Unveiling the distinct formation pathways of the inner and outer discs of the Milky Way with Bayesian Machine Learning

Ciuc, Ioana; Kawata, Daisuke; Miglio, Andrea; Davies, Guy R.; Grand, Robert J. J.

DOI:
10.1093/mnras/stab639

MNRAS 000, 1–11 (2020) Preprint 9 March 2021 Compiled using MNRAS LATEX style file v3.0

Unveiling the Distinct Formation Pathways of the Inner and Outer Discs of the Milky Way with Bayesian Machine Learning

Ioana Ciucă[1],★ Daisuke Kawata[1], Andrea Miglio[2], Guy R. Davies[2], Robert J.J. Grand[3]

[1] *Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK*
[2] *Department of Physics, University of Birmingham*
[3] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We develop a Bayesian Machine Learning framework called BINGO (Bayesian INference for Galactic archaeOlogy) centred around a Bayesian neural network. After being trained on the APOGEE and *Kepler* asteroseismic age data, BINGO is used to obtain precise relative stellar age estimates with uncertainties for the APOGEE stars. We carefully construct a training set to minimise bias and apply BINGO to a stellar population that is similar to our training set. We then select the 17,305 stars with ages from BINGO and reliable kinematic properties obtained from *Gaia* DR2. By combining the age and chemo-kinematical information, we dissect the Galactic disc stars into three components, namely, the thick disc (old, high-[α/Fe], [α/Fe] $\gtrsim$ 0.12), the thin disc (young, low-[α/Fe]) and the Bridge, which is a region between the thick and thin discs. Our results indicate that the thick disc formed at an early epoch only in the inner region, and the inner disc smoothly transforms to the thin disc. We found that the outer disc follows a different chemical evolution pathway from the inner disc. The outer metal-poor stars only start forming after the compact thick disc phase has completed and the star-forming gas disc extended outwardly with metal-poor gas accretion. We found that in the Bridge region the range of [Fe/H] becomes wider with decreasing age, which suggests that the Bridge region corresponds to the transition phase from the smaller chemically well-mixed thick to a larger thin disc with a metallicity gradient.

Key words: Galaxy: formation – Galaxy: abundances – asteroseismology

1 INTRODUCTION

The Galactic disc is traditionally separated into the geometric thick and thin disc after Gilmore & Reid (1983) found from star counts that the vertical density profile of the Milky Way was better characterised by a superposition of two exponential profiles rather than one. High-resolution spectroscopic studies of the solar neighbourhood revealed also a bimodality in the chemistry of the disc, with the [α/Fe]-[Fe/H] distribution showing distinct high- and low-[α/Fe] components and a less prominent intermediate region (e.g., Fuhrmann 1998; Prochaska et al. 2000). Beyond the local disc, recent large-scale spectroscopic surveys, such as the Apache Point Observatory Galactic Evolution Experiment (APOGEE), confirmed the existence of a similar high-[α/Fe] sequence spanning a large radial and vertical extent of the Milky Way disc (e.g., Anders et al. 2014; Nidever et al. 2014; Hayden et al. 2015; Queiroz et al. 2019). The high-[α/Fe] disc also appears to be thicker and more centrally concentrated than its low-[α/Fe] counterpart (e.g., Bensby et al. 2011; Bovy et al. 2012; Cheng et al. 2012).

One of the first approaches to explain the chemical bimodality seen in the Galactic disc is the two-infall model, a numerical chemical evolution model developed by Chiappini et al. (1997), Chiappini et al. (2001), Grisoni et al. (2017), Spitoni et al. (2019). Chiappini et al. (2001) suggested that the high-[α/Fe], chemically homogenous disc forms early during an intense star formation period dominated by Type II supernovae (SNe II) following a rapid infall of primordial gas. After a brief cessation in star formation, the second episode of gas accretion takes place that lowers the metal content in the interstellar medium due to the continuous infall of low metallicity fresh gas. The low-[α/Fe] disc then builds up gradually from lower [Fe/H]. Bekki & Tsujimoto (2011) also follow a semi-analytical approach to explain the existence of two distinct populations. In their continuous star formation model, the high-[α/Fe] sequence up to around solar [Fe/H], i.e. the thick disc, forms early during a rapid, intense star formation period. The thin disc then proceeds to form gradually from the remaining gas with solar [Fe/H] and [α/Fe] mixed with the fresh primordial gas accreted after the formation

★ E-mail: ioana.ciuca.16@ucl.ac.uk

of the thick disc.A sequence of increasing [Fe/H] and decreasing [α/Fe] builds up gradually. Still, this sequence is lower in [α/Fe] as Type Ia SNe can already enrich the environment at this time. Once star formation reaches its peak and starts decreasing, a sequence with decreasing [α/Fe] and increasing in [Fe/H] follows along with the same low-[α/Fe] sequence.

More recent scenarios inspired by galactic dynamics proposed that radial migration of kinematically hot stars formed in the inner disc builds up a thick disc after moving outward in the disc (Schönrich & Binney 2009; Loebman et al. 2011; Roškar et al. 2013). Radial migration is successful in explaining the age-metallicity and metallicity-rotation velocity relation observed in the Milky Way. However, there is still considerable debate regarding the efficiency of radial migration in building a geometrically thick disc (e.g., Minchev et al. 2012; Grand et al. 2016; Kawata et al. 2017).

High-resolution numerical simulations also suggested several thick and thin disc formation scenarios, including violent gas-rich mergers at high-redshift (e.g., Brook et al. 2004; Grand et al. 2018, 2020), accretion of high-[α/Fe] stars (Abadi et al. 2003; Kobayashi & Nakasato 2011; Tissera et al. 2012), vertical heating from satellite merging events (e.g., Quinn et al. 1993; Villalobos & Helmi 2008) and turbulence in clumpy high-redshift gas-rich disc (Noguchi 1998; Bournaud et al. 2009; Beraldo e Silva et al. 2020). The recent popular view is that the thick disc formation precedes the thin disc formation and the earlier disc was smaller and thicker, i.e. an inside-out and upside-down formation of the disc (e.g., Brook et al. 2004, 2006; Bird et al. 2013). In Brook et al. (2012), the majority of the thick-disc stars form as gas originating from a gas-rich merger at high-redshift settles into a disc at the end of the merger epoch. This early disc is kinematically hot and radially compact. Once the chaotic phase of the star formation of the thick disk ends, the younger, lower [α/Fe] thin disc can gradually grow in an inside-out fashion (e.g., Matteucci & Francois 1989) as gas is smoothly accreting to the central galaxy. As in Brook et al. (2012), Noguchi (2018) and Grand et al. (2018) suggested that chemical evolution proceeds at different rates in the inner and outer disc, resulting in more chemically evolved stars in the inner regions. Radial migration can bring the thick disk stars formed in the inner disc to the outer disc at redshift $z \sim 0$, so that we can observe the thick disc stars at the solar neighbourhood (Brook et al. 2012; Minchev et al. 2013).

The *Gaia* mission (DR2, Gaia Collaboration et al. 2018) is providing information to obtain the accurate position and motion for more than a billion stars in the Milky Way. The APOKASC-2 catalogue (Pinsonneault et al. 2018), comprised of 6,676 evolved stars in the APOGEE DR14 survey observed by the *Kepler* mission (Borucki et al. 2010), provides the best asteroseismology information to infer the age for giant stars, which is crucial for Galactic archaeology. In this paper, we use a state-of-the-art machine learning method, a Bayesian neural network, trained on the APOKASC-2 data, to obtain reliable relative stellar age estimates for 17,305 carefully selected disc stars in the APOGEE data. We use the age, chemistry and kinematical information to examine the formation history of the Galactic disc by comparing our results with what is expected from the formation scenarios of the thick and thin disc suggested by the recent numerical simulations described above.

This paper is organised as follows. Section 2 describes the Bayesian Machine Learning framework, called BINGO (Bayesian INference for Galactic archaeOlogy), that we employ in the current analysis. We discuss here how the biases in the training dataset affect the performance of the neural network model and our approach to minimise the bias in the subsequent inferences. In Section 3, we present the results after applying BINGO to carefully selected stars in the APOGEE survey. A brief discussion of our results is given in Section 4. Finally, a summary of our findings is given in Section 5.

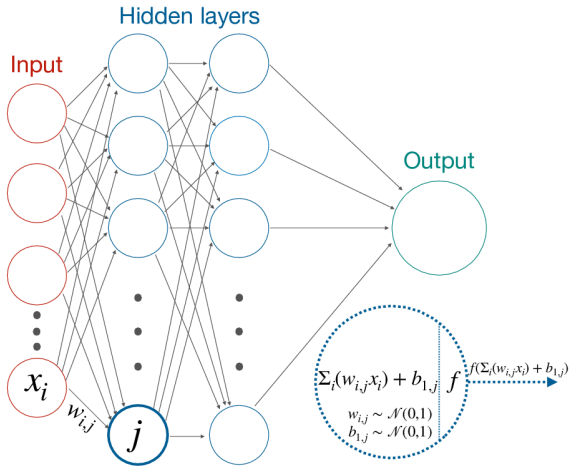

Figure 1. Schematics of a Bayesian neural network with 2 hidden layers. Each connection between neurons has an associated weight and the neurons in the the hidden and output layers have an associated bias. The connection between neurons i in the input layer and j in the first hidden layer has the associated weight $w_{i,j}$ and the neuron j has a bias $b_{1,j}$. Each weight and bias parameters have an associated prior $\mathcal{N}(0,1)$. The dotted circle is a zoom-in of the neuron j, and shows the transformation applied to the input data x_i in the first hidden layer of the network, namely $x_i \rightarrow f\,(\Sigma_i\,(w_{i,j}\,x_i) + b_{1,j})$, where f is the activation function. We use a rectified linear unit (ReLU) activation function in our analysis.

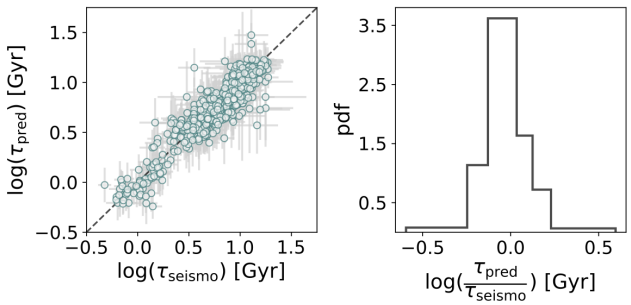

Figure 2. Comparison between observed (target) logarithmic age of stars, $\log(\tau_{\rm seismo})$, derived with asteroseismology and the predicted $\log(\tau_{\rm pred})$ by BINGO. The panels show the results when applying the model trained with the age data augmented training set with the distance shuffling (Model A) to the original test data (Test 1, see Sec. 2.2 for details). The light green circles in the left panel show the model prediction results against the observed target age. The standard deviation in the prediction and observation are shown as the grey lines. The right panel shows the difference between prediction and target, which peaks at 0 with a standard deviation of 0.1 dex.

2 METHOD

In this paper, we introduce BINGO which is a Bayesian Machine Learning framework to obtain stellar ages of evolved stars using photometric information from the second data release of the European Space Agency's (ESA) *Gaia* mission (*Gaia* DR2, Gaia Collaboration et al. 2018) and the stellar parameter information from the

fourteenth data release of the SDSS-IV APOGEE-2 (Majewski et al. 2017). BINGO consists of a Bayesian neural network trained using the asteroseismic age determined from $<\Delta\nu>$ from the individual radial-mode frequency from the *Kepler* light curve (Miglio et al. 2020).

Gaia DR2 provides astrometric information to obtain the position and proper motion for ~ 1.3 billion stars with unprecedented accuracy (Lindegren et al. 2018) as well as high-quality multi-band photometry for a large subset of these stars (Riello et al. 2018; Evans et al. 2018). For a selected type of stars with a G-band magnitude between about 4 and 13 magnitudes, the mean line-of-sight velocities measured with *Gaia* Radial Velocity Spectrometer (RVS), line-of-sight velocities have also been provided in *Gaia* DR2 (Cropper et al. 2018; Sartoretti et al. 2018; Katz et al. 2019). We use the photometric data from *Gaia* DR2 for BINGO, and the parallax and proper motion information to derive the kinematic properties for our sample of stars.

APOGEE is a spectroscopic survey in the near-infrared H-band (15,200 Å-16,900 Å) with a high resolution of R ~ 22,500, observing more than 200,000 stars (as of DR14) located primarily in the disc and bulge of the Milky Way. In this work, we employ the calibrated stellar parameters such as effective temperature and surface gravity as well as metal abundances obtained with the APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASCAP, García Pérez et al. 2016) in the APOGEE DR14 survey (Abolfathi et al. 2018). In addition, we use the 2MASS J, H, K photometry and their associated uncertainties (Skrutskie et al. 2006) reported in the APOGEE DR14 catalogue.

2.1 BINGO

Machine Learning has revolutionised the way we perform data analysis tasks in Astronomy, which has grown into a big data field with the emergence of large surveys such as SDSS and *Gaia*. Neural networks are Machine Learning methods that can, in principle, model any smooth map between a high-dimensional input data to a set of desirable outputs. Depending on their architecture, neural networks can consist of one or more fully-connected layers, each with a number of neurons that essentially take the input and transform it through linear activation functions to an output of interest (also known as feed-forward artificial neural networks). In supervised learning, which BINGO uses, the parameters of the neural network, e.g. weights that define the connection between neurons, are trained and optimised to best reproduce the training set where the input and output are known. Then, the trained neural network can be applied to the data whose output is unknown with much less computational cost than training.

In Bayesian Inference, the power of Bayes' Law is that it allows us to relate the probability of a model given the data to a quantity that is easier to understand, namely the probability we would observe the data given the model and any background information, I, i.e.,

$$p(model|data, I) \propto p(data|model, I)p(model|I), \quad (1)$$

where $p(model|data, I)$ is the posterior probability, $p(data|model, I)$ is the likelihood and $p(model|I)$ is the prior. The posterior encompasses our state of knowledge about a model given that we gather new data through the likelihood. Following equation (1), Bayes' Law can be applied to a neural network to come up with a probability distribution over its model parameters[1] and construct a Bayesian Neural Network as done in the pioneering work of Das & Sanders (2018) and Sanders & Das (2018). This powerful synergy between Bayesian Inference and Machine Learning allows us to naturally introduce uncertainty into our machine learning approach, i.e. we can get an estimate of how confident our neural network is of its predictions.

BINGO's architecture consists of 2 fully connected layers with 16 neurons each (Fig. 1). We use the probabilistic programming framework `pymc3` (Salvatier et al. 2015) and its Magic Inference Button, the No-U-Turn-Sampler (NUTS) as the MCMC sampler. We use a Gaussian prior of $\mathcal{N}(0, \sigma)$ for the weights and bias parameters in the neural network, which effectively acts as L2 regularisation. It is possible to optimise σ of the Gaussian prior, but it is computationally too expensive. Therefore, we adopt $\sigma = 1$ for simplicity. We use 4 chains that allow us to diagnose our samples and make sure the samples returned from the NUTS sampler are drawn from the target distribution. Once we have a posterior distribution over the neural network parameters, we can then compute a distribution over the network outputs by marginalising over the network parameters. We note that this Bayesian Neural Network scheme assumes that all the input features, such as the stellar parameters, are independent, and cannot take into account the covariance between the inputs. It is also worth noting that the neural network model depicted in Fig. 1 is not identifiable (Pourzanjani et al. 2017). Hence, the naive MCMC sampling of the network parameters suffers from the unidentifiability of the parameters. Still, we have confirmed that the posterior distribution of the target age prediction from the 4 different chains are consistent with each other. Therefore, we are confident that our age prediction, especially the mean of the prediction used in this paper, does not suffer severely from unidentifiability. These known challenges for Bayesian Neural Networks remain caveats of BINGO, upon which we hope to improve in a future study.

2.2 Building an effective training set

In this study, we employ a training set created from the APOKASC-2 dataset with our derived asteroseismic age (Miglio et al. 2020). The assumptions used to derive the asteroseismic ages that we are using are given as R11 in Table 1 of Miglio et al. (2020). We select only red clump stars (RC) with masses higher than 1.8 $M_\odot$ and the red giant branch (RGB) stars, for which the relative asteroseismic ages are reliable. To construct our base training set, we use only stars with high signal-to-noise (SNR) APOGEE spectra (SNR > 100), which leaves us with 2,915 stars. We then use the APOGEE stellar parameters and photometry data, $T_{\rm eff}$, log g, [α/M], [M/H] [2], [C/Fe], [N/Fe], G, BP, RP, J, H and K as the input features in BINGO to map them to the common logarithm of the asteroseismic age, $log(\tau)$, referred to as the target.

Because the original data comes from a limited *Kepler* field data, our original training set has a known dependence of age and metallicity on the distance (which affects photometry). Also, there are not many young or old stars in our selected RGB and RC data. To correct for the distance dependence, we randomly displace the distance of stars between 0 and 10 kpc and then adjust the apparent

[1] `https://twiecki.io/blog/2016/07/05/bayesian-deep-learning/`

[2] In APOGEE DR14, α-elements comprise of O, Mg, Si, S, Ca and Ti. We used the ASPCAP measurements of [α/M] and [M/H] as a proxy for [α/Fe] and [Fe/H], respectively. Correspondingly, we use the labels of [α/Fe] and [Fe/H] to refer to [α/M] and [M/H].

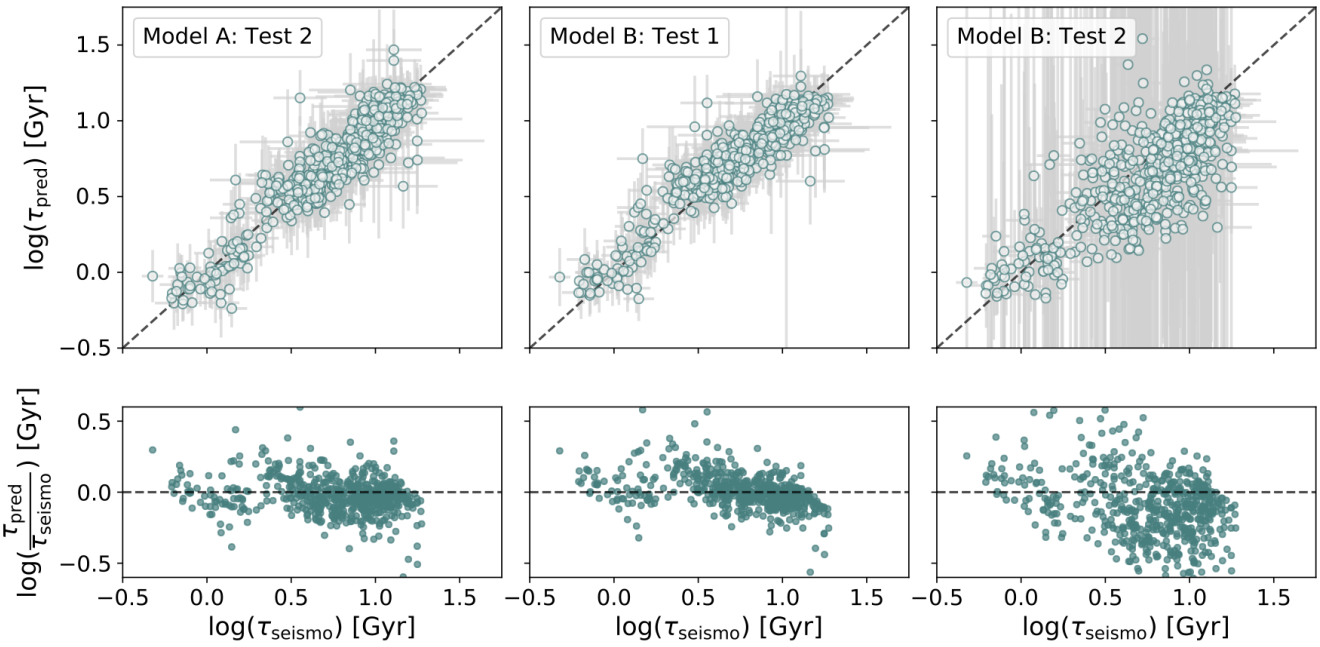

Figure 3. Predictions vs the target asteroseismic age in $\log(\tau)$. Left panel shows the result from a model trained on the age data augmented training set with the distance shuffling (Model A) and applied to the distance shuffled test set (Test 2). The middle and right panel show the predictions for a model trained on the original data (Model B) applied to the original (Test 1) and distance shuffled test data (Test 2), respectively. The standard deviation in the prediction and asteroseismic age are shown as the grey lines. Model A predictions for Test 2 performs better than Model B prediction for Test 1. The model trained on the original data and applied to the distance shuffle data, i.e. Model B prediction for Test 2, performs considerably worse as Model B has learned the distance dependence of age and metallicity, which is erased in Test 2.

magnitude of the stars depending on the difference between the new distance and the original distance. We do not change the extinction upon displacing the distance also to erase the dependence of extinction on the distance. We refer to this technique as distance shuffling.

Our training set contains a smaller number of young (age < 2 Gyr) stars and very old (age > 12 Gyr) stars, and this imbalance becomes more apparent when using $\log(\tau)$ as our target variable for BINGO. During training, the model learns to reproduce the target variable only for a majority of intermediate age stars, which biases the prediction toward the intermediate age irrespective of their true age, and consequently leads to overpredicting the age of younger stars and underpredicting the age of the very old stars, an effect also known as regression dilution. To minimize the effect of this bias and balance our training set, we effectively oversample the fewer young stars and very old stars to balance the number of stars at different $\log(\tau)$. To this end, we first examine the distribution of our original training set in $\log(\tau)$. We then use a Kernel Density Estimator (KDE) to approximate the distribution in $\log(\tau)$, and for each star, we find its probability under the KDE, which we refer to as $prob$. We then compute the inverse probability and round it the nearest integer, N = $[1/prob]$. Following the distance shuffling procedure described above, we randomly distance-shuffle each star N times. This approach leads to some of the stars in the original dataset to be sampled more than once. Since their distances and hence their apparent magnitude are different, these "artificial" stars become members of an augmented training set. Since we are using data augmentation, which is an established machine learning technique, we refer to our approach as age data augmentation. The final training set has 4,673 stars after performing the age data augmentation technique on the training set data (80% of the original data). Note that the data augmentation can reduce the uncertainties in our predictions, because we artificially increase the number of data points. Therefore, our uncertainties do not statistically reflect the uncertainty in the measurement of the stellar age. In this paper, however, our priority is to mitigate regression dilution with this simple data augmentation technique. This is another caveat of BINGO in addition to the assumed independence of the input features and the unidentifiability discussed above. In this paper, as described later, we use the uncertainties only as the metric of confidence of our prediction, and do not use the uncertainties for any quantitative discussion. Hence, the discussion of this paper is unlikely to be affected by these issues. We postpone the resolution of these issues to a future study.

To evaluate the prediction accuracy of BINGO, we split our original data of 2,915 stars into training (80 %, 2,331 stars) and testing (20 %, 583 stars) data. To demonstrate the importance of distance shuffling and age data augmentation, we consider two different trained models: Model A trained on the age data augmented training set of 4,673 stars with the distance shuffling and Model B trained on the original training set without the distance shuffling or the age data augmentation. Then, we create a testing set, Test 1, which is the 20% of the original data which are not used for training, and Test 2 which is the same data as Test 1, but the distance has been shuffled. Fig. 2 shows the predictions from Model A on Test 1 and the error associated with the prediction. The asteroseismic age is well reproduced by the prediction from BINGO Model A, with a standard deviation ~ 0.1 dex. Note that the ages of some of old stars are much older than the age of the Universe. This is because there

is no prior of the maximum age considered in our asteroseismic age measurement (Miglio et al. 2020).

Fig. 3 presents the predictions from Model A on Test 2 (left), from Model B on Test 1 (middle) and Test 2 (right). There is little difference between Model A on Test 1 (see Fig. 2 and Model A on Test 2. This means that BINGO Model A can recover the age well in application data which have no distance dependence in age or metallicity. The middle panel of Fig. 2 shows that Model B trained on the original dataset without the age data augmentation leads to a systematic overprediction for the age of stars with the asteroseismic age of $\log(\tau_{\rm seismo}) < 0.5$ dex and underprediction of the age for stars with $\log(\tau_{\rm seismo}) > 1.0$ dex. This is because Model B is trained mainly to reproduce the overwhelming number of stars with $0.5 < \log(\tau_{\rm seismo}) < 1.0$ dex and suffers from the regression dilution effect mentioned above. The right panel of Fig. 3 shows the age prediction of Model B on Test 2, which shows much worse recovery of the asteroseismic ages with large uncertainties. This is because Model B has learned the dependence of the age and metallicity on the distance in the original training set. These results demonstrate why it is important to erase the distance dependence in the training set and keep the balance of the number of sample in the output label, i.e., $\log(\tau)$. We therefore use Model A in this paper.

2.3 RGB and high mass RC selection

Our training set consists of the specific population of RC stars with a mass higher than 1.8 $M_\odot$ and RGB stars in the limited *Kepler* field. When we apply our trained model to the rest of APOGEE data, we select only the same population as the population of the training data. Hence, we train a 3-layer artificial neural network on the original APOKASC-2 data to classify RC stars with a mass higher than 1.8 $M_\odot$ and RGB stars. For this classification task, we train the model using `Keras` and `TensorFlow` (Abadi et al. 2016), which is much less computationally expensive than training a Bayesian Neural Network. The selection function of APOKASC-2 is not the same as the rest of the APOGEE data. However, because we need the stellar mass and the RC, RGB and AGB classification for the training and validation data, we use the APOKASC-2 data for training and validation. We have constructed a classification neural network to identify the RC stars with > 1.8 $M_\odot$ and RGB stars using our asteroseismic analysis of the APOKASC-2 data. We used the input features of $T_{\rm eff}$, log g, [α/Fe], [Fe/H], [C/Fe], [N/Fe], G, BP, RP, J, H and K, and 2,948 positive, i.e. the high mass RC or RGB, and 1,918 negative stars are used for the training. Similar strategies are employed in Hawkins et al. (2018) and Ting et al. (2018) to identify the RC stars.

We then use the trained neural network model to classify stars in the APOGEE cross-matched with *Gaia* DR2 dataset (Sanders & Das 2018). We also limit our data to having APOGEE spectra with SNR > 100 and the K-band extinction smaller than 0.1 mag in the APOGEE catalogue, because all of our training data has the K-band extinction < 0.1 mag. We only select stars that have a probability higher than 95 % of being classified as RC with higher mass than 1.8 $M_\odot$ or RGB. We apply the BINGO Model A to this selected data to get the posterior probabilities for $\log(\tau)$.

Our strategy in this paper is to use the most reliable data only. We therefore select stars with $\log(\tau)$ age uncertainties less than 10 %. Note that the age uncertainties from BINGO indicate epistemic uncertainties of the model prediction, which can be smaller than the observed uncertainty of the original asteroseismic age. Also, to obtain reliable kinematic properties from the *Gaia* data, we select the data with parallax uncertainties of $\pi/\sigma_\pi > 5.0$. We

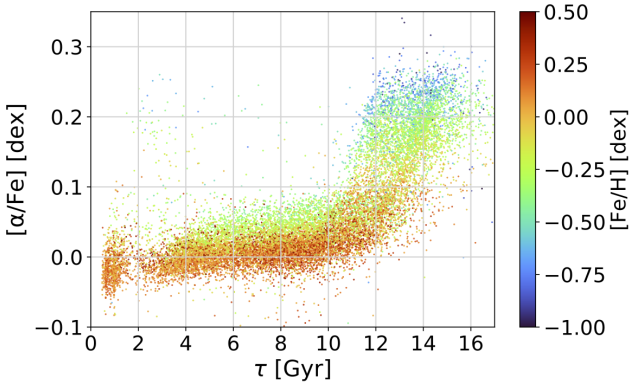

Figure 4. [α/Fe] as a function of age coloured by metallicity, [Fe/H]. The high-[α/Fe] population ([α/Fe]> 0.1 dex) is older and more metal-poor.

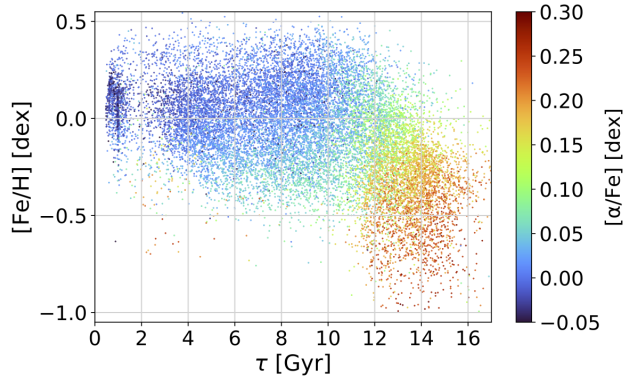

Figure 5. [Fe/H] as a function of age coloured by [α/Fe]. The younger population is more metal-rich than its older counterpart. Stars more metal-poor than −0.5 dex are considerably old.

compute the distance using the *Gaia* parallax with the additional systematic bias of parallax of 54 μas (Schönrich et al. 2019), and select the stars in the limited volume of $7 < R < 9$ kpc and $z < 2$ kpc, where we assume the solar position at the Galactocentric distance of 8 kpc and the vertical height of the Sun from the disc plane of 0.025 kpc. We obtain kinematic properties using `galpy` (Bovy 2015). We have confirmed that our derived age and kinematics are consistent with Sanders & Das (2018), except for the difference in the absolute age scales, because we use a different asteroseismic age scale for our training set (Miglio et al. 2020). As a result, we obtain 17,305 stars, which are used in the following sections.

3 RESULTS

In this section, we explore the relations between stellar age, chemistry and orbital properties for our sample of stars. Reliable relative age estimates for a large number of stars obtained with BINGO enable us to find that the inner and outer discs follow a different formation and chemical evolution pathway. Our results provide further evidence for an upside-down, inside-out formation of the Galactic disc.

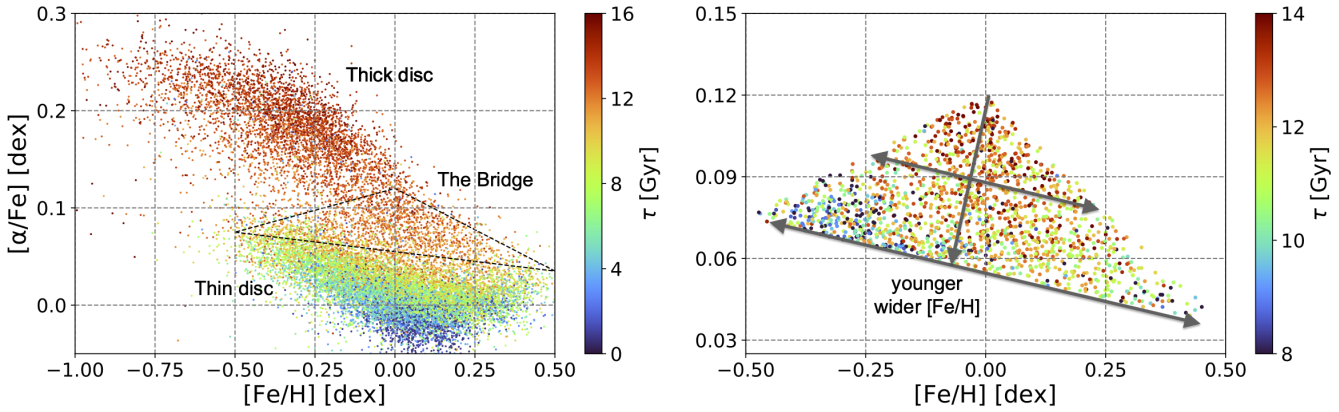

Figure 6. The distribution of [α/Fe] and [Fe/H] coloured by age for our sample of stars. We refer to the old high- and young low-[α/Fe] populations as the thick and thin disc, respectively, as highlighted in the left panel. The dotted triangle region in the left panel is referred to as the Bridge, and is a transition region between the thick (high-α sequence with [α/Fe] $\gtrsim$ 0.1 dex) and thin disc (low-[α/Fe] sequence). An age gradient is apparent, as indicated by the near vertical downward arrow, in a close-up of the Bridge region, shown in the right panel, and the range of [Fe/H] becomes wider, as indicated by the two near horizontal arrows for the younger stars in the Bridge region.

3.1 The chrono-chemical map of disc stars

We first investigate the evolution of α-abundances, [α/Fe], and metallicity, [Fe/H], with age, τ. Fig. 4 shows the enhancement in [α/Fe] as a function of age coloured by metallicity. The deficiency of stars with age $\sim$ 1.5 Gyr arises because we select the RC stars with mass > 1.8 $M_{\odot}$ and there are considerably fewer RGB stars with ages younger than 3 Gyr. The high-[α/Fe] "sequence" separates clearly from the low-[α/Fe] "sequence" in the age-[α/Fe] space at [α/Fe] $\sim$ 0.1 dex, where there seems to be a population gap extending approximately 0.02 dex. The majority of the high-[α/Fe] stars ([α/Fe] > 0.1 dex) are generally older and more metal-poor than the low-[α/Fe] population. [α/Fe] rapidly decreases with decreasing age up to $\sim$ 10 Gyr. The age-[α/Fe] relationship also appears to be broader in [α/Fe] at a fixed age for the high-[α/Fe] population, in qualitative agreement with Silva Aguirre et al. (2018). However, considering the uncertainties of the age, this relationship is considered to be tight (Haywood et al. 2013; Snaith et al. 2015). A striking feature of Fig. 4 is the young, low metallicity, high-α stars, also seen in Chiappini et al. (2015), Martig et al. (2015), Silva Aguirre et al. (2018). We discuss the origin of this population in more detail in Section 3.3.

Fig. 5 shows the age-metallicity relationship coloured with [α/Fe]. While the old, high-[α/Fe] stars exhibit a clear trend of decreasing [Fe/H] with age, the younger, low-[α/Fe] disc shows a flat age-[Fe/H] relation up to $\sim$ 11 Gyr. For stars with [Fe/H] > $-$0.5 dex, our results are qualitatively similar to those from previous studies, such as Casagrande et al. (2011), Silva Aguirre et al. (2018) and Mackereth et al. (2019). For the metal-poor and high [α/Fe] population, the tight trend observed between age and metallicity is consistent with Bensby et al. (2005) and Haywood et al. (2013), who analysed dwarf stars and used the isochrone age.

In Fig. 6 we examine the distribution in [α/Fe] and [Fe/H] coloured by age for the stars in our sample. Classically, this diagram is used to identify the chemically defined thick (high-[α/Fe]) and thin (low-[α/Fe]) disc stars. Using the age alongside the chemical information, we define three regions in the [α/Fe]-[Fe/H] space, namely an old thick disc (high-[α/Fe], low-[Fe/H]), young thin disc (low-[α/Fe], broad [Fe/H]) and the Bridge (high-[α/Fe], high-[Fe/H]). The Bridge stars[3] are selected to lie within the triangle region starting around ([Fe/H], [α/Fe]) = (0.0, 0.12) dex. Anders et al. (2018) suggested that the population of stars found in this region had a different origin and history to the thick and thin disc stars. Our results further suggest that the Bridge population appears to be a transition region connecting the old thick and the young thin disc. The right panel of Fig. 6 reveals a noticeable age gradient within this population from the oldest, more [α/Fe]-enhanced and metal-rich stars to a younger population spanning a broader distribution of metallicities from [Fe/H] = $-$0.5 to 0.5 dex. Although this age gradient in the Bridge is tentative, we notice that a similar trend is also seen in Delgado Mena et al. (2019) who studied 1,000 FGK dwarf stars from the HARPS-GTO programme and analysed the isocrhone ages of these stars. Therefore, it is reassuring that the trend shown in our asteroseismic-trained ages of giants is similar to that based on the isochrone ages of dwarfs.

3.2 Chemo-kinematical analysis

To connect the observed stellar chemical properties to kinematic properties, we compute the vertical action and the mean orbital radius for our sample of stars using `galpy` in the fixed `MWPotential2014` configuration (version 1.5, Bovy 2015). Although `MWPotential2014` is time-independent and axisymmetric and cannot capture disequilibrium effects present in the real disc, it is nonetheless a good and useful approximation for the purpose of this work. Fig. 7 shows the distribution of [α/Fe] and [Fe/H] as a function of age and vertical action J_z. The general trend is that J_z is decreasing with age, with the older population being significantly hotter than the younger population. As also inferred from Fig. 6, Fig. 7 clearly shows that the Bridge region starts appearing at age < 13 Gyr and it spreads to lower [α/Fe] and to a wider range of [Fe/H] with decreasing age as seen in the triangle features at [α/Fe] < 0.12 dex in the panels of 10 < τ < 12 Gyr. The lower panels of age lower than 10 Gyr show the dominant population of the low-[α/Fe] and

[3] There are 1,682 stars in the Bridge, which represents approximately 10% of the entire sample. Caution must be used when using this number as it can be sensitive to the selection function.

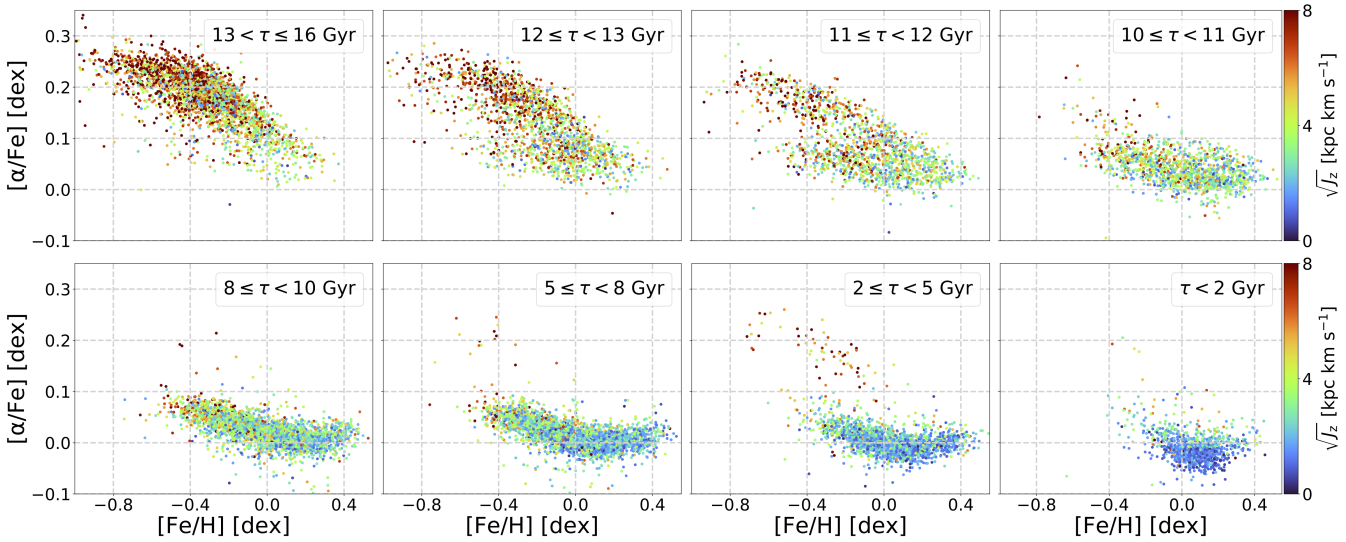

Figure 7. The distribution of [α/Fe] and [Fe/H] coloured by the square root of the vertical action, $\sqrt{J_z}$, for the samples of stars within 8 different age bins. The top four panels show [Fe/H]-[α/Fe] relationship for the older stars ($10 < \tau < 16$ Gyr), and the lower four panels present those for the younger stars ($\tau <$ 10 Gyr). There is a kinematically hot population of young high-[α/Fe] stars seen in the lower panels.

kinematically colder thin disc stars. The lower panels also reveal a small population of kinematically hot, young stars, and as high [α/Fe] as the thick disc population in Fig. 6. To understand their origin, we compare this population of stars with the thick disc stars (high [α/Fe] and old), and we discuss the results in Section 3.3.

To further examine whether or not there is a clear distinction between the high and metal-rich, low-[α/Fe] sequence, we select the high-metallicity ridge shown in the left panel of Fig. 8. The ridge is considered to represent the most advanced chemical evolution path of the stars born in the inner disc, $R < 6$ kpc, (e.g., Schönrich & Binney 2009), and in fact, as shown in the left panel of Fig. 9, their mean orbital radius is always smallest among the same [α/Fe] population. For the stars within this high-metallicity ridge, we divided the samples according to their [α/Fe]. We measure the scale height using a Markov-Chain Monte Carlo approach, where we fit the J_z distribution in each [α/Fe] bin with an isothermal profile, i.e., $p(J_z) \sim \exp(-J_z/h_{J_z})$ (Binney 2010; Binney & McMillan 2011; Ting & Rix 2019). We compute the scale height, h_{J_z}, and its uncertainty in 13 selected bins using [α/Fe]. The results, shown in the right panel of Fig. 8, reveal a smooth decrease of J_z with decreasing [α/Fe] and age indicated by colour. The derived h_{J_z} for the different [α/Fe] bins also show a smooth decrease with decreasing [α/Fe]. The oldest stars are kinematically hotter and higher [α/Fe]. This result is consistent with an upside-down formation of the Milky Way (Brook et al. 2012; Bird et al. 2013). Although it is subtle, our results also suggest that the decrease in h_{J_z} with decreasing [α/Fe] happens more rapidly for the high-[α/Fe], old population than for the young population as can be seen from the changing slope in the right panel of Fig. 8. The change of the slope happens roughly at [α/Fe] ~ 0.12, where the Bridge region starts. Overall, the high-[α/Fe] thick disc is smoothly connected to the low-[α/Fe] thin disc population. This result indicates that the chemo-dynamical evolution of the inner disc is smooth.

In Fig. 9, we show the relations between [α/Fe], [Fe/H] and age coloured by the mean radius of the stellar orbit, R_m. From the [α/Fe]-[Fe/H] relationship, shown in the left panel, we infer that the high-[Fe/H] ridge region selected in the left panel of Fig. 8 is mainly populated by small R_m stars, which is consistent with our view that this region is tracing the chemical evolution of the inner disc. On the other hand, it is known that the low [Fe/H] ([Fe/H] $\lesssim -0.1$), low-[α/Fe] population are not connected with the thick high-[α/Fe] population and show a distinct population (e.g., Hayden et al. 2015; Queiroz et al. 2019). However, as we discussed in Figs. 6 and 7, it is connected via the Bridge region. Interestingly, as seen in Fig. 7, the low-[Fe/H], low-[α/Fe] stars only appear at age < 11 Gyr. In addition, their R_m is predominantly larger ($R_\mathrm{m} > 9$ kpc). Hence, we consider that the low-[Fe/H], low-[α/Fe] stars formed at the outer disc and their star formation started when the disc grew large enough to develop a wide range of [Fe/H], i.e. the metallicity gradient, at the end of the transition period of the Bridge after the old thick disc formation. As a result, the star formation and chemical evolution path should be different from the inner disc, and the stars in the outer disc do not originate in the thick disc formation phase.

This different path of the disc formation in the inner disc and the outer disc is schematically described with the arrows in Fig. 9. The arrows highlighted with “inner", “local" and “outer" indicate the chemical evolution paths at the inner, local, i.e. solar radius, and outer discs, respectively, inferred from our data. The middle panel shows that low-[Fe/H] stars start forming later than the inner disc and are systematically younger than the thick disc, which is formed only in the inner disc. The right panel shows that the lower-[Fe/H] outer thin disc stars are higher [α/Fe]. This is seen as a positive [α/Fe] radial gradient in the thin disc (e.g., Hayden et al. 2015).

3.3 The young high-[α/Fe] stars

The lower panels of Fig. 7, consisting of stars younger than 10 Gyr, reveal the existence of a population of kinematically hot, high-[α/Fe] stars. To understand their origin, we look at the distribution in R_m and J_z between stars with [α/Fe] > 0.12 dex and old ($\log(\tau[\mathrm{Gyr}]) > 1.0$) and stars with [$\alpha$/Fe] > 0.12 dex and young ($\log(\tau[\mathrm{Gyr}]) < 0.8$). As shown in Fig. 10 the two groups of stars overlap significantly in both R_m and J_z distributions. Fig. 11, where we compared between [α/Fe] > 0.12 dex stars and young stars

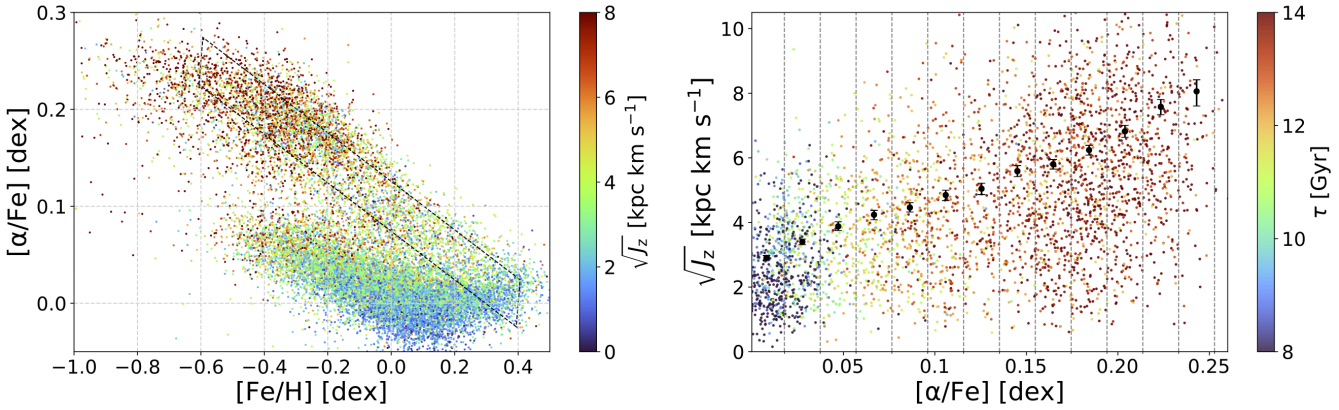

Figure 8. Left panel: the distribution of [α/Fe] and [Fe/H] coloured by the square root of the vertical action, $\sqrt{J_z}$. The ridge selection in the left panel represents the highest metallicity track in the [Fe/H]-[α/Fe] space. Right panel: the [α/Fe]-$\sqrt{J_z}$ relationship coloured by age in the ridge region highlighted in the left panel. We overlay the scale heights (black dots) and uncertainties (error bars) measured by fitting an isothermal profile to the distribution of $p(J_z)$ in 13 bins in [α/Fe]. Analysis of the high-[Fe/H] ridge shows that J_z smoothly decreases with [α/Fe] and age.

($0.2 < \log(\tau[\mathrm{Gyr}]) < 0.5$) having [$\alpha$/Fe] < 0.1 dex, shows that the two populations differ greatly in their kinematical properties.

These results indicate that the young high-[α/Fe] population originated from the old high-[α/Fe], i.e. old thick disc, population rather than the low-[α/Fe] thin disc population. Their hot kinematics implies that the high-[α/Fe] stars identified as young using the ages given by BINGO have originated from the star-forming gas disc which is as kinematically hot as the gas disc that created the old high-[α/Fe] stars. Such a highly turbulent gas disc only existed at the early epoch of the old high-[α/Fe] thick disc formation. Therefore, it is likely that these [α/Fe] stars identified as young formed at the early epoch, but they are identified as young stars most likely because they are the merger of old binary stars (Jofré et al. 2016), which has lowered their [C/N] abundance (Izzard & Halabi 2018) and biased the age estimator. Although we cannot establish when the merger took place, we can conclude from their hot kinematics that the high-[α/Fe] stars that appear young originated from the old high-[α/Fe] thick disc stars.

By combining the age information with the chemistry and kinematics, we can constrain the origin of the kinematically hot young high-[α/Fe] population. Our results are in agreement with Silva Aguirre et al. (2018), Miglio et al. (2020), who also found similar kinematical properties between young high-[α/Fe] stars and young low-[α/Fe]. We confirmed their results with a larger number of 69 young high-[α/Fe] stars. Miglio et al. (2020) also found a higher fraction of young (over-massive) high-[α/Fe] population in RC stars than RGB stars, and discussed that this is consistent with a scenario that these young [α/Fe] stars are the merged binaries, because more binaries are expected to have undergone an interaction around the tip of RGB than fainter RGB.

4 IMPLICATIONS FOR THE DISC FORMATION AND EVOLUTION

Our results suggest a formation scenario for the Galactic disc that involves distinct star formation and chemical evolution pathways of the inner and outer discs. In the inner disc, the thick disc forms early on from chemically well-mixed and turbulent gas, which can be, for example, associated with gas-rich mergers (e.g., Brook et al. 2004), cold gas flow accretion (e.g., Kereš et al. 2005; Dekel & Birnboim 2006; Brooks et al. 2009; Ceverino et al. 2010; Fernández et al. 2012) or, most likely, a complex interplay of both (e.g., Grand et al. 2018, 2020). Such a thick disc formation scenario can explain the clear and tight age-[α/Fe] (Fig. 4), age-[Fe/H] (Fig. 5) and [Fe/H]-[α/Fe] sequence (Fig. 6) for the old high-[α/Fe] thick disc stars.

After the formation of the old high-[α/Fe] thick disc, in the inner region there could be a smooth chemodynamical evolution from high-[α/Fe] to low-[α/Fe] and increasing metallicity as indicated by the “inner" pathway in Fig. 9. There is no distinct epoch of thick and thin disc formation, as seen in the ridge region of Fig. 8. Instead, the thicker to thinner disc transition happens in a smooth manner as stars continue to form with lower J_z from the dense cold gas continuously present at this radius (Brook et al. 2012; Grand et al. 2018). The smooth transition between the thick and thin discs in the inner region naturally arises in multi-zone semi-analytical chemo-dynamical evolution models (e.g., Schönrich & Binney 2009; Schönrich & McMillan 2017), where stars keep forming from the left-over gas of the high-[α/Fe] sequence.

A smooth transition between the formation of the thick and thin discs in the inner region is also suggested as the “centralised starburst pathway" in Grand et al. (2018). Using the high-resolution Auriga cosmological simulations of the Milky Way (Grand et al. 2017), Grand et al. (2018) propose the “centralised starburst pathway" model that can explain the single sequence of the [α/Fe]-[Fe/H] distribution in the inner disc seen in the APOGEE data of Hayden et al. (2015) (see also Palla et al. 2020). In their model, a major gas-rich merger and cold gas accretion at an early epoch initiates a short period of intense star formation in the inner region during which the thick disc forms with higher [α/Fe]. Once Type Ia SNe become significant in chemical enrichment after the peak of star formation in $\sim$ 1 Gyr timescale, more metal-rich low-[α/Fe] thin disc stars continuously form from the left-over less turbulent gas in the inner disc. As a result, there is no gap in the formation of the thick and thin disc and a single sequence of [α/Fe]-[Fe/H] is expected in the inner disc. Then, we can observe such inner disc stars in our data due to radial migration (e.g., Brook et al. 2012; Minchev et al. 2013; Kawata et al. 2018; Renaud et al. 2020), which brings them within $7 < R < 9$ kpc.

Our results also suggest that the star formation and chemical evolution in the outer disc starts after the thick disc phase. When

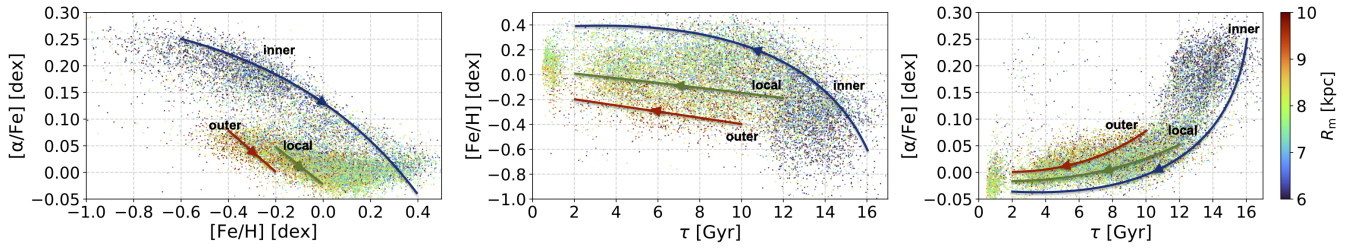

Figure 9. The distribution in [α/Fe] vs [Fe/H] (left panel), [Fe/H] vs age (middle panel), and [α/Fe] vs age (right panel) coloured by mean orbital radius, R_m. The "inner", "local" and "outer" arrows indicate the schematic chemical evolution paths at the inner ($R_m \sim 6$ kpc), local, i.e. solar radius ($R_m \sim 8$ kpc), and outer discs ($R_m \sim 10$ kpc), respectively. The metal-poor, outer disc stars follow a different chemical evolution pathway than the inner disc. These evolutionary paths are shown to describe qualitative trends of the chemical evolution at the different radii of the Galactic disc, and are not meant to indicate the chemical evolution paths quantitatively.

the thick-disc like, gas-rich merger- and/or cold accretion-dominant, turbulent star formation ends, the galactic halo may have grown enough for the hot gas accretion mode to become dominant (Brooks et al. 2009; Noguchi 2018). Then, the violent cold gas accretion stops, and the gas disc can grow in an inside-out fashion, as fresh low [Fe/H] gas is accreted smoothly from the hot halo gas. The disc rapidly grows large enough to develop a negative metallicity gradient as seen in the Bridge region of Fig. 6, unlike for the turbulent small thick disc phase, where the metals are well mixed, and no metallicity gradient can develop. Hence, the metal-poor outer disc developed after the thick disc formation, as indicated by the arrow of the "outer" disc chemical evolution pathway in Fig. 9.

The Bridge region in Fig. 6 shows that the range of [Fe/H] becomes wider for the younger stars, as indicated by the arrows in the left panel of Fig. 6. We consider that the Bridge region is where the thin disc formation begins, and the disc is developing a metallicity gradient with younger stars forming with a broader range of [Fe/H]. Radial migration brings stars formed in the inner disc and outer disc to the solar neighbourhood, which is where the stars in our samples lie (but see also Vincenzo & Kobayashi 2020; Khoperskov et al. 2021, who claim that radial migration is not very efficient). As a result, we can observe the mixed chemical distribution from pathways in the inner and outer discs (Schönrich & Binney 2009) . The high metallicity ridge highlighted in Fig. 8 represents the chemical evolution of the inner disc. The low-[Fe/H] and low-[α/Fe] stars came from the outer disc. As a result, we observe the two sequences of the high- and low-[α/Fe] stars in our sample (Brook et al. 2012; Grand et al. 2018). This is consistent with what is seen in the previous observational studies (e.g. Hayden et al. 2017; Feuillet et al. 2019). Our results provide more robust trends with more reliably measured relative ages.

Consequently, the stars currently around the solar radius follow the so-called two-infall model (Grisoni et al. 2017; Spitoni et al. 2019), although the high-[α/Fe] populations are formed in the inner disc and brought to the solar neighbourhood likely by radial migration. On the other hand, if we could select only the stars formed locally, the star formation starts at a later epoch from lower metallicity gas, when the star-forming disc becomes as large as the solar radius. In other words, the stars started forming locally around the solar radius from the second infall of gas, whose metallicity is lower, because of the dilution with the metal-poor gas accreted from the halo (Calura & Menci 2009; Spitoni et al. 2019; Buck 2020). The scenario is also consistent with what is suggested by Snaith et al. (2015) and Haywood et al. (2016, 2019). In fact, by private communication with Misha Haywood after the submission of the first version of this paper, we realised that our schematic view in Fig. 9 is consistent with Fig. 6 of Haywood et al. (2019), except that they considered that the outer low-[α/Fe] stars formed earlier than the inner low-[α/Fe] stars.

5 SUMMARY

In this paper, we determine precise relative stellar age estimates for 17,305 evolved stars in the APOGEE DR14 survey using a Bayesian Neural Network trained on the APOKASC-2 asteroseismic dataset. To minimize the bias in our age inference, we erase the distance dependence of metallicity and age in our training set by randomly displacing the distance of the stars. We also augment the dataset by over-sampling young and very old stars, to obtain a balanced training data and minimize the effect of regression dilution. Using the chemo-kinematical information, we separate the Galactic disc into three components, the thick and thin discs and the Bridge in the [Fe/H]-[α/Fe] distribution. The thick disc population is older and higher-[α/Fe] ([α/Fe] $\gtrsim 0.12$) than the thin disc. We argue that the Bridge population connects the thick disc and thin disc phases smoothly, rather than being part of the traditional thick disc. We also find an unusual population of young and high-[α/Fe] stars. However, we found that their kinematic properties are similar to the old high-[α/Fe] stars, which suggests that their origin must be the same as the old high-[α/Fe] stars. They are identified as young stars likely due to the merger of binary stars which decreased [C/N] and led to the predicted young ages.

To further investigate whether or not there is a smooth transition between the formation of the thick and thin disc in the inner region, we select a high-metallicity ridge region in the [Fe/H]-[α/Fe] plane that follows a continuously increasing [Fe/H] and decreasing [α/Fe] sequence. We examined the variation of J_z with [α/Fe] and age and concluded that, while there seems to be a hint of a sudden decrease in J_z around [α/Fe] ~ 0.12 dex, J_z smoothly decreases with [α/Fe] and also with age. We find that the oldest stars are kinematically hotter and enhanced in α-abundances than the younger stars. We found that the high-metallicity ridge is dominated by the stars from the inner disc and traces the continuous chemical evolution of the inner disc, $R < 6$ kpc. The formation of the thick disc is expected to happen in a compact disc, i.e. only in the inner disc, and a turbulent period of intense chemical mixing leads to the relatively tight sequence in the distribution of [α/Fe] and [Fe/H] for the old stars. From our results, we infer that the inner disc continuously forms stars from the

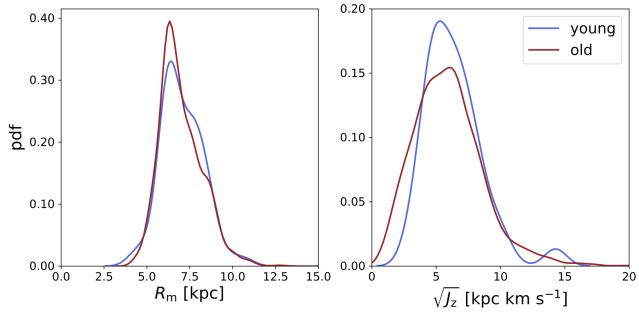

Figure 10. The distribution in R_m (left panel) and $\sqrt{J_z}$ (right panel) for old ($\log(\tau\,[\mathrm{Gyr}]) > 1.0$) stars with $[\alpha/\mathrm{Fe}] > 0.12$ dex, shown in red, and young ($\log(\tau\,[\mathrm{Gyr}]) < 0.8$) stars with $[\alpha/\mathrm{Fe}] > 0.12$ dex, shown in blue. The two populations have similar kinematics.

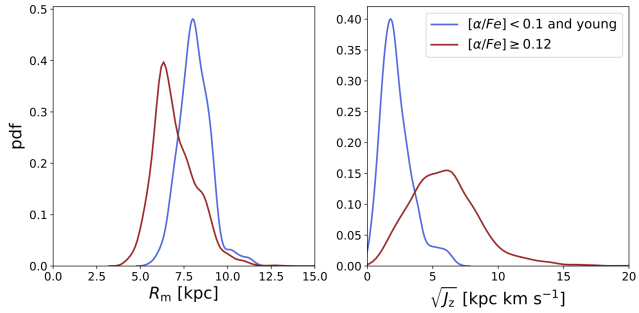

Figure 11. The distribution in R_m (left panel) and $\sqrt{J_z}$ (right panel) for stars with $[\alpha/\mathrm{Fe}] > 0.12$ dex, shown in red, and young ($0.2 < \log(\tau\,[\mathrm{Gyr}]) < 0.5$) stars with $[\alpha/\mathrm{Fe}] < 0.1$ dex, shown in blue. The kinematical properties differ between the two populations.

left-over gas after the thick disc formation phase and the subsequent accreting gas, and develop high-[Fe/H] and low-[α/Fe] stars.

We also found that the outer low-[Fe/H] and low-[α/Fe] stars are significantly younger than the inner high-[α/Fe] ([α/Fe] $\gtrsim$ 0.12 dex) stars. We argue that the outer metal-poor disc stars form after the end of cold-mode dominated violent thick disc formation phase. This likely corresponds to the transition from the cold to hot mode of the gas accretion due to the halo mass growth (Noguchi 2018; Grand et al. 2018).

In light of these results, we argue that the inner and outer discs of the Milky Way follow different chemical evolution pathways. After the violent thick disc formation phase ends, the thin disc formation starts with a smaller disc which is as small as the thick disc, and then the thin disc grows in an inside-out fashion. As the disc is growing with a supply of accreting low-[Fe/H] gas, metallicity gradients naturally arise, with the outer disc being more metal-poor than the inner disc. We found that the Bridge region shows a broader range of [Fe/H] with decreasing age, and suggest that the Bridge region is where the thin disc formation begins, and the disc is developing a metallicity gradient.

The recent work of Grand et al. (2020) suggested that the last significant merger of Gaia-Enceladus-Sausage (GES, Brook et al. 2003; Belokurov et al. 2018; Haywood et al. 2018; Helmi et al. 2018) was a gas-rich merger that was essential in forming the thick disc. This picture is also consistent with what we found in this paper because this gas-rich merger can induce a violent starburst in the inner disc due to the dissipation of the gas during the merger. If the GES merger was the last significant merger, then the thin disc phase could start after the GES merger settled. If this scenario is true, the end of the GES merger could correspond to the high-[α/Fe] tip ([α/Fe] $\sim$ 0.12 dex, [Fe/H] $\sim$ 0.0 dex) of the Bridge region of Fig. 6. After that, the thin disc grew inside-out from the smooth accretion of the low metallicity gas from the hot halo gas. Although this is admittedly pure speculation, we could test this hypothesis if we measured the relative difference in ages between the GES, the GES merger remnants (e.g., Chaplin et al. 2020; Montalbán et al. 2020), high-[α/Fe] thick disc and the Bridge. Measuring the age difference of stars precisely represents the holy grail of Galactic archaeology, as it allows us to improve our understanding of stellar evolution and probe deeper into the formation and evolution history of the Milky Way.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

ACKNOWLEDGEMENTS

We thank our anonymous referee for their helpful comments to improve the manuscript. IC and DK acknowledge the support of the UK's Science and Technology Facilities Council (STFC Grant ST/N000811/1 and ST/S000216/1). IC is also grateful for the STFC Doctoral Training Partnerships Grant (ST/N504488/1). IC thanks the LSSTC Data Science Fellowship Program, where their time as a Fellow has benefited this work. AM acknowledges support from the ERC Consolidator Grant funding scheme (project ASTEROCHRONOMETRY, grant agreement number 772293). GRD has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (CartographY GA. 804752). This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the *Gaia* Multilateral Agreement. This work was inspired from our numerical simulation studies used the UCL facility Grace and the Cambridge Service for Data Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The DiRAC component of CSD3 was funded by BEIS capital funding via STFC capital grants ST/P002307/1 and ST/R002452/1 and STFC operations grant ST/R00689X/1. DiRAC is part of the National e-Infrastructure.

REFERENCES

Abadi M. G., Navarro J. F., Steinmetz M., Eke V. R., 2003, ApJ, 597, 21
Abadi M., et al., 2016, arXiv e-prints, p. arXiv:1603.04467
Abolfathi B., et al., 2018, ApJS, 235, 42
Anders F., et al., 2014, A&A, 564, A115
Anders F., Chiappini C., Santiago B. X., Matijevič G., Queiroz A. B., Steinmetz M., Guiglion G., 2018, A&A, 619, A125
Bekki K., Tsujimoto T., 2011, ApJ, 738, 4

Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, MNRAS, 478, 611
Bensby T., Feltzing S., Lundström I., Ilyin I., 2005, A&A, 433, 185
Bensby T., Alves-Brito A., Oey M. S., Yong D., Meléndez J., 2011, ApJ, 735, L46
Beraldo e Silva L., Debattista V. P., Khachaturyants T., Nidever D., 2020, MNRAS, 492, 4716
Binney J., 2010, MNRAS, 401, 2318
Binney J., McMillan P., 2011, MNRAS, 413, 1889
Bird J. C., Kazantzidis S., Weinberg D. H., Guedes J., Callegari S., Mayer L., Madau P., 2013, ApJ, 773, 43
Borucki W. J., et al., 2010, Science, 327, 977
Bournaud F., Elmegreen B. G., Martig M., 2009, ApJ, 707, L1
Bovy J., 2015, ApJS, 216, 29
Bovy J., Rix H.-W., Liu C., Hogg D. W., Beers T. C., Lee Y. S., 2012, ApJ, 753, 148
Brook C. B., Kawata D., Gibson B. K., Flynn C., 2003, ApJ, 585, L125
Brook C. B., Kawata D., Gibson B. K., Freeman K. C., 2004, ApJ, 612, 894
Brook C. B., Kawata D., Martel H., Gibson B. K., Bailin J., 2006, ApJ, 639, 126
Brook C. B., et al., 2012, MNRAS, 426, 690
Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, ApJ, 694, 396
Buck T., 2020, MNRAS, 491, 5435
Calura F., Menci N., 2009, MNRAS, 400, 1347
Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, A&A, 530, A138
Ceverino D., Dekel A., Bournaud F., 2010, MNRAS, 404, 2151
Chaplin W. J., et al., 2020, Nature Astronomy, p. 7
Cheng J. Y., et al., 2012, ApJ, 752, 51
Chiappini C., Matteucci F., Gratton R., 1997, ApJ, 477, 765
Chiappini C., Matteucci F., Romano D., 2001, ApJ, 554, 1044
Chiappini C., et al., 2015, A&A, 576, L12
Cropper M., et al., 2018, Gaia Radial Velocity Spectrometer (arXiv:1804.09369)
Das P., Sanders J. L., 2018, Monthly Notices of the Royal Astronomical Society, 484, 294–304
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Delgado Mena E., et al., 2019, A&A, 624, A78
Evans D. W., et al., 2018, A&A, 616, A4
Fernández X., Joung M. R., Putman M. E., 2012, ApJ, 749, 181
Feuillet D. K., Frankel N., Lind K., Frinchaboy P. M., García-Hernández D. A., Lane R. R., Nitschelm C., Roman-Lopes A., 2019, MNRAS, 489, 1742
Fuhrmann K., 1998, A&A, 338, 161
Gaia Collaboration et al., 2018, A&A, 616, A11
García Pérez A. E., et al., 2016, AJ, 151, 144
Gilmore G., Reid N., 1983, MNRAS, 202, 1025
Grand R. J. J., Springel V., Gómez F. A., Marinacci F., Pakmor R., Campbell D. J. R., Jenkins A., 2016, MNRAS, 459, 199
Grand R. J. J., et al., 2017, MNRAS, 467, 179
Grand R. J. J., et al., 2018, MNRAS, 474, 3629
Grand R. J. J., et al., 2020, MNRAS, 497, 1603
Grisoni V., Spitoni E., Matteucci F., Recio-Blanco A., de Laverny P., Hayden M., Mikolaitis Ŝ., Worley C. C., 2017, MNRAS, 472, 3637
Hawkins K., Ting Y.-S., Walter-Rix H., 2018, ApJ, 853, 20
Hayden M. R., et al., 2015, The Astrophysical Journal, 808, 132
Hayden M. R., Recio-Blanco A., de Laverny P., Mikolaitis S., Worley C. C., 2017, A&A, 608, L1
Haywood M., Di Matteo P., Lehnert M. D., Katz D., Gómez A., 2013, A&A, 560, A109
Haywood M., Lehnert M. D., Di Matteo P., Snaith O., Schultheis M., Katz D., Gómez A., 2016, A&A, 589, A66
Haywood M., Di Matteo P., Lehnert M. D., Snaith O., Khoperskov S., Gómez A., 2018, ApJ, 863, 113
Haywood M., Snaith O., Lehnert M. D., Di Matteo P., Khoperskov S., 2019, A&A, 625, A105
Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, Nature, 563, 85–88
Izzard R. G., Halabi G. M., 2018, arXiv e-prints, p. arXiv:1808.06883
Jofré P., et al., 2016, A&A, 595, A60
Katz D., et al., 2019, Astronomy & Astrophysics, 622, A205
Kawata D., Grand R. J. J., Gibson B. K., Casagrande L., Hunt J. A. S., Brook C. B., 2017, MNRAS, 464, 702
Kawata D., et al., 2018, MNRAS, 473, 867
Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2
Khoperskov S., Haywood M., Snaith O., Di Matteo P., Lehnert M., Vasiliev E., Naroenkov S., Berczik P., 2021, MNRAS, 501, 5176
Kobayashi C., Nakasato N., 2011, ApJ, 729, 16
Lindegren L., et al., 2018, Astronomy & Astrophysics, 616, A2
Loebman S. R., Roškar R., Debattista V. P., Ivezić Ž., Quinn T. R., Wadsley J., 2011, ApJ, 737, 8
Mackereth J. T., et al., 2019, Monthly Notices of the Royal Astronomical Society, 489, 176–195
Majewski S. R., et al., 2017, AJ, 154, 94
Martig M., et al., 2015, MNRAS, 451, 2230
Matteucci F., Francois P., 1989, MNRAS, 239, 885
Miglio A., et al., 2020, arXiv e-prints, p. arXiv:2004.14806
Minchev I., Famaey B., Quillen A. C., Dehnen W., Martig M., Siebert A., 2012, A&A, 548, A127
Minchev I., Chiappini C., Martig M., 2013, A&A, 558, A9
Montalbán J., et al., 2020, arXiv e-prints, p. arXiv:2006.01783
Nidever D. L., et al., 2014, ApJ, 796, 38
Noguchi M., 1998, Nature, 392, 253
Noguchi M., 2018, Nature, 559, 585
Palla M., Matteucci F., Spitoni E., Vincenzo F., Grisoni V., 2020, MNRAS, 498, 1710
Pinsonneault M. H., et al., 2018, ApJS, 239, 32
Pourzanjani A., Jiang R., Petzold L., 2017, in NIPS Workshop on Bayesian Deep Learning.
Prochaska J. X., Naumov S. O., Carney B. W., McWilliam A., Wolfe A. M., 2000, AJ, 120, 2513
Queiroz A. B. A., et al., 2019, arXiv e-prints, p. arXiv:1912.09778
Quinn P. J., Hernquist L., Fullagar D. P., 1993, ApJ, 403, 74
Renaud F., Agertz O., Read J. I., Ryde N., Andersson E. P., Bensby T., Rey M. P., Feuillet D. K., 2020, arXiv e-prints, p. arXiv:2006.06011
Riello M., et al., 2018, A&A, 616, A3
Roškar R., Debattista V. P., Loebman S. R., 2013, MNRAS, 433, 976
Salvatier J., Wiecki T., Fonnesbeck C., 2015, arXiv e-prints, p. arXiv:1507.08050
Sanders J. L., Das P., 2018, Monthly Notices of the Royal Astronomical Society, 481, 4093–4110
Sartoretti P., et al., 2018, Astronomy & Astrophysics, 616, A6
Schönrich R., Binney J., 2009, MNRAS, 399, 1145
Schönrich R., McMillan P. J., 2017, MNRAS, 467, 1154
Schönrich R., McMillan P., Eyer L., 2019, MNRAS, 487, 3568
Silva Aguirre V., et al., 2018, MNRAS, 475, 5487
Skrutskie M. F., et al., 2006, AJ, 131, 1163
Snaith O., Haywood M., Di Matteo P., Lehnert M. D., Combes F., Katz D., Gómez A., 2015, A&A, 578, A87
Spitoni E., Silva Aguirre V., Matteucci F., Calura F., Grisoni V., 2019, A&A, 623, A60
Ting Y.-S., Rix H.-W., 2019, ApJ, 878, 21
Ting Y.-S., Hawkins K., Rix H.-W., 2018, ApJ, 858, L7
Tissera P. B., White S. D. M., Scannapieco C., 2012, MNRAS, 420, 255
Villalobos Á., Helmi A., 2008, MNRAS, 391, 1806
Vincenzo F., Kobayashi C., 2020, MNRAS, 496, 80

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