

A variational encoder–decoder approach to precise spectroscopic age estimation for large Galactic surveys

Henry W. Leung ¹★, Jo Bovy ^{1,2}, J. Ted Mackereth ^{1,2,3} and Andrea Miglio ⁴

¹David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4, Canada

²Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4, Canada

³Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St George Street, Toronto, Ontario M5S 3H8, Canada

⁴Dipartimento di Fisica e Astronomia, Università degli Studi di Bologna, Via Gobetti 93/2, I-40129 Bologna, Italy

Accepted 2023 April 24. Received 2023 April 11; in original form 2023 February 10

ABSTRACT

Constraints on the formation and evolution of the Milky Way Galaxy require multidimensional measurements of kinematics, abundances, and ages for a large population of stars. Ages for luminous giants, which can be seen to large distances, are an essential component of studies of the Milky Way, but they are traditionally very difficult to estimate precisely for a large data set and often require careful analysis on a star-by-star basis in asteroseismology. Because spectra are easier to obtain for large samples, being able to determine precise ages from spectra allows for large age samples to be constructed, but spectroscopic ages are often imprecise and contaminated by abundance correlations. Here we present an application of a variational encoder–decoder on cross-domain astronomical data to solve these issues. The model is trained on pairs of observations from APOGEE and *Kepler* of the same star in order to reduce the dimensionality of the APOGEE spectra in a latent space while removing abundance information. The low dimensional latent representation of these spectra can then be trained to predict age with just ~ 1000 precise seismic ages. We demonstrate that this model produces more precise spectroscopic ages (~ 22 per cent overall, ~ 11 per cent for red-clump stars) than previous data-driven spectroscopic ages while being less contaminated by abundance information (in particular, our ages do not depend on $[\alpha/M]$). We create a public age catalogue for the APOGEE DR17 data set and use it to map the age distribution and the age– $[\text{Fe}/\text{H}]$ – $[\alpha/M]$ distribution across the radial range of the Galactic disc.

Key words: methods: data analysis – techniques: spectroscopic – stars: fundamental parameters.

1 INTRODUCTION

Stars provide an important window into our Galaxy’s evolutionary history as all major events that occurred in the past leave imprints in the chemical abundances and kinematics of stars (Freeman & Bland-Hawthorn 2002). To improve our understanding of the formation history of the Milky Way and explore the evolution and chemodynamical structure of the Galaxy as a whole, we need to measure abundances and kinematics of stars as functions of age for stellar samples covering a large volume of our Galaxy from the bulge and the disc to the stellar halo (Rix & Bovy 2013; Bland-Hawthorn & Gerhard 2016). Age, chemical abundances, and kinematics are interconnected in complex ways (e.g. Edvardsson et al. 1993; Haywood et al. 2013; Bovy et al. 2019; Mackereth et al. 2019; Ness et al. 2019) and information on their distributions far away from the solar neighbourhood remains scant. To observe a large volume of stars for galactic archaeology purposes, low-mass giants are of particular interest, because they are common throughout the Galaxy, live relatively long and stable lives, and they are intrinsically luminous allowing them to be observed to large distances even in regions with high extinction such as the Galactic bulge.

Modern spectroscopic surveys such as SDSS-IV’s APOGEE (Blanton et al. 2017; Majewski et al. 2017), GALAH (De Silva et al. 2015), the ongoing SDSS-V’s Milky Way Mapper (MWM; Kollmeier et al. 2017), and *Gaia* (Gaia Collaboration 2016) provide accurate measurements of basic stellar parameters like T_{eff} , elemental abundances, and kinematics. These surveys, however, do not directly provide accurate stellar age measurements for low-mass giants, because stellar ages are not a directly observable quantity. Unlike subgiants, for which age can be measured fairly accurately with basic stellar parameters and using isochrones (Haywood et al. 2013; Xiang et al. 2017; Xiang & Rix 2022), stellar ages for giants are intrinsically difficult to measure, because giant evolutionary tracks are crowded together compared to subgiant isochrones, age and metallicity are to some extent degenerate observationally, and stellar evolutionary models have large uncertainties. While age correlates with kinematics and abundances, individual stellar ages cannot simply be inferred accurately from stellar kinematics (e.g. Beane, Ness & Bedell 2018), abundance, or kinematics-abundance alone. Even if ages could be inferred in this way, to investigate the relation between age, abundance, and kinematics, we cannot rely on pre-determined relations in this space.

Ages for giant stars can be obtained from determinations of their mass, because the age of a giant is almost directly given by its mass-dependent main-sequence (MS) lifetime. Because of the steep dependence of the MS lifetime on mass ($\tau \propto M^{-3.5}$),

* E-mail: henrysky.leung@mail.utoronto.ca

small uncertainties in mass determinations get amplified strongly and giant ages typically have large uncertainties. Spectroscopically, masses for giants can be determined using proxies such as the $[C/N]$ ratio (e.g. Masseron & Gilmore 2015; Martig et al. 2016), which is partially set by the mass-dependent dredge-up process. Alternatively, accurate masses for giants can be obtained with careful analysis on a star-by-star basis (Appourchaux 2020) using asteroseismic observations, which can determine masses from the properties of stochastically driven oscillations that can be observed by space telescopes such as *CoRoT* (Auvergne et al. 2009), *Kepler* (Borucki et al. 2010), and *TESS* (Ricker et al. 2015). Using stellar masses in combination with spectroscopically determined parameters like T_{eff} and $[Fe/H]$, we can derive stellar age using stellar models (e.g. Rodrigues et al. 2014). The APOKASC project (Pinsonneault et al. 2014, 2018) is an example of this, combining *Kepler* seismic observation with APOGEE spectroscopic parameters to derive stellar ages with uncertainties of ≈ 30 per cent. APOKASC has opened up the possibility of doing galactic archaeology with ages for thousands of giants, but it is limited to the relatively small *Kepler* field. While *TESS*' sky coverage is much bigger, its shorter observation span means that it is difficult to use *TESS* to determine asteroseismic ages for high-luminosity giants with their long oscillation time-scales.

Spectroscopic surveys such as APOGEE or the ongoing SDSS-V Milky Way Mapper (MWM) have all-sky coverage and obtain spectra for ≈ 1 million (for APOGEE) to 5 million (MWM) stars. Recent advances in machine learning methodology and algorithms allow us to do transfer learning, using which we can transfer knowledge obtained in one domain to another by training on pairs of data from the two domains. In the context of ages, this means that we can transfer age knowledge from the asteroseismic realm to the spectroscopic realm by training on stars with observations in both realms, without requiring any prior knowledge on how to map stellar spectra to ages. The APOKASC catalogue provides such a training set and it has been used to determine spectroscopic ages using APOGEE (e.g. Ness et al. 2016; Mackereth et al. 2019; Ciucă et al. 2021), LAMOST (e.g. Xiang et al. 2019), and other surveys. This has allowed for millions of data-driven spectroscopic ages to be determined. As Xiang et al. (2019) demonstrates, these spectroscopic ages provide great scientific value even if their precision falls short of what is ideally required.

Spectroscopic ages obtained through transfer-learning from asteroseismic data currently suffer from a series of limitations. First, the overlap between spectroscopic surveys and asteroseismic surveys is small with $\mathcal{O}(10^4)$ stars. This is a relatively small amount of data to train modern machine-learning methods (e.g. neural networks; Mackereth et al. 2019; Ciucă et al. 2021). However, many more pairs of, e.g. APOGEE/*Kepler* observations exist than we have asteroseismic ages for and these could in principle be used to improve the information transfer between the domains. Secondly, spectra contain information on abundances that are highly correlated with age (e.g. the alpha enhancement $[\alpha/Fe]$) and current methods provide no guarantee that the spectroscopically determined age is not solely or largely coming from age-abundance correlations present in the training sample rather than true spectral age information. This makes any inference of the age-abundance-kinematics correlations using current spectroscopic ages suspect.

In this paper, we present a novel transfer-learning method for determining spectroscopic ages for giants that solves these issues and furthermore allows for spectroscopic ages to be determined in the future using other small, high-quality samples of stellar

ages. We do this by splitting the spectroscopic-age determination task in two parts: (i) extracting the age information from stellar spectra while discarding abundance information and (ii) mapping the extracted age information to age using a small sample of accurate stellar ages. We achieve (i) by using a variational encoder-decoder neural network to map high-resolution spectra to asteroseismic power spectra using a small latent-space bottleneck connecting the two. Because asteroseismic power spectra contain information on mass and radius but not abundance, this effectively extracts age information from the stellar spectra while discarding abundance information. In step (ii) we then train a simpler machine-learning method with fewer free parameters to map the latent space to age using the small sample of high-quality ages. Predictions for new spectroscopic ages are obtained by encoding the spectrum in the latent space and then mapping its location in latent space to age. We then apply this method to the APOGEE data and map the age distribution and age-abundance correlations across the Galactic disc.

This paper is organized as follows. We have given an overview of the relevant deep-learning methodology in Section 2. Section 3 describes the actual machine-learning methods that we use: the encoder-decoder used in step (i) of the algorithm in Section 3.1 and a modified version of the random forest method used in step (ii) in Section 3.2. We then discuss the data from APOGEE, *Kepler*, and APOKASC that we use in Section 4. We give details on the training and validation steps of our algorithm in Section 5 and then describe the results in Section 6: in Section 6.1 we show how well we can reconstruct asteroseismic power spectra from high-resolution spectra using the encoder-decoder network, we discuss the latent space representation of the spectra in Section 6.2, we discuss the derived seismic parameters, ages, and evolutionary state classifications in Section 6.3. The detail of applying our method to generate APOGEE age catalogue for APOGEE DR17 is given in Section 7.1 and spatial age-abundance trends in the Galactic disc are shown in Section 7.2. We discuss the implications of our results and possible future applications in Section 8 and then conclude in Section 9.

2 DEEP LEARNING METHODOLOGY

In this Section we provide an overview of the deep-learning methodology that we use in this work: (variational) auto-encoders in Section 2.1, encoder-decoder networks in Section 2.2, and we end with a discussion of the rationale behind why we choose to use an encoder-decoder in this work in Section 2.3.

2.1 Variational auto-encoders

Deep learning using artificial neural networks is a versatile and flexible machine-learning method and it has been applied to problems in both supervised and unsupervised learning with data ranging from 2D images, voice recordings, video, and sequences of words (Lecun, Bengio & Hinton 2015). Auto-encoders are a specific type of neural network primarily developed to learn efficient, low-dimensional representations of input data that allow the input to be faithfully reproduced. As such, auto-encoders are essentially a non-linear version of Principal Component Analysis (PCA), that is an auto-encoder network without any non-linearity (e.g. not using non-linear activation functions such as Rectified linear units; Nair & Hinton 2010) using a squared-error loss function are equivalent to traditional PCA (Kramer 1991). The basic idea of using an

auto-encoder is having a network that takes an input image, compresses it to a low-dimensional middle layer (the *latent space*) in a sequence of layers (the *encoder*), and finally decompresses this low-dimensional representation to reconstruct the input image (the *decoder*). The low-dimensional latent space acts as a bottleneck that restricts the flow of information through the network and, thus, when trained well forms an efficient representation of the input data.

Auto-encoders have been used in astronomy in the past for applications such as denoising (training on pairs of noisy, near-noiseless, or augmented data; e.g. Frontera-Pons et al. 2017; Sedaghat & Mahabal 2018; Gheller & Vazza 2022), dimensionality reduction (e.g. Portillo et al. 2020), and representation learning (e.g. Cheng et al. 2020 for identifying strong lenses and de Mijolla et al. 2021 for chemical tagging). There have, to our knowledge, not been any astronomical applications of the type of encoder–decoder network with cross-domain data that we discuss in Section 2.2 below.

Variational auto-encoders are a type of auto-encoder that incorporates variational inference into the model (Kingma & Welling 2013). Variational inference is a maximum-likelihood estimation (MLE) method for situations in which the probability density is very complex, which in our case is the distribution of the latent variables that generate the output data. The addition of variational inference in practice acts as a regularization that forces the latent space distribution to follow a given statistical distribution. This is often a Gaussian distribution, as any distribution can be described as a set of normally distributed variables mapped via a complex function like a neural network. Moreover, forcing the latent space distribution to follow a known statistical distribution also makes generating new samples easier, because we can draw random samples from the latent space distribution and decode them to construction outputs. A detailed review of variational auto-encoders can be found in Doersch (2016).

Overall, a variational auto-encoder has the following major components:

The *encoder*: The encoder is a discriminative model that takes input data and compresses it to a latent space of much lower dimensionality. In variational auto-encoders, the encoder predicts the means and variances (in practice the encoder predicts logarithmic variance instead for numerical stability) of each latent-space parameter and then samples from a normal distribution with these parameters to obtain the final efficient representation.

The *latent Space*: The latent space is the layer that is populated by the latent variables. The latent space is fully unconstrained prior to training and the entire latent-variable representation is learnt during training. However, we do need to set the dimension of the latent space, analogous to setting the number of principal components to use in PCA.

The *decoder*: The decoder is a generative model that takes the latent variables and generates the (generally much higher dimensional) output. This is done by starting from the low-dimensional representation and building it out to the high-dimensional output through a sequence of steps (layers). In many ways, the decoder does the opposite of the encoder.

Variational auto-encoders are trained by minimizing an objective function that is composed of the sum of two loss terms. The first of these is the reconstruction loss, i.e. a measure of how well the model predicts the output. The second is a regularization term that shapes the distribution in the latent space, i.e. forcing the latent space to follow a certain distribution. We have adapted the mean squared error (MSE) as the reconstruction loss and the Kullback–Leibler (KL) divergence as the latent-space regularization loss to force the latent

space to follow a Gaussian distribution. The MSE reconstruction loss J_{MSE} is given by

$$J_{\text{MSE}}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N} \sum_{i=1}^N \mathbf{w}_i (\hat{y}_i - y_i)^2, \quad (1)$$

where \hat{y}_i is the predicted output, y_i is the true output, and \mathbf{w}_i is a weight for each pixel i in the output. The weight \mathbf{w} will be an array of ones if no pixel-level weighting applied. The KL-divergence regularization loss J_{KL} is

$$J_{\text{KL}}(\mu, \log \sigma^2) = \frac{1}{2} \left[- \sum_i (\log \sigma_i^2 + 1) + \sum_i \exp(\log \sigma_i^2) + \sum_i \mu_i^2 \right], \quad (2)$$

where $(\mu_i, \log \sigma_i^2)$ are the mean and logarithmic variance in each latent-space dimension for data point i from the encoder.

2.2 Encoder–decoder networks

An encoder–decoder network is very similar to an auto-encoder in terms of architecture and the same applies to variational encoder–decoders with respect to variational auto-encoders. They both consists of three major components, an encoder, a decoder, and a latent space between the encoder and the decoder as we have discussed in the previous subsection. The major difference between the two is in the input and output data used by the model. An auto-encoder uses data in the same domain (e.g. pairs of images of same objects) for the input and output node, while an encoder–decoder network employs data in different domains (e.g. image input and text output). Encoder–decoders are commonly used in neural machine translation (e.g. Cho et al. 2014; Vaswani et al. 2017), because different languages (hence different domains) are just different ways to express the same abstract ideas.

In our application of an encoder–decoder in this paper, we have two domains of data, which are APOGEE high-resolution spectra as inputs on the one hand and power spectral density (PSD) derived from *Kepler* light curves as outputs on the other hand. The goal of using an encoder–decoder is to extract the information about the *Kepler* PSDs that is contained in the APOGEE spectra, that is, the ‘asteroseismic’ information that is present in the spectroscopic data. We do not necessarily extract the true asteroseismic information from APOGEE spectra, but simply any spectral information that is correlated with the asteroseismic information. However, because *Kepler* PSDs do not contain abundance information, we emphatically do not extract abundance information at the same time (as we will explicitly demonstrate below). To be able to successfully construct *Kepler* PSDs, we expect that the latent space must contain information on the asteroseismic parameters ν_{max} and $\Delta\nu$ from which the mass can be derived using scaling relations (Brown et al. 1991; Kjeldsen & Bedding 1995) as

$$\frac{M}{M_{\odot}} \simeq \left(\frac{\nu_{\text{max}}}{\nu_{\text{max}, \odot}} \right)^3 \left(\frac{\Delta\nu}{\Delta\nu_{\odot}} \right)^{-4} \left(\frac{T_{\text{eff}}}{T_{\text{eff}, \odot}} \right)^{3/2}, \quad (3)$$

where $\nu_{\text{max}, \odot}$ and $\Delta\nu_{\odot}$ are the asteroseismic parameters of the Sun. Thus, because spectra also contain information about T_{eff} , we expect the latent space to contain information on the stellar mass, from which the age can be derived. Note that we do not use the scaling relations anywhere in our methodology, but we simply use them here to argue that the encoder–decoder should be able to extract mass information from the high-resolution spectra if it can successfully reconstruct *Kepler* PSDs from stellar spectra.

2.3 Rationale of using an encoder–decoder network

In this work, we use an encoder–decoder to map high-resolution infrared spectra from APOGEE to PSDs derived from *Kepler* light curves. This has the following advantages:

(i) *Ability to train on all APOGEE/Kepler pairs*: When training the encoder–decoder, we do not need labels (i.e. mass or age). Thus, all that we require are stars that have both APOGEE spectra and *Kepler* light curves and we can use *all* overlapping observations between the APOGEE spectra and *Kepler* (or, in the future, *TESS*) light curves. There are many more APOGEE–*Kepler* pairs than there are reliable asteroseismic measurements of mass and age for those pairs, so this allows for a large expansion of the available training data.

(ii) *Information extraction*: Current data-driven spectroscopic ages likely rely on proxies like the [C/N] ratio (Ness et al. 2016; Mackereth et al. 2019) and they may even use information such as that contained in $[\alpha/\text{Fe}]$. However, the [C/N] ratio is not only affected by mass-dependent mixing processes, but instead at least in part depends on galactic chemical evolution and this adversely affects age predictions. However, we have millions of spectra from surveys such as APOGEE and GALAH, so being able to determine ages without relying on abundance information beside [Fe/H] is important. By using an encoder–decoder, we force the method to extract only information necessary to predict the *Kepler* PSD without extracting unnecessary abundance information. Information such as [C/N] is not directly available in PSD, but its component that is due to stellar mixing may still emerge in the latent space through its correlation with mass.

(iii) *Application to large spectroscopic data sets*: Once trained, we can discard the decoder and apply the encoder plus the latent space to all available spectroscopic data, which cover the entire sky.

(iv) *Simplicity*: The dimensionality of the latent space is orders of magnitude smaller than that of the spectra and the PSD. Thus, we can train simpler regression models to predict the age from the latent space. Simple models can be trained using only a handful of very precise ages.

3 MODELS

In this Section we provide details on the encoder–decoder network that we use to map APOGEE spectra to *Kepler* PSDs and on the regressor that we employ to map the latent space to age. A schematic overview of our methodology is shown in Fig. 1. The encoder–decoder part is implemented as `ApokascEncoderDecoder()` using `tensorflow` (Abadi et al. 2015) in the `astroNN` package (Leung & Bovy 2019a).¹

3.1 Encoder–decoder model

The encoder in our model consists of a convolutional neural network with two convolutional layers, a max-pooling layer, and a dense layer that outputs the mean and variance for each dimension of the latent space starting with APOGEE stellar spectra. We use the ReLU activation function throughout the encoder except in the final layer where we employ the `tanh` activation function to prevent predicting extreme values in the latent space and improve training stability. The latent space is simply a normal distribution sampler that samples latent variables from the output of the encoder. The decoder is another convolutional neural network, but with transposed

convolutional layers to reconstruct the *Kepler* PSD as opposed to the regular convolutional layers used in the encoder. We again use the ReLU activation function throughout the decoder, except in the final layer where no activation is applied to the outputs in order not to limit the range of the reconstruction output.

The conventional convolutional layer (LeCun et al. 1989) used in convolutional neural networks excels in pattern recognition. It works by learning convolutional filters that recognize certain patterns, hence the output values are calculated as the dot product between filter and the input to the layer by moving the filter kernel across the input, which is the usual convolution operation. For the transposed convolutional layer (also known as deconvolution), the input value in the layer determines the filter values that will be written to the output. In other words, the input determines the weight of the filters as opposed to learning the weights as in the usual convolutional layers.

Hyperparameters such as the number of neurons, the dimension of the latent space, and the optimizer’s learning rate are optimized by hyperparameter search. For parameters like the latent space dimension, we have an educated guess of what value we want to use. We know at least three parameters are important to the reconstruction of the PSD: ν_{max} , $\Delta\nu$, as well as the evolutionary state. From this minimum number of dimensions, the latent space dimension is increased until there is no increase in the performance for the validation set. This gives a 5D latent space.

3.2 Probabilistic random forest model

To map the latent space to stellar age, we employ a version of the random forest implementation from `scikit-learn` (Pedregosa et al. 2012). In `scikit-learn`, random forests are built with individual trees in the forests in which each tree is built from bootstrapped samples from the training set; the prediction is simply the mean of the value returned by each tree in the forest. We have modified the implementation such that during training, we sample the input and output data with their corresponding uncertainty on top of the bootstrap sampling where the input data uncertainties are taken directly from the encoder (i.e. variance predicted by the encoder) while the output data uncertainties are given in training set. Each star is then weighted by the age uncertainty of the training set by $\sqrt{1/\sigma_{\text{age}}^2}$. At testing time, the predicted age is the mean of all of the trees and we interpret the standard deviation of the trees as the age uncertainty.

4 DATA SETS AND DATA REDUCTION

We use spectroscopic data from APOGEE, light curves and derived PSDs from *Kepler*, and ages for stars in the APOKASC catalogue derived by Miglio et al. (2021b). We discuss these different data sets in this section and describe the relevant data reduction processes for each data set.

4.1 APOGEE

We use high-resolution spectra from the APO Galactic Evolution Experiment (APOGEE; Blanton et al. 2017; Majewski et al. 2017), specifically from its seventeenth data release (DR17; Abdurro’uf et al. 2022). APOGEE DR17 is a high signal-to-noise (>100 per pixel typically), high resolution ($R \sim 22\,000$) panoptic spectroscopic survey in the near infrared *H*-band wavelength region of 1.5–1.7 μm . APOGEE spectra are the input of our neural network model and we

¹<https://github.com/henrysky/astroNN>

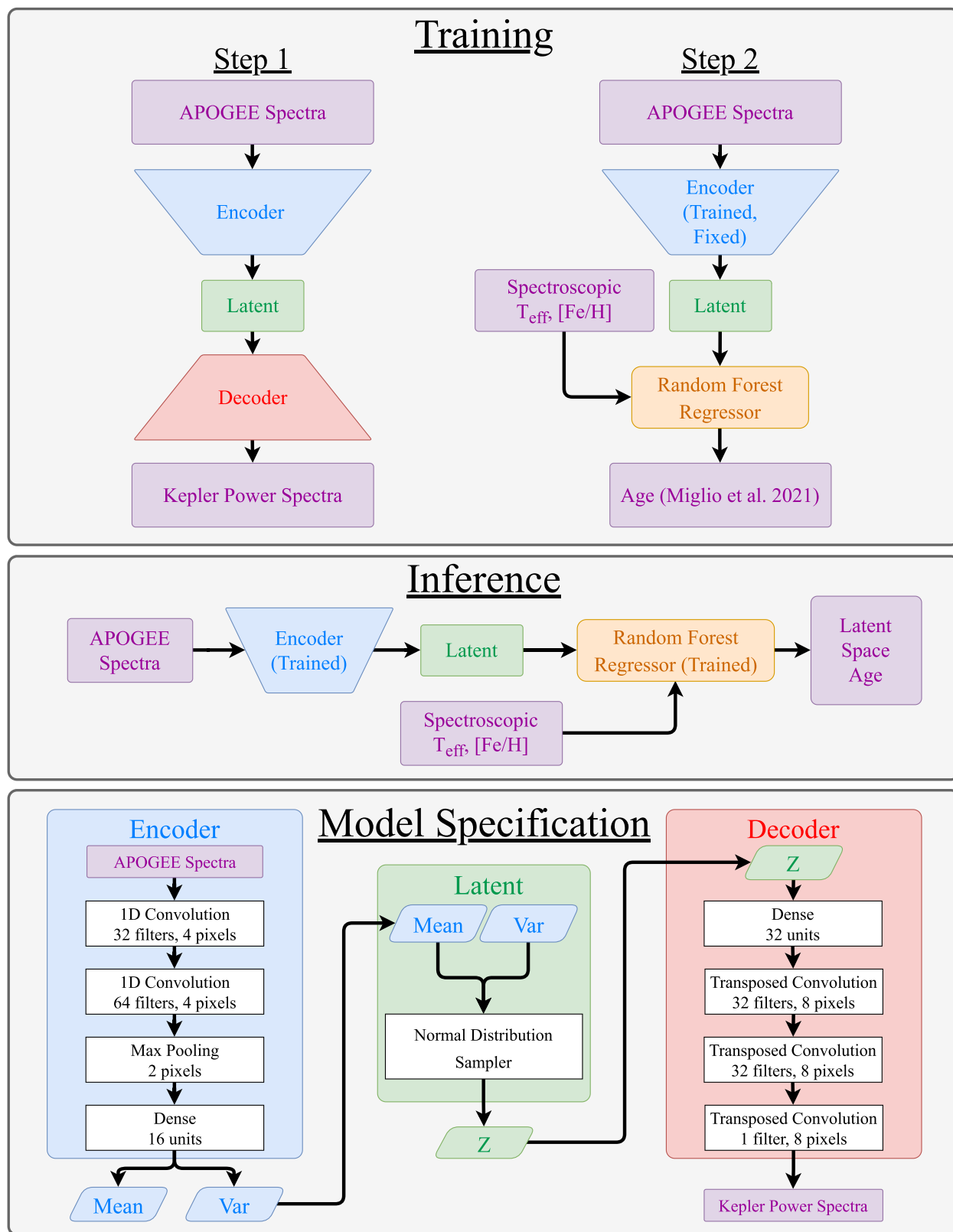


Figure 1. Schematics of the methods and models that we use to obtain spectroscopic ages from a latent-space representation of APOGEE spectra. The top panels show the training and inference phase. In step 1 of the training phase, the encoder and decoder act as a single model that takes APOGEE spectra as inputs with the objective to reconstruct *Kepler* PSDs as outputs. In step 2, we train a random forest regressor to determine age from the latent space of the encoder–decoder along with T_{eff} and [Fe/H]. During inference, we discard the decoder and use the encoder to compress APOGEE spectra to the latent space, and then on through the trained random forest regressor to get the age prediction. The bottom panel specifies the detailed architectures of the encoder and the decoder. Both the encoder and the decoder are essentially convolutional neural networks.

employ the same continuum normalization procedure as in Leung & Bovy (2019a). In addition to APOGEE spectra, we use stellar parameters and elemental abundances derived by the APOGEE Stellar Parameter and Chemical Abundances Pipeline (ASPCAP; García Pérez et al. 2016). These are not used by the encoder–decoder part of our model, but we add some of them to the input data given to the random-forest regressor when mapping the latent space to age. Specifically, we add the effective temperature T_{eff} and the metallicity $[\text{Fe}/\text{H}]$, because stellar models require these to determine stellar ages from asteroseismic observations. We use the ASPCAP parameters rather than external data-driven stellar parameters and elemental abundances such as those from Leung & Bovy (2019a) or Ting et al. (2019) that are proven to be robust to lower signal-to-noise ratio spectra, to be consistent with the stellar parameters used to derive the training ages.

4.2 Kepler

We use light curves for giant stars obtained by the *Kepler* telescope (Borucki et al. 2010) in the original Kepler field near the Cygnus and Lyra constellations, to compute the PSDs that are the output node of our encoder–decoder model. To download, manage, and manipulate *Kepler* data, we make use of the `lightkurve` Python package (Lightkurve Collaboration 2018). We compute the PSD from the observed light curve using the Lomb–Scargle periodogram (Lomb 1976; Scargle 1982) as implemented in *astropy* (Astropy Collaboration 2022). We adopt a minimum frequency of 2 μHz and maximum frequency of 270 μHz (roughly corresponding to half of the inverse of the *Kepler* 30 min cadence) with a frequency spacing of 0.009 μHz (roughly corresponding to the inverse of *Kepler*'s 3.5 yr baseline for our sample). These values are similar to those used in the Kepler Light Curves Optimized For Asteroseismology (KEPSEISMIC; García et al. 2011) work. We divide the PSD by a low-pass background filter largely corresponding to the noise background coming from stellar activity, granulation, and photon noise; the low-pass filter consist of a moving median filter with a width of 0.01 in logarithmic μHz frequency space.

In this work, we employ the PSD instead of the autocorrelation function (ACF) used in works like McQuillan, Aigrain & Mazeh (2013) and Angus et al. (2018) for asteroseismic data analysis as well and by Ness et al. (2018) for determining data-driven stellar parameters. The reason why the latter work prefers to use the ACF over the PSD is the smooth gradient with respect to parameters such as age of each pixel in the ACF, which is essential for models that require smooth gradients. Although our decoder is a generative model for the PSD, we do not use true stellar labels to generate the PSD but instead rely on latent variables. Moreover, Blancato et al. (2020) demonstrated that using the PSD rather than the ACF in a discriminative model improves performance. In this paper, using the PSD improves interpretability of our model compared to the ACF as we can directly check whether the decoder reconstructs the expected asteroseismic modes in the PSD which will be shown and discussed in Section 6.1.

The PSDs constructed using the method above have a large number of pixels and to reconstruct them with a neural network would require a large number of parameters that would likely be overfit due to the limited number of APOGEE–*Kepler* pairs available for training the encoder–decoder. Moreover, it is unlikely that the stellar spectra contain such extremely precise asteroseismic information that they can predict the PSD at 0.009 μHz resolution. The resolution of interest for our application is proportional to the large frequency separation $\Delta\nu$ – the average frequency spacing between modes of

adjacent radial order of a given angular degree l . In particular, we want to resolve mass-dependent variations in $\Delta\nu$ at a given ν_{max} . These are a few per cent of the $\Delta\nu$ expected for a typical star at a given ν_{max} (see equation 7 below). Thus, we rebin the PSD using a logarithmic spacing chosen such that $\Delta\nu$ is resolved by about 50 pixels at any given ν_{max} (i.e. on average there are 50 pixels between two oscillation modes of consecutive overtones of the same angular degree for stars with any ν_{max}). This gives 2092 pixels in each power spectrum.

The frequency values f_i of our final PSDs (that is, the frequency f_i in μHz represented at every pixel i) are approximately given by the expression

$$f_i = 77.35 \left(\frac{i}{2092} \right)^4 + 102.55 \left(\frac{i}{2092} \right)^3 + 68.96 \left(\frac{i}{2092} \right)^2 + 18.47 \left(\frac{i}{2092} \right) + 2.00. \quad (4)$$

The exact solution that we use is computed using the following recurrence relation

$$f_{i+1} = \frac{0.263 \mu\text{Hz} (f_i^{0.772})}{50} + f_i, \quad (5)$$

where f_0 is 2 μHz . The maximum deviation between the approximation and full solution is 0.008 μHz while median absolute deviation is 0.002 μHz .

4.3 APOKASC and ages

For asteroseismic data, we use data from the APOGEE–*Kepler* Asteroseismology Science Consortium (APOKASC; Pinsonneault et al. 2014) catalogue, the Yu et al. (2018) *Kepler* red-giant seismology catalogue, and ages from Miglio et al. (2021b). The APOKASC catalogue consists stars observed by both the APOGEE survey and the *Kepler* telescope. We have adopted the second data release of the catalogue (APOKASC-2; Pinsonneault et al. 2018) based on APOGEE DR14 (Abolfathi et al. 2018). The APOKASC-2 catalogue consist of 6676 giant stars, 85 per cent of which have *Kepler* light curves with time baseline of >3.5 yr. APOKASC first estimates global seismic parameters ν_{max} and $\Delta\nu$ from power spectra generated from the *Kepler* light curves using multiple pipelines that suffer from different systematics and apply different constraints on the model parameters. To expand the sample size of APOGEE–*Kepler* pairs, we cross-match the Yu et al. (2018) catalogue with APOGEE DR17 to obtain 10 526 stars. We use the union between the APOKASC-2 and Yu et al. (2018) catalogue to obtain 10 672 stars with solar-like oscillations.

The global parameters ν_{max} and $\Delta\nu$ are highly correlated and we adopt the following relation that describes this correction when relevant

$$\Delta\nu = 0.263 \mu\text{Hz} \times (\nu_{\text{max}} / \mu\text{Hz})^{0.772}, \quad (6)$$

which is a relation that is empirically determined using results from the SYD asteroseismic pipeline (Huber et al. 2009, 2010). We parametrize deviations from this relation for individual stars using what we call ‘excess $\Delta\nu$ ’, defined as

$$\text{Excess } \Delta\nu = \Delta\nu / (0.263 \mu\text{Hz} [\nu_{\text{max}} / \mu\text{Hz}]^{0.772}), \quad (7)$$

where ν_{max} and $\Delta\nu$ are the measured values for a given star. Excess $\Delta\nu$ can be shown to be sensitive to stellar mass using the asteroseismic scaling relations (a review on asteroseismology and seismic scaling relations can be found in García & Ballot 2019).

To train the latent-space-to-age part of our model, we adopt training ages from Miglio et al. (2021b). Specifically, we use a

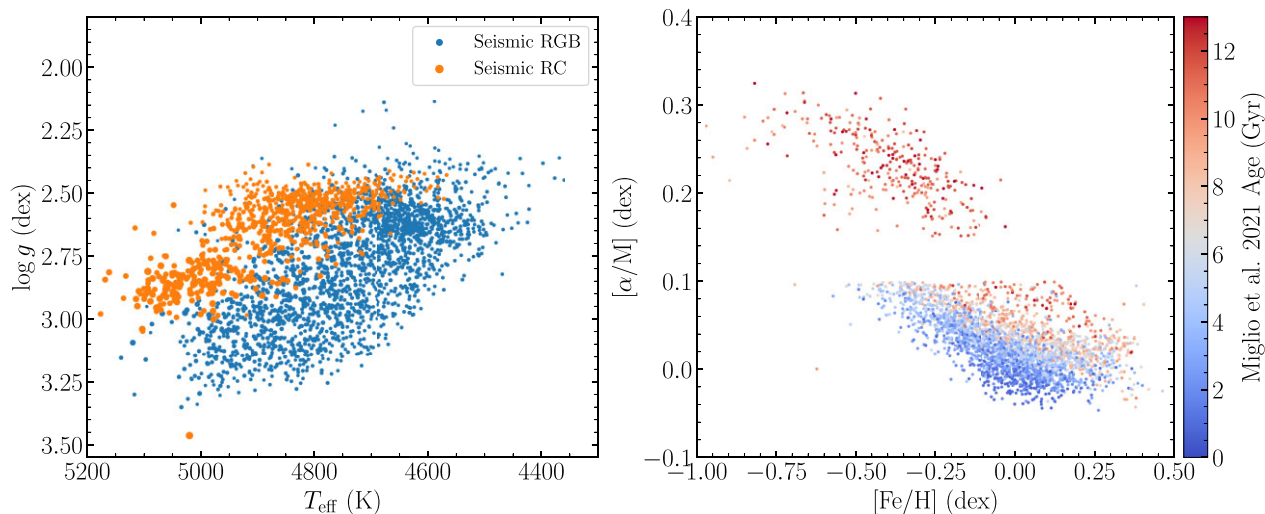


Figure 2. Age measurements from Miglio et al. (2021b) which are used for training in this paper. The left-hand panel shows the distribution of stars in $T_{\text{eff}}\text{-log } g$ space coloured by their seismic classification of red-giant branch (RGB) or red-clump (RC) stars with the marker size corresponding to their mass. The $\log g$ has a limited range with the majority of the stars lying within $2.5 \lesssim \log g \lesssim 3.3$ dex. The right-hand panel shows the distribution of stars in $[\text{Fe}/\text{H}]\text{-}[\alpha/\text{M}]$ space coloured by age. There is an artefact at $[\alpha/\text{Fe}]$ at 0.1 dex to separate α -rich and α -poor population, as Miglio et al. (2021b) separate their analysis into α -rich and α -poor stellar samples.

version of the method from this paper that is updated to use the DR17 ASPCAP stellar parameters and abundances as opposed to those from DR16 in the original work. The Miglio et al. (2021b) catalogue consist of 3078 evolved stars with measured ages. The distribution in the $\log g\text{-}T_{\text{eff}}$ plane and in the $[\alpha/\text{M}]\text{-}[\text{Fe}/\text{H}]$ plane of the sample is given in Fig. 2. The Miglio et al. (2021b) method uses a similar approach as the APOKASC catalogue to determine masses, radii, and ages using the observed light curves, spectroscopic parameters from APOGEE, and stellar models, but it has small implementation differences. Miglio et al. (2021b) uses PARAM (da Silva et al. 2006; Rodrigues et al. 2017) to infer masses and ages while APOKASC-2 uses BeSPP (Serenelli et al. 2017). Another difference is that Miglio et al. (2021b) determines the global seismic parameters using peak-bagging (Davies et al. 2016), where individual radial-mode ($\ell = 0$) frequencies are fit for a subset of stars to get ν_{max} and $\Delta\nu$, allowing a better measurement of those parameters. This in turn leads to improved stellar ages. The ages from Miglio et al. (2021b) that we use have a standard deviation of ~ 10 per cent compared to ages in the APOKASC-2 catalogue. In general, the Miglio et al. (2021b) age uncertainty is ~ 5 per cent larger for RGB but ~ 15 per cent less for RC stars when compared to APOKASC-2.

5 TRAINING AND TESTING

Before training, we further standardize the APOGEE spectra by subtracting the pixel-level mean and dividing by the pixel-level standard deviation after the data reduction steps discussed in Section 4.1. We restrict output PSDs to have $4 < \nu_{\text{max}} < 250 \mu\text{Hz}$ to ensure that the whole p-mode power envelope can be seen as a whole in the PSD with $\sigma_{\nu_{\text{max}}}$ less than 10 per cent. We restrict the sample of APOGEE-*Kepler* pairs to those with PSDs with evolutionary state determinations from APOKASC-2 or Yu et al. (2018), as this provides a good indication of the quality of the PSD (some PSDs without evolutionary state determinations have ν_{max} that appear far off visually). We predict the logarithmic amplitude of the PSD. The first step of the training process is to train the encoder–decoder (as shown in Fig. 1) as a whole

with 9869 pairs of APOGEE spectra and *Kepler* PSDs randomly selected from all APOKASC pairs. We optimize the model using the ADAM optimizer (Kingma & Ba 2014). To accelerate training as well as to make the training process more stable, we weight pixels near the observed ν_{max} higher when calculating the objective function in equation (1). We generate a Gaussian normalized to a height of one for each PSD centred at ν_{max} with a width of twice the $\Delta\nu$ expectation from equation (6) and we add this Gaussian to the existing sample weights for each pixel which were arrays of ones. It is not necessary to know ν_{max} in advance and apply this pixel-level weighting to obtain convergence to a good model, but the addition of this weighting makes the model converge much faster. The introduction of this weighting scheme does not prevent our model from learning features outside of the p-mode power envelope as shown in the bottom right-hand panel of Fig. 3.

After training the encoder–decoder, APOGEE spectra in the APOKASC sample are run through the trained encoder (but not the decoder) to get their latent representations. We then train random forest models to go from these latent representations to other labels, primarily age, but we also predict other quantities from the latent space to aid in understanding the latent space; this training step is shown in the middle panel of Fig. 1. When predicting chemical abundances and global seismic parameters from the latent space, 2000 stars are randomly selected to be the training set. When predicting stellar ages, about 1200 stars are randomly selected among the 2000 stars with ages to be the training set. Our fiducial model predicts age from the latent space augmented with T_{eff} and $[\text{Fe}/\text{H}]$; we refer to these as ‘latent space ages’. In practice, we predict the logarithm of the age and we do not apply any bounds on this logarithm.

6 RESULTS

The ultimate goal of our method is to determine ages from the high-resolution APOGEE data using the encoder, the latent space, and the random-forest regressor from the latent space to age. Because our

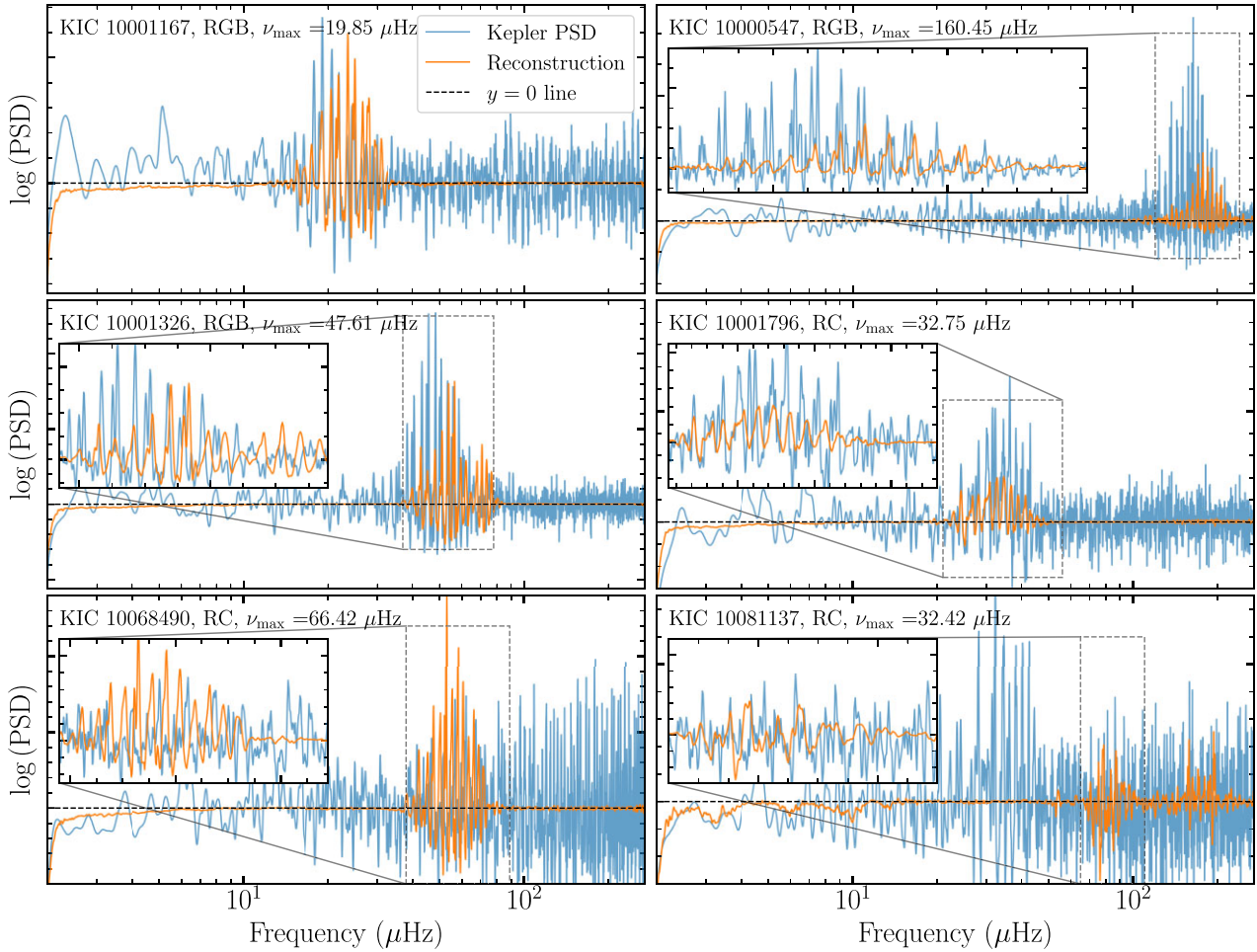


Figure 3. Reconstruction of the power spectral density (PSD) from APOGEE spectra with our encoder–decoder model (orange line) compared to the PSD generated directly from the *Kepler* light curve (blue line) for stars not included in training set, such that the model has never seen those spectrum–PSD pairs before (except for the bottom right-hand panel example, which is from the training set). Our encoder–decoder model successfully reconstructs the area of interest where the p-mode envelope is. The top left-hand panel demonstrates that even in the case of low ν_{\max} , where the amplitude of the p-mode envelope relative to the background is low, the model still successfully reconstructs the envelope at approximately the correct frequency location. The bottom two panels show two RCs with peculiar PSD reconstructions. The bottom left one shows one RC star that is not included in the training set where the p-mode envelope is reconstructed at the correct ν_{\max} , but the overall shape does not resemble the ground truth PSD. The bottom right-hand panel shows an RC star included in the training set where the model fails to reconstruct the p-mode envelope at the expected ν_{\max} . In general, the neural network correctly identifies and reconstructs the p-mode envelope and it does not reconstruct the noise, which is expected as noise is unpredictable from the APOGEE spectra.

methodology consists of multiple important components that build on each other to provide the final ages, we discuss the results from the components in order in this section. We commence with a discussion of the performance of the encoder–decoder in reconstructing the PSD from spectra in Section 6.1, then we take a detailed look at the latent space in Section 6.2, as well as describe how well we can predict different stellar parameters, including age, from the latent space representation of the spectra in Section 6.3. Finally, we discuss the application of our model to the whole APOGEE DR17 data set to get ages for a large sample of stars.

6.1 Power spectral density reconstruction

After training the encoder–decoder on the APOGEE/*Kepler* pairs, we can directly check the performance of the model by generating the reconstructed PSDs for spectra in a test set of pairs that was not used during the training. We expect that any information in

the PSD that is predictable from (or correlates with) the APOGEE spectra to be present in the output. PSD information that is not contained in the APOGEE spectra cannot be obtained by the decoder–encoder, instead the model will predict the mean value of those PSD pixels in the sample to minimize the objective function (equation 1). Information in the APOGEE spectra that is not necessary to reconstruct the PSD will, similarly, not be part of the latent space (this crucial last point is discussed in detail in Section 6.2).

In Fig. 3, we compare the actual power spectra computed from *Kepler* light curves and the encoder–decoder reconstruction from the APOGEE spectrum for pairs in the set of APOGEE/*Kepler* pairs that are not included in training or validation set and, thus, the model has never seen those pairs before. The figure includes stars with large and small ν_{\max} and of different evolutionary states. We see that, overall, the encoder–decoder reconstructs the p-mode envelope remarkably well. The p-mode envelope is the area of

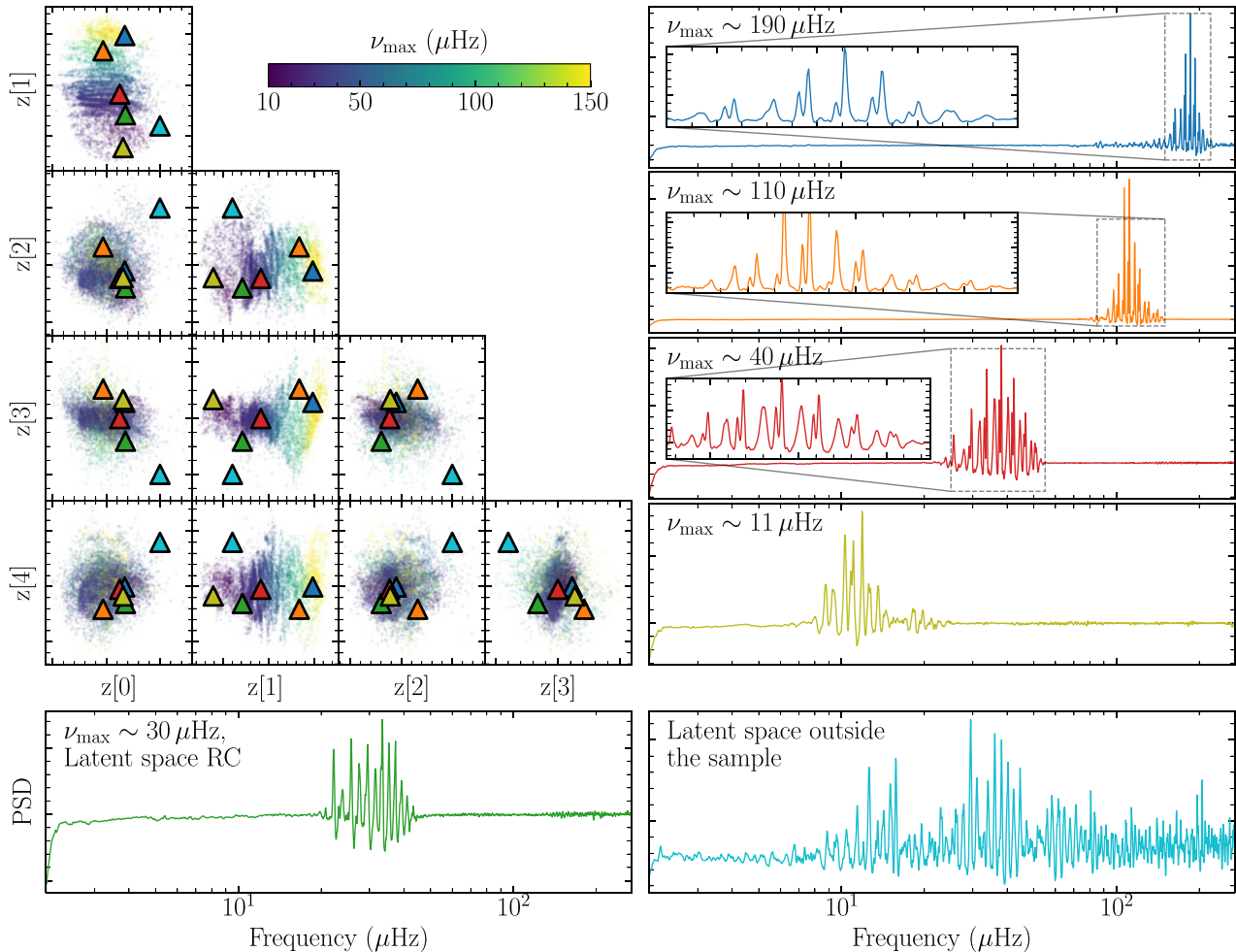


Figure 4. Reconstruction of power spectral densities from a few locations in the latent space. The big panel (top left-hand panel) shows the latent space coloured by ν_{\max} for the encoder–decoder training sample (i.e. the top right-hand panel of Fig. 5). The reconstructed PSDs at the locations of the triangles are shown in the six smaller panels, where the colours of the lines correspond to the colour of the triangle markers in the big panel. The four top right-hand panels show reconstructions for PSDs with ν_{\max} around 190, 110, 40, and 11 μHz . The bottom left-hand panel displays a PSD reconstruction from a region associated with the RC. The bottom right-hand panel shows a reconstruction from a region not populated by any real-world APOGEE spectra; this PSD is much noisier than the others. All PSD panels have arbitrary units on the y-axis and logarithmic frequency on the x-axis.

interest for our purposes, because it contains most of the seismic information relevant to age determination. The encoder–decoder is also able to reconstruct the locations and heights of individual oscillation modes to high precision considering the information comes from spectroscopic observations that only contain information on the surface condition of stars. Thus, the encoder–decoder appears to have learnt the global seismic properties like ν_{\max} and $\Delta\nu$.

The encoder–decoder is able to reconstruct the p-mode envelope well regardless of the value of ν_{\max} . The top left-hand panel of Fig. 3 demonstrates that even for a star with low ν_{\max} ($< 20 \mu\text{Hz}$), where the amplitude of the p-mode envelope relative to the background is low, the model still successfully reconstructs the envelope with an overall peak at approximately the correct location and with individual oscillation modes at the approximately correct locations as well (note that the absolute amplitude of the p-modes are higher at low ν_{\max} , but lower relative to the background granulation and activity noise so PSDs for low ν_{\max} stars seem ‘noisier’). In general, the encoder–decoder correctly identifies the location of the p-mode envelope and of the individual modes for a wide range of stars and reconstructs

them correctly. The model does not reproduce the noise in the PSD (e.g. the photon noise instrumental systematics, and stochastic granulation), which is expected because the noise is unpredictable from the APOGEE spectra. Increasing the latent space dimension from the five dimensions used by the model does not improve the PSD reconstruction for the testing set. The reconstructions are likely limited by the spectroscopic information available in APOGEE spectra.

The use of a variational method in the encoder–decoder also allows us to generate new PSD samples directly from the latent space (i.e. without starting from an APOGEE spectrum). In Fig. 4, we show PSDs generated from latent space locations not directly associated with APOGEE spectra, but within the parameter space of the training set, and one location outside of the training set’s parameter distribution. The PSD samples generated look reasonable across a wide range of ν_{\max} , except the one generated in latent space region associated with the RC, which does not have the correct envelope shape, and the one (cyan) that is generated outside of the parameter space of the training set, which is very noisy.

6.2 The latent space

The latent space of the trained encoder–decoder contains the information learnt by the encoder in its quest to reconstruct the PSD from the near-infrared APOGEE spectrum. It consists of the information that the model found to be crucial to predicting the PSD. Unlike checking the performance of the encoder–decoder as a whole by validating the quality of the PSD reconstruction using test data, the latent space is unsupervised in that we do not use any data to directly inform or restrict its structure. The only constraint placed on the latent space is the variational constraint that pushes it towards having a Gaussian distribution (see Section 2.1). Thus, the best we can do is to study the latent space and make sure that it has properties that we expect and desire when going from spectra to PSD: (i) that it clearly encodes crucial seismic information such as ν_{\max} and $\Delta\nu$ and (ii) that it does not contain extraneous spectral information such as elemental abundance ratios that are unnecessary for reconstructing the PSD but that have galactic evolution correlations with age.

We find that we can satisfactorily predict the PSD using a 5D latent space. Fig. 5 displays the distribution of the data in the 5D latent space values for all APOKASC stars whether they are in the training set or not (we can use all data, because we are simply trying to understand what the model has learned as opposed to directly validating the model, and the large number of stars in the full sample makes it easier to spot clear trends). The figure contains four sub-figures that are colour-coded by (i) the evolutionary state – whether a star is a red-giant branch (RGB) or a red-clump (RC) star – in the top, left, (ii) ν_{\max} in the top, right, (iii) [Fe/H] in the bottom, left, and (iv) excess $\Delta\nu$ in the bottom, right (see equation 7).

In the top, left-hand panel of Fig. 5, we see that the RC stars in red are clearly clustered in most of the latent space dimensions. This is expected for the following two reasons. First, the PSDs for RC stars look very different from those of RGB stars due to effects like mode-mixing (e.g. Grosjean et al. 2014; Mosser et al. 2014), such that one can distinguish RC from RGB with PSDs alone. The effect of these mixed modes can be more clearly seen when converting the PSD to an échelle diagram – transforming the PSD to a 2D image by stacking parts of the PSD separated by the large frequency separation $\Delta\nu$ – such as those shown in Metcalfe et al. (2014). Secondly, pure spectroscopic separation of RC and RGB stars is also possible (e.g. Bovy et al. 2014; Hawkins, Ting & Walter-Rix 2018; Ting, Hawkins & Rix 2018; He, Luo & Chen 2022) and it is possible to obtain samples of RC stars with very high purity (5 per cent contamination) at low completeness. In the latent space, the RC cluster is not entirely separated from the RGB cluster, showing that there is a fundamental limit on the spectroscopic separability of the RC and the RGB for stars at the edges of both clusters. A clean seismic separation of RGB and RC requires the use of the period spacings expected from gravity modes (Bedding et al. 2011). Because we are not primarily interested in classifying RC/RGB stars, we do not pursue that here, but a better spectroscopic classification may be possible using a similar encoder–decoder approach applied to period spacings.

The latent space colour-coded by ν_{\max} in the top, right-hand panel of Fig. 5 shows a smooth colour gradient in some of the latent space dimensions. This is not surprising, as we have already demonstrated that our model can reconstruct power spectra fairly well in Fig. 3 as discussed in Section 6.1. The parameter ν_{\max} scales with the acoustic cut-off frequency ν_{ac} , which in turn scales with surface parameters such as T_{eff} and $\log g$ (Stello et al. 2009) that can easily be determined from APOGEE spectra.

When colour-coding the latent space by [Fe/H] in the bottom, left-hand panel of Fig. 5, we find no strong trend, even though there are numerous [Fe/H] lines in the APOGEE spectrum and [Fe/H] is one of the easiest parameters to obtain from APOGEE spectra. There is information in the PSD that potentially correlates with metallicity. For example, Corsaro et al. (2017) demonstrates that the amplitude of the granulation activity is significantly affected by metallicity. However, we are not sensitive to this, because we explicitly remove the background level of the PSD as discussed in Section 4.2. If the model is working as expected, we should not see metallicity information in the latent space and this is exactly what we see in the bottom, left-hand panel of Fig. 5. We perform a more detailed assessment of the amount of abundance information in the latent space in Section 6.3 below.

The bottom, right-hand panel of Fig. 5 colour-codes the latent space by excess $\Delta\nu$, which is the basic seismic parameter that is most sensitive to stellar mass. Excess $\Delta\nu$ manifests itself in the PSD as small shifts in the separations of the oscillation modes compared to the expectation from equation (7). We see clear trends with excess $\Delta\nu$ in the latent space, indicating that the encoder has learnt to extract seismic information related to mass and to predict the detailed location of the oscillation peaks, rather than simply generating visually good-looking power spectra.

To further understand the latent space, we show in Fig. 6 the Jacobian of the fourth dimension $z[3]$ of the latent space with respect to each pixel in the APOGEE spectra (i.e. how changes in each pixel in the APOGEE spectra affects the value of $z[3]$) for RGB stars with low ν_{\max} and high ν_{\max} . Regions like the hydrogen lines (the Brackett series) in the APOGEE spectral range are known to be sensitive to $\log g$ and we see that $z[3]$ for RGB stars with low ν_{\max} is sensitive to the hydrogen lines, while this is not the case for stars with high ν_{\max} . Overall though, there is no one region in the infrared spectrum that contains the seismic information encoded in the latent space, instead it appears to be spread throughout the APOGEE spectral range.

6.3 Abundances, seismic parameters, and ages

As we discussed in Section 3.2, we use a modified version of the random forest method to map from the latent space to physical parameters such as abundances and stellar ages. Our primary goal is to determine ages from the latent space, but we also train random-forest regressors to determine abundances, seismic parameters, and the evolutionary state to better understand the information content of the latent space. We train separate random forest models for all of these cases.

Stellar parameters and abundances: To quantitatively determine the abundance information content of the latent space, we train different random forest regressors to predict [Fe/H], $[\alpha/M]$, [N/H], and [C/N] from the latent space. We compare these predictions to the APOGEE ASPCAP values in Fig. 7. We see in particular that the latent space is entirely unable to predict $[\alpha/M]$, as the latent-space prediction is simply the mean of the sample regardless of the true $[\alpha/M]$. Thus, the latent space has no information about $[\alpha/M]$. This in turn means that when we use the latent space to determine ages, these ages are entirely uninformed by $[\alpha/M]$. The [Fe/H] and [N/H] abundances show a small trend that still falls short of the 1-to-1 trend and the information in the latent space about these abundances is small. While there is only a weak trend in [N/H] (and similarly, in [C/H]), the latent space does appear to be able to predict [C/N]. This is consistent with the fact that the [C/N] ratio is a good mass proxy for giants (Masseron & Gilmore 2015; Martig

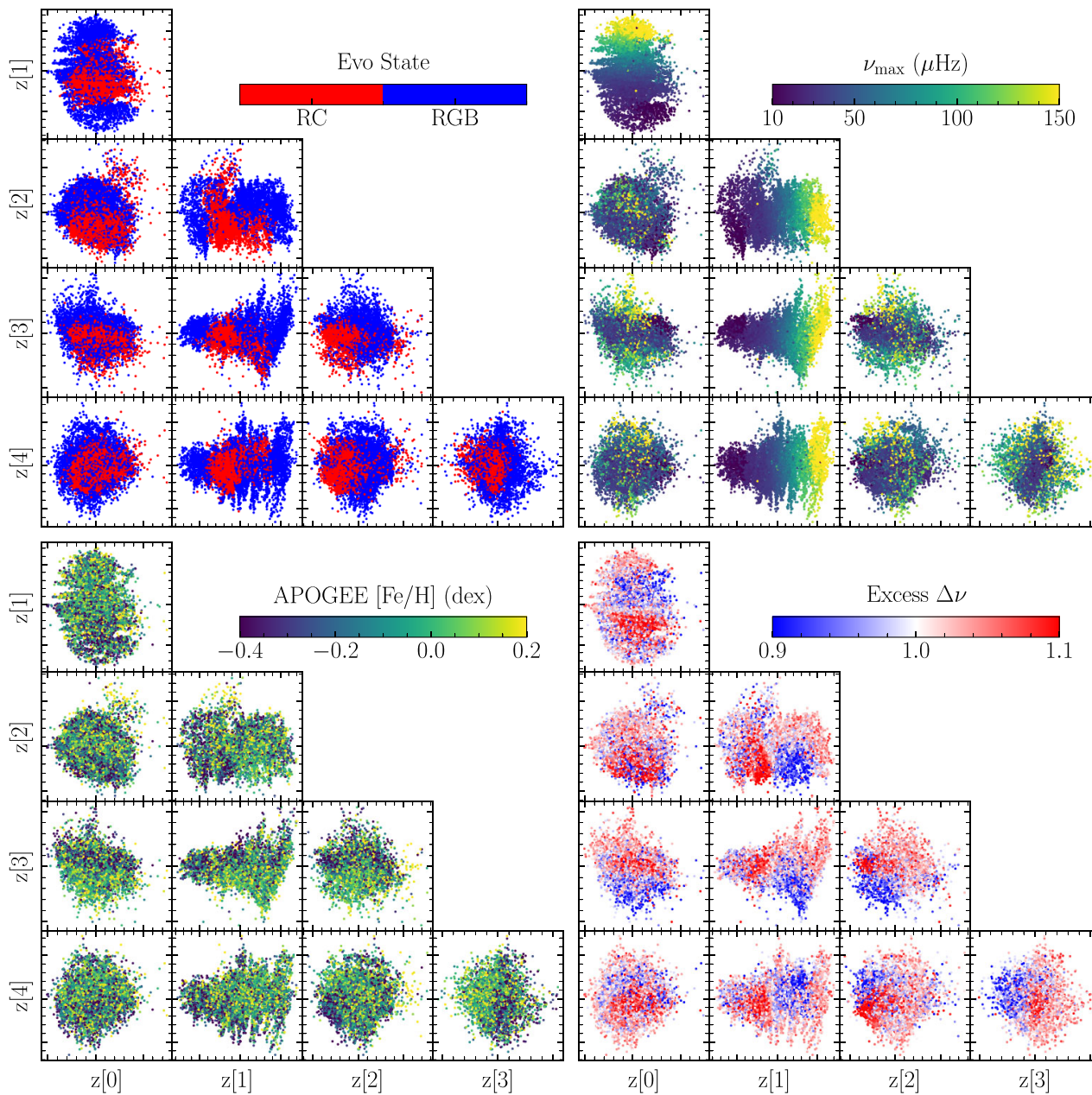


Figure 5. This figure shows the latent space prediction of our encoder–decoder model running through all the APOGEE spectra in the encoder–decoder training set and coloured by different labels (taken from external catalogues, not predicted by our neural networks). The top left-hand panel shows the latent space coloured by evolutionary state, with blue markers for RGB stars and red markers for RC stars. RC stars clearly cluster together in most of the latent space dimension. The top right-hand panel show the same latent space but coloured by ν_{\max} . Smooth trends in ν_{\max} are clearly seen in most of the dimensions, which is expected as the encoder–decoder is able to reconstruct the PSD so well as shown in Fig. 3. The bottom right-hand panel displays the latent space coloured by excess $\Delta\nu$, demonstrating that the model indeed learns the tiny shift in those oscillation peaks related to stellar mass instead of just generating visually good-looking PSDs. The bottom left-hand panel shows the latent space coloured by [Fe/H]. There is no clear [Fe/H] trend in any dimension, showing that the model did not learn [Fe/H]. This figure is discussed in detail in Section 6.2.

et al. 2016), because mass correlates with the seismic parameters. However, the precision to which the latent space can predict [C/N] is still much worse than the precision to which it can be measured from the APOGEE spectra directly (~ 0.05 dex). Thus, the [C/N] prediction here is not limited by how well one can determine [C/N] directly, but is instead limited by the fact that the encoder–decoder did not actually learn [C/N] directly. Rather, it learns other seismic parameters that correlate with [C/N]. In addition to the abundances,

we check how well we can recover T_{eff} from the latent space alone as T_{eff} is needed by traditional asteroseismic age determination. We recover T_{eff} from the latent space at a precision of ~ 110 K, similar to the precision to which one can recover T_{eff} from simply predicting it from $\log g$ for giants. So the latent space does not contain much information on T_{eff} .

Global seismic parameters: We check how accurately we can predict the global seismic parameters ν_{\max} , $\Delta\nu$ and excess $\Delta\nu$ from

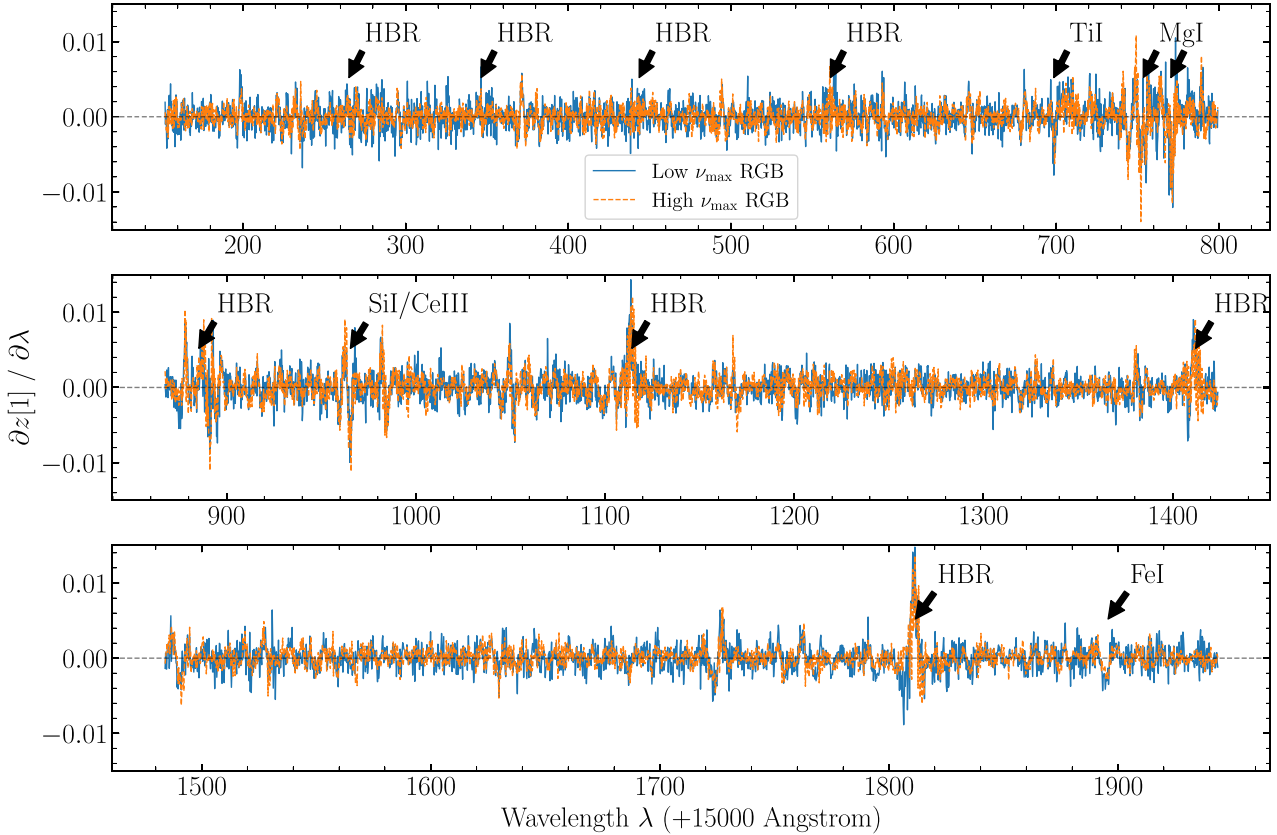


Figure 6. Jacobian of the second dimension $z[1]$ of the latent space for each pixel in the spectra (i.e. how changes in each pixel in the APOGEE spectrum changes the value of $z[3]$) for RGB stars with low ν_{\max} (blue line; $15 \mu\text{Hz} \lesssim \nu_{\max} \lesssim 20 \mu\text{Hz}$) and high ν_{\max} (orange dashed line; $180 \mu\text{Hz} \lesssim \nu_{\max} \lesssim 220 \mu\text{Hz}$) using the encoder part only in the encoder–decoder model. There are a few regions, especially around the hydrogen lines (the Brackett series) in the green and red regions of the APOGEE spectra that are known to be sensitive to $\log g$, with large sensitivity for the low ν_{\max} RGB stars but little sensitivity for the high ν_{\max} RGB stars in this particular latent space dimension. Overall, the information about the latent space appears to be spread over the entire spectrum.

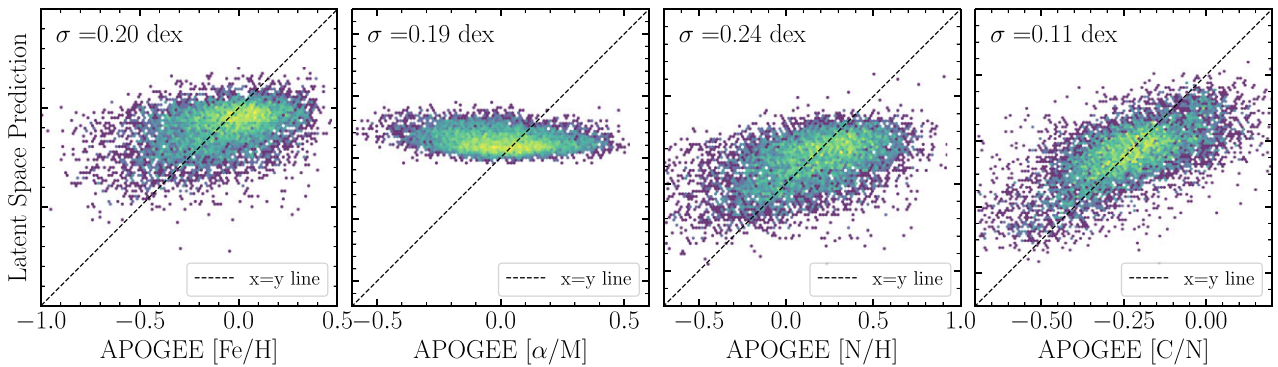


Figure 7. Latent-space abundance predictions. This figure shows density plots on a logarithmic scale of APOGEE abundance prediction from the latent space compared to the ground truth for $[\text{Fe}/\text{H}]$, $[\alpha/\text{M}]$, $[\text{N}/\text{H}]$, $[\text{C}/\text{N}]$, from left to right, respectively, with the overall scatter around the one-to-one line shown in the top left of each panel. It is clear that the latent space contains no information on $[\alpha/\text{M}]$, a small amount of information on $[\text{N}/\text{H}]$ and $[\text{Fe}/\text{H}]$, and some information on $[\text{C}/\text{N}]$.

the latent space. A comparison between the seismic parameters taken from the APOKASC catalogue and those predicted from the latent space is shown in Fig. 8. It is clear that in all three cases we can predict the global seismic parameters to high fidelity from the latent space. Comparing to previous work on getting data-driven seismic parameters from PSDs, our work here has a comparable accuracy, although the nature of the previous work is very different. For

example, Ness et al. (2018) predicts ν_{\max} and $\Delta\nu$ at an accuracy level of ~ 15 per cent from the autocorrelation function (ACF) of the light curve using a very simple but interpretable polynomial model. Hon, Stello & Zinn (2018) develops a ‘quick-look’ deep learning method using convolutional neural networks to detect the presence of solar-like oscillations from 2D images of PSDs and then estimate ν_{\max} at a level of ~ 5 per cent. It is worth noting

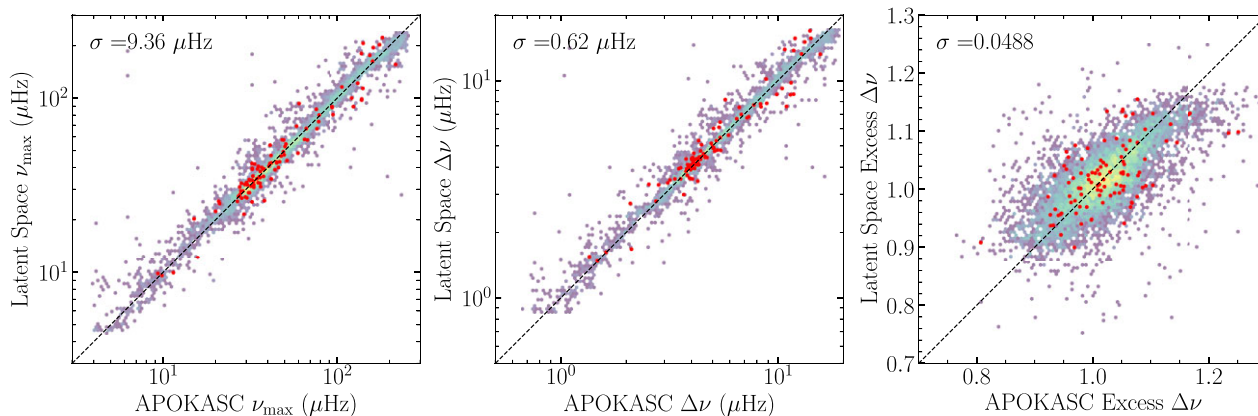


Figure 8. Latent-space seismic parameter predictions. This figure shows density plots on a logarithmic scale of the global seismic parameters predicted using the latent space compared to the ground truth for ν_{\max} in the left-hand panel, $\Delta\nu$ in the middle panel, and excess $\Delta\nu$ in the right-hand panel. These stars are not contained in the training set for the latent space model, but some are contained in the encoder–decoder training set. The red points show additional stars not included in the training set for either the latent space model or the encoder–decoder model; these red points follow approximately the same distribution as the density plot. We are able to determine the important seismic parameters from the latent space to high precision.

		Latent Space Prediction		Total
		RC	RGB	
APOKASC2 Evolutionary State	RC	True Positive 1,733	False Negative 73	1,806
	RGB	False Positive 222	True Negative 2,683	2,905
Total		1,955	2,756	4,711

Figure 9. Latent-space RC-versus-RGB classification. This figure shows the confusion matrix of a naive random forest model trained on the evolutionary state predicting whether a given APOGEE star is an RC or RGB star based on the latent space only. The model performs very well, only misclassifying ~ 10 per cent of the stars. The RC misclassification rate (RC being misclassified as RGB) is ~ 4 per cent while the RGB misclassification rate is ~ 8 per cent.

that converting a PSD to a low-resolution 2D image results in the loss of positional information of the oscillation modes, thus making the performance worse than it could in principle be. Hence, it is not surprising that our seismic parameter predictions from spectra are comparable to those from previous data-driven methods applied to PSDs. Comparing to the values from APOKASC, our predictions for ν_{\max} and $\Delta\nu$ are at the accuracy level of ~ 10 per cent and ~ 7 per cent using the latent space while APOKASC’s ν_{\max} and $\Delta\nu$ precisions are at ~ 0.02 per cent and ~ 0.17 per cent, respectively. This shows that the PSDs contains far more precise seismic information than can be obtained from the APOGEE spectra.

Evolutionary state classification: We saw in Fig. 5 that RC and RGB stars cluster separately in the latent space, which means that we should be able to classify giants as RC or RGB stars using the latent space. Thus, we train a naive random forest classifier to perform this classification using the latent space. The confusion matrix of the resulting classification is shown in Fig. 9. Unlike works

such as Bovy et al. (2014) and Ting et al. (2018), our classifier does not aim at high purity at the expense of completeness, so it is expected that our purity is lower, while our completeness should be higher. Overall, we are able to obtain good classification results based on the latent space, only misclassifying ~ 10 per cent of the stars. We see that our classifier is indeed able to obtain high completeness (only ~ 4 per cent of RC stars are misclassified as RGB stars) at the expense of somewhat higher contamination (~ 11 per cent) than obtained in studies that focus on purity. We can improve the classification’s performance along all axes by ~ 1 to 2 per cent by augmenting the latent space with T_{eff} and $[\text{Fe}/\text{H}]$.

Ages: Finally, we check how well we can predict age, which is the primary goal of this study. To obtain our fiducial age determinations, we augment the latent space with T_{eff} and $[\text{Fe}/\text{H}]$, because these are the spectroscopic parameters that a traditional asteroseismic age determination needs in addition to quantities derived from the light curves (i.e. ν_{\max} and $\Delta\nu$) and we have shown above that they are not available in the latent space. In Fig. 10, we compare the ages that we obtain in this way with the ages from Miglio et al. (2021b). We see that we are able to predict ages well over the entire age range of the sample: the predicted ages cluster around the one-to-one line. The overall bias is ~ 3 per cent with an overall dispersion of ~ 22 per cent. However, the accuracy of our age predictions depend strongly on the evolutionary state: for RC stars, which are generally younger than RGB stars (ages typically in the range 0–5 Gyr), have highly accurate age predictions with a scatter of only ≈ 11 per cent. For RGB stars, on the other hand, the age accuracy is worse with a scatter of ≈ 22 per cent. Nevertheless, this number represents a clear improvement to other data-driven spectroscopic ages including those from Mackereth et al. (2019), which uses a Bayesian neural network that is directly trained on APOKASC ages to obtain an age precision of ~ 30 per cent, and those from Lu et al. (2022), which uses the Cannon (Ness et al. 2015; Casey et al. 2016). Compared to Mackereth et al. (2019), we also do not observe any plateauing at ~ 8 Gyr, but we are instead able to obtain precise ages for old stars. Additionally, unlike the previous methods, we have demonstrated above that our ages are not informed by information coming from abundance ratios like $[\alpha/\text{M}]$, because this information is not contained in the latent

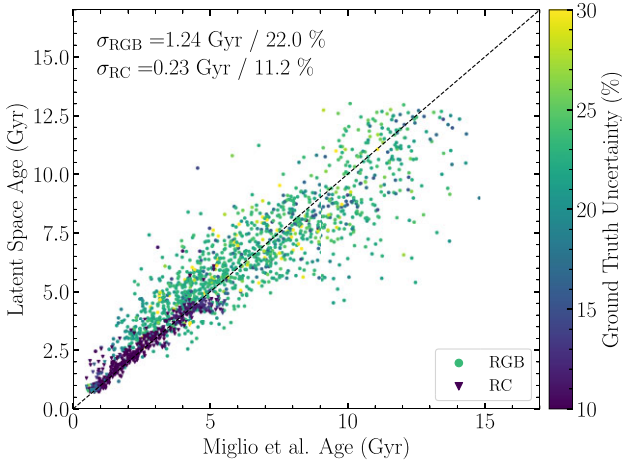


Figure 10. Latent-space age predictions. This figure shows the age prediction for stars based on the latent space augmented by T_{eff} and $[\text{Fe}/\text{H}]$ compared to the ground-truth value from Miglio et al. (2021b) for stars in the test set. Points are coloured by the uncertainty in the Miglio et al. (2021b) age with the marker shape indicating whether a star is an RC (triangle) or RGB (circle) star. The model is only trained on ~ 1200 stars selected randomly from Miglio et al. (2021b). Stars with very low ground truth uncertainty around 0–5 Gyr old are RC stars and we are able to predict these ages to high precision. The standard deviation around the one-to-one line for RC and RGB stars separately are given in the top left-hand corner and are ~ 11 per cent and ~ 22 per cent, respectively.

space. We also obtain uncertainties on our predicted ages from the random forest regressor. To check the quality of these uncertainties, we compute the distribution of $(\hat{y} - y) / \sqrt{\sigma_y^2 + \sigma_{\hat{y}}^2}$, where y is the ground truth age, \hat{y} is the model age, and σ_y and $\sigma_{\hat{y}}$ are their respective uncertainties. If the uncertainties are correct, this distribution should be a standard normal distribution. Instead, we find that the distribution is normal with a width ~ 0.75 , indicating that we are overestimating the uncertainty in our predicted ages. We provide further discussion of the predicted latent-space ages in Section 8.1 below.

Table 1. Data Model of `nn_latent_age_dr17.csv` that is available here https://github.com/henrysky/astroNN.ages/blob/main/nn_latent_age_dr17.csv.gz which is the data file generated from this paper, row-matched to the official `allStar-dr17-synspec.rev1.fits` with 733 901 rows in total while 308 119 stars with latent representation computed with a subset of 142 257 stars with latent space age measured. Missing data are represented by NaN. We recommend to use a latent space age uncertainty cut of 40 per cent, stars without any STARFLAG and ASPCAPFLAG (see Section 6.3 and Section 7.1 for discussion).

Label	Physical Units	Sources	Descriptions
<code>apogee_id</code>	n/a	APOGEE DR17	APOGEE ID
<code>telescope</code>	n/a	APOGEE DR17	Telescope used for APOGEE observation (<code>apo25m</code> or <code>lco25m</code>)
<code>field</code>	n/a	APOGEE DR17	APOGEE field name
<code>STARFLAG</code>	n/a	APOGEE DR17	APOGEE star flag bitmask
<code>ASPCAPFLAG</code>	n/a	APOGEE DR17	APOGEE ASPCAP flag bitmask
<code>z0</code>	n/a	This paper	1st dimension of latent representation
<code>z1</code>	n/a	This paper	2nd dimension of latent representation
<code>z2</code>	n/a	This paper	3rd dimension of latent representation
<code>z3</code>	n/a	This paper	4th dimension of latent representation
<code>z4</code>	n/a	This paper	5th dimension of latent representation
<code>LogAge</code>	$\log_{10}(\text{Gyr})$	This paper	Logarithmic latent space age
<code>LogAge_Error</code>	$\log_{10}(\text{Gyr})$	This paper	Logarithmic latent space age uncertainty
<code>Age</code>	Gyr	This paper	Latent space age
<code>Age_Error</code>	Gyr	This paper	Latent space age uncertainty

7 THE APOGEE DR17 LATENT-SPACE AGE CATALOGUE AND THE AGE STRUCTURE OF THE MILKY WAY DISC

7.1 Age catalogue

We have applied our models to the whole APOGEE DR17 catalogue to obtain latent space ages and produce a publicly available catalogue. First, we compute the latent-space representation of APOGEE DR17 stars using the trained encoder–decoder for stars within the parameter space of the training set, that is, for stars with $1.5 < \log g < 3.6$ and $\text{SNREV} > 30$, where SNREV is an alternative signal-to-noise measurement recommended by APOGEE that takes detector persistence issues in account. This allows 308 119 stars to have their latent representation computed. To compute latent space ages, we further restrict the sample to stars with T_{eff} and $[\text{Fe}/\text{H}]$ available as required by our pipeline, as well as lying in the same $2.5 < \log g < 3.6$ range as the Miglio et al. (2021b) training set used in this work. A data model and download link are available in Table 1. For science analyses, we recommend using stars with no STARFLAG and ASPCAPFLAG flag set as well as requiring a latent space age uncertainty less than 40 per cent.

When compared to our previous work (Mackereth et al. 2019) of data-driven spectroscopic ages with neural networks as shown in Fig. 11, the latent space ages from this work extend to significantly older ages for old stars, because unlike in our previous work, the latent space ages do not exhibit a plateau at around 8 Gyr. Otherwise the two works are quite consistent with each other with a scatter of 25 per cent. When plotting $[\text{Fe}/\text{H}] - [\alpha/\text{M}]$ coloured by latent space age as well as latent space age versus $[\alpha/\text{M}]$ as shown in Fig. 12, our latent space age shows the expected trends of low $[\alpha/\text{M}]$ sequence stars being young and high $[\alpha/\text{M}]$ sequence stars being old. The oldest stars in the low $[\alpha/\text{M}]$ sequence are as old as the youngest high $[\alpha/\text{M}]$ sequence stars. Although we have a number of young high $[\alpha/\text{M}]$ stars, some of them can be removed by further restricting $\log g > 2.55$ dex instead of 2.5 dex to avoid the edge of training set parameter space and using our recommended $\sigma_{\text{age}} < 40$ per cent instead of the 50 per cent used in the figure for completeness.

For the purpose of studying spatial age–abundance trends in the Milky Way disc in the next subsection we use the spectrophotometric distances from Leung & Bovy (2019b) and convert from heliocentric

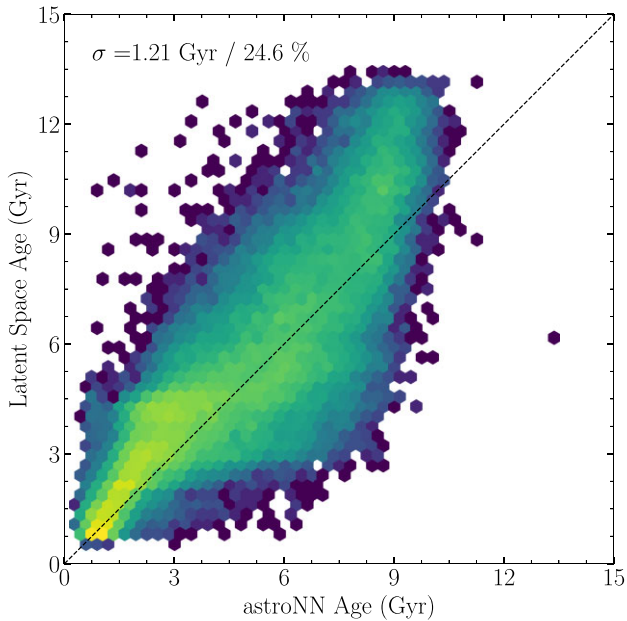


Figure 11. Latent-space ages compared to spectroscopic ages from astroNN (Mackereth et al. 2019). This figure shows the logarithmic density of a comparison of the latent-space ages from this work to the spectroscopic ages obtained using a Bayesian neural network trained directly on APOGEE spectra without physical constraints against using abundance information. The overall scatter between these ages predictions is ~ 25 per cent. Unlike the previous ages, the latent-space ages do not plateau around 8 Gyr.

to Galactocentric coordinates assuming $R_0 = 8.23$ kpc and $v_\odot = 249.44$ km s $^{-1}$ (Leung et al. 2023) and $z_\odot = 20.8$ pc (Bennett & Bovy 2019). Orbital parameters are calculated using `galpy` (Bovy 2015) using the standard `MWPotential2014` potential.

7.2 The age structure of the Milky Way disc

As a first application of our new age catalogue, we present a brief investigation into spatial age-abundance trends in the Milky Way disc here. In Fig. 12, the left-hand panel shows the $[\text{Fe}/\text{H}]$ – $[\alpha/\text{M}]$ distribution coloured by mean latent space age in each bin. We see a clear separation between old high and young low $[\alpha/\text{M}]$ sequences. The right-hand panel shows the age– $[\alpha/\text{M}]$ distribution coloured by angular momentum L_z with the cyan dashed line representing the median relation between age and $[\alpha/\text{M}]$; the $[\alpha/\text{M}]$ scatter around this relation is ≈ 0.05 dex for any age. The overall trend is a slowly increasing $[\alpha/\text{M}]$ with age for the low $[\alpha/\text{M}]$ sequence and a steeper trend when transitioning to the high $[\alpha/\text{M}]$ sequence, similar to what is seen for local samples (Haywood et al. 2013). The trend with angular momentum shows that the $[\alpha/\text{M}]$ –age trend is steeper in the high-angular momentum, outer disc than it is in the inner disc.

In the right-hand panel of Fig. 12, we also see that there are young, high $[\alpha/\text{M}]$ stars with ages ~ 6 Gyr, similar to previous works (e.g. Chiappini et al. 2015; Martig et al. (2015)), which are likely high $[\alpha/\text{M}]$ stars with unusual elemental abundance ratios (e.g. Hekker & Johnson 2019) that seem to be overmassive (and thus appear young) due to binary stellar evolution (Zhang et al. 2021; Jofre et al. 2023). It is worth noting that when we adopt the definition of young, high $[\alpha/\text{M}]$ stars with a flat cut of $[\alpha/\text{M}] > 0.15$ dex and age younger than 6 Gyr used in previous literature (e.g. Martig et al. 2015), we find that 9 per cent of the high $[\alpha/\text{M}]$ sequence population is young. This can

be compared to ≈ 6 per cent in Martig et al. (2015) and Zhang et al. (2021). Further restricting to $\log g > 2.55$ dex, the fraction of young high $[\alpha/\text{M}]$ stars approaches 7 per cent. Alternatively, defining high $[\alpha/\text{M}]$ as $[\alpha/\text{M}] > 0.18$ dex, similar to values adopted in asteroseismic analyses, the fraction is 6 per cent. These results are interesting, because there are no young high $[\alpha/\text{M}]$ stars in our training set, yet we recover similar fractions of them as independent, previous analyses.

The spatial distribution of stars in Galactocentric radius R and vertical height z is shown in Fig. 13. We clearly see the vertical flaring of the disc in age with radius. The outer disc is uniformly younger than ≈ 5 Gyr.

The age– $[\text{Fe}/\text{H}]$ – $[\alpha/\text{M}]$ distribution of stars in Galactocentric radius R and vertical height $|z|$ bins is displayed in Fig. 14. The dashed black line that we use to separate the low and high $[\alpha/\text{M}]$ stars is given by the combination of the following equations

$$[\alpha/\text{M}] > -0.2211 \times [\text{Fe}/\text{H}] + 0.0442 \text{ dex},$$

$$[\alpha/\text{M}] > 0.05 \text{ dex} \quad (8)$$

We see that the high $[\alpha/\text{M}]$ sequence is old wherever it appears, with age declining from ≈ 12 Gyr at its low-metallicity end to ≈ 8 Gyr at its high-metallicity end; the smattering of young high $[\alpha/\text{M}]$ stars is uniformly mixed up with these, again demonstrating that these are likely anomalously massive rather than anomalously young stars. The low $[\alpha/\text{M}]$ stars are generally younger than 8 Gyr. While the age–abundance trends seen in this figure are generally what has been found before, the precision of our ages for a large sample of stars sharpens the picture significantly compared to previous work.

A more detailed view of the age distribution and its overall radial and abundance trends is presented in Fig. 15. While the giant sample selected by our cuts does not sample age uniformly, calculations similar to those in Section 5 of Bovy et al. (2014) show that the relative age bias of ≈ 2 to 5 Gyr to ≈ 10 Gyr is a factor of ≈ 2.5 , with the age bias being relatively flat between 1 and 5 Gyr and then slowly decreasing towards larger ages. Given this, the overall age distribution in the left-hand panel is indicative of a roughly uniform intrinsic age distribution, or a flat star-formation history. Also from the left-hand panel, we see that the inner disc is more heavily weighted towards old ages, while the outer disc barely extends beyond 5 Gyr (see also the rightmost panels of Fig. 14). The middle and right-hand panels split these radial age trends by $[\alpha/\text{M}]$ abundance, defining high and low $[\alpha/\text{M}]$ sequences using the separation from equation (8). The age distribution of the high $[\alpha/\text{M}]$ population is approximately the same regardless of radius and peaks at ≈ 10 Gyr. Accounting for the age bias, there is a tail towards younger ages, but most of these are actually the likely anomalously massive high $[\alpha/\text{M}]$ stars discussed above. Fig. 14 demonstrates that the high $[\alpha/\text{M}]$ sequence ends at ≈ 8 Gyr. The low $[\alpha/\text{M}]$ sequence has a clear radial age trend, with the inner disc being older than the outer disc. Regardless of radius, the low $[\alpha/\text{M}]$ disc is younger than ≈ 9 Gyr, or likely younger than ≈ 8 Gyr once we account for age errors (we, however, do not attempt a deconvolution of the age distribution in the first look here). This all indicates that the high and low $[\alpha/\text{M}]$ sequences formed at different times, with the high $[\alpha/\text{M}]$ sequence corresponding to the first ≈ 4 Gyr of the disc’s existence and the low $[\alpha/\text{M}]$ sequence corresponding to the last ≈ 8 Gyr. There does not appear to be a large hiatus between the two, although properly understanding the transition would require proper deconvolution of the age distribution and a good understanding of the anomalous young high $[\alpha/\text{M}]$ stars. A similar picture has emerged

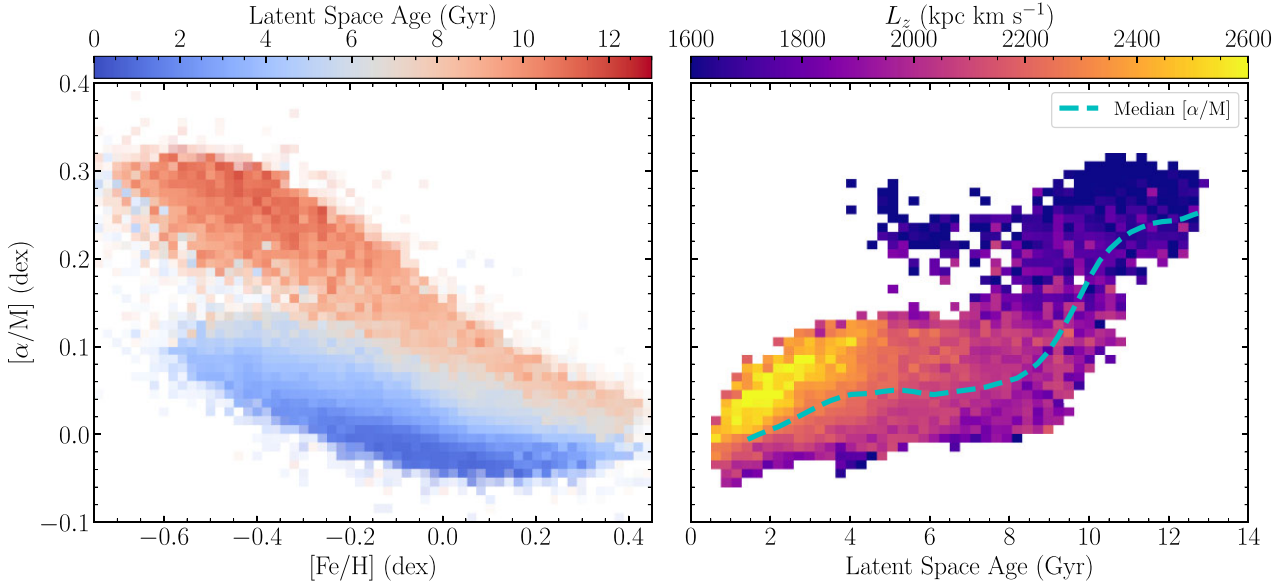


Figure 12. Ages inferred from the latent space model applied to the whole APOGEE DR17 data set. After various quality cuts of spectral signal-to-noise ratio $\text{SNREV} > 30$, latent space age $\sigma_{\text{age}} < 50$ per cent, $2.5 < \log g < 3.3$, which is the range covered by the training set, as well as cutting on the STARFLAG and ASPCAPFLAG flags, $\sim 56\,000$ giants are included in both panels. The left-hand panel displays the $[\text{Fe}/\text{H}]$ – $[\alpha/\text{M}]$ bimodality coloured by average latent space age in each bin. The right-hand panel shows the latent space age versus $[\alpha/\text{M}]$ coloured by average angular momentum L_z in each bin, with a cyan dashed line representing the median age– $[\alpha/\text{M}]$ relation. Both panels clearly demonstrate that high $[\alpha/\text{M}]$ sequence stars are significantly older than low $[\alpha/\text{M}]$ sequence stars. In the right-hand panel, there is a very small population of young ($\lesssim 6$ Gyr) $[\alpha/\text{M}]$ -enriched stars that is not presented in the training set (see Fig. 2) and that was previously observed using asteroseismic ages, but not usually in data-driven spectroscopic ages.

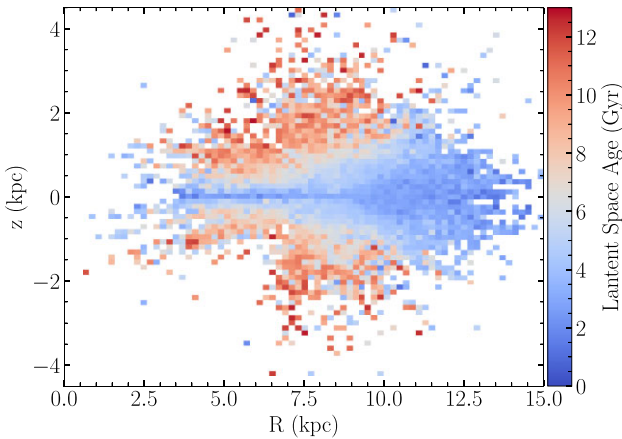


Figure 13. Spatial distribution of stars in APOGEE DR17 in Galactocentric radius R and vertical height z coloured by median latent space age. The spatial distribution clearly shows the vertical flaring in age of the disc with radius.

previously from observations near the Sun (e.g. Haywood et al. 2013; Bonaca et al. 2020).

8 DISCUSSION

8.1 Latent space ages

In the previous section and in Fig. 10, we have demonstrated that we can obtain precise ages from the latent space augmented by T_{eff} and $[\text{Fe}/\text{H}]$. To understand how important each of these three ingredients is to the prediction of the age, we have tested a few

other combinations of these ingredients. In Fig. 16, we use a three separate models to predict age. The first model, shown in the left-hand panel, uses the latent space only, the second model, in the middle, uses the latent space and T_{eff} , and the third model, in the right-hand panel, uses the latent space and $[\text{Fe}/\text{H}]$. These three models allow us to assess the relative importance of the latent space and the augmented parameters in the age prediction. The age predictions in all three of these models are significantly worse than our fiducial latent-space age. Interestingly, the predicted age plateaus at ~ 8 Gyr similar to what we observed in Mackereth et al. (2019) when directly predicting ages from APOGEE spectra. When only using the latent space, the age prediction still follows the one-to-one relationship relatively well, albeit with large scatter, indicating that the seismic information in the latent space is rich enough to provide a rough estimate of the age. While the scatter in the predicted ages decreases when adding T_{eff} and $[\text{Fe}/\text{H}]$ separately, it is clear from comparing the middle and right-hand panels to the left-hand panel that the addition of these parameters separately does not qualitatively improve the age prediction over that from the latent space alone. Thus, adding both T_{eff} and $[\text{Fe}/\text{H}]$ to the latent space is crucial to the good performance in Fig. 10.

In Fig. 16, it is clear that RC stars get good ages regardless of which combination of latent space, T_{eff} , and $[\text{Fe}/\text{H}]$ we use. This is a direct consequence of the scaling relations combined with the fact that RC stars have a narrow range of absolute magnitudes. The mass scaling relation in equation (3) can be expressed in terms of the luminosity instead of T_{eff} as shown in equation (7) of Miglio et al. (2012). Because of the narrow luminosity spread of RC stars (Paczynski & Stanek 1998), good ages for RC stars can thus be obtained with seismic data only assuming no significant mass-loss. This means that we can obtain good RC ages just using the latent space.

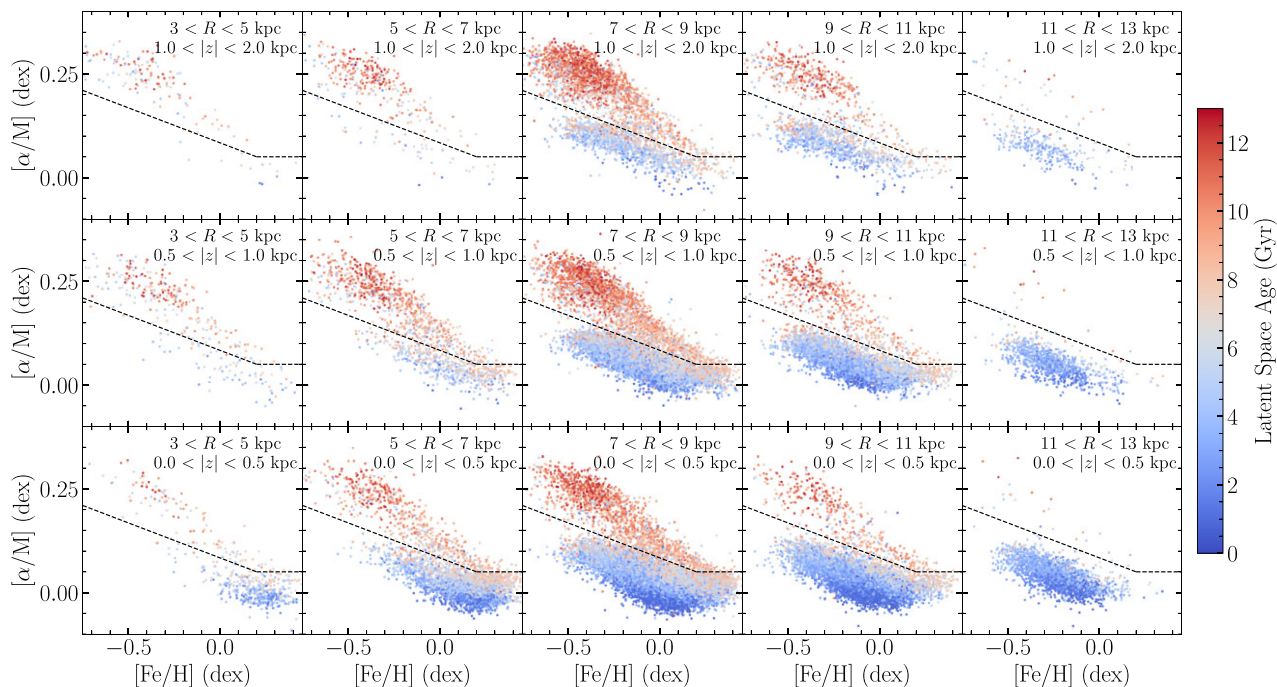


Figure 14. The age-[Fe/H]-[α/M] distribution of stars in the Milky Way disc. This figure – similar to fig. 4 in Hayden et al. (2015) – shows the [Fe/H]-[α/M] distribution colour-coded by age of stars in spatial R - z bins spanning $3 \text{ kpc} < R < 13 \text{ kpc}$ in Galactic radius and $|z| < 2 \text{ kpc}$ in height from the Galactic midplane. The dashed line roughly divides the high- and low-[α/M] sequences and is the same for all subplots. It is clear that the high-[α/M] sequence is older than the low-[α/M] sequence with little age overlap between the two.

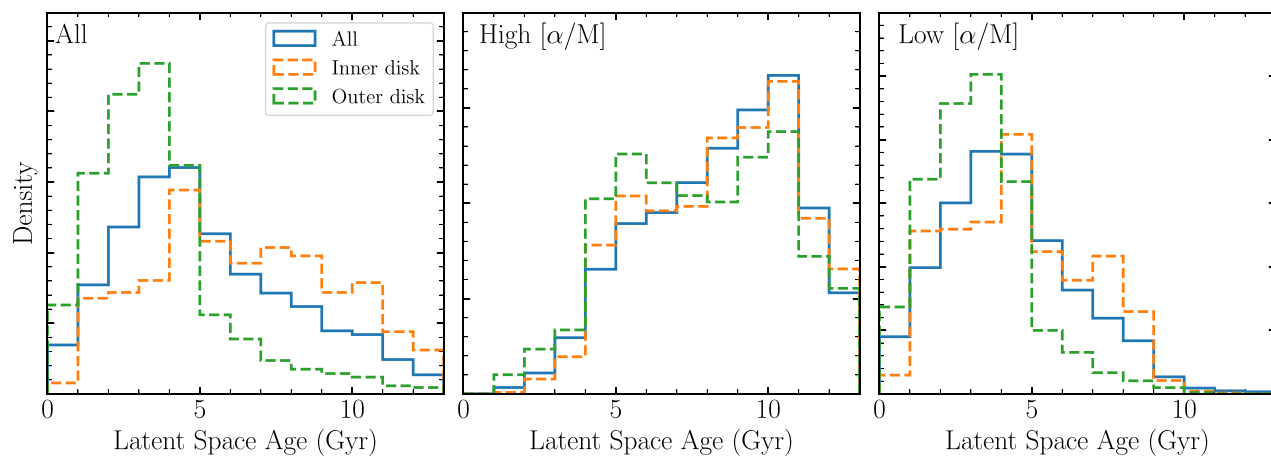


Figure 15. Stellar age distributions in the Milky Way. Each panel in this figure displays the age distribution stars across the entire disc (‘All’) and those in the inner and outer discs, defined as $3 \text{ kpc} < R < 6 \text{ kpc}$ and $10 \text{ kpc} < R < 13 \text{ kpc}$, respectively. The left-hand panel shows all stars, regardless of their abundances, while the middle and right-hand panels split the stars into high [α/M] and low [α/M] using the dashed line in Fig. 14. The left-hand panel demonstrates that outer disc stars are significantly younger than inner disc stars. The middle panel shows that high [α/M] stars are old, but extend down to ages of 5 Gyr. The right-hand panel demonstrates that low [α/M] sequence stars are young across the disc and do not exceed 9 Gyr (accounting for uncertainties, the low [α/M] upper age cut-off is ≈ 8 Gyr).

That not using both T_{eff} and [Fe/H] degrades the age predictions is expected, because traditional asteroseismic age pipelines require at least the following parameters to predict the age: seismic information (typically in the form of ν_{max} and $\Delta\nu$), T_{eff} , and [Fe/H]. This is because we can obtain the stellar mass using the seismic information and T_{eff} (equation 3) and converting mass to age mostly relies on the [Fe/H]-dependent main-sequence lifetime of the star. While we expect the latent space to contain all the seismic information that

can be extracted from the APOGEE spectra, it does not contain T_{eff} as we have discussed in Section 6.3 or [Fe/H] as we have shown in Fig. 5. Thus, it is not surprising that the age precision degrades when dropping one of T_{eff} and [Fe/H].

Finally, we investigate how the latent space ages change when we also include [α/M] in the prediction. Fig. 17 shows the age prediction for a model that includes [α/M] directly in addition to the latent space, T_{eff} , and [Fe/H]. Although including [α/M] does

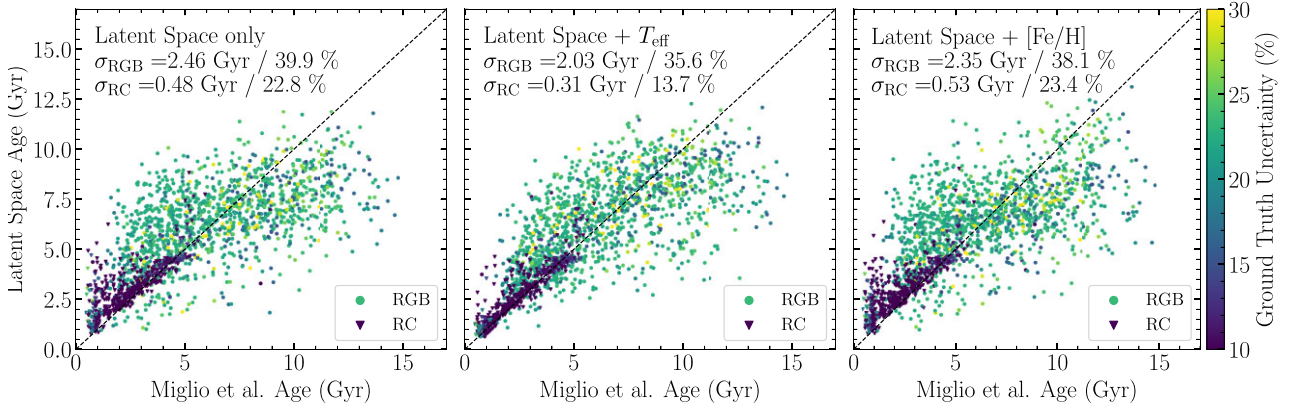


Figure 16. Latent space age prediction using different combinations of parameters. The left-hand panel shows the prediction based on the latent space only without any additional parameters. The middle panel displays the prediction based on the latent space augmented with T_{eff} and the right-hand panel’s age is based on the latent space with $[\text{Fe}/\text{H}]$. These can be compared to our fiducial latent space ages, which are based on the latent space augmented with both T_{eff} and $[\text{Fe}/\text{H}]$ and for which this comparison is shown in Fig. 10. In all three cases, the age prediction is significantly worse than in our fiducial model, with much larger scatter compared to the ground-truth ages and with a plateau arising at old ages ($\gtrsim 10$ Gyr).

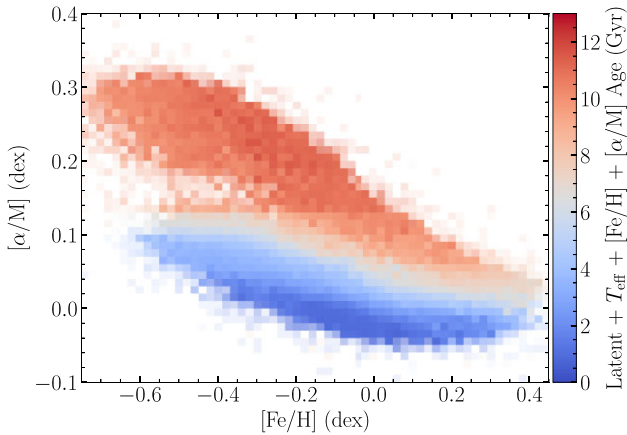


Figure 17. The impact of including $[\alpha/\text{M}]$ in the latent-space age model. This figure is similar to the left-hand panel of Fig. 12, but the latent space age here is obtained with model using the combination of the latent space, T_{eff} , $[\text{Fe}/\text{H}]$, and $[\alpha/\text{M}]$. The additional $[\alpha/\text{M}]$ on top of the other parameters smooths the $[\alpha/\text{M}]$ bi-modality plot significantly compared to the left-hand panel of Fig. 12, with additionally artifacts appearing at $[\alpha/\text{M}] \sim 0.13$ dex because of the $[\alpha/\text{M}]$ selection from our training set (see Fig. 2). This demonstrates that including $[\alpha/\text{M}]$ information in spectroscopic age determination strongly impacts the derived age–abundance trends.

not cause age to correlate much more strongly with $[\alpha/\text{M}]$, the inclusion of $[\alpha/\text{M}]$ does significantly smooth the age trend in the $[\text{Fe}/\text{H}]$ – $[\alpha/\text{M}]$ space while only improving the performance of our model by about 2 per cent. Therefore, even though $[\alpha/\text{M}]$ is also a parameter sometimes used by stellar models to get ages, we choose not to include $[\alpha/\text{M}]$ to avoid the adverse effects that result from its inclusion.

Our current latent space age model suffer from the limitations of the age, because the model can only be applied to new stars in the same parameter space as that covered by the training sets of the encoder–decoder and the age. In this case, the most limiting factor is the narrow $2.5 < \log g < 3.6$ range of the age training set.

8.2 Application with *TESS* and beyond

The NASA *Transiting Exoplanet Survey Satellite* (*TESS*) mission is an all-sky survey that allows for the detection of solar-like oscillations in at least an order of magnitude more giants than *Kepler* (Silva Aguirre et al. 2020; Hon et al. 2021; Mackereth et al. 2021; Hon et al. 2022). Compared to *Kepler* or even the K2 mission (Howell et al. 2014), *TESS* light curves are typically much shorter and less precise, making it more difficult to do precise asteroseismology.

We can exploit the flexibility of the encoder–decoder model to include *TESS* light curves. To train our method, we require pairs of APOGEE spectra and light curves that cover a larger parameter space than that spanned by the training set of the ages (i.e. we cannot train the latent space \rightarrow age regression with stars outside of the parameter space of the training set of the encoder–decoder; the parameter space spanned by the training set of the encoder–decoder is much larger than that of the latent space age). Including light curves from *TESS* can help enlarge the parameter space with stars with a wider range of chemical abundance patterns from different parts of the Galaxy, especially stars with low metallicity ($\lesssim -1.5$ dex) that are not abundant in the *Kepler* field.

Data-driven models, and in particular deep neural networks, are susceptible to small changes in the data. For example, a trained neural network to predict abundances from spectra cannot typically be applied to data taken by different instruments, even if they cover the same wavelength range. Also, it is difficult to train a model on synthetic data (e.g. synthetic spectra) and then apply it to observational data (e.g. observed spectra); often this type of analysis will show a ‘synthetic gap’ between the synthetic and true data that adversely affects the performance of the method (e.g. Fabbro et al. 2018). In the application at hand here, *TESS* and *Kepler* have different instrumental noise properties, different photometric precision, different observing cadences, etc. Naively combining *TESS* and *Kepler* light curves will likely make our model worse (see e.g. Hon et al. 2021 for a discussion of the difference between *TESS* and *Kepler* for machine learning models). Making use of better algorithms to generate the PSD (e.g. the multitaper algorithm of Patil et al. 2022) may result in more uniform PSDs derived from observational data. We could also use different styles of transfer learning that can map data taken with different instruments onto a uniform scale (e.g. O’Brian et al.

2021). For example, we could employ another encoder–decoder to map *TESS* PSDs to *Kepler* PSDs trained on overlapping observations between *TESS* and *Kepler*. Future missions like PLATO (Rauer et al. 2014) and HAYDN (Miglio et al. 2021a) will also have their own photometric precisions, baselines, and cadences and would therefore also benefit from such an approach.

8.3 Prospects for SDSS-V

The Milky-Way Mapper (MWM) in SDSS-V is an ongoing panoptic survey similar to the APOGEE survey that uses the same spectrograph, but combines it with a robotic focal plane system (FPS) with 300 robotic fibre positioners allowing flexible, high-cadence targeting and obtaining high target densities in a small field of view (Pogge et al. 2020). The Galactic Genesis Survey (GGS) programme within MWM will produce a densely sampled panoptic spectroscopic stellar map covering a large volume of the Milky Way by targeting very luminous giants with low $\log g$ (typically lower $\log g$ than APOGEE observations). As we have discussed in Fig. 12 and Section 8.1, currently the latent space age model only applies to the parameter space covered by the ages of Miglio et al. (2021b), with the major limitation coming from the narrow range of $\log g$ present in that sample. Because many stars in GGS have lower $\log g$ than this, our model will not be directly applicable to GGS/MWM observations. In particular, the $\log g$ limitation makes it difficult to reach the bulge with current APOGEE-like observations (see Fig. 13).

The minimum useful frequency we can obtain from the Lomb–Scargle PSD of *Kepler* light curves is about $3 \mu\text{Hz}$ and this minimum has been adopted by many previous papers (e.g. Hon et al. 2018; Ness et al. 2018). While we might be able to determine ν_{max} down to this minimum frequency, realistically we can only obtain precise determinations of additional global seismic parameters as $\Delta\nu$ at $\Delta\nu \gtrsim 7 \mu\text{Hz}$. This cut-off corresponds to giants with $\log g \sim 1.8$ dex using standard scaling relations. Aside from the fact that the estimating mass and age using the empirical scaling relations might break down at very low ν_{max} , the minimum frequency sets a lower $\log g$ limit on applying our encoder–decoder methodology to luminous giant stars. The $\log g$ limitation along with the lack of low-metallicity stars in the *Kepler* field currently makes it difficult to train and apply our method to obtain age estimates for interesting objects such as the Gaia–Enceladus merger remnant (Helmi et al. 2018), which shows up in APOGEE in significant numbers only at low $\log g$ and low metallicity.

8.4 Using alternative mass measurements to train

The two-component nature of our methodology, where we first extract seismic information that contains mass information into the latent space using the encoder–decoder and we then map the latent space to age, means that we have flexibility in how we train the latent space \rightarrow age regression. In the second step, we do not need to use ages derived from the PSDs used to train the encoder–decoder, but we could instead use ages obtained from other parts of the sky (e.g. the *TESS* continuous viewing zone), as long as the age sample has similar stellar populations as the sample used to train the encoder–decoder, because we need the latent-space representation of the age sample. We do not even need to use seismic ages at all in the second step, although we do so here because currently asteroseismology is the only method that provides precise ages for large-ish samples of stars. As we have shown in this work, we only need about a thousand

accurate age measurements as the training set for the latent space age.

Because we can convert mass to age relatively precisely along the giant branch, we could use alternative mass determinations to train the latent space age. For example, we could use stars in a cluster with masses determined using stellar evolution models and we could even require that they return the same age for all stars in the cluster. We could also make use of masses determined for eclipsing binaries using Kepler’s laws, as these allow mass determinations at the few per cent level compared with the ~ 8 per cent masses for giants from asteroseismology. That these methods can obtain such highly accurate masses and ages has been demonstrated by Brogaard et al. (2012) in estimating a cluster’s age using binary members as well as by Gaulme et al. (2013) and Brogaard et al. (2018) in estimating age with eclipsing binary systems.

As long as we can get APOGEE spectra for giants with such alternative mass measurements, we can include them in the training to go from the latent space to age, because we can determine their latent space parameters using the encoder (we do not need the decoder). Precise astrometry from *Gaia* can potentially detect tens of thousands of resolved wide binaries from which precise masses can be determined (e.g. Andrews, Chanamé & Agüeros 2017; Kochoska et al. 2017; Mowlavi et al. 2017; El-Badry, Rix & Heintz 2021; Gaia Collaboration 2022) for an all-sky sample.

9 CONCLUSIONS

We are living in an era of abundance asteroseismic and spectroscopic data. Surveys such as *TESS* and the upcoming PLATO mission (Rauer et al. 2014) will allow for large sets of ages for giant stars to be determined, which are essential for Galactic archaeology. At the same time, even larger spectroscopic surveys are ongoing, like SDSS-V and the upcoming 4MOST (de Jong et al. 2012) survey, which collect stellar spectra densely sampled across the sky and covering large parts of the Milky Way disc and halo. Thus, being able to determine ages using spectroscopic data allows for detailed investigations into our Galaxy’s formation and evolution.

In this paper, we have applied a well-developed methodology in deep learning called a variational encoder–decoder to create a data-driven model that can determine more precise ages from APOGEE spectra compared to other data-driven methods by leveraging available asteroseismic data, provided that there is a large sample of spectrum–light-curve pairs (which do not require age determinations). We train a model on $\sim 10\,000$ pairs of APOGEE spectra and *Kepler* light curves to reduce the dimensionality of APOGEE spectra while simultaneously extracting mass and age information without contamination from abundance information beside [Fe/H]. Reducing the dimensionality of APOGEE spectra in a latent space is crucial for being able to train an age model, because precise giant ages are rare and it is, therefore, difficult to train a complex model to infer spectroscopic age. We then trained a simple random forest model to determine ages from the latent space of the encoder–decoder model.

We showed that we are able to reduce the dimensionality of APOGEE spectra to just five dimensions for the purpose of reconstructing the relevant information in the light-curve’s PSD. The decoder is able to reconstruct the PSD in the region where pressure modes are located. From the resulting latent space, we are able to infer ages precise to ~ 22 per cent by training only with ~ 1200 stars with good ages using the latent space augmented by T_{eff} and [Fe/H]; for red clump stars we approach a precision of ~ 10 per cent. We have applied our methods to the whole APOGEE DR17 catalogue

for stars that fall within the parameter set of the training data. The resulting latent space ages are overall consistent with the data-driven spectroscopic ages from Mackereth et al. (2019), except that old stars are much older (there is no plateau at the old end as in previous work) and the $[\alpha/M]$ rich stars are significantly older than the oldest $[\alpha/M]$ poor stars. Because our latent space does not include information on $[\alpha/M]$ and only weak information about other abundance ratios, our latent space ages are independent of abundance ratios, yet we will obtain precise ages.

Using the APOGEE DR17 age catalogue that we create in this paper, we have mapped the age-abundance distribution across the radial range spanned by the Galactic disc. We find that the high $[\alpha/M]$ sequence has the same age distribution at any radius, extending from ages of ≈ 12 Gyr to ≈ 8 Gyr; at younger ages, we find a small fraction of young, high $[\alpha/M]$ stars similar to what has previously been found. The low $[\alpha/M]$ disc is younger than ≈ 8 Gyr everywhere, with a radial gradient in the oldest low $[\alpha/M]$ stars: the outer disc ($R \gtrsim 10$ kpc) is almost entirely low $[\alpha/M]$ and younger than ≈ 5 Gyr. The high and low $[\alpha/M]$ discs appear to be temporally separated, with the high $[\alpha/M]$ disc representing the early evolution of the disc that transition to the later low $[\alpha/M]$ evolution ≈ 8 Gyr ago.

The PSD reconstruction provided by our encoder–decoder has interesting applications of its own. For example, it can be used as a ‘sanity check’ for the light curve, because we can quickly check whether the directly observed PSD deviates greatly from that reconstructed from the APOGEE spectra. Any deviations could be due to observational or reduction issues in the light curves or they could result from information in the PSD that is not predictable from photospheric observations like stellar spectra. Examples of the latter are mode mixing or strong internal magnetic fields (e.g. Fuller et al. 2015).

In the near future, the all-sky *TESS* light curves provide a great opportunity to expand this model especially with the two ecliptic continuous viewing zones. Proposed mission like PLATO and HAYDN (Miglio et al. 2021a) can provide even more accurate ages to use in the latent space age training set, as can more detailed asteroseismic modelling of individual modes in current data (e.g. Montalbán et al. 2021). More generally, advances in machine learning using artificial neural network allows more opportunities for training with cross-domain data with few available labels. In astronomy, it is very common to have observations across multiple-domain, for example in multimessenger astronomy (e.g. the famous multimessenger gravitational wave event GW170817; Abbott et al. 2018). Many of the objects on the sky are observed in multiple large-scale surveys; stellar spectra and light curves as an example in this paper. The large amount of overlap between surveys in different domains can be exploited using methods such as the one used in this paper, because they contain more information than using one survey in one domain alone.

ACKNOWLEDGEMENTS

We thank the anonymous referee for helpful and constructive comments. HL and JB acknowledge financial support from NSERC (funding reference number RGPIN-2020-04712) and an Ontario Early Researcher Award (ER16-12-061). AM acknowledges support from the ERC Consolidator Grant funding scheme (project ASTER-OCHRONOMETRY, G.A. n. 772293).

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 website is www.sdss.org.

DATA AVAILABILITY

The basic data described in Section 4 is publicly available, with some data available in the `astroNN` DR17 VAC available as part of SDSS-IV’s DR17. All of the code underlying this article and neural network models are available in a GitHub repository at https://github.com/henrysky/astroNN_ages. Tutorials are provided in the GitHub repository on how to load and use the models trained for this paper, including predicting a corresponding PSD from an APOGEE spectrum, getting a latent representation of an APOGEE spectrum, sampling a new PSD from the latent space, and performing the age determination from the latent space.

REFERENCES

- Abadi M. et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available at: <https://www.tensorflow.org/>
- Abbott B. P., et al., 2018, *Living Rev. Relat.*, 21, 3
- Abdurro’uf et al., 2022, *ApJS*, 259, 35
- Abolfathi B. et al., 2018, *ApJS*, 235, 42
- Andrews J. J., Chanamé J., Agüeros M. A., 2017, *MNRAS*, 472, 675
- Angus R., Morton T., Aigrain S., Foreman-Mackey D., Rajpaul V., 2018, *MNRAS*, 474, 2094
- Appourchaux T., 2020, *A&A*, 642, A226
- Astropy Collaboration 2022, *ApJ*, 935, 167
- Auvergne M. et al., 2009, *A&A*, 506, 411
- Beane A., Ness M. K., Bedell M., 2018, *ApJ*, 867, 31
- Bedding T. R. et al., 2011, *Nature*, 471, 608
- Bennett M., Bovy J., 2019, *MNRAS*, 482, 1417
- Blancato K., Ness M., Huber D., Lu Y., Angus R., 2020, preprint (arXiv:2005.09682)
- Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
- Blanton M. R., et al., 2017, *AJ*, 154, 28
- Bonaca A. et al., 2020, *ApJ*, 897, L18
- Borucki W. J. et al., 2010, *Science*, 327, 977
- Bovy J., 2015, *ApJS*, 216, 29
- Bovy J. et al., 2014, *ApJ*, 790, 127
- Bovy J., Leung H. W., Hunt J. A. S., Mackereth J. T., García-Hernández D. A., Roman-Lopes A., 2019, *MNRAS*, 490, 4740
- Brogaard K. et al., 2012, *A&A*, 543, A106
- Brogaard K. et al., 2018, *MNRAS*, 476, 3729
- Brown T. M., Gilliland R. L., Noyes R. W., Ramsey L. W., 1991, *ApJ*, 368, 599
- Casey A. R., Hogg D. W., Ness M., Rix H.-W., Ho A. Q. Y., Gilmore G., 2016, preprint (arXiv:1603.03040)
- Cheng T.-Y., Li N., Conselice C. J., Aragón-Salamanca A., Dye S., Metcalf R. B., 2020, *MNRAS*, 494, 3750
- Chiappini C. et al., 2015, *A&A*, 576, L12
- Cho K., van Merriënboer B., Bahdanau D., Bengio Y., 2014, preprint (arXiv:1409.1259)
- Ciučă I., Kawata D., Miglio A., Davies G. R., Grand R. J. J., 2021, *MNRAS*, 503, 2814
- Corsaro E. et al., 2017, *A&A*, 605, A3
- Davies G. R. et al., 2016, *MNRAS*, 456, 2183
- da Silva L. et al., 2006, *A&A*, 458, 609
- de Jong R. S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460T
- de Mijolla D., Ness M. K., Viti S., Wheeler A. J., 2021, *ApJ*, 913, 12
- De Silva G. M. et al., 2015, *MNRAS*, 449, 2604
- Doersch C., 2016, preprint (arXiv:1606.05908)
- Edvardsson B., Andersen J., Gustafsson B., Lambert D. L., Nissen P. E., Tomkin J., 1993, *A&A*, 275, 101

- El-Badry K., Rix H.-W., Heintz T. M., 2021, *MNRAS*, 506, 2269
- Fabbro S., Venn K. A., O’Brian T., Bialek S., Kieley C. L., Jahandar F., Monty S., 2018, *MNRAS*, 475, 2978
- Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
- Frontera-Pons J., Sureau F., Bobin J., Le Floc’h E., 2017, *A&A*, 603, A60
- Fuller J., Cantiello M., Stello D., Garcia R. A., Bildsten L., 2015, *Science*, 350, 423
- Gaia Collaboration 2022, preprint (arXiv:2206.05595)
- Gaia Collaboration 2016, *A&A*, 595, A1
- García R. A., Ballot J., 2019, *Living Rev. Sol. Phys.*, 16, 4
- García Pérez A. E. et al., 2016, *AJ*, 151, 144
- García R. A. et al., 2011, *MNRAS*, 414, L6
- Gaulme P., McKeever J., Rawls M. L., Jackiewicz J., Mosser B., Guzik J. A., 2013, *ApJ*, 767, 82
- Gheller C., Vazza F., 2022, *MNRAS*, 509, 990
- Grosjean M., Dupret M. A., Belkacem K., Montalbán J., Samadi R., Mosser B., 2014, *A&A*, 572, A11
- Hawkins K., Ting Y.-S., Walter-Rix H., 2018, *ApJ*, 853, 20
- Hayden M. R. et al., 2015, *ApJ*, 808, 132
- Haywood M., Di Matteo P., Lehnert M. D., Katz D., Gómez A., 2013, *A&A*, 560, A109
- He X.-J., Luo A. L., Chen Y.-Q., 2022, *MNRAS*, 512, 1710
- Hekker S., Johnson J. A., 2019, *MNRAS*, 487, 4343
- Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, *Nature*, 563, 85
- Hon M., Stello D., Zinn J. C., 2018, *ApJ*, 859, 64
- Hon M. et al., 2021, *ApJ*, 919, 131
- Hon M., Kuzslawicz J. S., Huber D., Stello D., Reyes C., 2022, *AJ*, 164, 135
- Howell S. B. et al., 2014, *PASP*, 126, 398
- Huber D., Stello D., Bedding T. R., Chaplin W. J., Arentoft T., Quirion P. O., Kjeldsen H., 2009, *Commun. Asteroseismol.*, 160, 74
- Huber D. et al., 2010, *ApJ*, 723, 1607
- Jofre P. et al., 2023, *A&A*, 671, A21
- Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
- Kingma D. P., Welling M., 2013, *Auto-Encoding Variational Bayes*. Available at: <https://arxiv.org/abs/1312.6114>
- Kjeldsen H., Bedding T. R., 1995, *A&A*, 293, 87
- Kochoska A., Mowlavi N., Prša A., Lecoœur-Taïbi I., Holl B., Rimoldini L., Süveges M., Eyer L., 2017, *A&A*, 602, A110
- Kollmeier J. A. et al., 2017, preprint (arXiv:1711.03234)
- Kramer M. A., 1991, *AICHE J.*, 37, 233
- LeCun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D., 1989, *Neural Comput.*, 1, 541
- Lecun Y., Bengio Y., Hinton G., 2015, *Nature*, 521, 436
- Leung H. W., Bovy J., 2019a, *MNRAS*, 483, 3255
- Leung H. W., Bovy J., 2019b, *MNRAS*, 489, 2079
- Leung H. W., Bovy J., Mackereth J. T., Hunt J. A. S., Lane R. R., Wilson J. C., 2023, *MNRAS*, 519, 948
- Lightkurve Collaboration 2018, Lightkurve: *Kepler* and *TESS* time series analysis in Python, Astrophysics Source Code Library, record ascl:1812.013
- Lomb N. R., 1976, *Ap&SS*, 39, 447
- Lu Y. L., Ness M. K., Buck T., Zinn J. C., Johnston K. V., 2022, *MNRAS*, 512, 2890
- Mackereth J. T. et al., 2019, *MNRAS*, 489, 176
- Mackereth J. T. et al., 2021, *MNRAS*, 502, 1947
- Majewski S. R. et al., 2017, *AJ*, 154, 94
- Martig M. et al., 2015, *MNRAS*, 451, 2230
- Martig M. et al., 2016, *MNRAS*, 456, 3655
- Masseron T., Gilmore G., 2015, *MNRAS*, 453, 1855
- McQuillan A., Aigrain S., Mazeh T., 2013, *MNRAS*, 432, 1203
- Metcalfe T. S. et al., 2014, *ApJS*, 214, 27
- Miglio A. et al., 2012, *MNRAS*, 419, 2077
- Miglio A. et al., 2021a, *Exp. Astron.*, 51, 963
- Miglio A. et al., 2021b, *A&A*, 645, A85
- Montalbán J. et al., 2021, *Nat. Astron.*, 5, 640
- Mosser B. et al., 2014, *A&A*, 572, L5
- Mowlavi N. et al., 2017, *A&A*, 606, A92
- Nair V., Hinton G. E., 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML’10. Omnipress, Madison, WI, USA, p. 807
- 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
- Ness M. K., Silva Aguirre V., Lund M. N., Cantiello M., Foreman-Mackey D., Hogg D. W., Angus R., 2018, *ApJ*, 866, 15
- Ness M. K., Johnston K. V., Blancato K., Rix H. W., Beane A., Bird J. C., Hawkins K., 2019, *ApJ*, 883, 177
- O’Brian T., Ting Y.-S., Fabbro S., Yi K. M., Venn K., Bialek S., 2021, *ApJ*, 906, 130
- Paczyński B., Stanek K. Z., 1998, *ApJ*, 494, L219
- Patil A. A., Eadie G. M., Speagle J. S., Thomson D. J., 2022, preprint (arXiv:2209.15027)
- Pedregosa F. et al., 2012, preprint (arXiv:1201.0490)
- Pinsonneault M. H. et al., 2014, *ApJS*, 215, 19
- Pinsonneault M. H. et al., 2018, *ApJS*, 239, 32
- Pogge R. W. et al., 2020, in Evans C. J., Bryant J. J., Motohara K., eds, Proc. SPIE Conf. Ser. Vol. 11447, Ground-based and Airborne Instrumentation for Astronomy VIII. SPIE, Bellingham, p. 1144781
- Portillo S. K. N., Parejko J. K., Vergara J. R., Connolly A. J., 2020, *AJ*, 160, 45
- Rauer H. et al., 2014, *Exp. Astron.*, 38, 249
- Ricker G. R. et al., 2015, *J. Astron. Telesc. Instrum. Syst.*, 1, 014003
- Rix H.-W., Bovy J., 2013, *A&AR*, 21, 61
- Rodrigues T. S. et al., 2014, *MNRAS*, 445, 2758
- Rodrigues T. S. et al., 2017, *MNRAS*, 467, 1433
- Scargle J. D., 1982, *ApJ*, 263, 835
- Sedaghat N., Mahabal A., 2018, *MNRAS*, 476, 5365
- Serenelli A. et al., 2017, *ApJS*, 233, 23
- Silva Aguirre V. et al., 2020, *ApJ*, 889, L34
- Stello D., Chaplin W. J., Basu S., Elsworth Y., Bedding T. R., 2009, *MNRAS*, 400, L80
- Ting Y.-S., Hawkins K., Rix H.-W., 2018, *ApJ*, 858, L7
- Ting Y.-S., Conroy C., Rix H.-W., Cargile P., 2019, *ApJ*, 879, 69
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I., 2017, preprint (arXiv:1706.03762)
- Xiang M., Rix H.-W., 2022, *Nature*, 603, 599
- Xiang M. et al., 2017, *ApJS*, 232, 2
- Xiang M. et al., 2019, *ApJS*, 245, 34
- Yu J., Huber D., Bedding T. R., Stello D., Hon M., Murphy S. J., Khanna S., 2018, *ApJS*, 236, 42
- Zhang M. et al., 2021, *ApJ*, 922, 145

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