

# MADE: a spectroscopic mass, age, and distance estimator for red giant stars with Bayesian machine learning

Payel Das<sup>1</sup>★ and Jason L. Sanders<sup>2</sup>

<sup>1</sup>Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

<sup>2</sup>Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

Accepted 2018 October 10. Received 2018 October 9; in original form 2018 April 23

## ABSTRACT

We present a new approach (MADE) that generates mass, age, and distance estimates of red giant stars from a combination of astrometric, photometric, and spectroscopic data. The core of the approach is a Bayesian artificial neural network (ANN) that learns from and completely replaces stellar isochrones. The ANN is trained using a sample of red giant stars with mass estimates from asteroseismology. A Bayesian isochrone pipeline uses the astrometric, photometric, spectroscopic, and asteroseismology data to determine posterior distributions for the training outputs: mass, age, and distance. Given new inputs, posterior predictive distributions for the outputs are computed, taking into account both input uncertainties, and uncertainties in the ANN parameters. We apply MADE to  $\sim 10\,000$  red giants in the overlap between the 14th data release from the APO Galactic Evolution Experiment (APOGEE) and the *Tycho-Gaia* astrometric solution (TGAS). The ANN is able to reduce the uncertainty on mass, age, and distance estimates for training-set stars with high output uncertainties allocated through the Bayesian isochrone pipeline. The fractional uncertainties on mass are  $< 10$  per cent and on age are between 10 to 25 per cent. Moreover, the time taken for our ANN to predict masses, ages, and distances for the entire catalogue of APOGEE-TGAS stars is of a similar order of the time taken by the Bayesian isochrone pipeline to run on a handful of stars. Our resulting catalogue clearly demonstrates the expected thick- and thin-disc components in the  $[M/H]$ – $[\alpha/M]$  plane, when examined by age.

**Key words:** methods: data analysis – surveys – Galaxy: evolution – Galaxy: kinematics and dynamics.

## 1 INTRODUCTION

Knowledge of the distance, age, and chemical composition of stars is fundamental for understanding the Galaxy’s history of formation, enrichment, and dynamical evolution (Bland-Hawthorn & Gerhard 2016). When interpreted through chemodynamical models such as ‘extended distribution functions’ (Sanders & Binney 2015; Das & Binney 2016), they can constrain Galactic evolution as well as the gravitational potential and dark matter content. However, distance, age, and chemical composition are not measured directly and need to be inferred from stellar models.

The oldest method of measuring distance is through the measurement of a star’s parallax. This method has traditionally been limited to local Galactic studies, but the second data release from the *Gaia* mission (Gaia Collaboration 2016, 2018) has provided parallaxes for a billion stars down to 20th magnitude. Variable

stars, such as Cepheids, can be used as standard candles and so their apparent magnitudes can be mapped to a narrow range of intrinsic luminosities, allowing distances to be easily derived. This method can be used to large distances but only with a limited range of stellar types. Another example of a model-independent method estimates distances for spectroscopic twins of nearby stars with known parallaxes (Jofré et al. 2015, 2017). This method does not degrade with increasing stellar distance, but is limited by the availability of high-resolution stellar spectra and parallaxes for a range of stellar types.

The age of an individual star can be estimated through a range of almost entirely model-dependent methods (for a review, see Soderblom 2010). Pont & Eyer (2004) developed a Bayesian method to estimate the ages of dwarf stars from temperature, luminosity and metallicity data using isochrones. Several other authors (e.g. Burnett & Binney 2010; Carlin et al. 2015; Santiago et al. 2016) have since developed similar procedures to estimate the mass, age, distance, and metallicity of a star from photometric and spectroscopic data, and where available, luminosity. Parallaxes, which coupled with photometric data lay strong constraints on the

\* E-mail: [payel.das@physics.ox.ac.uk](mailto:payel.das@physics.ox.ac.uk)

intrinsic luminosity, have also been incorporated into these pipelines where available (e.g. Jørgensen & Lindegren 2005; da Silva et al. 2006; McMillan et al. 2017; Mints & Hekker 2018; Queiroz et al. 2018). Without an indication of the star's intrinsic luminosity from variable star relations or parallax, the inferred ages are only reliable in the turn-off region. The approach works well for stellar parameter estimation over a large range of distance.

In recent years, asteroseismology has revealed considerable potential in determining ages for giant stars. Surveys such as the Microvariability and Oscillations of Stars (MOST) mission (Guenther et al. 2005), the Convection, Rotation et Transits planétaires (CoRoT) mission (De Ridder et al. 2009), and Kepler (Bedding et al. 2010) have been extremely successful in probing the stellar interiors for red giants (e.g. Chaplin & Miglio 2013; Silva Aguirre et al. 2015). These missions take high-cadence, high-precision stellar photometry that shows oscillation modes in Fourier space. The frequency of maximum intensity,  $\nu_{\max}$ , and the frequency spacing of the modes,  $\Delta\nu$ , are related to the density and mass of the stars through simple scaling relations (Ulrich 1986; Kjeldsen & Bedding 1995). As giants spend most of their life in the main-sequence phase, the age at which they entered the giant branch is approximately their present age. This age is fixed by the mass of the star, i.e. their mass is a strong indicator of their age. Determination of ages for red giants with asteroseismology data has already been demonstrated to be a powerful probe of Milky Way chemodynamics, in particular when combined with spectroscopy (Miglio et al. 2013; Rodrigues et al. 2014; Stello et al. 2015; Casagrande et al. 2016; Anders et al. 2017).

At present, these asteroseismological surveys have yielded thousands of masses rather than millions. However, they have made it possible to determine the significance of mass-driven differences in the spectra of giants. This has significantly broadened the scope for asteroseismology through the development of ‘spectroscopic mass estimators’, which use spectral parameters or spectra to predict mass  $m$ . Masseron & Gilmore (2015) showed that masses of red giants could be derived from spectra given the [C/N] change due to dredge-up in the giant phase. Martig et al. (2016) developed simple linear regression models to estimate masses of giants from  $\log g$ ,  $T_{\text{eff}}$ , [Fe/H], carbon abundance, and nitrogen abundance using data from the APOGEE-Kepler Asteroseismology Science Consortium (APOKASC) catalogue (Pinsonneault et al. 2014). This catalogue contains stars in the cross-match between stars with spectroscopic data from the APO Galactic Evolution Experiment (APOGEE; Majewski et al. 2017) and asteroseismology from the Kepler mission (Haas et al. 2010). Ness et al. (2016) trained the *Cannon* (Ness et al. 2015) on this catalogue, a regression model that fits each pixel of a star's spectrum with a polynomial function of the spectroscopic labels ( $\log g$ ,  $T_{\text{eff}}$ , and [Fe/H],  $[\alpha/\text{Fe}]$ ,  $m$ ). The mass estimates can then be combined with isochrones to infer ages for all  $\sim 60\,000$  stars in APOGEE DR12 (Alam et al. 2015), rather than just the  $\sim 5000$  in APOKASC, clearly revealing relations such as variation of age with Galactic position in mono-abundance populations.

The potential of spectroscopic mass estimators in light of the second data release from *Gaia* is massive. The number of stars in spectroscopic surveys with accurate parallaxes has dramatically increased, and includes a significant number of red giants. However, to then derive ages and distances for these stars, a Bayesian isochrone pipeline needs to be applied using the spectroscopic estimate of mass. This paper expands the potential of a spectroscopic mass estimator to ages and distances using the Bayesian Mass, Age, and Distance Estimator (MADE). The estimator is an artificial neural network (ANN), which has been demonstrated to be highly flexible in reproducing non-linear relations (Gelder et al. 2014). It is trained on

stars with asteroseismology data as well as astrometric, photometric, and spectroscopic data, for which masses, ages, and distances are estimated using a Bayesian isochrone pipeline. The trained ANN completely replaces the need for stellar models. The ANN then calculates posterior predictive distributions for mass, age, and distance that account for measurement uncertainties in new data, and uncertainties in model parameters. All the codes for the routines in the paper can be cloned from <https://github.com/payeldas/MADE>.

Section 2 introduces the components of MADE and Section 3 describes how we develop the new Bayesian spectroscopic mass, age, and distance estimator. Section 4 presents an application of MADE to observations and Section 5 summarizes the work and looks towards future data releases.

## 2 THE MADE APPROACH HERE

Here, we introduce the components of MADE and the notation used to represent inputs, outputs, and model parameters.

### 2.1 MADE components

The core of MADE is a PYTHON class that builds a Bayesian ANN given training data, and then applies this ANN to unseen data. Here, the training sample comprises red giants (i.e.  $\log g < 3.5$ ) for which there are asteroseismology estimates for the current mass as well as parallaxes from astrometry, photometry, and spectroscopy. A Bayesian isochrone pipeline is applied to the training sample to generate posterior distributions for the desired outputs of mass, age, and distance. The ANN then learns the relationship between these outputs and input astrometric, photometric, and spectroscopic data. Given unseen astrometric, photometric, and spectroscopic input data of a red giant star, the ANN is used to generate predictive posterior distributions for the mass, age, and distance (see Section 3). In this way, the ANN completely replaces the reliance on isochrones for estimating mass, age, and distance from the set of inputs.

### 2.2 Random variables in MADE

In Bayesian statistics, everything (i.e. input labels, outputs, and model parameters) is considered to be a random variable (RV), i.e. is described by a distribution rather than a single value. A distinction is made between an observed RV (i.e. has predicted and measured values that are linked through a likelihood depending on a measurement uncertainty), and an unobserved RV (i.e. model parameters), which is assigned a prior. The RVs in MADE are given in Table 1 and explained in more detail below.

A number of observed RVs and model parameters are associated with the Bayesian isochrone pipeline used to generate the training set (Section 3.1). The predicted observed RVs are the properties of the star  $i$  generated by the isochrones and distance model

$$v^i = (m^i, H^i, (J - K_s)^i, \varpi^i), \quad (1)$$

where  $m$  is the current mass,  $H$  is the  $H$ -band magnitude,  $J - K_s$  is a colour, and  $\varpi$  is the parallax. The accompanying measured observed RV stellar properties are denoted by  $\tilde{v}^i$ . These are associated with measurement uncertainties,  $\tilde{\rho}^i$ . The parameters,  $\phi^i$ , of the star  $i$  estimated from the Bayesian isochrone pipeline are given by

$$\phi^i = (\mathcal{M}^i, \tau^i, [\text{M}/\text{H}]^i, s^i), \quad (2)$$

where  $\mathcal{M}$  is the initial mass,  $\tau$  is the age,  $[\text{M}/\text{H}]$  is the metallicity, and  $s$  is the distance. The priors on these parameters are informed

**Table 1.** Random variables in MADE. The predicted inputs and outputs in the training sample, and the predicted inputs for new stars, also have measured equivalents that are denoted with a tilde. The measured quantities are given by the predicted quantities convolved with uncertainties.

Component	Notation	Description
Isochrone pipeline observed RVs	$v_j^i$	$j$ -th property of $i$ -th star predicted by the isochrone pipeline
Isochrone pipeline unobserved RVs	$\tilde{v}_j^i$ $\phi^i$ GP	Uncertainty in $j$ -th measured property of $i$ -th star Parameters of $i$ -th star predicted by the isochrone pipeline Milky Way model that provides the prior on parameters of the isochrone pipeline
ANN observed RVs (training)	$x_j^i$ $y_j^i$ $u_j^i$ $\tilde{u}_j^i$	$j$ -th predicted input of $i$ -th star in training sample $j$ -th predicted output of $i$ -th star in training sample $j$ -th predicted property of $i$ -th star in training sample Uncertainty in $j$ -th measured property of $i$ -th star in training sample
ANN observed RVs (prediction)	$x_{N,j}^i$ $y_{N,j}^i$ $\tilde{\sigma}_{N,j}^i$	$j$ -th predicted input for $i$ -th new star $j$ -th predicted output for $i$ -th new star Uncertainty on the $j$ -th measured input for the $i$ -th new star
ANN unobserved RVs	$\theta$ $\hat{\theta}$	Parameters of the ANN Priors on parameters of the ANN

by the Galaxy Prior (GP), which is a model for the Milky Way (Section 3.1.2).

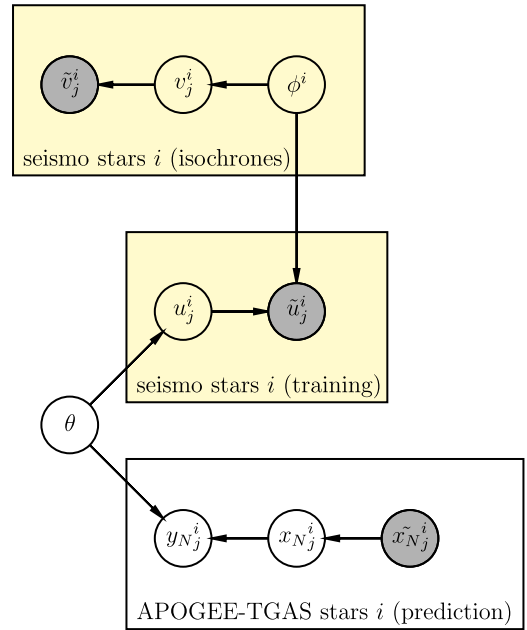
The observed RVs of the ANN (Section 3.2) consist of those associated with the training sample and those associated with new data.  $x_j^i$  represents the  $j$ -th predicted input of the  $i$ -th repeated unit (which in practice is a star) in the training sample.  $y_j^i$  represents the  $j$ -th predicted output of the  $i$ -th star in the training sample. Measurements of the predicted inputs and outputs are  $\tilde{x}_j^i$  and  $\tilde{y}_j^i$ , respectively. We further introduce  $u_j^i$  and  $\tilde{u}_j^i$ , which are the predicted and measured properties of the star in the training sample, respectively, where ‘properties’ combine input and outputs into a single vector. We associate the measured properties of the star in the training sample with a measurement uncertainty  $\tilde{\sigma}_j^i$ . Predicted inputs and outputs for new stars are denoted by  $x_{N,j}^i$  and  $y_{N,j}^i$ , respectively, with measurements of the input labels denoted by  $\tilde{x}_{N,j}^i$ . The measurement uncertainty on the outputs is  $\tilde{\sigma}_N^i$  and model uncertainty on the predicted outputs is  $\sigma_N^i$ . We assume all observed RVs of the ANN are scaled by the mean and standard deviation of the corresponding variable of the training sample (i.e. subtract the mean and then divide by the standard deviation). The unobserved RVs are the parameters of the ANN,  $\theta$ , and its priors  $\hat{\theta}$ .

### 3 A BAYESIAN SPECTROSCOPIC MASS, AGE, AND DISTANCE ESTIMATOR

Here, we discuss how we (i) build a training set for the estimator using a Bayesian isochrone pipeline, (ii) train the ANN using a Bayesian method, and (iii) use the ANN to generate posterior predictive distributions. The connection between these three components is illustrated in Fig. 1.

#### 3.1 Building a training set

We select a training sample for which there are current mass constraints from asteroseismology in addition to parallax constraints from astrometry, photometry, and spectroscopy. We then apply a Bayesian isochrone pipeline, similar to many in the literature (e.g. Burnett & Binney 2010; Carlin et al. 2015; McMillan et al. 2017) to estimate the posterior distributions of mass, age, and distance for the stars in the training sample. The main difference however with these pipelines is that we do not estimate the line-of-sight



**Figure 1.** A graphical model of the relationships between RVs in the development of the training set using the Bayesian isochrone pipeline with the seismo stars (top panel), the training of the ANN with the seismo stars (middle panel), and the application of the ANN to the APOGEE-TGAS stars (bottom panel). Each panel represents a joint probability, where arrows indicate a correlation. Filled circles indicate observed RVs and open circles indicate unobserved RVs, i.e. parameters of the ANN or the isochrone model. Variables within a box are repeated, and therefore represent a product of probabilities. The yellow colour highlights the training components.

(LOS) extinction as an additional output here. It is an output with an irregular dependence on Galactic location and therefore difficult for an ANN to capture without a sufficiently large training sample. Most of the asteroseismology data we use currently only probe one LOS. We leave this exercise of predicting the LOS extinction to a future project. We instead make a point estimate of the LOS extinction in each photometric band (see below for details).

### 3.1.1 The Bayesian isochrone pipeline

Applying Bayes' law to each star gives

$$p(\phi^i | \tilde{v}^i, l^i, b^i) = \frac{p(\tilde{v}^i | \phi^i, l^i, b^i) p(\phi^i | l^i, b^i)}{p(\tilde{v}^i)}, \quad (3)$$

where  $(l^i, b^i)$  are the predicted sky positions of star  $i$  in Galactic coordinates (which we assume to be the same as the observed sky positions) and  $p(\tilde{v}^i)$  is an unimportant normalization. The first panel in Fig. 1 shows the correlations between the variables in the Bayesian isochrone pipeline and how the outputs feed into the training of the ANN (Section 3.2). The apparent magnitudes of star  $i$  are extinction-corrected using the state-of-the-art `Combined15` map compiled by Bovy et al. (2016) in the `mwDust` package. This is a composite three-dimensional map of the integrated extinction. It was constructed from a number of extinction maps produced using stellar photometry but with differing sky coverage. The core is constructed from the maps of Marshall et al. (2006) and Green et al. (2015). The remaining parts of the map are filled in using the map of Drimmel, Cabrera-Lavers & López-Corredoira (2003), which is normalized to the map of Schlegel, Finkbeiner & Davis (1998).

The likelihood of the star's observed properties,  $p(\tilde{v}^i | \phi^i, l^i, b^i)$ , is assumed to be the product of the separate likelihoods. Each likelihood is represented by a Gaussian distribution

$$G(\tilde{v}_j^i, v_j^i, \tilde{\rho}_j^i) = \frac{1}{\sqrt{2\pi}\tilde{\rho}_j^i} \exp\left(-\frac{(\tilde{v}_j^i - v_j^i)^2}{2(\tilde{\rho}_j^i)^2}\right), \quad (4)$$

thus

$$p(\tilde{v}^i | \phi^i, l^i, b^i) = \prod_j G(\tilde{v}_j^i, v_j^i, \tilde{\rho}_j^i). \quad (5)$$

The stellar properties,  $v_j^i$ , are predicted from the model parameters,  $\phi^i$ , using

$$(v_0^i, v_1^i, v_2^i) = \mathcal{I}(\phi^i) \\ v_4^i = 1/\phi_3^i, \quad (6)$$

where  $\mathcal{I}$  denotes the isochrones and the second line refers to the distance model. When we refer to the Bayesian isochrone pipeline, we allude to both the isochrones and the distance model. Here, we employ PARSEC isochrones v1.1 ( $\eta = 0.2$ ; Bressan et al. 2012) evaluated for 57 metallicities ranging between  $-2.192$  and  $0.696$ , and 353 ages ranging between  $\log_{10}\tau = 6.60$  and  $10.12$  (i.e. a spacing of  $\Delta\log_{10}\tau = 0.01$ ) for which we create a dictionary of interpolants in PYTHON that returns luminosity,  $\log g$ ,  $T_{\text{eff}}$ , and apparent magnitudes given the metallicity, age, and mass of a star.

### 3.1.2 The Milky Way prior

We base the prior  $(\phi^i | l^i, b^i)$  on the Milky Way model of Binney et al. (2014) but add a component for the bulge (the superscript  $i$  is omitted in the following)

$$p(\mathcal{M}, \tau, [\text{M}/\text{H}], s | l, b) = s^2 \epsilon(\mathcal{M}) \sum_{k=1}^4 p_k([\text{M}/\text{H}]) p_k(\tau) p_k(R, z), \quad (7)$$

where  $k = 1, 2, 3, 4$  correspond to a bulge, thin disc, thick disc, and stellar halo, respectively, the  $s^2$  term accounts for the Jacobian

of the transformation of spatial coordinates, and  $\epsilon(\mathcal{M})$  is the initial mass function. We consider that of Kroupa, Tout & Gilmore (1993)

$$\epsilon(\mathcal{M}) = \begin{cases} 0.035\mathcal{M}^{-1.5} & \text{if } 0.08 \leq \mathcal{M}/\text{M}_\odot < 0.5 \\ 0.019\mathcal{M}^{-2.2} & \text{if } 0.5 \leq \mathcal{M}/\text{M}_\odot < 1.0 \\ 0.019\mathcal{M}^{-2.7} & \text{if } \mathcal{M}/\text{M}_\odot \geq 1.0. \end{cases} \quad (8)$$

The spatial prior for the bulge component is based on Dehnen, McLaughlin & Sachania (2006) and the metallicity prior approximately reflects the distributions found in the Abundances and Radial velocity Galactic Origins Survey (ARGOS) of the bulge (Ness et al. 2013). We update the truncated age prior of Binney et al. (2014) for the thick and thin discs, which assigns zero probability to young, high-latitude distant stars, to a smooth one. We also adopt a Gaussian for stellar halo ages, based on the work of Jofré & Weiss (2011). For the bulge, we assume a broad age distribution based on Bensby et al. (2013).

In summary, for the four components we assume:  
Bulge ( $k = 1$ ):

$$p_1([\text{M}/\text{H}]) = G([\text{M}/\text{H}], \mu_{[\text{M}/\text{H}],1}, \sigma_{[\text{M}/\text{H}],1}), \\ p_1(\tau) = G(\tau, \mu_{\tau,1}, \sigma_{\tau,1}), \\ p_1(R, z) \propto \frac{(1+m)^{(\gamma-\delta)}}{m^\gamma} \exp[-(mr_0/r_t)^2],$$

$$\text{where } m(R, z) = \sqrt{(R/r_0)^2 + (z/q r_0)^2}. \quad (9)$$

Thin disc ( $k = 2$ ):

$$p_2([\text{M}/\text{H}]) = G([\text{M}/\text{H}], \mu_{[\text{M}/\text{H}],2}, \sigma_{[\text{M}/\text{H}],2}), \\ p_2(\tau) \propto \begin{cases} \exp(\frac{\tau}{8.4}) & \text{if } \tau/\text{Gyr} \leq 8 \\ 2.6 \exp\left(-0.5\frac{(\tau-8)^2}{1.5^2}\right) & \text{if } \tau/\text{Gyr} > 8 \end{cases} \\ p_2(R, z) \propto \exp\left(-\frac{R}{R_{d,2}} - \frac{|z|}{z_{d,2}}\right). \quad (10)$$

Thick disc ( $k = 3$ ):

$$p_3([\text{M}/\text{H}]) = G([\text{M}/\text{H}], \mu_{[\text{M}/\text{H}],3}, \sigma_{[\text{M}/\text{H}],3}), \\ p_3(\tau) = G(\tau, \mu_{\tau,3}, \sigma_{\tau,3}), \\ p_3(R, z) \propto \exp\left(-\frac{R}{R_{d,3}} - \frac{|z|}{z_{d,3}}\right). \quad (11)$$

Stellar halo ( $k = 4$ ):

$$p_4([\text{M}/\text{H}]) = G([\text{M}/\text{H}], \mu_{[\text{M}/\text{H}],4}, \sigma_{[\text{M}/\text{H}],4}), \\ p_4(\tau) \propto G(\tau, \mu_{\tau,4}, \sigma_{\tau,4}) \\ p_4(R, z) \propto r^{-3.39}. \quad (12)$$

The parameters of the prior are given in Table 2. The thin disc is normalized to have a local density of  $0.04 \text{ M}_\odot \text{ pc}^{-3}$  (Bovy 2017). The thick disc and stellar halo are normalized so that the ratios of their local densities with that of the thin disc are 0.04 and 0.005, respectively (Bland-Hawthorn & Gerhard 2016; Bovy 2017). Finally, the bulge disc is normalized to have a central density of  $35.45 \text{ M}_\odot \text{ pc}^{-3}$  (Robin et al. 2012). An overall prior is imposed that constrains metallicities and ages to the range covered by the isochrones.

We calculate  $p(\phi^i | u^i)$  directly on a grid of all 353 ages,  $\tau^i$ , all metallicities,  $[\text{M}/\text{H}]^i$ , lying within  $3\sigma$  of the measured metallicity, 2000 initial masses,  $\mathcal{M}^i$ , ranging between the minimum and maximum masses of the relevant isochrone, and 30 distances,  $s^i$ , based on a linear grid of parallaxes ranging between  $3\sigma$  below and above the measured parallax.  $\tau^i$  and  $[\text{M}/\text{H}]^i$  define the isochrone and  $\mathcal{M}^i$  defines the position along the isochrone. The position on the isochrone returns a prediction of mass,  $m^i$ , colours, and



**Table 2.** Parameters of the four-component Milky Way model.

Component	Parameter	Value
Bulge	$\mu_{[M/H], 1/\text{dex}}$	−0.3
	$\sigma_{[M/H], 1/\text{dex}}$	0.3
	$\mu_{\tau, 1/\text{Gyr}}$	5.0
	$\sigma_{\tau, 1/\text{Gyr}}$	5.0
	$q$	0.5
	$\gamma$	0.0
	$\delta$	1.8
	$r_0/\text{kpc}$	0.075
	$r_i/\text{kpc}$	2.1
Thin disc	$\mu_{[M/H], 2/\text{dex}}$	0.0
	$\sigma_{[M/H], 2/\text{dex}}$	0.2
	$R_d, 2/\text{kpc}$	2.6
	$z_d, 2/\text{kpc}$	0.3
Thick disc	$\mu_{[M/H], 3/\text{dex}}$	−0.6
	$\sigma_{[M/H], 3/\text{dex}}$	0.5
	$\mu_{\tau, 3/\text{Gyr}}$	10.0
	$\sigma_{\tau, 3/\text{Gyr}}$	2.0
	$R_d, 3/\text{kpc}$	3.6
	$z_d, 3/\text{kpc}$	0.9
Stellar halo	$\mu_{[M/H], 4/\text{dex}}$	−1.6
	$\sigma_{[M/H], 4/\text{dex}}$	0.5
	$\mu_{\tau, 4/\text{Gyr}}$	11.0
	$\sigma_{\tau, 4/\text{Gyr}}$	2.0

absolute magnitudes, i.e. the predicted stellar properties  $v^i$ . The choice of  $s^i$  tells us the apparent magnitudes. We calculate the first and second moments of the logarithm of age,  $\log_{10}\tau$ , metallicity,  $[M/H]$ , logarithm of the current mass,  $\log_{10}m$ , and the distance modulus,  $\mu$ , as our outputs and output uncertainties, e.g.

$$\langle \log_{10} \tau \rangle = \left( \int d\phi \log_{10} \tau p(u|\phi, l, b) p(\phi|l, b) \right) / \left( \int d\phi p(u|\phi, l, b) p(\phi|l, b) \right). \quad (13)$$

### 3.2 The Bayesian ANN

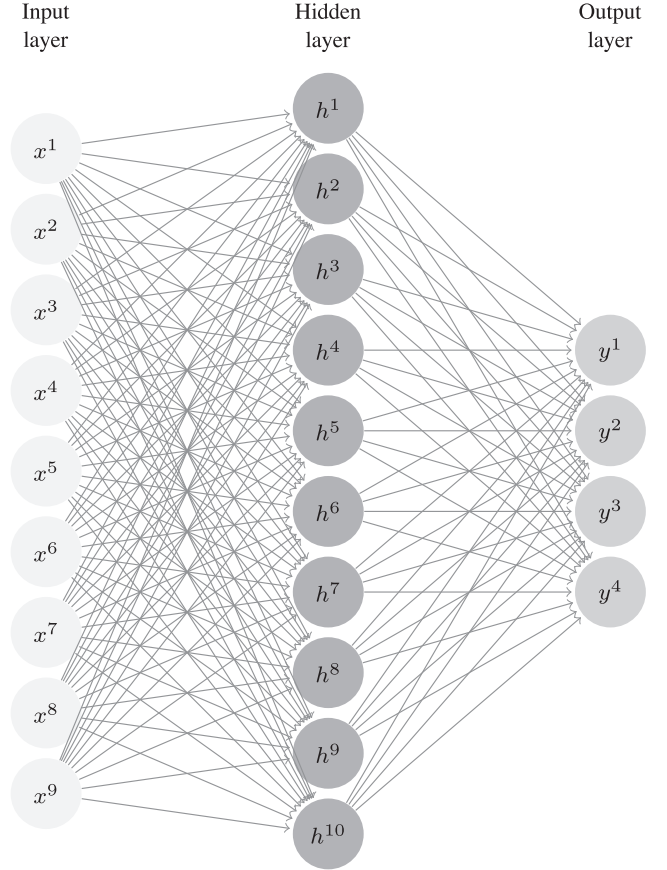
An ANN consists of interconnected layers of neurons, which represent linear or non-linear transformations by an ‘activation’ function. The first layer is the input layer comprising the same number of neurons as the number of inputs,  $n_{\text{in}}$ . The central layers are hidden layers, each with potentially a different number of neurons per hidden layer,  $n_{\text{hid}}$ . The final layer is the output layer with the same number of neurons as the number of outputs,  $n_{\text{out}}$ .

#### 3.2.1 ANN architecture

The universality theorem<sup>1</sup> tells us that even ANNs with a single hidden layer can be used to approximate any continuous function to any desired precision (given a sufficient number of hidden neurons). We expect a continuous mapping from one finite space to another and therefore consider a simple ANN architecture that contains a single hidden layer.

In a feed forward ANN, only neurons in adjacent layers are connected to one another. We assume linear activation functions for the input and output layers and a tanh activation function for the hidden layer, which maps variables ranging from  $-\infty$  to  $\infty$  to a domain

<sup>1</sup> A visualization of this theorem can be found at <http://neuralnetworksanddeeplearning.com/chap4.html>.

**Figure 2.** Architecture of ANN assuming  $n_{\text{in}} = 9$  (light grey),  $n_{\text{out}} = 4$  (middle grey), and one hidden layer with  $n_{\text{hid}} = 10$  neurons (dark grey).

extending between  $-1$  and  $1$ . One can imagine easily replicating non-linear relations by stacking shifted and differentially stretched versions of this sigmoidal function. Therefore, the predicted outputs  $y_i$  are calculated from the predicted inputs  $x_i$  by

$$y^i = (w_{h,\text{out}} \tanh(w_{h,\text{in}} x^i + b_{h,\text{in}})) + b_{h,\text{out}}, \quad (14)$$

where  $w_{h,\text{in}}$  is a  $n_{\text{hid}} \times n_{\text{in}}$  matrix of weights,  $b_{h,\text{in}}$  is a length- $n_{\text{hid}}$  vector of biases,  $w_{h,\text{out}}$  is a length- $n_{\text{hid}}$  vector of weights, and  $b_{h,\text{out}}$  is a length- $n_{\text{out}}$  vector of biases. This architecture has  $n_{\theta} = n_{\text{hid}}(n_{\text{in}} + 2) + n_{\text{out}}$  model parameters. Fig. 2 shows the structure of the ANN in the case of  $n_{\text{hid}} = 10$ .

#### 3.2.2 Training the ANN

The posterior distributions of model parameters,  $\theta$ , can be estimated using Bayes’ law

$$p(\theta|\tilde{\mathbf{u}}) = \frac{p(\tilde{\mathbf{u}}|\theta)p(\theta)}{p(\tilde{\mathbf{u}})}, \quad (15)$$

where  $p(\tilde{\mathbf{u}}|\theta)$  is the joint likelihood of the measured stellar properties given the model parameters,  $p(\theta)$  is the prior on the model parameters, and  $p(\tilde{\mathbf{u}})$  is the distribution of the measured stellar properties.  $p(\tilde{\mathbf{u}})$  is the same for every model and can be ignored. The correlations between the variables are illustrated in the middle panel of Fig. 1. The figure also shows how the Bayesian isochrone pipeline feeds into the training of the ANN, and how the parameter distributions determined at this stage are used to generate posterior predictive distributions in Section 3.3. We do not explicitly include

measurement and model uncertainties in the representation of our probabilities throughout the paper as they only appear in likelihoods. The likelihood of the star's measured properties  $p(\tilde{\mathbf{u}}|\theta)$  is assumed to be the product of the likelihoods of each measured stellar property. Assuming Gaussian measurement uncertainties, we can represent each likelihood as

$$G(\tilde{u}_j^i, u_j^i, \tilde{\sigma}_j^i) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_j^i} \exp\left(-\frac{(\tilde{u}_j^i - u_j^i)^2}{2(\tilde{\sigma}_j^i)^2}\right). \quad (16)$$

Thus,

$$p(\tilde{\mathbf{u}}|\theta) = \prod_i \prod_j G(\tilde{u}_j^i, u_j^i, \tilde{\sigma}_j^i). \quad (17)$$

The predicted outputs are those generated by the ANN. The predicted inputs are initially unknown and therefore assigned Gaussian priors with zero mean and a large standard deviation of 10. We assume Gaussian priors with mean zero and standard deviation 1 for the ANN parameters.

The No-U-Turn Sampler (NUTS, i.e. adaptive Hamiltonian Monte Carlo, Hoffman & Gelman 2014) in PyMC3 is used to train the ANN on 80 per cent stars randomly selected from a sample of stars for which both inputs and outputs have been measured. The remaining 20 per cent is used as an independent test of the ANN.

### 3.3 Generating posterior predictive distributions

Once we have obtained posterior distributions for the parameters of the ANN,  $p(\theta|\tilde{\mathbf{u}})$ , we can calculate posterior predictive distributions for selected predicted stellar properties of new stars,  $\mathbf{y}_N$ , given the training sample and the new measured inputs, i.e.

$$p(\mathbf{y}_N|\tilde{\mathbf{u}}, \tilde{\mathbf{x}}_N) = \int p(\mathbf{y}_N|\theta) p(\theta|\tilde{\mathbf{u}}) p(\mathbf{y}_N|\mathbf{x}_N) p(\mathbf{x}_N|\tilde{\mathbf{x}}_N) d\theta d\mathbf{x}_N, \quad (18)$$

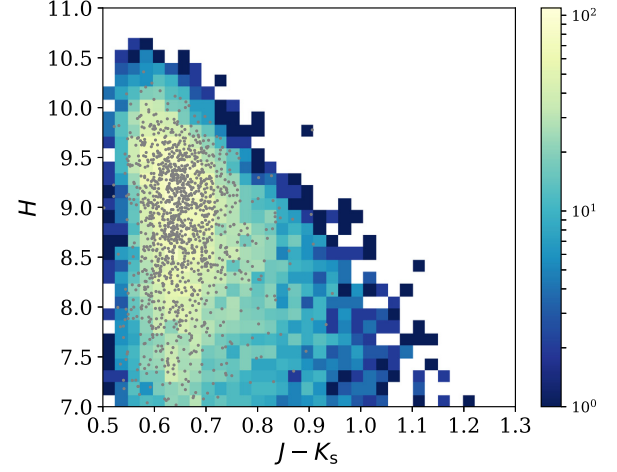
where  $p(\mathbf{y}_N|\theta)$  is the probability of the new predicted outputs given some set of model parameters,  $p(\theta|\tilde{\mathbf{u}})$  is the posterior distributions of the ANN model parameters evaluated in Section 3.2.2,  $p(\mathbf{y}_N|\mathbf{x}_N)$  gives the distribution of new predicted outputs given new predicted inputs, which are simply the predictions of the trained ANN.  $p(\mathbf{x}_N|\tilde{\mathbf{x}}_N)$  is the distribution of new predicted inputs given the new measured inputs, which is the likelihood as given in equation (17). Marginalizing the product of these probabilities over the parameters of the ANN and the new inputs give the posterior predictive distributions on the new outputs. Generating the posterior predictive distributions therefore does not require one to engage with the isochrones. The correlations between the variables are illustrated in the bottom panel of Fig. 1. The figure also shows how the parameter distributions found in the training process are linked to the new predicted outputs.

## 4 APPLICATION

Here, we introduce the APOGEE-TGAS sample of stars and a further subsample for which current masses have been determined from asteroseismology.

### 4.1 The APOGEE-TGAS catalogue

APOGEE spectra are taken in the  $H$  band with a resolution  $R \sim 22\,500$ . The first phase of the APOGEE Survey (APOGEE-1) was carried out between 2011 September and 2014 July (Majewski et al. 2017). Observations were performed with the APOGEE-North spectrograph on the Sloan Foundation 2.5 m Telescope of Apache



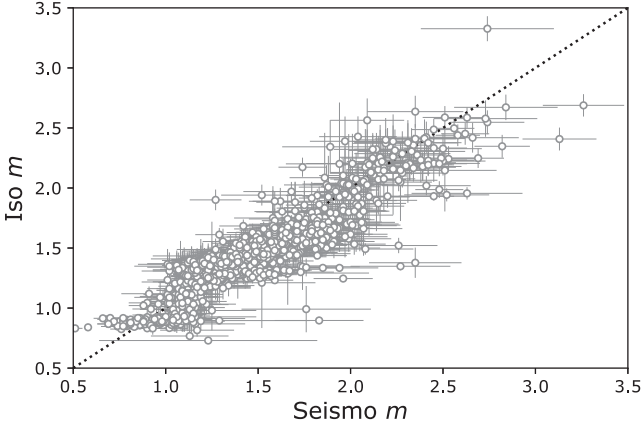
**Figure 3.** Colour–magnitude distribution of stars in the APOGEE-TGAS sample. The grey dots show the location of the seismo sample.

Point Observatory (APO). The second phase, APOGEE-2, started in 2014 July and will be completed in the summer of 2020 (Zasowski et al. 2017). Observations will also be taken with the APOGEE-South spectrograph on the Irénée du Pont 2.5 m Telescope of Las Campanas Observatory (LCO). Data Release 14 (DR14; Abolfathi et al. 2018) contains  $\sim 263\,000$  of mostly red giants, but with a significant contribution from red dwarf stars. APOGEE DR14 releases spectra (Nidever et al. 2015) and derived spectroscopic properties that include 20 individual chemical abundances (García Pérez et al. 2016). Surface gravities are calibrated using independent asteroseismology determinations with the Kepler mission (Haas et al. 2010). Kepler is described in more detail below.

The *Gaia* DR1 TGAS catalogue (Michalik et al. 2015) provides positions, parallaxes, and proper motions for the 2.5 million *Tycho*-2 stars (Høg et al. 2000). APOGEE DR14 includes a cross-match with these stars that results in  $\sim 46\,000$  stars. We select unique stars by keeping those with the highest signal-to-noise ratio, and remove stars with unreliable abundances and those that are cluster or calibration targets. We then select stars within the designed APOGEE colour–magnitude box, i.e.  $0.5 \leq (J - K_s)_0 \leq 1.3$  (we imposed the upper bound) and  $7.0 < H < 13.8$ . These operations reduce the sample to 10 074 stars. The distribution of the final sample in the space of colours and magnitudes is shown in Fig. 3. Assuming  $\log g \leq 3.5$  for giant stars (Hekker et al. 2011), 10 016 are very likely to be giants indicating a very small contamination of dwarf stars. We therefore assume the catalogue to comprise only giants.

### 4.2 The seismo-APOGEE-TGAS (seismo) sample

The Kepler spacecraft was launched in 2009 March (Haas et al. 2010) and spent a little over four years monitoring the brightness of more than 150 000 stars in the Cygnus-Lyra region, with the primary science objective of detecting transit-driven exoplanet stars. This was succeeded by the K2 mission (Howell et al. 2014) in 2014 June. We construct a catalogue of asteroseismology masses by combining the catalogue of Vrard, Mosser & Samadi (2016) with the first APOKASC catalogue (Pinsonneault et al. 2014). Both initially determine mode spacings in the Kepler photometric data and then apply scaling relations to determine masses and radii. 1235 of these stars also have TGAS parallaxes, creating the seismo-APOGEE-TGAS sample, or seismo sample for short. The location



**Figure 4.** Masses determined from asteroseismology (seismo  $m$ ) against masses determined from the Bayesian isochrone pipeline (iso  $m$ ). The uncertainty in the isochrone mass is estimated from  $10^{\mu \log_{10} m} \sigma_{\log_{10} m}$ .

of these stars in the colour–magnitude diagram is also shown in Fig. 3. They have a similar distribution to the APOGEE-TGAS stars in this space.

#### 4.3 A new Bayesian spectroscopic mass, age, and distance estimator

We apply the Bayesian isochrone pipeline described in Section 3.1.1 to the seismo sample of stars and determine the first and second moments of  $\log_{10} \tau$ ,  $[M/H]$ ,  $\log_{10} m$ , and  $\mu$ . Fig. 4 compares the asteroseismology masses with the isochrone masses. We do expect some differences between the masses as the isochrone masses are informed by astrometric, photometric, and spectroscopic data, as well as the asteroseismology data. In general though they agree well. The uncertainty in isochrone masses is generally smaller due to a larger number of observational constraints being used, and priors at the minimum and maximum ages. The prior on age in the Bayesian isochrone pipeline results in a cut-off at the lowest and maximum masses, due to the mass–age relation for red giant stars. The scaling relations used to derive the asteroseismology masses do not impose such a prior. The isochrone masses therefore systematically deviate from the asteroseismology masses for the lowest and highest masses.

As metallicity is an additional output to age, distance, and mass, we consider the following ANN outputs

$$y^i = (\log_{10} \tau^i, [M/H]^i, \log_{10} m^i, \mu^i). \quad (19)$$

To build the Bayesian spectroscopic estimator, we first investigate which inputs can be used as predictors of these outputs. Calculating the Spearman Rank Correlation Coefficient and associated  $p$ -value between a set of potential predictors and the desired outputs (Table 3), we find all of them to be significantly correlated with at least one of the outputs. We therefore take the ANN inputs to be

$$x^i = (H^i, (J - K)^i, \varpi^i, \log g^i, T_{\text{eff}}^i, [M/H]^i, [\alpha/M]^i, [C/M]^i, [N/M]^i). \quad (20)$$

As discussed earlier, age and mass are closely related for red giant stars, and therefore they have significant correlations with the same set of inputs. The apparent magnitude combined with parallax tells us the absolute magnitude of the star, which reflects the age, metallicity, and mass of the star. Although the distance of the star is primarily driven by its parallax, it is also modified

**Table 3.** Correlation between each input and output for the ANN. The correlation is measured with the Spearman Rank Correlation Coefficient. If the associated  $p$ -value is less than 0.05, then we take the correlation to be significant, and the cell is coloured grey.

	$\log_{10} \tau$	$[M/H]$	$\log_{10} m$	$\mu$
$H$				
$J - K$				
$\varpi$				
$\log g$				
$T_{\text{eff}}$				
$[M/H]$				
$[\alpha/M]$				
$[C/M]$				
$[N/M]$				

**Table 4.** Performance of different ANN architectures: Columns (1) architecture, (2) total number of parameters in ANN [i.e.  $n_{\text{hid}}(n_{\text{in}} + 2) + n_{\text{out}}$ ], (3) reduced  $\chi^2$  for each output of the testing set (1- $\log_{10} \tau$ , 2- $[M/H]$ , 3- $\log_{10} m$ , and 4- $\mu$ ). The selected model is highlighted in grey.

Architecture	$n_{\theta}$	$\chi_{r,1}^2$	$\chi_{r,2}^2$	$\chi_{r,3}^2$	$\chi_{r,4}^2$
A. $n_{\text{hid}} = 5$	59	0.706	0.213	0.682	0.793
B. $n_{\text{hid}} = 10$	114	0.705	0.184	0.719	0.745
C. $n_{\text{hid}} = 15$	169	0.781	0.197	0.785	0.703

by the absolute magnitude of the star, which is related directly to the apparent magnitude. The mass of the star also sets  $\log g$ , the radius of the star, and therefore  $T_{\text{eff}}$  of the star. The age of a star correlates with  $[M/H]$  and  $[C/N]$  as a result of both stellar evolution and Galactic chemical evolution.

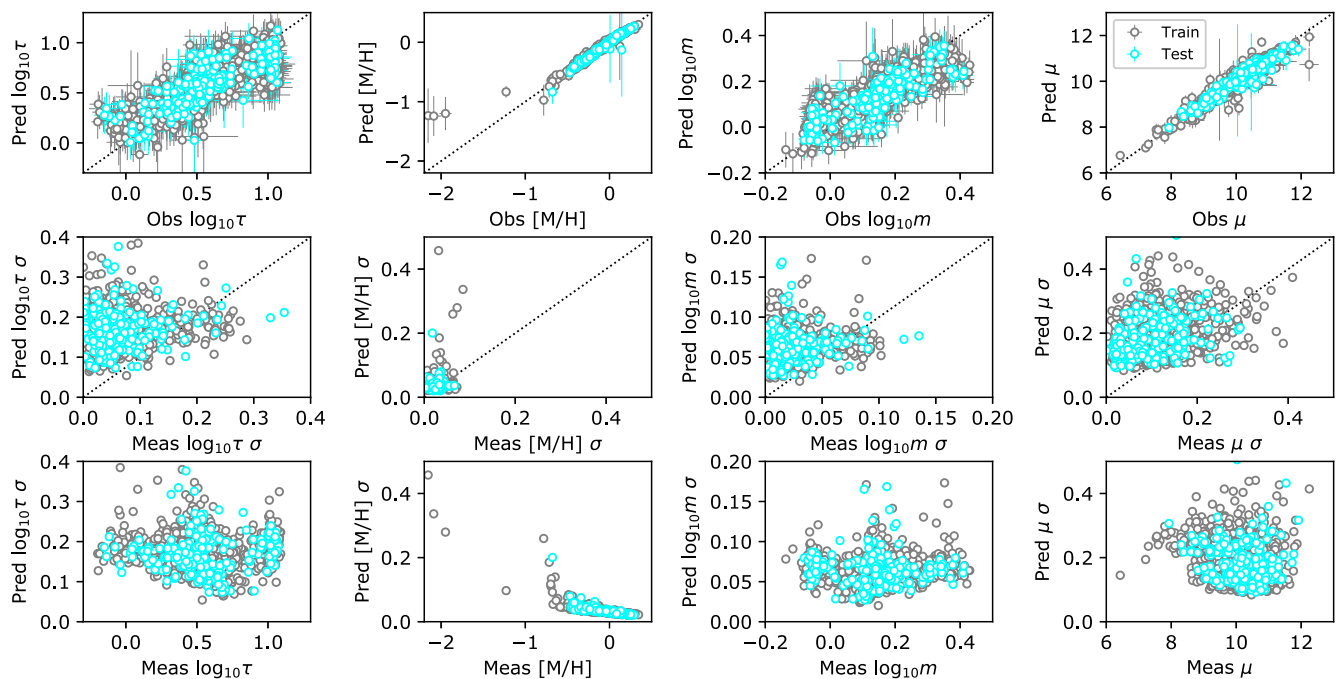
Considering successful outputs from the Bayesian isochrone pipeline, and valid measurements for  $x^i$ , our seismo sample reduces to 1214 stars. We explore three ANN architectures each with a single hidden layer, and either 5, 10, or 15 neurons in the hidden layer. The top part of the graphical model in Fig. 1 summarizes the training of the ANN.

The best ANN architecture is estimated by calculating a reduced  $\chi$ -squared,  $\chi_{r,j}^2$ , for each predicted output. This statistic is formulated to consider both uncertainty in the prediction from the ANN and the uncertainty in the measurements of the outputs

$$\chi_{r,j}^2 = \frac{1}{n_{\text{test}}} \sum_k \frac{(y_{N,j}^k - \tilde{y}_{N,j}^k)^2}{(\sigma_{N,j}^k)^2 + (\tilde{\sigma}_{N,j}^k)^2}, \quad (21)$$

where  $n_{\text{test}}$  is the number of stars in the testing sample. Table 4 shows  $\chi_{r,j}^2$  for each output and for the range of explored architectures. We select model B as the best model though note that  $\chi_r^2$  is always below 1, and so all models perform well.

Fig. 5 compares the means and standard deviations of the output distributions predicted by our favoured model B against those of the measured output distributions for the combined training and testing samples. As expected, the ANN performs almost perfectly in the prediction of  $[M/H]$  and  $\mu$ .  $[M/H]$  is also an input to the ANN and only marginally corrected by the other input variables.  $\mu$  is primarily driven by  $\varpi$ , with spectro-photometric variables contributing more for distant stars. For the three most metal-poor stars however, the ANN systematically overestimates the metallicity, possibly due to a lack of stars at the metal-poor end in the training sample. Age and mass are generally recovered well, but there is significantly more



**Figure 5.** A comparison between predicted and measured means (top panel), predicted and measured standard deviations (middle panel), and predicted standard deviation and measured mean (bottom panel). The predicted output distributions are generated by model B and the measured output distributions are generated by the Bayesian isochrone pipeline for the training (grey) and testing (cyan) samples.

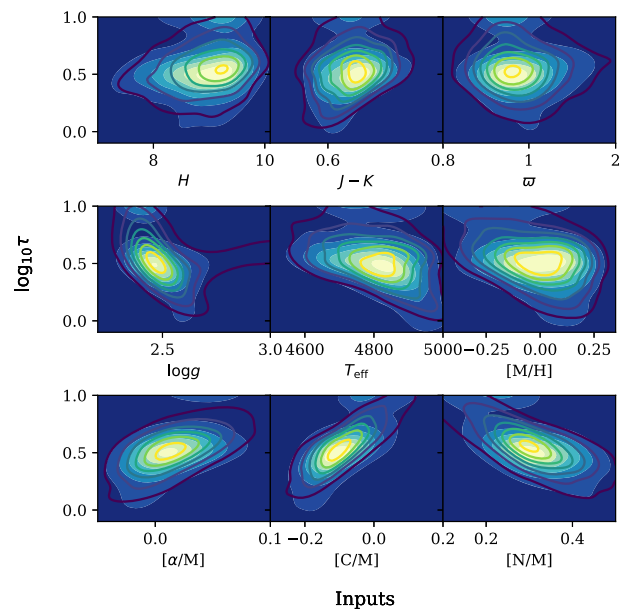
scatter. The lowest ages are systematically overestimated and the highest ages systematically underestimated, as a result of the effect of the age prior on generating isochrone ages and therefore masses.

The range of uncertainties recovered by the ANN is smaller than the uncertainties in the isochrone estimates. For stars with the smallest uncertainties in the isochrone estimates, the ANN estimates are generally higher. This probably reflects a lack of sufficient flexibility in the ANN architecture. However many stars with the highest isochrone uncertainties have lower ANN uncertainties, showing the powerful possibility of using the ANN to ‘denoise’ our data.

The uncertainties on the predicted outputs are flat with the means of the observed outputs, except in the case of  $[M/H]$ , for which the prediction capability gently degrades towards lower metallicities. Beyond a metallicity of  $-1$ , the ANN does not perform well.

The predicted joint probability distributions between a selection of the inputs (parallax, photometry, and spectroscopy) and one of the outputs (age) for the combined training and testing sets for model B are shown compared to the measured joint distributions in Fig. 6. The plot shows that the Bayesian ANN in general does a good job of reproducing the joint distributions, except for the oldest stars where the age prior in the Bayesian isochrone pipeline forces a pile-up towards the oldest and youngest ages.

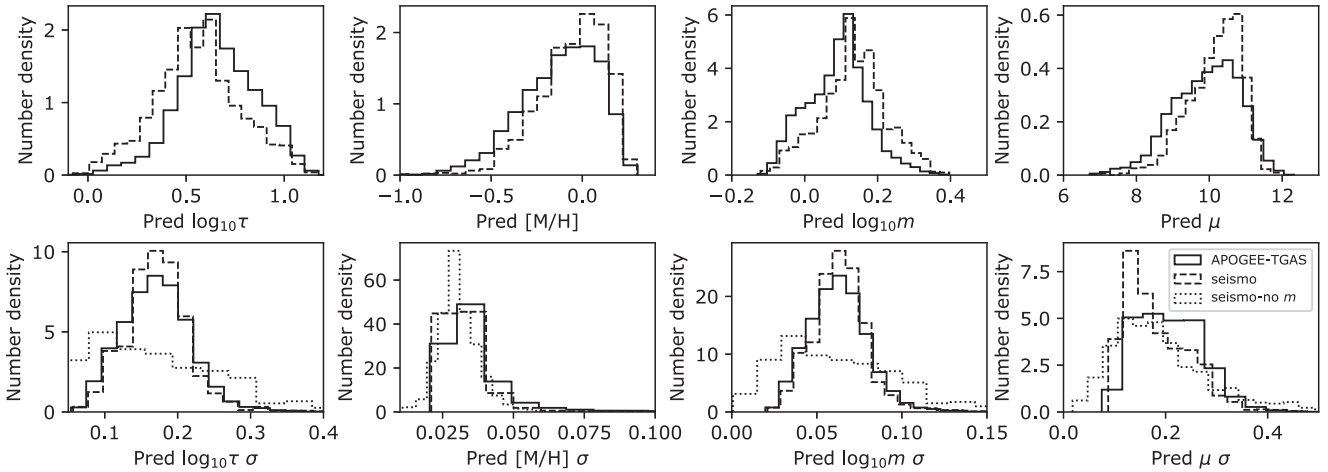
As a final illustration of the power of the Bayesian ANN, we redetermine masses, ages, distances, and metallicities for the training and testing samples, but assume there is no asteroseismology mass information. The uncertainty distributions in the bottom panel of Fig. 7 show that the ANN is not able to produce age and mass estimates with as low uncertainties as the Bayesian isochrone pipeline for some stars. As stated above, this can probably be improved by allowing more flexibility in the ANN. It could also be due to the isochrones giving precise mass estimates that are biased



**Figure 6.** Contour plots of the joint distributions of a selection of the inputs (parallax, photometry, and spectroscopy) and one of the outputs (age) for the combined training and testing sets. The measurements are shown by filled contours and the coloured contour lines present joint distributions of model B.

with respect to the masses that would be predicted by carbon and nitrogen only. As the ANN considers both estimates, the associated uncertainty will be larger. The ANN also performs much better than the Bayesian isochrone pipeline for many stars, partly because of





**Figure 7.** Histograms of the predicted mean and uncertainty in age, metallicity, mass, and distance derived by the ANN for the APOGEE-TGAS sample (solid, black), compared with that evaluated by the ANN for the seismo sample with the ANN (dashed, black). The bottom panel also shows uncertainty distributions when applying the Bayesian isochrone pipeline to the seismo sample but not using the asteroseismology mass information (dotted, black).

the carbon and nitrogen information, and partly due to the capability of the ANN to ‘denoise’ the noisiest isochrone outputs.

#### 4.4 New masses, ages, distances, and metallicities for APOGEE-TGAS stars

Selecting the APOGEE-TGAS stars within the same input domain spanned by the seismo stars, and further cutting to limit the sample to metallicities above  $-1.0$ , reduces the sample from 10 074 to 9403 stars. Applying the ANN takes just over a minute on a single core to calculate posterior predictive distributions for mass, age, distance, and metallicity. This is comparable to the time taken for the Bayesian isochrone pipeline to run on a handful of stars. Fig. 7 shows the distributions in the mean and standard deviations predicted for the APOGEE-TGAS stars. The age distribution peaks at 4–5 Gyr (and mass at around  $1.3 M_{\odot}$ ), the metallicity peaks at solar metallicity, and the distance modulus at 10.5 mag (a distance of 1.3 kpc). The uncertainty distributions show that mass can be estimated to almost always better than 10 per cent and ages to between 10 and 25 per cent. The uncertainties on metallicity and distance modulus are, as expected, small.

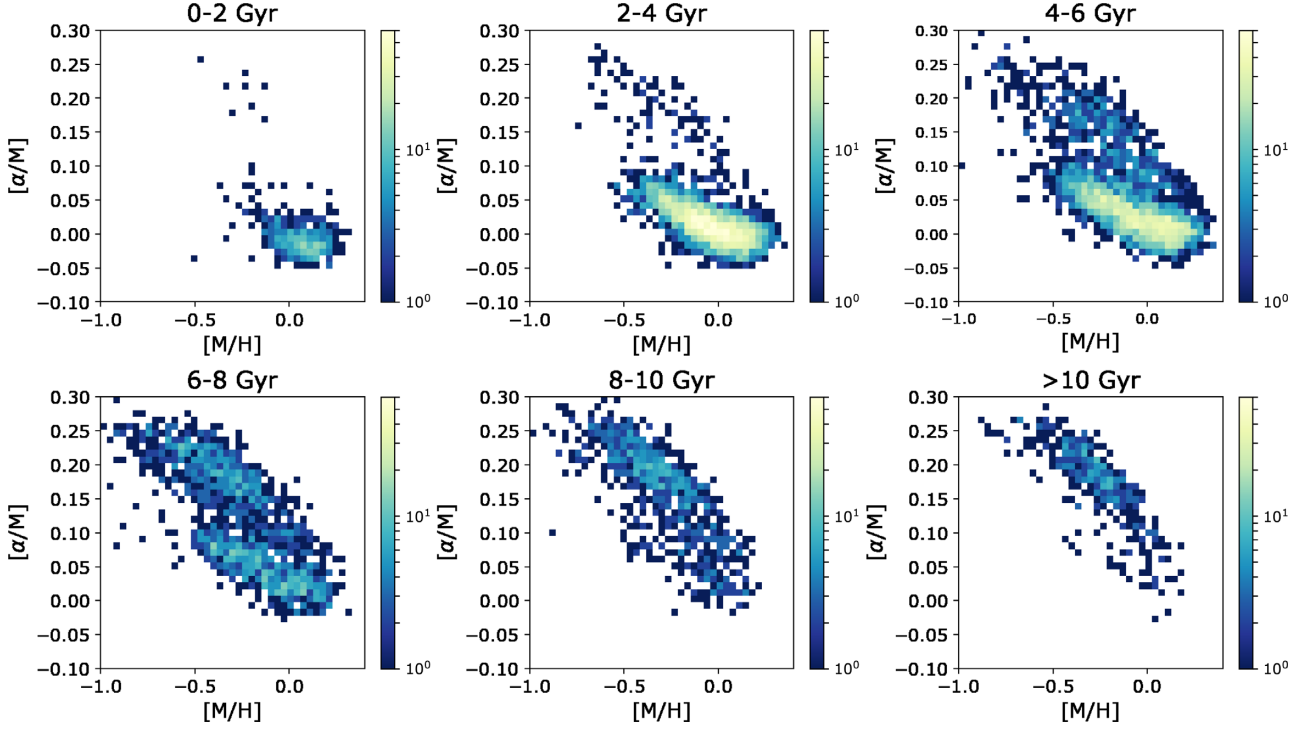
To check whether the ages correlate with abundances as expected, we consider the density of stars in age slices in the  $[M/H]$ – $[\alpha/M]$  plane (Fig. 8). Between 0 and 2 Gyr, most stars are relatively metal rich with a low  $\alpha$  abundance. From 2 to 4 Gyr, the sequence of metal-rich, low- $\alpha$  stars has become more evident, and a sequence of high- $\alpha$ , metal-poor stars appears. The metal-rich, low- $\alpha$  component starts to fade for larger ages and has almost disappeared for stars older than 10 Gyr. The metal-poor, high- $\alpha$  component is most dominant between 6 and 8 Gyr. Two sequences have been discussed in the literature several times (e.g. Hayden et al. 2014). The high- $\alpha$ , metal-poor stars can be considered to be the ‘thick disc’, formed  $\sim 8$  Gyr ago (Haywood et al. 2013), while the low- $\alpha$ , metal-rich stars can be considered to be the ‘thin disc’, which has an extended star formation history. The thin disc is therefore expected to have stars ranging from those born in the present time to those as old as  $\sim 8$  Gyr, while the thick disc stars are all expected to be  $\sim 8$  Gyr old. This is very similar to what we see. It should be noted that uncertainties in age can cause the stars to scatter between the age slices. This could result in thick-disc stars appearing at younger ages and thin-disc stars appearing at older ages.

## 5 CONCLUSIONS AND FUTURE OUTLOOK

We demonstrate the potential for ANNs for precise predictions of masses, ages, and distances for red giants using the Bayesian spectroscopic mass, age, and distance estimator. A Bayesian isochrone pipeline was applied to a training sample of giant stars with apparent magnitudes and colours,  $\log g$  and  $T_{\text{eff}}$  from APOGEE, asteroseismology masses from Kepler, and parallaxes from TGAS to derive posterior distributions for mass, age, distance, and metallicity. We then use this sample, complementing the input data with APOGEE carbon and nitrogen abundance estimates, to train a Bayesian ANN to learn the relationship between the inputs and our desired outputs. The ANN on average reproduces the isochrone estimates for mass, age, distance, and metallicity with similar uncertainties. For training stars with high output uncertainties, the ANN reduces the uncertainty probably as a result of using the extra information in carbon and nitrogen.

The ANN is then used to estimate posterior predictive distributions for masses, ages, distances, and metallicities for the whole APOGEE-TGAS sample. This procedure takes about a minute, and so is significantly quicker than applying the Bayesian isochrone pipeline. The fractional uncertainties on mass are at worst  $\sim 10$  per cent and on age range between 10 and 25 per cent.

The regression presented here is inspired by similar regressions developed by Martig et al. (2016) and Ness et al. (2016) to estimate spectroscopic masses and ages for the same stars in APOGEE. Martig et al. (2016) developed a polynomial regression between carbon abundance, nitrogen abundance,  $\log g$ ,  $T_{\text{eff}}$ ,  $[\text{Fe}/\text{H}]$ , and mass using 1475 giants with asteroseismology mass estimates from Kepler. They predicted stellar masses with fractional uncertainties of about 14 per cent and ages with fractional uncertainties of about 40 per cent. Ness et al. (2016) use the same asteroseismology mass estimates with the *Cannon*, a sophisticated data-driven model that uses the full spectrum. This also estimated ages with fractional uncertainties of about 40 per cent. The ANN developed here uses spectral parameters rather than the full spectrum. However, it employs a Bayesian approach unlike the methods of Martig et al. (2016) and the *Cannon*, deriving posterior distributions for the model parameters and ANN outputs. It is able to estimate age, distance, and metallicity as well as mass, therefore omitting the expensive Bayesian isochrone pipeline.



**Figure 8.** Distribution of APOGEE-TGAS stars in the  $[M/H]$ – $[\alpha/M]$  plane for different age slices.

The work presented here is exploratory and there is significant room for improvement. In light of parallaxes that have just arrived for over a billion stars from *Gaia* DR2 and the second APOKASC catalogue, a much larger training sample can be built. New isochrones (Rodrigues et al. 2017; PARAM) exist that provide  $\Delta\nu$  and  $\nu_{\max}$ , removing the reliance on scaling relations that may have been one of the sources of discrepancies between the asteroseismology masses and the isochrone masses. The age prior in the Bayesian isochrone pipeline may need to be examined in more detail. The ANN architecture may lack sufficient flexibility as it cannot reproduce the small uncertainties achieved for some stars with the Bayesian isochrone pipeline. Finally, the potential for the ANN to also estimate the LOS extinction is a powerful future extension.

To summarize, the method presented here can calculate posterior predictive distributions for masses, ages, distances, and metallicities quickly for a large number of stars. Several spectroscopic surveys now have parallaxes for all stars in their samples. These include APOGEE (DR14; Abolfathi et al. 2018), the Radial Velocity Experiment survey (RAVE DR5; Kunder et al. 2017), the Galactic Archaeology with HERMES survey (GALAH DR2; Buder et al. 2018), *Gaia*-ESO, and The Large sky Area Multi-Object Fiber Spectroscopic Telescope survey (LAMOST DR4; He et al. 2016). These surveys and many more future surveys will greatly benefit from the application of MADE to efficiently calculate masses, ages, distances, and metallicities.

## ACKNOWLEDGEMENTS

The research leading to these results has received funding from the ERC under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 321067 and the ‘Extracting science from Galaxy surveys’ project on the Science and

Technology Facilities Council (STFC) consolidated grant ‘Astrophysics at Oxford’ (grant agreement no. ST/N000919/1).

This project was inspired by the 2016 NYC *Gaia* Sprint, hosted by the Center for Computational Astrophysics at the Simons Foundation in New York City. PD would like to thank Professor Suzanne Aigrain at the University of Oxford for her expertise regarding the construction of graphical models and Bayesian statistics. It is a pleasure to thank the Galactic Dynamics group in Oxford for many fruitful discussions, in particular Professor James Binney for a thorough read of the manuscript. The manuscript also benefitted greatly from many recommendations from the anonymous referee.

We thank GitHub for providing free private repositories for educational use. We are indebted to a number of PYTHON packages such as Astropy (The Astropy Collaboration 2018) and packages in the *scipy* ecosystem (Jones et al. 2001; Hunter 2007; Pérez & Granger 2007; McKinney 2010).

This work has made use of data from the European Space Agency (ESA) mission *Gaia* ([https://www.cosmos.esa.int/web/gaia/iow\\_20180316](https://www.cosmos.esa.int/web/gaia/iow_20180316)), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>).

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-

Performance Computing at the University of Utah. The SDSS web site is [www.sdss.org](http://www.sdss.org).

## REFERENCES

- Abolfathi B. et al., 2018, *ApJS*, 235, 42  
 Alam S. et al., 2015, *ApJS*, 219, 12  
 Anders F. et al., 2017, *A&A*, 597, A30  
 Bedding T. R. et al., 2010, *ApJ*, 713, L176  
 Bensby T. et al., 2013, *A&A*, 549, A147  
 Binney J. et al., 2014, *MNRAS*, 437, 351  
 Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529  
 Bovy J., 2017, *MNRAS*, 470, 1360  
 Bovy J., Rix H.-W., Green G. M., Schlafly E. F., Finkbeiner D. P., 2016, *ApJ*, 818, 130  
 Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, 427, 127  
 Buder S. et al., 2018, *MNRAS*, 478, 4513  
 Burnett B., Binney J., 2010, *MNRAS*, 407, 339  
 Carlin J. L. et al., 2015, *AJ*, 150, 4  
 Casagrande L. et al., 2016, *MNRAS*, 455, 987  
 Chaplin W. J., Miglio A., 2013, *ARA&A*, 51, 353  
 da Silva L. et al., 2006, *A&A*, 458, 609  
 Das P., Binney J., 2016, *MNRAS*, 460, 1725  
 De Ridder J. et al., 2009, *Nature*, 459, 398  
 Dehnen W., McLaughlin D. E., Sachania J., 2006, *MNRAS*, 369, 1688  
 Drimmel R., Cabrera-Lavers A., López-Corredoira M., 2003, *A&A*, 409, 205  
 Gaia Collaboration, 2018, *A&A*, 616, A1  
 Gaia Collaboration, 2016, *A&A*, 595, A1  
 García Pérez A. E. et al., 2016, *AJ*, 151, 144  
 Gelder L. V., Das P., Janssen H., Roels S., 2014, *Simul. Modelling Pract. Theory*, 49, 245  
 Green G. M. et al., 2015, *ApJ*, 810, 25  
 Guenther D. B. et al., 2005, *ApJ*, 635, 547  
 Haas M. R. et al., 2010, *ApJ*, 713, L115  
 Hayden M. R. et al., 2014, *AJ*, 147, 116  
 Haywood M., Di Matteo P., Lehnert M. D., Katz D., Gómez A., 2013, *A&A*, 560, A109  
 He B., Fan D., Cui C., Li S., Li C., Mi L., 2016, in Nuria P. F. L., Keith S., Randall W., eds, ASP Conf. Ser. Vol. 512, Astronomical Data Analysis Software and Systems XXV. Astron. Soc. Pac., San Francisco, p. 153  
 Hekker S. et al., 2011, *MNRAS*, 414, 2594  
 Hoffman M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, 15, 1593  
 Howell S. B. et al., 2014, *PASP*, 126, 398  
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90  
 Høg E. et al., 2000, *A&A*, 355, L27  
 Jofré P., Weiss A., 2011, *A&A*, 533, A59  
 Jofré P., Mädler T., Gilmore G., Casey A. R., Soubiran C., Worley C., 2015, *MNRAS*, 453, 1428  
 Jofré P. et al., 2017, *MNRAS*, 472, 2517  
 Jones E. et al., 2001, SciPy: Open Source Scientific Tools for Python, Available at: <http://www.scipy.org/>  
 Jørgensen B. R., Lindegren L., 2005, *A&A*, 436, 127  
 Kjeldsen H., Bedding T. R., 1995, *A&A*, 293, 87  
 Kroupa P., Tout C. A., Gilmore G., 1993, *MNRAS*, 262, 545  
 Kunder A. et al., 2017, *AJ*, 153, 75  
 Majewski S. R. et al., 2017, *AJ*, 154, 94  
 Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, *A&A*, 453, 635  
 Martig M. et al., 2016, *MNRAS*, 456, 3655  
 Masseron T., Gilmore G., 2015, *MNRAS*, 453, 1855  
 McKinney W., 2010, in van der Walt S., Millman J., eds, Proceedings of the 9th Python in Science Conference. p. 51  
 McMillan P. J. et al., 2017, *MNRAS*, 477, 5279  
 Michalik D., Lindegren L., Hobbs D., 2015, *A&A*, 574, A115  
 Miglio A. et al., 2013, *MNRAS*, 429, 423  
 Mints A., Hekker S., 2018, *A&A*, 618A, 54M  
 Ness M. et al., 2013, *MNRAS*, 430, 836  
 Ness M., Hogg D. W., Rix H.-W., Ho A. Y. Q., Zasowski G., 2015, *ApJ*, 808, 16  
 Ness M., Hogg D. W., Rix H.-W., Martig M., Pinsonneault M. H., Ho A. Y. Q., 2016, *ApJ*, 823, 114  
 Nidever D. L. et al., 2015, *AJ*, 150, 173  
 Pérez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21  
 Pinsonneault M. H. et al., 2014, *ApJS*, 215, 19  
 Pont F., Eyer L., 2004, *MNRAS*, 351, 487  
 Queiroz A. B. A. et al., 2018, *MNRAS*, 476, 2556  
 Robin A. C., Marshall D. J., Schultheis M., Reylé C., 2012, *A&A*, 538, A106  
 Rodrigues T. S. et al., 2014, *MNRAS*, 445, 2758  
 Rodrigues T. S. et al., 2017, *MNRAS*, 467, 1433  
 Sanders J. L., Binney J., 2015, *MNRAS*, 449, 3479  
 Santiago B. X. et al., 2016, *A&A*, 585, A42  
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525  
 Silva Aguirre V. et al., 2015, *MNRAS*, 452, 2127  
 Soderblom D. R., 2010, *ARA&A*, 48, 581  
 Stello D. et al., 2015, *ApJ*, 809, L3  
 The Astropy Collaboration, 2018, *AJ*, 156, 123  
 Ulrich R. K., 1986, *ApJ*, 306, L37  
 Vrad M., Mosser B., Samadi R., 2016, *A&A*, 588, A87  
 Zasowski G. et al., 2017, *AJ*, 154, 198

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