



**HAL**  
open science

## Stellar age determination using deep neural networks

T. Boin, L. Casamiquela, M. Haywood, P. Di Matteo, Y. Lebreton, M. Uddin, D. Reese

► **To cite this version:**

T. Boin, L. Casamiquela, M. Haywood, P. Di Matteo, Y. Lebreton, et al.. Stellar age determination using deep neural networks. *Astronomy & Astrophysics - A&A*, 2026, 708, pp.A215. <10.1051/0004-6361/202558436>. <hal-05599914>

**HAL Id: hal-05599914**

**<https://hal.science/hal-05599914v1>**

Submitted on 22 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Stellar age determination using deep neural networks

## Isochrone ages for 1.3 million stars, based on BaSTI, MIST, PARSEC, Dartmouth, and SYCLIST evolutionary grids

T. Boin<sup>1</sup>, L. Casamiquela<sup>1</sup>, M. Haywood<sup>1</sup>, P. Di Matteo<sup>1</sup>, Y. Lebreton<sup>1,2</sup>, M. Uddin<sup>1</sup>, and D. R. Reese<sup>1</sup>

<sup>1</sup> LIRA, Observatoire de Paris, PSL Research University, CNRS, Univ Paris Diderot, Sorbonne Paris Cité, Place Jules Janssen, 92195, Meudon, France

e-mail: tristan.boin@obspm.fr

<sup>2</sup> Univ Rennes, CNRS, IPR (Institut de Physique de Rennes) – UMR 6251, 35000 Rennes, France

Received XXX; accepted XXX

### ABSTRACT

**Context.** Recent spectroscopic surveys provide element abundances for large samples of Milky Way stars, from which stellar parameters can be inferred. Stellar ages, among them, are both a notoriously difficult parameter to estimate and a fundamental property for Galactic archaeology studies.

**Aims.** We aim to develop a model-driven deep learning approach to age determination by training neural networks on stellar evolutionary grids. Contrary to the usual data-driven deep learning approach of using prior age estimates as training data, our method has the potential for a wider and less biased range of application. The low computational cost of deep learning methods compared to, for example, Bayesian isochrone fitting enables a broad analysis of large spectroscopic catalogues.

**Methods.** We trained multilayer perceptrons on different stellar evolutionary grids to map  $[M/H]$ ,  $M_G$ ,  $(G_{BP} - G_{RP})$  to stellar age  $\tau$ . We combined *Gaia* photometry and parallaxes, metallicities, and  $\alpha$  elements from spectroscopic surveys and extinction maps, which are passed through neural networks to estimate stellar ages.

**Results.** We applied our method to the LAMOST DR10, GALAH DR3 & DR4, and APOGEE DR17 spectroscopic surveys, estimating ages using the BaSTI tracks and other stellar evolutionary models. We leveraged this novel technique to study, for the first time, differences in age estimates from several evolutionary grids applied to very large datasets. In addition, we dated 13 open clusters and one globular cluster, finding a median absolute deviation with literature ages of 0.20 Gyr. Along with the stellar age catalogues from our estimates, we release NEST (Neural Estimator of Stellar Times), a python package to estimate stellar age based on this work, as well as a web interface.

**Conclusions.** We show that, when using the same evolutionary grid, our method retrieves the same ages as a Bayesian approach similar to SPInS, for only a fraction of the computational cost, with a 60,000 speed-up factor for a typical star. This model-driven deep learning technique thus opens up the way for broad galactic archaeology studies on the largest datasets available today and in the near future with upcoming surveys such as 4MOST.

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

## 1. Introduction

Galactic archaeology relies on a combination of precise abundance measurements and age estimates to study stellar populations and their evolution. On the chemical abundances side, large spectroscopic surveys such as the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) (Cui et al. 2012; Zhao et al. 2012), Apache Point Observatory Galactic Evolution Experiment (APOGEE) (Abdurro'uf et al. 2022), and Galactic Archaeology with HERMES (GALAH) (Buder et al. 2021; Buder et al. 2025) provide invaluable amounts of data on elemental abundances, which, coupled with the astrometric and photometric data provided by the *Gaia* DR3 catalogue (Gaia Collaboration et al. 2023), open up the door to a deeper understanding of our Galaxy.

Concomitantly, recent advances in stellar physics modelling have been made by different research groups, providing numerous stellar evolution models, spanning a wide range of parameters and set of underlying physical assumptions (see Lebreton et al. (2014a) for a review). Coupled with Bayesian (Jørgensen

& Lindegren 2005; Casagrande et al. 2011; Haywood et al. 2013; von Hippel et al. 2014; Lebreton & Reese 2020; Cerqui et al. 2025) or probabilistic (Bensby et al. 2011) approaches to isochrone fitting, these new datasets and models have yielded millions of stellar ages.

In parallel, missions such as *Kepler* (Borucki et al. 2010; Bedding et al. 2010; Chaplin et al. 2010), CoRoT (Baglin et al. 2006), TESS (Ricker et al. 2014; Hon et al. 2021), and in the near future PLATO (Rauer et al. 2014; Goupil et al. 2024), have and will provide data used for asteroseismic age determinations. This method of dating is often complementary to isochrone fitting techniques, as the latter are most often restricted to the main-sequence turn off (MSTO) and sub-giant branch (SGB) of the Hertzsprung-Russell diagrams (HRDs), whereas the former applies to solar-like oscillations of red giant stars, but also main sequence and sub-giants (Lebreton et al. 2014b; García & Ballot 2019). Asteroseismic ages have been determined for field stars (Anders et al. 2017; Pinsonneault et al. 2025) as well as clusters (Reyes et al. 2025); however, discrepancies arise when comparing them to ages from isochrone fitting techniques (Tayar

& Joyce 2025). For a review of the different age determination methods, see Soderblom (2010).

With the advent of machine learning methods in recent years, age determination techniques have evolved towards more and more data-driven artificial intelligence methods (e.g. Anders et al. 2023a; Bu et al. 2020; Moya et al. 2021; Van-Lane et al. 2023; Boulet 2024; Tamames-Rodero et al. 2025). Indeed, machine learning appears well suited for the amount of data made available by large surveys, providing ample training sets. Compared to traditional isochrone fitting methods, which are relatively computationally expensive as they often need to load the full evolutionary grids to run their fitting procedures, neural networks (NNs) have a much lower computational cost. Moreover, once the training procedure has been run and the network has built a faithful representation of the model, the data they have been trained on do not need to be accessed.

Most (but not all; see the discussion in Section 5) of the machine learning approaches to age determinations are data-driven, meaning they rely on observable quantities, as well as age estimates to be used as a training set. These estimates can be asteroseismic (Das & Sanders 2018; Ciucă et al. 2021; Anders et al. 2023a; Wang et al. 2024), gyrochronologic (Van-Lane et al. 2023), isochrone-fitted (Cantat-Gaudin et al. 2020), or chemical-clock (Hayden et al. 2022) ages. Depending on the size, coverage, and quality of the training set, this can lead to biases in the resulting predictions, compounded with the biases already present from the methodology that provided these training ages, as well as the biases introduced by the deep learning methodology using this training set. These can be difficult to track down, isolate, and thus deal with. In the best case scenario, if the biases are known, the range of applicability of the deep learning approach can be reduced to a subset where the estimates are reliable. On the contrary, a model-driven deep learning approach would only be dependent on said model and on the methodology to leverage it. In this work, we introduce a model-driven approach to NNs for stellar age determination. Our aim is to apply this technique to a wide range of datasets, and compare our results with other age estimates from the literature. This general approach generates different age distributions, which can also be used to study abundance-age relations, the common structures found between datasets, and their differences.

This paper is structured as follows. In Section 2, we introduce the deep learning methodology, the NN architecture, and the training data and provide a comparison between NNs trained on different stellar evolution models. Section 3 covers different data samples of cluster and field stars for which the NNs produce age predictions. In Section 4, we analyse the age predictions for each sample, as well as age-chemistry relations. In Section 5, we discuss the advantages and shortcomings of our methodology and its place in the landscape of age estimation methods.

## 2. Methods and stellar models

The general procedure employed is to train an NN on a stellar evolution model grid, taking as inputs observables such as magnitude, colour, and metallicity, and using the age they provide as the output expected from the network. We describe the stellar evolution models used and the architecture of the NN below.

### 2.1. Method

We modeled our NNs as multilayer perceptrons, with four hidden layers having 32, 64, 64, and 32 neurons, respectively (see

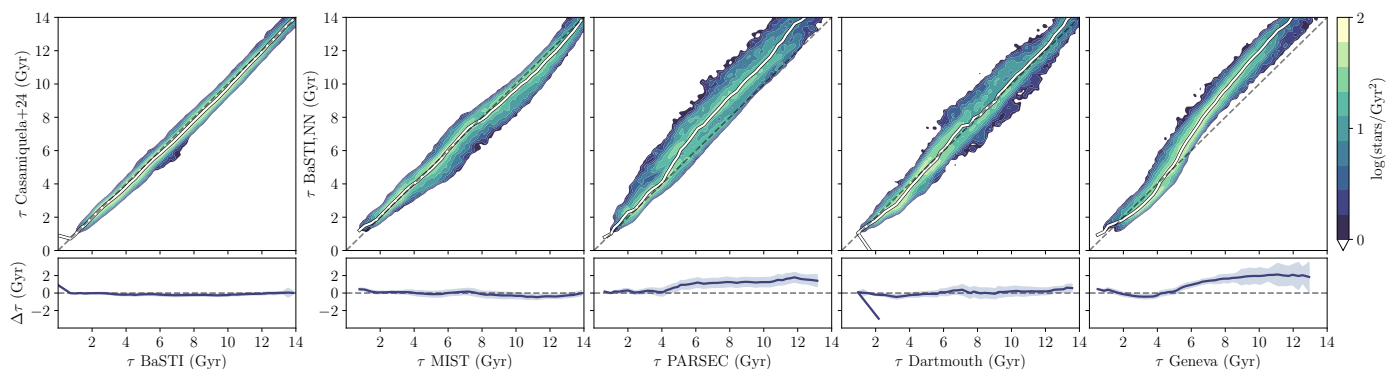
Appendix A for a representation of the adopted architecture). A lower number of hidden layers as well as a lower number of neurons per hidden layer both lead to decreased performance, as measured by the median absolute deviation (MAD) on the test split, while a higher number of neurons per layer and a higher number of hidden layers both lead to no gain in performance. The networks use a ReLU activation function, and the training was done using the `MLPRegressor.fit` function of the scikit-learn (Pedregosa et al. 2011) Python package, using the Adam solver (Kingma & Ba 2017) and a constant learning rate of  $1 \times 10^{-3}$ . The number of epochs needed to converge to a constant loss varied depending on the theoretical grid used, with a mean of about 300 epochs.

Similar yet superior results were achieved when the NNs were trained with only three inputs ( $\mathbf{x} = \{[M/H], M_G, (G_{BP} - G_{RP})\}$ ), as opposed to five inputs ( $\mathbf{x} = \{[M/H], M_G, (G_{BP} - G_{RP}), G_{BP}, G_{RP}\}$ ) (as measured by the MAD on the test split). Thus, we conclude that the addition of separate  $G_{BP}$  and  $G_{RP}$  magnitudes does not help differentiate between isochrones and that the colour-magnitude space (in our case  $M_G, (G_{BP} - G_{RP})$ ) provides most, if not all, of the information needed to infer the age. Different architectures were also tried, with the presented one offering the best compromise between accuracy on the training data and avoiding over-fitting.

Using  $T_{\text{eff}}$  as an additional input generally worsened the age estimations. This is at first counter-intuitive as we expect more input data to improve the NN performance. However, in our stellar evolutionary grid,  $T_{\text{eff}}$  is strongly correlated to colour, but because our observational data do not have an exact one-to-one relation between the estimated  $T_{\text{eff}}$  and colour, our NNs struggle to reconcile observational values of colour and  $T_{\text{eff}}$  that do not agree with each other (see Appendix B for an example of the theoretical and observed  $T_{\text{eff}} - (G_{BP} - G_{RP})$  relation). We note that increasing the number of input parameters also complexifies the NN architecture, which might also impact its ability to converge to a faithful representation of the models. Additionally,  $T_{\text{eff}}$  uncertainties are generally high, leading to an age uncertainty being largely dominated by them.

The input parameters  $\mathbf{x}$  were rescaled before being passed to the NNs by subtracting the mean value and dividing by the standard deviation of each of the training set features. The same rescaling, using the mean and standard deviation of the training set, was performed on the input data passed to the trained NN to estimate ages. The dataset used for training, i.e. the grid points of the stellar evolution models was separated into a training set (90%) and a testing set (10%). After training, the network converged to a reasonable representation of the training model, as the MAD on the testing set after the training procedure was about 35 Myr, with variations depending on the grid used.

For a given star, the age estimate  $\tau$  is given by the value of the single neuron of the last layer of the NN. This is, however, a single value that does not account for observational uncertainties and, as such, might not provide a good estimate of the true age. Some probabilistic approaches to machine learning, such as Bayesian NNs (Tishby et al. 1989), exist that estimate the uncertainty in their output. However, our training dataset does not include uncertainties, and we are interested in a full age probability description for each star. Ideally, we want to obtain a probability density function (PDF) of the age of the star, given its observable properties  $\mathbf{x}_{\text{obs}}$  and associated uncertainties  $\Delta\mathbf{x}_{\text{obs}}$ . To this end, we followed a Monte Carlo (MC) sampling method, as used in, for example, Jørgensen & Lindegren (2005), where we generated a large number of samples  $\mathbf{x}_i = \mathcal{N}(\mathbf{x}_{\text{obs}}, \Delta\mathbf{x}_{\text{obs}})$  (i.e. assuming Gaussian error distributions) and passed them to the NN



**Fig. 1.** *Top panels, leftmost panel:* Age comparison of Casamiquela et al. (2024)’s 35K sub-giant sample age estimates and results from the NN trained on BaSTI. *Top panels, other panels:* Age estimates from the NN trained on BaSTI compared to those from NNs trained on (from left to right) MIST, PARSEC, Dartmouth, and SYCLIST evolutionary grids. Running medians are shown with white lines, as well as a one-to-one dashed black line to guide the eye. *Bottom panels:* Running median of the difference  $\Delta\tau$  between the age estimates  $\tau$  of C24 and our BaSTI NN (leftmost panel), between our BaSTI NN and other NNs (other panels) in solid lines, and the associated standard deviations in shaded areas.

to obtain age estimates  $\tau_i$ . The normalised distribution of the age estimates  $\tau_i$  thus represents a discretised PDF  $p(\tau|x_{\text{obs}}, \Delta x_{\text{obs}})$ , given the observed properties of the star and their uncertainties. From this PDF, we estimated the age uncertainty by computing the standard deviation of the  $\tau_i$  age distribution. Different estimators can then be used to obtain a single-valued age from the PDF, namely the median, mean, or mode of the distribution. For a comparison of these three estimators, we refer to Jørgensen & Lindgren (2005).

In this work, the median of the distribution was used as our single-valued age estimate, and we used 10 000 MC samples for each star. As we applied the NNs to a variety of large catalogues, covering a wide range of the colour-magnitude diagram (CMD), a choice of 10 000 Monte Carlo samples ensured a faithful representation of the age PDF, even in regions of the CMD with high isochrone degeneracy and overlap. However, in more constrained, well-defined regions, 1000 Monte Carlo samples are sufficient to generate a representative age PDF. In Appendix B, we discuss how the median ages and their associated uncertainties computed in this work did not significantly change when using 1000 Monte Carlo samples. We verified that using the mode or the mean instead did not affect our conclusions: individually, these estimators can lead to different age estimates, but they have little consequence on the analysis of the large datasets used in this work.

## 2.2. Stellar evolution models

Neural networks were trained on different stellar evolution models, whose modelling physics can vary; however, none include the effects of stellar rotation on the structure, i.e. all models have  $v/v_{\text{crit}} = 0$ . This work uses the following grids of stellar models: Bag of Stellar Tracks and Isochrones (BaSTI)<sup>1</sup> (Hidalgo et al. 2018), MESA Isochrones and Stellar Tracks (MIST)<sup>2</sup> (Choi et al. 2016), PAdova and TRieste Stellar Evolution Code (PARSEC)<sup>3</sup> (Bressan et al. 2012; Nguyen et al. 2022; Costa et al. 2025), Dartmouth Stellar Evolution Program (DSEP)<sup>4</sup> (Dotter et al. 2007, 2008), and the Geneva Stellar Evolution Code<sup>5</sup> (Ek-

ström et al. 2012; Georgy et al. 2013a,b; Georgy, C. et al. 2014), specifically the large grids of Yusof et al. (2022). The parameter spaces for each model are given in Appendix C. We chose solar-scaled grids, with non- $\alpha$ -element-enhanced abundances, as we corrected our stellar samples for non-zero  $\alpha$  element abundances to build a homogeneous set of models. All models include the overshooting of convective cores and microscopic diffusion.

All stellar evolution phases are included in the grids, except pre-MS (pre-Main Sequence). We note that in the rest of this work, this entire set of points has been kept as training data for the NNs, including the MS phase. Neural networks trained only on specific regions of the CMD, restricted to regions where the isochrones are well separated (e.g. MSTO and SGB) have also been tested, but we found no significant gain in their accuracy, as measured by the MAD on the test set (although reducing the training data size and complexity greatly reduces the training time). Thus, we decided on training the NNs on the grids including the MS, so as to retain a more general tool of age estimation. In the MS sequence region, the age estimations come, as expected, with large uncertainties such that a suitable a posteriori cut on the age uncertainty inevitably removes most stars below the MSTO, due to the high density of isochrones in this region.

We investigated the effects that different evolutionary models have on our age estimates. To this end, we trained NNs with the same common architecture, but on each of the grid of models presented in Section 2.2. We used Casamiquela et al. (2024)’s (thereafter C24) sample of 35K sub-giant stars as a control group, with cuts on  $\log g$  ( $3.5 < \log g < 4.45$ ) and on age uncertainty ( $\sigma_\tau/\tau < 10\%$  and  $\sigma_\tau < 1$  Gyr) so as to restrict our comparison to regions of low isochrone degeneracy and overlap. The results are summarised in Fig. 1. The age estimates in C24 are based on the SPInS tool (Lebreton & Reese 2020), which uses a Markov chain Monte Carlo (MCMC) approach on BaSTI isochrones. The ages used in Fig. 1 are the median ages of the age PDFs it provides.

Because evolutionary grids might use different solar mixtures as reference for their metallicity, it is important to take into account the different zero points adopted so as to make consistent comparisons between grids. To this end, we took the BaSTI grid as a reference metallicity (which uses Caffau et al. (2011)’s solar abundances), computed, and applied the following metallicity offsets. For the MIST grid, which uses Asplund et al. (2009)’s solar abundances, we applied a -0.057 dex offset; for the Dartmouth grid, which uses Grevesse & Sauval (1998)’s solar abun-

<sup>1</sup> <http://basti-iac.oa-abruzzo.inaf.it>

<sup>2</sup> <https://waps.cfa.harvard.edu/MIST>

<sup>3</sup> <https://stev.oapd.inaf.it/PARSEC>

<sup>4</sup> <https://rcweb.dartmouth.edu/stellar>

<sup>5</sup> <https://www.unige.ch/sciences/astro/evolution/en/database/syclist>

dances, we applied a 0.046 dex offset; and for the Geneva grids, which uses [Asplund et al. \(2005\)](#) solar abundances, we applied a -0.075 dex offset. We did not apply any offset to the PARSEC grid, as it uses the same solar mixture as the BaSTI grid.

The NN trained on the BaSTI grid gives the best agreement with the C24 results out of all the NNs, as it uses the same evolutionary grid (see the leftmost subplot of Fig. 1). The only noticeable offset is at very young ages, which have a lower estimate with our method. Such a tight one-to-one relation highlights two results: first, a deep learning approach, trained on theoretical models instead of on other age estimates of samples, can provide at least as accurate stellar ages as other age dating methods. Second, the discrepancies between different age estimation methods applied on the same data is, in a non-negligible part, due to the underlying theoretical evolutionary model used (as seen in the other subplots of Fig. 1). This conclusion holds at least for isochrone fitting procedures, whose result heavily depends on the model used. Indeed, our NN-based age estimates, compared to C24 age estimates using the SPInS MCMC approach, only display appreciable dispersion once different models are used and display a very low dispersion otherwise, when using the same input data. That is to say, under the same input data and stellar evolution model, a MCMC method and an NN will converge towards the same age estimates. As the same data and model were used, the low dispersion seen is a result of the different Bayesian and deep learning approaches. The slightly higher dispersion seen for ages  $\sim 5$  Gyr is a result of the metallicity correction we applied, which was not applied in C24. We verified that using uncorrected metallicities yielded better agreement and did not produce the aforementioned feature.

Other age estimates compare differently. Both the MIST and Dartmouth ages show very little offset of the median age versus true age, with median standard deviations of 0.31 and 0.40 Gyr. The final deviation at ages  $\tau_{\text{MIST}} > 12.6$  Gyr in the MIST ages is due to the age range of the MIST grid, which stops at 12 Gyr (see Appendix C), making the prediction of ages older than this unlikely (but possible, as the NN can extrapolate). The PARSEC ages are mostly offset by about 1.5 Gyr. We also note that the distribution of age estimates is not symmetrical around the running median but instead has a longer tail of estimated old ages, with almost no ages being estimated younger by the NN than those found with the BaSTI NN. Finally, the Geneva grid shows younger ages having a low offset, and older ages being lower than BaSTI ones by up to 3 Gyr for the oldest ones.

Notwithstanding training domain restrictions, it is interesting to note that both the PARSEC and Geneva models tend to give systematically lower ages than BaSTI for nearly all ages, mostly incompatible within  $1\sigma$  (see the bottom panels of Fig. 1). We find that correcting for different assumed solar mixtures improves the agreement between models, as we compared ages without such corrections and found that for all models, BaSTI ages would be systematically higher than all other models, for nearly all ages.

Comparisons with other models suggest that input physics in stellar models still produces systematic and random uncertainties of the order of 1-2 Gyr in the age determinations. For stars older than about 4 Gyr and younger than 1-2 Gyr, BaSTI ages are systematically older than the PARSEC and Geneva models considered here. MIST and Dartmouth ages show no systematic offset. We emphasise that the differences are not only systematic, with dispersions sometimes very significant at all ages, in particular with the PARSEC and Dartmouth models. For a discussion on stellar modelling shortcomings and their consequences on age determination, we refer to [Howes et al. \(2019\)](#). We note

that for this comparison, we did not compute the colours from  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  as could also be done, but instead relied on the colours provided by each model. This can lead to slight differences, as the PARSEC grid, for example, assumes slightly different bolometric corrections than other grids. Comparisons of age estimates using different grids will also depend on the dataset used and its CMD coverage. To highlight this fact, we compare evolutionary models in Appendix D as in Fig. 1, but using GALAH DR4 as the comparative dataset. In this case, we find better agreement between grids, as stars populate regions of lower isochrone degeneracy and overlap compared to C24's dataset.

To estimate the stochasticity of the training procedure of NNs, we trained ten NNs on the BaSTI grid and estimated ages with each of them using C24's dataset. The resulting median standard deviation is 0.22 Gyr, compared to 0.37 Gyr for the median standard deviation of each star, as computed by the MC sampling method described in Section 2.1. Thus, in the vast majority of cases, the age estimate uncertainty budget is dominated by the measurement-derived uncertainties.

### 3. Data samples

#### 3.1. Field stars

For this study, multiple large-scale spectroscopic catalogues were considered, each providing metallicity and  $\alpha$  elements (as defined by each catalogue, and as  $[\text{Mg}/\text{Fe}]$  for GALAH DR4) abundances crossmatched with apparent magnitudes and parallaxes from the *Gaia* DR3 catalogue ([Gaia Collaboration et al. 2023](#)). The following selection criteria were applied:

```
phot_g_mean_flux_over_error > 50,
phot_bp_mean_flux_over_error > 20,
phot_rp_mean_flux_over_error > 20,
parallax_over_error > 10,
ipd_frac_multi_peak < 2,
ipd_gof_harmonic_amplitude < 0.1,
ruwe < 1.1,
parallax > 0.33.
```

The last condition was added to restrict our samples to stars within a 3 kpc volume around the Sun, as this is the region where we can estimate interstellar extinction with our method. Parallaxes were corrected for zero-point following [Lindgren et al. \(2021\)](#). The following spectroscopic catalogues were used:

- The LAMOST survey Data Release 10<sup>6</sup> ([Cui et al. 2012](#); [Zhao et al. 2012](#)), containing 7.4 million stars in the Low-Resolution Survey (LRS) and 2.2 million stars in the Medium-Resolution Survey (MRS). After filtering stars excluding those with  $\sigma_{[\text{Fe}/\text{H}]} > 0.1$  dex and  $\sigma_{v_r} > 5$  km s<sup>-1</sup>, and once crossmatched with our *Gaia* dataset, this survey provides us with  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$  for 935K stars in the LRS and 88K stars in the MRS.
- The GALAH survey, both the Data Release 3<sup>7</sup> ([De Silva et al. 2015](#); [Buder et al. 2021](#)) and 4<sup>8</sup> ([De Silva et al. 2015](#); [Buder et al. 2025](#)), containing 588K and 917K stars, respectively. Both data releases were used, as they often provide different metallicity estimates for the same star, leading to different age estimates, as well as to highlight

<sup>6</sup> <https://www.lamost.org/dr10>

<sup>7</sup> <https://www.galah-survey.org/dr3/overview>

<sup>8</sup> <https://www.galah-survey.org/dr4/overview>

the jump in the number of datable stars in the DR4. After applying selection cuts ( $\text{snr\_c3\_iraf} > 30$ ,  $\text{flag\_sp} == 0$ ,  $\text{flag\_fe\_h} == 0$ , and  $\text{flag\_alpha\_fe} == 0$  for GALAH DR3 and  $\text{snr\_px\_ccd3} > 30$  and  $\text{flag\_sp} == 0$  for GALAH DR4),  $\sigma_{[\text{Fe}/\text{H}]} < 0.1$  dex, and crossmatching with our *Gaia* sample, we end up with 13K and 251K stars with spectroscopic, astrometric and photometric data.

- The Apache Point Observatory Galactic Evolution Experiment (APOGEE) Data Release 17 survey<sup>9</sup> (Abdurro’uf et al. 2022) containing spectroscopic information for 657K stars, of which we recover 147K after crossmatching with our *Gaia* dataset and filtering out stars with a bad bitmask ASPCAPFLAG flag (bit 23 for STAR\_BAD, bit 19 for M\_H\_BAD, and bit 20 for ALPHA\_M\_BAD), signal-to-noise ratio (S/N) < 50, and  $\sigma_{[\text{Fe}/\text{H}]} > 0.1$  dex.

To correct for interstellar extinction, we used the 3D extinction maps from Vergely et al. (2022), which cover a 3 kpc x 3 kpc x 0.8 kpc region around the Sun at a 10 pc resolution. We integrated these maps along the line of sight of each star to compute the total  $A_V$  extinction value. Because extinction both dims and reddens light, it is vital to implement it as faithfully to the ground truth as possible, as any deviation will inevitably affect age estimation. For a discussion of these effects and a comparison with Edenhofer et al. (2024) extinction maps, see Appendix F. The V-Band  $A_V$  values computed were translated into *Gaia* passband interstellar extinction values  $A_G$ ,  $A_{BP}$ , and  $A_{RP}$  using the relations given in Danielski et al. (2018), and *Gaia* eDR3 coefficients<sup>10</sup>. Outside the range of applicability for this relation, we used Wang & Chen (2019)’s relation. The CMDs and galactic coverages of each survey are presented in Appendix E.

As the grids used in this work all have solar  $\alpha$  element abundances, the global metallicity [M/H] was corrected to take the non-zero  $\alpha$  element abundances of stars into account by applying the Salaris et al. (1993) relation, with coefficients corresponding to the Caffau et al. (2011) solar abundances used by the BaSTI grid:

$$[\text{M}/\text{H}] = [\text{Fe}/\text{H}] + 0.76[\alpha/\text{Fe}].$$

For a comparison of ages with and without this rescaling step to study the influence of  $\alpha$  elements on age determination, see Appendix H. We note that we did not correct for different solar mixtures used by the spectroscopic catalogues, as we did for the different evolutionary grids in Section 2.2. This is motivated by the fact that most age catalogues we compare our results to do not apply these corrections – even those using stellar evolutionary models that do not share the same solar mixtures as the spectroscopic surveys they use. Moreover, the LAMOST survey is based on a data-driven training approach, making a metallicity correction more difficult. Thus, we decided not to apply these corrections and instead made comparisons with metallicities as is.

The final uncertainties on the extinction-corrected  $M_G$  and  $(G_{BP} - G_{RP})_0$  were obtained by propagating the uncertainties on the apparent magnitudes in the  $G$ ,  $BP$ , and  $RP$  bands, the uncertainty of the *Gaia* parallaxes, and the uncertainty of the coefficients of the relations of Danielski et al. (2018) and Wang & Chen (2019). We note that these are lower bounds as no uncertainty on  $A_V$  is provided by the interstellar extinction maps. This set of uncertainties, along with  $\sigma_{[\text{M}/\text{H}]}$ , forms the vector  $\Delta\mathbf{x}_{\text{obs}}$

used to sample Monte Carlo realisations around the mean position  $\mathbf{x}_{\text{obs}}$  of each star, as described in Section 2.1.

### 3.2. Stellar clusters

In addition to field stars, we also estimate ages of open and globular clusters and compare them to other age estimates in the literature. To this end, we take the selection of clusters given in C24, which contains 13 open clusters and one globular cluster. We assume the same extinction values, metallicities, and distances, as defined in their work (see their Table 1). The literature ages are taken from Gaia Collaboration et al. (2018); Cantat-Gaudin et al. (2020); Netopil et al. (2022); Tsantaki et al. (2023).

## 4. Results

### 4.1. Age estimation of stellar clusters

Because clusters are usually thought of as coeval stellar populations, they offer a wide coverage in the CMD of what is considered as a single age isochrone sequence. Most isochrone fitting methods benefit from this fact and use each star member as a unique data point along a shared sequence to constrain the cluster age, although it is worth noting that this deceptively simple assumption does not hold for some globular clusters, which may contain multiple stellar populations of different ages and chemical compositions (see Bastian & Lardo (2018) for a review). Our method, in contrast, only considers, by construction, single independent stars when dating clusters, as our NNs have been built to receive single star inputs. While this is useful if one is interested in investigating separate populations within a cluster, we need a way to combine the individual age estimates produced to form a cluster age estimate.

Mathematically speaking, the global cluster age PDF is defined as the product of the individual star age PDFs. However, in practice, not all PDFs have a non-zero contribution at the true cluster age, such that a few individual PDFs can make the global product of PDFs vanish. This problem was mentioned in Jørgensen & Lindegren (2005), who solved it by adding a small  $\varepsilon$  contribution to all PDFs so that their product did not involve multiplication by zero:

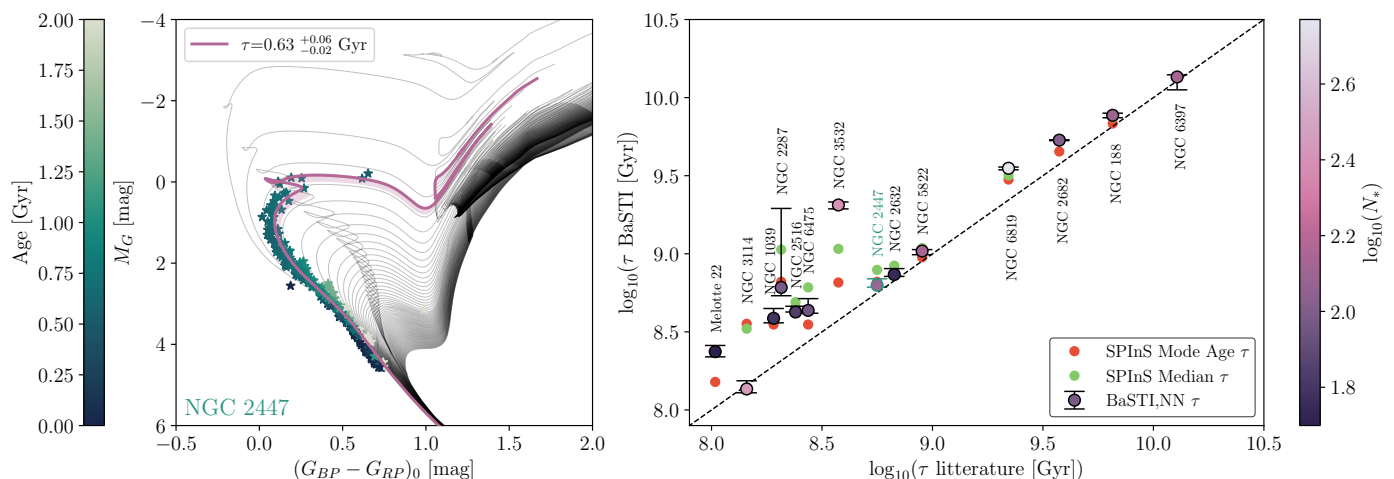
$$p_{\text{Cluster}}(\tau|\mathbf{x}_{\text{obs}}) = \prod_{n=1}^{N_{\star}} (p_i(\tau|\mathbf{x}_{\text{obs},i}) + \varepsilon). \quad (1)$$

In some pathological cases, if the cluster CMD sequence is broad enough or has enough outliers (e.g. binaries or blue stragglers), the computed PDF can still vanish. To remedy this, we employed a bootstrapping approach, where we computed  $p_{\text{Cluster},j}$  from a random sample of our  $N_{\star}$  cluster members, with replacement. This step was iterated a number of times (in our case 100) to produce a set of  $p_{\text{Cluster},j}$  containing some non-zero PDFs when enough outliers had been discarded. The cluster age was then taken to be the median of the peaks of these well-defined  $p_{\text{Cluster},j}$ .

The age PDFs of each bootstrap step generally peak at a single value, such that  $p_{\text{Cluster},j} \sim \delta(\tau - \tau_{\text{Cluster},j})$ , providing no information on the global uncertainty of this estimate. Thus, we estimated the age uncertainty of the cluster by computing the standard deviation of the distribution of the peaks of our  $p_{\text{Cluster},j}$ . This value, providing an estimate of the random variations in the cluster age given a subsample of its members, is usually very small and should not be interpreted as a total error estimate. In

<sup>9</sup> <https://www.sdss4.org/dr17/irspec>

<sup>10</sup> <https://www.cosmos.esa.int/web/gaia/edr3-extinction-law>



**Fig. 2.** *Left panel:* CMD of cluster NGC 2447 members (star symbols, coloured by our age estimates). Superimposed are BaSTI isochrones in black and the isochrone corresponding to the cluster age given by our method in pink, with its given uncertainties represented by the filled area. *Right panel:* Age comparison of our (pink symbols) and Casamiquela et al. (2024)’s age estimates (mode in red and median in green) and literature ages as defined in C24, coloured by the number of members. NGC 2447 is highlighted in cyan.

fact, the biggest source of error would be systematic, from the physics of the chosen evolutionary grid, the extinction and distance computations, as well as metallicity and  $\alpha$  elements abundance estimations, all of which were not included in this random error estimate.

The results of the cluster age estimation are presented in Fig. 2. In the left panel, we give an example of a cluster (NGC 2447) sequence in a CMD, where each cluster member is coloured by its estimated age. We over-plot the isochrone age at  $\tau=0.63$  Gyr. This age is very close to other ages given for this cluster in the literature (e.g. Cantat-Gaudin et al. (2020) give an age of 0.575 Gyr; Hunt & Reffert (2023) an age of 0.74 Gyr; Cavallo et al. (2024) an age of 0.63 Gyr; and Bossini et al. (2019) an age of 0.56 Gyr). Indeed, the isochrone closely follows the cluster sequence, from the MS to the MSTO.

In the right panel of Fig. 2, our age estimates are compared to those of C24, using their mode (red) and median (green) age estimates, and to ages from the literature, as defined in C24. Overall, we find good agreement between our ages and literature ages, particularly for clusters with  $\tau \geq 1$  Gyr. For younger clusters, we find the same trends as in C24, i.e. our ages are higher than those found in the literature, for which the relative difference of estimates can be higher. We also find that this effect can be explained by stars in the MS, which tend to have overestimated ages compared to the cluster age and MSTO stars. The former occupy a more isochrone-dense region of the CMD, where age estimates produce more uniform PDFs during the Monte Carlo sampling process, leading to higher individual ages.

NGC 2287 has a higher uncertainty than other clusters, as its CMD sequence does not follow a single isochrone. In the MS, its members overlap with isochrones of age 1 to 1.5 Gyr, while in the MSTO, its members more closely follow isochrones of age  $\sim 0.4$  Gyr. The literature age is in fact  $\sim 200$  Myr, much younger than our age estimates for MS members, but closer to our age estimate for its MSTO members. A similar situation is found in NGC 3532, where the MS aligns with isochrones of  $\sim 2$  Gyr, whereas MSTO align with isochrones of age  $< 0.5$  Gyr, closer to the literature age of  $\sim 350$  Myr. In this case, the low uncertainty is driven by two probably spurious stars on the  $\tau = 14$  Gyr isochrone, far from the cluster sequence, driving the computed  $p_{\text{Cluster}}(\tau|\mathbf{x}_{\text{obs}})$  to a narrow peak at 2 Gyr. This discrepancy can

also be seen in C24 (Fig. 5), where both clusters have a long tail in their PDF towards higher ages.

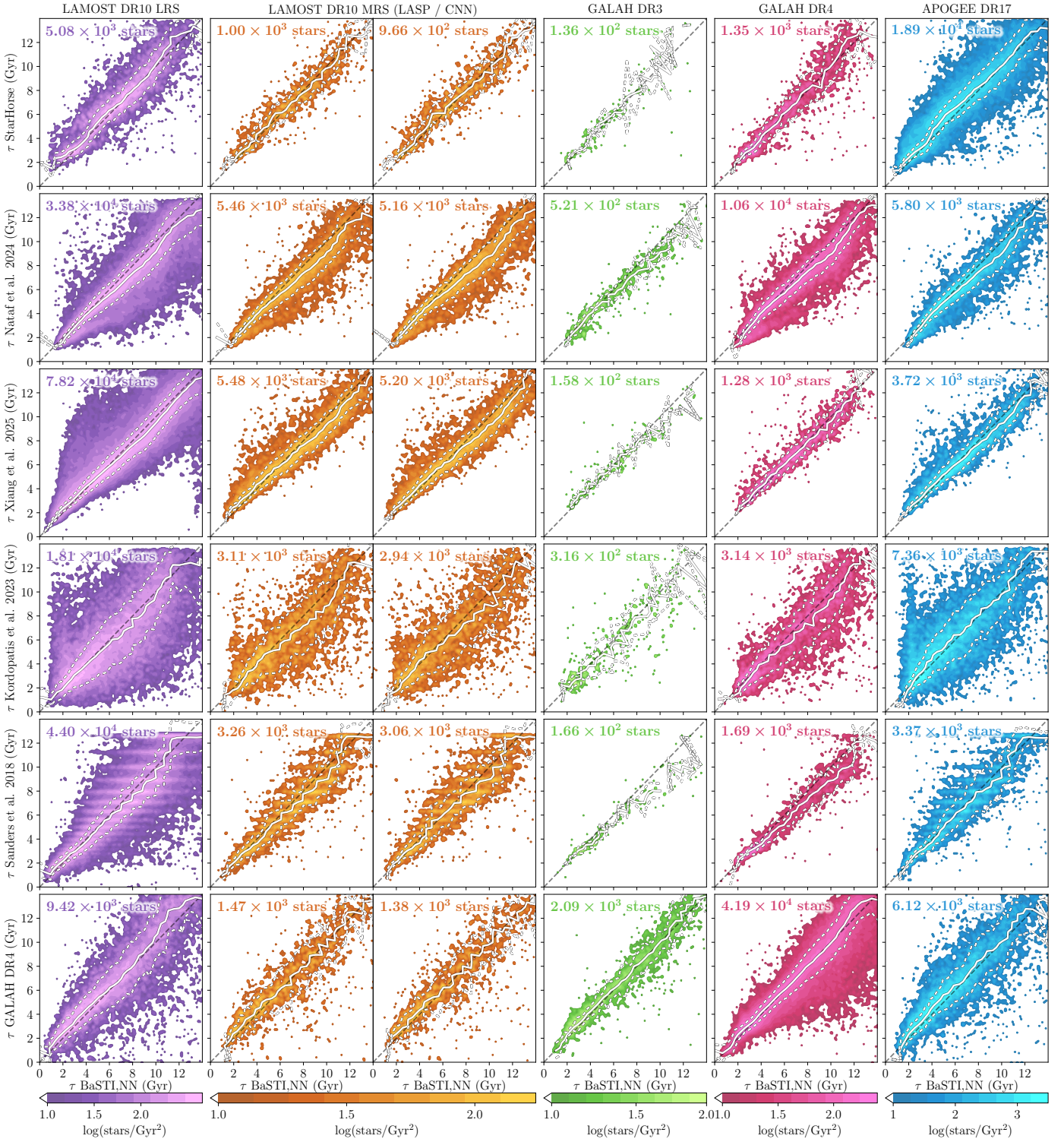
For clusters of ages  $\log_{10}(\tau[\text{Gyr}]) > 9$ , our age estimates agree with both C24’s and literature results, even if they are still all larger. On all clusters, we find a MAD between our estimates and the mean literature age of 0.20 Gyr, compared to a MAD of 0.27 Gyr for C24’s results with SPInS.

#### 4.2. Age estimation of field stars

We then applied our NNs to the data samples of field stars described in Section 3. This is a more challenging task than dating clusters, as field stars are not part of a coeval population but rather individual, unrelated points in the CMD. Nevertheless, the NN provide an age value, whose precision can then be appreciated through the uncertainty estimated by the method described in Section 2.1. Henceforth, we restricted each sample of ages obtained to ‘golden’ samples by imposing the following criteria:

- The star’s position in the CMD, within its uncertainties ( $\sigma_{M_G}$ ,  $\sigma_{(G_{BP}-G_{RP})_0}$ ,  $\sigma_{[M/H]}$ ), should lie within the training domain of the NN, i.e. the set of evolutionary curves of the model used for training<sup>11</sup>.
- The relative age uncertainty  $\frac{\sigma_\tau}{\tau}$  should be less than 10%.
- The absolute age uncertainty  $\sigma_\tau$  should be less than 1 Gyr.
- The age estimate  $\tau$  should be between 0 and 15 Gyr.
- The star’s surface gravity  $\log g$  should be between 3.5 and 4.45. This criterion is enforced to restrict our comparison to stars for which we expect the isochrones to be separated and reliable enough for dating stars. On the lower end, we remove red giant stars, for which we know general isochrone fitting methods struggle, and on the higher end, we removed dwarfs in the MS where isochrone density was too high to provide a reliable age information. The bounds were set manually to minimise contamination in our samples and maximise the number of remaining stars.

<sup>11</sup> The training domain is computed by building a grid where each cell is populated by stellar tracks passing through it. Convex hulls are not suitable, as tracks have concave shapes, and we experimented with  $\alpha$  shapes but found the discretized representation gave better results in thin areas such as the MS.



**Fig. 3.** Age comparison of the ‘golden’ samples as defined in Section 4.2 used in this work with age estimates from the literature. From top to bottom: StarHorse (Queiroz et al. 2023), Nataf et al. (2024), Xiang et al. (2025), Kordopatis et al. (2023), and Sanders & Das (2018). The number of stars is shown in each panel in the top left corner, a one-to-one dashed line is provided to guide the eye. The running median and  $\pm 1\sigma$  are plotted in solid white and dashed lines, respectively.

For each NN trained on a different evolutionary model, we thus obtained an age catalogue for our stellar samples. In this section, we compare the ages we obtained with the NN trained on the BaSTI grids to other age estimates from the literature:

- The StarHorse catalogue (Queiroz et al. 2023), which uses stellar photometry and PARSEC isochrones to estimate ages

based on a Bayesian approach. It provides ages for stars in the APOGEE, *Gaia*-ESO, GALAH, and RAVE catalogues, for a total of more than 10 million stars.

- The Nataf et al. (2024) (N24) catalogue of 401K stars dated through isochrone fitting using the MIST isochrone set.
- The Xiang et al. (2025) (X25) catalogue of 320K stars, which extends the Xiang & Rix (2022) catalogue of 247K stars,

- also dated through an isochrone fitting procedure, using the  $Y^2$  isochrone (Yi et al. 2001; Demarque et al. 2004) set.
- The Kordopatis et al. (2023) (**K23**) catalogue of  $\sim 5.5$  million stars dated through isochrone fitting using the PARSEC and COLIBRI (Marigo et al. 2013; Rosenfield et al. 2016; Pastorelli et al. 2020) sets of isochrones.
  - The Sanders & Das (2018) (**S18**) catalogue of 3.7 million stars from the APOGEE, *Gaia*-ESO, GALAH, LAMOST, RAVE, and SEGUE spectroscopic surveys, through a Bayesian approach and PARSEC isochrones.
  - The GALAH DR4 (Buder et al. 2025) (**GDR4**) catalogue of stellar ages, providing ages for 900K stars through a Bayesian approach and PARSEC isochrones.

The results are shown in Fig. 3, where each column represents one of our data samples and each row is one of the age catalogues used for the age comparison. The number of stars cross-matched between each sample and each catalogue are displayed in each panel. The running median is shown as a white line and was computed as  $\text{med}(\tilde{x} = \frac{x-y}{\sqrt{2}})$  in diagonal bins  $\tilde{y}_i < \tilde{y} \leq \tilde{y}_{i+1}$  with  $\tilde{y} = \frac{x+y}{\sqrt{2}}$ . The same was done for the 16th and 84th percentiles, shown in dashed white lines. Indeed, computing the running median in bins of  $x$  or  $y$  instead can result in boundary issues. For example, at  $13 \text{ Gyr} < x \leq 14 \text{ Gyr}$ , the age distribution along the  $y$  axis is not symmetrical, with most ages between 13-14 Gyr – some below but very few above. Thus, the median is biased towards younger ages; similarly, but in the opposite direction, for running medians computed in  $y$  bins. Computing the median along the diagonal direction instead ensures the result is independent of the bins chosen to compute it and unbiased close to the boundaries.

Overall, our method provides age estimates in good agreement with ages from the literature. **K23** ages, while being in good agreement in the most dense regions for all of our samples shows the highest standard deviation, with a median standard deviation of 1.61 Gyr marginalised on all our samples. **GDR4**'s age estimates also show a relatively high standard deviation on some samples with a median standard deviation of 1.21 Gyr, marginalised on all ages and samples. For **S18**, there are clear differences at the high end of the age range, with an absence of any ages  $\tau > 13 \text{ Gyr}$ . **X25** shows the lowest standard deviation, with a median value of 0.79 Gyr. All of our catalogues show good agreement with our results. This is especially true for the **X25** catalogue, which displays a very low to zero offset and the lowest standard deviation. LAMOST DR10 MRS (both LAMOST Stellar Pipeline (LASP) and convolutional neural network (CNN)), consistently display the highest dispersion across all of the samples and the greatest divergence from the one-to-one relation – with age estimates for the oldest ages higher by a few gigayears than most catalogues. For **X25** and **N24**, the median diverges at ages  $> 5 \text{ Gyr}$ . While for other catalogues, the median has very little offset from the one-to-one relation. Overall, the fact that the age catalogues and the different samples generally agree and have a low systematic difference is surprising, as one would expect the variety of methods, evolutionary tracks, extinction and distance estimates used to have a strong effects on the age estimates.

### 4.3. Age-chemistry relations

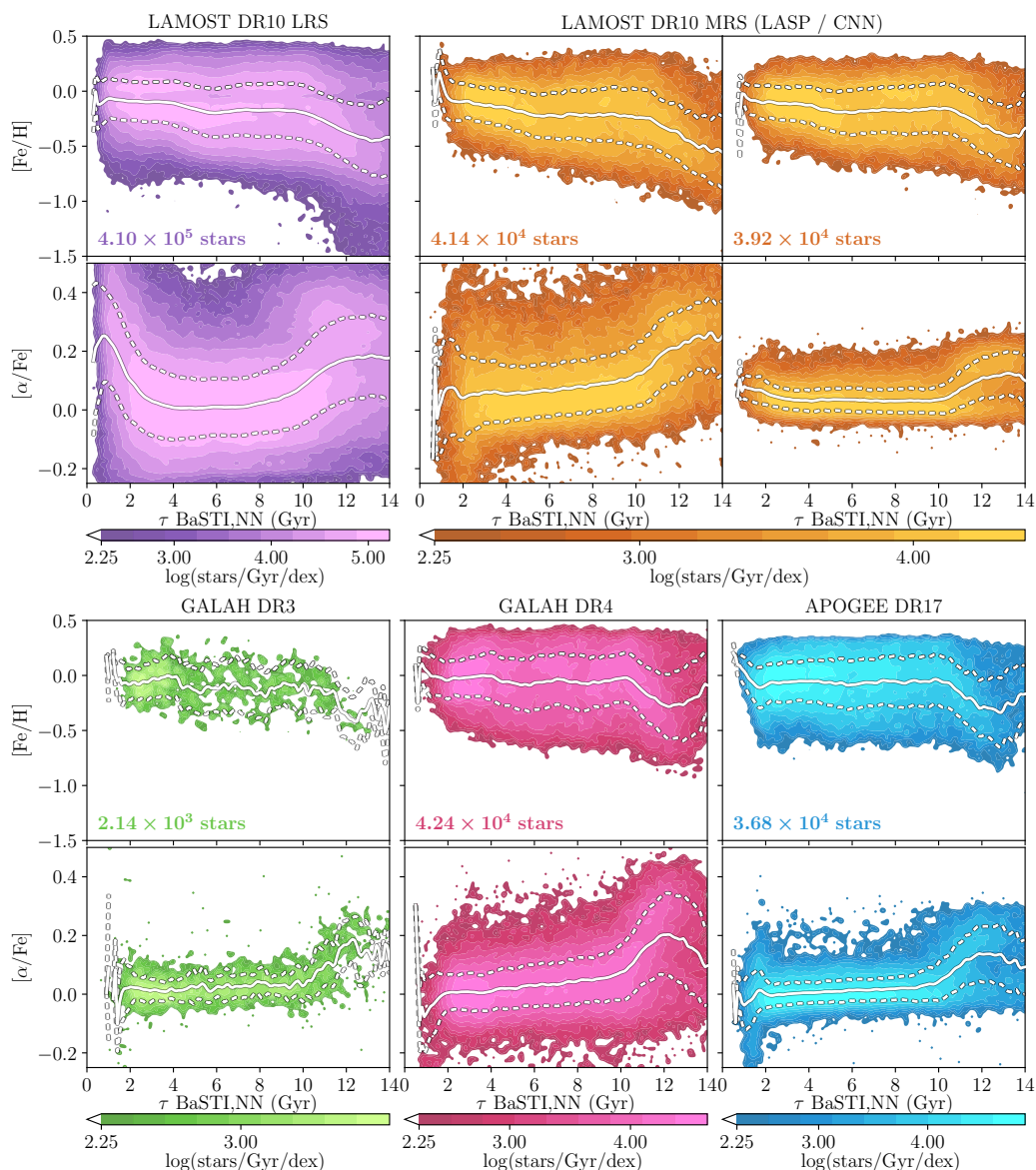
Using the golden samples defined in the previous section and the spectroscopic abundances, we plot the age-abundance diagrams in Fig. 4, with age versus metallicity in the top row and age versus  $\alpha$  element abundance in the lower row. Interestingly,

the age-metallicity distributions do not share the same patterns. Some, such as LAMOST DR10 MRS (LASP and CNN) and GDR4, display two sequences at ages  $\tau \lesssim 10 \text{ Gyr}$ , commonly seen in other studies of our Galaxy's age-metallicity relation (AMR) (Haywood et al. 2013; Bensby et al. 2014; Xiang & Rix 2022; Cerqui et al. 2025). For ages  $\tau \gtrsim 10 \text{ Gyr}$ , they show a sharp decline compared to the flat metallicity for younger ages, more prominently shown in LAMOST DR10 MRS (LASP and CNN) and GDR4, seen albeit more faintly in LAMOST DR10 LRS. We note that the under-density at all ages at  $[\text{Fe}/\text{H}] \sim -0.1$  for LAMOST DR10 LRS is not a result of our selection cuts but seems to be inherent to this specific data release of the LAMOST survey. The distribution for GALAH DR3 is more noisy on account of the lower number of stars; however, the two sequences are still visible.

Similarly, the age- $\alpha$  distributions, while sharing similar features, vary from sample to sample. In all our samples, a clear thick-disk sequence can be seen at ages  $\tau \gtrsim 10 \text{ Gyr}$ , with a sharp increase in  $[\alpha/\text{Fe}]$  from  $\sim 0$  up to 0.4 – particularly for a number of stars in the GDR4 sample, which show the sharpest increase in  $\alpha$  element abundance. We note a similar over-density of  $\alpha$ -rich stars ( $> 0.1$ ) in LAMOST DR10 LRS, MRS LASP, and APOGEE DR17 for stars with ages  $0 \text{ Gyr} \lesssim \tau \lesssim 4 \text{ Gyr}$ . We note that the definition of the  $\alpha$ -element content, which we took directly from the catalogues and as  $[\text{Mg}/\text{Fe}]$  for GDR4, varies from survey to survey, explaining the stark differences in  $[\alpha/\text{Fe}]$  distributions (e.g. between APOGEE DR17 and GDR4 or LAMOST DR10 LRS). Nevertheless, the difference between LAMOST DR10 MRS LASP and LAMOST DR10 MRS CNN is striking (as also seen in C24), considering the two share the same set of observed stars and differ only in their metallicity estimates.

Imposing stricter cuts on the age uncertainty by for example filtering out stars with  $\sigma_\tau > 0.3 \text{ Gyr}$  reveals clearer sequences for LAMOST DR10 (LRS and MRS), as well as for APOGEE DR17. The number of stars with solar-value  $\alpha$  element abundances with ages  $\tau \gtrsim 10 \text{ Gyr}$  is greatly reduced, making the sequences both in the  $[\text{Fe}/\text{H}]$ -age and the  $[\alpha/\text{Fe}]$ -age relations thinner. However, as the age error generally increases with age, a strict criterion on absolute rather than relative age uncertainty greatly limits the analysis of older populations. For example, when imposing the  $\sigma_\tau > 0.3 \text{ Gyr}$  criterion, we lose most stars with  $\tau > 10 \text{ Gyr}$  in our GALAH DR3 and DR4 sample.

The distributions in the age-abundance planes of our six samples seem to vary drastically, for example, between the broad LAMOST DR10 LRS and the thin LAMOST DR10 MRS CNN distribution. However, taking the mean  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$  for each age bin reveals similar trends. In Fig. 5, we plot the mean  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$  of each of our samples, with age estimates from the NN trained on the BaSTI grid. We find good agreement between the different samples, both in the AMR and in the  $[\alpha/\text{Fe}]$ -age relation. At young ages  $\tau < 1 \text{ Gyr}$ , the samples diverge from each other, mostly due to the very low number of stars. For ages  $\tau > 1 \text{ Gyr}$ , the mean relations closely follow each other, with a flat to shallow sequence from 1 to 10 Gyr in the AMR, with an initial metallicity of -0.1 for all samples at 2 Gyr, and a final metallicity between -0.2 and -0.1 at 10 Gyr. For older ages  $\tau > 10 \text{ Gyr}$ , the AMRs show a steeper slope, reaching metallicities between -0.2 and -0.5 at 14 Gyr. For GALAH DR3 and DR4, metallicity drops for ages between 5 and 9 Gyr; this can also be seen – albeit faintly – in the LAMOST DR10 samples. The six AMRs share the same knee at 10 Gyr, separating the two sequences. The same behaviour is seen in the  $[\alpha/\text{Fe}]$ -age relations, with an initial divergence of the samples at young ages



**Fig. 4.** Abundance vs age planes for our samples, with ages estimated with the NN trained on the BaSTI grid. *Odd rows:* Age vs  $[M/H]$ . *Even rows:* Age vs  $[\alpha/Fe]$ . The number of stars in each sample is annotated in the corners of the upper panels. The lighter colours correspond to higher density regions, and the means and  $\pm 1\sigma$  of the distributions are plotted as solid white and dashed lines, respectively.

$\tau < 1$  Gyr, followed by a shallow to flat sequence between 1 and 10 Gyr. All samples except LAMOST DR10 MRS LASP follow a very tight sequence centred at  $[\alpha/Fe] = 0.025$ , while LAMOST DR10 MRS LASP is offset by  $\sim 0.05$  dex. At older ages  $\tau > 10$  Gyr, the relations show steeper slopes, reaching  $[\alpha/Fe]$  abundances between 0.1 and 0.2. For a comparison of the  $[Fe/H]$ -age and  $[\alpha/Fe]$ -age relations from our APOGEE DR17 sample (with ages estimated by our NNs trained on different models), see Appendix G.

## 5. Discussion

Our approach is one of many to provide stellar age estimates. Compared to traditional isochrone fitting procedures and Bayesian methods, ours is generally much faster, owing to the efficient representation of evolutionary models by a small number of neurons. The advantage of building an embedding space of the model, rather than relying on the model itself, is twofold:

first, this leads to fewer overall operations to estimate ages from observational inputs; second the NNs were trained to build this embedding, and as such the initial models are not needed for age determination once the networks are trained. This results in significantly faster estimations, even when accounting for the N Monte Carlo needed to compute the related uncertainties. Compared to SPInS (Lebreton & Reese 2020), which takes about 20 seconds to calculate the age of a single star, our approach can estimate the ages of 60,000 stars in the same amount of time (together with 1,000 Monte Carlo samples for each star). We stress that the comparison made is not one-to-one as SPInS provides full PDFs for each of the inferred parameters. However, if one is only interested in an age estimate, for which our method gives similar results, deep learning techniques have a clear advantage. In addition, the computation time is dependent on the HRD region in SPInS, as the convergence time of the MCMC walkers towards a final age PDF varies. This runs contrary to our deep learning approach, which has a constant number of numerical

calculations, independently of the inputs and uncertainties. To our knowledge, this is the first time a general age comparison has been made between different stellar evolutionary models, mainly thanks to the speed of our deep learning approach.

As we demonstrated in Section 2.2 and in the first panel of Fig. 1, our models are in fact able to build a faithful embedding of evolutionary grids. By construction, the NNs are able to interpolate between the learned tracks, which, in traditional isochrone fitting procedures, is an additional non-trivial task (Jørgensen & Lindegren 2005; Lebreton & Reese 2020; Aguirre Børsen-Koch et al. 2022) for which several approaches are possible. In this work, out of all fundamental stellar parameters, we focused solely on age, contrary to other methods based on evolutionary grids, which can also provide estimates for stellar mass and radius, as well as other properties such as metallicity, extinction, and distance. As this study aims to show deep learning’s potential for handling ever-larger datasets quickly and reliably, we focused on age, as it is arguably the most important fundamental parameter in Galactic archaeology. However, as mass and radius are also provided by evolutionary tracks, one could extend our approach to build an NN trained to predict those parameters as well, with a minor impact on computational time.

Due to our choice of extinction maps, the age estimates of this work are confined to stars within a 3 kpc cylindrical radius around the Sun, which is the volume of the high-resolution maps of Vergely et al. (2022). Our analysis could be extended to higher distances from the Sun by using the lower-resolution maps of, for example, Vergely et al. (2022), or Edenhofer et al. (2024). We stress that the results would vary, as ages depend on  $A_V$  estimates (see Appendix F); however, the conclusions should remain similar as we did not observe any overall offset between age estimates using different extinction maps.

Upcoming surveys such as WHT Enhanced Area Velocity Explorer (WEAVE) (Jin et al. 2024), Sloan Digital Sky Survey-V (SDSS-V) (Kollmeier et al. 2025), and 4-metre Multi-Object Spectroscopic Telescope (4MOST) (de Jong et al. 2019) will provide an-order-of-magnitude increase in stellar element abundances, enabling age estimates and more comprehensive studies of our Galaxy’s history and evolution. A deep learning approach will provide fast and reliable age estimates of the large datasets these surveys will deliver.

Other catalogues, such as Zhang et al. (2023); Anders et al. (2022); Andrae et al. (2023); Khalatyan et al. (2024); Martin et al. (2024); Guiglion et al. (2024); Li et al. (2024), provide stellar atmospheric parameters for a very large number of stars by leveraging machine learning methods for their estimation (Anders et al. 2023b). As an example, Zhang et al. (2023) trained their model on data from the LAMOST survey to model *Gaia* XP spectra, which they then applied to the 220 million available GDR3 XP spectra to obtain estimations of their stellar atmospheric parameters. Notably, their catalogue does not contain  $[\alpha/\text{Fe}]$  estimates. However, as shown in Appendix H,  $\alpha$  element abundances can have a non-negligible effect on age estimations, especially for older stars for which Salaris et al. (1993)’s rescaling relations of metallicity can induce a shift of several 0.1 dex. For younger stars, whose  $\alpha$  element contents are close to solar, we expect our method to provide reliable age estimates using  $[\text{Fe}/\text{H}]$  rather than the rescaled  $[\text{M}/\text{H}]$ . As such, open clusters, which contain young stellar populations, can benefit from catalogues such as Zhang et al. (2023), which provide broad isochrone sequences at well-known metallicities.

Finally, we highlight the work of Garraffo et al. (2021), which also developed a model-driven deep-learning age-determination method based on MIST evolutionary tracks. How-

ever, theirs was based on the  $T_{\text{eff}}$ -luminosity-age relation and was only applied to solar-metallicity stars as it did not account for metal content. Hon et al. (2024) also trained generative models on MESA stellar tracks (Paxton et al. 2011, 2013, 2015, 2018, 2019). They used asteroseismic observables rather than photometric data as input, aiming to produce an emulator for evolutionary tracks.

The NNs developed for this work are made available through NEST, a Python package, as well as through a web interface. The NEST package includes individual age estimations computed using the method described in Section 2.1, as well as cluster age estimations using the method described in Section 4.1. The left panel of Fig. 2 was plotted using its `HR_diagram()` method, plotting a set of isochrones as well as an interpolated isochrone of the cluster age.

## 6. Conclusions

In this paper, we introduced a model-driven deep-learning approach to stellar age determination. We trained deep NNs on stellar evolution models using *Gaia* photometry and metal and  $\alpha$  element abundances to predict ages.

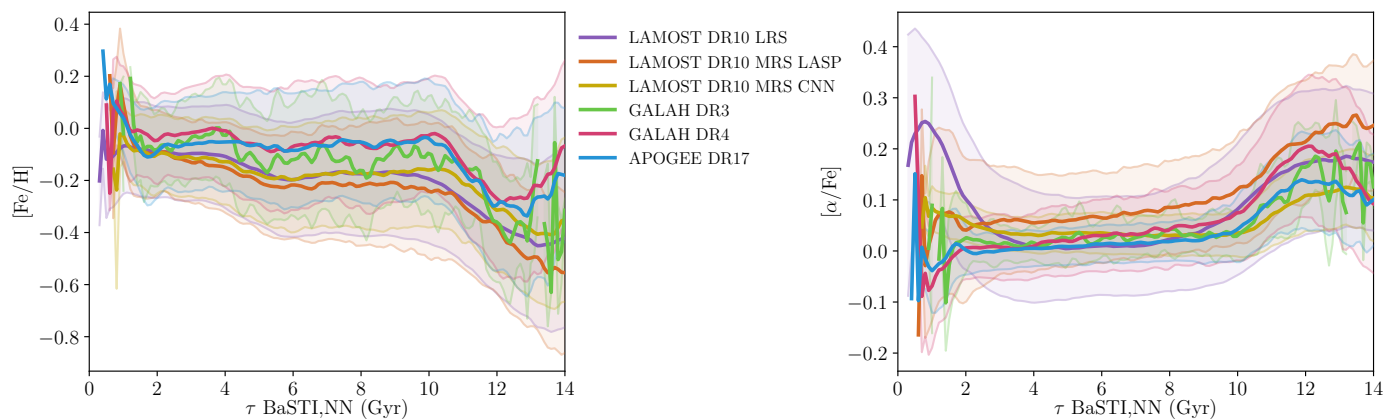
We trained multilayer perceptrons on different stellar grids (BaSTI, MIST, PARSEC, Dartmouth, and Geneva) until their loss function converged. The resulting MAD on the test data is  $\sim 35$  Myr, confirming the faithful representation of the model by the network. Following Jørgensen & Lindegren (2005), we used Monte Carlo sampling to estimate uncertainties on our age estimates.

Contrary to traditional isochrone fitting methods, ours is generally much faster, owing to the low computational cost of NNs, with an estimated 60,000 speed-up compared to an MCMC procedure used in, for example, the SPInS pipeline. Other deep learning approaches, which are generally data-driven (i.e. trained on a dated subset of stars), typically suffer from biases inherited from the training sample’s age estimation method and sample size and can be restricted to a subset of the available star samples. In comparison, our method only depends on the evolutionary tracks used and offers a broader range of application.

We compared age estimates made with NNs trained on different evolutionary tracks and find some disagreements between tracks, with PARSEC tracks yielding lower ages than BaSTI by  $\sim 1 - 2$  Gyr, but relatively low overall dispersion. Geneva tracks show a much higher divergence, with old ages being lower than SPInS ones by as much as 4 Gyr. When comparing our method using BaSTI tracks with the results of C24 using the SPInS pipeline (Lebreton & Reese 2020) and BaSTI tracks, we find no offset for all ages except at very low ages, confirming the reliability of our method. Thus, we find that differences in age estimates appear mostly from using different evolutionary tracks and/or from using different input data, such as extinction estimates (see Appendix F).

We applied our NNs to stellar cluster populations, from which we derived a global cluster age by combining the individual age PDFs. We find that our cluster age estimates are in agreement with the literature for older ages, while our ages are systematically higher than the literature age for younger clusters. This effect was also seen in the analysis of Casamiquela et al. (2024), who also used the BaSTI grid. We find a MAD between our ages and literature values of 0.20 Gyr, compared to 0.27 Gyr for Casamiquela et al. (2024) using SPInS.

We then applied our NNs to large spectroscopic catalogues, namely LAMOST DR10, GALAH DR3, GDR4, and APOGEE DR17, crossmatched with *Gaia* photometry and parallaxes. The



**Fig. 5.** Abundance vs age trends for each of our samples, with ages estimated with the NN trained on the BaSTI grid. The coloured lines and shaded areas represent the means and  $\pm 1\sigma$  of the distributions, as shown in white in Fig. 4.

photometry was corrected for extinction using Vergely et al. (2022) 3D extinction maps. The resulting ages were compared to the other age catalogues of StarHorse (Queiroz et al. 2023), Nataf et al. (2024) (N24), Xiang et al. (2025) (X25), Kordopatis et al. (2023) (K23), Sanders & Das (2018) (S18), and Buder et al. (2025) (GDR4). We find overall good agreement, especially for the X25, N24, and StarHorse catalogues, where the running median of our age versus theirs shows low to no offset and low dispersion. K23 and GDR4 both show a relatively low offset, giving lower ages compared to ours for older ages, as well as higher median dispersions of 1.61 and 1.21 Gyr, respectively. The X25 catalogue shows the least dispersion, with a median value of 0.79 Gyr when comparing our ages and theirs, and very low differences in median age estimates. We find that LAMOST DR10 MRS displays the highest dispersion for all catalogues. Aside from divergences at older ages in the X25 and N24 catalogues, the running median closely follows the one-to-one relation for all other catalogues and samples.

Finally, we analysed the abundance-age relations of the aforementioned spectroscopic catalogues and find similar structures, with a thin and thick-disk sequence visible both in the  $[\text{Fe}/\text{H}]$ -age and  $[\alpha/\text{Fe}]$ -age planes. For stars younger than  $\sim 10$  Gyr, we find a flat metallicity and  $\alpha$  element distribution with age, and for older stars, we find a sharp decrease in metallicity and a sharp increase in  $[\alpha/\text{Fe}]$ . The epoch separating these two sequences is also noticeable when comparing ages estimated using metallicity scaling relations based on  $\alpha$  element content with those obtained without such scaling relations (see Appendix H). Given its flexibility and speed, this code opens unprecedented possibilities for estimating ages of millions of stars in upcoming spectroscopic surveys such as 4MOST (de Jong et al. 2019), SDSS-V (Kollmeier et al. 2025), and possible future ones such as WST (Mainieri et al. 2024).

Parallel to this work, the trained NNs are released as part of NEST<sup>12</sup>, a Python package to infer individual and population ages and estimate uncertainties. We also release a web interface<sup>13</sup> that incorporates the same trained NNs and an interactive CMD for dating stars, as well as tools for importing star populations to be dated.

<sup>12</sup> <https://github.com/star-age/NEST>

<sup>13</sup> <https://star-age.github.io>

## Data availability

Full catalogues are available at the CDS via anonymous ftp to [cdsarc.cds.unistra.fr](https://cdsarc.cds.unistra.fr) (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/000/A000>. The NEST python package is available at <https://github.com/star-age/NEST> and the web interface at <https://star-age.github.io>.

*Acknowledgements.* T.B. thanks the Ecole Doctorale Astronomie et Astrophysique d’Île de France for funding their thesis project. The authors warmly thank J. Bland-Hawthorn for discussions on surveys and their uncertainties.

This research or product makes use of public auxiliary data provided by ESA/Gaia/DPAC/CU5 and prepared by Carine Babusiaux, for computing the extinction in the Gaia passbands.

This work has made use of data from the European Space Agency (ESA) mission Gaia (<https://www.cosmos.esa.int/gaia>), processed by the Gaia Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

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.

This work made use of the Third Data Release of the GALAH Survey (Buder et al. 2021). The GALAH Survey is based on data acquired through the Australian Astronomical Observatory, under programs: A/2013B/13 (The GALAH pilot survey); A/2014A/25, A/2015A/19, A2017A/18 (The GALAH survey phase 1); A2018A/18 (Open clusters with HERMES); A2019A/1 (Hierarchical star formation in Ori OB1); A2019A/15 (The GALAH survey phase 2); A/2015B/19, A/2016A/22, A/2016B/10, A/2017B/16, A/2018B/15 (The HERMES-TESS program); and A/2015A/3, A/2015B/1, A/2015B/19, A/2016A/22, A/2016B/12, A/2017A/14 (The HERMES K2-follow-up program). We acknowledge the traditional owners of the land on which the AAT stands, the Gamilaraay people, and pay our respects to elders past and present. This paper includes data that has been provided by AAO Data Central ([datacentral.org.au](http://datacentral.org.au)).

This work made use of the Fourth Data Release of the GALAH Survey (Buder et al. 2021). The GALAH Survey is based on data acquired through the Australian Astronomical Observatory, under programs: A/2013B/13 (The GALAH pilot survey); A/2014A/25, A/2015A/19, A2017A/18 (The GALAH survey phase 1); A2018A/18 (Open clusters with HERMES); A2019A/1 (Hierarchical star formation in Ori OB1); A2019A/15, A/2020B/23, R/2022B/5, R/2023A/4, R2023B/5 (The GALAH survey phase 2); A/2015B/19, A/2016A/22, A/2016B/10, A/2017B/16, A/2018B/15 (The HERMES-TESS program); A/2015A/3, A/2015B/1, A/2015B/19, A/2016A/22, A/2016B/12, A/2017A/14, A/2020B/14 (The HERMES K2-follow-up program); R/2022B/02 and A/2023A/09 (Combining asteroseismology and spectroscopy in K2); A/2023A/8 (Resolving the chemical fingerprints of Milky Way mergers); and A/2023B/4 (s-process variations in southern globular clusters). We acknowledge the traditional owners of the land on which the AAT stands, the Gamilaraay people, and pay our respects to elders past and present. This paper includes data that has been provided by AAO Data Central ([datacentral.org.au](http://datacentral.org.au)).

This work has used data from the APOGEE DR17 catalogue. Funding for the

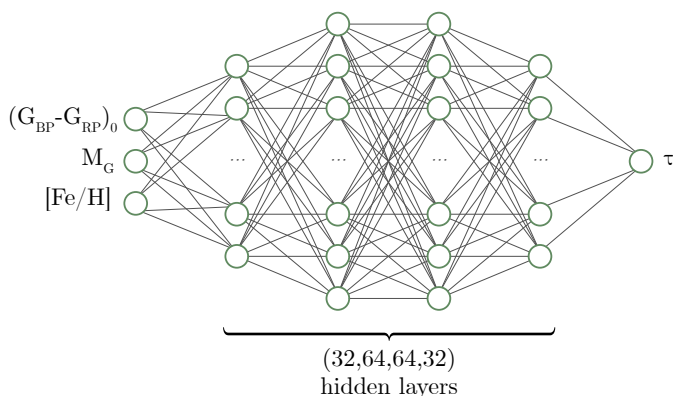
Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is [www.sdss4.org](http://www.sdss4.org). SDSS 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, Center for Astrophysics | Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University. The following python libraries were used for this study: *Astropy* (*Astropy Collaboration et al. 2013, 2018, 2022*), *scipy* (*Virtanen et al. 2020*), *numpy* (*Harris et al. 2020*), (*Pedregosa et al. 2011*) and *matplotlib* (*Hunter 2007*).

## References

- Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, *ApJS*, **259**, 35
- Aguirre Børsen-Koch, V., Rørsted, J. L., Justesen, A. B., et al. 2022, *MNRAS*, **509**, 4344
- Anders, F., Chiappini, C., Rodrigues, T. S., et al. 2017, *A&A*, **597**, A30
- Anders, F., Gispert, P., Ratcliffe, B., et al. 2023a, *A&A*, **678**, A158
- Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, *A&A*, **658**, A91
- Anders, F., Khalatyan, A., Queiroz, A. B. A., Nepal, S., & Chiappini, C. 2023b, *arXiv e-prints*, [arXiv:2302.06995](https://arxiv.org/abs/2302.06995)
- Andrae, R., Rix, H.-W., & Chandra, V. 2023, *ApJS*, **267**, 8
- Asplund, M., Grevesse, N., & Sauval, A. J. 2005, in *Astronomical Society of the Pacific Conference Series*, Vol. 336, *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis*, ed. Barnes, III, T. G. & Bash, F. N., **25**
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, **935**, 167
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Baglin, A., Auvergne, M., Boisnard, L., et al. 2006, in *36th COSPAR Scientific Assembly*, Vol. 36, **3749**
- Bastian, N. & Lardo, C. 2018, *ARA&A*, **56**, 83
- Bedding, T. R., Huber, D., Stello, D., et al. 2010, *ApJ*, **713**, L176
- Bensby, T., Adén, D., Meléndez, J., et al. 2011, *A&A*, **533**, A134
- Bensby, T., Feltzing, S., & Oey, M. S. 2014, *A&A*, **562**, A71
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, **327**, 977
- Bossini, D., Vallenari, A., Bragaglia, A., et al. 2019, *A&A*, **623**, A108
- Boulet, T. 2024, *A&A*, **685**, A66
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
- Bu, Y., Kumar, Y. B., Xie, J., et al. 2020, *ApJS*, **249**, 7
- Buder, S., Kos, J., Wang, X. E., et al. 2025, *PASA*, **42**, e051
- Buder, S., Sharma, S., Kos, J., et al. 2021, *MNRAS*, **506**, 150
- Caffau, E., Ludwig, H.-G., Steffen, M., Freytag, B., & Bonifacio, P. 2011, *Sol. Phys.*, **268**, 255
- Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, *A&A*, **640**, A1
- Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, *A&A*, **530**, A138
- Casamiquela, L., Reese, D. R., Lebreton, Y., et al. 2024, *A&A*, **692**, A243
- Cavallo, L., Spina, L., Carraro, G., et al. 2024, *AJ*, **167**, 12
- Cerqui, V., Haywood, M., Snaith, O., Di Matteo, P., & Casamiquela, L. 2025, *A&A*, **699**, A277
- Chaplin, W. J., Appourchaux, T., Elsworth, Y., et al. 2010, *ApJ*, **713**, L169
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
- Ciucă, I., Kawata, D., Miglio, A., Davies, G. R., & Grand, R. J. J. 2021, *MNRAS*, **503**, 2814
- Costa, G., Shepherd, K. G., Bressan, A., et al. 2025, *A&A*, **694**, A193
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Res. Astron. Astrophys.*, **12**, 1197
- Danielski, C., Babusiaux, C., Ruiz-Dern, L., Sartoretti, P., & Arenou, F. 2018, *A&A*, **614**, A19
- Das, P. & Sanders, J. L. 2018, *MNRAS*, **484**, 294
- de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, *The Messenger*, **175**, 3
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, **449**, 2604
- Demarque, P., Woo, J.-H., Kim, Y.-C., & Yi, S. K. 2004, *ApJS*, **155**, 667
- Dotter, A., Chaboyer, B., Jevremović, D., et al. 2007, *AJ*, **134**, 376
- Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, *ApJS*, **178**, 89
- Edenhofer, G., Zucker, C., Frank, P., et al. 2024, *A&A*, **685**, A82
- Ekström, S., Georgy, C., Eggenberger, P., et al. 2012, *A&A*, **537**, A146
- Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018, *A&A*, **616**, A10
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, *A&A*, **674**, A1
- García, R. A. & Ballot, J. 2019, *Living Reviews in Solar Physics*, **16**, 4
- Garraffo, C., Protopapas, P., Drake, J. J., Becker, I., & Cargile, P. 2021, *AJ*, **162**, 157
- Georgy, C., Ekström, S., Eggenberger, P., et al. 2013a, *A&A*, **558**, A103
- Georgy, C., Ekström, S., Granada, A., et al. 2013b, *A&A*, **553**, A24
- Georgy, C., Granada, A., Ekström, S., et al. 2014, *A&A*, **566**, A21
- Goupil, M. J., Catala, C., Samadi, R., et al. 2024, *A&A*, **683**, A78
- Grevesse, N. & Sauval, A. J. 1998, *Space Sci. Rev.*, **85**, 161
- Guiglion, G., Nepal, S., Chiappini, C., et al. 2024, *A&A*, **682**, A9
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Hayden, M. R., Sharma, S., Bland-Hawthorn, J., et al. 2022, *MNRAS*, **517**, 5325
- Haywood, M., Di Matteo, P., Lehnert, M. D., Katz, D., & Gómez, A. 2013, *A&A*, **560**, A109
- Hidalgo, S. L., Pietrinferni, A., Cassisi, S., et al. 2018, *ApJ*, **856**, 125
- Hon, M., Huber, D., Kuszewicz, J. S., et al. 2021, *ApJ*, **919**, 131
- Hon, M., Li, Y., & Ong, J. 2024, *ApJ*, **973**, 154
- Howes, L. M., Lindegren, L., Feltzing, S., Church, R. P., & Bensby, T. 2019, *A&A*, **622**, A27
- Hunt, E. L. & Reffert, S. 2023, *A&A*, **673**, A114
- Hunter, J. D. 2007, *Computing in Science & Engineering*, **9**, 90
- Jin, S., Trager, S. C., Dalton, G. B., et al. 2024, *MNRAS*, **530**, 2688
- Jørgensen, B. R. & Lindegren, L. 2005, *A&A*, **436**, 127–143
- Khalatyan, A., Anders, F., Chiappini, C., et al. 2024, *A&A*, **691**, A98
- Kingma, D. P. & Ba, J. 2017, Adam: A Method for Stochastic Optimization
- Kollmeier, J. A., Rix, H.-W., Aerts, C., et al. 2025, *AJ*, **171**, 52
- Kordopatis, G., Schultheis, M., McMillan, P. J., et al. 2023, *A&A*, **669**, A104
- Lebreton, Y., Goupil, M. J., & Montalbán, J. 2014a, in *EAS Publications Series*, Vol. 65, *EAS Publications Series*, ed. Lebreton, Y., Valls-Gabaud, D., & Charbonnel, C. (EDP), **99–176**
- Lebreton, Y., Goupil, M. J., & Montalbán, J. 2014b, in *EAS Publications Series*, Vol. 65, *EAS Publications Series*, ed. Lebreton, Y., Valls-Gabaud, D., & Charbonnel, C. (EDP), **177–223**
- Lebreton, Y. & Reese, D. R. 2020, *A&A*, **642**, A88
- Li, J., Wong, K. W. K., Hogg, D. W., Rix, H.-W., & Chandra, V. 2024, *ApJS*, **272**, 2
- Lindegren, L., Bastian, U., Biermann, M., et al. 2021, *A&A*, **649**, A4
- Mainieri, V., Anderson, R. I., Brinchmann, J., et al. 2024, *arXiv e-prints*, [arXiv:2403.05398](https://arxiv.org/abs/2403.05398)
- Marigo, P., Bressan, A., Nanni, A., Girardi, L., & Pumo, M. L. 2013, *MNRAS*, **434**, 488
- Martin, N. F., Starkeburg, E., Yuan, Z., et al. 2024, *A&A*, **692**, A115
- Moya, A., Recio-Martínez, J., & López-Sastre, R. J. 2021, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1971–1981
- Nataf, D. M., Schlauffman, K. C., Reggiani, H., & Hahn, I. 2024, *ApJ*, **976**, 87
- Netopil, M., Oralhan, İ. A., Çakmak, H., Michel, R., & Karataş, Y. 2022, *MNRAS*, **509**, 421
- Nguyen, C. T., Costa, G., Girardi, L., et al. 2022, *A&A*, **665**, A126
- Pastorelli, G., Marigo, P., Girardi, L., et al. 2020, *MNRAS*, **498**, 3283
- Paxton, B., Bildsten, L., Dotter, A., et al. 2011, *ApJS*, **192**, 3
- Paxton, B., Cantiello, M., Arras, P., et al. 2013, *ApJS*, **208**, 4
- Paxton, B., Marchant, P., Schwab, J., et al. 2015, *ApJS*, **220**, 15
- Paxton, B., Schwab, J., Bauer, E. B., et al. 2018, *ApJS*, **234**, 34
- Paxton, B., Smolec, R., Schwab, J., et al. 2019, *ApJS*, **243**, 10
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, **12**, 2825
- Pinsonneault, M. H., Zinn, J. C., Tayar, J., et al. 2025, *ApJS*, **276**, 69
- Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2023, *A&A*, **673**, A155
- Rauer, H., Catala, C., Aerts, C., et al. 2014, *Experimental Astronomy*, **38**, 249
- Reyes, C., Stello, D., Hon, M., et al. 2025, *MNRAS*, **538**, 1720
- Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, *JATIS*, **1**, 014003
- Rosenfield, P., Marigo, P., Girardi, L., et al. 2016, *ApJ*, **822**, 73
- Salaris, M., Chieffi, A., & Straniero, O. 1993, *ApJ*, **414**, 580
- Sanders, J. L. & Das, P. 2018, *MNRAS*, **481**, 4093
- Soderblom, D. R. 2010, *ARA&A*, **48**, 581
- Tamames-Rodero, V., Moya, A., Sarro, L. M., & López-Sastre, R. J. 2025, *Astronomy and Computing*, **52**, 100957

- Tayar, J. & Joyce, M. 2025, *ApJ*, 984, L56
- Tishby, Levin, & Solla. 1989, in International 1989 Joint Conference on Neural Networks, 403–409 vol.2
- Tsantaki, M., Delgado-Mena, E., Bossini, D., et al. 2023, *A&A*, 674, A157
- Van-Lane, P., Speagle, J. S., & Douglas, S. 2023, *arXiv e-prints*, [arXiv:2307.08753](https://arxiv.org/abs/2307.08753)
- Vergely, J. L., Lallement, R., & Cox, N. L. J. 2022, *A&A*, 664, A174
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261
- von Hippel, T., van Dyk, D. A., Stenning, D. C., et al. 2014, in EAS Publications Series, Vol. 65, EAS Publications Series, ed. Lebreton, Y., Valls-Gabaud, D., & Charbonnel, C. (EDP), 267–287
- Wang, H.-F., Carraro, G., Li, X., et al. 2024, *ApJ*, 967, 37
- Wang, S. & Chen, X. 2019, *ApJ*, 877, 116
- Xiang, M. & Rix, H.-W. 2022, *Nature*, 603, 599–603
- Xiang, M., Rix, H.-W., Yang, H., et al. 2025, *Nature Astronomy*, 9, 101
- Yi, S., Demarque, P., Kim, Y.-C., et al. 2001, *ApJS*, 136, 417
- Yusof, N., Hirschi, R., Eggenberger, P., et al. 2022, *MNRAS*, 511, 2814
- Zhang, X., Green, G. M., & Rix, H.-W. 2023, *MNRAS*, 524, 1855
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *Res. Astron. Astrophys.*, 12, 723

## Appendix A: Neural network architecture



**Fig. A.1.** Architecture of the neural networks used in this work.

The general architecture of the neural networks used in this work is represented in Fig. A.1.

## Appendix B: Effective temperature-colour relation

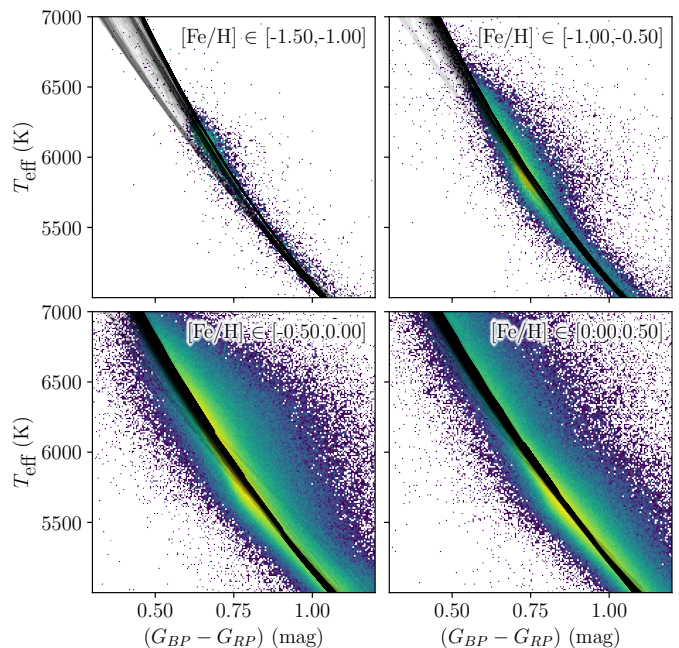
Fig. B.1 shows the effective temperature-colour relation, at different metallicities, for the LAMOST DR10 LRS sample, compared to the theoretical relation given by the BaSTI grid. The colour index is given by Gaia photometry ( $G_{BP} - G_{RP}$ ) and the effective temperature comes from the LAMOST catalogue. While the observational relation matches the theoretical one without any offset, in all metallicity bins, the spread is significant, meaning a non-negligible part of our sample has an estimated effective temperature that differs from the one our theoretical grid expects given the same colour index. At lower metallicities, some of the other catalogues can also display offsets from the theoretical relations (e.g. the GALAH DR4 catalogue at metallicities lower than 0.3).

### Dependence of age estimations on Monte Carlo samples

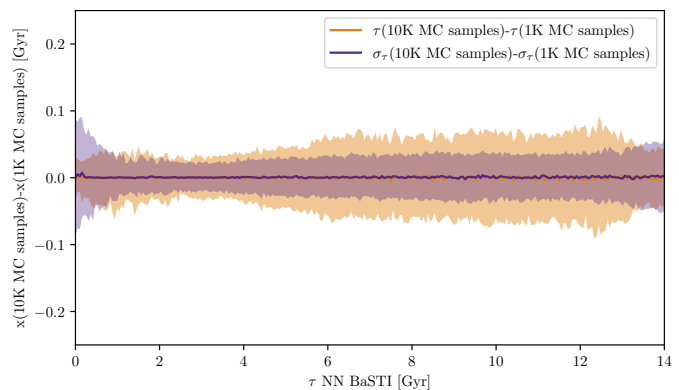
The choice of number of Monte Carlo samples to estimate the age PDF of a star is arbitrary, with a higher number resulting in a better convergence towards the true age distribution, as well as a higher computational cost. In Fig. B.2, we compare the age medians and associated standard deviations of GALAH DR4 stars, when using 10 000 and 1000 Monte Carlo samples. We find that the median difference of median ages, marginalized over all ages, is of  $-3.3 \times 10^{-5} \pm 0.06$  Gyr, while the median difference of standard deviations, marginalized over all ages, is of  $4.8 \times 10^{-4} \pm 0.03$  Gyr. Thus, for the vast majority of cases, 1000 Monte Carlo samples is enough to faithfully represent the underlying true age PDF, whereas some extreme cases necessitate more samples.

## Appendix C: Evolutionary model parameters

The different parameter spaces of each model used in this work are given in Table C.1. We note that the number of samples taken for metallicities, ages and masses are not the same in the different models, leading to a different number of points in each grid. The provided ranges of parameters help compare the domains of applicability of each model.



**Fig. B.1.** Effective temperature–colour index relation for the BaSTI stellar evolution grid (black curves) and stars from the LAMOST DR10 LRS catalogue (colourmap).



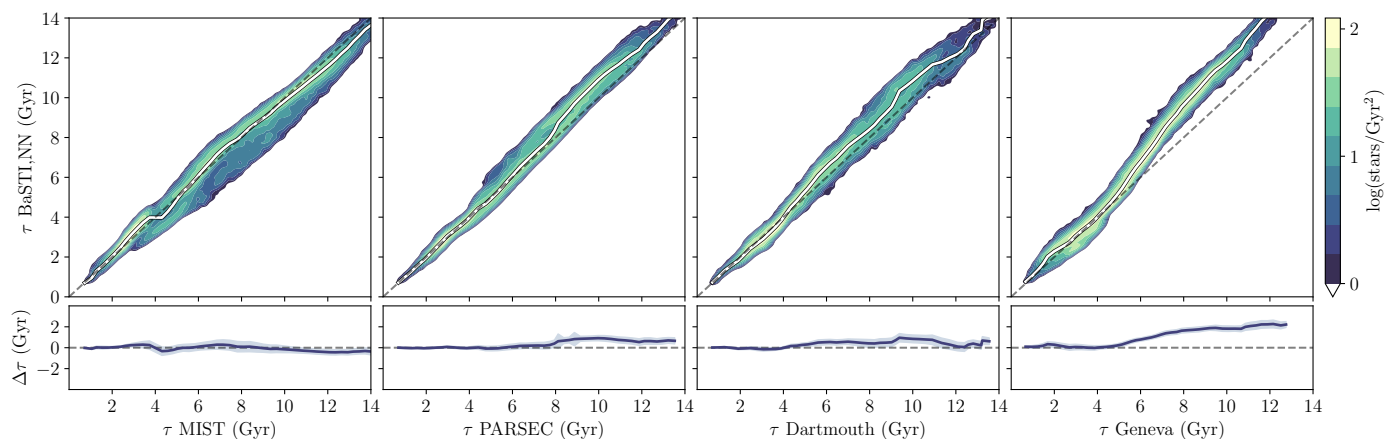
**Fig. B.2.** Age versus difference in median age (orange) and standard deviation (purple) estimates, using 1000 and 10000 Monte Carlo samples. Shaded areas show the 16th and 84th percentiles.

**Table C.1.** Parameter spaces for each stellar evolution model.

Model name	[M/H]	Age (Gyr)	M ( $M_{\odot}$ )
BaSTI	[-3.2,0.3]	$[5 \times 10^{-6}, 14]$	[0.1,10]
MIST	[-5.0,0.5]	[0.1,12.6]	[0.1,5.2]
PARSEC	[-0.9,0.3]	[0.1,13.9]	[0.1,5]
Dartmouth	[-2.6,0.7]	[0.3,13.5]	[0.1,3.8]
Geneva	[-1.5,0.2]	$[1 \times 10^{-2}, 12.6]$	[0.8,500]

## Appendix D: Comparison of different grid age estimates using GALAH DR4 stars

In this section, we compare age estimates made using neural networks trained on different stellar evolutionary models, as was done in Fig. 1, but this time using GALAH DR4 stars instead of C24's dataset. The resulting distributions are plotted in Fig. D.1, where age estimates made with a BaSTI-trained neural network are compared to age estimates made with neural net-



**Fig. D.1.** *Top panels:* Age estimates of GALAH DR4 stars from the NN trained on BaSTI compared to those from NNs trained on (from left to right) MIST, PARSEC, Dartmouth and SYCLIST evolutionary grids. Running medians are shown in white lines, as well as a one-to-one black dashed line to guide the eye. *Bottom panels:* Running median of the difference  $\Delta\tau$  between the age estimates of our BaSTI-trained NN and other NNs in solid lines, and associated standard deviations in shades areas.

works trained on the MIST, PARSEC, Dartmouth and Geneva grids. We find that BaSTI ages still agree with MIST ages over the whole age range. For Dartmouth ages, the standard deviation is greatly reduced from a median value over the whole age range of 0.40 Gyr for C24’s dataset to 0.27 Gyr for GALAH DR4. The highest deviation is seen at ages  $\sim 10$  Gyr, that was not seen on C24’s dataset, although still compatible with it within  $1\sigma$ . For Geneva ages, the trends are similar with GALAH DR4 and C24 datasets, although the former displays lower standard deviations, from a median value of 0.31 Gyr to 0.22 Gyr. The PARSEC ages show the biggest difference with C24’s dataset. With GALAH DR4 stars, the BaSTI and PARSEC ages show a better agreement, with very little offset for ages  $< 8$  Gyr. For older ages, the difference is of about 1 Gyr, and 0.5 Gyr for ages  $> 12$  Gyr. The standard deviation is also lower, with a median value of 0.26 compared to 0.52 Gyr using C24’s dataset. The difference in grid comparisons can mainly be explained by the CMD coverage of the sample used. In the case of C24’s dataset, used in Fig. 1, older stars are distributed closer to the Zero Age Main Sequence, where isochrone degeneracy is high, while in the case of GALAH DR4 stars, used here, older stars tend to populate further regions along those isochrones, where degeneracy is lower.

## Appendix E: Spectroscopic surveys used in this work

The samples defined in this work based on several spectroscopic surveys are displayed in a  $(X_0, Y_0)$  Galactic face-on view and as CMDs in Fig. E.1. Each sample covers a wide range of the MS, MSTO, Red Giant Branch (RGB) and SGB phases, with varying degrees of spread. Our LAMOST DR10 LRS sample offers the most stars, but it also has the most spread out CMD (as a consequence of the high number of stars covering more regions, but also because of the bigger spatial coverage, reaching distances of higher extinctions, where extinction maps are less precise), while the GALAH DR4 and APOGEE DR17 samples offer more thin sequences. Together, the samples cover a wide and diverse region of the solar neighbourhood. In all samples, the most densely populated regions of the CMD are the lower-magnitude end of the MS, and the MSTO, along with a notable over-density in the red clump.

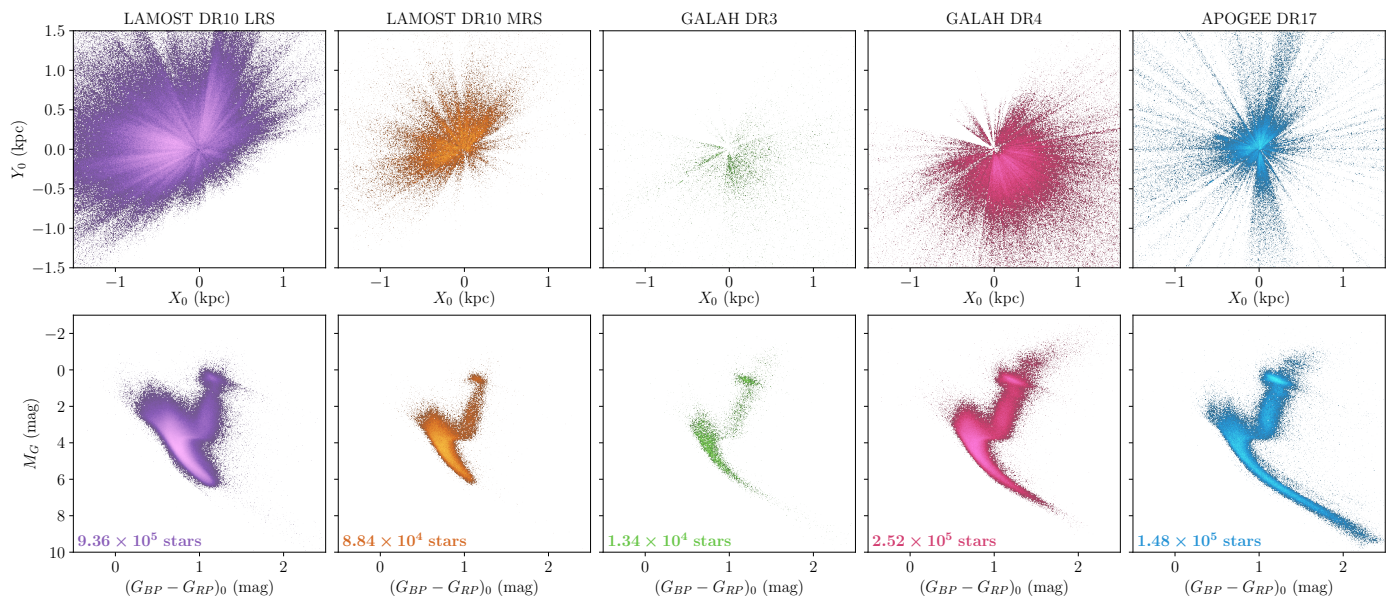
## Appendix F: Influence of $A_V$ on age estimations

Stellar extinction effects offset the observed position of a star in the CMD compared to its true position. Thus, age estimates based on this position have to rely on precise  $A_V$  values. To demonstrate this effect, we plot in Fig. F.1 age determinations based on the same observed properties, but corrected for extinction using two different extinction maps (one from Vergely et al. (2022) and one from Edenhofer et al. (2024)), for stars of C24’s sample.

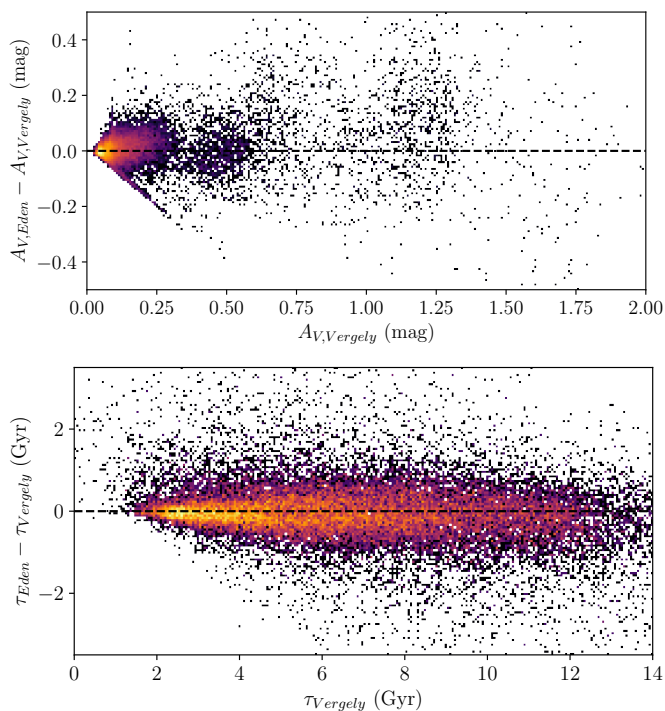
The top panel of Fig. F.1 shows the difference in  $A_V$  values computed for individual stars by integrating the extinction maps along the line of sight from the star to the sun, when the extinction maps used are different. This, in turn, will affect the age estimation, which is shown in the bottom panel of Fig. F.1. In some cases, the effect of a different value of  $A_V$  can lead to several Gyr of difference, highlighting the need for precise extinction maps for stellar age estimation. We note that no error on this extinction value is provided by both maps used here.

## Appendix G: Age-abundance planes for APOGEE DR17 and different evolutionary models

In this section, we compare the AMRs and  $[\alpha/\text{Fe}]$ -age relations given for our APOGEE DR17 sample, and with ages estimated using our NNs trained on different evolutionary models. As was discussed in Section 2.2, the BaSTI, MIST, PARSEC and Dartmouth trained NNs produce similar age estimates, and thus similar age-abundance relations. The two sequences of flat/shallow slopes at ages  $< 10$  Gyr and steeper slopes at older ages, both seen in the AMR (left subplot of Fig. G.1) and the  $[\alpha/\text{Fe}]$ -age relation (right subplot), are similar. We observe a slight deviation at older ages in the  $[\alpha/\text{Fe}]$ -age relation for the PARSEC-trained NNs, with a peak at ages  $\sim 11$  Gyr, while the peak is at ages  $\sim 12$  Gyr for the BaSTI and Dartmouth-trained NNs. The Geneva-trained NN gives different age estimates, and the corresponding relations diverge from the others, especially at older ages, with a drop/peak positioned at younger ages  $\sim 9$ -10 Gyr.



**Fig. E.1.** Overview of the surveys used in this work, coloured by density. From left to right: LAMOST DR10 (Cui et al. 2012; Zhao et al. 2012), LRS (purple) and MRS (orange), GALAH DR3 (De Silva et al. 2015; Buder et al. 2021) (green), GALAH DR4 (De Silva et al. 2015; Buder et al. 2025) (pink) and APOGEE DR17 (Abdurro’uf et al. 2022) (blue). *Top panel:* Face-on ( $X_0, Y_0$ ) galactic maps, centred on the Sun. *Bottom panel:* Hertzsprung-Russell Diagrams, coloured by density. The number of stars in each sample is annotated in the bottom left panel corner. Lighter colours correspond to higher density regions.



**Fig. F.1.** Comparison of  $A_V$  (top panel) values from Vergely et al. (2022) and Edenhofer et al. (2024). From these values, different extinction corrections are applied, which give different age estimations (bottom panel).

## Appendix H: Influence of $\alpha$ elements on age estimations

In this section, we compare age estimates made with a metallicities scaled using Salaris et al. (1993) relation (see Section 3) and with those that do not use such scaling relation. We estimate

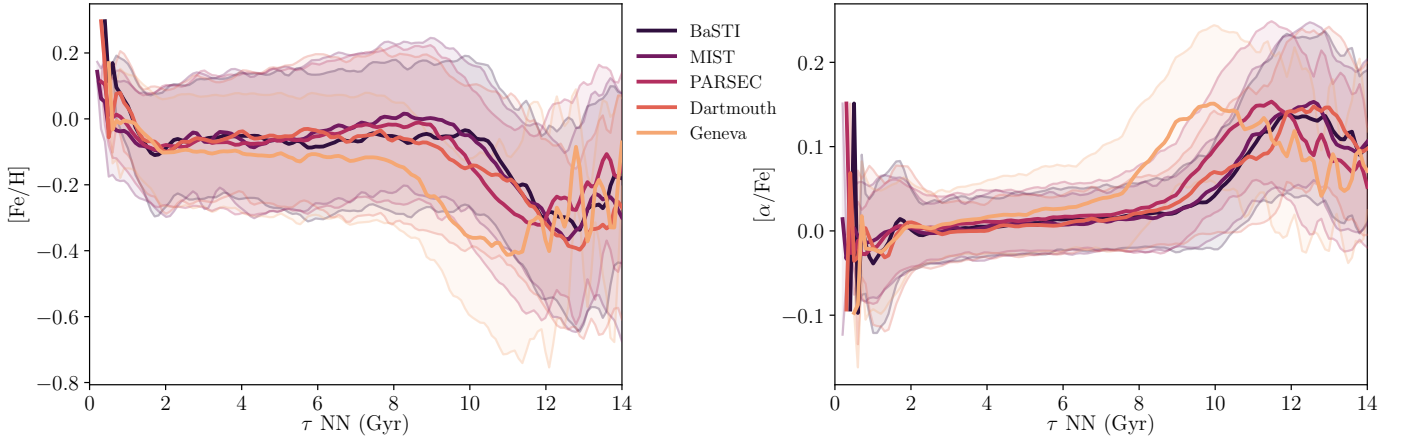
the ages for the LAMOST, GALAH and APOGEE catalogues, and compare them in Fig H.1. The GALAH catalogues show a very tight relation with low dispersion, and offset up to ages  $\sim 10$  Gyr, whereas LAMOST DR10 and APOGEE DR17 show a similar agreement for this age range, but with higher dispersion. The APOGEE DR17 panel seems to show a secondary sequence seen between 0 and 6 Gyr, for which unscaled metallicity ages reach ages higher than 10 Gyr. For ages older than 10 Gyr, all catalogues show a bump of greater ages for the unscaled-metallicity ages. This behaviour is expected as, for ages  $< 10$  Gyr, the  $\alpha$  element abundance distribution is centred on 0.0, while for older ages, the thick-disk  $[\alpha/\text{Fe}]$  increases with age (see Fig. 4 and Section 4.3). Hence, for young stars, on average, we have:

$$[\text{M}/\text{H}] \approx [\text{Fe}/\text{H}]$$

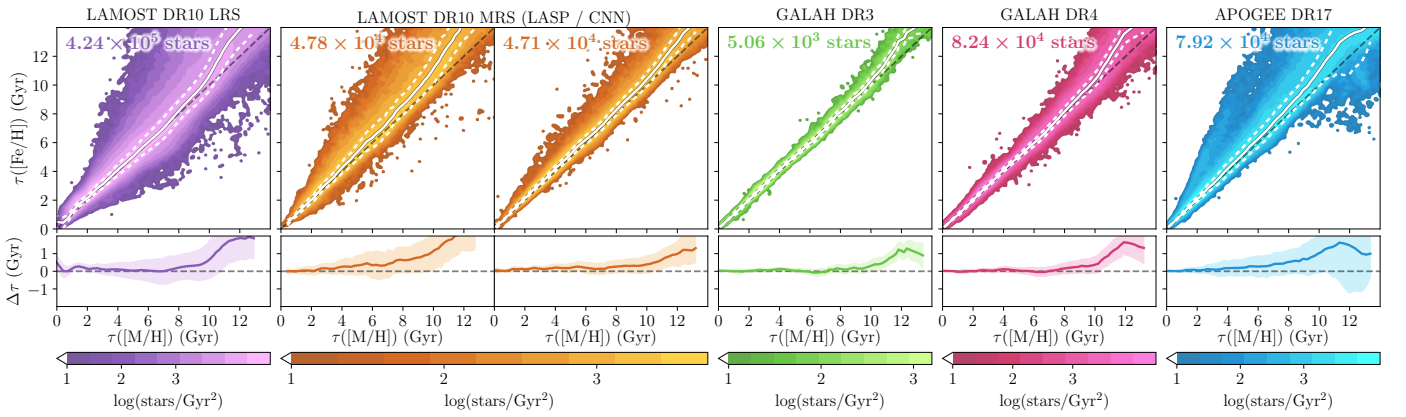
On the other hand, for an average old star having  $[\alpha/\text{Fe}] \approx 0.2$ , we have:

$$[\text{M}/\text{H}] \approx [\text{Fe}/\text{H}] + 0.76 * 0.2 = [\text{Fe}/\text{H}] + 0.152$$

This increase in the scaled metallicity will generally lead to a decrease in the age estimate, explaining the observed trends in Fig. H.1. The age at which  $\Delta\tau$  begins to differ is thus a proxy for the start of the thick disk sequence.



**Fig. G.1.** Abundance versus age trends for our APOGEE DR17 sample, with age estimated using NN trained on the BaSTI, MIST, PARSEC, Dartmouth, and Geneva grids. Coloured lines and shaded areas are the means and  $\pm 1\sigma$  of the distributions.



**Fig. H.1.** Age comparison with scaled ( $[M/H]$ ) and unscaled ( $[Fe/H]$ ) metallicities, from an NN trained on the BaSTI grid. *Top panel:* Age from scaled metallicity  $\tau([M/H])$  versus age from unscaled metallicity  $\tau([Fe/H])$ . The number of stars in each sample is annotated in the upper panel corners, and running medians and  $\pm\sigma$  lines are shown in solid and dotted white lines, as well as a one-to-one black dashed line to guide the eye. *Bottom panel:* Running median of the difference  $\Delta\tau$  between the two age estimates in solid lines, and associated standard deviation in shaded areas.