0.96 ^{+0.05} _{-0.05}	0.27 ^{+0.03} _{-0.02}	0.009 ^{+0.001} _{-0.001}	-0.17 ^{+0.06} _{-0.06}	8.5 ^{+1.3} _{-1.2}	5808 ⁺³⁸ ₋₃₉	1.53 ^{+0.03} _{-0.03}	71.3 ^{+1.2} _{-1.2}	-0.28 ^{+0.07} _{-0.07}
KIC4446300	1.88 ^{+0.02} _{-0.03}	1.05 ^{+0.04} _{-0.04}	0.27 ^{+0.03} _{-0.02}	0.014 ^{+0.001} _{-0.001}	0.03 ^{+0.04} _{-0.04}	8.5 ^{+0.5} _{-0.4}	5294 ⁺⁵⁰ ₋₄₈	1.99 ^{+0.03} _{-0.03}	50.0 ^{+0.5} _{-0.4}	0.00 ^{+0.04} _{-0.04}
KIC9025370	0.21 ^{+0.16} _{-0.13}	1.10 ^{+0.03} _{-0.04}	0.26 ^{+0.02} _{-0.01}	0.015 ^{+0.002} _{-0.002}	0.06 ^{+0.06} _{-0.06}	1.3 ^{+1.2} _{-0.8}	5940 ⁺⁵⁵ ₋₅₅	1.05 ^{+0.01} _{-0.01}	132.6 ^{+0.4} _{-0.4}	0.03 ^{+0.06} _{-0.07}
KIC8938364	1.29 ^{+0.16} _{-0.18}	0.98 ^{+0.05} _{-0.05}	0.28 ^{+0.03} _{-0.02}	0.014 ^{+0.002} _{-0.002}	0.03 ^{+0.06} _{-0.06}	8.9 ^{+1.6} _{-1.5}	5756 ⁺⁵³ ₋₅₄	1.36 ^{+0.02} _{-0.02}	85.7 ^{+0.1} _{-0.1}	-0.08 ^{+0.07} _{-0.07}
KIC10963065	0.69 ^{+0.14} _{-0.21}	1.09 ^{+0.05} _{-0.05}	0.27 ^{+0.03} _{-0.02}	0.012 ^{+0.002} _{-0.002}	-0.01 ^{+0.06} _{-0.06}	3.5 ^{+1.2} _{-1.2}	6091 ⁺⁵³ ₋₅₄	1.23 ^{+0.02} _{-0.02}	103.2 ^{+0.1} _{-0.1}	-0.15 ^{+0.07} _{-0.07}
KIC10971974	0.87 ^{+0.17} _{-0.17}	1.00 ^{+0.06} _{-0.06}	0.28 ^{+0.03} _{-0.02}	0.015 ^{+0.003} _{-0.002}	0.07 ^{+0.07} _{-0.07}	6.8 ^{+2.7} _{-2.1}	5805 ⁺⁶⁹ ₋₆₇	1.18 ^{+0.03} _{-0.03}	106.0 ^{+3.4} _{-3.2}	-0.03 ^{+0.08} _{-0.08}

Table E2. 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_{vol}	M (M_{\odot})	Y_{init}	Z_{init}	[M/H]init (dex)	τ (Gyr)	T_{eff} (K)	R (R_{\odot})	$\Delta\nu$ (μHz)	[M/H]surf (dex)
KIC10079226	0.22 ^{+0.10} _{-0.09}	1.16 ^{+0.02} _{-0.03}	0.28 ^{+0.01} _{-0.01}	0.020 ^{+0.003} _{-0.002}	0.19 ^{+0.06} _{-0.06}	1.2 ^{+0.6} _{-0.5}	5962 ⁺⁴⁴ ₋₄₂	1.17 ^{+0.01} _{-0.01}	115.9 ^{+0.7} _{-0.7}	0.15 ^{+0.07} _{-0.07}
KIC10215584	0.37 ^{+0.15} _{-0.13}	1.14 ^{+0.03} _{-0.03}	0.27 ^{+0.01} _{-0.01}	0.018 ^{+0.002} _{-0.002}	0.14 ^{+0.06} _{-0.06}	2.1 ^{+1.0} _{-0.8}	5941 ⁺⁵⁷ ₋₅₆	1.18 ^{+0.02} _{-0.02}	112.5 ^{+2.6} _{-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 ⁺⁷⁸ ₋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 E3. 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_{vol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (Gyr)	T_{eff} (K)	$R (R_\odot)$	$\Delta\nu$ (μHz)	[M/H] _{surf} (dex)
KIC10079226	0.35 ^{+0.11} _{-0.12}	1.17 ^{+0.02} _{-0.03}	0.27 ^{+0.01} _{-0.01}	0.020 ^{+0.003} _{-0.002}	0.20 ^{+0.06} _{-0.06}	2.1 ^{+0.8} _{-1.2}	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} _{-1.7}	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 ^{+0.04} _{-0.04}	103.2 ^{+0.1} _{-0.1}	-0.15 ^{+0.07} _{-0.07}
KIC10971974	0.83 ^{+0.10} _{-0.13}	1.03 ^{+0.04} _{-0.03}	0.26 ^{+0.01} _{-0.01}	0.015 ^{+0.003} _{-0.002}	0.06 ^{+0.07} _{-0.07}	6.5 ^{+1.7} _{-1.0}	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 ^{+0.9} _{-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}

Table E4. 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_{vol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (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 ⁺⁸⁹ ₋₈₈	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} _{-0.03}	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 ⁺⁵⁶ ₋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 E5. 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_{vol}	$M (\text{M}_\odot)$	Y_{init}	Z_{init}	[M/H] _{init} (dex)	τ (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} _{-3.2}	-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}

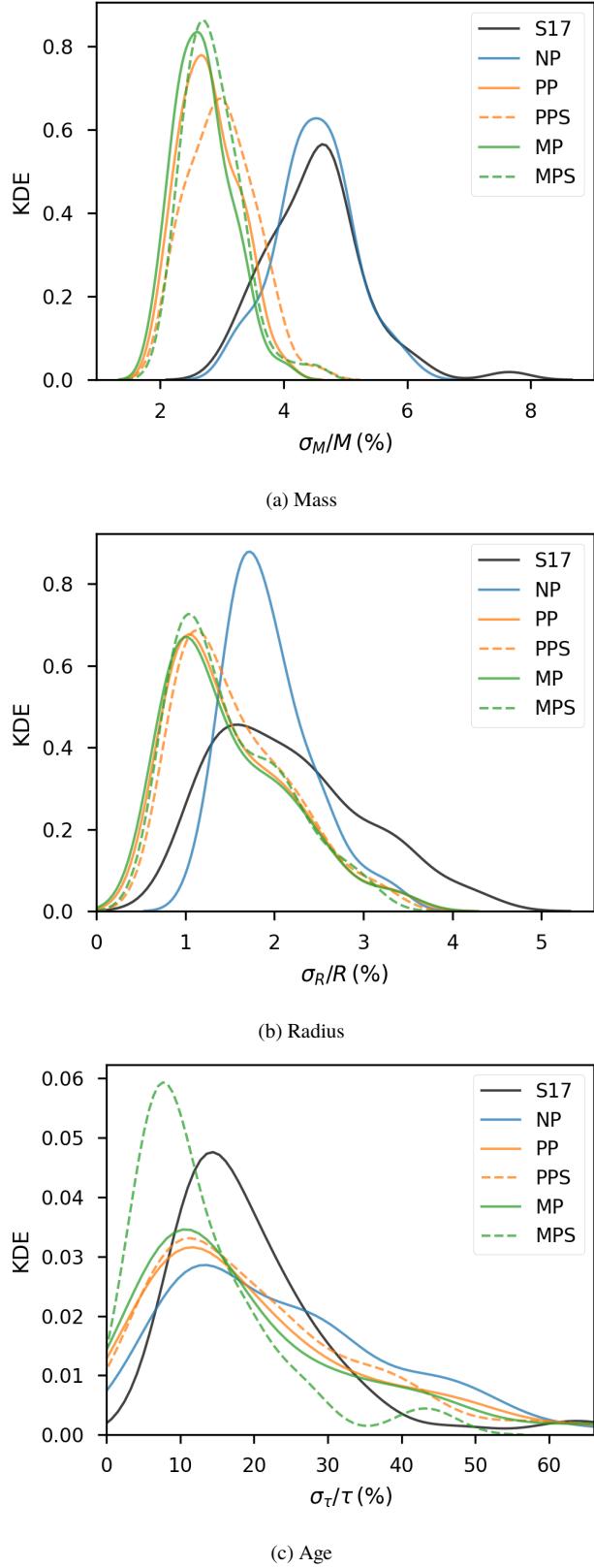


Figure 6. Kernel density estimates (KDEs) of the uncertainties in the results from each model compared with that of (S17).

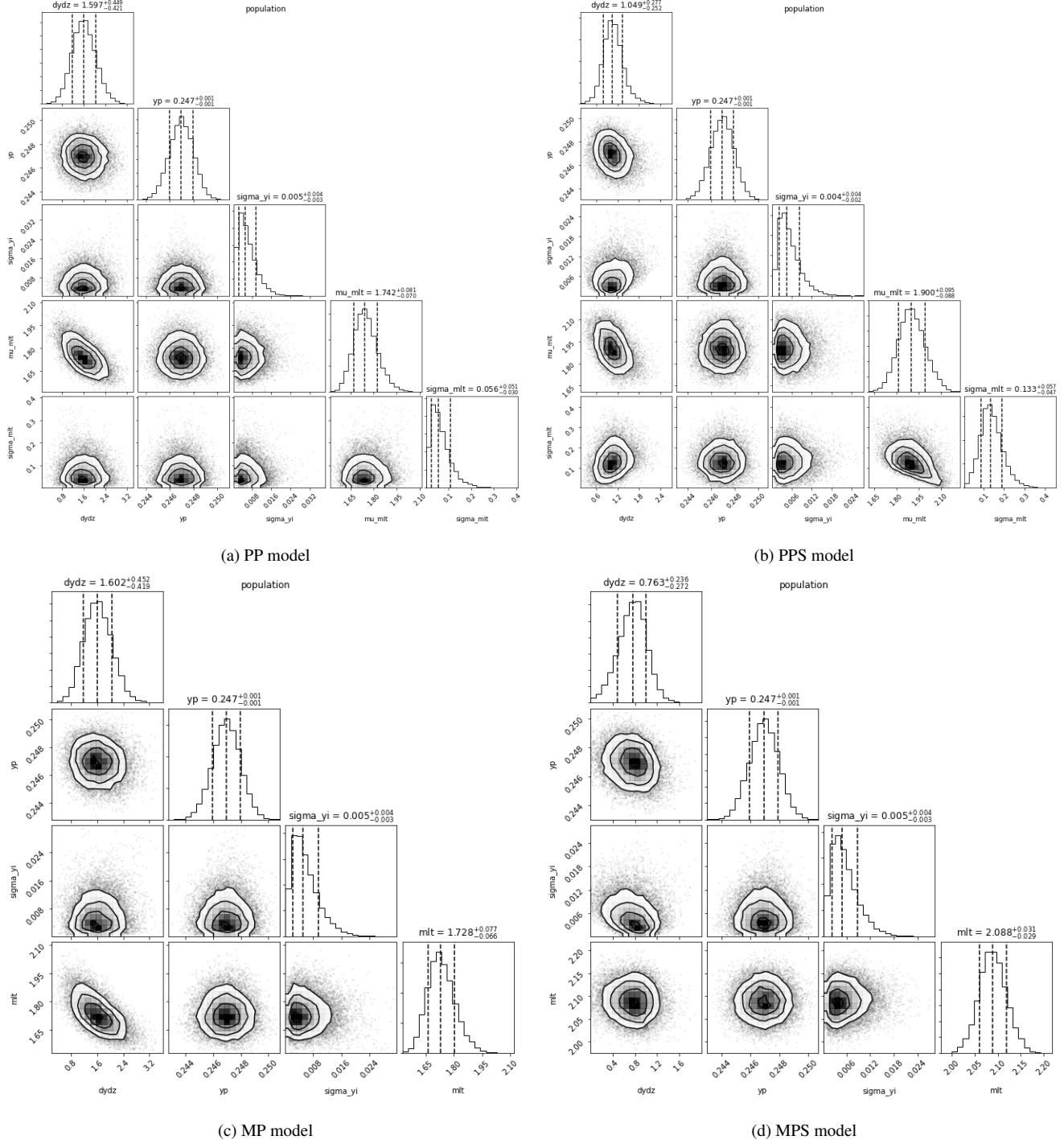


Figure 7. Corner plots showing the joint and marginalised sampled posterior distributions for the hyperparameters for both the PP and MP models, with and without the Sun. The vertical dashed lines give the 16th, 50th and 84th percentiles.

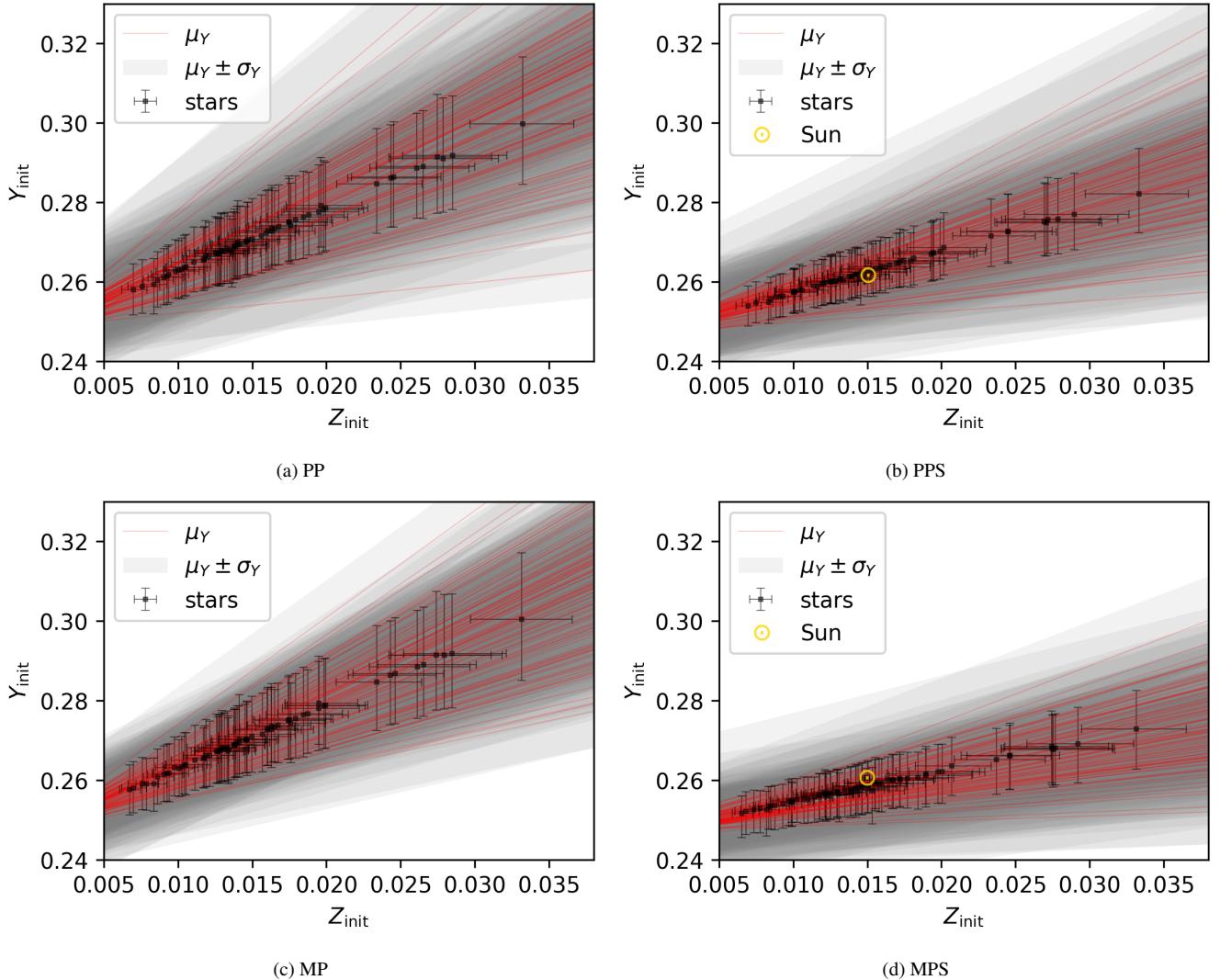


Figure 8. The results for initial helium against initial heavy-element fraction for each star. 100 random samples from the posterior for the population mean, $\mu_Y = Y_P + (\Delta Y / \Delta Z) Z_{\text{init}}$ and spread, $\mu_Y \pm \sigma_Y$ are shown in red and grey respectively. The Sun is shown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

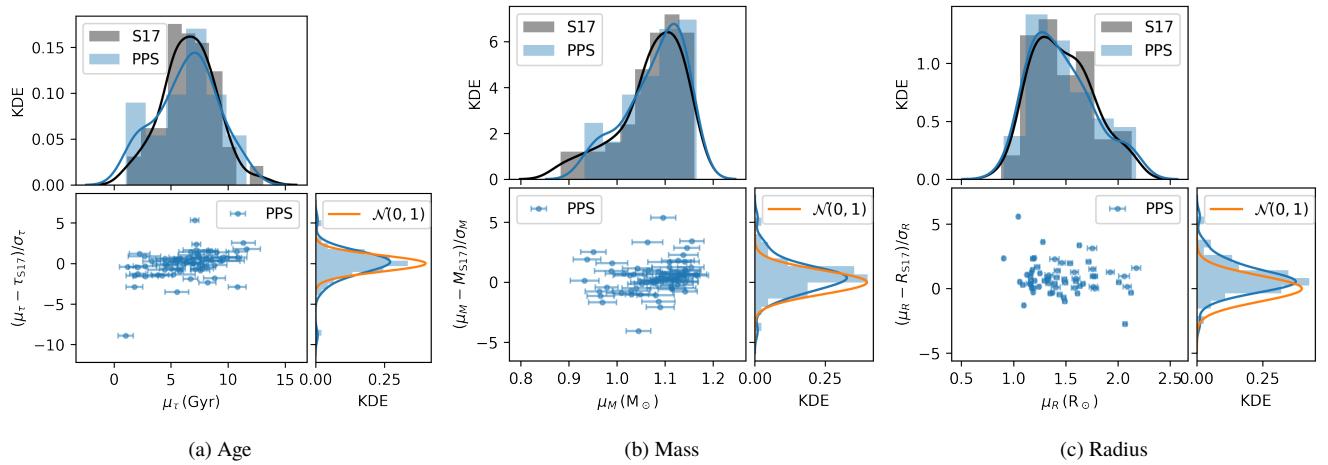


Figure 9. The mean and standard deviation in age, mass and radius results from the PPS model compared with the results (using the photometric temperature scale) from S17.

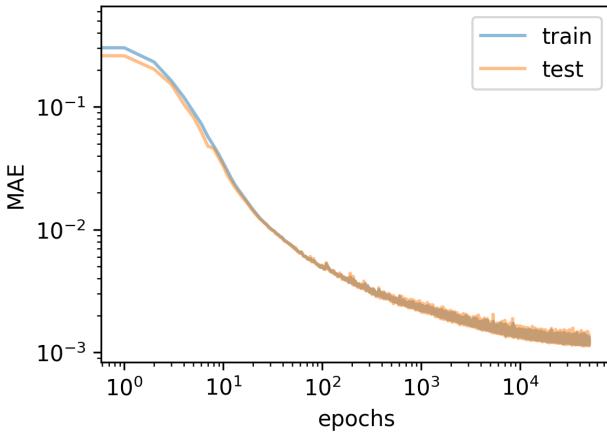


Figure B1. The MAE as a function of epochs for the train and test dataset.

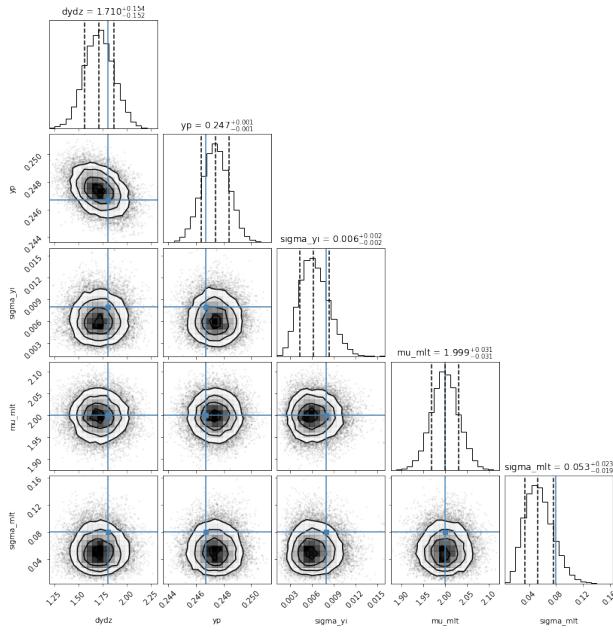


Figure C1. Corner plot showing the marginalised and joint posterior distributions between the NP model parameters for the synthetic stars. The true values are shown by the blue lines.

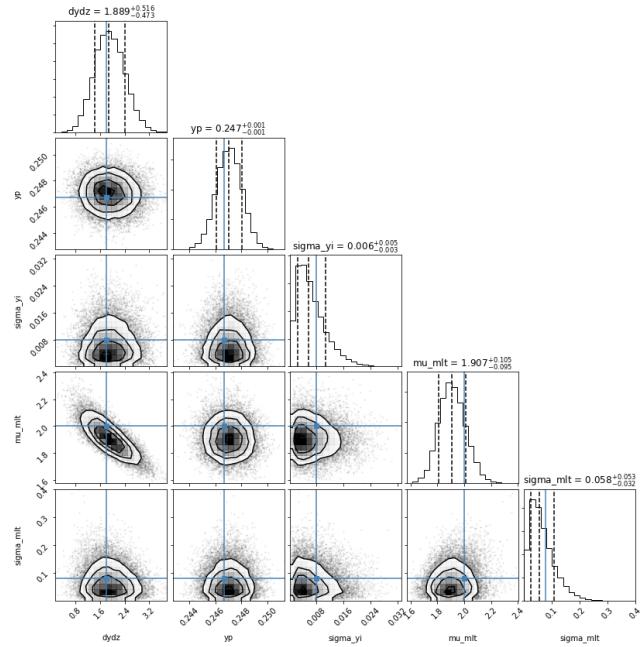


Figure C2. Corner plot showing the marginalised and joint posterior distributions between the NP model parameters for the synthetic stars. The true values are shown by the blue lines.

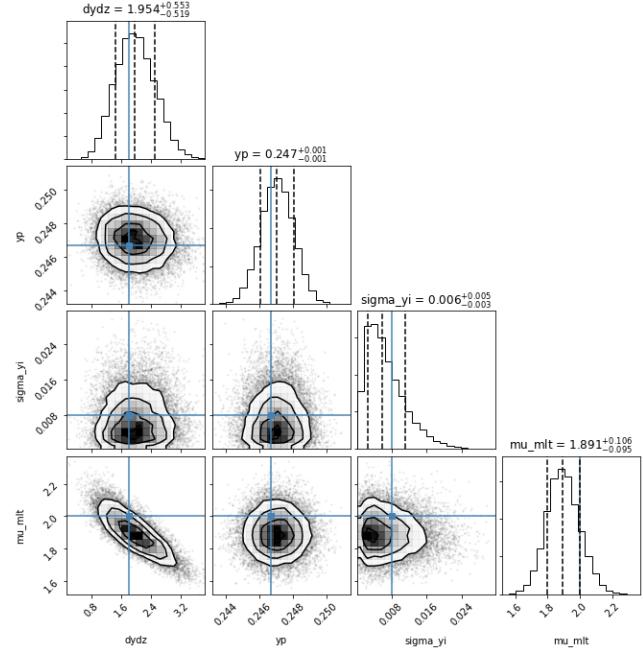


Figure C3. Corner plot showing the marginalised and joint posterior distributions between the MP model parameters for the synthetic stars. The true values are shown by the blue lines.

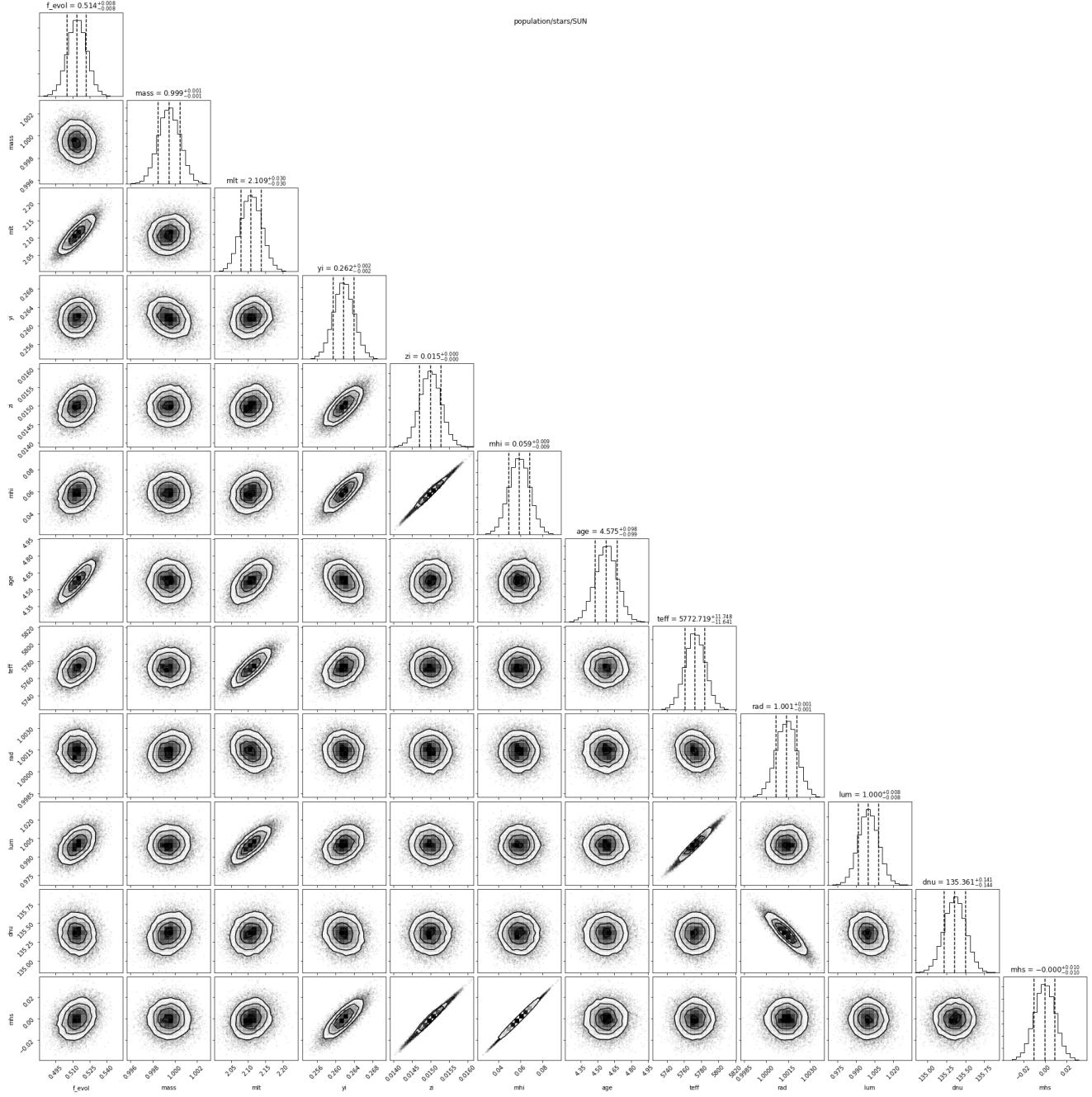


Figure D1. A corner plot showing the sampled marginal and joint posterior distributions for the Sun as a part of the PP model.