Dynamical heating across the Milky Way disc using APOGEE and Gaia


Affiliations are listed at the end of the paper

Accepted 2019 May 30. Received 2019 May 15; in original form 2019 January 11

ABSTRACT

The kinematics of the Milky Way disc as a function of age are well measured at the solar radius, but have not been studied over a wider range of Galactocentric radii. Here, we measure the kinematics of mono-age, mono-[Fe/H] populations in the low and high $\alpha$/Fe discs between $4 \lesssim R \lesssim 13$ kpc and $|z| \lesssim 2$ kpc using 65 719 stars in common between APOGEE DR14 and Gaia DR2 for which we estimate ages using a Bayesian neural network model trained on asteroseismic ages. We determine the vertical and radial velocity dispersions, finding that the low and high $\alpha$/Fe discs display markedly different age–velocity dispersion relations (AVRs) and shapes $\sigma_z/\sigma_R$. The high $\alpha$/Fe disc has roughly flat AVRs and constant $\sigma_z/\sigma_R = 0.64 \pm 0.04$, whereas the low $\alpha$/Fe disc has large variations in this ratio that positively correlate with the mean orbital radius of the population at fixed age. The high $\alpha$/Fe disc component’s flat AVRs and constant $\sigma_z/\sigma_R$ clearly indicate an entirely different heating history. Outer disc populations also have flatter radial AVRs than those in the inner disc, likely due to the waning effect of spiral arms. Our detailed measurements of AVRs and $\sigma_z/\sigma_R$ across the disc indicate that low $\alpha$/Fe, inner disc ($R \lesssim 10$ kpc) stellar populations are likely dynamically heated by both giant molecular clouds and spiral arms, while the observed trends for outer disc populations require a significant contribution from another heating mechanism such as satellite perturbations. We also find that outer disc populations have slightly positive mean vertical and radial velocities likely because they are part of the warped disc.

Key words: Galaxy: disc – Galaxy: evolution – Galaxy: formation – Galaxy: kinematics and dynamics – Galaxy: stellar content.

1 INTRODUCTION

The present-day kinematics of stars in the Milky Way is immutably tied to the dynamical history of the Galaxy. As a result, the measurement of the kinematics of the stellar disc as a function of age or element abundances offers tight constraints on models for the formation and evolution of the Galaxy (Rix & Bovy 2013). In the new era of Gaia (Gaia Collaboration 2016), our access to the kinematics of stars has increased by orders of magnitude. In addition, the recent advent of large-scale spectroscopic surveys (e.g. Gaia–ESO, Gilmore et al. 2012; APOGEE, Majewski et al. 2017; GALAH Martell et al. 2017) has delivered stellar spectra, and therefore element abundance measurements, on an unprecedented scale. However, the measurement of accurate stellar ages, especially those of red-giant stars, has been problematic (see Soderblom 2010, and references therein). On this front, recent advances, combining spectroscopic surveys with high-quality asteroseismic data (e.g. Pinsonneault et al. 2014, 2018), have led to great improvements in age measurements (e.g. Martig et al. 2016; Silva Aguirre et al. 2018), leading Galactic astrophysics into a truly multidimensional mode of operation.

Pioneering work that studied the variation of stellar velocities with indicators of their age demonstrated that there exists a positive correlation between age and velocity dispersion in the solar vicinity (Strömgren 1946; Roman 1950a,b). This property of the Galactic disc has since been well measured and characterized in the solar neighbourhood (e.g. Wielen 1977; Seabroke & Gilmore 2007;
Soubiran et al. 2008; Casagrande et al. 2011). Observational limitations have meant that this relationship is not yet well constrained throughout the disc that would allow a detailed assessment of its dynamical history and the heating processes that have shaped it.

Barbanis & Woltjer (1967) were among the first to show that transient spiral arms could contribute to disc heating, while Spitzer & Schwarzschild (1951, 1953) demonstrated similar effects from small-scale irregularities in the disc potential, such as those caused by giant molecular clouds (GMCs). Jenkins & Binney (1990) considered the combined effect of both these heating agents, finding that increasingly prominent spiral perturbations (over those from GMCs) tend to flatten the axis ratio of velocity dispersions (e.g. decreasing $\sigma_{\alpha}/\sigma_{\beta}$) overtime. However, these models all struggled to reconcile the predicted slope of the vertical age–velocity dispersion relation (AVR; usually modelled as a power law $\sigma \propto \tau^{\beta}$, where models predict $\beta \sim 0.25$; e.g. Lacey 1984; Hänninen & Flynn 2002) with that observed at the solar vicinity (where $\beta_{\odot}$ is closer to $\sim 0.5$; e.g. Wielen 1977; Seabroke & Gilmore 2007; Soubiran et al. 2008). Increased heating from GMCs early in the history of the disc can be invoked to overcome these discrepancies (e.g. Aumer, Binney & Schönrich (2016a, b), and a recent study of the vertical disc can also be invoked to overcome these discrepancies (e.g. Aumer, Binney & Schönrich (2016a, b).)

High-resolution cosmological zoom simulations have also revealed that other non-axisymmetric features such as bars may be a dominant heating agent, alongside (rare) perturbations from relatively massive ($M \gtrsim 10^{10} \text{M}_\odot$) satellite mergers (Grand et al. 2016). This is particularly prescient given that it has become apparent that the Milky Way likely underwent such a merger $\sim 10$ Gyr ago (e.g. Belokurov et al. 2018; Deason et al. 2018; Helmi et al. 2018; Krujssen et al. 2018; Mackereth et al. 2018a). The effect of this accretion event on the dynamics of the disc itself is yet to be fully considered.

Understanding the effect of different heating agents is of course instructive, but any constraints therein on galaxy formation are limited without consideration of the velocity dispersion that the stars formed with. It is clear, from cosmological simulations at least, that discs do not tend to form thin and cool, but instead follow stars formed with. It is clear, from cosmological simulations at least, that discs do not tend to form thin and cool, but instead follow the subsequent mapping of them through the galaxy (Adibekyan et al. 2017) will lead to the greatest insights into the evolutionary process of the Milky Way.

Whilst it has been informative to look at how the disc kinematics change with either age or abundances in isolation, it has become clear that stellar ages and element abundances in the solar neighbourhood (and beyond) do not simply correlate with one another, with perhaps the clearest example of this being [$\alpha$/Fe]–[Fe/H] as a function of age (e.g. Martig et al. 2016; Ness et al. 2016). The discovery of the separated high and low [$\alpha$/Fe] disc components in the solar vicinity (e.g. Fuhrmann 1998; Prochaska et al. 2000), the subsequent mapping of them through the galaxy (Adibekyan et al. 2012; Anders et al. 2014; Nidever et al. 2014; Hayden et al. 2015), and the characterization of their structure (Bovy, Rix & Hogg 2012a; Bovy et al. 2012b; Haywood et al. 2013; Bovy et al. 2016) has led to a new view of our Galaxy. In particular, it seems that the commonly assumed linkage between the high and low [$\alpha$/Fe] disc and the thick and thin disc is not as clear cut (Bovy et al. 2012a).

In fact, any structural dichotomy between the high and low [$\alpha$/Fe] discs is more apparent radially, with the high [$\alpha$/Fe] disc being centrally concentrated, and the low [$\alpha$/Fe] populations occupying donut-shaped annuli (Bovy et al. 2016), which change shape as a function of age and metallicity (Mackereth et al. 2017). Haywood et al. (2016) present similar arguments, proposing an inner/outer disc divide to be more constraining to the evolution of the disc.

Models for the formation of the bimodality in [$\alpha$/Fe] generally predict that high [$\alpha$/Fe] stars form in rapid and intense star formation, precluding enrichment by Type Ia SNe. These high star formation efficiency environments are either brought about by rapid infall of gas (e.g. Chiappini, Matteucci & Gratton 1997; Chiappini, Matteucci & Romano 2001; Chiappini 2009), or are present in the innermost, high-density regions of modelled discs (e.g. Schönrich & Binney 2009a, b). Alongside such analytic models, [$\alpha$/Fe] bimodality has also been recently realized in fully self-consistent cosmological simulations. Grand et al. (2018) found that [$\alpha$/Fe] bimodality in the discs of the AURIGA zoom-in simulations (Grand et al. 2017) arises due to a double-peaked star formation history, interjected by a shrinking of the gas disc. They also showed that [$\alpha$/Fe] bimodality can appear without this process, via an initial rapid burst of star formation in the inner galaxy, but noted that the bimodality was not ubiquitous among the simulations. Mackereth et al. (2018b) further showed that [$\alpha$/Fe] bimodality in the EAGLE simulations (Crain et al. 2015; Schaye et al. 2015) is very rare (occurring in $\sim 6$ per cent of Milky Way mass galaxies), and driven largely by the atypical assembly history of the haloes hosting the bimodal galaxies, which accreted mass faster at earlier times (and slower at late times) than their non-bimodal counterparts. Importantly, they showed that low and high [$\alpha$/Fe] populations form chemically separated from each other and usually overlap in age, becoming co-spatial later in the history of the galaxies.

These models for the formation of the high and low [$\alpha$/Fe] components make qualitative predictions for the kinematic structure of these populations, which are becoming more easily testable as the available data becomes more sophisticated. For example, the ‘radial migration’ model of Schönrich & Binney (2009a) predicts that outwardly migrating high [$\alpha$/Fe] stars, formed in the inner disc at higher $\sigma$, act to increase the slope of the AVR in the solar neighbourhood (relative to models where stars do not migrate). Therefore, isolating the low [$\alpha$/Fe] disc AVR should yield a reconciliation with the relation predicted by models of time-dependent scatterers (e.g. Lacey 1984; Hänninen & Flynn 2002). ‘Two-infall’-type models require a very rapid infall of gas onto the proto-Milky Way to form the high [$\alpha$/Fe] stars, which may correspond to the clumpy galaxies observed at high redshift (e.g. Noguchi 1998, and see also the recent preprint by Clarke et al. 2019), and rapid merging at similar times seen in simulations (e.g. Brook et al. 2004; Bournaud, Elmegreen & Martig 2009), consistent with the Milky Way-like galaxies from Mackereth et al. (2018b). Rapid, early accretion, and subsequent intense star formation would likely produce a centrally concentrated and very kinematically hot high [$\alpha$/Fe] population. Such models require that the high and low [$\alpha$/Fe] populations become co-spatial, and so require that the high [$\alpha$/Fe] populations are heated radially somehow, bringing stars to the solar radius. The distinction in kinematics between radial migration and heating may be slight, but would place a very strong constraint on these models. At present, the main distinction between radial migration and heating can be made by measuring the flaring of the disc (e.g. Bovy et al. 2012b), as it is expected that the flaring resultant from each process is different (e.g. Minchev et al. 2012). Understanding whether the flaring, the disc surface density, and
its kinematics are commensurate with one another would help to disentangle these processes.

Since the second Gaia data release (DR2 Gaia Collaboration 2018a), there has been a rapid development in our understanding of the kinematics and dynamics of the disc. Gaia Collaboration (2018b) mapped the kinematics of the largest portion of the Milky Way to date, revealing a rich structure and unveiling the clear non-axisymmetries of the disc, providing the first insight into our new view of the Galactic disc. Antoja et al. (2018) showed that Gaia DR2 data reveals a ’phase-spiral’ feature in the solar neighbourhood, which is now widely considered to be the lasting harmonic relic of a recent perturbation by a satellite flyby (e.g. Binney & Schönrich 2018; Bland-Hawthorn et al. 2018; Laporte et al. 2019). Such a wave-like oscillation is also apparent as an asymmetry in number counts as a function of height above the Galaxy mid-plane (Bennett & Bovy 2018). It is worth noting also that the effect of phase-mixing and warping on phase space in the disc was studied prior to Gaia by a number of works (e.g. Minchev et al. 2009; Quillen et al. 2009; Gómez et al. 2012; de la Vega et al. 2015). The general picture that emerges is that the disc is that is likely somewhat out of equilibrium due to recent events in its history.

As mentioned briefly above, a study of the vertical kinematics of RC giant stars in common with APOGEE and Gaia – DR2 recently suggested that gradual orbit scattering by small perturbations like those from GMCs may be enough to explain the observed trends between vertical actions and age – showing that a model of birth temperature scaling with star formation rate did not fit the data well (Ting & Rix 2018). That study focused mainly on the younger, low [$\alpha$/Fe]-dominated populations, whereas we extend the modelling of the vertical kinematics to the high [$\alpha$/Fe], old stellar populations, as well as examining the radial kinematics. Ting & Rix (2018) found that the slope of vertical action–age relationship (which can be understood as the adiabatically invariant counterpart to the AVR) increased with Galactocentric radius, but found this increase to also be in-line with expectations from GMC scattering.

In this paper, we present a dissection of the disc kinematics in age, [$\alpha$/Fe], and [$\alpha$/Fe], exploiting an unprecedented, multidimensional data set consisting of stars with element abundances and ages based on APOGEE spectra, whose distances are estimated to high precision using neural network modelling of these spectra trained on Gaia DR2 data, which also offers high-quality proper motion information. In Section 2, we present and describe this data set, and describe the procedure by which we estimate ages for a large portion of the APOGEE catalogue, using a Bayesian Convolutional Neural Network trained on APOGEE spectra and asteroseismic data from the APOKASC catalogue. We complete that section by describing the features seen in the data in the age–[$\alpha$/Fe]–[$\alpha$/Fe] plane. In Section 3, we present the procedure used to model the velocity dispersions in the vertical and radial direction in the Galactic disc for mono-age, mono-[Fe/H] populations in the low and high [$\alpha$/Fe] discs. Section 4 presents the main results from modelling the velocity dispersion of the mono-age, mono-[Fe/H] populations. In Section 5, we assess the results in the context of previous work on constraining the effects of disc heating, and discuss the implications of our findings on models for the origin of the high and low [$\alpha$/Fe] populations. We summarize and conclude the paper in Section 6.

2 DATA

We use a catalogue of stellar positions, velocities, element abundances, and estimated ages, comprising a cross-match of the catalogues from the 14th data release (DR14; Abolfathi et al. 2018) of the SDSS-IV APOGEE-2 survey, and the second data release (DR2; Gaia Collaboration 2018a) of the ESA–Gaia mission. Ages are estimated from a neural network based model (described next) trained on data from the APOKASC catalogue, which contains stars observed both spectroscopically by APOGEE and asteroseismically by the Kepler mission.

2.1 The APOGEE-2 DR14 catalogue

APOGEE (Majewski et al. 2017) is a spectroscopic survey of the Milky Way in the near-infrared $H$ band (1.5–1.7 $\mu$m), which has observed over 200 000 stars at high signal-to-noise ratio (SNR > 100 pixel$^{-1}$) and high resolution ($R \sim 22 500$), measuring over 15 element abundances. We use the DR14 data (Abolfathi et al. 2018), which consists of a combination of stars observed between the first and second iterations of the survey, APOGEE-1 (forming part of SDSS-III; Eisenstein et al. 2011) and 2 (in SDSS-IV; Blanton et al. 2017).

In this paper, we refer to these surveys collectively as APOGEE. All APOGEE data products used in this paper are those output by the standard data analysis pipeline, the APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP; García Pérez et al. 2016), which uses a pre-computed spectral library (Zamora et al. 2015), synthesized using a customised $H$-band linelist (Shetrone et al. 2015), to measure stellar parameters and element abundances. A full description and examination of the analysis pipeline is given in Holtzman et al. (2018). The individual element abundances are well tested against samples from the literature (Jönsson et al. 2018), and found to agree very well.

We refer here only to the abundances of the $\alpha$-elements included in DR14: Oxygen, Magnesium, Silicon, Sulphur, and Calcium, and the abundance of Iron. All abundances are those calculated by ASPCAP, and included in the APOGEE catalogue. We combine the $\alpha$ elements to attain the ratio of the mean $\alpha$ element abundance to that of Iron, [$\alpha$/Fe]. The element abundances are determined by ASPCAP via a two-step process. The best-fitting stellar parameters $T_{\text{eff}}, \log(g), v_{\text{turb}}$ (microturbulent velocity), [M/H], [$\alpha$/M], [C/M], and [N/M] are determined by a global fit to the grid of synthetic stellar spectra. Individual element abundances are then estimated by fitting windows in the spectrum to synthetic spectra with varying [M/H] (or [$\alpha$/M], [C/M] or [N/M], for $\alpha$-elements, Carbon and Nitrogen, respectively). For DR14, a small external calibration is applied to all the abundances, which forces the abundance ratios of solar metallicity stars in the solar vicinity to be equal to solar (Holzman et al. 2018).

We do not use Gaia parallaxes directly as distance indicators, but instead use spectro-photometric distances obtained by training a neural network to predict the luminosity of a star from its infrared spectrum using luminosities for a training set obtained from Gaia DR2. This procedure is described in detail in Leung & Bovy (2019a), but we re-iterate it here for clarity. The neural network has a similar architecture to the $\text{Apo} \text{gene} \text{BCNN}$ network used in Leung & Bovy (2019b): it is composed of two convolutional layers and two dense layers with rectifier activation (which introduce non-linearity in the hidden layers and prevent non-physical negative luminosity in the output); the network is optimized with the ADAM optimizer (Kingma & Ba 2014) and is implemented in $\text{astroNN}$ (Leung & Bovy 2019b), which itself relies on the TensorFlow framework (Abadi et al. 2016). To train the network in a way that allows us to take the Gaia parallax uncertainty into account, we follow Anderson et al. (2018) and define the output to be the following transformation
\[ L_{\text{fakemag}} = 10^{\frac{1}{2}M_K + 2}, \]
\[ = \sigma 10^{\frac{1}{2}K_{\odot}}, \]

where \( \sigma \) is the parallax in units of mas and \( K_{\odot} \) is the extinction-corrected apparent magnitude in the \( K \) band (for all stars, we obtain extinction corrections directly from the APOGEE catalogue). We then train the neural network to predict the ‘pseudo-luminosity’ \( L_{\text{fakemag}} \) from the continuum-normalized APOGEE spectrum using a \( \chi^2 \) objective function that takes the uncertainty in the \( L_{\text{fakemag}} \) of the training set (due to the uncertainty in \( \sigma \)), into account. When training the neural network, we adopt a 0.0562 mas \( Gaia \) DR2 parallax offset because this value leads to approximately unbiased luminosities at small \( \sigma \) (see Leung & Bovy 2019a, for more details on this procedure). We train the network using APOGEE DR14 (Holtzman et al. 2018) spectra that have SNR > 200, no APOGEE_FLAG flags set, radial velocity scatter smaller than 1 km s\(^{-1}\), and \( \sigma_{Gaia} / \sigma_{\sigma, Gaia} > 1 \).

Once trained, we use the network to predict luminosities \( L_{\text{fakemag}} \) for the entire APOGEE DR14 sample and, combined with extinction-corrected \( K_{\odot} \), we obtain spectro-photometric parallaxes using equation (2.1). We obtain uncertainties on the predicted \( L_{\text{fakemag}} \) using dropout variational inference as an approximation of a Bayesian neural network (Leung & Bovy 2019b) and these uncertainties are propagated to the spectro-photometric parallax. The distances are well tested in Leung & Bovy (2019a) against the APOGEE RC star catalogue (Bovy et al. 2014), alternative spectrophotometric distances from Santiago et al. (2016) and also those from (Hogg, Eilers & Rix 2018), and found to be in very good agreement, if not with improved precision. We also test the spectrophotometric distances against the recently published catalogue of distances of Schönrich, McMillan & Eyer (2019). We find that the distances agree to within ~10 per cent, compatible with the uncertainties quoted in the Leung & Bovy (2019a) catalogue.

### 2.2 Gaia DR2

\( Gaia \) DR2 contains astrometric parameters for over 1.3 billion sources, collected over 22 months from 2014 July. Many major improvements were made over the initial data processing, released in the DR1 and Tycho-\( Gaia \) Astrometric Solution (TGAS) catalogues, which are summarized in Gaia Collaboration (2018a). In particular, \( Gaia \) DR2 is the first data release that is not tied to any external catalogue to provide proper motion measurements (DR1 was tied to the HIPPARCOS and Tycho-2 catalogues), and uses its own reference frame based on quasars (\( Gaia \)-CRF-2; Gaia Collaboration 2018c). Other improvements included better modelling of the spacecraft attitude, and a significant improvement to the source detection algorithm that limits the number of spurious sources, and removes many duplicate sources that were included in the DR1 source catalogue (for this reason, the catalogues between data releases are treated separately). As well as removing spurious sources, these improvements have led to significant increases in the astrometric precision, affording typical residuals on the astrometric solution of ~0.2–0.3 mas in the middle of the magnitude range, decreasing to ~2 mas for the faintest sources.

We cross-match the full APOGEE DR14 catalogue with the \( Gaia \) DR2 source catalogue, searching in a cone of radius 0.5 arcsec around each source for its corresponding object in the \( Gaia \) catalogue using the CDS X-match service.\(^1\) We find that there are 254 789 objects in common between APOGEE and \( Gaia \) DR2. After removing stars from APOGEE that have warning or bad flags set, selecting only stars that have proper-motion measurements from \( Gaia \), and removing duplicate entries from the APOGEE catalogue, 83 189 stars with 6D phase space information remain. In the following analysis, we only use stars with 1.8 < log(g) < 3.0 [removing dwarf and subgiant stars, and upper red giant branch (RGB) stars] and those with [Fe/H] > −0.5, below which the APOKASC catalogue (used in the following section for training our age estimation model) has poor coverage. In addition to this, it was recently shown that extra mixing becomes significant at approximately this metallicity, which would substantially change the relationship between mass (and therefore the inferred ages) and [C/N] (Shetrone et al. 2019), which is likely to be the parameter that drives the relationship between the spectroscopic data and age. This final sample contains 65 719 red giant stars, whose spectra are measured with a minimum SNR of ~80, and a median SNR of ~170. The median uncertainty on the proper motion measurements from \( Gaia \) is ~0.5 mas, corresponding to a median uncertainty (after transformation) on the Galactocentric radial and vertical velocities of ~2 and ~1 km s\(^{-1}\), respectively. The median uncertainty on the distances from \( a_{\text{Gaia}} = 0.24 \) kpc, better than 10 per cent in most cases.

We transform the positions, proper motions, and radial velocities into the Galactocentric cylindrical frame, adopting the radial and vertical solar motion relative to the local standard of rest of Schönrich, Binney & Dehnen (2010) and the tangential motion of the sun relative to the Galactic centre of 245.6 km s\(^{-1}\) computed using the proper motion and distance to Sgr A* (Reid & Brunthaler 2004; Gravity Collaboration 2018). We propagate the observational uncertainties and their covariance matrix into this frame also. The data extend in Galactocentric radius roughly between 4 \( \lesssim R \lesssim 13 \) kpc, and between \(-1 \lesssim z \lesssim 2 \) kpc above and below the plane (here, and throughout the paper, we assume \( R_0 = 8 \) kpc and \( z_0 = 0.025 \) kpc). We show the distribution of the data in \( R \) and \( z \) in Fig. 1. This figure demonstrates the strong spatial selection biases that are present when using APOGEE data. For example, there is a large overdensity of stars above the plane, just inside the solar radius.

\(^1\)http://cdsxmatch.u-strasbg.fr/xmatch

---

Figure 1. The distribution of the APOGEE–\( Gaia \) sample used in this paper in Galactocentric \( R \) and \( z \). The pencil-beam selection of APOGEE is clearly apparent. The data cover a region of the Galaxy between 4 \( \lesssim R \lesssim 13 \) kpc and \(-1 \lesssim z \lesssim 2 \) kpc.
that corresponds to the heavily observed Kepler field. Correcting for these biases is possible, but not necessary for the kinematic modelling performed here, because neither the APOGEE nor Gaia data are kinematically biased. As we are interested only in modelling the velocity distributions as a function of position, these spatial biases do not affect our procedure.

### 2.3 Asteroseismic data and age estimates for APOGEE red giants

To make estimates of the ages of stars in the APOGEE–Gaia data described above, we use data from the APOKASC–2 catalogue (Pinsonneault et al. 2018) of asteroseismic data for stars in common between APOGEE and Kepler (Borucki et al. 2010). We use the APOKASC–2 data set to train a Bayesian convolutional neural network (BCNN) model, as implemented in the astroNN PYTHON package (Leung & Bovy 2019b), to predict stellar ages from the APOGEE spectra. Our technique for modelling ages using the BCNN is described fully in Appendix A. Briefly, we expect that the ability of the BCNN model to predict ages from the spectra to a good degree of accuracy stems from the presence of molecular bands of Carbon and Nitrogen in the APOGEE spectra. It is well documented in the literature that there is likely to be a relationship between these element abundances and the stellar mass (and therefore age; e.g. Salari et al. 2015), and this has been exploited in previous work to estimate ages from APOGEE abundances and spectra (Martig et al. 2016; Ness et al. 2016; Das & Sanders 2019). The BCNN uses the full information content of the APOGEE spectra to estimate the ages, and allows for the proper propagation of uncertainties into the analysis, to predict ages with good accuracy and with reasonable error estimation. Using the method, we generate a catalogue of ages with a median uncertainty between ~30 and 35 per cent across the full range of ages. We fully discuss the limitations of the predicted ages in Appendix A, but note here for clarity that we expect that ages above ~10 Gyr are likely to be underestimated, and subject to very large errors, such that it becomes difficult to distinguish, for example, an 11 Gyr old star from one at 13 Gyr old.

### 2.3.1 The age–[α/Fe]–[Fe/H] relation of disc stars

For completeness, in Figs 2 and 3 we demonstrate the complexity of the sample in age, [α/Fe], and [Fe/H] space. Fig. 2 shows the [α/Fe]–[Fe/H] plane as a 2D histogram, coloured by the mean age of the stars in each bin. The opacity of the bins reflects the density of stars in that region of [α/Fe]–[Fe/H] space. The solid and dashed lines indicate the ‘by-eye’ cut used to define high and low [α/Fe] populations in the disc. The stars inside the dashed region are excluded from the analysis to avoid contamination of the two populations. The high [α/Fe] population is generally old, whereas the low [α/Fe] population has a spread in age.
in the z-direction (as in Bovy et al. 2012c) such that

\[
\sigma_{(R,z)}(R, z) = \sigma_{(R,z)}(z, R) |a_{(R,z)}, b_{(R,z)}, \sigma_{(R,z)}(R_0, z_1/2)| \times \exp \left(\frac{-(R - R_0)}{h_{(R,z)}}\right) + \delta_{(R,z)},
\]

where \(a_{(R,z)}\) and \(b_{(R,z)}\) are, respectively, linear and quadratic coefficients in the quadratic function

\[
\sigma_{(R,z)}(z, R_0) = a_{(R,z)}(z - z_1/2)^2 + b_{(R,z)}(z - z_1/2)
\]

\[
+ \sigma_{(R,z)}(R_0, z_1/2),
\]

which is centered on the median z value, \(z_1/2\). Therefore, \(\sigma_{(R,z)}(R_0, z_1/2)\) gives the velocity dispersion at \(z_1/2\) and \(R_0\) (which here we define as the solar radius, assumed to be 8 kpc). The \(R\) dependence is presumed to be exponential, with a scale length \(h_{(R,z)}\). The addition of \(\delta_{(R,z)}\) accounts for the convolution of the model with the error on the velocity measurements. These assumptions on the \(R\) and \(z\) dependence of the velocity dispersion are not intended to be prescriptive of the actual underlying dependence, for which we have little prior expectation, but are intended to approximate it with smooth functions.

We assume that the tilt angle \(\alpha\) is a function of \(R\) and \(z\) such that

\[
\tan(\alpha) = \alpha_0 + \alpha_1 \frac{z}{R},
\]

which includes the cases of a velocity ellipsoid that is always aligned with the Galactic cylindrical coordinate system (\(\alpha_0 = \alpha_1 = 0\)) or the Galactic spherical coordinate system (\(\alpha_0 = 0\) and \(\alpha_1 = 1\)). The covariance between the radial and velocity dispersion is then

\[
\sigma_{(R,z)}^2(R, z) = \frac{\sigma_R^2(R, z) - \sigma_z^2(R, z)}{1 - \tan^2(\alpha)} + \delta_{(R,z)}^2.
\]

where we again convolve the model with the covariance between the radial and vertical velocity \(\delta_{(R,z)}^2\). This then allows the construction of the covariance matrix that describes the velocity ellipsoid in the \(R\)- and \(z\)-directions:

\[
\Delta = \begin{bmatrix}
\sigma_R^2(R, z) & \sigma_{R,z}^2(R, z) \\
\sigma_{R,z}^2(R, z) & \sigma_z^2(R, z)
\end{bmatrix}.
\]

The likelihood of the parameters \(O = [\sigma_{(R,z)}(R_0, z_1/2), a_{(R,z)}, b_{(R,z)}, h_{(R,z)}, \alpha_0, \alpha_1, v_{R,0}, v_{z,0}, \epsilon]^{-1} \) given the data for a given age-[Fe/H] bin can then be expressed as

\[
\ln \mathcal{L}(O|R, z, v_R, v_z) = \sum_i \ln \left(\frac{1 - \epsilon}{2\pi|\Delta|^{1/2}} \exp\left(-\frac{1}{2} \left|\epsilon_\mathcal{L}\right| \Delta^{-1} \epsilon_\mathcal{L}^t\right) + \epsilon p_{\text{backgr}}(z, R)\right).
\]

where \(p_{\text{backgr}}(R, z)\) describes the velocity distribution of an interploer model that is contributed by a fraction of stars \(\epsilon\). We use an interploer model that is also a normalized multivariate Gaussian, but with no covariance terms and with \(\sigma_R = \sigma_z = 100\) km s\(^{-1}\). This interploer model is also convolved with the observational uncertainty for each star. We find that for all mono-age, mono-[Fe/H] bins, \(\epsilon\) is less than a few per cent. The velocity distribution is a multivariate Gaussian centred on \(v_{(R,z),0}\), where \(v_{(R,z),0}\) are allowed to vary freely. We find that allowing this variation in \(v_{(R,z),0}\) provides a better fit to the data, and also reveals information regarding the non-axisymmetric velocity structure of the disc.
4 RESULTS

4.1 Velocity dispersion profiles

We first examine the velocity dispersion profiles for mono-age mono-[Fe/H] populations. The radial and vertical velocity dispersion profiles of a function of height above the mid-plane $|z|$ at $R_0$ for each mono-age mono-[Fe/H] population are shown in Fig. 4. We show the variation of $\sigma_z$ (left) and $\sigma_R$ (right) for each mono-age mono-[Fe/H] population between the 5th and 95th percentile of the $|z|$ distances for that bin. The median height $z_{1/2}$ is indicated for each population by the scatter points. Although these profiles were not constrained to be flat, and $\sigma_R$ and $\sigma_z$ could vary smoothly with $|z|$, we find that a generally flat profile is fit in almost all cases, within the uncertainties, meaning that for most bins, the parameters $a_{R,R}$, $\sigma_R$, and $b_{R,z}$ are fit to be very small, and with low uncertainty. This means that these populations are well approximated as isothermal discs. While this was also shown to be true in $\sigma_z$ for mono-abundance populations (MAPs) in SDSS/SEGUE by Bovy et al. (2012c), here, we extend that finding to show that mono-age, mono-[Fe/H] populations (for which MAPs are near analogues, aside from a few caveats; see Minchev et al. 2017) show very little variation in $\sigma_z$ and $\sigma_R$ as a function of $|z|$. It is worth noting at this point that Bovy et al. (2012c) used the isothermality of MAPs to demonstrate the precision of the SEGUE element abundances. While we do not rigorously perform a similar test here, our finding that most mono-age bins demonstrate isothermality over a large range in $|z|$ is likewise indicative that the precision in our age measurements is likely similar to the size of bins that we use (given the assumptions made in Bovy et al. 2012c).

An interesting exception to the isothermality is apparent in the younger populations shown in Fig. 4. Both $\sigma_z$ and $\sigma_R$ in these populations does appear to show some variation as a function of $|z|$, such that the velocity dispersion in both directions is slightly increased at higher $|z|$. This is unlikely to be due to uncertainties in the age measurement, as the ages are well constrained at the youngest ages. A slightly increasing $\sigma_z$ as a function of $|z|$ was also seen in the lowest [α/Fe] MAPs in Bovy et al. (2012c), which are likely roughly equivalent to these young populations. We return to this in Section 5.4.

Further to the generally isothermal nature of the mono-age, mono-[Fe/H] populations, we also find that the dependence of $\sigma_R$ and $\sigma_z$ on $R$ is generally very weak. The scale lengths of the assumed exponential dependencies are long, with median values across all age and [Fe/H] bins of $h_{R_R} = 15^{+1}_{-1}$ kpc and $h_{R_z} = 16^{+3}_{-3}$ kpc. However, the constraints on the scale lengths are relatively poor in comparison to the stronger constraints placed on the other parameters of the model. This is likely due to the fact that in any given age–[Fe/H] bin, the radial extent of the data is relatively small. Regardless of the poor constraints on the scale length of the $R$ dependence, we note that although we allowed the data to be fit with either increasing or decreasing $\sigma$ as a function of $R$, we find that it is best fit by a slowly decreasing profile in all cases.

Finally, we find that the data are fit in most of the age–[Fe/H] bins by models where the tilt angle of the velocity ellipsoid $\alpha$ is aligned between the Galactocentric spherical and cylindrical coordinate system (i.e. we find that $\alpha_1$ lies between 0 and 1, with $\alpha_1 = 0$ corresponding to cylindrical and $\alpha_1 = 1$ corresponding to spherical), with a large uncertainty, such that most bins are consistent with $\alpha_1 = 0$. We compared our fits to the work of Bödenbender, van de Ven & Watkins (2015), who found that stars between 0.5 $\lesssim |z| \lesssim 2.0$ kpc were well fit by a relationship close to a spherical alignment, finding that $\alpha = 0.9 \arctan(|z|/R_\odot) = 0.01$. On directly comparing our fits with their results (plotting $\alpha(z_{1/2})$ against $z_{1/2}$), we find that the mono-age, mono-[Fe/H] populations at $z_{1/2} \gtrsim 0.5$ kpc agree well with Bödenbender et al. (2015), albeit with very large uncertainties. Younger bins, at lower $z_{1/2}$, also appear to lie on the Bödenbender et al. (2015) relation. Binney et al. (2014) found similarly consistent results for the tilt angle in the RAVE data.

4.2 Age–velocity dispersion relations in the high and low [$\alpha$/Fe] discs

We now turn to examining the trends between $\sigma_R$ and $\sigma_z$ with age. Quoted and displayed values of $\sigma_{R,z}$ in the following are those evaluated at $R_0$ and the median height $z_{1/2}$ (age, [Fe/H]) which, due to the weak dependence of $\sigma_{R,z}$ with $z$, essentially corresponds to $\sigma_{R,z}(R_0)$. Fig. 5 shows $\sigma_z$ (left) and $\sigma_R$ (right) against age for the mono-age, mono-[Fe/H] populations in the low and high [$\alpha$/Fe] populations. At each age, we show all mono-[Fe/H] bins in which there were more than 200 stars. We colour the points by the median of the mean orbital radii of the stars in each bin ($R_{mean}$), calculated
Figure 4. The vertical (left) and radial (right) velocity dispersion of mono-age, mono-[Fe/H] populations fit to the low (the solid lines) and high (the dashed lines) $[\alpha/Fe]$ sub-samples as a function of height above the midplane. The colour of each profile gives the age of the population, for which the profiles are only shown at $|z|$ between where 25 and 75 percent of the observed population are. The median $|z|$ of each population is indicated by the point on each line, and the coloured bands show the $1\sigma$ uncertainties at each $|z|$. In general, the profiles are flat as a function of $|z|$, indicating that mono-age populations are roughly isothermal, except for the youngest, coldest populations. The radial and vertical velocity dispersion increases as a function of age.

Figure 5. Velocity dispersions $\sigma_R$ (right) and $\sigma_z$ (left) at $R_0$ and the median height $z_{\frac{1}{2}}$(age, [Fe/H]) as a function of age in mono-age, mono-[Fe/H] bins. At each age, we display all the mono-[Fe/H] bins that have more than 200 stars and we apply a small random Gaussian jitter in age with a scale of 0.2 Gyr to the points, to make the variation (or lack thereof) in $\sigma_{[R,z]}$ with [Fe/H] at fixed age more clear. The points are coloured by the median of the mean orbital radius $\langle R_{\text{mean}} \rangle$ in each bin, to demonstrate the Galactocentric radius at which the stars reside. The dispersions $\sigma_R$ and $\sigma_z$ increase with age. At ages $<6$ Gyr, there is a large spread in $\langle R_{\text{mean}} \rangle$ at fixed age, where the mono-age populations with greater $\langle R_{\text{mean}} \rangle$ also have higher vertical velocity dispersion $\sigma_z$, while the $\sigma_R$ is roughly the same.

\[ R_{\text{mean}} = \frac{(r_{\text{peri}} + r_{\text{ap}})}{2} \]  

(9)

where $r_{\text{peri}}$ and $r_{\text{ap}}$ are the pericentre and apocentre radius of the orbits, which are determined using the orbital parameter estimation method of Mackereth & Bovy (2018), as implemented in the galpy PYTHON package for galactic dynamics (Bovy 2015). The mono-age, mono-[Fe/H] populations in the high $[\alpha/Fe]$ disc are displayed as the open points, whereas the low $[\alpha/Fe]$ populations are the solid points.

The velocity dispersions $\sigma_z$ and $\sigma_R$ both increase roughly monotonically as a function of age, with $\sigma_z$ going from $\sim 10$ to $\sim 40$ km s$^{-1}$ and $\sigma_R$ increasing from $\sim 30$ to a maximum of
~70 km s$^{-1}$ in the oldest, high [α/Fe] bins. The spread in $\sigma_z$ at fixed age is greater than that in $\sigma_R$ for most age bins, but these become similar at the oldest ages. For the younger, low [α/Fe] populations, with age < 6 Gyr, the bins with higher $\sigma_z$ correspond to those with the largest $\langle R_{\text{mean}} \rangle$, indicating that these stars are typically orbiting outside the solar circle. This indicates that $\sigma_z$ increases with R at all ages. Given that this is a somewhat unexpected finding, we briefly discuss its implications in Section 5.

The increasing $\sigma_z$ versus $\langle R_{\text{mean}} \rangle$ trend does not appear in the high [α/Fe] populations, although the spread in $\sigma_z$ is similar. For low [α/Fe] bins, the spread in $\sigma_R$ at fixed age is very small, such that the bins with high $\langle R_{\text{mean}} \rangle$ have an excess of $\sigma_z$, as discussed further below. The high [α/Fe] populations have lower $\langle R_{\text{mean}} \rangle$ than the low [α/Fe], but have the higher $\sigma_R$ and $\sigma_z$. Importantly, the considerable spread in $\sigma_R$ at fixed age in the high [α/Fe] populations (and lack thereof for low [α/Fe] bins) is very clear.

We approximate the shape of the radial and vertical A VRs by fitting simple power-law relationships $\sigma_{R,z}(\text{age}) \propto \text{age}^{\beta_{R,z}}$ (similar to those fit by; e.g. Aumer et al. 2016b; Ting & Rix 2018) to the run of $\sigma_R$ and $\sigma_R$ as a function of age in each [Fe/H] bin. We only fit the relationship in [Fe/H] bins where there are three or more age bins with more than 200 stars, and use a simple maximum likelihood procedure to estimate $\beta_R$ and $\beta_z$. The resulting $\beta$ for the vertical and radial A VRs as a function of $\langle R_{\text{mean}} \rangle$ is shown for the low [α/Fe] mono-[Fe/H] bins in Fig. 6. The left-hand panel shows that low [α/Fe] populations have a similarly shaped vertical A VR as a function of [Fe/H] (and therefore roughly constant with $\langle R_{\text{mean}} \rangle$ also), such that the mean $\beta_z = 0.5 \pm 0.1$ when these populations are combined. We also fit the high [α/Fe] populations (but do not show these in Fig. 6). For those that were fit, we find that although less well constrained, these populations have seemingly flat vertical A VRs, with $\beta_z = 0.02^{+0.12}_{-0.05}$. We strongly emphasize, however, that age uncertainties are much larger for high [α/Fe] stars. Therefore, contamination between these age bins is likely and this contamination would flatten any intrinsic A VR. Furthermore, it is likely that the age spread in the high [α/Fe] stars is artificially enhanced by these uncertainties, and so the A VR for these populations has little significance.

While $\beta_z$ is roughly constant as a function of [Fe/H] (and therefore $\langle R_{\text{mean}} \rangle$ in the low and high [α/Fe] discs, the shapes of the radial A VRs for the low [α/Fe] populations change significantly. We find that the high [α/Fe] population radial A VRs are consistent with one another, and again with being nearly flat, having $\beta_R = 0.2 \pm 0.2$. However, for the low [α/Fe] disc populations we find that the radial A VRs do appear to change shape with [Fe/H]. The values of $\beta_R$ against $\langle R_{\text{mean}} \rangle$ (coloured by [Fe/H]) fit for the low [α/Fe] disc are shown in the right-hand panel of Fig. 6. It is clear from this figure that the outer disc populations (those with the lowest [Fe/H]) have significantly flatter radial A VRs than those populations that reside closer to the solar radius. The index $\beta_R$ appears to increase to a maximum $\sim 0.35$ at $\langle R_{\text{mean}} \rangle \sim 8$ kpc, and then decline slightly again at $\langle R_{\text{mean}} \rangle$ inside the solar radius. However, within the uncertainties, all the mono-[Fe/H] populations with $\langle R_{\text{mean}} \rangle \lesssim 9$ kpc are roughly consistent with the same $\beta_R$. We discuss the implication of this finding in Section 5.

4.3 The shape of the velocity ellipsoid as a function of age

To examine the kinematic structure of the disc as a function of age, in Fig. 7 we display the axis ratio $\sigma_z/\sigma_R$ as a function of age for the mono-age, mono-[Fe/H] populations. As in Fig. 5, the points are coloured by $\langle R_{\text{mean}} \rangle$, and high [α/Fe] populations are plotted as the open points. In the two oldest age bins (≥6 Gyr), where the high [α/Fe] stars reside, the relation between $\sigma_z/\sigma_R$ and age is roughly flat, and has very little scatter. Combining all the MCMC samples for these bins, we find $\sigma_z/\sigma_R = 0.64 \pm 0.04$. This value is in rough agreement with the measurements of $\sigma_R/\sigma_z$ in LAMOST-TGAS stars by Yu & Liu (2018).

For low [α/Fe] populations, the relation between age and $\sigma_z/\sigma_R$ is more complex. The apparent excess of $\sigma_z$ seen in Fig. 5 is very clear: bins at the same age with different $\langle R_{\text{mean}} \rangle$ have different $\sigma_z/\sigma_R$, such that the mono-age, mono-[Fe/H] populations residing outside the solar radius have $\sigma_z/\sigma_R \gtrsim 0.7$, with many bins having higher $\sigma_z/\sigma_R$ than the oldest, high [α/Fe] bins. At fixed age in the low [α/Fe] populations, greater values of $\sigma_z/\sigma_R$ correspond to larger $\langle R_{\text{mean}} \rangle$. This trend is not apparent in the high [α/Fe] disc, where all populations have a similar $\langle R_{\text{mean}} \rangle$, and correspondingly similar $\sigma_z/\sigma_R$ (as discussed above). That outer disc populations are those with the highest $\sigma_z/\sigma_R$ due to an excess of $\sigma_z$ may seem unsurprising at first, given that it is well known that at young ages (and low [α/Fe]), the disc is strongly flared (Bovy et al. 2016; Mackereth et al. 2017) and that the disc has a significant warp in its outer regions (e.g. Poggio et al. 2018). However, we discuss the possible effects of this warping and other disc heating agents further in Section 5. The more novel result here is that there is an age (between ~6 and 8 Gyr ago) at which mono-age, mono-[Fe/H] populations exhibit a significant change in their velocity structure, which coincides with the age that roughly divides the high and low [α/Fe] disc populations. This may have important implications for the history of formation and assembly of the Galaxy, which we will also discuss in Section 5.

4.4 The mean velocity–$\langle R_{\text{mean}} \rangle$ relationship of populations as a function of age

By allowing the mean velocities of the model velocity distributions, $v_{R,z}$ and $v_{\beta,z}$, to be free parameters in the fits, we can assess the degree by which the velocity distribution is subject to non-axisymmetries in the disc. We find that the general results pertaining to $\sigma_R$ and $\sigma_z$ are robust to holding the mean velocities fixed at 0 km s$^{-1}$, and that the majority of the mono-age, mono-[Fe/H] populations are well fit by models with $v_{R,0} = v_{z,0} = 0$. However, we find that some mono-age, mono-[Fe/H] populations are better fit by models that have a significant departures from $v_{R(z),0} = 0$.

In Fig. 8, we show that there are clear trends between the $v_{R,0}$ and $v_{z,0}$ of mono-age, mono-[Fe/H] populations, and the typical orbital radius ($\langle R_{\text{mean}} \rangle$) of the populations: young (≤4 Gyr), low [α/Fe] populations have a small but significant increasing trend of their mean vertical velocity with $\langle R_{\text{mean}} \rangle$, increasing to ∼5 km s$^{-1}$ for the outermost populations. These populations retain the same high $\sigma_z/\sigma_R$ values (within the uncertainties) as when fit with the fixed mean velocity model. This trend of $v_{z,0}$ with $\langle R_{\text{mean}} \rangle$ in young low [α/Fe] stars is roughly consistent with the position and magnitude of the warp seen in TGAS by Schönrich & Dehnen (2018) and in Gaia DR2 by Poggio et al. (2018; their fig 3). The scale of the warping found in our best-fitting models is also roughly consistent with that found in the 2MASS data (Momany et al. 2006).

We find a different trend of the mean radial velocity with $\langle R_{\text{mean}} \rangle$ for the young, low [α/Fe] populations, such that $v_{R,0}$ has a wave-like pattern with increasing $\langle R_{\text{mean}} \rangle$, dipping down to ∼3 km s$^{-1}$ at $\langle R_{\text{mean}} \rangle ~ 8.5$, before rising again to ∼5 km s$^{-1}$ at the
Figure 6. The power-law index of the age–velocity dispersion relations $\beta_R$ and $\beta_z$ obtained by fitting a power-law behaviour to the $\sigma_R$ and $\sigma_z$ trends as a function of age for each mono-[Fe/H] bin that has $>3$ mono-age bins with $N > 200$ stars in the low [$\alpha$/Fe] disc. The $x$-axis is the median of the mean orbital radii in each of the mono-[Fe/H] bins and the points are coloured by the [Fe/H] of the bin. The outer disc, low [Fe/H] stars have flatter radial AVRs than the stellar populations in the inner disc (with higher [Fe/H]). The vertical AVRs do not change shape significantly as a function of $\langle R_{\text{mean}} \rangle$.

Figure 7. The ratio of vertical to radial velocity dispersion $\sigma_z/\sigma_R$ in mono-age, mono-[Fe/H] populations as a function of age. The same jitter as in Fig. 5 is applied to each point and the points are again coloured by the median of the mean orbital radii of the stars in each bin, $\langle r_{\text{mean}} \rangle$. Bins older than $\sim6$ Gyr have $\sigma_z/\sigma_R \sim 0.6$. At ages $<6$ Gyr, there appear to be two separated tracks in age-$\sigma_z/\sigma_R$, where bins with stars mainly in the outer disc have $\sigma_z/\sigma_R > 0.6$, and bins with stars nearer to the solar radius have lower values. At younger ages, there seems to be a positive correlation between $\sigma_z/\sigma_R$ and age, which is roughly the same for both tracks.

5 DISCUSSION

We have shown that when regions beyond the solar vicinity are considered, the relationship between the age and the different components of the velocity dispersion of stars in the disc is complex, and co-dependent on the stars’ mean orbital positions and their element abundances. In the following section, we contextualize our results within the existing work on the velocity structure of the disc and discuss the constraints that these results place on models for the formation and evolution of the Galactic disc.

5.1 The increasing trend of $\sigma_z$ with $R_{\text{mean}}$ in the low [$\alpha$/Fe] disc

Given that our finding that $\sigma_z$ shows a clear increase with $R_{\text{mean}}$ at all ages is somewhat surprising, as the expectation is that discs are most likely to show a decreasing trend, we briefly discuss here some possible explanations for this finding, and present some comparison to other observational results.

For a disc in equilibrium,

$$\Sigma(R) \propto \frac{\sigma_z^2}{h_z},$$

where $\Sigma(R)$ is the disc surface density, and $h_z$ is the disc scale height. It follows that if $h_z$ is not a function of $R$, then $\sigma_z$ should trace $\Sigma(R)$ and therefore be a declining function of $R$. This relationship between $\sigma_z$ and $\Sigma(R)$ is commonly used to measure the surface density of external galaxy discs, making an assumption of a radially constant $h_z$, which is difficult to measure in face-on galaxies where $\sigma_z$ is more readily measurable. It is commonly seen in these studies that $\sigma_z$ declines with $R$ (see e.g. Martinsson et al. 2013; Aniyan et al. 2018). The assumption of constant $h_z(R)$ is usually based on observations of edge-on spiral galaxies with small bulges (e.g.
5.2 Heating agents and the velocity dispersion history throughout the disc

Our understanding of heating in the Milky Way disc has thus far largely been informed by observational constraints that were confined relatively near to the Sun. Our data set extends well beyond the solar vicinity, out to \( R \gtrsim 12 \text{kpc} \) in the outer disc and into the inner disc down to \( R \lesssim 5 \text{kpc} \). This allows us to build a picture of heating and its connection to Galactic evolution over the extent of the disc. While a more robust understanding of these results will likely only be afforded by detailed numerical simulations, we discuss the implications of our findings using the extensive body of existing work on this topic. We attempt to understand trends in our results from the context of the more classical results on this topic, which, although outdated in some regards by numerical simulations, provide some key insights into the processes at hand.

5.2.1 Velocity ellipsoid shape

Analytical formulations of the (vertical) heating due to GMCs by Lacey (1984) showed that the velocity dispersion should increase as \( \sigma_z \propto R^{0.25} \), and that the axis ratio \( \sigma_z/\sigma_R \) should approach a value of \( \sim 0.8 \). Later work, using more sophisticated calculations, found that the axis ratio resulting from GMC heating is more likely somewhere between \( \sim 0.5 \) and 0.6, but found a similar power-law index of \( \sim 0.3 \) (Ida, Kokubo & Makino 1993; Hänninen & Flynn 2002). We find that all high \([\alpha/Fe]\) populations are consistent with having \( \sigma_z/\sigma_R = 0.64 \pm 0.04 \), in rough agreement with this revised value for GMC heating (we return to the AVR later in Section 5.2.2). In addition to this, it is interesting to note that all high \([\alpha/Fe]\) populations have similar, low \( \langle R_{\text{mean}} \rangle \), but large...
variations in velocity dispersion with [Fe/H] in a fixed age bin. That these populations were presumably formed at similar times and have different radial and vertical velocity dispersions but a similar ratio of these may indicate that these stars in fact formed in a turbulent interstellar medium (ISM; external galaxy observations between redshift 0.7 and 2.3 indicate that the ISM is indeed turbulent at these early times; e.g. Wisnioski et al. 2015), and have simply retained their kinematics. Understanding this will, however, require further modelling of the formation of these stellar populations in self-consistent cosmological simulations.

In the low [α/Fe] disc, the ratio $\sigma_z/\sigma_R$ varies strongly as a function of [Fe/H] (and therefore $R_{\text{mean}}$) and age, with many of the low [Fe/H] (outer disc) populations having $\sigma_z/\sigma_R$ exceeding that predicted by GMC models, and the higher [Fe/H] (inner disc) populations falling well below the predicted value. The fact that each [Fe/H] bin in the low [α/Fe] disc also shows an increasing trend between $\sigma_z/\sigma_R$ and age also indicates that the kinematics of this population cannot be explained by GMC heating alone because models predict a flat $\sigma_z/\sigma_R$ with age if a disc is heated by a single stationary process (this is seen, for example, in the work of Jenkins & Binney 1990). Predictions of $\sigma_z/\sigma_R$-age relations from more recent numerical simulations that include a changing mass fraction of GMCs overtime show clearly similar increasing trends to those seen in our results (Aumer et al. 2016b). In those simulations, GMC scattering is important for the early disc, and slowly gives way to spiral arm heating, as the disc grows in mass. Our result that $\sigma_z/\sigma_R$ increases with age in the low [α/Fe] disc is clearly in-line with this picture of heating by non-stationary processes. A similar scenario was proposed also by Yu & Liu (2018), whose results on the velocity ellipsoid shape are consistent with those presented here.

Jenkins & Binney (1990) showed that increasing the importance of spirals over GMCs reduced $\sigma_z/\sigma_R$, such that models that have the most dominant spiral arm perturbations have the lowest $\sigma_z/\sigma_R$. This is also borne out in numerical simulations (e.g. Aumer et al. 2016b), where spirals heat generally only in the plane (increasing $\sigma_R$), whereas GMCs heat both in-plane and vertically. By this logic, $\sigma_z/\sigma_R$ becomes a good measure of the relative importance of these processes in shaping the disc, but it is important to bear in mind also the external effect of satellite interactions that should also perturb and heat the disc (such interactions are not included, for example, in Aumer et al. 2016b). We find that in the low [α/Fe] disc, the outermost disc populations (at low [Fe/H]) have the highest $\sigma_z/\sigma_R$ at fixed age. This is consistent with the picture of declining importance of spiral arm heating with Galactocentric radius. However, an important finding here is that the highest values of $\sigma_z/\sigma_R$ in the outer disc far exceed that expected from GMC or spiral heating. The resulting $\sigma_z/\sigma_R$ for the low [α/Fe] populations in the outer disc also exceeds that found in many of the idealised simulations of Aumer et al. (2016b). It is logical to assume that the number of GMCs in the outer disc is less than that in the inner disc (indeed, this is readily observed in the Milky Way; e.g. Miville-Deschênes, Murray & Lee 2017), and likely even less at earlier times, while the disc was still growing. Therefore, it seems unlikely that GMCs or spiral arms could have heated these outer disc populations to their very high velocity dispersion. It is notable that this increasing vertical velocity dispersion with R is also realized in the results of Sanders & Das (2018; their fig. 9). Fully understanding this excess of vertical velocity dispersion will almost certainly require additional modelling of the spatial density structure of the disc, which can then be compared with the kinematic modelling performed here.

5.2.2 The age–velocity dispersion relation

It is also notable that we find that $\sigma_z \propto \tau^{0.5}$ for the low [α/Fe] disc populations, in general agreement with previous observational work (e.g. Wielen 1977; Seabroke & Gilmore 2007; Sobirian et al. 2008) and in disagreement the classical theoretical value for GMC heating. This discrepancy has been widely noted in the literature, and was cast by Aumer et al. (2016b) as arising from the discrepancy between AVRs and true heating histories, e.g. the evolution of velocity dispersion in co-evol populations that generates the AVR. They find that heating histories (described by a power-law index $\beta_R$), which are mainly set in the few Gyr after stellar birth, tend to depend on the birth time of stars itself, such that stars born at different times in the Galaxy’s history attain different kinematics due to the changing nature of perturbers with time. This leads to a discrepancy in the AVR from simple analytical values, which assume heating to be a stationary process.

Considering radial heating, the simulations of Aumer et al. (2016b) predict that $\beta_R$ should vary with Galactocentric radius, and also predict that the heating histories of younger populations should have greater $\beta_R$, likely due to the fact that in their simulations the GMC heating gives over to spiral arm perturbations (this was also predicted by Jenkins & Binney 1990, although they failed to predict correct indices, for a variety of reasons). We find that $\beta_R$ shows a clear trend with $R_{\text{mean}}$ in the low [α/Fe] disc, such that for the outer disc stars $\beta_R \sim 0.15$. As $(R_{\text{mean}})$ decreases to $\lesssim 10$ kpc, $\beta_R$ rapidly increases to values between ~0.3 and 0.4. The decreased $\beta_R$ in the outer disc stellar populations may be a manifestation of the waning influence of spiral arm heating on the kinematics of low [α/Fe] stellar populations that reside in the outermost part of the disc. If spiral arm heating is acting strongly inside $\sim 10$ kpc, peaking at 8 $< (R_{\text{mean}}) < 9$ kpc, then this may also suggest that radial migration of stars is most efficient in this region of the disc. We also find, at relatively low significance, that $\beta_R$ increases slightly between the innermost populations and those at ~8 kpc.

5.3 Warping of the disc as a heating agent

In the absence of extensive GMC or spiral arm heating, there are few remaining possible causes for the large velocity dispersions that we observe in the outer disc. One interesting possibility is heating from warping of the Galactic disc, caused by satellite interactions (e.g. Kim et al. 2014), bending instabilities (e.g. Khoperskov & Bertin 2017), or misalignment of the angular momenta of the disc and dark halo (e.g. Debattista & Sellwood 1999; Jiang & Binney 1999), to name a few proposed processes. Such a warping of the Milky Way disc is well documented (e.g. Momany et al. 2006; Schönrich & Dehnen 2018) and has been recently confirmed in the Gaia DR2 data (Poggio et al. 2018). Modelling of such bending modes has suggested that this process may act to increase $\sigma_z$ (as shown in the early work of Hunter & Toomre 1969) therefore increasing $\sigma_z/\sigma_R$ (e.g. Martig, Minchev & Flynn 2014). Khoperskov & Bertin (2017) suggested heating by these processes can cause $\sigma_z/\sigma_R$ to approach values as high as unity. Therefore, it is at least possible that these stars have been heated by warping or bending of the disc.

That the amplitude of warping in the low [α/Fe] disc is stronger at low ages may suggest that the warp is likely stronger in the star-forming gas disc, causing more recently formed stars to retain the warp kinematics more readily. This is somewhat consistent with the finding of Mackereth et al. (2017) that young stars in the low [α/Fe] disc are those that have a stronger flare in the outer parts of the disc.
The warping present in the kinematics combined with the flattened radial AVR and excess $\sigma_z/\sigma_R$ of the outer disc populations, strongly suggests that the kinematics of these populations is affected by the warping of the Galactic disc.

5.4 Non-isothermality in young populations

We saw in Section 4.1 that the vertical dependence of the radial and vertical kinematics departs from isothermality for the youngest populations in the disc. This manifests itself in the $\sigma_R$ profiles as a function of $|z|$ as an increasing trend with increased height above the mid-plane. This feature was first hinted at low significance in the dependence of $\sigma_z$ on $z$ for low [$\alpha$/Fe] populations by Bovy et al. (2012c), but is now well established from our analysis in this paper. We now investigate this further by also studying the shape of the velocity ellipsoid $\sigma_z/\sigma_R$ (as shown in Fig. 7) as a function of $|z|$. These profiles for mono-age, mono-[Fe/H] populations in the low and high [$\alpha$/Fe] disc are shown in Fig. 9.

It is immediately clear from Fig. 9 that the younger, low [$\alpha$/Fe] $\sigma_z/\sigma_R$ profiles increase with $|z|$, whereas the high [$\alpha$/Fe], old population profiles are flat over a large range of $|z|$. Following from the discussion of the work of, e.g. Aumer et al. (2016b) in Section 5.2, a decreased $\sigma_z/\sigma_R$ is likely a signature of stronger heating by spiral structure. This presents a compelling explanation for the departure from isothermality of these young populations, and may also indicate the differing time-scale of these heating agents. For example, if spirals can radially heat the stellar population quickly, this would reduce the $\sigma_z/\sigma_R$ in the plane for young populations. As GMCs likely redirect this radial heating into vertical heat more slowly than the spirals heat the disc, then one might expect also a trend in $\sigma_z$ with $|z|$ for the younger populations, as the stars that have been vertically heated are at this point more likely to be above the plane. For older ages, this trend becomes eroded as the populations approach isothermality due to prolonged exposure to this combination of heating agents. It is, however, apparent in Fig. 9 that some of the older low [$\alpha$/Fe] populations still show an increasing $\sigma_z/\sigma_R$ profile. As these populations are those residing in the outermost regions of the disc, we propose that this may be due again to the waning effect of the spiral structure, reducing the flattening at old ages of these profiles. Such trends are not seen at all in the high [$\alpha$/Fe] populations further suggests a distinct evolutionary history for these populations, which we will discuss further in the following section.

5.5 Kinematic distinction of the high and low [$\alpha$/Fe] discs

Our results demonstrate that there are many kinematic differences between the high and low [$\alpha$/Fe] discs. Not only do these populations differ in their velocity dispersions $\sigma_R$ and $\sigma_z$, but they appear to have an entirely different kinematic structure, displaying different trends in their kinematics with age, [Fe/H], and position in the Galaxy.

It is well known from previous work that the high [$\alpha$/Fe] disc populations are old, have the lowest $\langle R_{\text{mean}} \rangle$, and the highest velocity dispersions. In this paper, we have determined that the high [$\alpha$/Fe] disc populations are furthermore characterized by near-flat radial and vertical AVRs and that they are consistent with a single $\sigma_R/\sigma_K = 0.64 \pm 0.04$, which does not vary with age, [Fe/H], or $|z|$. These trends are markedly different from those displayed by the younger, low [$\alpha$/Fe] populations that show complex trends in $\sigma_R/\sigma_K$ with age and [Fe/H], and have increasing values of this ratio with $|z|$. High [$\alpha$/Fe] stellar populations appear to be more centrally concentrated, with smaller scale lengths and larger scale height than their low [$\alpha$/Fe] counterparts (Cheng et al. 2012; Bovy et al. 2012b, 2016; Mackereth et al. 2017), and have the smallest $\langle R_{\text{mean}} \rangle$ of all the mono-age mono-[Fe/H] populations. A centrally concentrated and kinematically distinct high [$\alpha$/Fe] component is a key prediction in models that invoke an early, rapid formation of the high [$\alpha$/Fe] population, induced by rapid gas inflow or high frequency gas-rich mergers (Chiappini et al. 1997, 2001; Brook et al. 2004; Mackereth et al. 2018). The slight discontinuity in age versus velocity dispersion between the high and low [$\alpha$/Fe] discs (seen in Fig. 5) is also suggestive of a different, more abrupt origin of the high [$\alpha$/Fe] populations than that predicted by, e.g. Schönrich & Binney (2009a,b). It was also predicted by Aumer et al. (2016a) that thick disc components do not arise during quiescent disc growth.

The low [$\alpha$/Fe] populations are extended in $R$, and have a rich kinematic structure in age and [Fe/H], consistent with a gradual formation over a longer time-scale, and a subsequent evolution that was likely affected by spiral structure, GMCs, and satellite interactions in ways that strongly depend on position in the Galaxy. As discussed in Section 5.2, we find that the low [$\alpha$/Fe] populations have vertical AVRs that agree well with those measured previously, where $\sigma_z \propto z^{0.25}$, and disagree with theoretical predictions that $\sigma_z \propto z^{0.75}$ for GMC and spiral heating. The Schönrich & Binney (2009a,b) model suggests that the increased slope of the AVR may be due to biases from high $\sigma_z$, old, high [$\alpha$/Fe] stars that have migrated to the solar radius. This bias should therefore be removed when fitting the AVR of only low [$\alpha$/Fe] stars. We do not find that the AVR slope is decreased when fitting these populations separately. As pointed out in Ting & Rix (2018), the AVR may be steepened by invoking a larger number of scatterers (GMCs) in the past. This, contrasted with the relatively flat vertical AVR of the high [$\alpha$/Fe] populations, further suggests that these populations had very different origins and subsequent evolutionary histories.

The emergent picture then from the kinematics of the low and high [$\alpha$/Fe] disc components is that they are quite different, and therefore likely formed and evolved differently and certainly over very different time-scales.

6 SUMMARY AND CONCLUSIONS

In this paper, we have performed a detailed dissection of the Milky Way’s low and high [$\alpha$/Fe] disc kinematics as a function of [Fe/H] and age. We derive a new set of ages for APOGEE DR14 red giant branch stars by applying a BCNN model, trained on stars with APOGEE spectra and asteroisomeric ages from the APOKASC catalogue. The improved age precision, alongside the exquisite positions, and kinematics with which to place constraints on models for the formation and evolution of the Galaxy. Here, we have focused on determining the kinematic structure of the disc in age–[Fe/H]–[\alpha/Fe] space. The large cross-over with the Gaia DR2 catalogue affords us an unparalleled catalogue of stars with measured ages, abundances, positions, and kinematics with which to place constraints on models for the formation and evolution of the Galaxy. Therefore, we have focused on determining the kinematic structure of the disc in age–[Fe/H]–[\alpha/Fe] space. Our main findings can be summarized as follows:

(i) The radial and vertical velocity dispersions of mono-age, mono-[Fe/H] populations in the high and low [$\alpha$/Fe] discs generally show little variation with $|z|$ – i.e. most are close to isothermal – and have slowly decreasing exponential dependencies on $R$. The youngest mono-age populations do, however, display a departure from isothermality as they slightly increase in $\sigma_R$ and $\sigma_z$ with $|z|$ (see Figs 4 and 9).

The low {$\sigma_z/\sigma_R$} uncertainties on age in this regime.

Our main findings place novel constraints on the history of dynamical heating across the disc. The finding that the velocity ellipsoid shape is strongly dependent on Galactocentric radius for low [Fe/H] populations, and that the AVR becomes flatter at greater radii is strongly suggestive that the Milky Way disc has been heated by external perturbers. We also propose that the changing AVR shape with {$R_{\text{mean}}$} seems to be better fit by models with a slightly positive {$v_{R,0}$}, and have a wave-like trend in {$v_{R,0}$} with {$R_{\text{mean}}$} (See Fig. 8). Old stars in the low [Fe/H] disc do not appear to be part of this warping. We contend that these shifts are likely due to the Galactic warp, which may be partially responsible for the excess heating in the outer disc.

Our main findings place novel constraints on the history of dynamical heating across the disc. The finding that the velocity ellipsoid shape is strongly dependent on Galactocentric radius for low [Fe/H] populations, and that the AVR becomes flatter at greater radii is strongly suggestive that the Milky Way disc has been heated by external perturbers. We also propose that the changing AVR shape with {$R_{\text{mean}}$} seems to be better fit by models with a slightly positive {$v_{R,0}$}, and have a wave-like trend in {$v_{R,0}$} with {$R_{\text{mean}}$} (See Fig. 8). Old stars in the low [Fe/H] disc do not appear to be part of this warping. We contend that these shifts are likely due to the Galactic warp, which may be partially responsible for the excess heating in the outer disc.

Figure 9. The profile of {$\sigma_z/\sigma_R$} with {$|z|$} at $R_0$ for mono-age, mono-[Fe/H] populations in the low and high [α/Fe] disc. Each line represents the best-fitting profile for that mono-age, mono-[Fe/H] bin, with the coloured band indicating the 95 per cent confidence interval. The colour indicates the age of the mono-age, mono-[Fe/H] population. The {$\sigma_z/\sigma_R$} of young, low [α/Fe] populations show a slight, but significant trend with {$|z|$}, increasing at greater {$z$}. The decreased {$\sigma_z/\sigma_R$} at lower {$z$} is indicative of spiral heating acting more strongly in the plane. The high [α/Fe] populations have very flat {$\sigma_z/\sigma_R$} with {$|z|$}, further indicating that these populations are kinematically distinct from the low [α/Fe] population and that their velocity dispersion is unlikely to be due to perturbers confined to the mid-plane.
These results pose many new constraints that must be considered in any model that attempts to explain the kinematics of the Milky Way disc. Importantly, they strongly suggest that external effects may play an important role in setting the kinematic structure of the outer disc. Furthermore, our results point to the need for the inclusion of non-axisymmetricities in any modelling, as the excess heating seen in the data may be caused by the induced warping of the disc, and the kinematics of the low [α/Fe] population may be affected by perturbations from spiral structure. Indeed, our results also suggest that fitting the observed disc kinematics with axisymmetric models (as we do here) may be limiting the inferences that can be made from the data.

Particularly, these results make clear the need for self-consistent modelling using large volume numerical simulations in a cosmological context, which provide statistical samples of galaxies evolving in different environments and with different assembly histories, but also where the stellar dynamics and kinematics are adequately resolved. In this regard, the data are somewhat leading the theory, as simulations of large volumes tend to poorly resolve (or not model) the cold gas phases required to form stars in disc galaxies with realistic kinematics and subsequently realistic dynamical heating processes. However, the most recent zoom-in simulations are now beginning to resolve colder gas phases, and self-consistent model heating from GMCs (e.g. Wetzel et al. 2016; Grand et al. 2017). An advantage of using such simulations is that non-axisymmetries are also modelled self-consistently, allowing the combined effects of bars, spiral arms, warps and waves on the galaxy kinematics to be assessed.

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for a useful report. We also thank Ivan Minchev, Jerry Sellwood, Chao Liu, and Michael Hayden for helpful discussions on the originally submitted manuscript. We also thank Leandro Silva and Martin Smith for highlighting necessary typographic corrections. JTM is funded by an (0: funding-source ) (0:funding-source ) STFC/(0:funding-source ) /0:funding-source ) studentship, and is grateful for funding from the Royal Astronomical Society and the Dunlap Institute for Astronomy and Astrophysics at the University of Toronto and thanks them for their hospitality during an extended visit while preparing this work. JTM and AM acknowledge support from the ERC Consolidator Grant funding scheme (project ASTEROCHRONOMETRY, G.A. n. 772293). JB acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2015-05235, and from an Alfred P. Sloan Fellowship. JT acknowledges that support for this work was provided by NASA through the NASA Hubble Fellowship grant #51424 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NASS-26555. This project was developed in part at the 2017 Heidelberg Gaia Sprint, hosted by the Max-Planck-Institut für Astronomie, Heidelberg, and the 2018 NYC Gaia Sprint, hosted by the Center for Computational Astrophysics of the Flatiron Institute in New York City.

This research has used the cross-match service provided by CDS, Strasbourg. Analyses and plots presented in this article used iPython, and packages in the SciPy ecosystem (Jones et al. 2001; Hunter 2007; Perez & Granger 2007; van der Walt, Colbert & Varoquaux 2011). The study has used high performance computing facilities at Liverpool John Moores University, partly funded by the Royal Society and LJMU’s Faculty of Engineering and Technology.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This work has used data from the European Space Agency (ESA) mission Gaia (http://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, http://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.

REFERENCES

Antoja T. et al., 2018, Nature, 561, 360
Borucki W. J. et al., 2010, Science, 327, 977
APPENDIX A: ESTIMATING AGES FOR RED GIANT STARS USING BAYESIAN
CONVOLUTIONAL NEURAL NETWORKS

The APOKASC-2 catalogue (Pinsonneault et al. 2018) consists of asteroseismic and spectroscopic data measured for 6676 stars in common between the APOGEE and NASA–Kepler mission. This represents a significantly larger catalogue than the previous release, as well as a number of improvements over the original analysis (presented and examined in Pinsonneault et al. 2014). APOKASC DR2 derives stellar properties by combining spectroscopically measured parameters from APOGEE DR14 with light curves from Kepler, the improved reduction of which is described by Elsworth et al. (in preparation). The asteroseismic analysis is performed using a set of five distinct pipelines, where the mean results of these are used to generate the final catalogue. Pipelines estimate the asteroseismic parameters, the large frequency separation \( \Delta \nu \), and the frequency of maximum power, \( v_{\text{max}} \), which are then used in semi-empirical scaling relations that give the mass and radius of the star, given a known \( T_{\text{eff}} \) (the spectroscopic data). Scatter about the ensemble mean of the pipelines is used to infer the random uncertainties on the asteroseismic parameters \( \Delta \nu \) and \( v_{\text{max}} \), which are propagated through to the stellar properties. Any differences between the pipelines that depend on \( \Delta \nu \) and \( v_{\text{max}} \) contribute to a systematic error budget. A key difference between APOKASC-1 and 2 lies in the treatment of the \( \Delta \nu \) scaling relation, which is no longer considered to be an exact relation, and is corrected on a star-by-star basis, through the approach described by Serenelli et al. (in preparation). This theoretical correction also relies on the APOGEE DR14 [Fe/H] and [\( \alpha/\text{Fe} \)] values. The property of most importance to this work, the stellar ages, are finally estimated a star-by-star basis, through the approach described by Serenelli et al. (in preparation). This theoretical correction also relies on the APOGEE DR14 [Fe/H] and [\( \alpha/\text{Fe} \)] values. The property of most importance to this work, the stellar ages, are finally estimated using the derived mass, surface gravity, and abundances with their uncertainties estimated through propagation of the uncertainties on those properties. We use the ‘recommended’ age estimates and uncertainties as tabulated and available in Pinsonneault et al. (2018), without applying any external corrections. The APOKASC-2 ages and their corresponding combined APOGEE spectra form the basis of the training set that we use here to train a model to predict ages for the remainder of the DR14 red-giant sample. This method relies mainly on the relation between the surface abundances of red giants and their main-sequence mass, which is set when the stars undergo first dredge up (FDU). Stars on the main sequence undergo hydrogen fusion by the CNO and CN cycles, in which C, N and O atoms catalyse burning, and so undergo evolution in their own relative abundances. As shown by Salaris et al. (2015), the abundance profile of Nitrogen going inwards to the stellar core (before FDU) reflects first that of the CN cycle, followed by the CNO cycle. The FDU mixes the envelope with the CN layer. Stellar models with increased mass tend to have

SUPPORTING INFORMATION

Supplementary data are available at MNRAS online. astroNNBayes_ages_goodDR14.csv

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.
The spectra and the stellar mass and therefore the implied age. In such that there should be some (likely complicated) relation between surface abundances of C, N, and O are reflected in APOGEE spectra, increasing RGB mass. We assume here that such differences in the stellar interior, meaning that the surface \([\text{C/N}]\) decreases with larger convective envelopes, meaning that they reach deeper into the stars in the APOKASC-2 catalogue with \(-0.05 < [\text{Fe/H}] < 0.05\) dex. We overplot the age–[C/N] relationships as obtained for solar metallicity stars from the BASTI (Salaris et al. 2015) and STAREVOL (Lagarde et al. 2017) stellar evolution models. Both models roughly match the data at young ages (\(\leq 5\) Gyr) but begin diverging at older ages. The Lagarde et al. (2017) models appear to have a systematically low [C/N] at any given age. The Salaris et al. (2015) agree well with the asteroseismic ages (the black points) and test samples (shown in blue, with their associated uncertainties, as estimated through the Bayesian dropout variational inference). These data are included as supplementary material. The bottom panel gives the residual \(\Delta\text{age} \equiv \text{age}_{\text{out}} - \text{age}_{\text{in}}\). The standard deviation of residuals across the whole range is 36 per cent, corresponding to a median \(\Delta\text{age} = 1.26\) Gyr. For input ages less than 5 Gyr, the scatter is reduced to 30 per cent. It is clear, however, that there is a significant trend such that young ages are overestimated, and old ages are underestimated, with a stronger effect at old age. Stars with input ages of \(\sim 10\) Gyr are underestimated in the test set by as much as 3.5 Gyr, with a large scatter, such that no stars are correctly predicted or overpredicted. This behaviour has been seen in other methods that exploit asteroseismology, using simpler polynomial regression method based on abundance measurements to measure ages (e.g. Martig et al. 2016). It is noteworthy that it persists even when the full scope of information in the spectra are used. We also tested the effect on the predicted ages of adjustments to the training data, implying progressively larger cuts in the asteroseismic \(v_{\text{max}}\) in order to remove stars that lie high on the RGB, where the asteroseismic models are less well tested. We find that implying these cuts makes little difference to the final results of the training, and the output ages from the test set match well with those predicted from the full training set in all cases. Dividing the sample between RC and RGB stars (as determined by the asteroseismology) and training the BCNN separately on these populations also provides comparable results, implying that the neural network is able to capture this information from the spectra and account for it when training on the entire APOKASC-2 sample (the presence of such information in stellar spectra has been

![Figure A2](https://academic.oup.com/mnras/article-abstract/489/1/176/552689/1176552689)
discussed by Hawkins, Ting & Walter-Rix 2018; Ting, Hawkins & Rix 2018).

We further examine the issue with the absolute scaling of ages predicted from the neural network (and other) models by comparing the predicted age–[C/N] relationship with that obtained from stellar evolution models. In Fig. A2, we show the BCNN predicted ages (upper panel) and the training set APOKASC-2 ages (lower panel) against the APOGEE DR14 [C/N] for stars in APOKASC-2 with $-0.05 < \text{[Fe/H]} < 0.05$ dex. The age–[C/N] relationship for solar metallicity stars, as obtained from the BASTI (Salaris et al. 2015) and STAREVOL (Lagarde et al. 2017) models, is overplotted in the dotted and dashed lines. It is apparent that the models roughly agree with the training and predicted data at young ages ($\lesssim 5$ Gyr). The Salaris et al. (2015) model appears to agree slightly better than that of Lagarde et al. (2017), which has a systematic offset to lower [C/N] at any given age. At older ages, the spectroscopic ages have higher [C/N] than is predicted by the models, whereas the asteroseismic input ages seem to agree well over the full range of abundance ratios. This test provides a further indication that the oldest ages are underpredicted by the neural network model, as the [C/N] ratios of these stars (as measured by APOGEE) should correspond an older age, given the stellar evolution models. This implies one of two possible explanations, either (1) the ages of old stars provided in the training set are underestimated, or (2) that the spectroscopic information is not representative of the true mass (as measured by asteroseismology) at old age. As the training set appear to match the models relatively well, it would seem that the second explanation is more likely. Fully understanding this effect is beyond the scope of this paper, but will be essential for accurately determining the ages of the oldest stars in the Milky Way using spectroscopy.

In a previous paper in this series, Mackereth et al. (2017), we chose to correct for the underprediction of old ages using a non-parametric LOWESS fit to the ageout–agein relation. Here, we leave underpredicted ages in the full data set, and simply warn readers that results pertaining to any trends in age at values greater than $\sim 8$ Gyr likely extend from that age to the oldest stars in the sample. The above tests demonstrate that the sample likely contains stars older than the $\sim 10$ Gyr limit seen in the final predicted ages, above which it appears the spectroscopic information is no longer useful to predict ages. The final age catalogue for DR14 stars, as predicted by the BCNN, is included in a supplementary data table.

APPENDIX B: EXAMPLE POSTERIOR PDF OF VELOCITY DISTRIBUTION MODEL

To demonstrate the overall quality of the fit to the velocity distribution that can be achieved using the APOGEE–Gaia sample, we provide here an example of the posterior probability distribution gained from the MCMC sampling of the parameters for the velocity distribution of a single exemplary bin in age–[Fe/H]. The bin that we choose to show here is at intermediate age ($4.0 < \text{age} < 4.5$ Gyr old) and subsolar [Fe/H] ($-0.1 < \text{[Fe/H]} < 0.$) in the low [$\alpha$/Fe] population, which has 2178 stars. The posterior distribution of each of the parameters and their covariance is shown in a corner plot, in Fig. B1. The posteriors are in most cases well behaved and approximately Gaussian. As expressed in Section 3, decreasing the number of stars used for the fit broadens the distributions, but the median values remain the same within the uncertainties. Checking the posterior distributions for all the mono-age, mono-[Fe/H] bins.
Figure B1. Posterior distributions of the parameters for a single, exemplary, mono-age, mono-[Fe/H] bin in the low [α/Fe] population with 4.0 < age < 4.5 Gyr old, and −0.1 < [Fe/H] < 0. The posteriors are generally well behaved and approximately Gaussian, with little covariance between most of the parameters.