

Isochrone ages for ~ 3 million stars with the second *Gaia* data release

Jason L. Sanders¹★ and Payel Das²

¹*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

²*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK*

Accepted 2018 September 7. Received 2018 September 7; in original form 2018 May 25

ABSTRACT

We present a catalogue of distances, masses, and ages for ~ 3 million stars in the second *Gaia* data release with spectroscopic parameters available from the large spectroscopic surveys: APOGEE, *Gaia*-ESO, GALAH, LAMOST, RAVE, and SEGUE. We use a Bayesian framework to characterize the probability density functions of distance, mass, and age using photometric, spectroscopic, and astrometric information, supplemented with spectroscopic masses where available for giant stars. Furthermore, we provide posterior extinction estimates (A_V) to every star using published extinction maps as a prior input. We provide an appendix with extinction coefficients for *Gaia* photometry derived from stellar models, which account for variation with intrinsic colour and total extinction. Our pipeline provides output estimates of the spectroscopic parameters, which can be used to inform improved spectroscopic analysis. We complement our catalogues with Galactocentric coordinates and actions with associated uncertainties. As a demonstration of the power of our catalogue, we produce velocity dispersion profiles of the disc separated by age and Galactocentric radius (between 3 and 15 kpc from the Galactic centre). This suggests that the velocity dispersion profiles flatten with radius in the outer Galaxy (> 8 kpc) and that at all radii the velocity dispersion follows the smooth power law with age observed in the solar neighbourhood.

Key words: stars: fundamental parameters – Galaxy: evolution – Galaxy: kinematics and dynamics – Galaxy: stellar content – Galaxy: structure.

1 INTRODUCTION

The second *Gaia* data release (*Gaia* DR2; Gaia Collaboration 2018a) heralds a revolution in the study of the Milky Way. *Gaia* (Gaia Collaboration 2016) has provided parallaxes and proper motions for ~ 1.3 billion Milky Way stars (Lindgren et al. 2018) as well as highly accurate multiband photometry for many of these stars (Evans et al. 2018; Riello et al. 2018). For a subset of stars brighter than 12th magnitude in the narrow *Gaia* Radial Velocity Spectrometer (RVS) photometric band (G_{RVS}), radial velocities have been computed from the *Gaia* RVS spectra (Cropper et al. 2018; Sartoretti et al. 2018). Whilst in its own right this data set will lead to rapid advances in our knowledge of the structure of the Milky Way (Gaia Collaboration 2018c), when complemented with other photometric surveys and large-scale spectroscopic surveys, *Gaia* is the perfect tool for Galactic archaeology studies.

The primary goal of Galactic archaeology is to construct a census of the structure of the Milky Way in spatial, kinematic, and chemical space (Bland-Hawthorn & Gerhard 2016). Differences in the chemodynamical structure of populations reflects both differences between the chemodynamical structure of the populations at their

formation epochs as well as differences in their subsequent dynamical evolution. The secondary goal of Galactic archaeology is to use the census of the Galaxy to infer the properties of the Galaxy at different epochs of star formation and hence the evolution history of the Galaxy from this data. This must necessarily be informed by Milky Way-like simulations and models. To disentangle the impacts of dynamical evolution from the formation environment in this way, we require the rich, detailed extended picture of our Galaxy, provided by the synergy of *Gaia* with large-scale spectroscopic surveys. One question of interest for example is to understand whether the correlation between stellar age and velocity dispersion is a consequence of older stars being born hotter, or stars heating over time.

Before *Gaia*, reliable kinematics (from parallaxes) were only available in the solar neighbourhood (Holmberg, Nordström & Andersen 2009; Bensby, Feltzing & Oey 2014). Early extended studies of the Milky Way were limited to using large-scale photometric surveys (e.g. 2MASS, SDSS; Skrutskie et al. 2006; Abolfathi et al. 2018). In recent years, large-scale spectroscopic surveys have been the dominant tool in furthering the study of the Milky Way. Many spectroscopic surveys (APOGEE, LAMOST, RAVE, GALAH, *Gaia*-ESO, SEGUE; Steinmetz et al. 2006; Yanny et al. 2009; Deng et al. 2012; Gilmore et al. 2012; De Silva et al. 2015; Majewski et al. 2017) have been designed with the intention of com-

* E-mail: jls@ast.cam.ac.uk

plementing the astrometry from *Gaia* with radial velocity information as well as detailed chemical abundances. With full phase-space coordinates, the dynamical structure of the Galaxy well beyond the solar neighbourhood can now be mapped in detail. Going forward, future surveys such as WEAVE (Dalton et al. 2012), 4-MOST de Jong et al. (2016), Milky Way Mapper (Kollmeier et al. 2017), and MOONS (Cirasuolo et al. 2012) will further complement the *Gaia* data and provide a more complete picture of the populations of the Milky Way.

The combination of spectroscopy, photometry, and astrometry allows accurate characterization of stellar properties. Given a set of stellar models, this data combination can give accurate measurements of the distance, mass, age, and other spectroscopic properties of the stars. In time, the *Gaia* data will be used to produce improvements in stellar models. However, a first step is to inspect the properties of the stars given the currently available models. Measuring distances from photometry and spectroscopy via sets of stellar isochrones has been developed over a number of years (Pont & Eyer 2004; Jørgensen & Lindegren 2005; Burnett & Binney 2010; Binney et al. 2014) and has been widely applied to current spectroscopic surveys (e.g. Mints & Hekker 2018; Queiroz et al. 2018). Utilizing additional information from parallaxes (e.g. da Silva et al. 2006; McMillan et al. 2018) and/or masses (Das & Sanders 2018) is a natural extension. We highlight two synergies of the methodology. For a typical star with $G < 15$, *Gaia* DR2 provides a median parallax uncertainty of ~ 0.04 mas giving parallaxes precise to 20 per cent out to ~ 5 kpc. For more distant giants observed by spectroscopic surveys, spectro-photometric distances can be superior. The combination of spectro-photometry and astrometric information is particularly powerful (McMillan et al. 2018) – for instance, *Gaia* parallaxes can cleanly distinguish between nearby dwarfs and distant giants, producing improved spectroscopic parameter estimates. These estimates are useful initial guesses for improved analyses of the spectra.

The real advantage, however, of comparing astrometry and spectro-photometry to sets of stellar models is estimating stellar ages. The age of a star is a fundamental parameter in understanding the formation and evolution of the Galaxy, but, except in a few limited cases, can only be inferred via models (Soderblom 2010). The most reliable age estimates are estimated for (near) co-eval cluster stars. Ages of co-eval populations are primarily constrained by the location of the main-sequence turn-off (e.g. Marín-Franch et al. 2009) as a star’s initial mass determines the time at which it leaves the main sequence. Such an approach can be utilized in the single field star case for those stars that are transitioning from the main sequence to the giant branch (e.g. Xiang et al. 2015, 2017). However, even with parallaxes from *Gaia* there is a strong degeneracy between metallicity and age that must be broken with the use of spectroscopic metallicities (Howes et al. 2018). Other populations of stars (e.g. lower main sequence dwarfs and giants) are much more difficult to accurately date due to only subtle differences in their observed spectro-photometric properties. However, recent theoretical and empirical work (Mason & Gilmore 2015; Martig et al. 2016) has demonstrated that there are clear spectroscopic indicators of a giant star’s mass (carbon and nitrogen atomic and molecular lines) and hence age (e.g. Ho et al. 2017; Wu et al. 2018). When further combined with information on a giant’s luminosity from *Gaia* parallaxes, it is also possible to accurately date giant stars. Das & Sanders (2018) have recently demonstrated the power of including masses for giant stars in the standard isochrone pipeline.

In this paper, we provide distances, masses, ages, extinctions, and spectroscopic parameters (T_{eff} , $\log g$, and $[M/H]$) for stars from spectroscopic surveys combined with the *Gaia* data. In Section 2, we

describe the method employed, focussing on details of the extinction priors and extinction law, the scheme for estimating masses from spectroscopic parameters (‘spectroscopic mass estimates’) and the adopted Milky Way prior. In Section 3, we describe the data sets to which we have applied our algorithm, along with the *Gaia* data employed. In Section 4, we discuss the results of our procedure concentrating on the quality of the catalogue as well as indicating the possible chemo-dynamical studies that such a catalogue makes possible. We present our conclusions in Section 5.

2 METHOD

We lay out the framework used to construct a probability density function for a star’s distance. We assume the stars are single (see Coronado, Rix & Trick 2018, for handling binary stars) and are solely the products of single star evolution (which does not encompass the significant blue straggler population). We will see that assuming single stars can lead to identification of potential multiple systems a posteriori. We first present the framework using the isochrones in this subsection, discuss extinction in Section 2.1, discuss complementing our data with spectroscopic mass estimates in Section 2.2, present our choice of Galaxy prior in Section 2.3 and discuss the outputs of our approach in Section 2.4.

We use the Bayesian method presented in Burnett & Binney (2010). As discussed in Bailer-Jones et al. (2018) and Luri et al. (2018), when working with *Gaia* parallaxes it is necessary to employ a full Bayesian scheme including a reasonable distance prior to yield meaningful results (particularly in the cases of parallax measurements of order the uncertainty or negative parallaxes). We assess the probability of the i th star being at a distance s (with distance modulus $\mu = 10 + 5 \log_{10}[s/\text{kpc}]$) given the data D_i , which consists of observed spectroscopic data $S_i = (\log g, T_{\text{eff}}, [M/H])_i$, photometric data $m_i = (J, H, K_s, \dots)_i$, astrometric data (Galactic coordinates ℓ_i, b_i , and parallax ϖ_i) and in some cases mass estimates (\mathcal{M}_i from spectroscopic mass estimators calibrated with asteroseismology) with associated uncertainties. The uncertainties in the spectroscopic parameters are given by the spectroscopic parameter covariance matrix Σ . The uncertainties in the photometry are assumed to be uncorrelated and given by $\sigma_{m,i}$, whilst the parallax and mass have associated uncertainties $\sigma_{\varpi,i}$ and $\sigma_{\mathcal{M},i}$ respectively. We drop the subscript i in the following, for clarity. The resulting pdf is given by

$$p(s|D) = \frac{p(D|s)p(s)}{p(D)} \propto \int dI p(D|I, s)p(s|I)p(I), \quad (1)$$

where the integral is performed over a set of isochrones $\{I\}$ indexed by metallicity, $[M/H]$, and log-age, $\log_{10}\tau$. We work with a set of PARSEC isochrones (v1.2S – excludes thermally pulsing AGB phase and using Reimers mass loss $\eta = 0.2$, Bressan et al. 2012; Chen et al. 2014; Tang et al. 2014; Chen et al. 2015) spaced by 0.01 dex in $[M/H]$ and 0.05 dex in $\log_{10}\tau$ (up to a maximum age of 12.6 Gyr). Each isochrone gives a set of observed properties as a function of the initial mass \mathcal{M}_0 . The integral over isochrones can therefore be written as

$$\int dI = \int d[M/H] d\tau d\mathcal{M}_0 = \sum_I \Delta[M/H] \Delta\tau \Delta\mathcal{M}_0, \quad (2)$$

where $\Delta[M/H] \Delta\tau \Delta\mathcal{M}_0$ is the volume occupied by each isochrone point I . Given an isochrone point and a distance s , the model spectroscopic $S' = (\log g, T_{\text{eff}}, [M/H])$, photometric (absolute magnitudes M') properties, and current mass (\mathcal{M}') may be computed and compared with the observed properties given the reported errors.

We write the likelihood term

$$p(D|I, s) = p(m|M', s)p(S|S')p(\varpi|s)p(\mathcal{M}|S') \quad (3)$$

and expand the first term as

$$p(m|M', s) = \int d \ln A_V p(m|M', s, \ln A_V) p(\ln A_V),$$

$$p(\ln A_V) = \mathcal{N}(\ln A_V - \langle \ln A_V \rangle, \sigma_{\ln A_V}), \quad (4)$$

to marginalize over the unknown V -band extinction A_V . The four likelihood terms are then given by

$$p(m|M', s, A_V) = \prod_k \mathcal{N}(m_k - M'_k - \mu(s) - A_k, \sigma_{mk}),$$

$$p(S|S') = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left(-\frac{1}{2} (S - S') \Sigma^{-1} (S - S') \right),$$

$$p(\varpi|s) = \mathcal{N}(\varpi - 1/s, \sigma_\varpi),$$

$$p(\mathcal{M}|S') = \mathcal{N}(\mathcal{M} - \mathcal{M}', \sigma_{\mathcal{M}}), \quad (5)$$

where we have introduced the notation $\mathcal{N}(\mu, \sigma)$ for a Gaussian with mean μ and standard deviation σ , and A_k is the extinction in the k th photometric band. The other term in our posterior is the prior $p(s|I)p(I)$, which we detail in Section 2.3.

2.1 Extinction

When using photometry in our pipeline, we have to correct for the line-of-sight extinction. We choose to characterize this by the parameter A_V , the extinction in the Johnson V band. The total extinction A_V is related to the selective extinction $E(B - V) = A_B - A_V$ by $R(V) = A_V/E(B - V)$. Therefore, the extinction coefficient for the V band, $R(V)$, is dependent on the assumed extinction law. For photometric bands other than V the extinction is given by $A_i = R(i)E(B - V)$ so is simply related to our parameter A_V via $A_i = A_V R(i)/R(V)$. The set of coefficients $R(i)/R(V)$ describe the adopted extinction law whilst A_V scales the total extinction. We first describe our choice of $R(i)/R(V)$ before discussing priors on A_V .

2.1.1 Extinction coefficients

Here, we briefly describe the adopted extinction coefficients, $R(i)$. Full details are given in Appendix A. We use the extinction curve $A(\lambda)$ (total extinction at wavelength λ) from Schlafly et al. (2016) calibrated using the Apache Point Observatory Galactic Evolution Experiment (APOGEE) data. We scale $A(\lambda)$ to the units E' adopted in the extinction maps of Green et al. (2018), E' is slightly different from $E(B - V)$. Therefore, the provided coefficients $R(i)$ can be used in combination with Green et al. (2018) extinction map to find the extinction in band i as $A_i = R(i)E'$.

Using the stellar model with $T_{\text{eff}} = 4500$ K and $\log g = 4.5$ from Castelli & Kurucz (2004), we compute $R(i)/R(V)$ for all photometric bands of interest (SDSS u, g, r, i, z , Pan-STARRS g_P, r_P, i_P, z_P , 2MASS J, H, K_s , APASS B, V , WISE W_1, W_2 , Gaia $G, G_{BP}, G_{RP}, G_{RVs}$) by integrating the photometric band response over the spectrum. For the broad Gaia G band we consider variation of $R(G)$ with effective temperature by tabulating $R(G)$ over the full range of Castelli & Kurucz (2004) models. In Appendix A, we provide expressions for $R(G)$ as a function of effective temperature and intrinsic colours in different bands.

There is also variation of $R(G)$ with A_V . We evaluate equation (A1) at a range of monochromatic extinctions and measure the gradient $dR(i)/dA_V$. The results are provided in Appendix A. For computational reasons, we do not consider variation of the extinction law

(characterized by the parameter R_V) despite Schlafly et al. (2016) demonstrating that there is variation of ~ 0.2 across the APOGEE fields.

2.1.2 Extinction prior

Our prior $p(\ln A_V)$ uses extinction measurements from a combination of three extinction maps (c.f. Bovy et al. 2016). We preferentially use the extinction maps from Green et al. (2018, using the `dustmaps bayestar` interface). For each star that falls in the Pan-STARRS footprint, we draw 10 samples of the extinction E' (in the `bayestar` 2017 units) at a set of discrete distances along the line-of-sight. At each distance, we use the coefficients from the previous section to compute the mean $\langle \ln A_V \rangle$ and the standard deviation $\sigma_{\ln A_V}$. For any star that falls outside the Pan-STARRS footprint (which only extends to $\delta \sim -30$ deg), we next attempt to use the extinction map from Marshall et al. (2006) which is confined to $|l| < 100$ deg and $|b| < 10$ deg and provides mean K_s extinction and its uncertainty. For each star, we use $R(V)/R(K_s)$ from the previous section to find $\langle \ln A_V \rangle$ and $\sigma_{\ln A_V}$ on a grid in distance. Finally, where neither of these maps are available we use the 3D extinction map (expressed in A_V) from Drimmel, Cabrera-Lavers & López-Corredoira (2003) using the interface from Bovy et al. (2016). Again we tabulate $\langle \ln A_V \rangle$ on a grid in distance and assume 30 per cent uncertainty in A_V .

2.2 Mass estimates

For giant stars, a mass measurement is a near direct measurement of age. Asteroseismology provides an estimate of stellar mass given measurements of the frequency spectrum of a star's oscillations. However, such observations require high-quality photometry over a long time baseline so are limited to small subsets of stars confined to limited regions of the Galaxy (e.g. the Kepler field). Recent work (Masseron & Gilmore 2015; Martig et al. 2016; Ness et al. 2016) has demonstrated empirically that the $[C/N]$ ratio in giant stars is an indicator of mass. Theoretical expectation (Charbonnel 1994) is that both the equilibrium position of core CNO burning and strength of dredge-up are functions of stellar mass. Higher mass stars produce more nitrogen than carbon leading to a suppression in $[C/N]$. We wish to use the carbon and nitrogen abundances as constraints within our framework by converting these spectroscopic parameters into estimates of the stellar mass.

We adopt the procedure in Das & Sanders (2018, DS18) for relating spectroscopic parameters to mass. This involves constructing a Bayesian artificial neural network for the input spectroscopic parameters $X = (T_{\text{eff}}, \log g, [M/H], [\alpha/M], [C/M], [N/M])$ and their uncertainties σ_X to the mass \mathcal{M} and its uncertainty $\sigma_{\mathcal{M}}$ (DS18 use the procedure to estimate mass, age and distance). We scale both the input X and output parameters \mathcal{M} to approximate unit gaussians. The neural network architecture presented here differs slightly with that in DS18. Here, the neural network contains two hidden layers (with 24 hidden nodes each) rather than one. Although this is a more complex architecture, marginalizing over the model parameters does not result in significant overfitting. We again assume each layer (except the output) uses a sigmoid function σ

$$f_\theta(X) = w_o \sigma(w_2 \sigma(w_1 X + b_1) + b_2) + b_o. \quad (6)$$

The weights w_i are matrices and the biases b_i vectors. To find the posterior distributions on the neural network parameters $\theta = (b_i,$

w_i), we evaluate

$$p(\theta|X, \mathcal{M}) = p(X, \mathcal{M}|\theta)p(\theta). \quad (7)$$

The likelihood term is given by

$$p(X, \mathcal{M}|\theta) = \int dX' \mathcal{N}(X - X', \sigma_X) \mathcal{N}(\mathcal{M} - f_\theta(X'), \sigma_{\mathcal{M}}), \quad (8)$$

and we choose normal priors on w_i with zero mean and standard deviations given by hyperparameters r – one for each of the six parameters $w_1, w_2, w_o, b_1, b_2, b_o$. The inclusion of these hyperparameters is a development from DS18 and are important as, although the inputs and outputs have been scaled to approximate normal distributions, the hidden layers can have considerably larger dynamic range.

We implement this model in PyMC3 (Salvatier, Wiecki & Fonnesbeck 2016) and train using automatic differentiation variational inference (ADVI). As our training set we use the overlap of APOGEE DR14 (Abolfathi et al. 2018, described later) with asteroseismic results (Pinsonneault et al. 2014; Vrad, Mosser & Samadi 2016). We preferentially work with the Vrad et al. (2016) sample and supplement with Pinsonneault et al. (2014) APOKASC results for those stars not in Vrad et al. (2016). We remove stars with ASPCAPFLAG $\neq 0$, those without all elements of X measured and duplicates. We further remove those stars that have been identified as rotating by Tayar et al. (2015).

For unseen data, the posterior distribution on the mass is given by

$$p(\mathcal{M}|X) = \int dX' d\theta p(\mathcal{M}|X', \theta) \mathcal{N}(X - X', \sigma_X) p(\theta), \quad (9)$$

where the first term is a δ -function involving $f_\theta(X')$. We generate samples from this pdf by using the trace output from PyMC3 for the model parameters θ and sample from a Gaussian for the spectroscopic parameters. We reduce the resulting pdf to the first two moments: mean and standard deviation. We only apply this method if the inputs satisfy $4000 < T_{\text{eff}}/\text{K} < 5250$, $1 < \log g < 3.3$ and $-1.5 < [\text{M}/\text{H}] < 0.5$, which is the approximate parameter range covered by the training set.

Here, we have used ADVI to characterize the pdf whereas DS18 fully sampled from the pdf using the No-U-Turn sampler (NUTS, Hoffman & Gelman 2011). The disadvantage of ADVI is that we fail to characterize the (potentially significant) correlations and multimodality of the pdf (although highly isolated modes will not be well sampled by NUTS). Additionally, DS18 found the behaviour of NUTS to be smoother when varying the architecture. However, ADVI is considerably faster and in practice the results obtained are quite similar.

The advantage of employing this method is that we need not perform quality cuts on our samples (provided we trust the reported uncertainties), and the uncertainty in the mass estimate reflects both the input uncertainty and the model uncertainty. Modelling the training set permits some level of de-noising of the output masses. The output mass uncertainty for unseen data is comparable (or better than) the mass uncertainty in the training set, significantly extending the power of asteroseismology to vast numbers of stars. Additionally, the flexibility of the model allows us to extrapolate into regions of spectroscopic parameter space that are sparsely populated where our model becomes more uncertain. A disadvantage is we are utilizing spectroscopic information e.g. T_{eff} to constrain the mass and both mass and T_{eff} are then used in the distance pipeline. We are therefore using some spectroscopic information twice. However, using solely T_{eff} , $\log g$ and $[\text{M}/\text{H}]$ produces poor constraints

on the mass, and most of the constraint on the mass comes from $[\text{C}/\text{M}]$ and $[\text{N}/\text{M}]$ which are not further utilized. The use of the other spectroscopic parameters can be thought of as weakly adjusting the relationship between $[\text{C}/\text{M}]$ and $[\text{N}/\text{M}]$, and mass. Furthermore, we must assume that the stars within our asteroseismic sample are representative of the stars in our entire spectroscopic sample. Finally, we are required to use input spectroscopic parameters calibrated on the same scale as the training set.

2.3 Galaxy prior

We choose to adopt a non-uniform prior distribution in distance, age and metallicity reflecting the prior knowledge that the considered stars belong primarily to the Galactic disc. For this we introduce the 3D prior distribution $p_{\text{gal}}(\mathbf{x}, [\text{M}/\text{H}], \tau)$ with which we must include a Jacobian factor s^2 to relate to the prior on distance. Additionally, we use the Kroupa initial mass function (Kroupa, Tout & Gilmore 1993) $p(\mathcal{M}_0)$ as a prior on the initial mass:

$$p(s|I)p(I) = s^2 f_{\text{prior}}(\mathbf{x}, [\text{M}/\text{H}], \tau, \mathcal{M}_0) = s^2 p_{\text{gal}}(\mathbf{x}, [\text{M}/\text{H}], \tau) p(\mathcal{M}_0). \quad (10)$$

We decompose our Galaxy prior into multiple components

$$p_{\text{gal}}(\mathbf{x}, [\text{M}/\text{H}], \tau) = \sum_i p_{\text{gal},i}(\mathbf{x}, [\text{M}/\text{H}], \tau). \quad (11)$$

Following Binney et al. (2014), we use a three-component (thin disc, thick disc, and halo prior) but supplemented by a bulge prior (c.f. Queiroz et al. 2018). As we are interested in producing reliable age estimates, we follow Queiroz et al. (2018) and adopt a smooth age prior [as opposed to the truncated age prior from Binney et al. (2014) that e.g. assigns zero probability to young ages for high latitude distant stars]. Each component can be written in the separable form

$$p_{\text{gal},i}(\mathbf{x}, [\text{M}/\text{H}], \tau) = p_{\text{gal},i}(\mathbf{x}) p_{\text{gal},i}([\text{M}/\text{H}]) p_{\text{gal},i}(\tau). \quad (12)$$

We detail each component in turn. All age distributions are truncated at 12.6 Gyr (the largest isochrone age we consider) and normalized appropriately.

(i) Thin disc: Double exponential profile (exponential in both R and z) with scalelength $R_d = 2.6$ kpc and scaleheight $z_d = 0.3$ kpc normalized with local density $0.04 \text{ M}_\odot \text{ pc}^{-3}$ (Bovy 2017), Gaussian distribution in metallicity (mean -0.1 dex, standard deviation 0.3 dex), age distribution

$$p_{\text{gal,thin}}(\tau) = \begin{cases} \exp(\tau/(8.4 \text{ Gyr})) & \tau < 8 \text{ Gyr}, \\ 2.6 \exp(-\frac{1}{2}(\tau - 8 \text{ Gyr})^2/(1.5 \text{ Gyr})^2) & \tau > 8 \text{ Gyr}. \end{cases} \quad (13)$$

(ii) Thick disc: Double exponential profile with scalelength $R_d = 2.0$ kpc and scaleheight $z_d = 0.9$ kpc normalized with local density $0.04 \times 0.04 \text{ M}_\odot \text{ pc}^{-3}$ (Bland-Hawthorn & Gerhard 2016; Bovy 2017), Gaussian distribution in metallicity (mean -0.6 dex, standard deviation 0.5 dex), truncated gaussian in age (mean 10 Gyr, standard deviation 2 Gyr)

(iii) Halo: spherical power law with $r^{-3.39}$ (Binney et al. 2014) normalized with local density $0.005 \times 0.04 \text{ M}_\odot \text{ pc}^{-3}$ (Bland-Hawthorn & Gerhard 2016; Bovy 2017), Gaussian in metallicity (mean -1.6 dex, standard deviation 0.5 dex), truncated gaussian in age (mean 12 Gyr, standard deviation 2 Gyr),

(iv) Bulge: Besançon density profile (Robin et al. 2012):

$$p_{\text{gal,bulge}}(\mathbf{x}) = \text{sech}^2 \left\{ \left[\left(\frac{x'}{x_b} \right)^{c_\perp} + \left(\frac{y'}{y_b} \right)^{c_\perp} \right]^{c_\parallel/c_\perp} + \left(\frac{z}{z_b} \right)^c \right\} f(R), \quad (14)$$

where $f(R) = e^{-(R-R_c)^2/R_s^2}$ if $R > R_c$. R is cylindrical polar radius and primed coordinates are aligned with the bar (at an angle 19.57 deg relative to $l = 0$, Simion et al. 2017). We normalize the profile such that the central density is $35.45 \text{ M}_\odot \text{ pc}^{-3}$ (Robin et al. 2012), and we set $R_c = 2.54 \text{ kpc}$ and $R_s = 0.5 \text{ kpc}$ from Sharma et al. (2011) and all other parameters from Simion et al. (2017). S model fits to VVV data, Gaussian in metallicity (mean 0 dex, standard deviation 0.5 dex), truncated gaussian in age (mean 10 Gyr, standard deviation 3 Gyr).

Importantly, we place the Sun at radius $R_0 = 8.2 \text{ kpc}$ and height above the plane $z_0 = 15 \text{ pc}$ (Bland-Hawthorn & Gerhard 2016).

2.4 Outputs

With our specification, the full pdf of distance can be constructed using equation (1). In practice, we simply compute the first two moments of the distance distribution, e.g.

$$\langle \mu \rangle = \left(\int dI ds \mu p(D|I, s)p(I, s) \right) / \left(\int dI ds p(D|I, s)p(I, s) \right). \quad (15)$$

For this we sum over a range in distance modulus at each isochrone point given by $m - M - A_m \pm N\sigma_m$, where σ_m is the uncertainty in the measured apparent magnitude m and M is the absolute magnitude of the isochrone point (we use $m = J$ or $m = J_{\text{VISTA}}$). For speed, the extinction prior $p(\ln A_V)$ is only computed at the central value $m - M - A_m$. Additionally, the sum in equation (2) is only performed over the isochrone points within N standard deviations of the observed metallicity and then we only consider points within N standard deviations of the reported $\log g$ and T_{eff} . We choose $N = 5$ but increase this (iteratively by a factor of two) if there is no reported overlap with any isochrone points on a first pass. The range of integration for $\ln A_V$ is found from $(\ln A_V)(s_0) \pm 3\sigma_{\ln A_V}(s_0)$ evaluated at a first estimate of the distance s_0 neglecting extinction (using only infra-red photometry).

In a similar fashion to equation (15), we can compute the moments of the log-age distribution for each star, and the covariance between distance modulus and log-age. We also compute the first and second moments of the effective temperature, surface gravity, metallicity, initial mass, and the logarithm of the extinction.

The pipeline will fail if $N \geq 20$ meaning there is no overlap of the data with any isochrone point within 20 times the uncertainties. In this case, the star is flagged in the output catalogue (see Section 3.8). For each spectroscopic data set, we thin the isochrone grid in metallicity to the median metallicity uncertainty of the data set. This results in some failures (see Section 2.4) which we reanalyse using the finest spacing of 0.01 dex.

With full 6D data for the stars (on-sky position, distance, line-of-sight velocity, and proper motion) we compute the Galactocentric velocity (using the peculiar solar velocity from Schönrich, Binney & Dehnen 2010), as well as the action coordinates and guiding-centre radius (using the Stäckel fudge method, Binney 2012; Sanders & Binney 2016) in the potential of McMillan et al. (2018). We first draw 50 samples from the multivariate distance modulus, line-of-sight velocity, and equatorial proper motion distribution. We account for the covariance in the *Gaia* parallax and proper motions by first drawing samples using the *Gaia* measurements and then using rejection sampling to generate 50 samples with acceptance proportional to $\mathcal{N}(\langle \mu \rangle, \sigma_\mu) / \mathcal{N}(\varpi, \sigma_\varpi)$ where $\langle \mu \rangle$ and σ_μ are the mean and uncertainty in the distance modulus from our pipeline and ϖ and σ_ϖ are the mean and standard deviation of the parallax reported by *Gaia*. For each sample we compute the derived quantities and

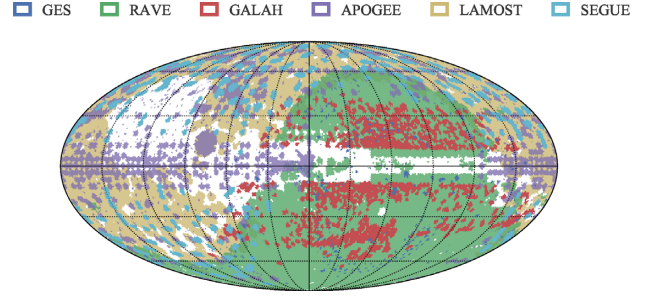


Figure 1. On-sky galactic distribution of the stars processed by our pipeline coloured by survey.

report the mean and standard deviation as the best estimate and uncertainty (ignoring unbound samples – if no samples are bound, the resulting actions are undefined). Note that the uncertainty in the actions does not reflect the uncertainty in the solar position or the potential.

3 SPECTROSCOPIC DATA SETS

In this section, we describe the spectroscopic data sets to which we apply our algorithm. We give a brief overview of each spectroscopic survey along with the specific data we use. In Fig. 1, we show the on-sky distributions of the different surveys we use. For APOGEE, LAMOST, and GALAH giants, we apply the spectroscopic mass estimator. For Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), we first build a data-driven model of the spectra (the Cannon, Casey et al. 2016) to find spectroscopic parameters calibrated to the APOGEE scale. For GALactic Archaeology with HERMES (GALAH) we build a data-driven model of the spectroscopic labels directly using a neural network to put GALAH spectroscopic parameters on the APOGEE scale.

3.1 APOGEE

As part of SDSS IV (Blanton et al. 2017), APOGEE has targeted primarily stars confined to the disc and bulge with infrared spectroscopy. The North survey uses the APOGEE 300-fibre spectrograph (Wilson et al. 2010) on the Sloan 2.5 m telescope (Gunn et al. 2006) at Apache Point Observatory. APOGEE spectra are taken in the H band (15 200–16 900 Å) with a resolution of $R \sim 22\,500$. We use the APOGEE DR14 catalogue (Abolfathi et al. 2018) removing duplicates (retaining the highest signal-to-noise). We use the calibrated T_{eff} , $\log g$, and $[\text{M}/\text{H}]$ (García Pérez et al. 2016) along with the reported covariances, and the 2MASS J , H , K_s photometry with uncertainties (Skrutskie et al. 2006). We use the provided radial velocities with uncertainties (stars with only a single visit have $\text{VSCATTER}=0$ so we assign these stars the median VSCATTER for all stars). Although APOGEE was designed to target giant stars, there are many nearby dwarf stars. These stars do not have provided $\log g$ in DR14 as they differed significantly from stellar models, but T_{eff} and $[\text{Fe}/\text{H}]$ is reported. For all stars with valid T_{eff} and $[\text{Fe}/\text{H}]$, but no $\log g$ we assign $\log g = 4.5 \pm 2$. Note, we do not perform any cuts on quality of spectroscopic parameters (flagged with ASPCAPFLAG) instead leaving this to be done at a later stage of analysis. We use the procedure of Section 2.2 to assign masses and uncertainties given the reported $[\alpha/\text{M}]$, $[\text{C}/\text{M}]$, and $[\text{N}/\text{M}]$.

3.2 LAMOST

LAMOST (Large Sky Area Multi-Object Fiber Spectroscopic Telescope, Cui et al. 2012; Zhao et al. 2012) Experiment for Galactic Understanding and Exploration (LEGUE, Deng et al. 2012) is a low-resolution ($R \sim 1800$) optical (3650–9000 Å) spectroscopic survey designed to study the Milky Way disc and halo. From the third data release A, F, G, K catalogue (<http://dr3.lamost.org/>), we use the reported T_{eff} , $\log g$ and $[\text{Fe}/\text{H}]$ (as a proxy for metallicity) and their uncertainties (approximately 3.1 million stars). We complement the catalogue with 2MASS J , H , K_s filtering on quality (different photometric catalogues were used in the LAMOST target selection). We use photometry provided:

- (i) the photometric quality flag `ph_qual` is A, B, C, or D,
- (ii) the contamination flag `cc_flg` is 0.
- (iii) if the read flag `rd_flg` is 2 (i.e. the magnitude is from profile fitting), `x_psf_chi` (where X is the magnitude) must be less than 3.

When 2MASS photometry does not satisfy these requirements, we instead use Pan-STARRS photometry (g_P , r_P , and i_P) along with G or just G , G_{RP} , and G_{BP} when Pan-STARRS is not available (Magnier et al. 2013, or potentially saturated g_P , r_P and $i_P < 13.5$ mag). Comparisons with the *Gaia* RVS sample (Sartoretti et al. 2018) and APOGEE Anguiano et al. (2018) has demonstrated the radial velocities from LAMOST are too small by 4.5 km s^{-1} so we correct all radial velocities by this amount.

3.2.1 Spectroscopic mass estimates

We further complement the LAMOST giants with mass measurements using a two-stage procedure. First, we apply the Cannon method of Ho et al. (2017) to a sample of $\sim 20\,000$ stars cross-matched between APOGEE and LAMOST (with `ASPCAPFLAG`, `C_M_FLAG`, `N_M_FLAG` = 0). We repeat the analysis of Ho et al. (2017) as (i) we wish to use our framework of Section 2.2 which has been calibrated using updated APOGEE parameters and (ii) we are using LAMOST DR3 instead of DR2. We build a seven label (T_{eff} , $\log g$, $[\text{M}/\text{H}]$, $[\alpha/\text{M}]$, $[\text{C}/\text{M}]$, $[\text{N}/\text{M}]$, A_K) using the DR14 calibrated APOGEE parameters using the code of Casey et al. (2017).¹ All spectra are normalized using a smoothed version of the spectrum (with FWHM 50 Å). Unlike Ho et al. (2017), we do not use any photometry (resulting in poorer estimates of A_K).

As Ho et al. (2017) found, there is a weak correlation between the output $[\alpha/\text{M}]$ and line-of-sight velocity due to sky features in the spectrum. We do not correct for this by omitting regions contaminated by sky features as it does not seem a significant issue for measuring the mass (which is a weak function of $[\alpha/\text{M}]$). We employ a leave-10 per cent-out scheme to measure the accuracy of our approach. The formal errors from the Cannon are significantly smaller than the scatter with respect to the test set. In a similar approach to Ho et al. (2017), we measure the scatter σ_X in label X in bins of signal-to-noise (S/N) and fit three parameter functions of the form

$$\sigma_X = \frac{p_0}{(\text{S/N})^{p_1} + p_2^2}. \quad (16)$$

We apply our model to all stars with $\log g < 3.9$ and $T_{\text{eff}} < 6000 \text{ K}$ in the LAMOST catalogue and deem the parameters satisfactory if the reduced chi-squared < 3 and the results for T_{eff} , $\log g$, and $[\text{M}/\text{H}]$ lie within the training set. We assign uncertainties using the

S/N relation (of equation 16). The resulting sets of parameters (T_{eff} , $\log g$, $[\text{M}/\text{H}]$, $[\alpha/\text{M}]$, $[\text{C}/\text{M}]$, $[\text{N}/\text{M}]$) are used in the mass estimator of Section 2.2 to obtain masses and uncertainties.

3.3 RAVE

The RAdial Velocity Experiment (RAVE, Steinmetz et al. 2006) was designed as a predecessor to the *Gaia* RVS. $\sim 500\,000$ medium resolution ($R \sim 7500$) spectra were taken in the spectral range $\sim 8400 \text{ Å} - \sim 8800 \text{ Å}$ (around the Ca II triplet) on the 1.2 m UK Schmidt Telescope at the Australian Astronomical Observatory (AAO) using a multi-object spectrograph. With a targeting magnitude range of $9 < I < 12$, RAVE has primarily observed local disc stars. Spectroscopic parameters were extracted from the spectra using two different pipelines. The first (denoted `RAVE_DR5`) utilizes the parameters provided in the latest RAVE data release (DR5; Kunder et al. 2017) using the analysis method from Kordopatis et al. (2013), whilst the second (denoted `RAVE_Cannon`) utilizes the Cannon results for the RAVE spectra from Casey et al. (2017). We remove duplicate observations retaining the higher S/N spectra. For both data sets, we use 2MASS J , H , K_s photometry and associated uncertainties. For `RAVE_Cannon` we compute the metallicity from the reported $[\text{Fe}/\text{H}]$ combined with an inverse-variance-weighted estimate of $[\alpha/\text{Fe}]$ using the formula from Salaris, Chieffi & Straniero (1993). We also use covariances between T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ for this data set.

3.4 GES

The *Gaia*-ESO survey (GES, Gilmore et al. 2012) is a public spectroscopic survey on the Very Large Telescope (VLT) utilizing the FLAMES (Fiber Large Array Multi-Element Spectrograph) spectrograph (both medium resolution GIRAFFE $R \sim 20\,000$ spectra and high-resolution UVES $R \sim 50\,000$). One of the goals of *Gaia*-ESO is to study the systematics in spectroscopic parameters by comparison of the results from multiple *nodes* and through a combination of field and cluster fields. The field sample primarily focusses on thick disc and halo stars. We use the reported T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ (as a proxy for metallicity) with the associated errors from the DR3 public data release.² The target selection for field stars in GES was performed using VISTA photometry (Dalton et al. 2006; Emerson, McPherson & Sutherland 2006) so we preferentially use VISTA photometry J_V , H_V , K_V with associated uncertainties and default to 2MASS J , H , K_s where unavailable (we transform the 2MASS bands provided from the PARSEC isochrones to VISTA bands using the relations from <http://casu.ast.cam.ac.uk/surveys-projects/vista/technical/photometric-properties>).

3.5 GALAH

GALactic Archaeology with HERMES (the High Efficiency and Resolution Multi-Element Spectrograph; De Silva et al. 2015; Martell et al. 2017) is a medium resolution ($R \sim 28000$) optical (four windows) multifibre spectroscopic survey. It was designed to provide a rich set of chemical abundances for the purpose of chemical tagging distinct star formation events (Freeman & Bland-Hawthorn 2002) and has a simple selection function ($12 < V < 14$) targeting primarily disc stars. The second GALAH data release

¹<https://github.com/andycasey/AnnieLasso>

²<https://www.gaia-eso.eu/data-products/public-data-releases/gaia-eso-survey-data-release-3>

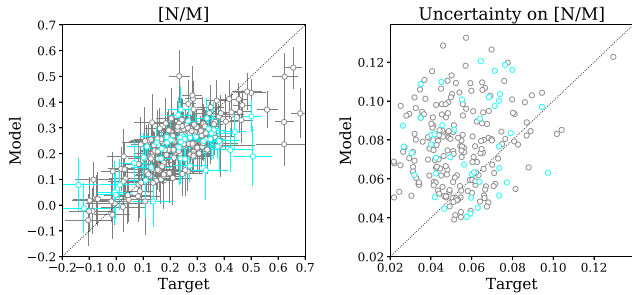


Figure 2. Recovery of nitrogen abundance (left) and its uncertainty (right) from the Bayesian neural network for 193 stars in the training sample (grey) and 44 stars in the unseen testing sample (cyan) from the overlap of APOGEE DR14 and GALAH DR2 data.

(Buder et al. 2018) contains ~340 000 stars. The spectroscopic analysis is performed in two stages with a high-quality test-set analysed in detail with line synthesis which is then used as a training set for a Cannon model (Casey et al. 2016; Ness et al. 2016). As a result, there are stars that fall outside the training set producing unreliable results (flagged by `flag_cannon`). We ignore these stars resulting in a catalogue of 264 227. We use 2MASS magnitudes J , H , K_s and compute metallicity from a combination of $[\text{Fe}/\text{H}]$ and $[\alpha/\text{Fe}]$ using Salaris et al. (1993) (if $[\alpha/\text{Fe}]$ is available, else we use $[\text{Fe}/\text{H}]$).

3.5.1 Spectroscopic mass estimates

We wish to complement the GALAH giant stars with spectroscopic mass estimates. We cannot employ the same procedure as we performed for LAMOST as the GALAH spectra are not publicly available. Instead, we build a model to put the reported spectroscopic parameters from GALAH on to the APOGEE scale. GALAH DR2 has measured $[\text{C}/\text{M}]$ from the spectra (Buder et al. 2018), but not $[\text{N}/\text{M}]$. Without $[\text{N}/\text{M}]$, the spectroscopic mass estimator is of limited power as the strongest correlation is between $[\text{C}/\text{N}]$ and mass. However, we can estimate the missing $[\text{N}/\text{M}]$ measurement from the other spectroscopic measurements. In particular, the conservation of total CNO in the CNO cycle produces a simple relationship between $[\text{C}/\text{M}]$, $[\text{O}/\text{M}]$ and $[\text{N}/\text{M}]$.

We take the 237 stars that lie in the overlap between APOGEE and GALAH (`ASPCAPFLAG! = 2**23` and `flag_cannon = 0`) and build a Bayesian neural network (with one hidden layer and 20 nodes) to relate the GALAH (T_{eff} , $\log g$, $[\text{M}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{O}/\text{Fe}]$) to APOGEE (T_{eff} , $\log g$, $[\text{M}/\text{H}]$, $[\text{C}/\text{M}]$, $[\text{N}/\text{M}]$, $[\text{O}/\text{M}]$). The procedure is similar to that described in Section 2.2 except we use NUTS to find the posterior distributions of the neural network parameters on a training subsample of 193 stars, and use the remaining 44 test stars to assess the performance of the neural network.

Fig. 2 shows how well the Bayesian neural network is able to predict the APOGEE DR14 nitrogen abundance from the GALAH DR2 spectral parameters for the training and testing samples. The model only starts failing for $[\text{N}/\text{M}] > 0.5$, beyond which the model underestimates the APOGEE DR14 nitrogen abundance. Nitrogen has a negative correlation with oxygen and carbon; however the five stars with $[\text{N}/\text{M}] > 0.5$ are found over the whole range of oxygen and carbon abundances. There may be a genuine physical reason underpinning this, but it's difficult to look beyond the Poisson noise obviously affecting the estimates for higher nitrogen abundances.

3.6 SEGUE

The Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009) is a low-resolution ($R \sim 2000$) optical spectroscopic survey designed to complement the SDSS catalogues with radial velocities. The survey was conducted in two key stages (SEGUE-1 and SEGUE-2) where SEGUE-2 focussed primarily on the outer halo of the Galaxy. Spectroscopic parameters have been computed by the SEGUE Stellar Parameter Pipeline (SSPP; Allende Prieto et al. 2008; Lee et al. 2008a,b; Smolinski et al. 2011). We adopt the external error estimates from Lee et al. (2008a) of 141 K, 0.23 dex, and 0.23 dex in T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$, respectively, which we add in quadrature to the reported uncertainties. We take all SDSS DR12 stars observed in the programmes SEGUE, SEGUE-2, or SEGUE-faint with valid T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$, that are science primary, have `zwarning = 0` or 16 and are flagged as normal ('nnnnn'), resulting in a catalogue of 187 152 stars). We use $[\text{Fe}/\text{H}]$ as a proxy for metallicity. We complement with the g , r , i photometry from SDSS.

3.7 Complementary Gaia data

For each catalogue, we perform a 5 arcsec radius cross-match to the second *Gaia* data release (Gaia Collaboration 2018a) by utilizing the *Gaia* proper motions and accounting for the respective epochs of the surveys (we assume the epoch of all the spectroscopic catalogue observations is 2000). From the *Gaia* DR2 source catalogue, we extract the parallax, proper motion, and the uncertainty covariance matrix for the astrometry. Despite the reported global zero-point parallax offset of $\sim 30 \mu\text{as}$ (Gaia Collaboration 2018d; Lindegren et al. 2018), we use the reported parallax and uncertainty as is. For studying large-scale Galactic structure, the zero-point is only important for more distant stars where other uncertainties (e.g. in the choice of prior) are also significant. Additionally, the zero-point is also a strong function of on-sky location, magnitude, and colour (Gaia Collaboration 2018d; Lindegren et al. 2018; Riess et al. 2018) so a global offset will only fix some systematic issues. For the *Gaia* G photometry we assume a 0.02 mag systematic uncertainty floor in the photometry that reflects systematics in the photometry (Riello et al. 2018) as well as intrinsic uncertainty in the isochrones. We use the parallax and G photometry as additional inputs in our pipeline. When the cross-match fails, we still process the stars without any *Gaia* data.

3.8 Output catalogue description

We provide a catalogue of the combined results from all surveys complete with the *Gaia* DR2 `source_id` for each entry (where available).³ We provide a full description of the columns (including their units) in Table B1. For each survey, we include a unique identifier (APOGEE: `APOGEE_ID`, LAMOST: `obsid`, RAVE: `raveid`, GES: `CNAME`, GALAH: `subject_id`, SEGUE: `specobjid`) and we provide a field `survey` with a string detailing which survey the entry comes from. Note that we have not removed duplicate stars that were observed by two separate surveys (e.g. RAVE and RAVE-On). For each star we include the cross-matched *Gaia* DR2 `source_id`, the angular separation of the cross-match and the photometric bands used in the pipeline. The table is saved as

³The full catalogue is available at https://www.ast.cam.ac.uk/jls/data/gaia_spectro.hdf5

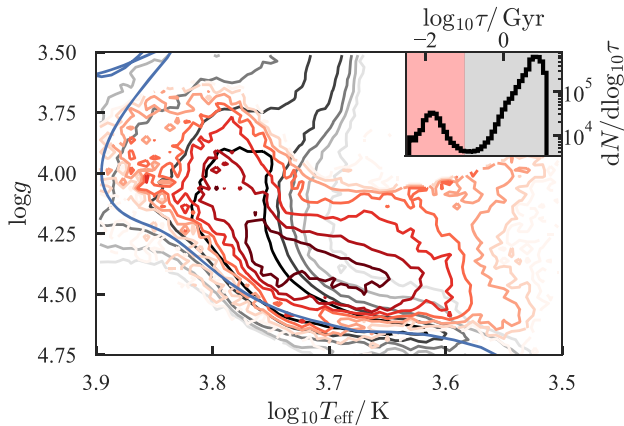


Figure 3. Spectroscopic HR diagram of stars assigned a young age by our pipeline: black contours show the distribution of stars with $\tau > 100$ Myr and red $\tau < 100$ Myr which are suspected binaries. The inset shows the distribution of $\log_{10} \tau$ shaded by the separation. The blue line shows a 1 Gyr solar metallicity isochrone.

an astropy table (The Astropy Collaboration 2018) in hdf format, and can be simply read with `from astropy.table import Table; data=Table.read('file')`. A description of the column and structure of the table is available in `data.meta['COMMENT']`.

We provide a `flag` entry in the table. If the isochrone pipeline has failed (e.g. there is no overlap between isochrones and data within 20 times the uncertainties due to a bad cross-match), `flag=1`. This flag is also non-zero if there is a problem with the input spectroscopy (`flag=2`), photometry (3), astrometry (4), or mass (5) (no entries in the catalogue have `flag=4,5`). A small fraction (~ 2 per cent) of the processed stars appeared to have overlap with only a single isochrone (point) leading to zero or undefined second moments (uncertainties). We flag these stars with `flag = 6`.

We found a non-negligible fraction of processed red stars had small ages. In Fig. 3, we show the 1D distribution of $\log_{10} \tau$ which exhibits a clear peak at -2 . We also display the $T_{\text{eff}} - \log g$ distribution of these stars which reveals that the pipeline has assigned these as pre-main-sequence objects. The assigned T_{eff} and $\log g$ for these stars form a clear sequence running alongside the main sequence, such that it is highly likely that many of these stars are binary. This is also evident in colour–magnitude space. We flag the stars for which $\tau < 100$ Myr and $\log g > 3.5$ and $\log_{10} T_{\text{eff}} < 3.9$ with `flag = 7`. For more accurate parameters, these stars require a special pipeline such as that presented in Coronado et al. (2018).

Some surveys contain duplicate observations of the same stars and some surveys have observed the same stars more than once. We identify duplicates using the *Gaia* source IDs and preferentially keep the entries flagged as `flag = 0` from different surveys in the order: APOGEE, GALAH, GES, RAVE-ON, RAVE, LAMOST, and SEGUE, and for duplicates within surveys we keep the result with the smallest vertical velocity uncertainty [to retain both the most precise distance estimates (from the spectroscopy) and the most precise radial velocities]. We provide a duplicated flag which is 1 for the rejected duplicates and zero otherwise. Finally, we provide a best flag which is 1 if `flag=0`, `duplicated=0`, and the star has a *Gaia* match. We provide statistics for all processed stars in Table 1. Our ‘best’ sample consists of ~ 3 million stars. It should be noted that additional quality cuts may be necessary for

certain analyses to remove stars that have been flagged as unusual (for instance, in RAVE stars with $c^*! = 'n'$).

Caveats: Our provided catalogue has a number of caveats and features that we should highlight. First, we note that when inspecting and analysing this catalogue one should be aware of our choice of Galaxy prior. This choice imposes some level of structure on the results. When attempting to fit the models to the provided data set, this must be accounted for by ‘dividing out’ our choice of prior. We are using carbon and nitrogen abundances to inform giant age estimates which requires good knowledge of the initial carbon and nitrogen abundance of each star before any processing and dredge-up. Although we have tried to only apply this procedure to stars within the convex hull of our training set, it may be inappropriate for more extreme populations, e.g. accreted halo stars. When inspecting the catalogue in age and metallicity, the imprint of the isochrone gridding is visible with a weak preference for stars to bunch at the isochrone points. Some of these stars will have very small uncertainty in age and metallicity and so in practice we recommend a minimum age uncertainty be used. The minimum spacing in metallicity and log-age of the isochrones is 0.01 dex and 0.05 dex, respectively, although from inspecting the age uncertainties discreteness effects are only visible when $\sigma_{\log_{10} \tau} \lesssim 0.015$ dex. Finally, our choice of prior restricts distant and metal-poor stars to be old and no star can be older than our final isochrone point. This biases the results for large ages and there is a correlation between uncertainty and age for these old stars.

Our catalogue could be further improved with additional data and modelling improvements. For instance, we have not explicitly included $[\alpha/M]$ abundance which is now regularly provided by large spectroscopic surveys. Furthermore, with access to the stellar spectra (from for instance RAVE and GALAH), we could improve age estimates from the giant stars (using $[N/M]$ or CN bands in the case of RAVE). Finally, the hard truncation in age and our adopted prior produces a number of undesirable features which could be improved through further refinements.

4 RESULTS

Our catalogue represents the largest, most homogeneous catalogue of distances, ages, masses, and spectroscopic parameters available. In this section, we demonstrate the properties of the catalogue, give a number of checks of its quality, and highlight its possible power in studies of the dynamical structure of the Galaxy.

Our catalogue has three key advantages: (1) we constrain the uncertain distances from *Gaia* parallaxes alone for distant stars using spectroscopy and photometry, (2) the extinction estimates use spectroscopy to better pin down the intrinsic spectral type, and (3) we provide age estimates, crucial for Galactic archaeology. There are two key subsamples of stars that have precise ages from our pipeline. First, the combination of parallaxes and spectroscopic metallicities break the metallicity–age degeneracy for turn-off stars (e.g. Howes et al. 2018). Secondly, for giant stars accurate ages are possible due to the employed spectroscopic mass estimates combined with *Gaia* parallaxes which further constrain the luminosity and hence age. We conservatively define these two subsamples as 1. giants: $\log g < 3$ dex and $\log_{10}(T_{\text{eff}}/K) < 3.73$, and 2. turn-off: $3.6 < \log g < 4.5$ dex and $\log_{10}(T_{\text{eff}}/K) < 4.1$.

4.1 Output uncertainties

In Fig. 4, we show the output uncertainties from our pipeline for all stars, the giant stars and the turn-off stars. We opt not to show

Table 1. Numbers of stars processed by pipeline. *Total* gives the total number of stars in the survey, *Gaia matches* gives the number of matches in *Gaia* DR2, and *Success* gives the number of stars with `flag=0`. The middle section shows the number of stars with pipeline failure (along with the allocated flag in brackets). *Success with Gaia* gives the number of stars with `flag=0` with *Gaia* DR2 matches. In the *All no dupl.* column we only consider *Gaia* matches and have removed duplicate *Gaia* entries.

Survey	RAVEDR5	RAVEON	GALAH	APOGEE	LAMOST	GES	SEGUE	All	All no dupl.
Total	457 555	457 555	342 682	258 475	3177 995	25 332	187 152	4906 746	3706 733
<i>Gaia</i> matches	456 353	456 353	342 212	256 851	3168 545	25 313	187 100	4892 727	3706 733
Success (0)	415 200	376 316	260 233	203 417	2861 310	10 882	182 132	4309 490	3318 119
Pipeline failed (1)	1358	368	67	2778	15 503	105	393	20 572	16 928
Spec. problem (2)	20 602	67 439	78 455	42 553	67 524	12 513	0	289 086	185 066
Phot. problem (3)	498	498	0	810	7660	1181	1490	12137	4667
Unreliable errors (6)	3220	1093	46	460	7948	52	72	12 891	6444
Low age (7)	16 677	11841	3881	8457	207 577	599	3065	252 097	166 986
Success with <i>Gaia</i>	414 238	375 488	259 877	203 127	2858 287	10 881	182 087	4303 985	3318 119

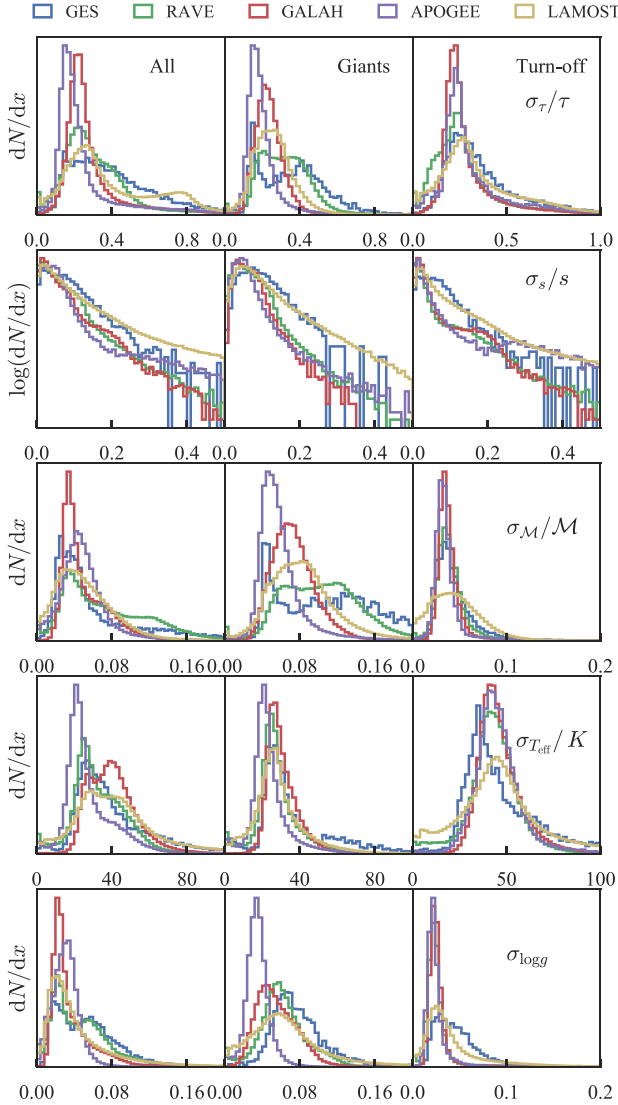


Figure 4. Distribution of output uncertainties from our pipeline: each row corresponds to a different quantity labelled in the right plots, and each column to a different sample (defined in the text) labelled in the top panels.

the results for SEGUE as the majority of these stars are distant and metal-poor so the uncertainties (particularly in age) strongly reflect the prior. We see that for giant stars we obtain uncertainties in age of ~ 16 per cent for APOGEE (which has the most accurate spectroscopic mass estimates), ~ 21 per cent for GALAH (for which we have inferred $[N/M]$ from other abundances before computing a spectroscopic mass estimate), ~ 25 per cent for LAMOST (which also has spectroscopic mass estimates), and ~ 40 per cent for RAVE and GES (for which no spectroscopic mass estimates are used). However, there are a number of metal-poor distant giants in these surveys that have ~ 20 per cent mass estimates due partly to the age prior employed. The mass uncertainties for the giant stars are $\sim 1/3$ the age uncertainties. For the turn-off stars, all surveys yield ages accurate to ~ 25 per cent with GES and LAMOST producing a slightly fatter tail to large uncertainties due to larger metallicity errors.

The distribution of distance errors peaks very close to zero due to the accuracy of the *Gaia* parallaxes. For the giant stars (which are at larger distance) the peak moves to higher σ_s/s . The tails of the relative distance uncertainty distributions are a combination of the different survey selection functions and the quality of the input parameters. We see the largest errors arise from LAMOST and GES (which observe the faintest stars).

Our pipeline takes as inputs the spectroscopic parameters, T_{eff} and $\log g$, but also provides these as outputs. In the lower panels of Fig. 4, we show the output uncertainties in these parameters. For all surveys, we find uncertainties in T_{eff} of 20–30 K and ~ 50 K for giants and turn-off stars, respectively. The effective temperature is strongly constrained with photometry, hence the independence with survey. However, the degree to which the output uncertainties in effective temperature can be trusted is unclear as they are dependent on systematics in the bandpasses and the assumed extinction law. For $\log g$, the uncertainties are survey dependent – we see the two surveys with spectroscopic mass estimates for giants (APOGEE and LAMOST) peak at smaller $\log g$ uncertainties (~ 0.04 , 0.06 dex, respectively) than the other surveys (~ 0.08 dex). For the turn-off stars, the difference in $\log g$ accuracy is due to the selection functions of the surveys, where LAMOST and GES target more distant stars with less accurate distances from *Gaia* and hence less accurate $\log g$.

The output T_{eff} and $\log g$ from our method can be compared with the inputs determined entirely spectroscopically. We show a comparison of the approaches in Fig. 5. We clearly see the power of enforcing an isochrone prior on the spectroscopic parameters. In particular, the main sequences are very tight due to the quality of

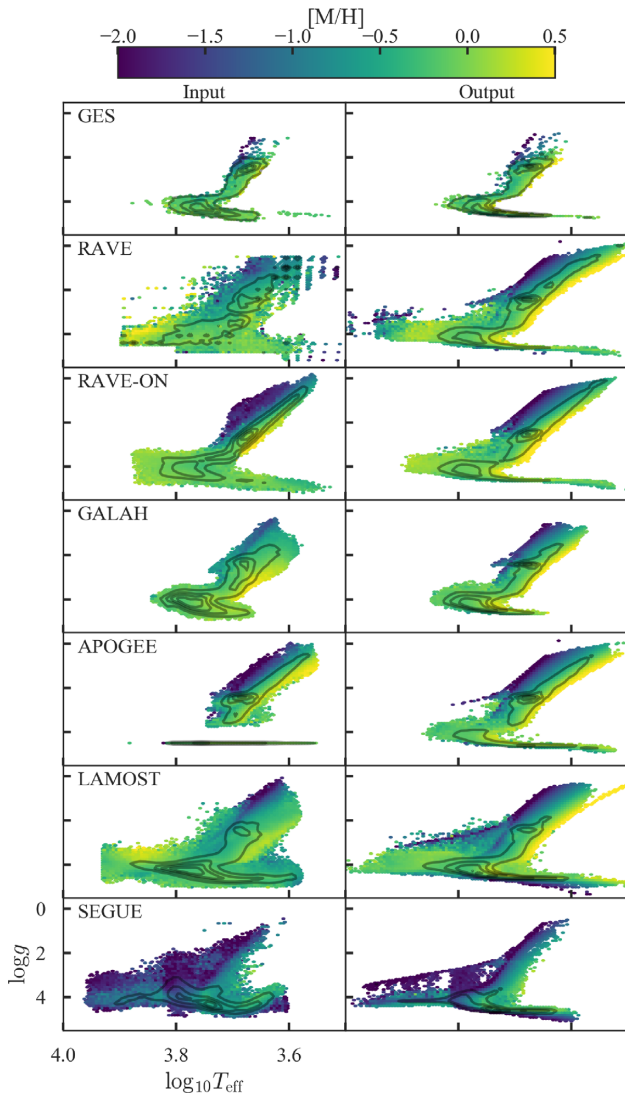


Figure 5. Input and output spectroscopic parameters (T_{eff} against $\log g$ coloured by $[M/H]$). The left column shows the spectroscopic parameters used as inputs in the pipeline and the right column the parameters output by our pipeline. Each row corresponds to a different survey labelled in the left plot. The colours correspond to the median metallicity (provided there are at least two stars in a bin) and the contours are spaced equally in the square-root of the number of stars in each bin.

the *Gaia* parallaxes. We also note that our procedure has introduced some discreteness to the diagrams due to our chosen isochrone spacing. However, each parameter estimate also has an associated error that smooths over this discreteness.

4.2 Extinction maps

The combination of photometric, spectroscopic, and astrometric data is powerful for mapping the 3D dust extinction throughout the Galaxy. Spectroscopy allows us to identify intrinsically similar stars, photometry gives a measure of the differential reddening between these stars, and astrometry tells us how the reddening varies along the line-of-sight. A_V measurements for all stars are an output of our pipeline. Our measurements are informed by the adopted extinction prior but stellar colours and effective temperatures provide additional information producing stronger constraints on the

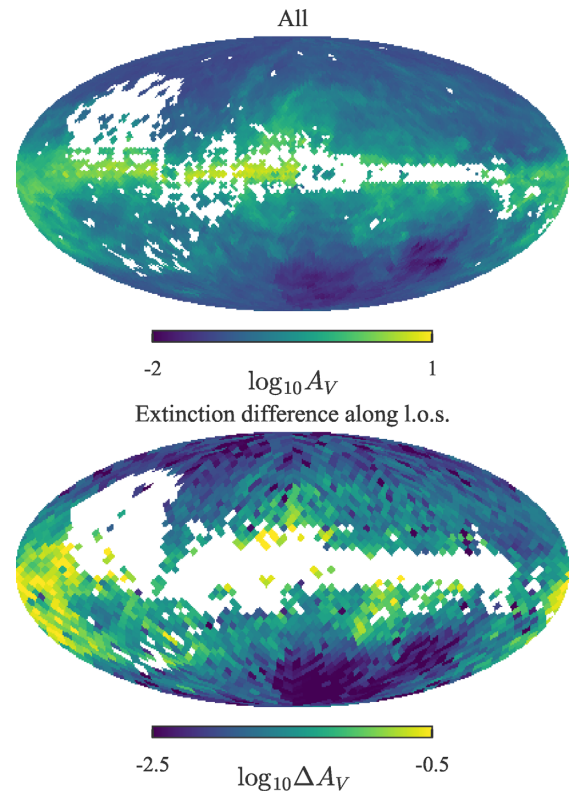


Figure 6. Extinction maps in galactic coordinates for V-band extinction, A_V . The top panel shows all stars with $\text{best}=1$, and the bottom panel shows the logarithm of the difference in median A_V in each bin between a slice with $0.15 < s/\text{kpc} < 0.25$ and a slice with $0.25 < s/\text{kpc} < 0.35$.

extinction than provided by the prior. In Fig. 6, we show on-sky A_V for all stars and the logarithm of the difference in median extinction between two distance slices of $0.15 < s/\text{kpc} < 0.25$ and $0.25 < s/\text{kpc} < 0.35$ for the ‘best’ data set. We see the expected large-scale dust structure and the extinction increasing as a function of distance. The combination of both distances and extinctions allow full 3D extinction maps of the Galaxy to be built.

In Fig. 7, we compare our A_V measurements with A_G provided in the *Gaia* DR2 source catalogue, calculated using a combination of the *Gaia* photometry and parallaxes. We see that above $A_G \approx 0.3$ there is a clear linear relationship between our output A_V and A_G with approximate gradient 0.77 in agreement with the expectation from Table A1. Below $A_V \approx 0.3$, A_G is poorly constrained, possibly due to limitations in the methodology (Andrae et al. 2018).

4.3 Age maps

In Fig. 8, we present all-sky age maps for the giant and turn-off subsamples. The entire giant subsample exhibits the expected gradients with galactic coordinates with younger stars confined primarily to the disc plane and higher latitudes dominated by older stars. As we slice through in distance we see the increasing contribution of old stars at high latitude (due in part to our age prior).

The turn-off subsample exhibits a less clean picture of on-sky age distribution and for the entire sample some of the observed features are produced by the different selection functions of the surveys. In particular, the southern sky surveys (GALAH and RAVE) observe brighter stars than the primary northern sky survey (LAMOST). If we restrict ourselves to observing the nearby bright turn-off stars

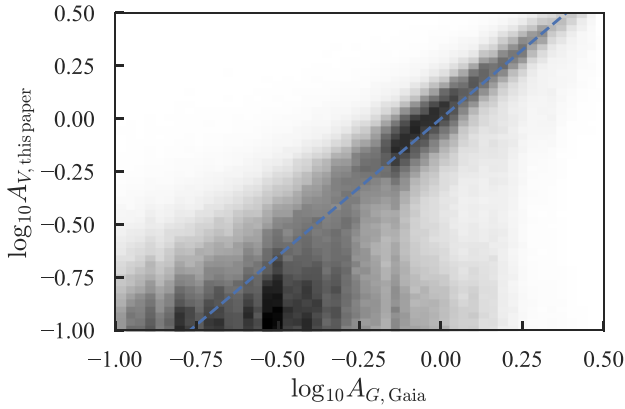


Figure 7. Comparison of extinction estimates A_V to *Gaia* DR2 catalogue results (2d histogram is log-scaled). The line has gradient 0.77 equal to the anticipated value for red stars from Table A1.

($0.2 < s/\text{kpc} < 0.3$) we find that there is essentially no on-sky age gradients. Over these distance scales, all age populations are contributing so we end up observing the average age of the solar neighbourhood. For a more distant bin ($0.8 < s/\text{kpc} < 1$) we observe the anticipated age gradient with latitude and the southern sky surveys smoothly match into the northern sky. For an even more distant bin ($1.8 < s/\text{kpc} < 2$), we lose southern sky coverage in the turn-off sample and are dominated by LAMOST which shows very strong age gradients on these large Galactic scales.

4.4 Age–kinematic relations

As a demonstration of the power of our catalogue, we briefly investigate some age–kinematic correlations in the disc. Studies of disc populations subdivided by age (or $[\alpha/\text{M}]$) (e.g. Bovy et al. 2016; Mackereth et al. 2017) have tended to be restricted to spatial–chemical correlations due to the limitations of pre-*Gaia* proper motions. Our catalogue opens up the possibility of inspecting not just spatial–chemical correlations but full chemo–kinematic correlations. We reserve a full analysis to a future project and instead here give a flavour of what is possible.

We compute the velocity dispersion of populations separated into radial and age bins. When computing the velocity dispersion in each bin, we account for the uncertainty in each velocity measurement σ_{v_i} in the following way. We first perform a 3σ clip of the raw velocities v_i to remove outliers and halo contaminants. We then seek to maximize the log-likelihood

$$\ln \mathcal{L} = -\frac{1}{2} \sum_i \left(\frac{(v_i - m)^2}{\sigma^2 + \sigma_{v_i}^2} + \ln(\sigma^2 + \sigma_{v_i}^2) \right), \quad (17)$$

which can be found by solving

$$0 = \sum_i \left(\frac{(v_i - m)^2}{(\sigma^2 + \sigma_{v_i}^2)^2} - \frac{1}{\sigma^2 + \sigma_{v_i}^2} \right). \quad (18)$$

Here, m is the simple mean computed without accounting for the uncertainties. We find the value of σ that solves this equation by Brent’s method. The uncertainty in the resulting estimate can be found from the second derivative of the log-likelihood and is approximately $\sigma/\sqrt{2N}$ for N stars. We consider at least 50 stars per bin resulting in dispersion uncertainties better than 10 per cent.

In Fig. 9, we give the Galactocentric radial dispersion profile of the radial and vertical velocity separated into different age bins. We use all ‘best’ giant and turn-off stars with $|z| < 0.6$ kpc,

$[\text{M}/\text{H}] > -1$ dex and $\sigma_\tau/\tau < 0.45$ (approximately 1.2 million stars). Inwards of the solar radius, both velocity dispersions decline approximately exponentially with scale radii of ~ 5 kpc. Near and beyond the solar radius, the velocity dispersion profiles flatten with radius. The radial velocity dispersion continues to decline, whilst the vertical dispersion plateaus and begins to weakly increase. This intriguing signature could be linked to the theoretical claims that mono-age populations show significant flaring (Minchev et al. 2015), although it is also possibly due to the selection effect of seeing higher latitude stars at larger Galactocentric radii. A similar flattening of the vertical velocity dispersion profile is seen for the *Gaia* RVS sample studied by *Gaia* Collaboration (2018c). At all radii, we clearly see a gradient in age with lower age populations on kinematically colder orbits. The break in the behaviour of the velocity dispersion profiles occurs at decreasing radius with increasing age. A flattening of velocity dispersion can be caused by distance systematics which for the sample we consider more strongly affect the vertical dispersion. A reasonable parallax systematic of ~ 0.03 mas (Lindegren et al. 2018) changes the dispersion by $\sim 3 \text{ km s}^{-1}$ 7 kpc away and so is unlikely to fully explain the effect seen in Fig. 9. Additionally, the spectroscopy and photometry used in our distance measurements goes some way to correcting any *Gaia* parallax systematic issues at larger distances.

The behaviour of the velocity dispersions with age can be observed more clearly in Fig. 10 where we show the dispersion profiles with age split by Galactocentric radius. At all radii, the velocity dispersions grow approximately like a power-law with age, τ^β . We find coefficients of $\beta \sim 0.3$ for the radial dispersion and $\beta \sim 0.4$ for the vertical dispersion although this is not accounting for the (significant) age uncertainties which act to flatten the relations (Martig, Minchev & Flynn 2014; Aumer, Binney & Schönrich 2016b). However, these values are consistent with observations of the solar neighbourhood (Aumer & Binney 2009) and indicate the velocity dispersions throughout the Galaxy are consistent with heating by a combination of spiral arms and molecular clouds (Aumer, Binney & Schönrich 2016a). With our particular set of cuts (in particular, only considering low Galactic height stars), we find no real indication of a break in the dispersion profile at intermediate/large age. Inside the solar radius, there is the suggestion of the profiles flattening with age (although this is possibly due to increased age uncertainties). We note that our observations must all be considered within the context of the selection function of the sample. A future work will account for these effects in a full model of our presented catalogue, enabling us to disentangle observational biases from features of the Galaxy.

In our output catalogue, we also provide estimates of the actions (with associated errors). In Fig. 11 we show the action distributions (estimated using a kernel density estimate) split into six age bins using all ‘best’ giant and turn-off stars with $|z| < 2.5$ kpc, $[\text{M}/\text{H}] > -1$ dex, and $\sigma_\tau/\tau < 0.45$. Mirroring the results from Fig. 9, we see the steadily increasing mean radial and vertical action with increasing age. We observe that the oldest age bin in vertical action is significantly skewed indicating the presence of a thick disc component. This is not mirrored in the radial action distribution. The picture from the radial action distributions alone is of a smooth quiescent evolution of the disc. The z -component of angular momentum smoothly declines with age reflecting both the asymmetric drift in the populations and the spatial distribution due to inside-out growth. The angular momentum distribution also has more evidence of the selection function of the catalogue (as z -component of angular momentum is a proxy for radius for dynamically cool stars).

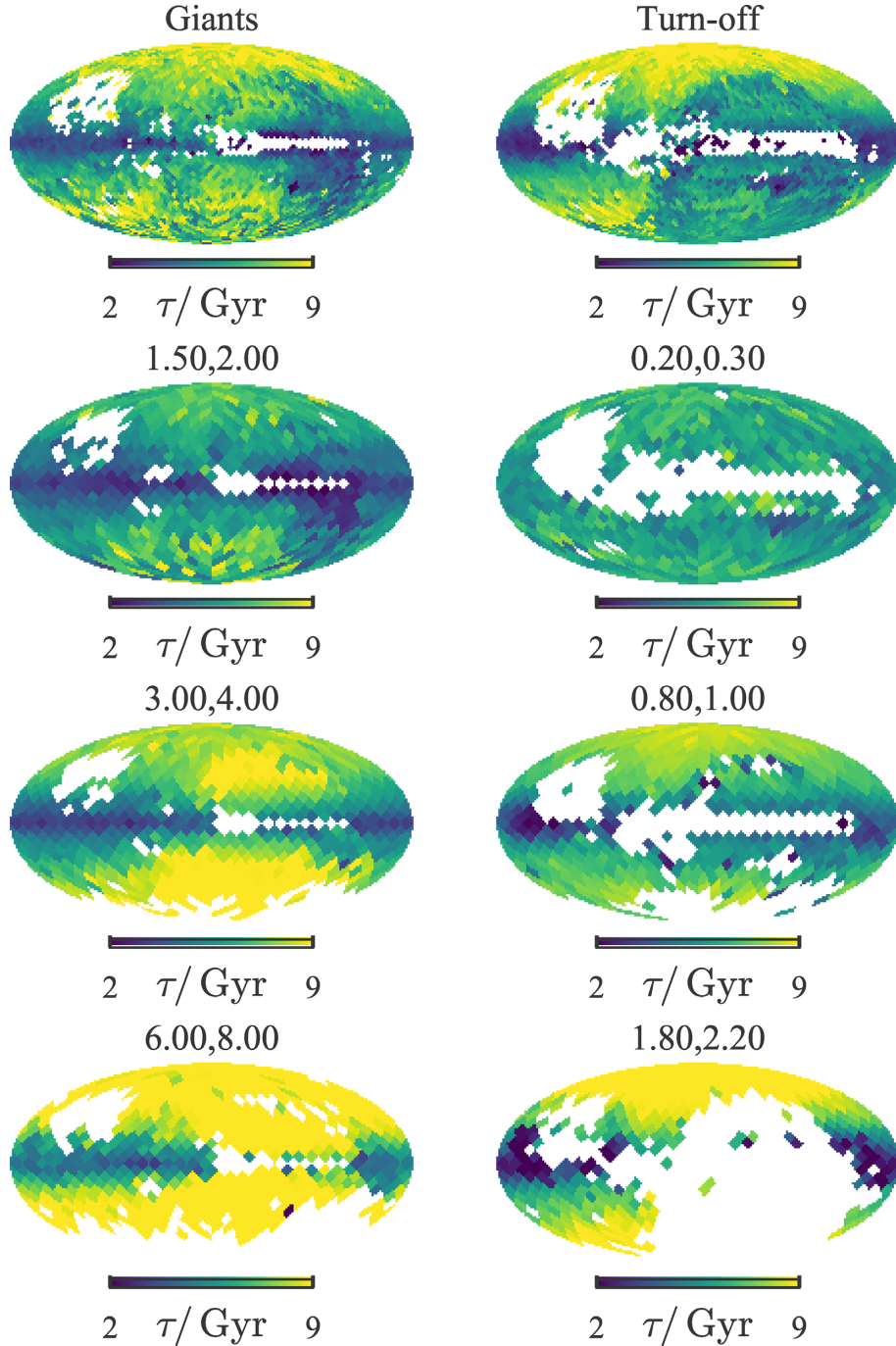


Figure 8. All-sky age maps in galactic coordinates: left-hand panels show giant stars and right-hand panels turn-off stars. Each row corresponds to different distance brackets (shown in kpc) above the plots with the top panels showing all stars. The different survey selection functions are clearly imprinted on the top right image.

5 CONCLUSIONS

We have presented a catalogue⁴ of approximately 3 million ages, masses, distances, extinctions, and spectroscopic parameters for stars in common with large spectroscopic surveys and the second

Gaia data release. We considered the results from APOGEE, LAMOST, SEGUE, *Gaia*-ESO, RAVE (using both the RAVE DR5 results and RAVE-On) and GALAH giving approximate all-sky coverage. We have complemented our catalogue with estimates of Galactocentric coordinates and actions along with associated (1D) uncertainties. As well as presenting details of our procedure, we have focussed on the quality of the output catalogue and presented some preliminary results demonstrating its power. Our conclusions are as follows:

⁴The catalogue is available from https://www.ast.cam.ac.uk/jls/data/gaia_spectro.hdf5 and the corresponding code at https://github.com/jls713/gaia_dr2_spectro. The format of the catalogue is given in Table B1.

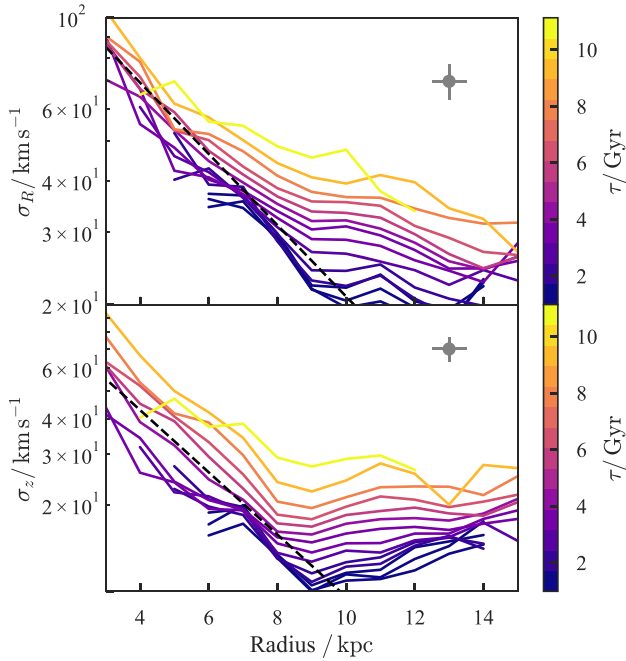


Figure 9. Velocity dispersions against Galactocentric radius for a series of age bins (for giant and turn-off stars with $|z| < 0.6$ kpc and $[M/H] > -1$ dex). Each line is coloured by the mean age of the bin. The top panel shows the radial velocity dispersion and bottom panel the vertical dispersion. The dashed black lines are to guide the eye and correspond to exponentials with scale radii of 5 kpc and 4 kpc, respectively. The y errorbar show the maximum error for each datapoint (the x errorbar is the bin width).

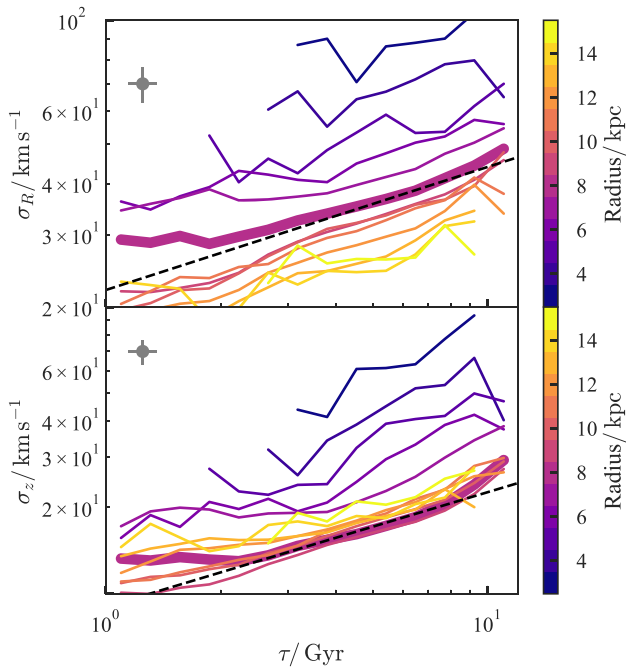


Figure 10. Velocity dispersions against age for a series of radial bins (for giant and turn-off stars with $|z| < 0.6$ kpc and $[M/H] > -1$ dex). Each line is coloured by the mean Galactocentric radius of the bin. The top panel shows the radial velocity dispersion and bottom panel the vertical dispersion. The thick line corresponds to bin centred around the solar neighbourhood. The dashed black lines are to guide the eye and correspond to power laws τ^β with coefficients $\beta = 0.3$ and $\beta = 0.4$, respectively. The y errorbar shows the maximum error for each datapoint (the x errorbar is the bin width).

(i) A non-negligible fraction of our catalogue is assigned pre-main sequence properties from our pipeline. Many of these stars lie along the binary sequence in colour-magnitude space so are suspected binary stars. Our catalogue can be used to assign binarity to the stars observed by the considered spectroscopic surveys.

(ii) We have investigated the output uncertainties produced by our catalogue. Two key subsamples for studying the age structure of the Galaxy are giant stars and turn-off stars. For the APOGEE, LAMOST, and GALAH subsamples, we have employed techniques to assign spectroscopic mass estimates to the giant stars (via carbon and nitrogen abundances). This results in ages accurate to 15 – 25 per cent. For turn-off stars, parallax is insufficient to determine an accurate age due to a degeneracy with metallicity. Employing spectroscopic metallicity measurements results in output uncertainties of $\sim 20 - 30$ per cent. We also have provided output spectroscopic parameters $\log g$ and T_{eff} which typically have uncertainties less than 0.1 dex and 80 K, respectively. These parameters can be used as initial guesses in improved analysis of the spectra.

(iii) We have provided output extinction estimates for all stars in our catalogue. These estimates are informed by prior extinction maps but the output uncertainties are typically smaller than the prior input uncertainties. When complemented with our output distances, the extinction values can be used to construct a 3D extinction map (over the volume probed by our spectroscopic samples).

(iv) We presented some first results on the correlations between kinematics and age for Milky Way stars using our new catalogue. We demonstrated that the Galactocentric radial profile of radial and vertical velocity dispersions appears to flatten beyond the solar radius. At all radii, we see a smooth power law increase of the (radial and vertical) velocity dispersions with age.

The presentation of our catalogue represents a first step in analysing the chemo-dynamic composition of the Galaxy with the *Gaia* data. We have provided an indication of the power of the catalogue in analysing the correlations between age and kinematics throughout the Galaxy. Although our catalogue is a combination of multiple surveys, the selection function of the total catalogue can be found, as most of the constituent catalogues have well defined selection functions. A future contribution will model the published catalogue.

ACKNOWLEDGEMENTS

JLS acknowledges the support of the Science and Technology Facilities Council (STFC). PD would like to acknowledge support from the STFC (ST/N000919/1). We thank Gyuchul Myeong for checking an early version of the catalogue, and Douglas Boubert and Eugene Vasiliev for useful conversations. We acknowledge the use of Sergey Koposov’s Whole Sky Database (WSDb), Andy Casey’s Cannon code, and Morgan Fouesneau’s *ezpadova* tool. We thank Andy Casey for providing the correlations between spectroscopic parameters for the RAVE-On catalogue. This research made use of Astropy, a community-developed core Python package for Astronomy (The Astropy Collaboration et al. 2018) and packages in the *scipy* ecosystem (Jones et al. 2001; Hunter 2007; Pérez & Granger 2007; McKinney 2010).

This project was developed in part at the 2016 NYC *Gaia* Sprint, hosted by the Center for Computational Astrophysics at the Simons Foundation in New York City.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/web/gaia/iow_2_0180316), processed by the *Gaia* Data Processing and Analy-

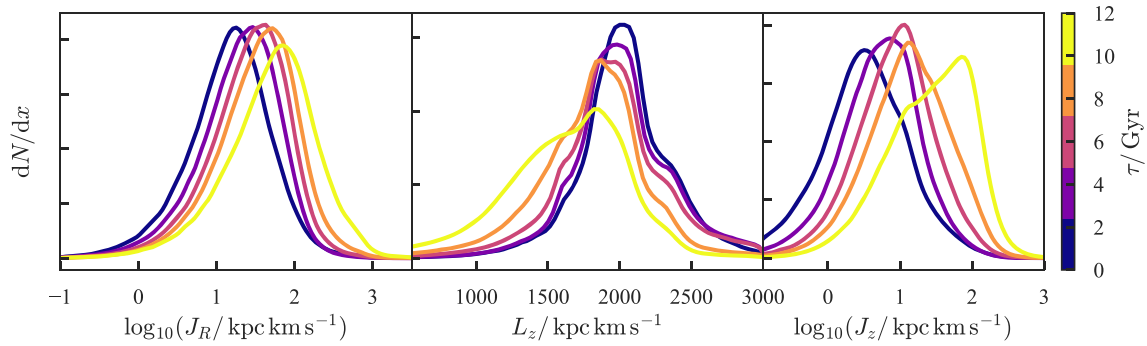


Figure 11. Normalized kernel density estimates for the action distributions in six evenly spaced age bins from 1 to 12 Gyr (for giant and turn-off stars with $|z| < 2.5$ kpc and $[M/H] > -1$ dex) – left-hand panel shows logarithm of the radial action, central panel the z -component of the angular momentum, and right-hand panel the logarithm of the vertical action.

sis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

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

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

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, 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.

Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

Funding for RAVE has been provided by: the Australian Astrophysical Observatory; the Leibniz-Institut fuer Astrophysik Pots-

dam (AIP); the Australian National University; the Australian Research Council; the French National Research Agency; the German Research Foundation (SPP 1177 and SFB 881); the European Research Council (ERC-StG 240271 Galactica); the Istituto Nazionale di Astrofisica at Padova; The Johns Hopkins University; the National Science Foundation of the USA (AST-0908326); the W. M. Keck foundation; the Macquarie University; the Netherlands Research School for Astronomy; the Natural Sciences and Engineering Research Council of Canada; the Slovenian Research Agency; the Swiss National Science Foundation; the Science & Technology Facilities Council of the UK; Opticon; Strasbourg Observatory; and the Universities of Groningen, Heidelberg and Sydney. The RAVE web site is at <https://www.rave-survey.org>.

REFERENCES

- Abolfathi B. et al., 2018, *ApJS*, 235, 42
- Allende Prieto C. et al., 2008, *AJ*, 136, 2070
- Andrae R. et al., 2018, *A&A*, 616, A8
- Anguiano B. et al., 2018, *A&A*, preprint ([arXiv:1807.07625](https://arxiv.org/abs/1807.07625))
- Aumer M., Binney J., Schönrich R., 2016a, *MNRAS*, 459, 3326
- Aumer M., Binney J., Schönrich R., 2016b, *MNRAS*, 462, 1697
- Aumer M., Binney J. J., 2009, *MNRAS*, 397, 1286
- Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, *AJ*, 156, 58
- Bensby T., Feltzing S., Oey M. S., 2014, *A&A*, 562, A71
- Binney J., 2012, *MNRAS*, 426, 1324
- Binney J. et al., 2014, *MNRAS*, 437, 351
- Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
- Blanton M. R. et al., 2017, *AJ*, 154, 28
- Bovy J., 2017, *MNRAS*, 470, 1360
- Bovy J., Rix H.-W., Green G. M., Schlafly E. F., Finkbeiner D. P., 2016, *ApJ*, 818, 130
- Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, 427, 127
- Buder S. et al., 2018, *MNRAS*, 478, 4513
- Burnett B., Binney J., 2010, *MNRAS*, 407, 339
- Casey A. R., Hogg D. W., Ness M., Rix H.-W., Ho A. Q., Gilmore G., 2016, *ApJ*, preprint ([arXiv:1603.03040](https://arxiv.org/abs/1603.03040))
- Casey A. R. et al., 2017, *ApJ*, 840, 59
- Castelli F., Kurucz R. L., 2004, preprint ([arXiv:astro-ph/0405087](https://arxiv.org/abs/astro-ph/0405087))
- Charbonnel C., 1994, *A&A*, 282, 811
- Chen Y., Bressan A., Girardi L., Marigo P., Kong X., Lanza A., 2015, *MNRAS*, 452, 1068
- Chen Y., Girardi L., Bressan A., Marigo P., Barbieri M., Kong X., 2014, *MNRAS*, 444, 2525

- Cirasuolo M. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, *Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV*, SPIE, Bellingham, p. 84460S
- Coronado J., Rix H.-W., Trick W. H., 2018, *MNRAS*, 481, 2970
- Cropper M. et al., 2018, preprint ([arXiv:1804.09369](https://arxiv.org/abs/1804.09369))
- Cui X.-Q. et al., 2012, *Res. Astron. Astrophys.*, 12, 1197
- Dalton G. B. et al., 2006, in McLean I. S., Iye M., eds, *Proc. SPIE Conf. Ser. Vol. 6269, The VISTA Infrared Camera*, SPIE, Bellingham, p. 62690X
- Dalton G. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, *Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV*, SPIE, Bellingham, p. 84460P
- da Silva L. et al., 2006, *A&A*, 458, 609
- Das P., Sanders J., 2018, *MNRAS*, preprint ([arXiv:1804.09596](https://arxiv.org/abs/1804.09596))
- de Jong R. S. et al., 2016, in Evans C. J., Simard L., Takami H., eds, *Proc. SPIE Conf. Ser. Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI*, SPIE, Bellingham, p. 990810
- Deng L.-C. et al., 2012, *Res. Astron. Astrophys.*, 12, 735
- De Silva G. M. et al., 2015, *MNRAS*, 449, 2604
- Drimmel R., Cabrera-Lavers A., López-Corredoira M., 2003, *A&A*, 409, 205
- Emerson J., McPherson A., Sutherland W., 2006, *Messenger*, 126, 41
- Evans D. W. et al., 2018, *A&A*, 616, A4
- Fitzpatrick E. L., 1999, *PASP*, 111, 63
- Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
- Gaia Collaboration, 2016, *A&A*, 595, A1
- Gaia Collaboration, 2018a, *A&A*, 616, A1
- Gaia Collaboration, 2018b, *A&A*, 616, A10
- Gaia Collaboration, 2018c, *A&A*, 616, A11
- Gaia Collaboration, 2018d, *A&A*, 616, A12
- García Pérez A. E. et al., 2016, *AJ*, 151, 144
- Gilmore G. et al., 2012, *Messenger*, 147, 25
- Green G. M. et al., 2018, *MNRAS*, 478, 651
- Gunn J. E. et al., 2006, *AJ*, 131, 2332
- Ho A. Y. Q. et al., 2017, *ApJ*, 836, 5
- Hoffman M. D., Gelman A., 2011, preprint ([arXiv:1111.4246](https://arxiv.org/abs/1111.4246))
- Holmberg J., Nordström B., Andersen J., 2009, *A&A*, 501, 941
- Howes L. M., Lindegren L., Feltzing S., Church R. P., Bensby T., 2018, *A&A*, preprint ([arXiv:1804.08321](https://arxiv.org/abs/1804.08321))
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jones E., Oliphant T., Peterson P. et al., 2001, SciPy: Open Source Scientific Tools for Python, available at: <http://www.scipy.org/>
- Jørgensen B. R., Lindegren L., 2005, *A&A*, 436, 127
- Kollmeier J. A. et al., 2017, preprint ([arXiv:1711.03234](https://arxiv.org/abs/1711.03234))
- Kordopatis G. et al., 2013, *AJ*, 146, 134
- Kroupa P., Tout C. A., Gilmore G., 1993, *MNRAS*, 262, 545
- Kunder A. et al., 2017, *AJ*, 153, 75
- Lee Y. S. et al., 2008a, *AJ*, 136, 2022
- Lee Y. S. et al., 2008b, *AJ*, 136, 2050
- Lindegren L. et al., 2018, *A&A*, 616, A2
- Luri X., et al., 2018, *A&A*, 616, A9
- Mackereth J. T. et al., 2017, *MNRAS*, 471, 3057
- Magnier E. A. et al., 2017, *ApJS*, 205, 20
- Majewski S. R. et al., 2013, *AJ*, 154, 94
- Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, *A&A*, 453, 635
- Martell S. L. et al., 2017, *MNRAS*, 465, 3203
- Martig M., Minchev I., Flynn C., 2014, *MNRAS*, 443, 2452
- Martig M. et al., 2016, *MNRAS*, 456, 3655
- Marín-Franch A. et al., 2009, *ApJ*, 694, 1498
- Masseron T., Gilmore G., 2015, *MNRAS*, 453, 1855
- Maíz Apellániz J., 2006, *AJ*, 131, 1184
- McKinney W., 2010, in van der Walt S., Millman J., eds, *Proceedings of the 9th Python in Science Conference, Data Structure for Statistical Computing in Python, Scipy*, p. 51
- McMillan P. J. et al., 2018, *MNRAS*, 477, 5279
- Minchev I., Martig M., Streich D., Scannapieco C., de Jong R. S., Steinmetz M., 2015, *ApJ*, 804, L9
- Mints A., Hekker S., 2018, preprint ([arXiv:1804.06578](https://arxiv.org/abs/1804.06578))
- Ness M., Hogg D. W., Rix H.-W., Martig M., Pinsonneault M. H., Ho A. Y. Q., 2016, *ApJ*, 823, 114
- Pinsonneault M. H. et al., 2014, *ApJS*, 215, 19
- Pont F., Eyer L., 2004, *MNRAS*, 351, 487
- Pérez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Queiroz A. B. A. et al., 2018, *MNRAS*, 476, 2556
- Riello M. et al., 2018, *A&A*, 616, A3
- Riess A. G. et al., 2018, *ApJ*, 861, 126
- Robin A. C., Marshall D. J., Schultheis M., Reylé C., 2012, *A&A*, 538, A106
- Salaris M., Chieffi A., Straniero O., 1993, *ApJ*, 414, 580
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, *PeerJ Comput. Sci.*, 2, e55
- Sanders J., Binney J., 2016, *MNRAS*, 457, 2107
- Sartoretti P. et al., 2018, *A&A*, 616, A6
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
- Schlafly E. F. et al., 2016, *ApJ*, 821, 78
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Schönrich R., Binney J., Dehnen W., 2010, *MNRAS*, 403, 1829
- Sharma S., Bland-Hawthorn J., Johnston K., Binney J., 2011, *ApJ*, 730, 3
- Simion I. T., Belokurov V., Irwin M., Koposov S. E., Gonzalez-Fernandez C., Robin A. C., Shen J., Li Z.-Y., 2017, *MNRAS*, 471, 4323
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Smolinski J. P. et al., 2011, *AJ*, 141, 89
- Soderblom D. R., 2010, *ARA&A*, 48, 581
- Steinmetz M. et al., 2006, *AJ*, 132, 1645
- Tang J., Bressan A., Rosenfield P., Slemmer A., Marigo P., Girardi L., Bianchi L., 2014, *MNRAS*, 445, 4287
- Tayar J. et al., 2015, *ApJ*, 807, 82
- The Astropy Collaboration, 2018, *AJ*, 156, 123
- Vrard M., Mosser B., Samadi R., 2016, *A&A*, 588, A87
- Wilson J. C. et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, *Proc. SPIE Conf. Ser. Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III*, SPIE, Bellingham, p. 77351C
- Wu Y. et al., 2018, *MNRAS*, 475, 3633
- Xiang M.-S. et al., 2015, *Res. Astron. Astrophys.*, 15, 1209
- Xiang M. et al., 2017, *ApJS*, 232, 2
- Yanny B. et al., 2009, *AJ*, 137, 4377
- Zhao G., Zhao Y., Chu Y., Jing Y., Deng L., 2012, preprint ([arXiv:1206.3569](https://arxiv.org/abs/1206.3569))

SUPPORTING INFORMATION

Supplementary data are available at: https://www.ast.cam.ac.uk/~jls/data/gaia_spectro.hdf5

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.

APPENDIX A: EXTINCTION COEFFICIENTS

In our Bayesian distance pipeline, we require an extinction law to deredden any photometry used. This amounts to computing the set of coefficients $R(i)/R(V)$ for the photometric bands i . Given the V -band extinction A_V , the extinction in band i is $A_i = R(i)/R(V)A_V$. The traditional definition of $R(V)$ is such that $A_V = R(V)E(B - V)$, where $E(B - V)$ is the selective extinction. As we work with the extinction maps of Green et al. (2018), who provide extinction in units of E' , we define $R(V) = A_V/E'$.

Here, we choose to adopt the extinction curve $A(\lambda)$ from Schlafly et al. (2016) that was determined from APOGEE data. The curve was chosen to reproduce the observed extinction coefficients at the iso-extinction wavelengths for the Pan-STARRS and 2-MASS bands. Schlafly et al. (2016) parametrizes the extinction curve in terms of x , where $x = 0$ corresponds approximately to $R_V = 3.3$. Schlafly et al. (2016) finds that on average $R_V = 3.3$ although there is

Table A1. Extinction coefficients from Schlafly et al. (2016) extinction curve for use with Green et al. (2018) extinction maps: λ_{eff} gives the effective wavelength, $R(\lambda_{\text{eff}})$ the extinction coefficient evaluated at the effective wavelength, and $R(\lambda)$ the extinction coefficient computed by integrating over a solar-metallicity stellar spectrum of effective temperature 4500 K and surface gravity $\log g = 4.5$. Given extinction in the E' extinction units reported by Green et al. (2018), the extinction in band i is $R(i)E'$.

Band	$\lambda_{\text{eff}}/\text{\AA}$	$R(\lambda_{\text{eff}})$	$R(\lambda)$	A_i/A_V
<i>u</i>	3607.7	4.781	4.869	1.636
<i>B</i>	4533.6	3.830	3.878	1.303
<i>g</i>	4811.9	3.576	3.613	1.214
<i>g_P</i>	4958.8	3.449	3.481	1.169
<i>G_{BP}</i>	5524.6	2.995	3.046	1.023
<i>V</i>	5560.0	2.969	2.977	1.000
<i>r</i>	6202.6	2.533	2.519	0.846
<i>r_P</i>	6239.8	2.510	2.499	0.840
<i>G</i>	6791.6	2.201	2.294	0.771
<i>i</i>	7503.5	1.868	1.824	0.613
<i>i_P</i>	7548.3	1.849	1.805	0.606
<i>G_{RP}</i>	7782.1	1.752	1.737	0.584
<i>G_{RVS}</i>	8597.2	1.449	1.393	0.468
<i>z_P</i>	8679.3	1.421	1.368	0.460
<i>z</i>	8957.7	1.329	1.285	0.432
<i>y_P</i>	9631.3	1.132	1.084	0.364
<i>J</i>	12365.1	0.662	0.632	0.212
<i>H</i>	16426.9	0.335	0.318	0.107
<i>K_s</i>	21526.0	0.170	0.158	0.053
<i>W₁</i>	33183.4	0.046	0.044	0.015
<i>W₂</i>	45550.9	−0.002	0.004	0.001

variation of ~ 0.2 across the APOGEE survey area. In the main body of the paper, we fix $R_V = 3.3$ and hence choose $x = 0$. The bluest band considered by Schlafly et al. (2016) was the Pan-STARRS g_P band. Therefore, computing the extinction coefficients for bluer bands (particularly *u*) is an extrapolation.

Following Green et al. (2018), we set the unknown grey component by insisting the extinction in the WISE W_2 band is zero. We set the scaling relative to the extinction E' provided by Green et al. (2018) (chosen such that one unit of $E(g_P - r_P)$ for the Pan-STARRS bands produces one unit of Schlegel et al. $E(B - V)$ by matching the coefficients for g_P and r_P provided in Table 1 of Green et al. (2018) at the iso-extinction wavelengths of Schlafly et al. (2016). We compute

$$R(i) = -\frac{2.5}{E'} \log_{10} \frac{\int d\lambda 10^{-0.4E' A(\lambda)} \lambda S_i(\lambda) f(\lambda)}{\int d\lambda \lambda S_i(\lambda) f(\lambda)}, \quad (\text{A1})$$

where $S_i(\lambda)$ is the response of bandpass i^5 and $f(\lambda)$ the stellar model flux. From these coefficients and the selective extinction reported by Green et al. (2018), the extinction in band i is $A_i = R(i)E'$.

In Table A1, we provide the extinction coefficients evaluated using equation (A1) with $E' = 0.01$ using for $f(\lambda)$ the Castelli & Kurucz (2004) stellar model with solar metallicity, $T_{\text{eff}} = 4500$ K and $\log g = 4.5$.

A1 Variation of $R(G)$ with intrinsic colour

As the *Gaia* G band is broad, the coefficient provided in Table A1 is not appropriate for high extinction and for stars with intrinsic colours significantly different from that of a $T_{\text{eff}} = 4500$ K star. To handle the latter of these issues, we evaluate equation (A1) as a function of effective temperature and R_V (we neglect any deviation with metallicity and surface gravity) using the solar metallicity and surface gravity $\log g = 4.5$ stellar models of Castelli & Kurucz (2004) (for $3500 \text{ K} < T_{\text{eff}} < 50\,000 \text{ K}$). To the resulting $R(G)$, we fit polynomials of the form (y^2 and y^3 terms were unnecessary)

$$R(G)(x, y) = \mathbf{P} \cdot (1, x, y, x^2, x^2y, xy^2, xy, x^3). \quad (\text{A2})$$

We use $x = \log_{10} T_{\text{eff}} - 4$ and $y = R_V - 3.3$. Additionally, using a solar metallicity PARSEC isochrone (Bressan et al. 2012; Chen et al. 2014; Tang et al. 2014; Chen et al. 2015) we can rewrite these polynomials with arguments $x = (J - K_s)$, $(g - r)$, $(g_P - r_P)$, $(B - V)$, $(G_{BP} - G_{RP})$, $(G - G_{RP})$, and $(G_{BP} - G)$. The coefficients for this polynomial is given in Table A2. Throughout the main body of the paper we use $y = 0$ so $R_V = 3.3$.

The run of $R(G)$ with $(J - K_s)$ is shown in Fig. A1 coloured by choice of R_V . We also show the polynomial result for $R_V = 3.3$ in black. We note that the variation of $R(G)$ with effective temperature is reasonably large (~ 40 per cent) where for blue stars $(J - K_s = 0)$ $R(G) \approx R(V) = 2.98$, for redder stars of $(J - K_s) = 0.4$ $R(G) \approx R(r) = 2.51$ and for very red stars $(J - K_s = 1)$ $R(G)$ is as low as 2.1.

As the *Gaia* G band is broad, we also consider variation in $R(G)$ with the monochromatic extinction. For two models $T_{\text{eff}} = (4500, 10\,000)$ K, $\log g = 4.5$, we compute equation (A1) for a range of $0.01 < E' < 1$. We fit a simple linear relation to A_G/A_V against A_V

$$A_G/A_V = R(G)(1 - c_G A_V), \quad (\text{A3})$$

finding that for the $T_{\text{eff}} = 4500$ K model $c_G = -0.031$ and for $T_{\text{eff}} = 10\,000$ K model $c_G = -0.038$. We adopt the simple relation $c_G = -0.03$. Similarly we find for both G_{BP} and G_{RP} , $c_{BP,RP} = -0.01$. For all other bands these gradients are negligible.

For completeness we provide $R(G)$ polynomial fits for use with Schlegel et al. (1998) extinction $E(B - V)_{\text{SFD}}$, i.e. $A_i = R(i)E(B - V)_{\text{SFD}}$. We utilize a Fitzpatrick (1999) extinction law corrected for the Schlegel et al. (1998) units using Schlafly & Finkbeiner (2011). Again, we consider a range of stellar models and fit polynomials of the form

$$R(G)(x, y) = \mathbf{P} \cdot (1, x, y, x^2, x^2y, xy^2, xy, x^3), \quad (\text{A4})$$

with $y = R_V - 3.1$. The G extinction is then computed using $A_G = R(G)E(B - V)_{\text{SFD}}$. The results are given in Table A3. For G_{RP} and G_{BP} , we find coefficients $R(G_{RP}) = 1.64$ and $R(G_{BP}) = 2.73$. Finally, we note that *Gaia* Collaboration (2018b) also provides relations for $R(G)$, $R(G_{BP})$, and $R(G_{RP})$ in terms of the intrinsic colour and absolute extinction.

⁵*Gaia* DR2 photometric response curves downloaded from www.cosmos.esa.int, Pan-STARRS from <http://ipp.ifa.hawaii.edu/ps1.filters/>, 2MASS from <https://www.ipac.caltech.edu/2mass/releases/second/> (provides $\lambda S(\lambda)$), WISE from <http://wise2.ipac.caltech.edu/docs/release/prelim/expSUP/>, Johnson *B* and *V* from Maíz Apellániz (2006) and SDSS from <http://classic.sdss.org/dr3/instruments/imager/filters/>.

Table A2. $R(G)$ extinction coefficient polynomial coefficients where $R(G)(x, y) = \mathbf{P} \cdot (1, x, y, x^2, x^2y, xy^2, xy, x^3)$, where $A_G = R(G)E'$ for the extinction E' reported by Green et al. (2018) using Schlafly et al. (2016) extinction curve. Each row corresponds to a different x given in the left column and $y = R_V - 3.3$.

$\log_{10}(T_{\text{eff}}) - 4$	2.9083	1.0268	0.0278	-1.571	0.2401	-0.0007	-0.247	0.9734
$(G - G_{\text{RP}})$	2.9228	-0.7571	0.022	-0.164	0.008	0.0007	0.1633	0.081
$(G_{\text{BP}} - G)$	2.9083	-0.9529	0.0285	-0.274	-0.0392	0.0007	0.1937	0.3879
$(G_{\text{BP}} - G_{\text{RP}})$	2.9176	-0.4347	0.0255	-0.0648	-0.0041	0.0004	0.0893	0.0394
$(J - K_s)$	2.8808	-1.0154	0.0302	0.2385	-0.0696	0.0008	0.2245	-0.0014
$(g - r)$	2.7549	-0.5898	0.0525	0.3843	-0.0425	0.0005	0.1456	-0.2209
$(g_{\text{P}} - r_{\text{P}})$	2.7433	-0.6583	0.0552	0.4938	-0.056	0.0006	0.165	-0.3346
$(B - V)$	2.908	-0.7492	0.0198	0.3956	-0.0402	0.0005	0.1621	-0.1553

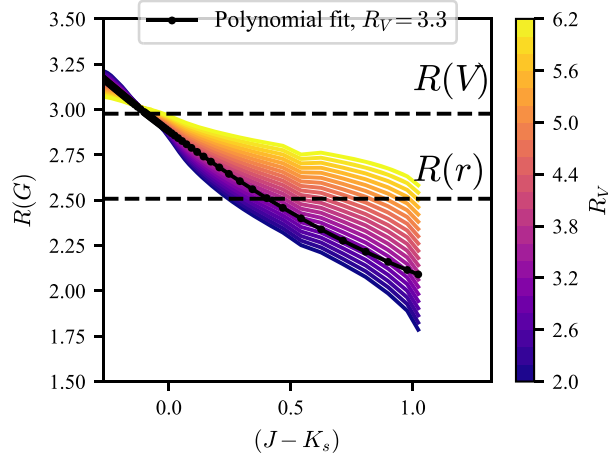


Figure A1. Extinction coefficient for *Gaia* G band: the extinction coefficient $R(G)$ is plotted as a function of $(J - K_s)$ coloured by the choice of Schlafly et al. (2016) extinction law coefficient R_V . The black dots show the polynomial fit for $R_V = 3.3$. The two horizontal dashed lines show the extinction coefficients for the narrow bands r and V . This extinction coefficient is for use with the extinctions E' reported by Green et al. (2018) as $A_G = R(G)E'$.

Table A3. $R(G)$ extinction coefficient polynomial coefficients where $R(G)(x, y) = \mathbf{P} \cdot (1, x, y, x^2, x^2y, y^2, xy^2, xy, x^3, y^3)$, where $A_G = R(G)E(B - V)_{\text{SFD}}$ for the extinction $E(B - V)_{\text{SFD}}$ reported by Schlegel, Finkbeiner & Davis (1998) using Fitzpatrick (1999) extinction curve and the correction of Schlafly & Finkbeiner (2011). Each row corresponds to a different x given in the left-hand column and $y = R_V - 3.1$.

$\log_{10}(T_{\text{eff}}) - 4$	2.6241	0.9392	-0.1496	-1.4703	0.1517	0.0525	0.0368	-0.2465	0.9321	-0.0092
$(G - G_{\text{RP}})$	2.6343	-0.698	-0.1641	-0.1484	-0.0126	0.0579	-0.0356	0.1954	0.0846	-0.0092
$(G_{\text{BP}} - G)$	2.6214	-0.881	-0.1601	-0.2573	-0.0424	0.0579	-0.0365	0.2183	0.3736	-0.0092
$(G_{\text{BP}} - G_{\text{RP}})$	2.6298	-0.4016	-0.1619	-0.0602	-0.0071	0.0579	-0.0182	0.1038	0.0382	-0.0092
$(J - K_s)$	2.5951	-0.9401	-0.1587	0.226	-0.0725	0.058	-0.0401	0.2513	0.014	-0.0092
$(g - r)$	2.4775	-0.5453	-0.1352	0.3646	-0.0411	0.0549	-0.0254	0.1604	-0.2039	-0.0092
$(g_{\text{P}} - r_{\text{P}})$	2.4668	-0.6083	-0.1323	0.4688	-0.0541	0.0545	-0.0289	0.1818	-0.3089	-0.0092
$(B - V)$	2.6197	-0.6955	-0.1708	0.3719	-0.0394	0.0602	-0.0259	0.1774	-0.1416	-0.0092

APPENDIX B: OUTPUT CATALOGUE FORMAT

In Table B1, we detail the columns in the output catalogue available at https://www.ast.cam.ac.uk/jls/data/gaia_spectro.hdf5. We give

the column name, description, and the unit of the quantity where applicable.

Table B1. Format of provided catalogue. We give the names, description, and unit of each provided quantity.

Name	Description	Unit
survey	String describing the spectroscopic survey from which the entry derives – APOGEE, GALAH, GES, RAVEON, RAVEDR5, LAMOST, SEGUE	–
raveid	unique ID for the RAVE (and RAVE-ON) catalogue	–
APOGEE_ID	unique ID for the APOGEE catalogue	–
obsid	unique ID for the LAMOST catalogue	–
subject_id	unique ID for the GALAH catalogue	–
CNAME	unique ID for the GES catalogue	–
specobjid	unique ID for the SEGUE catalogue	–
source_id	Cross-matched <i>Gaia</i> DR2 source ID	–
angular_separation	On-sky separation between spectroscopic catalogue entry and <i>Gaia</i> DR2 cross-match (epoch corrected if proper motion has been considered)	arcsec
ra	Right ascension from spectroscopic catalogue	degree
dec	Declination from spectroscopic catalogue	degree
mag_use	Comma-separated string giving photometry used in pipeline: J, H, K are 2MASS bands, G, GBP, GRP are <i>Gaia</i> bands, gP, rP, iP are Pan-STARRS bands, g, r, i are SDSS bands, Jv, Hv, Kv are VISTA bands	–
Primary outputs		
dm	Distance modulus	mag
dist	Distance	kpc
par	Parallax (not <i>Gaia</i> DR2 parallax)	mas
log10_age	Base-10 logarithm of age	log ₁₀ (Gyr)
mass	(Initial) Mass	M_{\odot}
Z	Metallicity	dex
log10_av	Base-10 logarithm of V-band extinction	log ₁₀ (mag)
log10_teff	Base-10 logarithm of effective temperature	log ₁₀ (K)
logg	Surface gravity	log ₁₀ (cm/s ²)
Auxiliary outputs		
l	Galactic longitude	rad
b	Galactic latitude	rad
s	Distance derived from dm column	kpc
vlos	Line-of-sight velocity	km s ^{−1}
mu_l	Proper motion in Galactic longitude	mas yr ^{−1}
mu_b	Proper motion in Galactic latitude	mas yr ^{−1}
R	Galactocentric cylindrical polar radius	kpc
phi	Galactocentric cylindrical polar angle (zero at solar position increasing in direction opposite to solar azimuthal velocity)	rad
z	Galactic height	kpc
vR	Galactocentric cylindrical radial velocity	km s ^{−1}
vphi	Galactocentric azimuthal velocity (positive for Sun and decreasing phi)	km s ^{−1}
vz	Galactocentric vertical velocity	km s ^{−1}
JR	Radial action	kpc km s ^{−1}
Lz	z-component of angular momentum (positive for Sun)	kpc km s ^{−1}
Jz	Vertical action	kpc km s ^{−1}
Rc	Galactocentric radius of circular orbit with angular momentum Lz	kpc
*_err	Corresponding uncertainty for each field	–
*_*_corr	Correlations in derived quantities	–
Flags		
flag	non-zero if pipeline has failed (see text for meaning of each value)	–
duplicated	1 if duplicate <i>Gaia</i> source_id (see text for how we select the source to keep)	–
best	1 if flag=0, duplicated=0 and valid <i>Gaia</i> source_id	–

This paper has been typeset from a \LaTeX file prepared by the author.