Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **517,** 5325–5339 (2022) https://doi.org/10.1093/mnras/stac2787
Advance Access publication 2022 October 12

The GALAH survey: chemical clocks

Michael R. Hayden,[1,2]★ Sanjib Sharma,[1,2] Joss Bland-Hawthorn,[1,2] Lorenzo Spina,[2,3,4] Sven Buder,[2,5] Ioana Ciucă,[2,5] Martin Asplund,[6] Andrew R. Casey,[2,3] Gayandhi M. De Silva,[7,8] Valentina D'Orazi,[3,4] Ken C. Freeman,[2,5] Janez Kos,[9] Geraint F. Lewis,[1] Jane Lin,[2,5] Karin Lind,[10] Sarah L. Martell,[2,11] Katharine J. Schlesinger,[5] Jeffrey D. Simpson,[2,11] Daniel B. Zucker,[2,8,12] Tomaž Zwitter,[9] Boquan Chen,[1,2] Klemen Čotar,[9] Diane Feuillet,[13] Jonti Horner,[14] Meridith Joyce,[2,5] Thomas Nordlander,[2,5] Dennis Stello,[11,2] Thor Tepper-Garcia,[1,2,15] Yuan-sen Ting (丁源森),[5,16,17,18] Purmortal Wang (王梓先),[1,2] Rob Wittenmyer[14] and Rosemary Wyse[19]

Affiliations are listed at the end of the paper

Accepted 2022 August 24. Received 2022 August 23; in original form 2021 May 18

ABSTRACT

We present the first large-scale study that demonstrates how ages can be determined for large samples of stars through Galactic chemical evolution. Previous studies found that the elemental abundances of a star correlate directly with its age and metallicity. Using this knowledge, we derive ages for 214 577 stars in GALAH DR3 using only overall metallicities and chemical abundances. Stellar ages are estimated via the machine learning algorithm *XGBoost* for stars belonging to the Milky Way disc with metallicities in the range $-1 < [\mathrm{Fe/H}] < 0.5$, using main-sequence turn-off stars as our training set. We find that stellar ages for the bulk of GALAH DR3 are precise to 1–2 Gyr using this method. With these ages, we replicate many recent results on the age-kinematic trends of the nearby disc, including the solar neighbourhood's age–velocity dispersion relationship and the larger global velocity dispersion relations of the disc found using *Gaia* and GALAH. These results show that chemical abundance variations at a given birth radius are small, and that strong chemical tagging of stars directly to birth clusters may prove difficult with our current elemental abundance precision. Our results highlight the need to measure abundances for as many nucleosynthetic production sites as possible in order to estimate reliable ages from chemistry. Our methods open a new door into studies of the kinematic structure and evolution of the disc, as ages may potentially be estimated to a precision of 1–2 Gyr for a large fraction of stars in existing spectroscopic surveys.

Key words: Galaxy: abundances – Galaxy: kinematics and dynamics – Galaxy: stellar content – Galaxy: structure.

1 INTRODUCTION

How the Milky Way formed and its evolution through time is one of the critical questions facing astrophysics today. Galactic Archaeology (Freeman & Bland-Hawthorn 2002) is the midst of a revolution in its attempt to answer these questions with the advent of *Gaia* (Gaia Collaboration 2016), enabling a plethora of new discoveries about the structure, formation, and evolution of the Galaxy. Even now, however, the origin of various Galactic substructures and the relative importance of secular processes such as blurring and migration are still a matter of great debate (e.g. Rix & Bovy 2013; Bland-Hawthorn & Gerhard 2016). Stars and stellar populations are one of the primary tracers by which the Galaxy can be studied, as stars contain the chemical imprint of the gas from which they formed, allowing the evolutionary history of the Galaxy to be traced through time (Freeman & Bland-Hawthorn 2002). However, our knowledge and understanding of the Milky Way has been hampered by the lack of large samples of stars for which reliable age estimates are available.

Traditionally, stellar ages can be determined only for small subsets of the Hertzspring–Russell (H-R) diagram. Isochrone matching can be used to determine ages for stars along the main-sequence turn-off (MSTO) and subgiant branch, where there is a large separation between stars of different ages in the T_{eff} versus luminosity plane. Ages can also be determined from high-quality studies of solar twins, where stellar parameters can be determined with much higher precision than other stellar types. Asteroseismology provides an additional avenue for age determination, particularly for giant stars, by providing accurate mass and radius determinations. However, this requires extremely accurate photometry and long baseline observations, and is generally restricted to specific areas of the sky like the *Kepler* (Borucki et al. 2010) fields. In giants, the C/N ratio can also be used as a proxy for age, as the first dredge up along the giant branch is mass dependant (e.g. Masseron & Gilmore 2015; Martig et al. 2016; Ness et al. 2016; Casali et al. 2019). This relation can be calibrated using asteroseismic observations. However, each of these methods is only able to provide ages for a relatively small number of stars, and require significant effort to provide those ages.

★ E-mail: michael.hayden@sydney.edu.au

None of these methods are therefore able to estimate ages for the full sample of stars that would be observed by a large-scale spectroscopic survey such as GALAH.

Studies of the chemical abundances of nearby stars have found that there are clear age–[α/Fe] relations (e.g. Haywood et al. 2013; Bensby, Feltzing & Oey 2014; Hayden et al. 2017), and that the [α/Fe] ratio is a good proxy for the age of a star in [α/Fe]-enhanced populations. For younger stellar populations belonging to the thin disc, this is no longer the case, as most stars have similar [α/Fe] abundances; therefore [α/Fe] ceases to be a direct measurement of stellar age for thin disc stellar populations. However, for the thin disc slow neutron capture process (s-process) abundances have been found to have a strong correlation with a stars age. In particular, the [Y/Mg] or [Y/Al] ratio has been observed to have a tight correlation with stellar age in studies of local MSTO stars (Nissen 2015, 2016; Spina et al. 2016, 2018; Feltzing et al. 2017; Bedell et al. 2018; Lin et al. 2020). Additionally, there have been differences observed between the age-abundance trends of the thin and thick discs, or as a function of metallicity (e.g. Feltzing et al. 2017; Titarenko et al. 2019; Casali et al. 2020; Lin et al. 2020; Nissen et al. 2020). Still, these studies highlight the potential of using chemical abundances such as yttrium to directly determine the age of a star, and potentially estimate the birth radius. Several studies have attempted to estimate the birth radii of stars by comparing their age and metallicities with models of Galaxy evolution (e.g. Minchev et al. 2018; Feltzing, Bowers & Agertz 2019), placing a star at a specific place and time in the Galaxy when it was born.

Tagging a star to a particular birth location is known as chemical tagging (Freeman & Bland-Hawthorn 2002). In strong chemical tagging, as outlined by Bland-Hawthorn, Krumholz & Freeman (2010), a star can be assigned to an individual birth cluster by measuring many abundances to high precision. Open clusters have been found to have very uniform chemical abundances, with scatter lower than 0.03 dex, which is the typical measurement uncertainty in such studies (De Silva et al. 2006; Bovy 2016). However, it is unclear to what degree each cluster has a unique chemical signature, or how this changes with time. For example, if the thick disc formed from many massive clumps (Clarke et al. 2019, but see Ting, Conroy & Rix 2016 also), these clumps may have unique chemical signatures enabling strong chemical tagging, while the smaller clusters forming in the thin disc today may not be very chemically distinct from other nearby star-forming regions. If the interstellar medium (ISM) is in general well-mixed and the mixing time-scale is short, it is possible that tagging to a specific cluster may be very difficult with the abundance precision in current spectroscopic surveys. Even in the case where strong chemical tagging is not possible, however, weak chemical tagging can still be extremely useful for Galactic Archaeology. Tagging a star to a general location in a Galaxy (i.e. a birth radius), rather than a specific star cluster, is known as weak chemical tagging. Weak chemical tagging allows the study of the spatial and kinematic structure of the Galaxy through time, and can aid in determining the effectiveness of secular processes such as blurring or migration (Sellwood & Binney 2002; Schönrich & Binney 2009).

Ness et al. (2019) used APOGEE (Majewski et al. 2017) data and found that the age and metallicity of a star alone could be used to estimate the initial orbits and abundances for stars, and that deviations from calculated age–abundance relations were quite small, within the measurement errors. These authors argue that this makes strong chemical tagging unrealistic for much of the disc, at least with current measurement precision. Casali et al. (2020) use high-resolution HARPS observations, and argue that the age–abundance relations, particularly those involving yttrium, are not universal, as they find that there are variations between the [Y/Mg]–age relation as a function of metallicity. These s-process elements were not observed by APOGEE. Casey et al. (2019) used GALAH (De Silva et al. 2015) to group stars by their chemical abundances, in particular by latent factors linked to the relative contribution of different nucleosynthetic production sites (e.g. Kobayashi, Karakas & Lugaro 2020); these authors identified six different latent factors linked to elemental production sites required to explain the abundance trends observed in GALAH. Sharma et al. (2022) also use GALAH data to measure age–abundance relations for nearly 30 elements and find that almost all of the elements studied show significant trends with age. They find, similar to Ness et al. (2019), that to first order age and metallicity alone can be used to estimate abundances across the disc, but do find minor deviations from these relations for some elements often related to the SNR or $T_{\rm eff}$.

The arguments laid out in Ness et al. (2019) and Sharma et al. (2022) have profound implications for the study of the kinematic structure and evolution of the disc. These authors argue that the age and metallicity alone can be used to predict chemical abundances. If correct, these relations can also be inverted: the age of a star can be inferred based on its overall metallicity and chemical abundances, if the abundance plane is unique for a given age. In essence, the chemical abundances of a star act as a clock for the chemical evolution of the Galaxy, from which we can date the age of a star based on its abundances. This means that if abundances can be measured to high precision, ages can potentially be estimated for every star in a stellar survey, expanding sample sizes with reliable ages by orders of magnitude compared to existing studies. In the ideal case, a survey would measure abundances for as many different metal production sites as possible (Type II Supernovae, Type Ia Supernovae, s-process, and rapid neutron capture process). In practice, not all surveys are able to observe elements from all production sites given their resolution or wavelength coverage. Additionally, some elements may be difficult to measure for different stellar types or due to non-LTE effects. Compromises must be made when selecting which chemical abundances to use between having more abundances that make an age estimate more precise, or using fewer abundances that are well measured for a larger fraction of the stars in a stellar survey.

In this paper, we estimate the ages of several hundred thousand stars directly from their chemical abundances as measured in GALAH DR33 (Buder et al. 2021). We use MSTO stars as a training set for Bayesian and machine learning models for age estimation, and attempt to estimate ages for stars across the H-R diagram based on chemical abundances alone. There are several potential obstacles with this approach, as the surface abundances of stars may vary due to mixing processes associated with stellar evolution, which may cause large systematic errors for some types of stars. This is further compounded by potential difficulties in measuring a uniform set of abundances to high fidelity across a large span of $T_{\rm eff}$, $\log g$, and [Fe/H]. With these ages, we reproduce recent age-kinematic observations of the disc with a sample size an order-of-magnitude larger than the previously available one. GALAH DR3 provides abundance determinations for nearly 30 elements, allowing us to be selective and choose elements that are well estimated for a large fraction of the sample while also covering the different nucleosynthetic production sites in the Milky Way.

This paper is organized as follows: in Section 2, we describe the sample selection, data cuts, and the MSTO training selection used in our analysis. In Section 3, we outline the methods we use to determine stellar ages for a large fraction of the GALAH DR3 sample,

using Bayesian analysis as well as *XGBoost* (Chen & Guestrin 2016) to generate a model for the age of a star based on its chemical abundances. In Section 4, we demonstrate the reliability of our age determinations and reproduce many of the recent kinematic studies of the Galaxy. In Section 5, we discuss our findings in the context of strong and weak chemical tagging, implications for future kinematic studies of the disc, the chemical evolution of the Galaxy, and future survey design.

2 DATA

Spectroscopic data are taken from GALAH DR3 (Buder et al. 2021), along with additional fields from the K2-HERMES (Wittenmyer et al. 2018) and *TESS*-HERMES surveys (Sharma et al. 2018). GALAH uses the High Efficiency and Resolution Multi-Element Spectrograph (HERMES, Sheinis et al. 2015) instrument, which is a high-resolution ($R \sim 28\,000$) multifibre spectrograph mounted on the 3.9 m Anglo Australian Telescope (AAT). HERMES covers four wavelength ranges (4713–4903 Å, 5648–5873 Å, 6478–6737 Å, and 7585–7887 Å), carefully selected to maximize the number of elemental abundances that are able to be measured. Observations are reduced through a standardized pipeline developed for the GALAH survey as described in Kos et al. (2017). Stellar atmospheric parameters and individual abundances are derived using Spectroscopy Made Easy (SME, Valenti & Piskunov 1996; Piskunov & Valenti 2017). The precision of individual abundances [X/Fe] is typically ∼0.05 dex, while the random errors in radial velocities are $\sim$100 m s^{-1} (Zwitter et al. 2018). In this analysis, we use a selection of 13 well-measured elemental abundances from GALAH DR3: [Fe/H], [Mg/Fe], [Ca/Fe], [Ti I/Fe], [Si/Fe], [O/Fe], [Mn/Fe], [Cr/Fe], [Na/Fe], [K/Fe], [Y/Fe], [Ba/Fe], and [Sc/Fe] which span a range of nucleosynthetic production sites.

Because of slight differences in both abundance precision and zero-point shifts between giants ($\log g < 3.5$) and dwarfs ($\log g > 3.5$), we report results for these subsamples as well as all of GALAH DR3 separately in this paper; however the same model trained on MSTO stars is applied to both subsamples. The training set is made up of stars with current radii in the range $6 < R < 10$ kpc, which is set by the absolute magnitude of stars belonging to the turn-off region and the magnitude limits of the GALAH survey. However, these stars span a much larger range in birth radius due to radial mixing processes (blurring, migration) than their current Galactocentric radial distributions, giving us much greater coverage of the Galaxy. Current measurements of the efficiency of migration (Sanders & Binney 2015; Frankel et al. 2020; Sharma, Hayden & Bland-Hawthorn 2021b) mean that nearly 50 per cent of stars observed within the solar neighbourhood have migrated more than 2 kpc from their birth position to their current location (see e.g. Sharma et al. 2021b, their Fig. 12). This is also reflected in the large spread of metallicities in the solar neighbourhood, with 25 per cent of stars being significantly more metal-rich than the local ISM and coming from the inner disc (e.g. Hayden 2020). We note that this does mean that some areas of the disc may not be present in our training set, particularly if the time-scale for radial mixing of stars from those radii is long (i.e. several Gyr), and the reliability of the chemical clock age determinations for these stellar populations would be poor.

We implement several quality cuts on the signal-to-noise ratio of the spectra, stellar parameter, and abundance flags, as well as a temperature cut-off of $T_{\rm eff} < 6200$ K, as outlined in Table 1. We require an SNR > 20, a quality flag = 0 for both the stellar parameters and the individual abundances (see the GALAH DR3 release paper, Buder et al. 2021). The reason for the $T_{\rm eff}$ cut at 6200 K is that we have found that there are significant issues for abundances measured for the hottest stars. We restrict our parameter space to that covered by the disc, $-1 < [{\rm Fe/H}] < 0.5$, as the age–abundance relations of any accreted objects, e.g. the halo, follow a different enrichment history. These cuts give us a sample of 155 519 dwarf stars and 92 645 giant stars for which we determine ages in this paper.

Table 1. Data selection criteria used to select stars in this paper.

Parameter	Value
sp_flag	0
abundance_flag [X/Fe]	0
SNR	>20
$T_{\rm eff}$	<6200 K
$\sigma_{T_{\rm eff}}$	<150 K
Spectral Fit χ^2	<4
[Fe/H]	$-1.0 <$ [Fe/H] < 0.5 dex

Table 2. Data cuts used on the GALAH DR3 sample to select training set of MSTO stars.

Parameter	Value
$\log g$	$3.5 < \log g < 4.1$
SNR	>45
τ (Gyr)	>1.75 Gyr
$\frac{\sigma_\tau}{\tau}$	<0.2

For the training set we use MSTO and subgiant stars, with ages computed with the code BSTEP (Sharma et al. 2018) with isochrone matching using PARSEC-COLIBRI isochrones (Marigo et al. 2017). BSTEP provides a Bayesian estimate of intrinsic stellar parameters by comparing observed parameters to those of the PARSEC-COLIBRI isochrones. For determining the age of an MSTO star through BSTEP, the relevant stellar and astrometric parameters are $T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], J, Ks, and parallax. Astrometric parameters are taken from *Gaia* DR2 (Gaia Collaboration 2018; Lindegren et al. 2018). The stellar distances and velocities used in this analysis are taken from the GALAH DR3 dynamics value added catalogue, see Buder et al. (2021) for details.

The accuracy for ages derived for MSTO stars in BSTEP is of the order of 10–15 per cent based on uncertainties in the spectroscopic parameters for MSTO stars in GALAH. With these low fractional uncertainties, MSTO stars make an ideal training set for age–abundance relations. In addition to the abundance and stellar parameter quality conditions applied to the rest of the GALAH sample, we apply additional criteria for the MSTO training set as outlined in Table 2. To isolate the MSTO stars, we use a cut in $\log g$ space with $3.5 < \log g < 4.1$, as well as stricter criteria on the SNR, with SNR > 45 for the training set. See Fig. 1 for the H-R diagram of the sample, which meets our selection criteria. The black box denotes the MSTO sample for which accurate ages are able to be estimated from isochrones. There are issues in determining the abundances for the youngest stars, $\tau < 1.75$ Gyr, and we elect to remove these stars from the training set. For stars belonging to the thick disc ([α/Fe]>0.15), we require these stars to have an age $\tau > 8$ Gyr (i.e. we remove the young α-rich stars from the training set, see Chiappini et al. 2015), see Fig. 2 for the age–α relation (top panel) and age distribution (bottom panel) found in our training set. Note that MSTO training set shown in the age–α relationship of Fig. 2 spans the metallicity range $-1 <$ [Fe/H] < 0.5, while the colourbar is set to maximize contrast for the bulk of the stellar sample. These young high-[α/Fe] populations make up less than 10 per cent of high-[α/Fe] sample, but

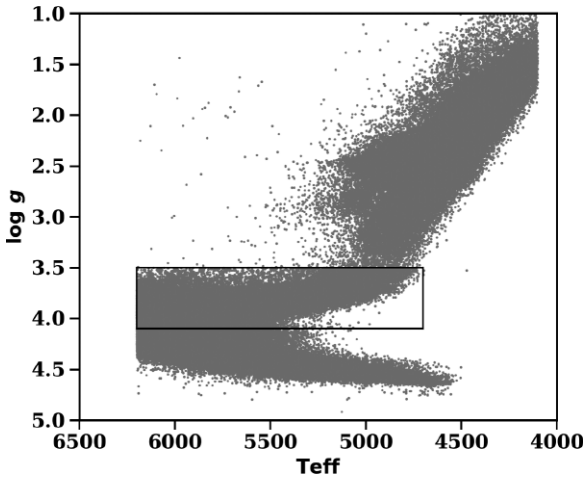

Figure 1. The H-R diagram for the sample presented in this paper. The MSTO selection criteria is shown by the black box.

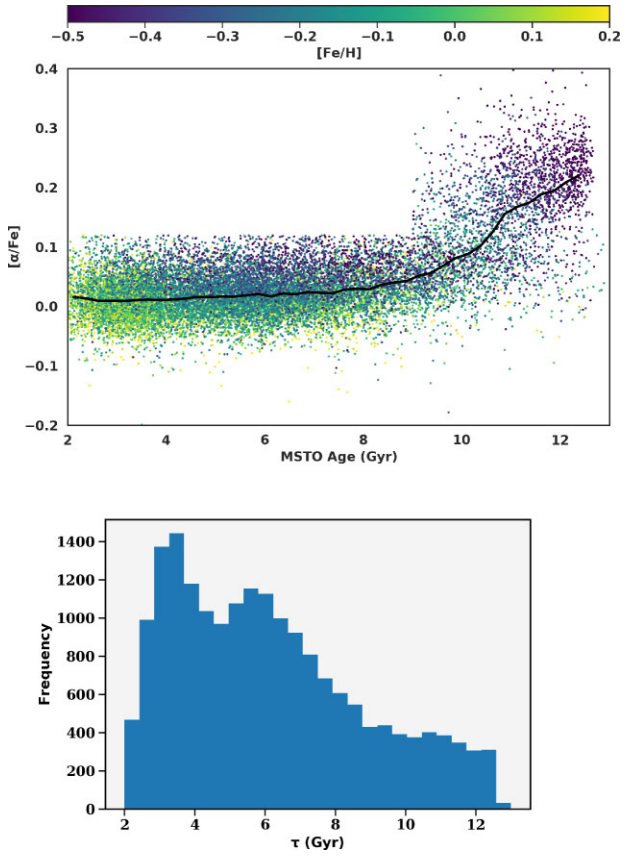

Figure 2. Top: The age–[α/Fe] relation for the MSTO training set. Note that the colourbar is set to maximize contrast; the MSTO sample spans the range of metallicity from $-1 <$ [Fe/H] < 0.5. Bottom: The age distribution derived from isochrone matching using BSTEP for the MSTO training set.

the origin of these stars is unclear in terms of Galactic evolution; they are likely blue stragglers (Jofré et al. 2016). Therefore, we remove these stars to avoid learning potentially spurious chemical trends. We also require the fractional uncertainty in the age determination from the isochrone matching to be better than 20 per cent, i.e. $\frac{\sigma_\tau}{\tau} < 0.2$. These restrictions, along with the quality cuts from Table 1, leave a sample of 15 424 MSTO stars belonging to the training and test sets. The age abundance trends for the input training set are shown in Fig. 3. These are the same trends found in Sharma et al. (2022).

3 METHODS

Ages can be estimated from chemical abundances in several ways. Initial studies of the [Y/Mg] ratio used linear fits (e.g. Nissen 2016; Bedell et al. 2018; Spina et al. 2018). However, variations have been found with metallcity in the age-abundance trends (e.g. Feltzing et al. 2017; Titarenko et al. 2019; Casali et al. 2020; Lin et al. 2020), and as our sample covers a large range in Galactocentric radii, more sophisticated methods are required.

3.1 Bayesian analysis

Sharma et al. (2022) determine age–abundance relations for various elements measured in GALAH DR3, under the assumption that an abundance can be determined given its age and metallicity. For elements which follow this assumption, this formalism can be inverted, and an age determined given a metallicity and a set of abundances, using Bayes Theorem. Let X_i denote the observed abundance [X/Fe] of the i-th element with measurement uncertainty σ_{Xi}. Let F be the observed metallicity [Fe/H] and σ_F its uncertainty. Given age τ and metallicity $F^{'}$ we can predict the abundance $X_{i,\,\mathrm{Model}}(\tau,\,F^{'})$ using a model with some intrinsic dispersion ϵ_{Xi}. Using Bayes theorem the probability distribution of age of a star given its metallicity and elemental abundances can then be written as

$$p(\tau|X_i, \sigma_{Xi}, F, \sigma_F) = \int \mathrm{d}F' p(\tau|F') p(F'|F, \sigma_F) \times \tag{1}$$

$$\prod_i p(X_i|\tau, F', \sigma_{Xi}), \tag{2}$$

where

$$p(X_i|\tau, F', \sigma_{Xi}) = \mathcal{N}\left(X_i|X_{i,\mathrm{Model}(\tau,F')}, \sigma_{Xi}^2 + \epsilon_{Xi}^2\right) \tag{3}$$

and

$$p(F'|F, \sigma_F) = \mathcal{N}\left(F|F', \sigma_F^2\right). \tag{4}$$

We apply this methodology with the age-abundance trends measured in Sharma et al. (2022), and results for the age determination using Bayesian analysis are shown in Fig. 4. There is in general a good agreement between the age determined from isochrone matching and the age determined using Bayes theorem, although there are significant number of outliers. The primary benefit of the Bayesian approach is that the errors in abundances are directly incorporated into the age estimation, along with the fact that missing or flagged abundances still allow an age to easily to be determined. Note that the chemical ages of Fig. 4 are derived solely based on Equations 1–4 using chemical abundances alone, and are derived separately compared to the isochrone matching ages determined with BSTEP (which use T_{eff}, log g, metallicity, [α/Fe], and luminosity) and were not used in the analysis of Sharma et al. (2022).

3.2 Machine learning: *XGBoost*

In addition to Bayesian analysis, ages can also be estimated using machine learning tools such as Neural Networks or gradient boosting algorithms. For this paper, we use the gradient boosting algorithm *XGBoost* (Chen & Guestrin 2016). *XGBoost* has a good mix of performance and speed, and is ideal for predictive modelling problems with structured data sets, such as how stellar age might

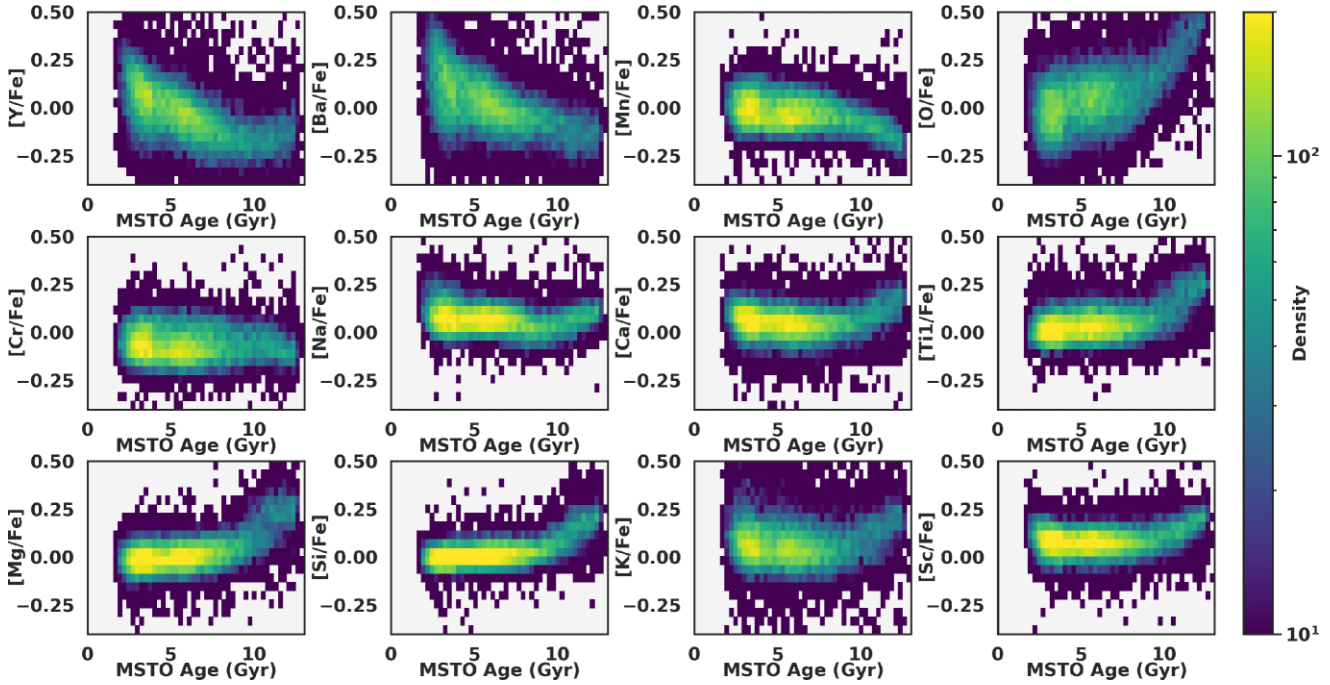

Figure 3. The age-abundance trends for MSTO using isochrone ages.

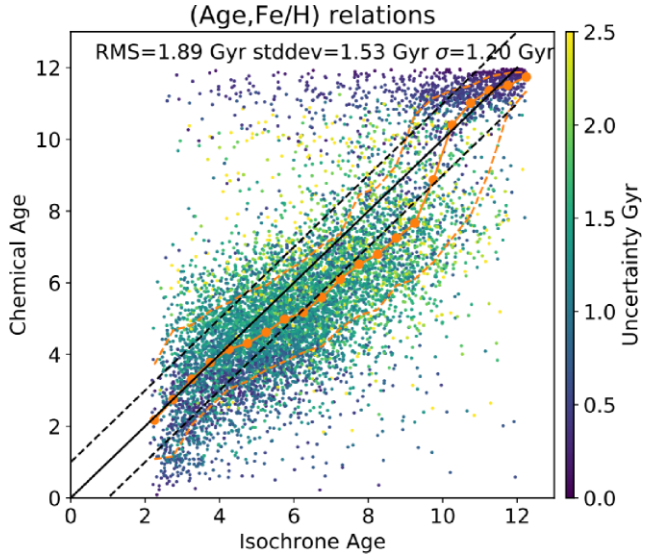

Figure 4. Comparison of chemical ages with those based on isochrones and stellar parameters.

vary with chemical abundance. Gradient boosting algorithms are a series of decision trees, where boosting is the process of adding additional models/trees to correct errors from the previous models. The gradient portion of the name comes in because it uses a gradient descent algorithm to minimize the loss when adding additional models. An example of a decision tree that a model might likely make for determining ages from abundances is as follows: is the star enhanced in [α/Fe] or is it solar-[α/Fe]? We know for instance that the high-[α/Fe] thick disc populations are almost uniformly old (ages >8 Gyr), so this tree could quickly split samples into old and young stars based on the [α/Fe] abundance. Further trees are added for additional abundance information, such a tree might be for example

Table 3. *XGBoost* parameters used in model generation.

Parameter	Value
Γ	5
Max depth of tree	7
Subsample ratio	0.7
Learning rate η	0.05
Minimum child weight	5
Column sampled by tree	0.8

the s-process abundance for the solar-[α/Fe] populations, which are then given weights in order to minimize the loss/errors (comparing the model predictions to the input data sets).

We split our initial sample of 15 424 MSTO stars into a training set (70 per cent of the original MSTO sample) and test set (30 per cent of the MSTO sample). The training set uses 13 chemical abundances described in the data section as input, along with the desired output parameter of age τ determined from MSTO isochrone fitting. We ran a fivefold cross validation grid of several hundred thousand *XGBoost* hyper parameters to obtain the model that best reproduced the trends of the test set while trying to minimize the overfitting of the training set. The parameters used for *XGBoost* are given below in Table 3.

The chemical age estimated from the best-fitting model relative to the input isochrone age determination is shown in Figs 5 and 6. The training set is slightly overfit relative to the test set, with a scatter ~1.1 Gyr in the training set compared to the test set ($\sigma_\tau \sim$ 1.35 Gyr). However, this overfitting problem is less severe than in other methods we attempted (e.g. neural networks), while still giving age determinations for the test set that are more robust than those determined using Bayesian analysis. This model is able to reproduce the general age trend found in the test set, with a slight overestimate of ages for stars $\tau < 6$ Gyr and a slight underestimate in ages for stars $\tau > 6$ Gyr, i.e. the slope of the chemical age versus isochrone age relation is generally shallower at intermediate ages relative to the 1:1 line, as shown by the running median (orange line) in Fig. 6. While

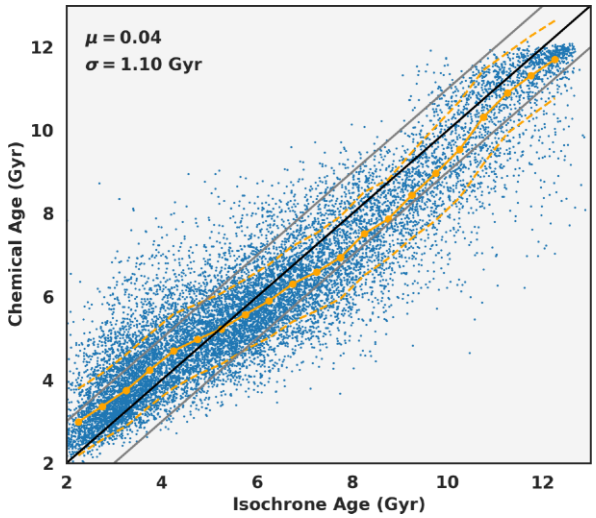

Figure 5. The chemical age versus age determined from isochrone matching for the training set of MSTO stars using *XGBoost*. The black line denotes a 1:1 correlation, while the grey lines denote ±1 Gyr. The orange line denotes the median and 1σ scatter about the relation.

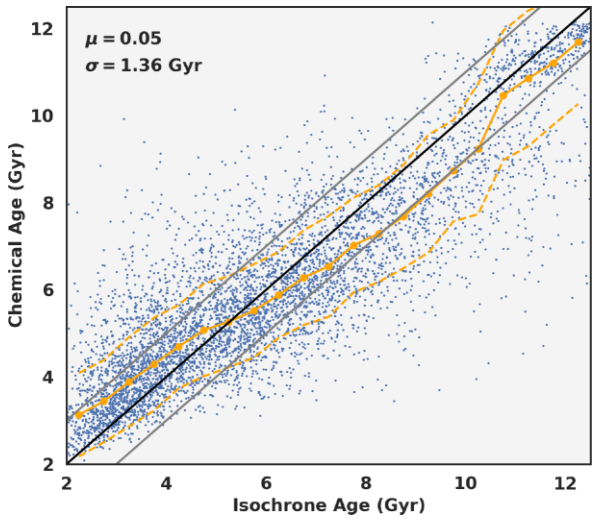

Figure 6. The chemical age versus age determined from isochrone matching for the test set of MSTO stars using *XGBoost*. The black line denotes a 1:1 correlation, while the grey lines denote ±1 Gyr. The orange line denotes the median and 1σ scatter about the relation. Note that the test set performs worse than the training set as *XGBoost* is slightly overfitting. However, even here the scatter is relatively small, and even the largest systematic age error is <1 Gyr for the worst-performing cases (age in the range 8–10 Gyr).

the spread in the recovered ages is only slightly smaller compared to that using the Bayesian chemical age determination method, the number of critical failures/outliers is significantly reduced when using *XGBoost*.

The relative importance of different abundances in determining the age using *XGBoost* is shown in Fig. 7. The feature gain, which measures how useful each element is in improving the accuracy of the age estimates, is shown in the top panel. Perhaps unsurprisingly, the [α/Fe] elements [Mg/Fe] and [Si/Fe] are at the top, followed by the s-process element [Y/Fe]. This makes sense from a qualitative perspective: we know the oldest stars are enhanced in α elements, so we have an immediate age indicator just from the [Mg/Fe] and

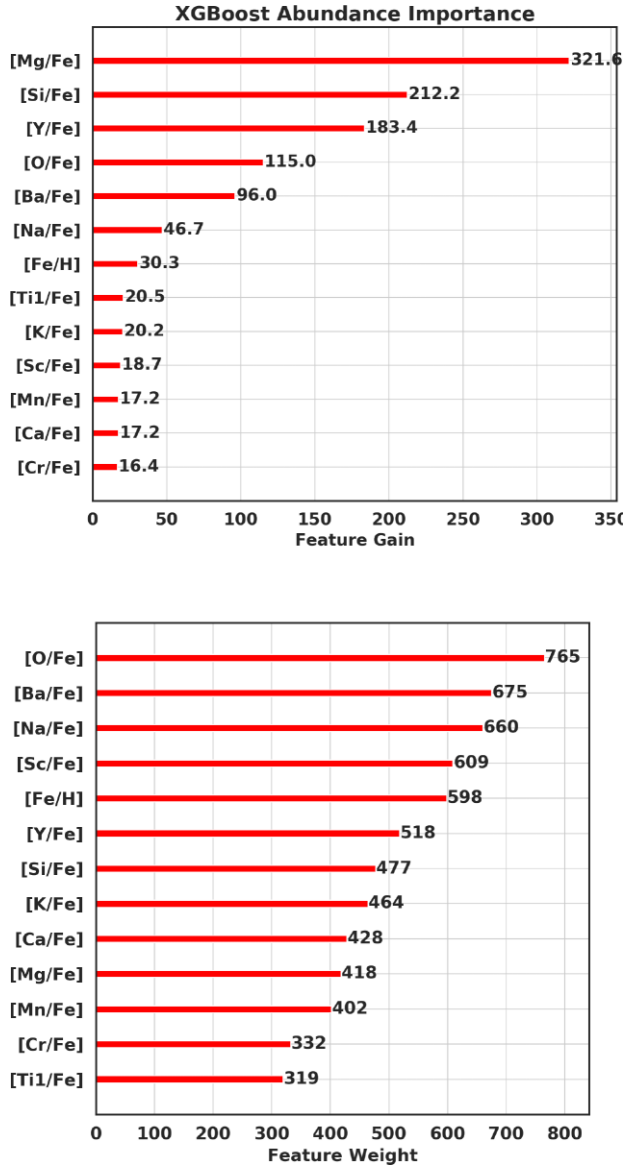

Figure 7. Top: The relative gain of different chemical abundances in determining the age of a star. Bottom: The relative weight of different chemical abundances in determining the age of a star.

[Si/Fe] abundance if a star is (relatively) young or old. For younger and intermediate-aged stars with lower [Mg/Fe] or [Si/Fe], [α/Fe] is less able to discriminate age. This is where the s-process elements become important in distinguishing young from intermediate-aged populations. The bottom panel of Fig. 7 is the relative weight of each element in *XGBoost*. The weight describes how often an element occurs in different decision trees in *XGBoost*. For this metric, the α elements are much lower down in the metric, with s-process, odd-Z, and iron peak elements being dominate, as they help distinguish between different locations within the Galaxy for a given age (i.e. for a given age there can be a range of metallicities which correspond to a different birth radius). In principal, we find that a fairly reliable age can be estimated with just three abundances: an iron peak element ([Fe/H]), an α element ([Mg/Fe]), and an s-process element ([Y/Fe]); increasing the number of elements used in the age determinations helps mitigate uncertainties in individual abundance measurements, so the more well-measured elements that can be used, the better.

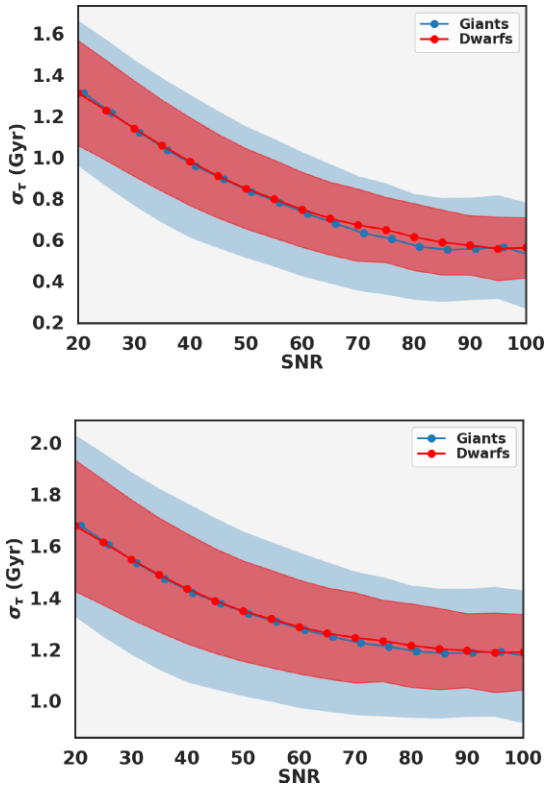

Figure 8. Top: The random age errors due to abundance uncertainties as a function of SNR. The giants and dwarfs have similar median trends, but the giants have significantly more scatter as a function of SNR than dwarfs. Bottom: The total uncertainty in the age estimates, combining the random and systematic errors.

Uncertainties in the age estimates are calculated by doing 1000 Monte Carlo runs of the input abundances through the model for each star, with the standard deviation of the age distribution being reported. The age distribution generated from the Monte Carlo run is generally a single-peaked Gaussian. However, there is an exception for stars with ages in the range 8–10 Gyr. These stars typically have an age distribution that is bimodal, and the ages are quite sensitive to errors in [α/Fe]. Stars in this age range generally have thin disc/solar-[α/Fe] measurements, but still have very low s-process abundances, similar to that of the thick disc population. This means that α elements are the dominant age indicator distinguishing them from the thick disc. For these older thin disc stars, an error in [α/Fe] of 0.1 dex, for example, might shift a star towards the thick disc as the other most age-sensitive abundances are not yet reliable for age determination.

The median age uncertainty due to random errors in the individual abundance measurements as a function of SNR is shown in the top panel of Fig. 8. The median random errors for dwarfs and giants as a function of SNR are comparable, and between 0.5 and 1.5 Gyr, with the error in the age decreasing as signal to noise increases. The giants have a larger scatter as a function of SNR, however, as the uncertainties in abundances are larger for giants. We estimate the systematic errors using the results of the MSTO test set. As we have calculated the random errors due to abundance uncertainties for the test set stars, the systematic errors inherent in the method can be estimated using the scatter about the relation found in Fig. 6 while taking into account the random errors in the test set stars based on their SNR. We determine the mean systematic error for the *XGBoost* method by subtracting in quadrature the standard deviation of the relation for the test set (1.36 Gyr) from the mean random uncertainties due to abundance errors of the test set MSTO stars. As the test set is made up of high-SNR spectra, the random errors due to abundance uncertainties are ∼0.7–1.0 Gyr, so the systematic errors inherent in the method is typically dominant. We find a mean systematic uncertainty of ∼1 Gyr for the test set. This systematic error may be due to errors in the input ages for the MSTO training set, as well as deviations from the age–abundance relations in the Milky Way itself. The total uncertainty for each star is then determined by adding in quadrature the random uncertainty from the MC run due to abundance uncertainties along with the systematic uncertainty of 1 Gyr inherent from our test set. The global uncertainty is shown in the bottom panel of Fig. 8; the typical age errors for each star are ∼1–2 Gyr. Our age errors are roughly independent of age (see Figs 5 and 6) and are primarily then a function of the systematic error budget and the SNR of each star. This means that younger stars will have fractionally larger age errors than for older stars. These results are more robust than those determined using Bayesian statistics, with smaller scatter and reduced rate of catastrophic failures, and we elect to use the ages derived from *XGBoost* going forward.

The age–metallicity relation for MSTO stars is shown in Fig. 9. The trends between the input training set from the isochrones are quite similar to the trends found using chemical ages; the age determined from chemistry has slightly less scatter as expected, given that the ages are determined directly from the chemistry in this case. The global features of the relations are present and in the same locations: the oldest stars are [α/Fe] enhanced, and there is a trend of decreasing [α/Fe] combined with increasing metallicity and decreasing age for these thick disc stars. Once $\tau \lesssim 9$ Gyr, however, the evolution of [α/Fe] with age is shallower, mirroring the results from previous studies (e.g. Tinsley 1980; Nordström et al. 2004; Casagrande et al. 2011; Haywood et al. 2013; Bergemann et al. 2014).

3.3 The impact of atomic diffusion and convective mixing length variation

The photospheric chemical abundances (i.e. those we estimate) can be different from the bulk chemical composition of a star due to atomic diffusion. Atomic diffusion is a catch-all term to describe the various transport processes that are most efficient in the radiative zones of stars, driven by gradients in pressure, temperature, and concentration. Differences between the surface and bulk composition can lead to age errors of the order of 20 per cent for MSTO stars (see e.g. Thoul, Bahcall & Loeb 1994; Dotter et al. 2017; Liu et al. 2019), becoming preferentially larger for older stars. Additionally, for MSTO stars in particular, atomic diffusion can cause abundance underestimates of ∼0.1 dex in [Fe/H] relative to the bulk composition (e.g. Michaud, Richer & Richard 2013). This means that atomic diffusion, if not taken into account, can cause an error in our calibration sample both in the age determination and in the chemical abundance relations used to derive the ages for the rest of the sample. The version of the PARSEC isochrones used in this analysis do take atomic diffusion into account, and we use the bulk composition in our age estimates, so the effect on age determination of MSTO stars is mitigated. Still, there remains the potential for diffusion to effect the abundances of our training set. The overall impact of this is likely small, given that the magnitude of the effect is only slightly larger than our abundance uncertainties, and most importantly because the abundance variation of [X/H] varies roughly in lock step with [Fe/H] (Michaud et al. 2013). This means that the [X/Fe] ratio used in our analysis will be relatively unaffected, and the primary error in a chemical age determination will then be due to any [Fe/H] offset between the surface and bulk composition.

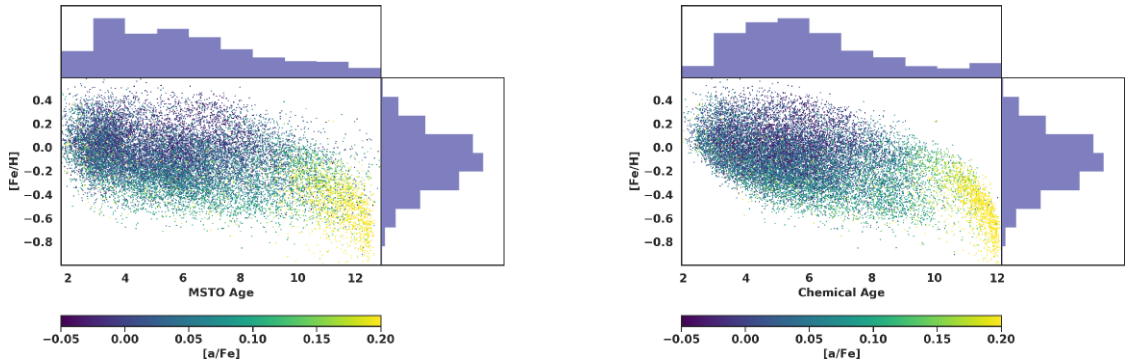

Figure 9. Left-hand panel: The age–metallicity relation for MSTO stars determined from isochrone ages. Right-hand panel: The age–metallicity relation determined from chemical ages.

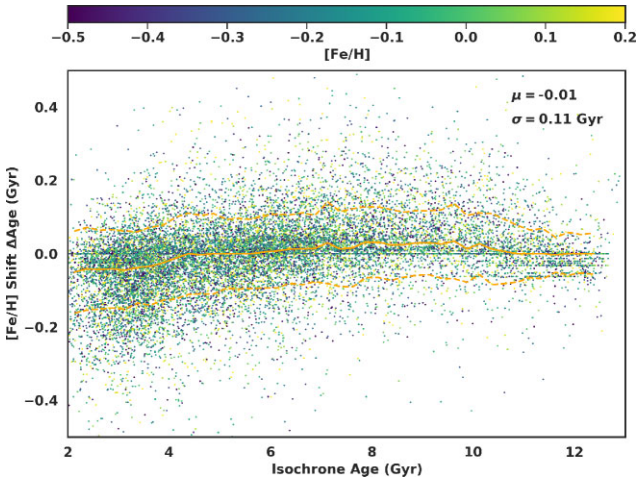

Figure 10. The difference in derived chemical age due to a shift of −0.05 dex in [Fe/H] for the MSTO training and test sets, as a function of the BSTEP age derived from isochrone matching. The training and test sets span the metallicity range Fe/H] < +0.5, but the colourbar has been truncated to maximize contrast.

This [Fe/H] offset could then yield a systematic error in age of up to 20 per cent in the chemical age of a star, but is in general much smaller than this value as [Fe/H] is not as important as α or s-process elemental abundances in the age estimations. To estimate the impact of a shift in [Fe/H] of −0.05 dex, we calculate the difference in derived chemical age in our training and test set stars due a shift of −0.05 dex, as shown in Fig. 10. We find that a shift of −0.05 dex in [Fe/H] results in an age error of 0.1 Gyr on average for the training and test sets, although there are stars for which the shift is much larger, up to 0.5 Gyr. There is no significant trend with derived age or original metallicity. However, for more than 90 per cent of our training and test sets, the difference in chemical age due to a shift of −0.05 dex in [Fe/H] is less than 0.2 Gyr. Increasing the [Fe/H] shift to −0.1 dex increases the age scatter up to 0.18 Gyr, on average. Overall, the impacts of these effects are relatively minor compared to the input uncertainties from isochrone matching, and contribute to the systematic error budget by making the age–abundance relations more difficult to estimate as additional scatter of the order of 0.1 Gyr is input into the relations.

Variations have been found in the convective mixing length with metallicity (e.g. Tayar et al. 2017), and there are also hints that there could be variations due to mass as well (Joyce & Chaboyer 2018b). The primary impact of this variation in the convective mixing length is to cause an offset between the observed effective temperature of a star and the temperature of a star of that mass and metallicity in model isochrones, if the variation in mixing length is not accounted for in the models. As temperature is a critical component to the age estimation of MSTO stars, any offsets between observations and isochrones will therefore cause an error in the age determination. The impact on MSTO stars is largest for older (lower mass) and more metal-poor stars; Viani et al. (2018) find for example a 0.3 difference in the mixing length parameter for an old metal-poor star relative to the solar mixing length, which causes a 200 K shift towards cooler temperatures of the models on some portions of the turn-off region. Similarly, in their study of metal-poor stars with [Fe/H] ∼−2, Joyce & Chaboyer (2018a) find variations of the mixing length of 0.3, which also causes a shift of ∼150 K cooler along the MSTO region. This means that the ages would be underestimated compared to reality. However, the areas of the turn-off region which are most effected by this are relatively small (the turn-off itself only, see Fig. 16 of Viani et al. (2018)), and the majority of the stars in our training set lie further along the MSTO or subgiant branch where the effective temperature difference between models is small. For the majority of our training set, our age determinations will be unaffected. For the older, more metal-poor stars, we might underestimate the age of an MSTO star by up to ∼15 per cent based on a temperature shift of 200 K in the isochrones. There is an additional potential issue with radiative acceleration for the youngest stellar populations ($M_{\odot} > 1.5$, corresponding to ages <2.7 Gyr for solar metallicity), which may cause variations in surface abundances for MSTO stars at these masses (Turcotte, Richer & Michaud 1998; Deal et al. 2018). However, even in the worst case, this will only impact age–abundance relations for these very young stellar populations, which do not make up a large portion of our sample.

As a whole, these effects are relatively minor compared to our global uncertainties, but do add to the systematic error budget. Particularly, these effects could be another possible explanation for the slight discrepancy between the chemical and isochrone ages for stars with 8 Gyr < τ < 10 Gyr, or for potential issues around stars with ages <3 Gyr.

4 RESULTS

4.1 Age–abundance relations

In the ideal case, the age-abundance trends for the chemical ages should match the training set age-abundance trends exactly, as the

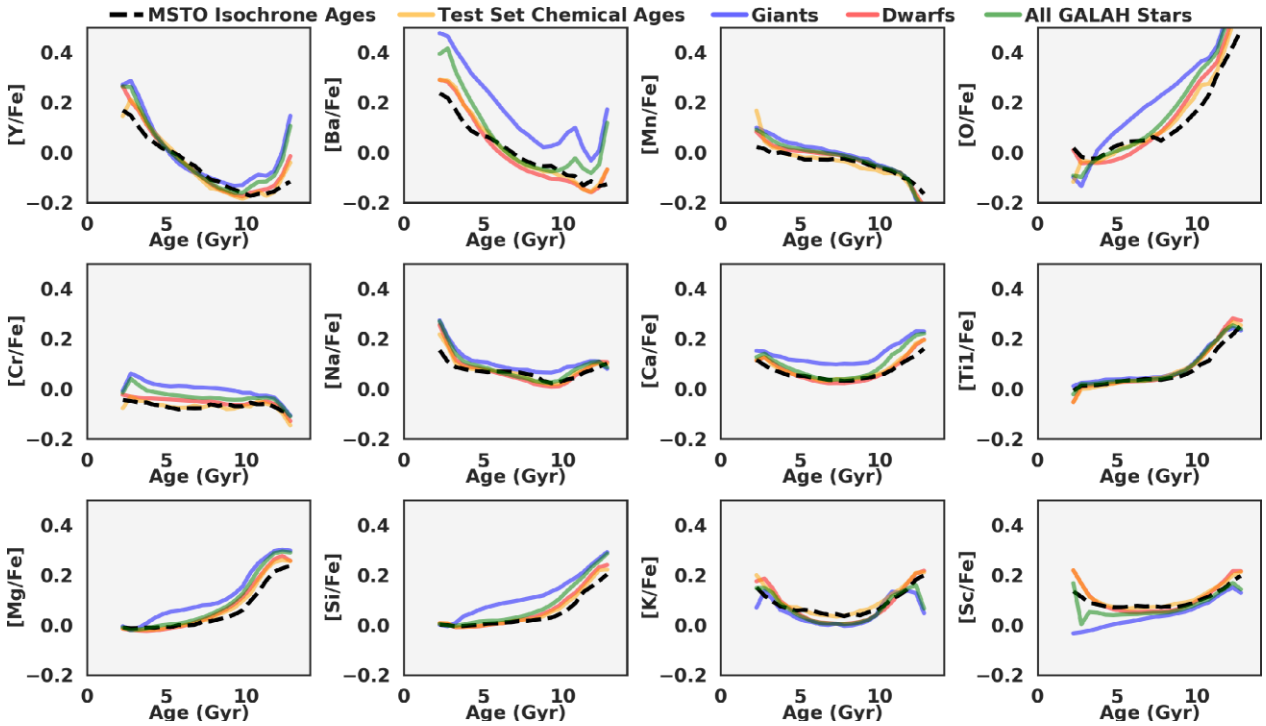

Figure 11. The age-abundance trends for the various GALAH DR3 samples analysed in this paper. Note the large discrepancy for the s-process elements [Y/Fe] and [Ba/Fe] for the giants.

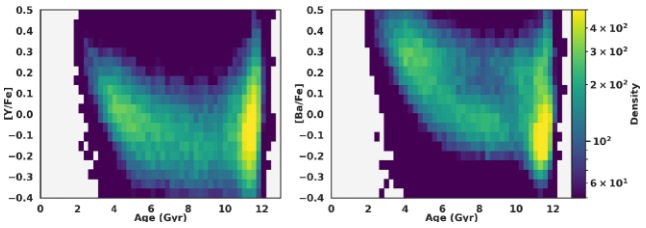

Figure 12. The age-abundance trends of [Y/Fe] and [Ba/Fe] for giant stars. Note the large scatter in these s-process abundances for old stellar populations. This highlights a potential issue in the s-process abundance estimates for the old, metal-poor thick disc populations.

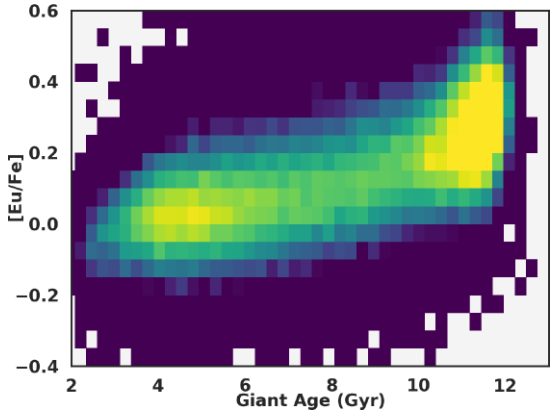

Figure 13. The age-[Eu/Fe] trend for giants. [Eu/Fe] is an r-process element and should have a strong trend with age, which we recover in our analysis. [Eu/Fe] was not one of the elements used in the age generation. This means that the age determinations for giants are robust, despite the issue with the s-process abundances.

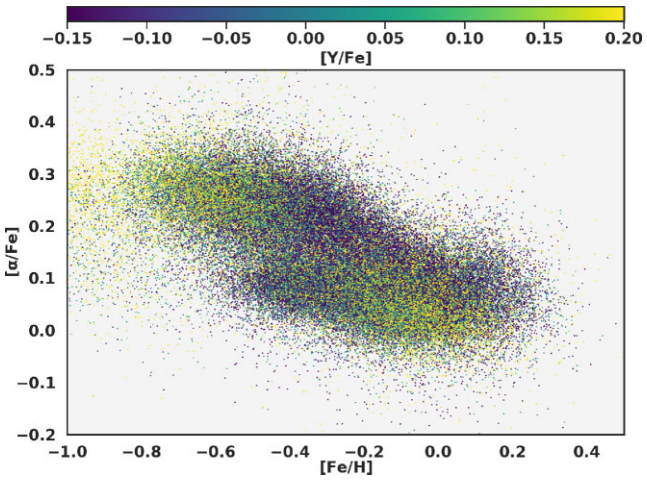

Figure 14. The [α/Fe] versus [Fe/H] plane for the giants. It is expected that young stars have higher [Y/Fe] than older stars, as [Y/Fe] is an s-process element. We identify a potential issue with the [Y/Fe] abundance for the metal-poor thick disc in the giant subsample ([Fe/H <-0.5, [α/Fe] > 0.15), where many stars have enhanced [Y/Fe]. These stars should be among the oldest in the disc based on their belonging to the metal-poor thick disc.

chemical abundances themselves are being used to determine the ages. For the most part, we find that this is true, and the same abundance trends found in the training set are well matched to those for the entire sample derived with chemical ages, as shown in Fig. 11. These results closely match the age-abundance trends measured in Sharma et al. (2022). However, there is a particular issue with the [Y/Fe] and [Ba/Fe] abundances for giants, where there is a strong discrepancy between the expected output and what is observed, with the [Y/Fe] and [Ba/Fe] abundances having large scatter for the oldest

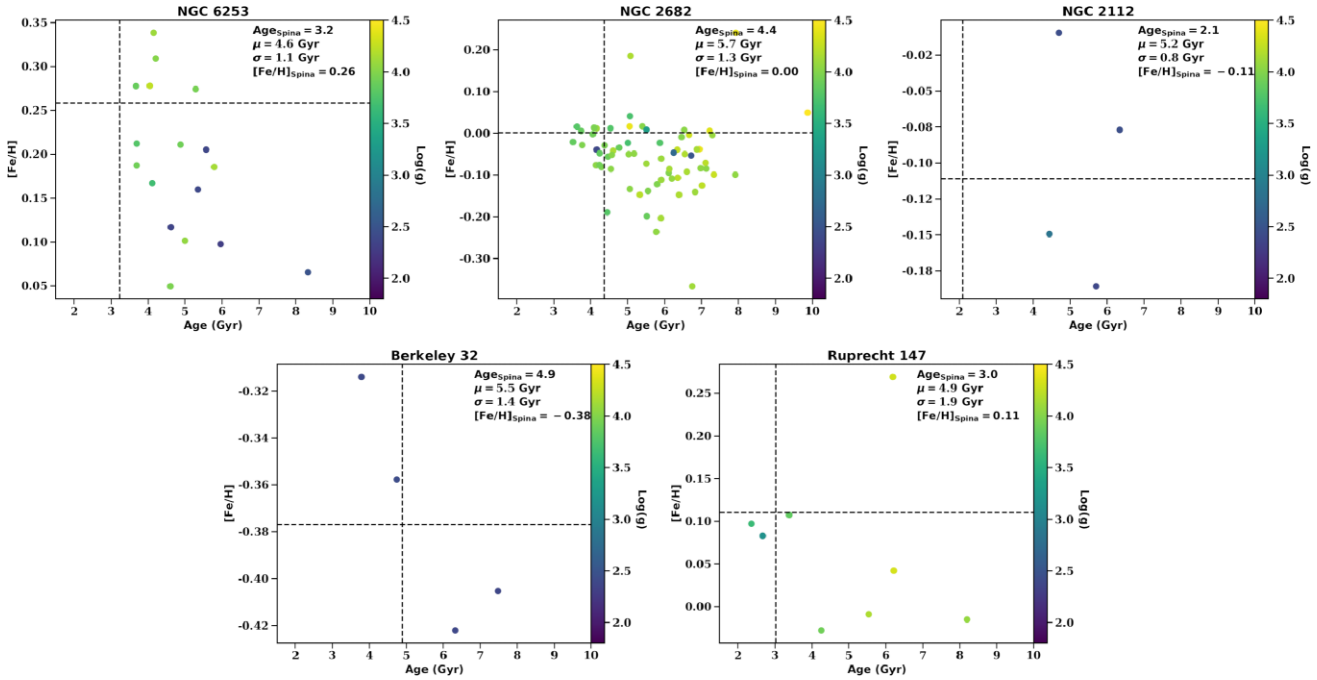

Figure 15. The age determinations for five old open clusters observed with GALAH, compared to the age determinations of Spina et al. (2021). The median chemical age and standard deviation for each cluster are shown in the top right corner of each panel. The dashed lines show the values from Spina et al. (2021) for each cluster. We find that for most clusters, the chemical ages recover the reference age to within 1 Gyr when the metallicity is measured accurately to within 0.075 dex. Metallicity errors larger than this have correspondingly larger age errors. The exception is NGC 2112, where the ages are significantly overestimated for all stars regardless of metallicity.

stellar populations, which is not present in the MSTO or dwarf subsamples. The large scatter in [Y/Fe] and [Ba/Fe] for the giant subsample is shown in Fig. 12.

The expectation for [Y/Fe] and [Ba/Fe], as s-process elements, is for low abundances at old ages, and high abundances at young ages, as is measured in the MSTO and dwarf samples; this trend is not reproduced in the giants highlighting an issue either with the abundance of [Y/Fe] and [Ba/Fe] or with the chemical age determination. We perform several tests to validate the reliability of the giant ages in light of the issue surrounding the s-process abundances for old stars. The [α/Fe] versus [Fe/H] plane for the giant sample is shown in Fig. 13, with the colour code being the [Y/Fe] abundance. The problem with the [Y/Fe] measurements for the oldest and most metal-poor stars is immediately apparent: there are a large fraction of stars with [α/Fe] and [Fe/H] belonging to the thick disc which have high [Y/Fe] measurements. These stars are likely some of the oldest in the Galaxy, based purely on their [α/Fe] and metallicity: this is the metal-poor end of the thick disc, so their [Y/Fe] should therefore be low. This hints at a problem in the [Y/Fe] and [Ba/Fe] abundances for metal-poor high-[α/Fe] stars. Luckily, for stars belonging to the thick disc, the [Y/Fe] and [Ba/Fe] are not useful tracers for age, as AGB stars have not produced large amounts of s-process elements, and the age can be derived almost entirely from the [α/Fe] and [Fe/H] abundances alone (see e.g. Haywood et al. 2013; Hayden et al. 2017). We do note, however, that studies of globular clusters have also found anomalous populations of s-process enhanced stars (e.g. Johnson & Pilachowski 2010); it is possible a similar enrichment process has happened for some of our metal-poor giant sample. This warrants further investigation, but will not impact the results of this paper.

The question of reliability for the [Y/Fe] and [Ba/Fe] and their impact on the age determination is then primarily for the thin disc, i.e. stars with age <9 Gyr. For these stars, [α/Fe] is potentially a less sensitive age indicator, as while there are age trends of [Fe/H] and [a/Fe] for the low-alpha disc (see Fig. 2), the narrow alpha range spanned by these populations is potentially within our measurement errors. The [Y/Fe] and [Ba/Fe] therefore become critically important for reliable age determination for these stellar populations. For younger stars, the giant age-abundance trends match closely (modulo some zero-point offsets) to those of the dwarfs and MSTO stars, which is an encouraging sign and hints that the [Y/Fe] and [Ba/Fe] are well measured for stellar populations where these abundances are important for the age measurement. Further, we have additional abundance tracers that were not used in the age determinations available for giants for which to test the age reliability. As mentioned in the data section, [Eu/Fe] is not well measured in the MSTO or dwarf samples, but is measured in the majority of giants. As an r-process element, [Eu/Fe] is expected to have a strong age-abundance trend (e.g. Sharma et al. 2022). This makes [Eu/Fe] a useful diagnostic in our chemical age determinations, as we can measure the [Eu/Fe] versus chemical age trend and compare this to the expected trends from other studies. We find a strong trend of increasing [Eu/Fe] as a function of chemical age as shown in Fig. 14, as expected for an r-process element. Our result closely matches the age-[Eu/Fe] trends measured in other studies (e.g. Sharma et al. 2022). This highlights that the ages derived for giants from chemical abundances are likely trustworthy, albeit with increased age errors on individual stars, and the issue with the [Y/Fe] and [Ba/Fe] abundances only affects thick disc stars for which these abundances are not critically important in age determination. There is one exception to giant ages being reliable for our sample: stars belonging to the red clump. The red clump has a highly biased age distribution, peaking in the range 1–2 Gyr (Girardi 2016). Our training set of MSTO stars lies outside of this age range, and so the ages for the young red clump

stars; when comparing to a subset of APOGEE observations, we find that the young RC has additional systematics with age underestimates of about 2 Gyr.

4.2 Open clusters

Spina et al. (2021) have created a catalogue of many open clusters throughout the disc which have been observed by the ongoing spectroscopic surveys. Several of these clusters have ages older than 2 Gyr and have been observed by GALAH, and thus make an ideal test of the chemical clock age estimates. The results for this test are shown in Fig. 15. In general, the ages of the open clusters are well recovered with the chemical clocks when the metallicity of the star is close to the reference value. For example, in Ruprecht 147 the three stars with metallicity estimates within 0.05 dex of the reference value are all 3 Gyr old, the measured age of the cluster. This also holds true for NGC 6253, where the stars at high metallicity close to the reference value are within 0.5–1 Gyr of the cluster; it is only where the measured metallicity is off by more than 0.1 dex that the age estimates begin to fail. The exception is NGC 2112, for which all of the stars are 3 Gyr older than the reference age at all metallicities. We note that the stars for which the metallicities are significantly different than the reference values also tend to have differences in their abundance ratios relative to the reference as well. This is why the chemical clocks perform worse here than in the test carried out in Fig. 10, as in that test for the impact of atomic diffusion, only [Fe/H] was varied and the abundance ratios were fixed. In this case, a poor measurement error of the metallicity has also impacted the other abundances, which has caused less precise age determinations from the chemical clocks.

4.3 Kinematic trends

We demonstrate both the utility and accuracy of ages derived from chemical abundances by reproducing some of the recent results on the kinematic structure of the disc (e.g. Mackereth et al. 2019; Sharma et al. 2021a). These results also highlight that the potential issue with the [Y/Fe] and [Ba/Fe] abundances for giants belonging to the metal-poor thick disc does not dramatically impact the age determination, as the recovered kinematic relations generally agree for both giants and dwarfs. As shown in Fig. 16, we reproduce the age–velocity dispersion relation for the solar neighbourhood (7.1 kpc $< R <$ 9.1 kpc, $|z| <$ 0.5 kpc), and find that all of the subsamples agree well to within the errors in both the measured ages and the velocities. This result also closely matches previous studies of the vertical velocity dispersion in the solar neighbourhood, who find a low velocity dispersion of ∼10–15 km s^{-1} for younger stellar populations, which increases with age up to values in the range 40–50 km s^{-1} for the oldest stars in the disc (Nordström et al. 2004; Haywood et al. 2013; Feuillet et al. 2016; Hayden et al. 2017; Yu & Liu 2018; Sharma et al. 2021a). Some authors have found a step function in the velocity dispersion between low- and high-[α/Fe] populations (e.g. Silva Aguirre et al. 2018; Miglio et al. 2021), but this is likely due to the definition of low- and high-[α/Fe] and imposing a hard [α/Fe] cut on the boundary between thin and thick discs, when the chemical distributions are actually rather continuous (see Sharma et al. 2021b). We find that the velocity dispersion increases smoothly with age and that there is no step function in the estimated dispersion, although there is a clear inflection point around ∼8–9 Gyr, with a steeper increase in dispersion with age for older stellar populations.

We also apply the data set to cover a larger fraction of the Galaxy, and reproduce the results shown in Sharma et al. (2021a). Sharma

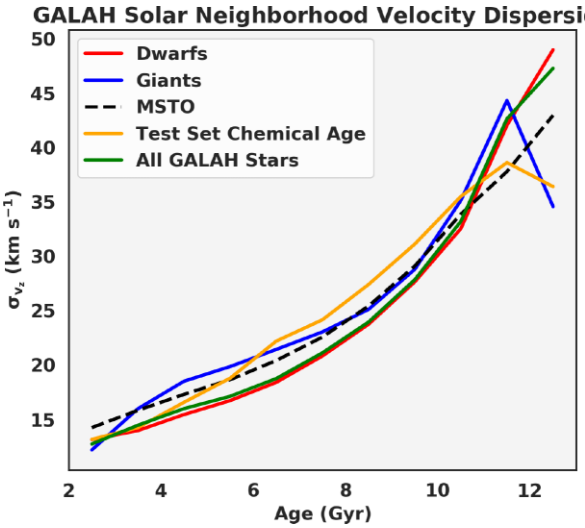

Figure 16. The vertical velocity dispersion for different samples as a function of age in the solar neighbourhood (7.1 kpc $< R <$ 9.1 kpc, $|z| <$ 0.5 kpc).

et al. (2021a) measure how the velocity dispersion of the disc varies with age, angular momentum, [Fe/H], and height about the plane ($|z|$). We apply the same formalism used in their paper to disentangle how the velocity dispersion depends on these parameters, as outlined below. The dispersion σ_v of velocity v (for either v_R or v_z), is assumed to depend on the stellar age τ, angular momentum L_z, metallicity [Fe/H], and vertical height from the disc mid-plane z, via the following multiplicatively separable functional form:

$$\sigma_v(X, \theta_v) = \sigma_v(\tau, L_z, [\mathrm{Fe/H}], z, \theta_v) = \sigma_{0,v} f_\tau f_{L_z} f_{[\mathrm{Fe/H}]} f_z. \quad (5)$$

Here, $X = \{\tau, L_z, [\mathrm{Fe/H}], z\}$ is a set of observables that are independent variables and

$$f_\tau = \left(\frac{\tau/\mathrm{Gyr} + 0.1}{10 + 0.1}\right)^{\beta_v}, \quad (6)$$

$$f_{L_z} = \frac{\alpha_{L,v}(L_z/L_{z,\odot})^2 + \exp[-(L_z - L_{z,\odot})/\lambda_{L,v}]}{1 + \alpha_{L,v}}, \quad (7)$$

$$f_{[\mathrm{Fe/H}]} = 1 + \gamma_{[\mathrm{Fe/H}],v}[\mathrm{Fe/H}], \quad (8)$$

$$f_z = 1 + \gamma_{z,v}|z|, \quad (9)$$

and $\theta_v = \{\sigma_{0,v}, \beta_v, \lambda_{L,v}, \alpha_{L,v}, \gamma_{[\mathrm{Fe/H}],v}, \gamma_{z,v}\}$ is a set of free parameters. We assume the same values used in their paper outlined in their table 2 for these free parameters.

Our results, shown in Fig. 17, match closely with the results of Sharma et al. (2021a) (fig. 1 in their paper) for all of our samples. Briefly, we find that the velocity dispersion increases with age following a shallow power-law slope. The velocity dispersion as a function of the angular momentum, represented by the guiding radius $R_g \sim \frac{L_z}{\Phi_\odot}$ (i.e. the angular momentum divided by the circular velocity at the solar position), is high in the inner Galaxy, while being lower in the solar neighbourhood with a flattening, or perhaps hints of an increase, in the outer disc. As GALAH is a Southern hemisphere survey, it is difficult to probe the outer Galaxy as compared to the analysis in Sharma et al. (2021a), who also utilized observations from LAMOST (Deng et al. 2012) to determine the relation in the outer disc. The dependence on metallicity is roughly linear, although there is an S shape present in the dispersions for the dwarf stars relative to the model; this is still similar to what is observed by Sharma et al. (2021a). The dependence on metallicity is the only panel that shows

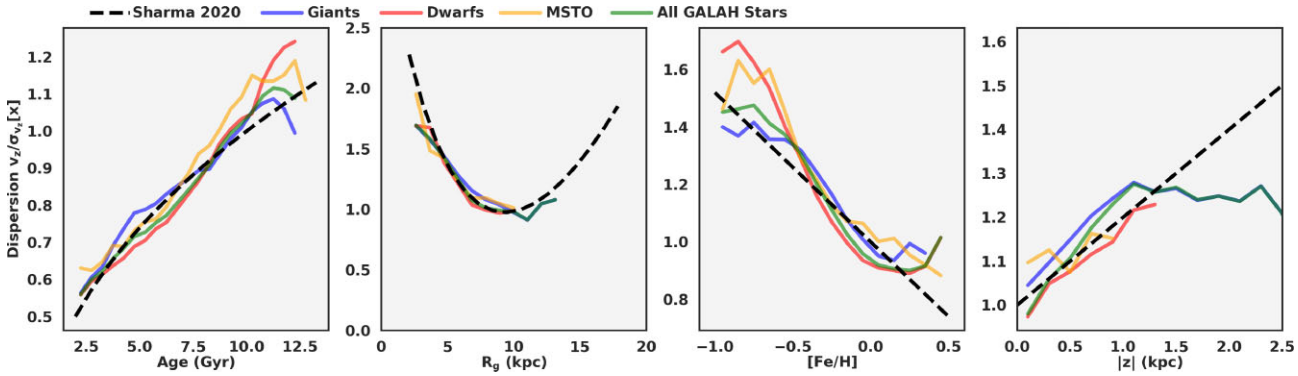

Figure 17. The vertical velocity dispersion for different samples as a function of age, angular momentum (represented by the guiding radius R_g), [Fe/H], and height above the plane. Note that the test set relation is omitted in this figure due to insufficient sampling.

any difference on the velocity dispersion relation between the various subsamples characterized in this paper, with the giants following the linear trend more closely, but even here the difference is relatively minor between dwarfs and giants. The only major difference in our results compared to Sharma et al. (2021a) is the vastly increased coverage of the thick disc in the sample presented in this paper, showing that while a linear dependence on $|z|$ is a good fit for stars up to $|z| < 1.5$ kpc, it may not hold for stars high above the plane. The vastly increased sample size in our analysis, in particular for the thick disc, is made possible by the use of chemical ages, rather than being forced to rely on MSTO or *Kepler* asteroseismic values for age estimates which have a more limited Galactic coverage.

5 DISCUSSION

We calculate the ages of stars in GALAH DR3 using their measured chemical abundances. With these chemical ages, we are able to reproduce the global kinematic properties of the disc, such as the age–angular momentum–velocity dispersion relation found by Sharma et al. (2021a). We have demonstrated that ages can be accurately predicted for stars with a large range of birth radii, metallicity, and chemical abundances. These results, as well as those in Ness et al. (2019) and Sharma et al. (2022), demonstrate that the age, metallicity, and overall chemical abundance do not vary dramatically at a given birth radius.

The results presented here, as well as those in Ness et al. (2019) and Sharma et al. (2022), highlight that the variation in abundances at a given place in time in the Galaxy is quite small, ∼0.02–0.04 dex. This means that strong chemical tagging, particularly for stars belonging to the thin disc, may prove to be difficult with our current abundance precision. The deviations from age-abundance trends measured by Ness et al. (2019) are roughly the same size as the abundance measurement uncertainties, and Sharma et al. (2022) find similar results using GALAH observations covering a wider range of chemical abundances and Galactic spatial coverage. To first order, this means that the dominant factor in the abundance of a star is its age and birth radius. Early strong chemical tagging efforts find many groups in chemical space, but have been unable to recover open clusters observed in major spectroscopic surveys relative to the background disc stars (e.g. Price-Jones et al. 2020). Verification of these groups is difficult, as our results demonstrate that by having similar chemical abundances, stars grouped in chemical space will by necessity have similar age. This means that having a similar age or falling along an isochrone cannot be used as validation for strong chemical tagging, as stars will by definition meet this criteria simply by having similar chemical abundances. Until it is demonstrated that open clusters can be chemically tagged reliably, claims of strong chemical tagging should be viewed with some scepticism, given our results and the arguments laid out in Ness et al. (2019) and Sharma et al. (2022).

However, it is possible that future improvements in the precision of abundance determination, as well as including additional elements, could make strong chemical tagging more tangible going forward. For instance, Bland-Hawthorn et al. (2010) estimate that abundance precision of 0.01 to 0.02 dex is needed to enable reliable strong chemical tagging. This is still likely lower than our current uncertainties, but we are approaching this value, particularly with data-driven machine learning methods such as the Cannon (Ness et al. 2015) or Astro-NN (Leung & Bovy 2019). Continued improvements in abundance determination, such as improved corrections for non-local thermal equilibrium effects (see e.g. Amarsi et al. 2020) as well as better reduction techniques to eliminate issues such as fibre cross-talk (Kos et al. 2018), could mean that we are close to the required abundance precision necessary to enable strong chemical tagging.

For the Milky Way, the fact that abundances can be determined merely by a stars age and metallicity (Ness et al. 2019; Sharma et al. 2022), or conversely accurate ages directly from abundances, likely means that the gas azimuthal mixing time-scales are quite short. Strong azimuthal abundance variation would yield a much larger scatter in these relations and hints that the Milky Way must be fairly uniform with azimuth. However, recent observations of H II regions in the Milky Way make claims of azimuthal metallicity variation across the Galaxy (Wenger et al. 2019). The results from the H II regions are difficult to mesh with the conclusions of Ness et al. (2019), Sharma et al. (2022), and this paper; it is possible that the results coming from H II regions is due to large uncertainties in the distances to Galactic H II regions. Observations of azimuthal abundance variation in external spiral galaxies are inconclusive: some galaxies have azimuthal variation, often correlating with metallicity peaks along the spiral arms or due to interactions with the bar, while other galaxies with seemingly similar properties do not (e.g. Kreckel et al. 2019). This presents a somewhat muddled view of azimuthal abundance variations and it is unclear how the Milky Way fits into this picture, other than that at least for most of its history the Milky Way seems to have been fairly well mixed azimuthally, as otherwise the age–abundance relations and the chemical age determinations would not work as well as they do. It may be that azimuthal abundance variations are a transient phenomena; one possibility for the apparent underestimate of the chemical ages for the training and test set stars with 8 Gyr $< \tau <$ 10 Gyr is that azimuthal variation was larger during

this time period of the Milky Way's evolution than for the rest of its history.

While strong chemical tagging may prove difficult with current spectroscopic studies, this is not necessarily bad news for Galactic Archaeology. The results presented in this paper highlight the capability of weak chemical tagging, which enables a host of new opportunities and pathways for studying the Milky Way, and in particular the kinematic structure and evolution of the Galaxy. The fact that we can use chemistry alone to estimate a robust age for a star has important implications for future studies of the Milky Way. Previously, kinematic studies were often limited to MSTO stars or asteroseismic targets (Sharma et al. 2021a), or use chemistry where ages were not available (e.g. Hayden et al. 2018, 2020). The ability to use abundances to estimate a reliable age provides not only a greatly increased sample size, but also vastly increases the Galactic coverage. MSTOs probe a limited volume given the combination of their absolute magnitudes and the magnitude limits imposed by spectroscopic surveys; for example, in GALAH MSTO stars probe a volume of $\sim$1.5 kpc around the solar position. Relying on asteroseismic targets for ages restricts you to the *Kepler* and *K2* fields, or the bright magnitude limits of the *TESS* mission. Additionally, asteroseismic ages also have much larger uncertainties than those in MSTO stars, and potentially the ages presented in this paper. For example, the age–[α/Fe] relation found in studies of asteroseismic giants has significantly more scatter than those of MSTO stars (see e.g. Silva Aguirre et al. 2018; Miglio et al. 2021).

Applying the methods outlined in this paper is also possible for other completed, ongoing, and future spectroscopic surveys. In particular, the existing data sets of LAMOST and APOGEE and the upcoming surveys 4-MOST (de Jong et al. 2014) and WEAVE (Dalton et al. 2014) are prime targets for such an analysis. In the ideal case, putting these surveys on the same abundance scale (and, by the methods outlined in this paper, the same age scale) would enable them to be used together, providing a sample of several million stars with reliable ages determined directly from chemical abundances. A data set of this magnitude would allow for precise constraints on important kinematic properties that govern the secular evolution of the disc such as the efficiency and time-scales of blurring and migration. It would also allow, for example, a detailed study of the age structure of the ridges observed in V_ϕ-R space (Antoja et al. 2018; Khanna et al. 2019), the phase-spiral (Antoja et al. 2018; Bland-Hawthorn et al. 2019), or Galactoseismology (Widrow et al. 2012; Bland-Hawthorn & Tepper-Garcia 2021).

The results presented here also highlight the need for future spectroscopic surveys to cover as many nucleosynthetic channels as possible, while also providing high-quality and high-SNR spectra. As clearly seen in Fig. 8, the age precision derived from chemical abundances is a direct function of the signal-to-noise ratio of the spectra. Higher quality observations will lead to a direct improvement in the age determinations. The need for precision abundances is also highlighted in Sharma et al. (2022): while the age-abundance trend for every element is dominated by the age and metallicity, there are still smaller second-order effects with signal to noise, $T_{\rm eff}$, and log g. The difficulty in measuring abundances to high precision across the H-R diagram is clearly demonstrated in the strange trends we recover with the s-process abundances for the giants in our analysis, as well as the small zero-point shifts in the age–abundance relations between dwarfs and giants. There is also significant room for improvement in our understanding and ability to correct for the impact of processes such as atomic diffusion, which primarily impact the reliability of our training set and increase the scatter in the age–abundance relations; it is possible that future improvements in this area could allow for even higher precision chemical clocks. There is also the potential for Galactic evolution effects on smaller scales, such as the offsets in the [Y/Mg] relations as a function of metallicity identified by Casali et al. (2020); these effects will be difficult to detect if other sources of uncertainty in the age–abundance relation dominate due to low-quality observations. Additional elements, particularly those coming from the s-process, will also provide tighter constraints on the age determination as well.

6 CONCLUSIONS

In this paper, we have determined ages for a large fraction of the stars in the GALAH survey with well-measured chemical abundances, using the *XGBoost* algorithm. The training set consists of high-SNR MSTO stars with precision ages determined from isochrone matching. This is the first time that the Galactic chemical evolution has been used to estimate the ages for a large sample of stars; previous attempts were limited to solar twins with very limited sample sizes and spatial coverage. The catalogue is publicly available and ideal for kinematic studies of the Galaxy. We demonstrate that the stellar ages derived from chemical abundances are accurate to 1–2 Gyr based on SNR and the location in the H-R diagram for individual stars, with giants having larger age uncertainties on average than dwarfs. We reproduce many of the age-kinematic results that have been observed since the advent of *Gaia*. In particular, we reproduce the age-velocity dispersion of the solar neighbourhood, and of the global relation found by Sharma et al. (2022), but with a greatly increased GALAH sample and much improved coverage of the thick disc. This form of weak chemical tagging, which enables an order-of-magnitude increase in the sample size as well as increasing the sample volume of stars which have reliable age estimates, enables more detailed studies of the kinematic structure of the disc.

When combined with the results from Ness et al. (2019) and Sharma et al. (2022), strong chemical tagging seems unlikely with our current abundance precision of $\sim$0.05 dex. The ability to measure precise ages and recreate the age-kinematic trends across the disc, as well as find very small dispersions in age–abundance relations, requires that the chemical abundance variation at a given radius and time is small. This is a boon for weak chemical tagging and the study of the kinematic structure of the Galaxy, but means that our current abundance precision is likely not high enough to engage in strong chemical tagging.

The reduced precision in abundances measured for giants in GALAH (and therefore the accuracy in the recovered ages) highlights the potential difficulty in applying this method for large samples spanning a large range of the H-R diagram. As improvements are made in our ability to measure abundances across the H-R diagram, as well as improvements in stellar modelling allowing for potential mixing processes due to stellar evolution to be quantified, the age precision presented in this paper maybe improved upon and differences between dwarfs and giants reduced. Going forward, the methods outlined in this paper can be applied to several of the existing and upcoming Galactic Archaeology surveys. If these surveys can be put on the same metallicity scale, reliable ages can be determined for millions of stars while also having an excellent coverage of the Galaxy spatially. This data set would be truly unprecedented, and when combined with *Gaia* DR3, our understanding of the formation and evolution of the Milky Way could undergo another gigantic leap forward.

ACKNOWLEDGEMENTS

This work is also based on data acquired from the Australian Astronomical Telescope. We acknowledge the traditional owners of the land on which the AAT stands, the Gamilaraay people, and pay our respects to elders past and present. This research was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work has used 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).

In addition to ASTRO3D, MRH received support from ARC DP grant DP160103747. LS acknowledges financial support from the Australian Research Council (discovery Project 170100521). JK and TZ acknowledge financial support of the Slovenian Research Agency (research core funding No. P1-0188) and the European Space Agency (PRODEX Experiment Arrangement No. C4000127986). KL acknowledges funds from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 852977). SLM and JDS acknowledge support from the UNSW Scientia Fellowship program and the Australian Research Council through grant DP180101791. YST is grateful to be supported by the NASA Hubble Fellowship grant *HST*-HF2-51425.001 awarded by the Space Telescope Science Institute.

DATA AVAILABILITY

This catalogue is available online through MNRAS, as well as locally at the University of Sydney here: http://physics.usyd.edu.au/~mhay5097/chemicalclocks/. The catalogue contains the object_id for a given star, the age estimate for that star, the total uncertainty in the age estimate, as well as flags on which stars were used during the training and testing phases.

REFERENCES

Amarsi A. M. et al., 2020, A&A, 642, A62
Antoja T. et al., 2018, Nature, 561, 360
Bedell M. et al., 2018, ApJ, 865, 68
Bensby T., Feltzing S., Oey M. S., 2014, A&A, 562, A71
Bergemann M. et al., 2014, A&A, 565, A89
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Bland-Hawthorn J., Tepper-Garcia T., 2021, MNRAS, 504, 3168
Bland-Hawthorn J., Krumholz M. R., Freeman K., 2010, ApJ, 713, 166
Bland-Hawthorn J. et al., 2019, MNRAS, 486, 1167
Borucki W. J. et al., 2010, Science, 327, 977
Bovy J., 2016, ApJ, 817, 49
Buder S. et al., 2021, MNRAS, 506, 150
Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, A&A, 530, A138
Casali G. et al., 2019, A&A, 629, A62
Casali G. et al., 2020, A&A, 639, A127
Casey A. R. et al., 2019, ApJ, 887, 73
Chen T., Guestrin C., 2016, preprint (arXiv:1603.02754)
Chiappini C. et al., 2015, A&A, 576, L12
Clarke A. J. et al., 2019, MNRAS, 484, 3476
Dalton G. et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9147, Ground-Based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham, p. 91470L
De Silva G. M., Sneden C., Paulson D. B., Asplund M., Bland-Hawthorn J., Bessell M. S., Freeman K. C., 2006, AJ, 131, 455
De Silva G. M. et al., 2015, MNRAS, 449, 2604
Deal M., Alecian G., Lebreton Y., Goupil M. J., Marques J. P., LeBlanc F., Morel P., Pichon B., 2018, A&A, 618, A10
de Jong R. S. et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9147, Ground-Based and Airborne Instrumentation for Astronomy V. SPIE, Bellingham, p. 91470M
Deng L.-C. et al., 2012, Res. Astron. Astrophys., 12, 735
Dotter A., Conroy C., Cargile P., Asplund M., 2017, ApJ, 840, 99
Feltzing S., Howes L. M., McMillan P. J., Stonkutė E., 2017, MNRAS, 465, L109
Feltzing S., Bowers J. B., Agertz O., 2019, MNRAS, 493, 1419
Feuillet D. K., Bovy J., Holtzman J., Girardi L., MacDonald N., Majewski S. R., Nidever D. L., 2016, ApJ, 817, 40
Frankel N., Sanders J., Ting Y.-S., Rix H.-W., 2020, ApJ, 896, 15
Freeman K., Bland-Hawthorn J., 2002, ARA&A, 40, 487
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2018, A&A, 616, A1
Girardi L., 2016, ARA&A, 54, 95
Hayden M. R., Recio-Blanco A., de Laverny P., Mikolaitis S., Worley C. C., 2017, A&A, 608, L1
Hayden M. R. et al., 2018, A&A, 609, A79
Hayden M. R. et al., 2020, MNRAS, 493, 2952
Haywood M., Di Matteo P., Lehnert M. D., Katz D., Gómez A., 2013, A&A, 560, A109
Jofré P. et al., 2016, A&A, 595, A60
Johnson C. I., Pilachowski C. A., 2010, ApJ, 722, 1373
Joyce M., Chaboyer B., 2018a, ApJ, 856, 10
Joyce M., Chaboyer B., 2018b, ApJ, 864, 99
Khanna S. et al., 2019, MNRAS, 489, 4962
Kobayashi C., Karakas A. I., Lugaro M., 2020, ApJ, 900, 179
Kos J. et al., 2017, MNRAS, 464, 1259
Kos J. et al., 2018, MNRAS, 480, 5475
Kreckel K. et al., 2019, ApJ, 887, 80
Leung H. W., Bovy J., 2019, MNRAS, 483, 3255
Lin J. et al., 2020, MNRAS, 491, 2043
Lindegren L. et al., 2018, A&A, 616, A2
Liu F., Asplund M., Yong D., Feltzing S., Dotter A., Meléndez J., Ramírez I., 2019, A&A, 627, A117
Mackereth J. T. et al., 2019, MNRAS, 489, 176
Majewski S. R. et al., 2017, AJ, 154, 94
Marigo P. et al., 2017, ApJ, 835, 77
Martig M. et al., 2016, MNRAS, 456, 3655
Masseron T., Gilmore G., 2015, MNRAS, 453, 1855
Michaud G., Richer J., Richard O., 2013, Astron. Nachr., 334, 114
Miglio A. et al., 2021, A&A, 645, A85
Minchev I. et al., 2018, MNRAS, 481, 1645
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., Johnston K. V., Blancato K., Rix H. W., Beane A., Bird J. C., Hawkins K., 2019, ApJ, 883, 177
Nissen P. E., 2015, A&A, 579, A52
Nissen P. E., 2016, A&A, 593, A65
Nissen P. E., Christensen-Dalsgaard J., Mosumgaard J. R., Silva Aguirre V., Spitoni E., Verma K., 2020, A&A, 640, A81
Nordström B. et al., 2004, A&A, 418, 989
Piskunov N., Valenti J. A., 2017, A&A, 597, A16
Price-Jones N. et al., 2020, MNRAS, 496, 5101
Rix H.-W., Bovy J., 2013, A&AR, 21, 61
Sanders J. L., Binney J., 2015, MNRAS, 449, 3479
Schönrich R., Binney J., 2009, MNRAS, 396, 203
Sellwood J. A., Binney J. J., 2002, MNRAS, 336, 785
Sharma S. et al., 2018, MNRAS, 473, 2004
Sharma S. et al., 2021a, MNRAS, 506, 1761
Sharma S., Hayden M. R., Bland-Hawthorn J., 2021b, MNRAS, 507, 5882

Sharma S. et al., 2022, MNRAS, 510, 734
Sheinis A. et al., 2015, J. Astron. Telesc. Instrum. Syst., 1, 035002
Silva Aguirre V. et al., 2018, MNRAS, 475, 5487
Spina L., Meléndez J., Karakas A. I., Ramírez I., Monroe T. R., Asplund M., Yong D., 2016, A&A, 593, A125
Spina L. et al., 2018, MNRAS, 474, 2580
Spina L. et al., 2021, MNRAS, 503, 3279
Tayar J. et al., 2017, ApJ, 840, 17
Thoul A. A., Bahcall J. N., Loeb A., 1994, ApJ, 421, 828
Ting Y.-S., Conroy C., Rix H.-W., 2016, ApJ, 816, 10
Tinsley B. M., 1980, Fundam. Cosm. Phys., 5, 287
Titarenko A., Recio-Blanco A., de Laverny P., Hayden M., Guiglion G., 2019, A&A, 622, A59
Turcotte S., Richer J., Michaud G., 1998, ApJ, 504, 559
Valenti J. A., Piskunov N., 1996, A&AS, 118, 595
Viani L. S., Basu S., Ong J. M. J., Bonaca A., Chaplin W. J., 2018, ApJ, 858, 28
Wenger T. V., Balser D. S., Anderson L. D., Bania T. M., 2019, ApJ, 887, 114
Widrow L. M., Gardner S., Yanny B., Dodelson S., Chen H.-Y., 2012, ApJ, 750, L41
Wittenmyer R. A. et al., 2018, AJ, 155, 84
Yu J., Liu C., 2018, MNRAS, 475, 1093
Zwitter T. et al., 2018, MNRAS, 481, 645

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

[1]*Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006 Sydney, Australia*
[2]*Centre of Excellence for Astrophysics in Three Dimensions (ASTRO-3D), Australia*
[3]*School of Physics and Astronomy, Monash University, Australia*
[4]*Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[5]*Research School of Astronomy & Astrophysics, Australian National University, ACT 2611 Canberra, Australia*
[6]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str 1, D-85741 Garching, Germany*
[7]*Faculty of Science and Engineering, Australian Astronomical Optics, Macquarie University, Macquarie Park, NSW 2113, Australia*
[8]*Macquarie University Research Centre for Astronomy, Astrophysics & Astrophotonics, Sydney, NSW 2109, Australia*
[9]*Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia*
[10]*Department of Astronomy, Stockholm University, AlbaNova University Centre, SE-106 91 Stockholm, Sweden*
[11]*School of Physics, UNSW, Sydney, NSW 2052, Australia*
[12]*Department of Physics and Astronomy, Macquarie University, 2109 Sydney, Australia*
[13]*Lund Observatory, Department of Astronomy and Theoretical Physics, Box 43, SE-221 00 Lund, Sweden*
[14]*Centre for Astrophysics, University of Southern Queensland, Toowoomba, QLD 4350, Australia*
[15]*Centre for Integrated Sustainability Analysis, School of Physics, The University of Sydney, NSW 2006 Sydney, Australia*
[16]*Institute for Advanced Study, Princeton, NJ 08540, USA*
[17]*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*
[18]*Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[19]*Johns Hopkins University, Dept of Physics & Astronomy, Baltimore, MD 21218, USA*

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