







## Research Article

# The GALAH survey: Data release 4

Sven Buder<sup>1,2,3</sup> , Janez Kos<sup>4</sup>, Xi Ella Wang<sup>1,2</sup>, Madeleine McKenzie<sup>1,2</sup>, Madeleine Howell<sup>2,5</sup>, Sarah Martell<sup>2,6</sup>, Michael R. Hayden<sup>2,7,8,9</sup>, Daniel B. Zucker<sup>2,10,11</sup>, Thomas Nordlander<sup>1,2,12</sup>, Benjamin Montet<sup>6,13</sup>, Gregor Traven<sup>4</sup>, Joss Bland-Hawthorn<sup>9,2</sup>, Gayandhi M. De Silva<sup>2,10,11</sup>, Kenneth Freeman<sup>1,2</sup>, Geraint Lewis<sup>9</sup> , Karin Lind<sup>14</sup>, Sanjib Sharma<sup>15</sup>, Jeffrey D. Simpson<sup>2,8</sup>, Dennis Stello<sup>2,8,9,16</sup>, Tomaz Zwitter<sup>4</sup>, Anish M. Amarsi<sup>12</sup>, Joseph J. Armstrong<sup>17</sup>, Kirsten Banks<sup>2,8</sup>, Mark Beavis<sup>18</sup>, Kevin-Luke Beeson<sup>4</sup>, Boquan Chen<sup>1,2</sup> , Ioana Ciucă<sup>1,2</sup>, Gary S. Da Costa<sup>1,2</sup> , Richard de Grijs<sup>10,11,19</sup>, Bailey Martin<sup>1</sup> , David Moise Nataf<sup>20</sup>, Melissa Ness<sup>1,2</sup>, Adam D. Rains<sup>12</sup>, Tim Scarr<sup>1</sup>, Rok Vogrinčič<sup>4</sup>, Zixian Purmortal Wang<sup>2,9,21</sup> , Rob A. Wittenmyer<sup>18</sup>, Yi Anne Xie<sup>1,2</sup> and The GALAH Collaboration<sup>22</sup>

<sup>1</sup>Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT, Australia, <sup>2</sup>ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia, <sup>3</sup>ACCESS-NRI, Australian National University, Canberra, ACT, Australia, <sup>4</sup>Faculty of Mathematics & Physics, University of Ljubljana, Ljubljana, Slovenia, <sup>5</sup>School of Physics and Astronomy, Monash University, Clayton, VIC, Australia, <sup>6</sup>School of Physics, University of New South Wales, Sydney, NSW, Australia, <sup>7</sup>Homer L. Dodge Department of Physics & Astronomy, University of Oklahoma, Norman, OK, USA, <sup>8</sup>School of Physics, UNSW, Sydney, NSW, Australia, <sup>9</sup>Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, Sydney, NSW, Australia, <sup>10</sup>School of Mathematical and Physical Sciences, Macquarie University, Sydney, NSW, Australia, <sup>11</sup>Astrophysics and Space Technologies Research Centre, Macquarie University, Sydney, NSW, Australia, <sup>12</sup>Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden, <sup>13</sup>UNSW Data Science Hub, University of New South Wales, Sydney, NSW, Australia, <sup>14</sup>Department of Astronomy, Stockholm University, AlbaNova University Centre, Stockholm, Sweden, <sup>15</sup>Space Telescope Science Institute, Baltimore, MD, USA, <sup>16</sup>Stellar Astrophysics Centre, Aarhus University, Aarhus C, Denmark, <sup>17</sup>Department of Space, Earth & Environment, Chalmers University of Technology, Gothenburg, Sweden, <sup>18</sup>Centre for Astrophysics, University of Southern Queensland, Toowoomba, QLD, Australia, <sup>19</sup>International Space Science Institute–Beijing, Zhongguancun, Beijing, China, <sup>20</sup>Department of Physics & Astronomy, University of Iowa, Iowa City, IA, USA, <sup>21</sup>Department of Physics and Astronomy, University of Utah, Salt Lake City, UT, USA and <sup>22</sup><https://www.galah-survey.org>

## Abstract

The stars of the Milky Way carry the chemical history of our Galaxy in their atmospheres as they journey through its vast expanse. Like barcodes, we can extract the chemical fingerprints of stars from high-resolution spectroscopy. The fourth data release (DR4) of the Galactic Archaeology with HERMES (GALAH) Survey, based on a decade of observations, provides the chemical abundances of up to 32 elements for 917 588 stars that also have exquisite astrometric data from the *Gaia* satellite. For the first time, these elements include life-essential nitrogen to complement carbon, and oxygen as well as more measurements of rare-earth elements critical to modern-life electronics, offering unparalleled insights into the chemical composition of the Milky Way. For this release, we use neural networks to simultaneously fit stellar parameters and abundances across the whole wavelength range, leveraging synthetic grids computed with Spectroscopy Made Easy. These grids account for atomic line formation in non-local thermodynamic equilibrium for 14 elements. In a two-iteration process, we first fit stellar labels to all 1 085 520 spectra, then co-add repeated observations and refine these labels using astrometric data from *Gaia* and 2MASS photometry, improving the accuracy and precision of stellar parameters and abundances. Our validation thoroughly assesses the reliability of spectroscopic measurements and highlights key caveats. GALAH DR4 represents yet another milestone in Galactic archaeology, combining detailed chemical compositions from multiple nucleosynthetic channels with kinematic information and age estimates. The resulting dataset, covering nearly a million stars, opens new avenues for understanding not only the chemical and dynamical history of the Milky Way but also the broader questions of the origin of elements and the evolution of planets, stars, and galaxies.

**Keywords:** Surveys; the Galaxy; methods: observational; methods: data analysis; stars: fundamental parameters; stars: abundances

(Received 22 October 2024; revised 28 January 2025; accepted 10 March 2025)

## 1. Introduction and workflow

### 1.1. Motivation

The history of our Milky Way galaxy is written in starlight. By capturing and analysing the light from millions of stars, which are now millions or billions of years old, we can uncover the chemical compositions embedded in their atmospheres since birth and use stars as time capsules into the past evolution of the Milky

**Corresponding author:** Sven Buder; Email: [sven.buder@anu.edu.au](mailto:sven.buder@anu.edu.au)

**Cite this article:** Buder S, Kos J, Wang XE, McKenzie M, Howell M, Martell S, Hayden MR, Zucker DB, Nordlander T, Montet B, Traven G, Bland-Hawthorn J, De Silva GM, Freeman K, Lewis G, Lind K, Sharma S, Simpson JD, Stello D, Zwitter T, Amarsi AM, Armstrong JJ, Banks K, Beavis M, Beeson KL, Chen B, Ciucă I, Da Costa GS, de Grijs R, Martin B, Nataf DM, Ness M, Rains AD, Scarr T, Vogrinčič R, Wang ZP, Wittenmyer RA, Xie YA and The GALAH Collaboration. (2025) The GALAH survey: Data release 4. *Publications of the Astronomical Society of Australia* 42, e051, 1–42. <https://doi.org/10.1017/pasa.2025.26>

Way. The light of stars can thus guide us to explore and map our environment and Country, just as it has guided Aboriginal and Torres Strait Islander peoples and their astronomers for tens of thousands of years.

With this fourth data release (DR4) from the Galactic Archaeology with HERMES (GALAH) Survey, we are proudly publishing the next set of measurements of stellar chemical abundances for almost a third of the elements in the periodic table that are created by stars. The initial motivation for measuring so many elemental abundances was laid out by De Silva et al. (2015) and included the major motivation – chemical tagging – with the aim to trace back stars that were born together through their (expected) similar chemical compositions. The recent and ongoing efforts of GALAH and other surveys like the SDSS/APOGEE surveys (e.g. Abdurro’uf et al. 2022; Kollmeier et al. 2017), LAMOST (Zhao et al. 2012), *Gaia*-ESO (Gilmore et al. 2022; Hourihane et al. 2023), RAVE (Steinmetz et al. 2020), and *Gaia* RVS (Recio-Blanco et al. 2023) have taught us that the chemical evolution of our Galaxy and stars is complex and it is difficult to recover stellar siblings on a large scale due to limitations in our observations, analysis methods, and intrinsic changes to chemical composition due to stellar evolution. New observations and innovations in the analysis that are presented in this data release will allow us to make significant progress towards chemical tagging.

The unique observational setup of GALAH allows us to deliver chemical abundance information for a powerful and substantial set of stars: those which have exquisite astrometric information from the revolutionary *Gaia* satellite (Gaia Collaboration et al. 2016) and for which we can estimate stellar ages either from empirical or theoretical models, like stellar isochrones or mass- and age-dependent relations of chemical compositions. By combining stellar ages, orbits, and chemistry, we have made major advances in the understanding of our Galaxy. In particular, the discovery of the major merger of the Milky Way with another slightly less massive galaxy between 8 and 10 Gyr ago (Belokurov et al. 2018; Helmi et al. 2018) was paradigm shifting and motivated a new rush to collect more (and more diverse) information about the stars in our Milky Way.

GALAH DR4 presents two major improvements over the previous data releases. We have increased the quantity as well as quality of observations and we have implemented a hybrid spectrum synthesis approach that allows us to fit 95% of the spectrum, including broad molecular absorption features from  $C_2$  and CN. This allows us to now infer up to 32 elements,<sup>a</sup> including N, with unprecedented precision for a larger number of stars. GALAH DR4 naturally continues both the observing program aimed at acquiring spectra of 1 million stars (De Silva et al. 2015), and our ongoing efforts to improve the spectrum reduction and analysis pipelines, including the novel and more accurate line modelling with non-local thermodynamics equilibrium. In GALAH DR1 and DR2 (Martell et al. 2017; Buder et al. 2018), we developed a novel, data-driven pipeline using the interpolation and fitting code *The Cannon* (Ness et al. 2015). However, for DR3 (Buder et al. 2021), we reverted to the more computationally expensive method of spectrum synthesis, applying it to a limited wavelength range to confirm the accuracy of our data-driven approach. In this data release, we are now implementing a hybrid approach. We create

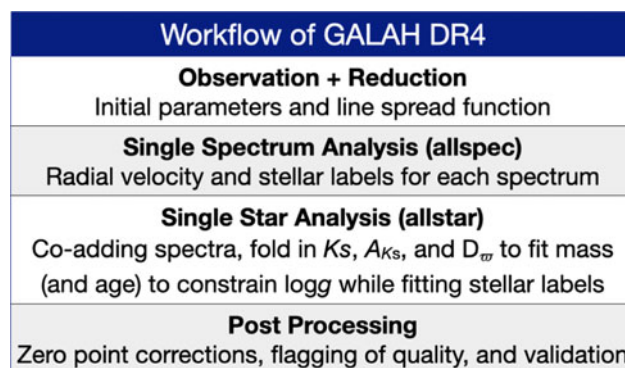


Figure 1. Workflow of GALAH DR4.

a training set of synthetic spectra across the full wavelength range using the same synthesis code as DR3, then train a neural network to interpolate the spectra efficiently in a high-dimensional space with up to 36 dimensions. By using neural networks, we can model the entire wavelength range, including broad molecular absorption features from  $C_2$  and CN, rather than focusing on narrow atomic line windows. This approach allows us to simultaneously model all stellar labels – global parameters and elemental abundances. Additionally, we can infer the shape of the interstellar spectrum from the differences between observed and synthetic spectra, while also incorporating non-spectroscopic information during the optimisation process.

In the following section, we outline our workflow and provide detailed explanations of our methodology throughout this manuscript, offering insights that upcoming surveys like WEAVE (Dalton et al. 2014), SDSS-V (Kollmeier et al. 2017), and 4MOST (de Jong et al. 2019) can readily utilise.

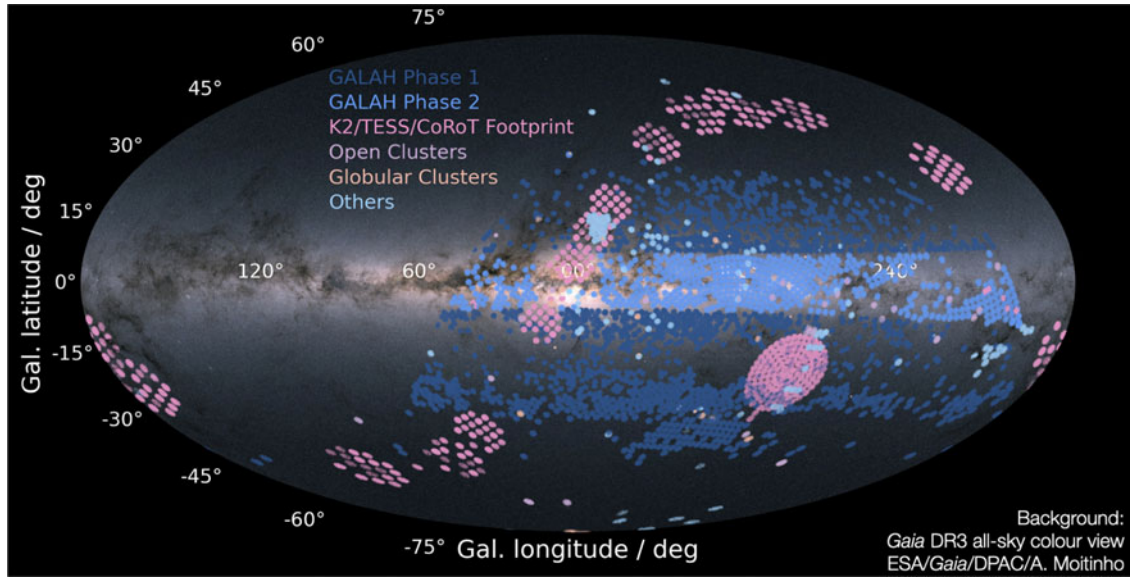
## 1.2. Workflow

The workflow of GALAH DR4 is depicted in Fig. 1 and will serve as a guideline for this manuscript: We first describe the collection of data in Section 2, most notably the observation of HERMES spectra. We explain how we create synthetic stellar spectra to compare with the observed ones in Section 3. This comparison is done in two consecutive steps. In Section 4, we explain how we extract stellar labels from individual observations (without non-spectroscopic information folded into the optimisation), while Section 5 describes how we co-add repeated observations and fold in non-spectroscopic information for each star. We describe the post-processing and validation of our data in Section 6. The data products of this data release are explained in Section 7. We describe identified caveats in Section 8 and make suggestions for minimising them in the future, before concluding this manuscript in Section 9.

## 2. Data

The GALAH Survey uses the 3.9-m Anglo-Australian Telescope at Siding Spring Observatory on Gamilaraay Country and its Two-Degree Field positioning system (2dF) top end (Lewis et al. 2002). 2dF magnetically places up to 400 fibre buttons on one of two metal field plates, which can be tumbled to allow observing with one set of fibres while configuring the other. Light is delivered through the fibres to the High Efficiency and Resolution Multi-Element Spectrograph (HERMES) spectrograph (Barden

<sup>a</sup>Li, C, N, O, Na, Mg, Al, Si, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Rb, Sr, Y, Zr, Mo, Ru, Ba, La, Ce, Nd, Sm, and Eu.



**Figure 2.** Overview of the distribution of stars included in this fourth GALAH data release in Galactic coordinates with the centre of the Galaxy at the origin and the *Gaia* DR3 all-sky colour view (Gaia Collaboration et al., 2023) as background. Shown are the targets of GALAH Phase 1 (dark blue) and Phase 2 (medium blue), the targets of the K2-HERMES follow-up along the ecliptic and TESS-HERMES in the TESS Southern Continuous Viewing Zone as well as CoRoT fields (pink). Both open and globular cluster points are shown in purple and orange, respectively. All other targets are shown in light blue across the Southern sky.

et al. 2010; Brzeski, Case, & Gers 2011; Heijmans et al. 2012; Farrell et al. 2014; Sheinis et al. 2015) and dispersed into four non-contiguous wavelength bands in the optical that cover  $\sim 1\,000\,\text{\AA}$  in the range of 4 713–4 903 (blue CCD or CCD1), 5 648–5 873 (green/CCD2), 6 478–6 737 (red/CCD3), and 7 585–7 887  $\text{\AA}$  (infrared IR/CCD4). The data used in this data release is primarily based on observations of stars with this setup, but also makes use of auxiliary photometric and astrometric information for the stars where available.

In this Section, we describe which stars we have targeted as part of configured fields (Miszalski et al. 2006) and observed with the 2dF-HERMES setup (Section 2.1), including the first description of the second phase of GALAH observations (GALAH Phase 2) which has a sharper focus on main-sequence turn-off stars to estimate more precise ages. In Section 2.2, we briefly summarise the properties of the spectroscopic data and how they were reduced to one-dimensional spectra. We also point out major changes in the observations and reductions with respect to the previous (third) data release (Buder et al. 2021). We further elaborate on the auxiliary information that was used for the analysis in Section 2.3.

### 2.1. Target selection and observational setup

GALAH DR4 is a combination of the main GALAH survey and additional projects to observe asteroseismic targets from the K2 (Howell et al. 2014) and TESS (Ricker et al. 2015) missions, that is, K2-HERMES (Sharma et al. 2019) and TESS-HERMES (Sharma et al. 2018), as well as numerous smaller programs and public HERMES data. Additional proposals with 2dF-HERMES have contributed targeted observations of globular cluster members (PI M. McKenzie and PI M. Howell), open clusters (PI G. De Silva and PI J. Kos), young stellar associations (PI J. Kos and J. Armstrong), and halo stars (PI S. Buder) in addition to their observation through the main surveys. The column survey\_name in

our catalogues denotes the origin. An all-sky view of GALAH DR4 is shown in Fig. 2.

#### 2.1.1. Target selection for GALAH Phase 1 and 2

For GALAH Phase 1 (DR1-DR3) and in the absence of a precise and volume-complete survey in the optical, we used the 2MASS photometric survey (Skrutskie et al. 2006) with its  $J$  and  $K_s$  filters as a precise and nearly volume-complete parent sample from which we selected stars based on approximated (De Silva et al. 2015) visual magnitudes

$$V_{JK_s} = K_s + 2(J - K_s + 0.14) + 0.382e^{((J-K_s)-0.2)/0.5}. \quad (1)$$

For GALAH Phase 1, a tiling pattern (with unique field\_id entries) with 2 deg fields of view below declination  $\delta \leq +10$  deg was created for regions with Galactic latitude  $|b| \geq 10$  deg to avoid crowding and strong extinction. For each tile, a selection of 400 stars within magnitudes  $9 \leq V_{JK_s} \leq 12$  for a bright magnitude cut and  $12 \leq V_{JK_s} \leq 14$  for the nominal magnitude cut is randomly selected from the complete parent sample of 2MASS. Of those, typically 350 stars are actually observed with around 2/3 main-sequence and turn-off stars and 1/3 evolved stars.

For GALAH Phase 2, a stronger focus on turn-off stars was implemented with the photometric and astrometric information of *Gaia* data release 2 as a parent sample. For each field, we therefore first allocate fibres to stars with absolute *Gaia* magnitude in the range of  $2 \leq M_G \leq 4$ , where

$$M_G = G + 5 \cdot \log_{10} \left( \frac{\varpi}{100 \text{ mas}} \right) \quad (2)$$

with apparent magnitude ( $G / \text{mag} \equiv \text{phot\_g\_mean\_mag}$ ) and parallax measurements ( $\varpi / \text{mas} \equiv \text{parallax}$ ) from *Gaia* DR2 (Gaia Collaboration et al. 2018; Evans et al. 2018; Lindegren et al. 2018). Remaining fibres are filled with targets as done with the



**Table 1.** Overview of stars observed for the programs included in GALAH DR4. Numbers of open and globular cluster observations were estimated after observations as described in Section 2.3.3. We have observed 30 globular clusters (23 with  $\geq 5$  stars) and 361 open clusters (109 with  $\geq 5$  stars).

| Program       | No. stars | Program           | No. stars |
|---------------|-----------|-------------------|-----------|
| galah_bright  | 67 680    | k2_hermes         | 117 736   |
| galah_main    | 434 901   | tess_hermes       | 37 129    |
| galah_faint   | 33 907    | globular clusters | 2 509     |
| galah_phase2  | 172 494   | open clusters     | 3 706     |
| commissioning | 2 625     | other             | 44 901    |

Phase 1 selection function. This leads to a different selection function for each phase. For science cases in which selection functions matter, we thus recommend to use the `survey_name` (Table 1) for a clean selection of phase and selection function.

### 2.1.2. Observational setup

We list the observations under various sub-programs in Table 1. Except for 2 935 spectroscopic observations with the high-resolution mode of HERMES ( $R \sim 42\,000$ ) on 7, 8, 10, 11 and 12 February 2014, all observations were made in the low-resolution mode ( $R \sim 28\,000$ ) with different total exposure times chosen for different programs, but typically between 60 and 90 min. Under sufficient conditions (no clouds and seeing below 2 arcsec), GALAH Phase 1 and TESS-HERMES observed 3 exposures for 6 min for bright targets ( $9 \leq V_{JKs} \leq 12$ ) and 3 exposures for 20 min for the majority of targets ( $12 \leq V_{JKs} \leq 14$ ).

GALAH Phase 2 extended these times to 3 exposures of 10 or 30 min, respectively, and included repeat observations of GALAH Phase 1 main targets with another 3 exposures for 15 min. K2-HERMES observations targeted stars with  $13 \leq V_{JKs} / \text{mag} \leq 15$  or even  $13 \leq V_{JKs} / \text{mag} \leq 15.8$  to complement the K2 Galactic Archaeology Program (Stello et al. 2015). These fields were observed for 2 h, similar to most globular and open cluster stars. Worse seeing conditions leading to increasing full-width at half maxima or thin clouds triggered between one ( $2 < \text{seeing} \leq 2.5$  arcsec) and 3 ( $2.5 < \text{seeing} \leq 3$  arcsec) additional exposures. In addition to the science frames, quartz fibre flat and ThXe arc observations were taken directly before or after each set of science exposures, and bias frames were taken at the beginning or end of each observing night.

## 2.2. Spectroscopic data from GALAH observations

Since the commissioning of the HERMES spectrograph in late 2013 until 6 August 2023, the GALAH collaboration and its partners have observed and successfully reduced 1 085 520 spectra of 917 588 stars. Each single observation is given a unique `subject_id` YYMMDDRRR01FFF that is based on its year (YY), month (MM), and day (DD) of observations, its exposure run number (RRRR), and the used fibre (FFF). A reduced example spectrum of the asteroid Vesta (observed on 15 January 2014 during run 22 through fibre 239 with `subject_id` 210115002201239) is shown in Fig. 3 and used as a reference for a Solar spectrum. The reduction process to create FITS files of reduced spectra from two-dimensional images from the cameras employs an updated and publicly available version 6 of the already well-tested reduction pipeline (Kos et al. 2017). The file extensions are listed in Table 2 and created as follows.

Science frames are corrected by removing the bias, dividing out the different gains (provided in the FITS headers) of the two readout amplifiers per CCD, flagging bad pixels, and dividing by master flat field frames, as well as removing cosmic rays and scattered light. Subsequently, apertures for each fibre trace are identified and used to extract the individual spectra.

Wavelength calibrations are performed via Chebyshev polynomial functions based on the up to 62, 52, 41, or 31 emission lines within the ThXe arc frames of CCDs 1–4, with wavelengths reported in air, and the spectra are interpolated onto a linearly increasing wavelength grid. Typical root mean square values for the wavelength solutions of CCDs 1–4 are 0.010, 0.015, 0.019, and 0.028 Å, respectively. The starting wavelength CRVAL1 and dispersion CDELTA1 are saved in the headers of each FITS file.

Finally, sky lines are subtracted and telluric features removed, before a barycentric correction is applied to create the ‘reduced’ spectra that are saved in extension 0 of the reduction pipeline FITS files and used for the subsequent analysis. Reduction pipeline spectra are normalised by an eleventh order Legendre polynomial fit and saved in extension 1 of the reduction products.

Fractional noise/uncertainties are saved in extension 2 and calculated from the square root of the sum of squared counts, sky features (extension 3), scattered light (extension 5), and crosstalk (extension 6) measurements as well as the squared readout noise.

The wavelength dependent line spread functions (LSFs) are measured from the arc calibration frames for each spectrum and CCD by fitting modified Gaussian distributions with one boxiness parameter  $b$  per CCD and full width half maxima  $fwhm$  for each wavelength point in the spectrum, that is

$$\exp(-0.693147 \cdot |2 \cdot \mathbf{x} / fwhm|^b) \quad (3)$$

The array  $\mathbf{x}$  then includes the pixels around each wavelength step that are used to apply the convolution from higher resolution to GALAH resolution spectra. The fitted values of  $fwhm$  are saved in extension 7 with  $b$  saved in the headers.

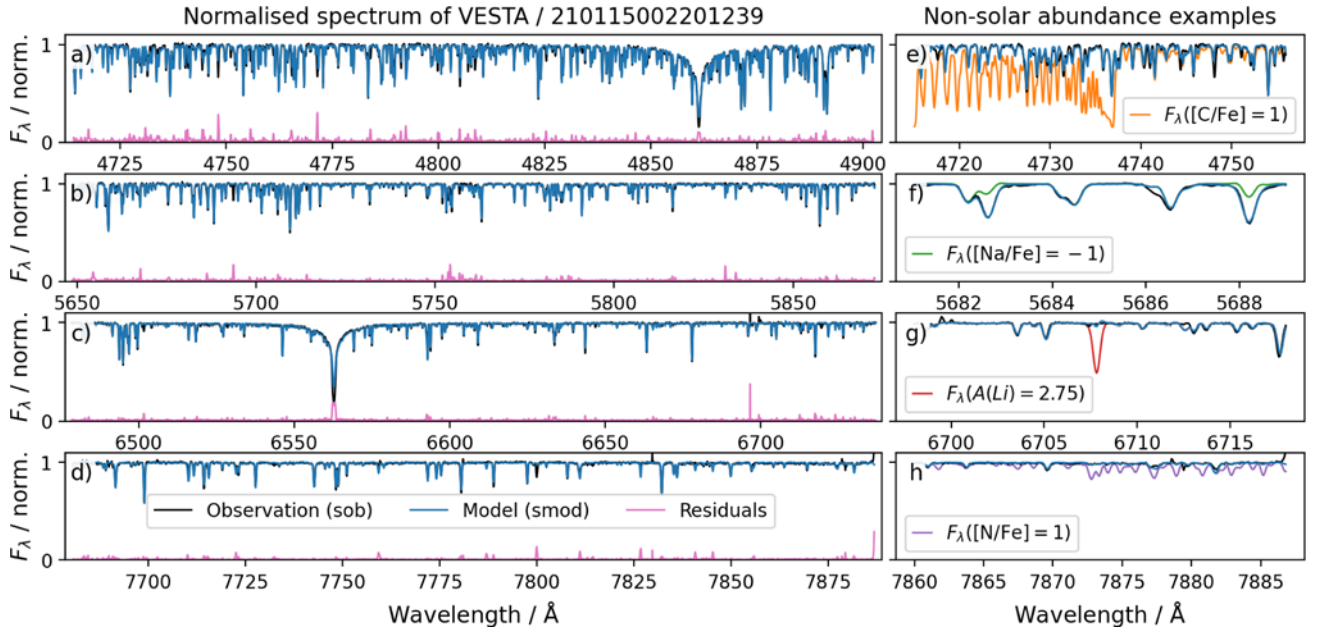
The achieved Signal-to-Noise Ratio (SNR) per pixel of the individual exposures depends on the spectral types, reddening, and observational conditions. In particular the repeat observations of previous pointings have increased the SNR for co-added spectra with respect to GALAH DR3. This can be appreciated from Fig. 4, where we plot the cumulative distribution function for all stars of GALAH DR3 (dashed lines) and GALAH DR4 (solid lines) for the different CCDs.

## 2.3. Auxiliary data from Gaia, 2MASS, and literature

To support our spectroscopic analysis, we make use of astrometric and photometric information from the *Gaia* satellite (Gaia Collaboration et al. 2016) and 2MASS survey (Skrutskie et al. 2006), which is available for essentially all our targets. We further use the value-added catalogues, like distance estimates for field stars by Bailer-Jones et al. (2021) as well as open and globular cluster membership probabilities from Cantat-Gaudin & Anders (2020) as well as Vasiliev & Baumgardt (2021) and Baumgardt & Vasiliev (2021).

### 2.3.1. Gaia DR3

We crossmatch our observations to the *Gaia* DR3 catalogue (Gaia Collaboration et al. 2021a, 2023) using the 2MASS ID, via the nearest neighbour crossmatches provided as part of *Gaia* DR3



**Figure 3.** Comparison of normalised observed (black) and synthetic spectra (blue) of the asteroid Vesta with solar composition as well as examples of synthetic spectra with non-solar abundances. Panels (a–d) show the observed and best-fitting synthetic spectrum as well as their absolute residual (pink) for the four wavelength channels of the HERMES spectrograph. Panel (e) shows the beginning of the blue CCD 1 (left most part of panel a) with an additional synthetic spectrum with ten times higher  $[C/Fe]$  in orange, for which the  $C_2$  Swan bands are prominent. Panel (f) shows the beginning of the green CCD 2 (left most part of panel b) and exemplifies with a synthetic spectrum in green that also has a ten times lower  $[Na/Fe]$  abundance (for example, in accreted stars) can still be reliably detected. Panel (g) shows the end of the red CCD 3 with a synthetic spectrum of primordial Li abundance of  $A(Li) = 2.75$  in red. While this abundance could be detected, the line for the Solar value  $A(Li) = 1.05$  is barely detectable. Panel (h) shows the end of the infrared CCD 4, which would show strong molecular absorption features of the CN molecule for  $[N/Fe] = +1$  dex (purple).

(Torra et al. 2021). 911 754 (99.0 %) also have astrometric information (Lindgren et al. 2021b) and 849 867 (93.0 %) have radial velocity estimates (Katz et al. 2023). We apply the corrections to both photometric (Riello et al. 2021) and astrometric (Lindgren et al. 2021a) information. Where possible we prefer the photo-geometric distances over the geometric distances from Bailer-Jones et al. (2021). Where neither are available, we further try to find parallaxes from van Leeuwen (2007). The average parallax uncertainty of the GALAH stars is  $\sigma_{\varpi}/\varpi = 1.6^{+2.6}_{-0.9}$  %. Only 2.3 % of GALAH stars have no parallax measurements<sup>b</sup> or parallax measurements beyond 20% uncertainty, for which the priors adopted by Bailer-Jones et al. (2021) start to dominate distance estimates.

### 2.3.2. 2MASS, WISE, and extinction

In addition to the excellent infrared photometry for 99.9 % of our sources from the 2MASS survey (Skrutskie et al. 2006), 98.7 % of them have far-infrared measurements from the WISE mission (Cutri et al. 2014). We therefore can estimate the extinction in the  $K_S$  band via the Rayleigh-Jeans colour excess (RJCE) method (Majewski, Zasowski, & Nidever 2011)  $A_{K_S} = 0.917 \cdot (H - W2 - 0.08)$  for most stars. We confirm this estimate by estimating the extinction in  $K_S$  via the extrapolation of the colour extinction of  $B - V$ , that is,  $A_{K_S} \sim 0.36 \cdot E(B - V)$  (Cardelli, Clayton, & Mathis 1989). We revert to this value if it is less than half the value of the RJCE estimate, or if either of the  $H$  and  $W2$

bands does not have an excellent quality flag ‘A’. For negative estimates by the RJCE method and very nearby stars ( $< 100$  pc) we null the value.

### 2.3.3. Open and globular cluster members and distances

We identify open cluster members using the membership catalogue from Cantat-Gaudin & Anders (2020) via crossmatch with the *Gaia* source\_id and adjust their parallaxes and distance estimates to the average cluster values if the latter are more precise. We identify globular cluster members (with more than 70% probability) via the membership catalogue from Vasiliev & Baumgardt (2021) by crossmatching with the *Gaia* source\_id. We then adjust the parallaxes and distances for the member stars to the mean values listed by Baumgardt & Vasiliev (2021).

## 3. Synthetic spectra for 2DF-hermes

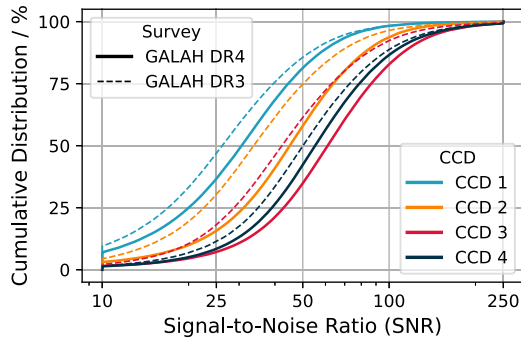
The goal of our spectroscopic analysis is to estimate the optimal set of stellar properties (labels) that influence a stellar spectrum by minimising the difference between observed stellar spectra and synthetic ones. In this data release, we push the analysis further by fitting up to 32 elemental abundances and stellar parameters simultaneously across the full GALAH wavelength range with the appropriate synthetic spectra.

To make this computationally feasible, we adopt a strategy inspired by Rix et al. (2016), where we create flexible models for smaller regions of the parameter space, utilizing only a limited number of *ab initio* synthetic spectra (see also Ting, Conroy, & Rix 2016). These synthetic spectra are calculated using Spectroscopy Made Easy (SME; Valenti & Piskunov 1996; Piskunov & Valenti

<sup>b</sup>For stars without parallaxes, we only perform an analysis without astrometric information.

**Table 2.** Data product 1: FITS files of reduced spectra.

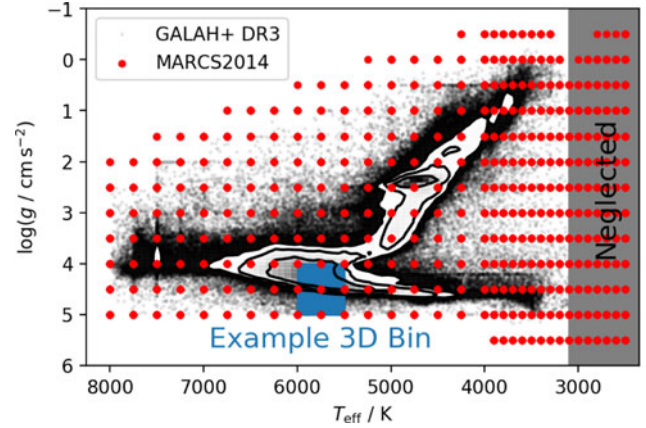
| FITS Ext. | Description                               |
|-----------|---|
| Ext. 0    | Un-normalised signal/counts               |
| Ext. 1    | Normalised signal (by reduction pipeline) |
| Ext. 2    | Relative uncertainty of signal            |
| Ext. 3    | Subtracted sky signal/counts              |
| Ext. 4    | Applied telluric correction               |
| Ext. 5    | Scattered light/counts                    |
| Ext. 6    | Cross-talk/counts                         |
| Ext. 7    | Resolution profile/FWHM                   |

**Figure 4.** Cumulative Distribution Function (CDF) of the logarithmic Signal-to-Noise Ratio (SNR) per pixel for the 4 CCDs of the HERMES spectrograph comparing GALAH DR4 (solid lines) and GALAH DR3 (dashed lines).

2017), covering the entire wavelength range and accounting for all visible atomic and molecular lines. The spectra are generated for random selections of elemental abundances and stellar parameters within the range allowed by MARCS atmospheric models (Gustafsson et al. 2008), at much higher resolution and sampling than our 2dF-HERMES spectra. From these, we select subsets of spectra corresponding to restricted regions of the parameter space defined by  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . This method is analogous to using Solar twins (see, e.g. Nissen 2015) or performing differential abundance analysis of globular cluster stars (e.g. Yong et al. 2013; Monty et al. 2023). By reducing the impact of systematic uncertainties in atomic data and atmospheric models, these approaches have proven to be highly effective (Nissen & Gustafsson 2018).

For each parameter subset, we train a neural network to map stellar fluxes to their corresponding stellar parameters and abundances, similar to *The Payne* (Ting et al. 2019). These models allow us to generate synthetic spectra across the full wavelength range for any combination of elemental abundances within the restricted parameter space in under a second – compared to the minutes or hours required by traditional physics-driven spectrum synthesis approaches.

Another key motivation for creating smaller training sets is the limited flexibility of interpolation methods when dealing with the full parameter space. Spectroscopic surveys like GALAH, RAVE, and APOGEE aim to fit all types of stellar spectra simultaneously, including Sun-like stars, red clump stars, metal-poor stars, evolved stars with strong molecular bands, and hot stars with shallow and broad lines. Attempting to model this vast range with a single model leads to systematic trends, particularly in extreme cases (Casey et al. 2016; Buder et al. 2018; Ting et al. 2019). To mitigate

**Figure 5.** Coverage in  $T_{\text{eff}}$  and  $\log g$  of the MARCS2014 grid (red) and GALAH DR3 (black, including density contours). Shown is also an example of one of the 3D bins used to create stellar sibling models with each neural network. MARCS grid points  $T_{\text{eff}} < 3\,100\text{ K}$  or  $[\text{Fe}/\text{H}] < -3$  dex are neglected for GALAH DR4.

these issues, we deliberately limit the complexity of the models by creating smaller, more focused models. For example, the model for hot stars does not need to predict the strong molecular absorption features seen in cooler stars. The potential caveats and limitations of this approach are discussed in detail in Section 8.

In the following sections, we describe our approach to dividing the parameter space into smaller bins for training (Section 3.2) and explain how we generate high-resolution synthetic spectra for this parent sample (Section 3.2). We also outline how we train neural networks to rapidly interpolate these synthetic spectra (Section 3.3).

### 3.1. Stellar twin training sets rather than one-fits-all

The base grid for our training set computation is the MARCS grid (Gustafsson et al. 2008), which is shown with red points in Fig. 5. Following the aforementioned idea of restricting ourselves to stellar siblings, we create multiple 3-dimensional bins in  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  within  $\pm 1$  grid points in  $T_{\text{eff}}$  (with either  $\pm 250$  or  $\pm 100\text{ K}$ ),  $\log g$  ( $\pm 0.5$  dex), and  $[\text{Fe}/\text{H}]$  ( $\pm 0.5$  or  $\pm 0.25$  dex). An example box is shown for Solar siblings as a blue box in Fig. 5, which is centred on  $T_{\text{eff}} = 5\,750 \pm 250\text{ K}$ ,  $\log g = 4.5 \pm 0.5$  dex and  $[\text{Fe}/\text{H}] = 0.0 \pm 0.25$  dex.

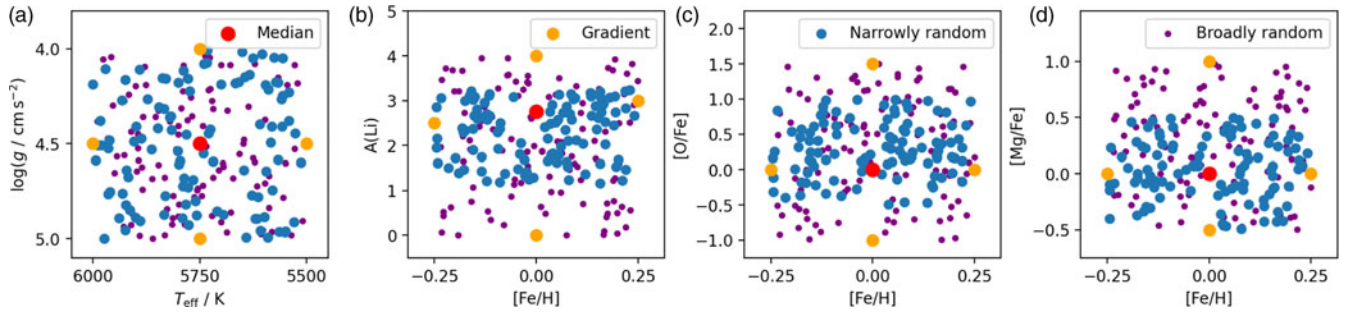
Within these bins we sample  $280^\circ$  synthetic spectra with no rotational broadening, which are later broadened with different rotational velocities  $v \sin i$  to create between 1 680 and 2 240 training set spectra for each bin. For clarity, we explain the parameter and abundance sampling for an example 3D bin centred on  $T_{\text{eff}} = 5\,750 \pm 250\text{ K}$ ,  $\log g = 4.5 \pm 0.5$  dex and  $[\text{Fe}/\text{H}] = 0.0 \pm 0.25$  dex (see blue box in Fig. 5).

Stellar parameters ( $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ,  $v_{\text{mic}}$ ) and elemental abundances  $[\text{X}/\text{Fe}]$  of all 32 elements are randomly sampled within reasonable limits (see examples in Fig. 6 and Table 3) and fed into SME to create self-consistent synthetic spectra over the full HERMES wavelength range for MARCS atmospheres.

Microturbulence velocity ( $v_{\text{mic}}$ ) values are sampled uniformly between the upper and lower limits of the empirical relation from GALAH DR3 (Eqs. 4 and 5 from Buder et al. 2021) and an adjusted

<sup>c</sup>This number is chosen to match the 28 CPUs of our computing nodes.





**Figure 6.** Coverage of stellar parameters and abundances for one of the 3D bins. Shown is the example of the Solar 3D bin ( $T_{\text{eff}} / \text{K} = 5750$ ,  $\log g / \text{dex} = 4.5$ ,  $[\text{Fe}/\text{H}] / \text{dex} = 0.0$ ). Panel a):  $T_{\text{eff}}$  and  $\log g$ , Panel (b):  $[\text{Fe}/\text{H}]$  vs.  $A(\text{Li})$ , Panel (c):  $[\text{Fe}/\text{H}]$  vs.  $[\text{O}/\text{Fe}]$ , Panel (d):  $[\text{Fe}/\text{H}]$  vs.  $[\text{Mg}/\text{Fe}]$ . While  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  are sampled randomly within the 3D bin, the abundances are sampled both narrowly (blue) and broadly (purple) within limits as described in the text. Red points indicate the median label values and orange points the adjusted label values (see Table 3) to test the gradient change of spectra with individual labels.

version of the relation of Dutra-Ferreira et al. (2016). The latter has been adjusted for  $T'_{\text{eff}} = T_{\text{eff}} / \text{K} - 5500$  as well as  $\log g' = \log g / \text{dex} - 4.0$  to return  $v_{\text{mic}} / \text{km s}^{-1}$ :

$$v_{\text{mic}} = 1.198 + 3.16 \times 10^{-4} \cdot T'_{\text{eff}} - 0.253 \cdot \log g' - 2.86 \times 10^{-4} \cdot T'_{\text{eff}} \cdot \log g' + 0.165 \cdot (\log g')^2 \quad (4)$$

### 3.2. High-resolution synthetic spectra with SME

We create training sets from high-resolution stellar spectra for each smaller 3D bin region of the parameter space. We compute oversampled synthetic intensity spectra at higher resolution and sampling than the typical GALAH resolution with SME for seven equal-area angles (see Fig. 7) of the plane-parallel or spherically symmetric stellar surfaces (Gustafsson et al. 2008).

For each spectrum, we first run a test on all available lines in the GALAH linelist. We use the same linelist as in GALAH DR3 (Buder et al. 2021). This linelist was adapted from the linelist of Heiter et al. (2021) and implements numerous updates to line data, such as updates or corrections of  $\log gf$  values in the GALAH wavelength range. The test is used to restrict the myriad of possible molecular lines to the visible ones with SME.depth above 0.001, while keeping all atomic lines for the final synthesis.

Spectra are computed at a resolution of  $R = 300\,000$  on a fine wavelength grid with 60 819 pixels spread over the extended wavelengths 4 675.1–4 949.9, 5 624.1–5 900.9, 6 424.1–6 775.9, and 7 549.1–7 925.9 Å. We note that these extend significantly beyond the actual GALAH wavelength range.

We use one-dimensional (1D) MARCS atmospheres from the MARCS grid (Gustafsson et al. 2008, version 2014) with a trilinear interpolation for combinations of  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . We use grids of non-LTE departure coefficients from Amarsi et al. (2020b), Amarsi, Liljegren, & Nissen (2022) for atomic lines of H, Li, C, N, O, Na, Mg, Al, Si, K, Ca, Mn, Fe, and Ba. For most elements, the non-LTE departure coefficient grids include isotropic and coherent scattering for lines from background atomic and ionic species (see Equation 7 of Amarsi et al. 2020b) as well as Thompson and Rayleigh scattering. The calculations for C include all background species in pure absorption (Equation 6 of Amarsi et al. 2020b), whereas for Fe, Thompson and Rayleigh scattering were included but all background lines were treated in pure absorption.

Our synthetic grid explicitly includes C and N abundances. C was previously included in the analysis of GALAH DR3, but limited to the atomic C line. The analysis thus neglected the molecular absorption features of  $\text{C}_2$  and CN at the beginning of CCD1 and end of CCD4, respectively. With the new self-consistent grid, we can include these features, as they hold valuable information for both C and N, as well as several other features through the molecular equilibrium in stars (see e.g. Ting et al. 2018).

To be able to test that the flux-label correlations found by our interpolation routine are limited to reasonable wavelength ranges, we also calculate one spectrum that is exactly in the middle of the parameter range and additional spectra, where we increase the value of one label at a time (e.g. increase  $[\text{O}/\text{Fe}]$  by 1 dex) to test the response in the synthetic spectrum.

To save computational costs, we compute synthetic spectra with no rotational or macroturbulence broadening ( $v_{\text{mac}} = v \sin i = 0 \text{ km s}^{-1}$ ), but save the continuum flux (sme.cmod) and the specific intensities (sme.sint) as a function of the equal-area midpoints of each equal-area annulus<sup>d</sup>  $\mu$  (see Fig. 7). We then apply the broadening of spectra due to rotation ( $v \sin i$ ) with the flux integration code of the python-implementation PYSME (Wehrhahn, Piskunov, & Ryabchikova 2023) of SME. Depending on the expected rotational velocities (increasing with temperature) we sample a range of

$$v \sin i / \text{km s}^{-1} \in \{1.5, 3, 6, 9, 12, 18, 24, 36\}. \quad (5)$$

Note that  $v \sin i = 24 \text{ km s}^{-1}$  is only included for bins with  $T_{\text{eff}} \geq 5\,000 \text{ K}$  and  $v \sin i = 36 \text{ km s}^{-1}$  for those with  $T_{\text{eff}} \geq 6\,000 \text{ K}$ .

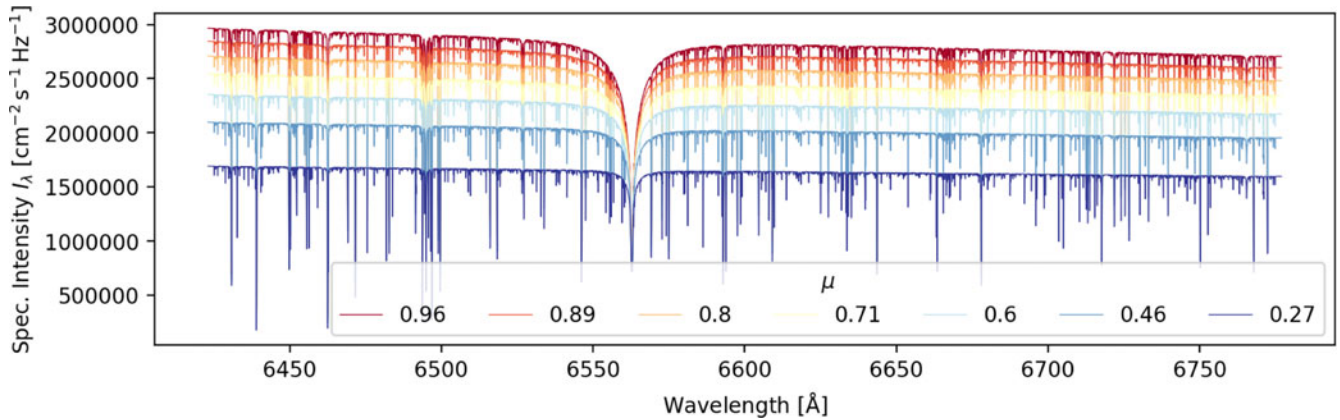
### 3.3. Interpolating synthetic spectra with neural networks

To allow a fast interpolation with new and different stellar labels, we use data-driven models to connect stellar fluxes at given pixels from a combination of stellar labels. This method is well established in stellar spectroscopy through the successful applications of quadratic models with *The Cannon* (see e.g. Ness et al. 2015; Ness et al. 2016; Casey et al. 2016; Casey et al. 2017; Ho et al. 2017; Buder et al. 2018) as well as neural networks with *The Payne* (see e.g. Ting et al. 2019; Xiang et al. 2019; Xiang et al. 2022). Because of the needed flexibility to predict synthetic spectra with 36 stellar labels for a large parameter space (for a detailed discussion

<sup>d</sup>  $\mu \in [0.96, 0.89, 0.8, 0.71, 0.6, 0.46, 0.27]$ .

**Table 3.** Example of boundaries for the uniform sampling of synthetic spectrum labels (stellar parameters and elemental abundances) for the 3-dimensional bin of Solar siblings 5750\_4.50\_0.00.

| Parameter                           | Sampling                          | Element                                       | Sampling narrow | Element                                       | Sampling broad |
|-------------------------------------|-----------------------------------|---|-----------------|---|----------------|
| $T_{\text{eff}} / \text{K}$         | 5500.5750.6000                    | A(Li)   | 1.05.2.75.3.26  | A(Li)   | 0.00.4.00      |
| $\log g / \text{dex}$               | 4.0.4.5.5.0                       | C, N, O                                       | −0.5.0.0.1.0    | C, N, O                                       | −1.0.1.5       |
| $[\text{Fe}/\text{H}] / \text{dex}$ | −0.25.0.0.0.25                    | Y, Ba, La, Ce, Nd                             | −0.5.0.0.1.0    | Y, Ba, La, Ce, Nd                             | −1.0.1.5       |
| $v_{\text{mic}} / \text{km s}^{-1}$ | 0.5,1.5,4.0, but see Equation (4) | $[\text{X}/\text{Fe}]$ for Mg, Si, Ti         | −0.5.0.0.0.5    | $[\text{X}/\text{Fe}]$ for Mg, Si, Ti         | −0.5.1.0       |
| $v \sin i / \text{km s}^{-1}$       | 0.0, but see Equation (5)         | $[\text{X}/\text{Fe}]$ for all other elements | −0.5.0.0.0.5    | $[\text{X}/\text{Fe}]$ for all other elements | −1.0.1.0       |



**Figure 7.** Example output of SME for a solar spectrum in HERMES CCD3 (around the Balmer  $H_\alpha$  line). Shown are the specific intensities (`sme.sint`) as a function of the viewing angle  $\mu = \cos \theta$ .

of advantages of neural networks over quadratic models see Ting et al. (2019), we choose neural networks to interpolate between our synthetic spectra in this data release.

In this work, we utilise the neural network architecture and training algorithms from the spectrum interpolation framework of *The Payne* (Ting et al. 2019). While we do not implement the full functionality of *The Payne*, we specifically adopt its spectrum interpolation capabilities. Unlike the version originally published by Ting et al. (2019), we use the architecture of the latest available version of *The Payne*. This modified architecture connects  $k$  stellar labels  $\ell$  with the flux  $f$  at each wavelength pixel  $\lambda$  via

$$f_\lambda = w \cdot \text{lReLU} \left( \tilde{w}_\lambda^i \cdot \text{lReLU} \left( w_{\lambda i}^k \ell_k + b_{\lambda i} \right) + \tilde{b} \right) + \tilde{f}_\lambda, \quad (6)$$

which encapsulates the so-called layers of a neural network with  $i = 300$  neurons with weights  $w$  and biases  $b$  as well as a leaky Rectified Linear Unit (lReLU)

$$\text{lReLU}(x) = \begin{cases} x & x \geq 0 \\ 0.01x & x < 0. \end{cases} \quad (7)$$

After optimising the mean absolute error loss function for  $10^4$  steps, we consider the network trained with an optimised combination of three sets of weights and biases within the minimum and maximum ranges of each label. We discuss the performance and caveats of this particular neural network architecture and training setup in Section 8.3. The trained networks can then be used with new input labels to quickly create synthetic spectra for the label optimisation. Computational resources could be conserved by training neural networks exclusively on spectra from non-rotating

stars and subsequently applying broadening through convolution with a center-to-limb darkening law. This method, while less accurate, could enable the fitting of broader velocity ranges and enhance neural network performance by simplifying the spectral shapes they must learn. However, shifting the broadening process from training to post-processing does not necessarily guarantee a reduction in computational costs.

#### 4. Single spectrum analysis (ALLSPEC)

As outlined in Section 1, the workflow of GALAH DR4 includes a first analysis step of all observed spectra without including non-spectroscopic information for the optimisation. This allows us to identify shifts in radial velocity between separate spectroscopic observations of the same star<sup>e</sup> and a better co-adding of spectra for the *allstar* analysis (see Section 5). Another motivation for this step is to get a first estimate of stellar labels without potentially biased photometric and astrometric information, for example for binary stars.

The optimisation of stellar labels thus aims to minimise the absolute difference between synthetic and observed spectra, weighted by their uncertainty. Starting from a set of initial labels (Section 4.1), we create high-resolution synthetic spectra and convolve them to the resolution and wavelength grid of each observed spectrum. We remind ourselves that in GALAH DR3,

<sup>e</sup>While repeat observations were only done for quality assurance in GALAH Phase 1, a significant number of repeat observations was performed as part of Phase 2.



we used a repeated combination of spectrum normalisation followed by stellar parameter optimisation and a subsequent fit of individual elements with fixed stellar parameters. In the analysis of GALAH DR4, we perform an on-the-fly re-normalisation of the observed spectrum for every change of the simultaneously fitted stellar parameters and elemental abundances. This allows a more accurate comparison of synthetic and observed spectra (Section 4.2) and thus a more accurate stellar label optimisation (see Section 4.3).

#### 4.1. Initial stellar labels

Initial values of all stellar labels are needed for creating a first synthetic spectrum. For  $v_{\text{rad}}$ ,  $T_{\text{eff}}$ ,  $\log g$ , and  $v \sin i$  we use a combination of sources. Where possible, we use the previous estimates from GALAH DR3 (Buder et al. 2021), and otherwise use estimates from the GALAH DR4 reduction pipeline (Section 2.2). Because of the limited accuracy of the latter for cool stars with  $T_{\text{eff}} < 4\,000$  K as well as the hot stars with  $T_{\text{eff}} > 6\,500$  K, we perform a consistency check with photometric information from *Gaia* DR3 (Gaia Collaboration et al. 2021a) and 2MASS (Skrutskie et al. 2006). For most of the aforementioned cool and hot stars, we therefore prefer the parameters from the *Gaia* DR3 photometric pipeline GSP-Phot (Andrae et al. 2023; Fouesneau et al. 2023) as initial values.

In selected cases, we further adjust the starting parameters toward reasonable limits. For example, hot stars are likely to be young and are adjusted to close to Solar metallicity. Furthermore, we recalculate the initial  $v_{\text{mic}}$  based on Equation (4) and limit rotational broadening values to  $3 \leq v \sin i \leq 10 \text{ km s}^{-1}$  for stars below  $T_{\text{eff}} = 5\,500$  K and  $3 \leq v \sin i \leq 20 \text{ km s}^{-1}$  for hotter stars. The explicit choices of starting values for  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ,  $v_{\text{mic}}$ , and  $v \sin i$  are described in our online repository<sup>f</sup> and are depicted in Fig. A1.

Based on the value of  $[\text{Fe}/\text{H}]$  we apply an offset to the  $\alpha$ -process elements O, Mg, Si, Ca, and Ti. The initial value is 0.4 for  $[\text{Fe}/\text{H}] < -1$ , 0.0 for  $[\text{Fe}/\text{H}] > 0$ , and  $-0.4 \cdot [\text{Fe}/\text{H}]$  for  $-1 \leq [\text{Fe}/\text{H}] \leq 0$ . All other abundances are initialised at  $[X/\text{Fe}] = 0$ .

#### 4.2. Comparison of synthetic spectra to observations

The major aim of our spectroscopic analysis is to predict the best set of stellar labels by minimising the uncertainty-weighted difference between observed and synthetic spectra. In this section, we describe several important steps to enable the pixel-level comparison of the higher resolution, oversampled synthetic spectra created with the neural networks from Section 3.3 and the observational data at actually measured resolution and sampling (presented in Section 2.2).

##### 4.2.1. Downgrading synthetic spectra to observed resolution

Because dedicated line-spread-function measurements are available for every spectrum (see Section 2.2), we use this information to downgrade our synthetic spectrum with Gaussian kernels on an equidistant velocity grid to the measured resolution of each observation. We then interpolate the oversampled synthetic spectrum onto the observed wavelength grid.

##### 4.2.2. On-the-fly re-normalisation of observed spectra

Measurements of the GALAH flux and flux uncertainty are reported in counts by the reduction pipeline. To compare with our synthetic spectra, which are normalised to the continuum, we fit an outlier-robust polynomial function to the ratio of observed and synthetic spectra and re-normalise our observed spectra and their uncertainties via this normalisation function.

This specific approach is similar to the internal routine of SME and has the important advantage that no continuum points have to be defined. This is advantageous because we try to cover the full parameter range of FGKM stars for which positions of continuum points – corresponding to 1 on a (pseudo-)continuum-normalised spectrum – differ significantly or for which continuum points may not even be present, or will be a strong function of  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  (as is the case for M stars).

We make two additional adjustments to the reduced spectra, which come in the form of counts and uncertainty per wavelength,  $f_{\lambda}$  and  $\sigma_{f,\lambda}$ .

As we compare the observation to model spectra, we do not have to restrict ourselves to an *a priori* normalisation, but can take into account the residual information on the continuum in parts of the spectrum. For each model spectrum that we compare to, we therefore perform a normalisation by fitting a fourth order Chebyshev polynomial with outlier clipping to the ratio of model and observation (see Fig. 8). This allows us to both overcome previous shortcomings of the synthetic analysis in GALAH+ DR3 (Buder et al. 2021), which had to be restricted to small wavelength segments and assumed a linear relation for those. Our new approach allows us to properly assess the structure of deep and steep molecular features that can dominate spectra of cool stars and carry significant information on  $T_{\text{eff}}$ ,  $v_{\text{rad}}$ , as well as abundances (Mann et al. 2012).

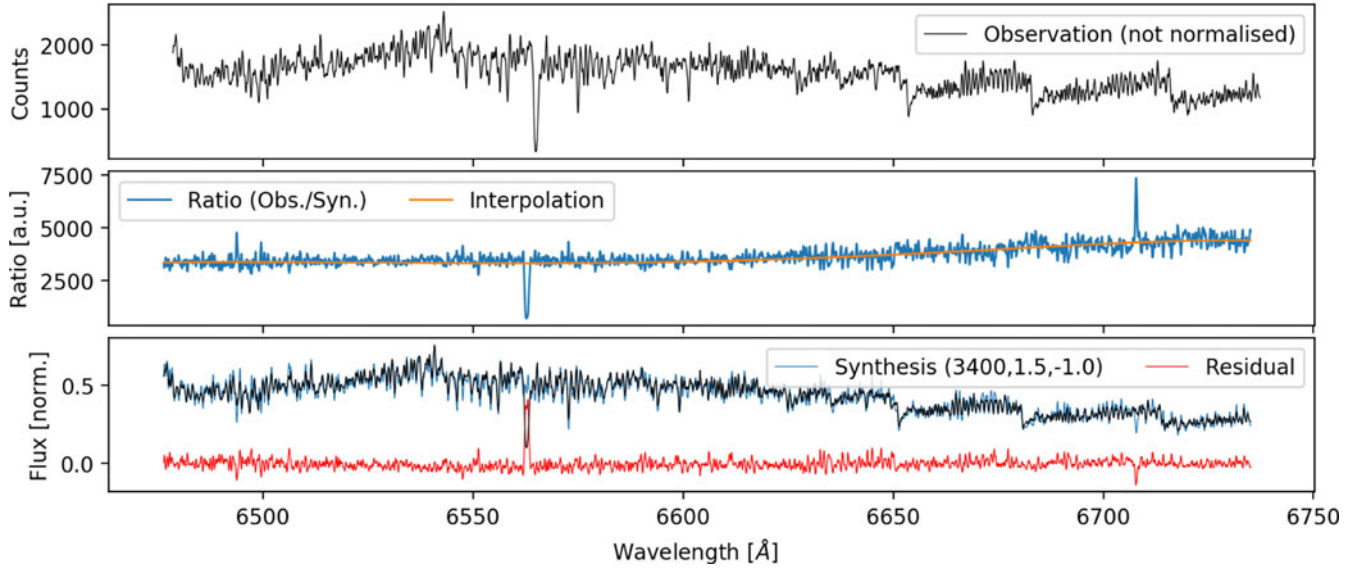
#### 4.3. Stellar label optimisation

In up to four major loops, we optimise the radial velocities and all other stellar labels and report (a) their values, (b) their covariances, (c) the best-fitting synthetic and re-normalised spectra along with (d) their uncertainties and (e) masks that indicate which pixels were used in the final optimisation.

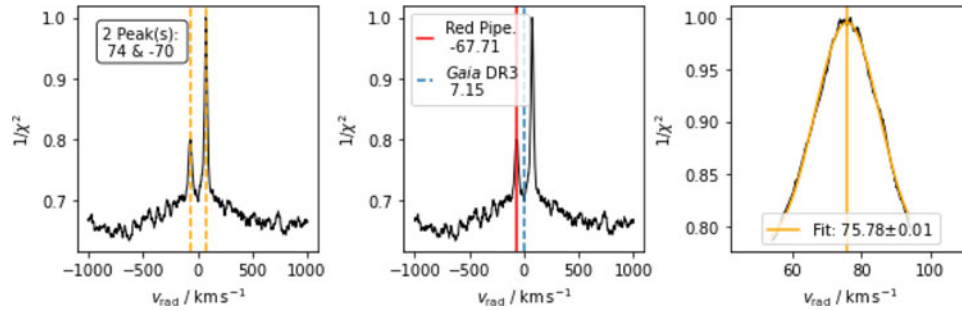
Starting from the initial values, a first synthetic spectrum is computed and compared with the observation in order to assess the initial radial velocity. This is done by applying the `SCIPY.SIGNAL.FIND_PEAKS` algorithm on the normalised inverse residuals of non-shifted observed and synthetic spectra, when the latter is shifted by  $v_{\text{rad}} = -1\,000..(2)..1\,000 \text{ km s}^{-1}$  (see Fig. 9a). If no peak is found, the initial  $v_{\text{rad}}$  value is used hereafter. If more than one peak is found (see Fig. 9 with *Gaia* DR3 agreeing with the systemic radial velocity), the two strongest peaks are reported. For the purpose of the single star analysis, a narrower search is conducted around the highest peak with a  $v_{\text{rad}}$  shift of  $-20.00..(0.04)..20.00 \text{ km s}^{-1}$  around said peak by fitting a Gaussian function to the inverse of the residuals that were normalised with the smallest residual values (see Fig. 9c). The mean of this fit and its uncertainty are reported by the pipeline.

The centerpiece of our optimisation is the `SCIPY.OPTIMIZE` module's `CURVE_FIT` function (Virtanen et al. 2020), which we call with counts and uncertainties (our absolute sigmas) as input for a placeholder function that self-consistently re-normalises the observed spectrum. We estimate the labels via the least squares

<sup>f</sup>GALAH\_DR4/spectrum\_analysis/galah\_dr4\_initial\_parameters.ipynb.



**Figure 8.** Example of normalisation for GALAH DR4 for a model spectrum ( $T_{\text{eff}} = 3\,400\text{ K}$ ,  $\log g = 1.5$ ,  $[\text{Fe}/\text{H}] = -1.0$  dexbest – fitting) that is selected during the label optimisation. Panel (a): Observed spectrum (counts). Panel (b): Ratio (blue) of observed spectrum and model spectrum as well as Chebyshev polynomial fit (orange). Panel (c): Normalised observed spectrum (black) compared to the model spectrum (blue). Residuals (red) can then be used as input for the likelihood function.



**Figure 9.** Output of the radial velocity fitting step. Panel (a) shows the initial broad search on a  $v_{\text{rad}}$  array of  $-1000..(2)..1000\text{ km s}^{-1}$ . In the case of 2MASS J060846577815235, two peaks (yellow, dashed) are visible for this double-lined spectroscopic binary. Panel (b) shows the same plot, but overlaid with the GALAH DR4 reduction pipeline (red) and *Gaia* DR3 (blue, dashed) estimates for  $v_{\text{rad}}$ . Panel (c) shows the narrow window of  $-20.00..(0.04)..20.00\text{ km s}^{-1}$  around the highest peak and its Gaussian fit (yellow). Despite their low resolution (26 KB), these on-the-fly created diagnostic images already occupy 50GB in total.

optimisation within less than  $10^4$  iterations and a desired relative error ( $x_{\text{tol}}$ ) below  $10^{-4}$ .

For each optimisation loop, a new, best-fitting 3D bin and neural network is identified via a grid search in the  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  dimensions with SKLEARN.CKDTree. If the stellar labels that are being fitted have changed (for example if an element is deemed not detectable for the new 3D bin during the neural network training), the label and its value are either set to or initialised with  $[\text{X}/\text{Fe}] = 0$ .

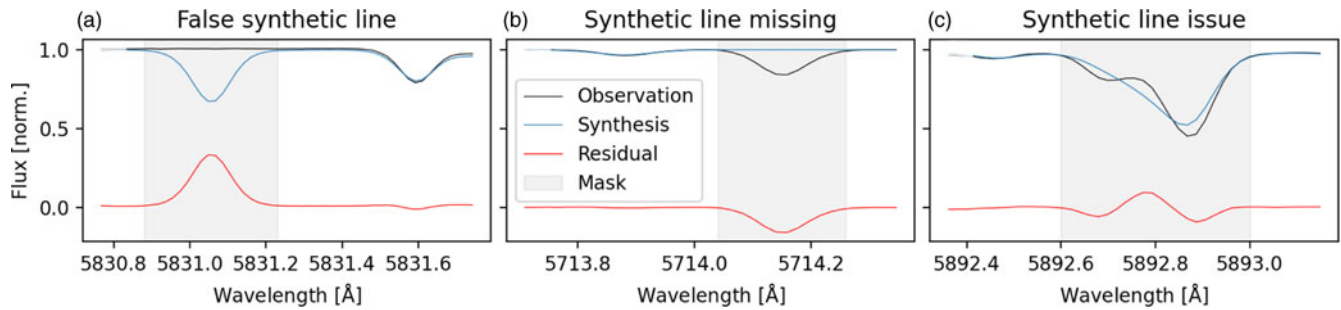
While the optimisation of the neural network selection has not converged (the final parameters  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  are not within the current 3D bin), the optimisation is repeated, starting with the previous best-fitting parameters as starting guesses.

We measured the time taken for the individual steps in the CURVE\_FIT function's execution to be approximately 80 ms. The total fitting process for stellar labels, including input/output

overheads, was timed at  $89^{+77}_{-29}\text{ s}$  for the allspec module, and  $125^{+81}_{-33}\text{ s}$  for the more complex allstar module.

#### 4.3.1. Which stellar labels are optimised?

As part of the synthetic grid computations, we have sampled each label of stellar parameters and elemental abundances individually between our chosen maximum and minimum ranges (see Section 3.1). This allows us to also judge which stellar labels to fit for each given star. We choose to fit a stellar label if either of these two cases applies to said label for the GALAH wavelength range when neglecting the cores of the Balmer lines: (i) The normalised spectrum between minimum and maximum label value at any pixel exceeds the threshold of 0.007 or (ii) The normalised spectrum between the minimum and maximum value changes by more than 0.005 for at least 25% of the pixels. While the first case



**Figure 10.** Examples of masks applied to unreliable pixels for the spectrum fitting, which is done by the minimisation of residuals (red) between observation (black) and synthesis (blue). Panel (a) shows a strong synthetic line, where no line is observed in the data. Panel (b) shows an observed line without any line being synthesised. Panel (c) shows significant disagreement between the two observed lines and the synthesis.

is constructed for atomic lines, such as Li I 6 708 Å, the second case addresses in particular molecular lines like the C<sub>2</sub> and CN lines. The pipeline can handle missing arms, for example in the case of readout issues of a CCD, and will fix abundances to the scaled-Solar values for elements with absorption features solely in the missing arm, for example N, O, K, and Rb for CCD4.

Initial tests of the pipeline have revealed that in cases where the initial parameter estimates deviate significantly from the final values, several elemental abundance estimates were shifted towards their boundaries, leading to a masking of their elemental abundance lines by the masking module (Section 4.3.2) at the beginning of each optimisation loop. To minimise this effect, we therefore shift the interim abundance values towards the narrow label boundaries. In practice, we limit the initial and interim abundances to 1.05:3.26 for A(Li),  $[X/Fe] = -0.5..1.0$  for C, N, O, Y, Ba, La, Ce, and Nd, and  $[X/Fe] = -0.5..0.5$  for all other elements before optimising them again. For warm and hot stars ( $T_{\text{eff}} > 6\,000$  K), this effect was seen to affect multiple abundances, such that we needed to implement a reset of all abundances except Li to their initial values for stars above 6 000 K, which would on average be expected to be young and have a Solar-like composition.

#### 4.3.2. Masking of unreliable wavelength regions

Not all pixels of the observed or synthetic spectra might prove useful for estimating reliable stellar labels. Observations can include bad pixels/patterns and incorrect corrections (for example of telluric or sky lines). Flux predictions of synthetic spectra are only as good as the input physics (limited for example for specific lines via uncertain oscillator strengths).

To minimise the influence of inaccurate synthetic pixel predictions, we have compared a 2dF-HERMES observation of the asteroid 4 Vesta and a high-quality Solar spectrum from Hinkle et al. (2000) with the flux that would be predicted by our pipeline for a star with Solar labels ( $T_{\text{eff}} = 5\,772$  K,  $\log g = 4.438$  dex,  $[Fe/H] = 0.00$  dex,  $v_{\text{mic}} = 1.06$  km s<sup>-1</sup>,  $v \sin i = 1.6$  km s<sup>-1</sup>,  $v_{\text{mac}} = 4.2$  km s<sup>-1</sup> Prša et al. 2016; Jofré et al. 2017, and  $[X/Fe] = 0.00$  dex for the default Solar abundance pattern for MARCS by Grevesse, Asplund, & Sauval 2007).

We have identified all lines that showed differences of the normalised flux of more than 0.1, lines where either a synthetic line or an observed one was completely missing, or lines that

were significantly misaligned. Examples of masks<sup>8</sup> are shown in Fig. 10. To avoid the influence of bad spectrum regions with an observational origin, we mask pixels where the synthetic and re-normalised observed spectra differ by more than  $5\sigma$  or a flux of 0.3 (0.4 before the initial optimisation). To avoid the masking of lines that are vital for our spectroscopic analysis, we have created a list<sup>h</sup> with segments of such lines that is mainly based on the previous element masks from GALAH DR3 (Buder et al. 2021). The final mask of pixels to use for the optimisation then includes all vital line regions, as well as those wavelengths that do not show a too strong disagreement between observation and synthesis and are not deemed unreliable in their synthesis.

In addition to this default masking, we exclude pixels for each major iteration, for which the flux of observation and synthesis differ by more than  $5\sigma$  and 30% of the normalised flux and by more than 100% of the normalised flux for the vital line regions.

We further indirectly take into account the currently less constrained molecular data for cool stars in optical spectra, in particular towards the blue (e.g. Rains et al. 2021; Rains et al. 2024). For presumably cool stars (with initial  $T_{\text{eff}} < 4\,100$  K), we therefore double the observational uncertainty of the blue arm.

## 5. Single star analysis (ALLSTAR)

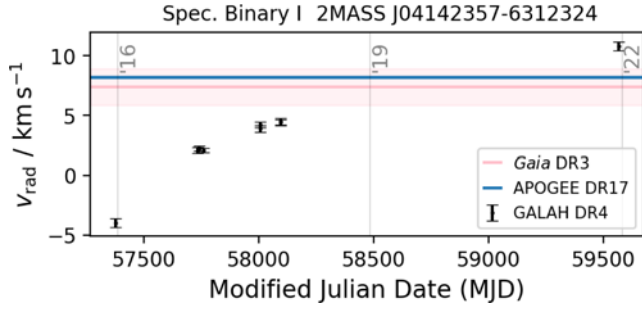
After the allspec module (Section 4) has been used to estimate spectroscopic labels for all spectra, we use the allstar module to co-add spectra and analyse one spectrum per star while taking into account photometric and astrometric information to constrain the surface gravities.<sup>i</sup> The optimisation of stellar spectroscopic parameters with the help of non-spectroscopic information was successfully applied for GALAH DR3 (Buder et al. 2021), using Gaia DR2 distances (Bailer-Jones et al. 2018) to overcome spectroscopic degeneracies. For the co-adding, we test whether the radial velocity estimates of individual exposures agree within  $2\sigma$ . Below this threshold, we apply no radial velocity correction and fit a global radial velocity. Above this threshold (which is useful for single-lined spectroscopic binaries as shown in Fig. 11), we apply a radial velocity correction before co-adding.

<sup>8</sup>Example masks can be found in the GALAH DR4 repository here.

<sup>h</sup>The list is available in the GALAH DR4 repository here.

<sup>i</sup>In line with Nissen (2015), Nissen et al. (2020), we refer to these non-spectroscopically constrained surface gravities as photometric ones.





**Figure 11.** Example of radial velocity evolution over modified Julian Date (vertical lines show the beginning of 2016, 2019, and 2022) for a single-lined spectroscopic binary (SB1).

To speed up computation, we use the mean results of the *allspec* analyses as initial stellar labels for the *allstar* analysis. All other methodology of the comparison of synthetic spectra to observations (Section 4.2) and label optimisation (Section 4.3) apply also to this module, with the exception of the optimisation of  $\log g$ . Contrary to the *allspec* approach, we do not fit  $\log g$  in this module, but estimate the logarithmic surface gravity  $\log g$  using a combination of its definition ( $g \propto \frac{M}{R^2}$ ) and the Stefan-Boltzmann law relative to the Solar values:

$$\log g = \log g_{\odot} + \log \frac{M}{M_{\odot}} + 4 \log \frac{T_{\text{eff}}}{T_{\text{eff},\odot}} - \log \frac{L_{\text{bol}}}{L_{\text{bol},\odot}} \quad (8)$$

While we can use our spectroscopically determined  $T_{\text{eff}}$  in Equation (8), the other values have to be estimated through models or non-spectroscopic information. The logarithmic bolometric luminosity,  $L_{\text{bol}}$ , can be estimated from the bolometric magnitude  $M_{\text{bol}}$ , such that  $\log \frac{L_{\text{bol}}}{L_{\text{bol},\odot}} = -0.4 \cdot (M_{\text{bol}} - M_{\text{bol},\odot})$ . The bolometric magnitude can be estimated from any given apparent magnitude, if we correct the latter for the distance modulus, bolometric correction, and extinction. Because essentially all stars in GALAH DR4 have high-quality infrared magnitudes available that suffer less from (uncertain) extinction corrections, we use  $K_S$  as the magnitude to estimate our bolometric magnitudes and luminosities via

$$M_{\text{bol}} = K_S - 5 \cdot \log \frac{D_{\varpi}}{10} + BC(K_S) - A(K_S). \quad (9)$$

While the values for  $K_S$ , curated distances  $D_{\varpi}$  (rather than raw parallaxes  $\varpi$ ), and  $A(K_S)$  are readily available (see Section 2.3), we need to estimate the bolometric correction from tabulated values using the routines provided by Casagrande & Vandenberg (2018):

$$BC(K_S) = f(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}]) \quad (10)$$

We choose to assume an extinction value of  $E(B - V) = 0$  mag for this particular interpolation and post-correct the value by  $A(K_S)$  based on the actual extinctions. The reason for this is that the latter values can exceed the maximum tabulated values of  $E(B - V) = 0.72$  mag of Casagrande & Vandenberg (2018).

Because of the appearance of  $\log g$  in Equation (10), we iterate the calculation of  $BC(K_S)$  and subsequently  $\log g$  up to four times or until the latter value changes less than 0.02 dex between iterations. Similarly, we need to estimate the stellar masses (and ages as a byproduct) from tabulated values, that is,

$$M, \tau = f(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], L_{\text{bol},\odot}) \quad (11)$$

For this on-the-fly estimate of masses and ages we use a likelihood-weighted estimate with default uncertainties of 100 K, 0.25 dex, 0.2 dex, respectively, and an average uncertainty of  $L_{\text{bol},\odot}$  from propagated uncertainties of Equation (9). We weigh the ages and masses via their likelihood of all isochrone grid points within these uncertainties of the PARSEC+COLIBRI isochrones (Bressan et al. 2012; Marigo et al. 2017), which cover the logarithmic ages of  $\log(\tau/\text{Gyr}) = 8.00..(0.01)..10.18$  by default and metallicities  $[\text{M}/\text{H}] = -2.75..(0.25)..-0.75$  as well as  $[\text{M}/\text{H}] = -0.6..(0.1)..0.7$ . We exclude hot stars above 10 000 K as well as extremely evolved white dwarf and extremely luminous giant stars ( $\log g > 6$  dex or  $J - K_S > 2$  mag) as they fall far outside our spectroscopic pipeline range. We convert between the theoretical  $[\text{M}/\text{H}]$  and our measured  $[\text{Fe}/\text{H}]$  as well as an assumed  $[\text{Fe}]$  enhancement<sup>1</sup> via the correlation of Salaris & Cassisi (2006),  $[\text{M}/\text{H}] = [\text{Fe}/\text{H}] + \log(10^{[\alpha/\text{Fe}]} \cdot 0.694 + 0.306)$ . For open clusters with age estimates below 1 Gyr as well as unevolved stars that are more luminous than expected from the oldest cool main-sequence isochrone with matching  $[\text{M}/\text{H}]$ , we sample  $\log(\tau/\text{Gyr}) = 6.19..(0.01)..10.18$ . For globular cluster stars identified in the crossmatch with Baumgardt & Vasiliev (2021), we limit the isochrones to a minimum age of 4.5 Gyr.

## 6. Post-processing

After the *allspec* and *allstar* modules have been run for a night's data (see Sections 4 and 5, respectively), a post-processing routine is used to estimate additional parameters from the residuals of the spectra (Section 6.1), estimate and validate accuracy and precision uncertainties (Section 6.2), and perform quality assurance tests on a global scale (*flag\_sp*, see Section 6.3) as well as for the individual abundances of elements X (*flag\_X\_fe*, see Section 6.4).

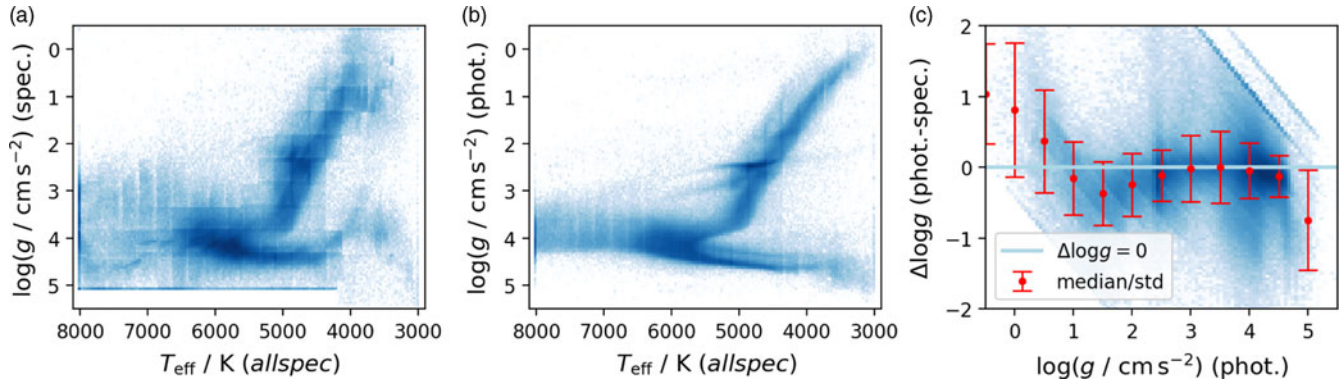
### 6.1. Analysis of spectral residuals

#### 6.1.1. Binary signatures

The residual spectrum of our best-fitting single star analysis can help us to identify a second flux contributor to the observed spectrum. In our case, there are two points in the analysis where we can identify such an influence. Firstly, the residuals are visible in the  $\chi^2$  distribution as a function of radial velocity shifts (see Fig. 9). While a single star would only show one peak (saved as *rv\_comp\_1*), a binary system like 2MASS J06084657-7815235 shows a second peak ( $-70$  km s<sup>-1</sup> in addition to 74 km s<sup>-1</sup>) that is saved as *rv\_comp\_2*. Secondly, we perform an automatic search for reoccurring residuals as a function of radial velocity for a few selected lines. We chose the combination of strong lines in the spectra (Balmer lines, Fe lines at 4 890 and 4 891 Å, Ni at 6 644 Å) as well as those with the largest expected wavelength shift in the infrared detector (O triplet at 7 772–7 775 Å as well as Mg at 7 692 Å). If we find several peaks with a reasonably similar radial velocity, the likely  $X \in 16, 50, 84$ th percentiles of this radial velocity are saved in *sb2\_rv\_X*.

Because radial velocities from the *Gaia* radial velocity spectrometer (Katz et al. 2023) are reported in *Gaia* DR3 for 94% (774 914) of the stars observed for GALAH DR4, we can also compare against those radial velocity estimates. For 6% (50 577) of

<sup>1</sup>We assume  $[\alpha/\text{Fe}] = 0.4$  for  $[\text{Fe}/\text{H}] < -1$ ,  $[\alpha/\text{Fe}] = 0.0$  for  $[\text{Fe}/\text{H}] > 0$  and linearly interpolate between these points for  $-1 \leq [\text{Fe}/\text{H}] \leq 0$ .



**Figure 12.** Comparison of spectroscopic and photometric  $\log g$  estimates in the *allspec* analysis. Panel (a) shows the distribution of spectroscopic  $\log g$  and  $T_{\text{eff}}$  from the *allspec* module. Panel (b) shows the distribution of the same  $T_{\text{eff}}$  and photometric  $\log g$ . Panel (c) shows the difference of photometric  $\log g$  and spectroscopic  $\log g$  as a function of photometric  $\log g$ . Red error bars indicate the  $1\sigma$  percentiles of this difference in 0.5 dex bins.

our stars, we find a difference with respect to *Gaia* DR3 larger than  $10 \text{ km s}^{-1}$ . For these stars, we often noticed unrealistically high  $v_{\text{mic}}$  and  $v \sin i$  or negative velocities in our *allspec* analysis. We note that the *allspec* analysis was run without boundary conditions for global parameters and thus also resulted in negative velocities, which are later flagged and might indicate binarity (Section 6.3). *allstar*, however, was run with  $v_{\text{mic}}$  and  $v \sin i$  forced to be above  $0 \text{ km s}^{-1}$ .

### 6.1.2. Post-correction of $\log g$ for *allspec* results

While we estimate logarithmic surface gravities  $\log g$  solely from spectra in the *allspec* results, we also perform a post-processing estimate where we employ the methodology of Section 5 while fixing all other stellar parameters. The approach of only using spectroscopic information confirmed the previous conclusions of GALAH DR1-DR3 that the spectroscopic information in HERMES spectra to estimate  $\log g$  is not sufficient for the majority of the parameter space for the given SNR. We show the spectroscopic  $\log g$  in Fig. 12a and the photometric  $\log g$  and their difference in Fig. 12b and c, respectively.

We see an overall good agreement of both  $\log g$  estimates for stars between  $4250 < T_{\text{eff}} < 6500 \text{ K}$ . Hotter stars show a strong dispersion of spectroscopic  $\log g$  due to limited information from fewer and shallower lines. Cooler stars show a significant trend towards much lower  $\log g$  for main-sequence stars and much higher  $\log g$  for cool evolved stars up to an order of  $\Delta \log g$  of 1 dex. This trend was previously seen in GALAH DR2 (Buder et al. 2018) and is believed to be caused by the onset of molecular absorption features which suppress the continuum for almost the entire HERMES wavelength range (see for example Fig. 8), thus introducing several degeneracies. In addition, we can notice a significantly lower precision of the spectroscopic  $\log g$  in comparison to the excellent precision of photometric  $\log g$ , for example in the red clump stars.

On closer inspection, we notice several trends in Fig. 12a. Most notably, we see nodding patterns along the  $T_{\text{eff}}$  and  $\log g$  grids where the *allspec* module switches between different neural network models. Our investigation of these nodding effects is addressed in Section 8. In comparison to Fig. 12b, where a clear equal-mass binary sequence is visible just above the cool main-sequence, we do not see such a sequence in Fig. 12a. The difference

between spectroscopic and photometric  $\log g$  will therefore be useful to identify photometric binaries with high quality spectra with  $\log g$  precisions below the single to binary system offset of up to  $\Delta \log g = 0.3$  dex, as discussed in Section 6.3. We caution, however, that the use of stellar structure models for the estimation of surface gravities can introduce systematic trends, as we discuss in Section 8.4.

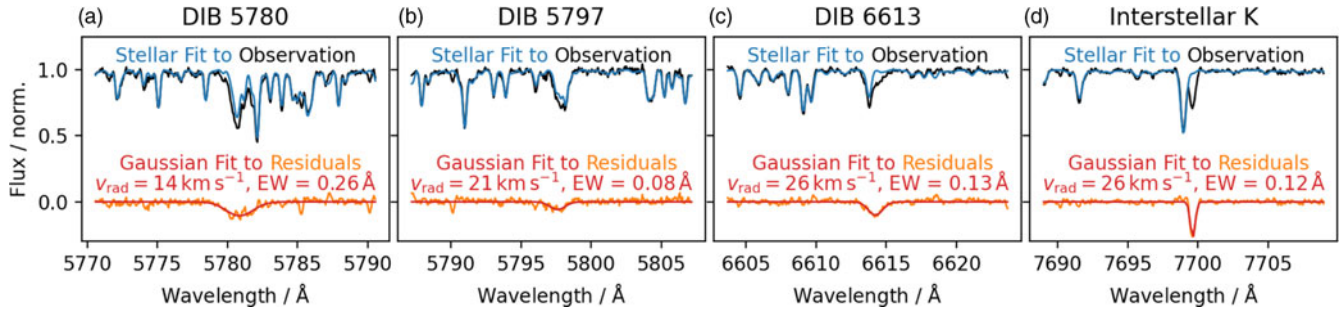
### 6.1.3. Interstellar absorption

Because we can create synthetic stellar spectra for the full wavelength range, we can now also trace interstellar absorption in the residuals of observed spectra. By default, we try to calculate the equivalent width via Gaussian fits to the three strongest diffuse interstellar bands (DIBs; 5 780.59, 5 797.19, 6 613.66 Å) with central wavelengths identified by Vogrinčič et al. (2023) as well as for interstellar K (7 698.9643 Å), see Fig. 13. We report the equivalent widths  $\text{eq\_x}$ , standard deviations  $\text{sigma\_x}$  and radial velocities  $\text{rv\_x}^k$  for  $\text{x}$  in  $\text{k\_is}$  for interstellar K and  $\text{x}$  in  $\text{DIB\_5780}$ ,  $\text{DIB\_5797}$ , and  $\text{DIB\_6613}$  for the DIBs. The coverage of interstellar material, estimated from DIB\_5780, within  $D_{\text{av}} < 5 \text{ kpc}$  is shown in an all-sky map in Fig. 14, with the GSPH extinction by Andrae et al. (2023) in the background.

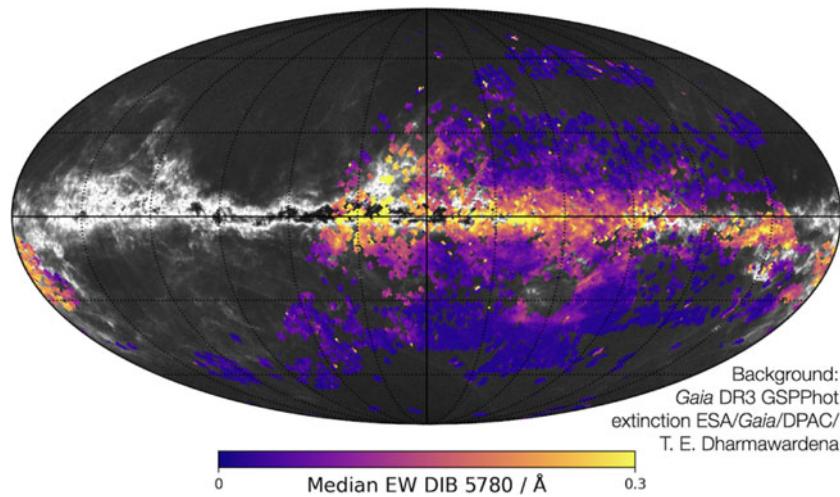
### 6.1.4. Emission estimates for the Balmer lines

The shape of the Balmer absorption lines holds valuable information on active stars as well as masses for evolved stars (Bergemann et al. 2016) and possibly even information on unresolved binary systems (Sayeed et al. 2024). Although the cores of these lines suffer from inaccuracies in the synthesis, the residuals of synthetic and observed lines can be used in relative analyses. We therefore perform a trapezoidal integration around the Balmer lines of each normalised spectrum at 4 861.3230 and 6 562.7970 Å whose values we report in  $\text{ew\_h\_beta}$  and  $\text{ew\_h\_alpha}$ . By default we integrate in a window of  $\pm 0.75$  and  $1.25 \text{ Å}$  for  $\text{H}\beta$  and  $\text{H}\alpha$ , respectively, and increase this window to  $5 \text{ Å}$  if the average observed, normalised flux within  $\pm 0.5 \text{ Å}$  of the Balmer line core exceeds 1. An example of such a star is shown in Fig. 15, for which we measure a residual EW of  $-1.09 \text{ Å}$ . Most emission line stars in the GALAH sample are found in the region of pre-main-sequence and hot stars

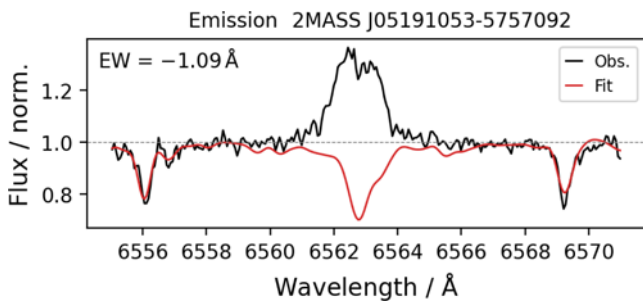
<sup>k</sup>In v240705,  $\text{rv\_comp\_1}$  has to be added to  $\text{rv\_k\_is}$  due to a bug.



**Figure 13.** Example of three diffuse interstellar bands (DIBs) and interstellar K absorption for 2MASS J06453479-0102137 with an  $E(B - V) = 0.84$  mag value from Schlegel et al. (1998). Shown are the observation (black) and stellar fit (blue) as well as a Gaussian fit (red) to the residual (orange), resulting in an estimate of the equivalent width (EW) as well as radial velocity.



**Figure 14.** All-sky map (l, b) of GALAH DR4 equivalent width measurements of the diffuse interstellar band around 5780 Å, with the GSPPhot extinction by Andrae et al. (2023) in the background.



**Figure 15.** Example of a flagged emission star with clear emission in the Balmer lines (here H $\alpha$ ).

(see Fig. AC6a). We conservatively only flag stars with a median normalised flux above 1 in H $\beta$  or H $\alpha$  as emission line stars.

## 6.2. Uncertainty estimation and validation

The uncertainties that we report for our spectroscopic data analysis are based on the comparison to literature measurements (see also Beeson et al. 2024) to estimate accuracy uncertainties and a combined precision uncertainty estimate from adjusted

covariance estimates from the fitting process and the scatter of repeat observations. Formally, we estimate the total variance of measurements as a combination of the accuracy and precision variance

$$\sigma_{\text{total}}^2 = \sigma_{\text{accuracy}}^2 + \sigma_{\text{precision}}^2 \quad (12)$$

Representative values of accuracy and precision for our stellar parameters are listed in Table 4. We lay out how we estimate and validate accuracy and precision uncertainties in Sections 6.2.1 and 6.2.2, respectively.

### 6.2.1. Accuracy estimation and validation

Estimating the accuracy of spectroscopic measurements has always been a complicated endeavour, because there are no universal benchmark sets for all parameters across all stellar types. Subsequently, we describe the numerous comparisons that we have performed for both stellar parameters ( $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ,  $v_{\text{mic}}$ ,  $v \sin i$ , and  $v_{\text{rad}}$ ) as well as the elemental abundance measurements. Consistent with GALAH DR3 (Buder et al. 2021), and caused by the limited coverage of benchmark literature, we continue to use a single accuracy estimate for each stellar parameter



**Table 4.** List of accuracy and representative precision uncertainties for stellar parameters in GALAH DR4. Accuracy values are estimated from comparisons with literature references (see Section 6.2.1), whereas precision estimates are estimated from covariance uncertainties and repeat observations (Section 6.2.2). Here, we list the median precision uncertainties for stars with  $SNR = 50 \pm 10$  on CCD2 (see Fig. 20).

| Parameter/Unit                    | Accuracy | Precision ( $SNR = 50$ ) |
|-----------------------------------|----------|--------------------------|
| $T_{\text{eff}}/\text{K}$         | 66       | $23 \pm 5$               |
| $\log(g)/\text{cm s}^{-2}$        | 0.042    | –                        |
| $[\text{Fe}/\text{H}]/\text{dex}$ | 0.051    | $0.025 \pm 0.004$        |
| $v_{\text{mic}}/\text{km s}^{-1}$ | 0.28     | $0.05 \pm 0.03$          |
| $v \sin i/\text{km s}^{-1}$       | 1.4      | $0.5 \pm 0.2$            |
| $v_{\text{rad}}/\text{km s}^{-1}$ | 0.15     | $0.17 \pm 0.02$          |

and ignore the possibly large accuracy uncertainties for the individual elemental abundances. In all cases, we estimate an overall bias with respect to literature values and then combine these estimates to a globally applied zero-point correction. Where not explicitly stated otherwise, we assume that the spread of stellar parameters residuals is indicative of the accuracy of either method and estimate our accuracy by dividing the parameter spread by  $\sqrt{2}$ . The applied shifts are listed in Table C1. We estimate the accuracy and bias correction for stellar parameters (including the iron abundance as a global parameter) and abundances separately.

Our primary reference source for parameter accuracy remains the *Gaia* FGK benchmark stars (Jofré et al. 2014; Jofré et al. 2015; Jofré et al. 2018; Heiter et al. 2015). Additionally, we use asteroseismic estimates from the K2 and TESS photometry (Zinn et al. 2020; Hon et al. 2021) to compare our surface gravities and perform a validation to higher quality observations of globular cluster stars with typically lower metallicities (Carretta et al. 2009a; Carretta et al. 2009b; Johnson & Pilachowski 2010). Because the overlap with APOGEE DR17 (Abdurro’uf et al. 2022) has increased from 41 941 stars in GALAH DR3 to 60 046 stars with 92 368 repeat observation matches in GALAH DR4, we also can assess systematic trends for a larger parameter space. For clarity, we discuss the stellar parameters separately, but show most accuracy estimates in a combined Fig. 16.

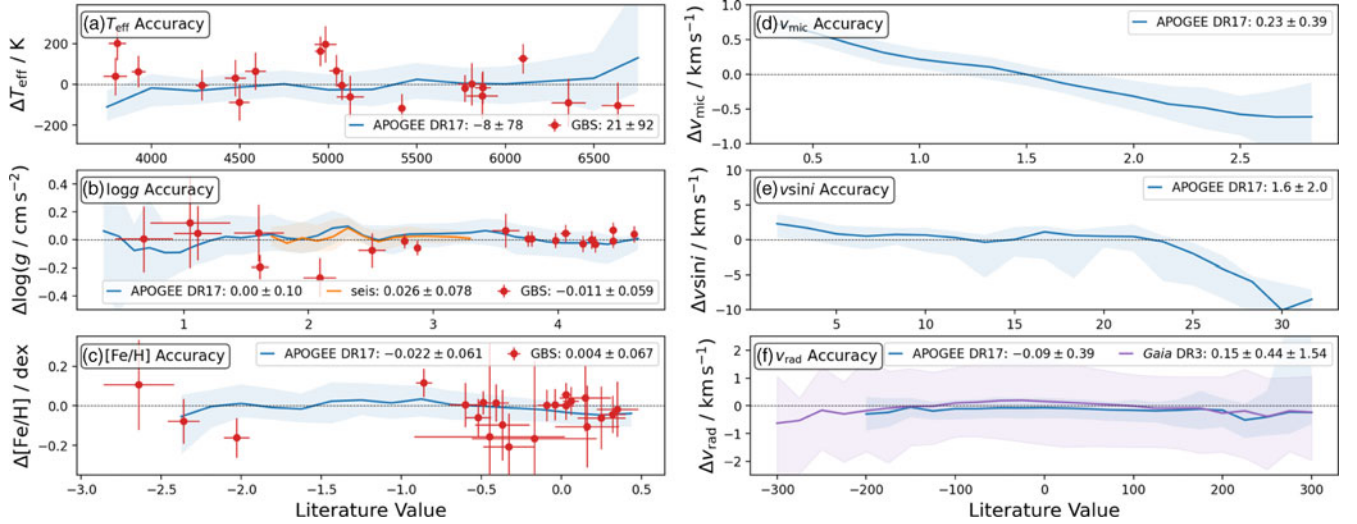
$T_{\text{eff}}$ : The effective temperature estimates from GALAH DR4 show good agreement with the *Gaia* FGK benchmark stars (Fig. 16a). Specifically, we find a mean difference of  $\Delta T_{\text{eff}} = 21 \pm 92$  K, indicating no significant bias between our temperatures and the benchmark values. Comparisons with APOGEE DR17 show an equally robust agreement, with  $\Delta T_{\text{eff}} = -8 \pm 78$  K. This small offset and uncertainty suggest that the GALAH DR4  $T_{\text{eff}}$  estimates are highly reliable across a wide range of at least G- and K-, but possibly also F- and M-type stars. Here, we use  $1/\sqrt{2}$  of the residual spread with respect to *Gaia* benchmark stars as our accuracy estimate.

$\log g$ : For surface gravity, we compared our  $\log g$  estimates to both the *Gaia* FGK benchmark stars, asteroseismic measurements from Zinn et al. (2020) and Hon et al. (2021), and APOGEE DR17. The asteroseismic  $\log g$  values are derived from  $v_{\text{max}}$  measurements for giant stars, and they show excellent agreement with our results, with a mean difference of  $\Delta \log g = 0.026 \pm 0.078$ . Both the asteroseismic comparison as well as the *Gaia* benchmark star comparison ( $\Delta \log g = -0.011 \pm 0.059$ ) and APOGEE DR17

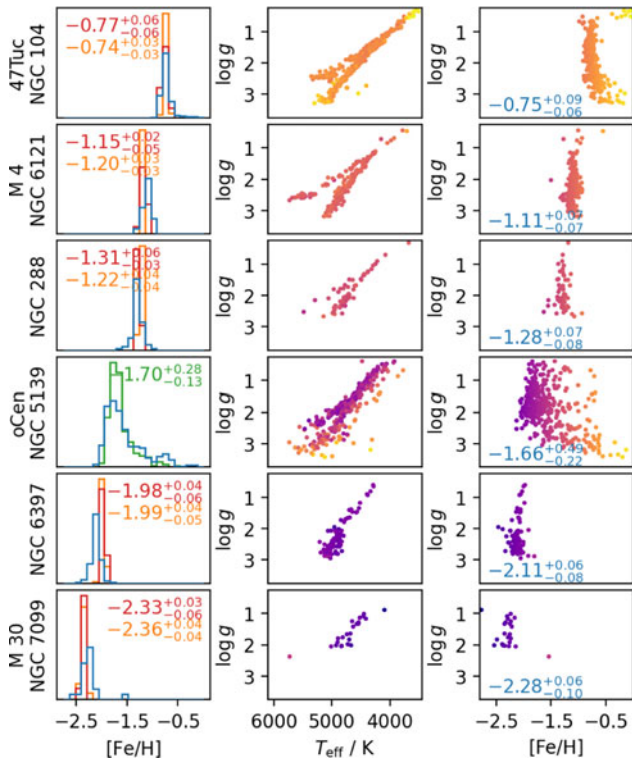
( $\Delta \log g = 0.00 \pm 0.10$ ) agree well and show no trends across the  $\log g$  range. For the low metallicity regime, we compare GALAH  $\log g$  values with asteroseismically derived values from Howell et al. (2022) for the globular cluster M 4 (NGC 6121). Stars from this cluster were observed as part of a dedicated survey (PI M. Howell) aimed at spectroscopically characterising their sample of stars observed by the K2 mission (Howell et al. 2014). Across the 75 overlapping targets, we find a  $\Delta \log g = 0.056 \pm 0.128$ . The comparison between independently derived light element abundance variations and asteroseismic masses will be presented in an upcoming paper (Howell et al., in preparation.). This is a significant improvement over GALAH DR3, where significant deviations were found for luminous giant stars – whose parameter estimates in GALAH DR3 suffered from less precise and systematically biased distance and thus  $\log g$  estimates. We find significant outliers, however, particularly for primary red clump stars, which were mistaken as secondary red clump stars, leading to larger deviations. We discuss this issue later in Section 8.4. Because this single group is driving the scatter in our disagreement with the asteroseismic estimates, we revert to the *Gaia* benchmark stars to estimate the accuracy.

[Fe/H]: The comparison of GALAH DR4 metallicities to the *Gaia* FGK benchmark stars initially showed the similar bias of GALAH towards more metal-poor values at the 0.049 dex level. The application of a zero-point correction (see Table C1) yields an excellent agreement, with  $\Delta[\text{Fe}/\text{H}] = 0.004 \pm 0.067$  for the benchmark stars and  $\Delta[\text{Fe}/\text{H}] = -0.022 \pm 0.061$  for APOGEE DR17, confirming the reliability of the GALAH DR4 metallicity estimates across a large range of metallicities. For the metal-poor regime, benchmark estimates are still rare. Luckily, a dedicated observing program – whose results are included in this data release – was performed and an overview of globular cluster Kiel diagrams is appended in Fig. AC5. We therefore only perform a comparison with globular cluster stars – often measured in 1D LTE – to get a quantitative impression of the agreement. We restrict ourselves to a few studies, namely those by Carretta et al. (2009a,b) for NGC 104, 6121, 288, 6397, and 7099 as well as Johnson & Pilachowski (2010) for NGC 5139. In all cases, we find a good agreement of the metallicity distribution function for overlapping stars within the uncertainties (see Fig. 17). While this does not necessarily confirm our accuracy, it shows consistency within this uncertain parameter regime. We note however, a specific region in NGC 104, where the metallicity of the most luminous giants ( $T_{\text{eff}} < 3\,750$  K and  $\log g < 0.5$ ) is incorrectly estimated near the Solar value. We discuss this problem in detail as a caveat in Section 8.7, since we have not been able to systematically flag these stars. More generally, we note that the strong and unexpected abundance trends with  $T_{\text{eff}}$  or  $\log g$  in globular clusters from GALAH DR3 have decreased for most elements. However, we still urge users to take caution when using globular cluster abundances, and we discuss this further in Section 8.6. A custom, by hand analysis of globular cluster abundances beyond [Fe/H] will be performed in a separate study (McKenzie et al., in preparation), as these observations have been part of a dedicated observing program (PIs M. McKenzie and M. Howell). Similarly, a dedicated verification of open cluster observations (PIs J. Kos and G. De Silva) will be performed in a separate study (Kos et al. 2025).

$v_{\text{mic}}$ : Microturbulence velocities show a more complex pattern when compared to APOGEE DR17. We find a mean difference of  $\Delta v_{\text{mic}} = 0.23 \pm 0.39$  km s<sup>−1</sup>. However, the comparison reveals a linear mismatch: APOGEE DR17 tends to measure lower  $v_{\text{mic}}$



**Figure 16.** Accuracy of the main stellar parameters  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ,  $v_{\text{mic}}$ ,  $v \sin i$ , and  $v_{\text{rad}}$  for GALAH DR4. Each panel shows the comparison to literature (DR4 – literature) with median values as lines and contours between 16th and 84th percentiles. Comparisons are performed for the *Gaia* FGK Benchmark stars (red), APOGEE DR17 (blue),  $\log g$  inferred from asteroseismic measurements (orange) and *Gaia* DR3 radial velocities (purple).

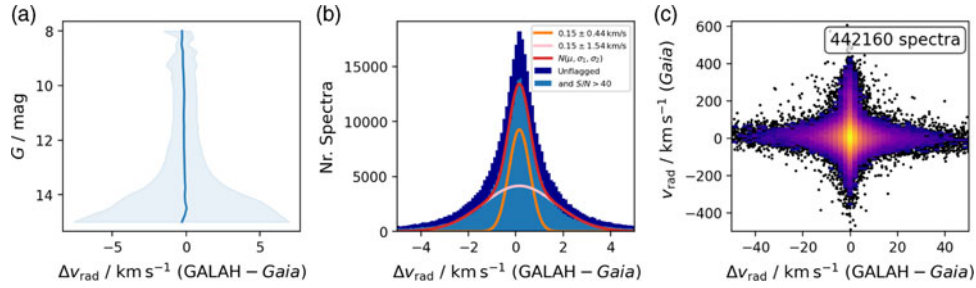


**Figure 17.** Comparison of iron abundances (16th, 50th and 84th percentiles) and overview of spectroscopic and photometric properties of globular cluster stars in GALAH DR4. Left panels show histograms of iron abundances from GALAH DR4 (blue) as well as literature estimates for the globular clusters from Giraffe (orange) and UVES (red) observations by Carretta et al. (2009a, b) as well as observations from Johnson & Pilachowski (2010). Middle panels show the spectroscopic  $T_{\text{eff}}$ - $\log g$  diagrams coloured by iron abundance  $[\text{Fe}/\text{H}]$ . Right panels show the trend of GALAH DR4  $[\text{Fe}/\text{H}]$  along the different  $\log g$  values.

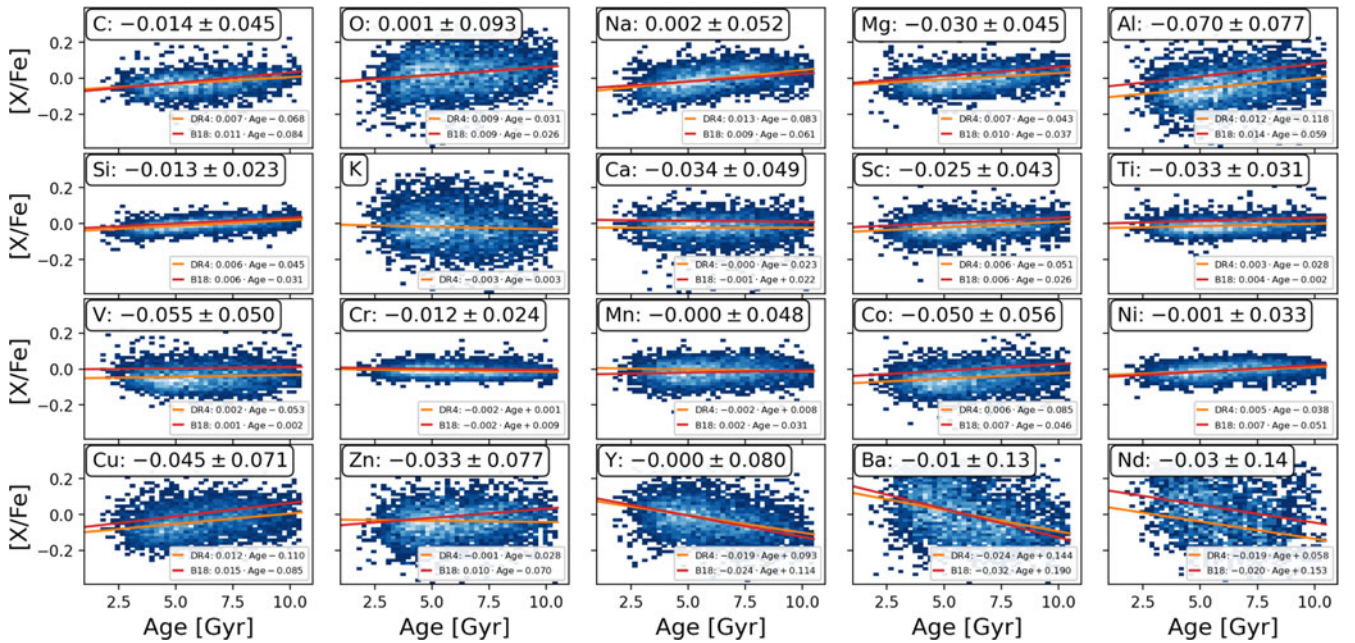
values for stars with low microturbulence and larger  $v_{\text{mic}}$  values for stars with higher microturbulence compared to GALAH DR4. This systematic trend suggests that the  $v_{\text{mic}}$  calibration between the two surveys may differ slightly, particularly at the extremes of the parameter range. We note, however, that the surveys agree much better than for GALAH DR3, where a fixed quadratic relation was used that did not allow for deviations, for example for red clump stars. Adding  $v_{\text{mic}}$  as free parameter returned a similar pattern as the empirical relation by Dutra-Ferreira et al. (2016) and shows a significantly different behaviour of  $v_{\text{mic}}$  for the hottest, coolest, and red clump stars (see Fig. AA1). This mismatch of  $v_{\text{mic}}$  could have indeed driven the metallicity mismatch of metal-rich red clump stars in GALAH DR2 and DR3 (Buder et al. 2018, 2021), since their metallicities are in agreement with other estimates now (e.g. APOGEE DR17).

$v \sin i$ : The rotational velocity estimates agree well with APOGEE DR17, with a mean difference of  $\Delta v \sin i = 1.6 \pm 2.0 \text{ km s}^{-1}$ . However, at higher rotational velocities (above approximately  $24 \text{ km s}^{-1}$ ), our neural networks start to extrapolate, leading to an upper limit in the estimates and returning significantly lower  $v \sin i$  values compared to APOGEE DR17. This issue highlights the limitations of the GALAH DR4  $v \sin i$  estimates for rapidly rotating stars.

$v_{\text{rad}}$ : For radial velocity we compared our results to both APOGEE DR17 and *Gaia* DR3. The comparison with APOGEE DR17 yields a small offset of  $\Delta v_{\text{rad}} = -0.09 \pm 0.39 \text{ km s}^{-1}$ , indicating excellent agreement between the two surveys. Accounting for the much lower SNR for faint *Gaia* targets and unidentified binaries, we fit two Gaussian distributions to the overall difference of GALAH and *Gaia* radial velocities (see Fig. 18). The comparison with *Gaia* DR3 shows a slightly larger offset of  $\Delta v_{\text{rad}} = 0.15 \pm 0.44 \pm 1.54 \text{ km s}^{-1}$ , which is expected due to the lower precision of the *Gaia* DR3 radial velocities (Katz et al. 2023). We use the median residual of  $0.15 \text{ km s}^{-1}$  with respect to *Gaia* DR3 rather than the spread as our accuracy estimate.



**Figure 18.** Comparison of radial velocities between GALAH DR4 allspec and *Gaia* DR3. Panel (a) shows the difference of radial velocities as function of *Gaia*  $G$  magnitude. Panel (b) shows a histogram of the difference with two Gaussian distributions (with same mean) fitted to them to estimate a more robust, binary independent, radial velocity difference. Panel (c) shows the difference of radial velocities as function of radial velocity, showing the systematic scatter introduced by binaries.



**Figure 19.** Chemical abundances  $[X/\text{Fe}]$  of Solar twin stars as a function of ages that were estimated as part of the mass and age estimation of the allstar spectrum analysis. We overplot linear fits to our age-abundance relations for Solar twins in orange and literature values from Bedell et al. (2018) in red. Panels also indicate the median and standard deviation with respect to Bedell et al. (2018) when assuming a correct age.

**Elemental abundances  $[X/\text{Fe}]$ :** While there is no model-independent benchmark for absolute abundance accuracy, we continually perform comparisons with literature values to assess the consistency of our results with other studies. In GALAH DR4, we evaluate the abundance zero-points using up to five different reference estimates (see Fig. AC1): (1) the spectroscopic analysis of a Solar-composition spectrum of the asteroid Vesta (subject\_id 210115002201239), (2) abundance estimates for Solar twins corresponding to a Solar age of 4.5 Gyr (see Fig. 19), (3) abundances of *Gaia* FGK benchmark stars (Jofré et al. 2015, 2018), (4) stars with Solar-like metallicity  $-0.1 < [\text{Fe}/\text{H}] < 0.1$  within 500 pc of the Sun (a method introduced by Jönsson et al. 2020), and (5) differences in abundance estimates for stars overlapping with the high-resolution, large-scale spectroscopic APOGEE DR17 survey (Abdurro'uf et al. 2022).

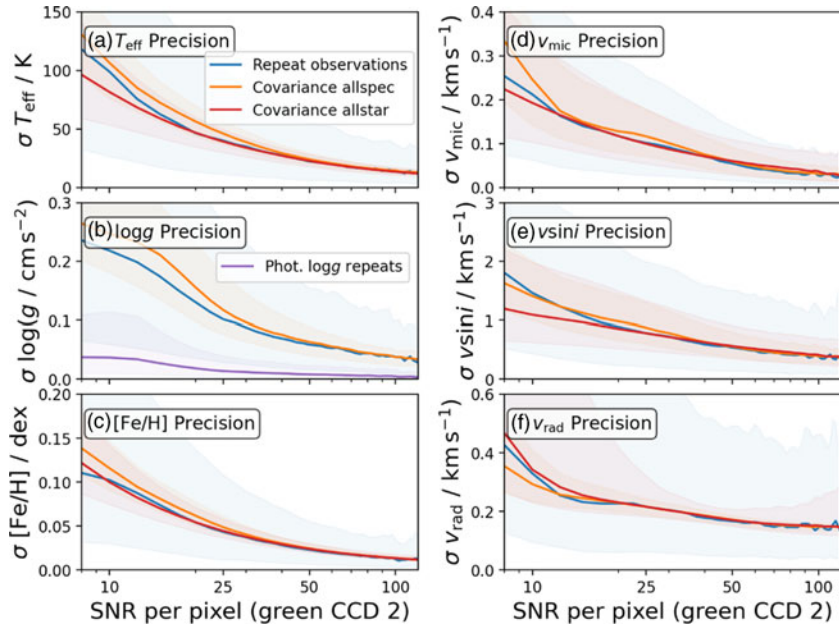
It is important to note that our abundance corrections, and consequently the Solar abundances presented in Table AC1, are

determined within the framework of 1D LTE or 1D NLTE models and are not intended to represent the most accurate Solar abundances. Instead, they reflect our best effort to minimise discrepancies across different comparison cases. Given the differences in line modelling, such as those between Grevesse et al. (2007) (who used 3D atmospheres) and our 1D models, and possible deviations in the reference abundance from our Vesta spectrum, we refer to these adjusted values as *zero-points*. For certain scientific applications, adjusting these abundance zero-points might be necessary to ensure consistency with other datasets.

While we are not able to include all of our validation plots, we refer the interested reader to the publicly available code in our code repository.<sup>1</sup> Generally speaking, we have found that a large number of systematic trends of abundances with temperature and

<sup>1</sup>[https://github.com/svenbuder/GALAH\\_DR4/tree/main/validation](https://github.com/svenbuder/GALAH_DR4/tree/main/validation).





**Figure 20.** Precision monitoring (with a median line and standard deviation shading) of stellar parameters as a function of SNR for the green CCD2 across GALAH DR4. Each panel shows the behaviour for bins of width 10 for the scatter of repeat observations of the *allspec* runs (blue), covariance uncertainties of *allspec* (orange) and *allstar* (red) setups as well as scatter of photometric  $\log g$  from repeat observations (purple).

surface gravity has decreased with respect to GALAH DR3, as can be appreciated from dedicated validation plots (online here) – with a similar appearance as the right hand panels of Fig. 17.

### 6.2.2. Precision estimation and validation

In addition to the accuracy uncertainty, we estimate the total uncertainty through the additional precision uncertainty (Equation 12). For this purpose, we mainly rely on the fitting uncertainties of the *curve\_fit* function, which we rescale based on repeat observations.

While we report the raw fitting covariance matrix for each spectrum and module (see Fig. AB2 for the covariance matrices of Vesta and Arcturus), their entries are not validated for reliability and have not been adjusted to incorporate a rescaling towards the final uncertainties. For the purposes of reporting stellar parameter and abundance fitting uncertainties, we restrict ourselves to the standard deviations of each feature, that is, the square root values of the diagonal covariance matrix entries.

Similar to GALAH DR3, we apply a precision adjustment of the fitting uncertainty towards consistency with the scatter of repeat observations only as a function of SNR of CCD2. Contrary to GALAH DR3, we have extended this rescaling function to be fitted in bins of SNR with both a constant, linear, and exponential term with  $\text{snr\_px\_ccd2}$  as the independent variable, that is,  $c_1 + c_2 \cdot \text{SNR} + c_3 \cdot \exp(c_4 \cdot \text{SNR})$ . We report the fitted constants for both *allspec* and *allstar* in the online repository<sup>m</sup> for each stellar parameter and abundance.

In Figs. 20 and AC2, we then confirm that the overall trends of fitting uncertainties for *allspec* and *allstar* are consistent with the repeat observation scatter of the *allspec*. The latter has to

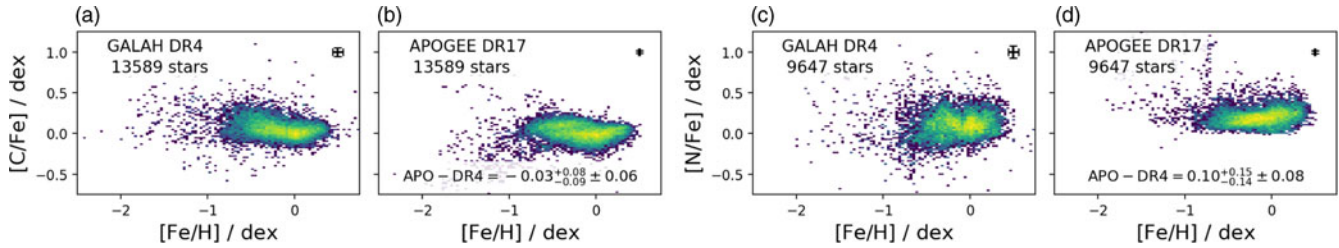
be used as reference, because the *allstar* module uses co-added spectra of repeat observations rather than the repeat observations themselves. While this might actually overestimate the precision uncertainty of stellar parameters, we do not expect a too strong overprediction for abundances.

While the precision levels of stellar parameters have on average actually remained similar to the estimates of GALAH DR3, we see notable improvement of the precision for multiple elements, such as C, Mg, V, Cr, Co, Ni, La, Ce, Nd, and Sm. The precision of Eu, however, seems to have decreased.

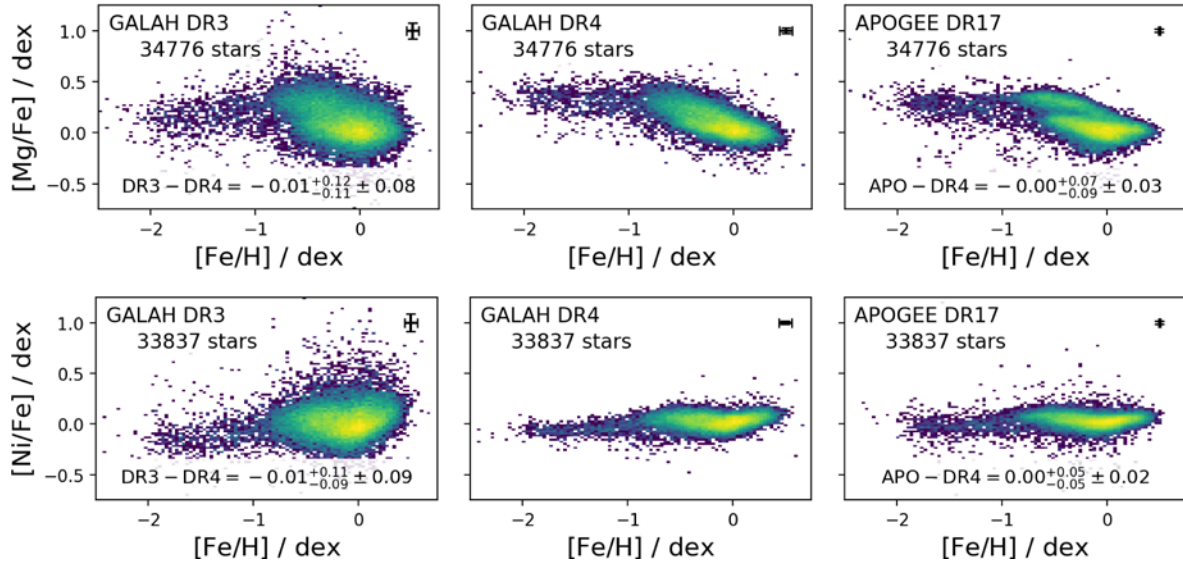
Separately from this work, we performed an extensive analysis of the precision and accuracy of spectroscopic parameters from the observation of star clusters, 43 open clusters of all ages and 10 globular clusters (Kos et al. 2025). In this work, we compare  $T_{\text{eff}}$ ,  $\log g$ , and stellar ages with the values obtained from cluster isochrone fitting. Ages show typical uncertainties of 10 to 50%, depending on the stellar type.  $T_{\text{eff}}$  and  $\log g$  match well for stars hotter than 4 000 K with a bias of  $\Delta T_{\text{eff}} = -68$  K (GALAH – Isochrones), and  $\Delta \log g = -0.03$ . For stars cooler than stars 4 000 K, GALAH DR4 temperatures are overestimated by up to 250 K at  $T_{\text{eff}} = 3$  000 K and we find a complicated pattern in  $\Delta \log g$ , with  $\log g$  being sometimes severely overestimated for the coolest dwarf stars.

Most interesting is the analysis of elemental abundances. Assuming that stellar clusters are chemically homogeneous, we can study the precision of the reported abundances over a large range of temperatures. We find that cold stars show consistent systematic trends, that can reach values of 0.5 dex for some elements. Dwarf stars are most affected at temperatures  $T_{\text{eff}} < 4$  600 K, while giants show much smaller systematics with strong trends only at  $T_{\text{eff}} < 4$  000 K. The results of this cluster validation (Kos et al. 2025), including a detrended set of elemental abundances, will be published as value-added catalogues in DR4.

<sup>m</sup>\*precision\_correction\_factors\* in <https://github.com/svenbuder/GALAH-DR4/tree/main/catalogs>.



**Figure 21.** Comparison of stars with available measurements in GALAH DR4 and APOGEE DR17 for [C/Fe] and [N/Fe].



**Figure 22.** Comparison of stars with available measurements in GALAH DR3 (left), GALAH DR4 (middle) as well as APOGEE DR17 (right) for [Mg/Fe] (top row) and [Ni/Fe] (bottom row).

### 6.2.3. Uncertainties in light of GALAH DR3 and APOGEE DR17

To get a better idea of the actual improvement of accuracy and precision, we have performed more elaborate comparisons than those in Fig. 16 and only showcase a few in this manuscript with reference to the online repository. We have found the comparison of GALAH DR4 with both GALAH DR3 and APOGEE DR17 highly informative.

Because GALAH DR3 did not include N measurements and only a limited amount of C measurements, our first comparison concerns the abundances of C and N between GALAH DR4 and APOGEE DR17 in Fig. 21. We attach the comparisons for the other overlapping elements O, Na, Al, Si, K, Ca, Ti, V, Cr, Mn, Co, and Ce in Figs. AC3 and AC4. While we see a generally good agreement of the shapes, we notice biases of  $-0.03$  dex and  $0.10$  dex for C and N, respectively. These can be, however, explained by the lower precision of GALAH and might, in part, be driven by the slightly different trends of C and N towards lower metallicities. In particular, [C/Fe] decreases to sub-Solar level in APOGEE DR17 for metal-poor stars, whereas it is Solar or even enhanced in GALAH DR4. Enhanced levels would be expected for metal-poor disk stars, whereas sub-Solar levels are expected for accreted stars (Amarsi, Nissen, & Skúladóttir 2019b), warranting a future population analysis to test the accuracy of either survey.

In addition to these novel abundances, we also showcase two previously measured abundances, namely [Mg/Fe] and [Ni/Fe] in Fig. 22. The  $\alpha$ -process element Mg has significant value for Galactic studies because it is predominantly produced by core-collapse supernovae (Kobayashi, Karakas, & Lugaro 2020). In GALAH DR3, only the Mg I 5711 Å line was used, whereas we now use a combination of several lines. This has led to a significant improvement in precision, as can be appreciated from the comparison of Fig. 22a and b. Even more positive, we see an improved agreement of the [Fe/H] vs. [Mg/Fe] measurements between GALAH DR4 (Fig. 22b) and APOGEE DR17 (Fig. 22c), with no abundance bias. One of the elements with the most significant precision improvement is Ni. For this element, our move to fitting the full wavelength range has increased the number of lines from two very reliable lines to several dozen lines. Albeit less reliable in their line data and possibly blended, the sheer increase in flux information used has improved the precision almost to the level of APOGEE DR17 – with no bias and a standard deviation of only 0.05 dex between APOGEE DR17 and GALAH DR4 (see Fig. 22c–f).

In addition to these instructive comparisons, we also return to the precision of Solar twins from Fig. 19. Here we specifically highlight the significant improvement of precision from GALAH DR3 to GALAH DR4 with respect to the linear estimate from

**Table 5.** List of major quality flag `flag_sp` listing the bit, description and how often the flag was raised for the *allstar* and *allspec* routines. Notes: Multiple bits can be raised for each of the 1 085 520 spectra of 917 588 stars.

| Raised Bit | Flag  | Description              | <i>allspec</i> | <i>allstar</i> |
|------------|-------|--------------------------|----------------|----------------|
|            | 0     | No flag                  | 700 125        | 663 075        |
| 0          | 1     | Emission                 | 9 568          | 7 646          |
| 1          | 2     | CCD missing              | 70 078         | 44 344         |
| 2          | 4     | Spectr. Binary 1         | 0              | 25 538         |
| 3          | 8     | Spectr. Binary 2         | 34 833         | 32 566         |
| 4          | 16    | $\chi^2 > 3\sigma$       | 66 859         | 20 544         |
| 5          | 32    | $v \sin i$ warning       | 138 317        | 95 990         |
| 6          | 64    | $v_{\text{mic}}$ warning | 99 692         | 78 686         |
| 7          | 128   | Triple Binary warning    | 0              | 0              |
| 8          | 256   | $T_{\text{eff}}$ warning | 0              | 0              |
| 9          | 512   | $\log g$ warning         | 19 863         | 10 900         |
| 10         | 1024  | [Fe/H] warning           | 0              | 0              |
| 11         | 2048  | S/N low                  | 123 736        | 71 154         |
| 12         | 4096  | Not converged            | 32 986         | 0              |
| 13         | 8192  | Model extrapolated       | 69 613         | 5 953          |
| 14         | 16384 | No Results               | 7 400          | 10 899         |

Bedell et al. (2018) for C (from 0.09 to 0.045), Si (from 0.04 to 0.023), Ca (0.07 to 0.049), Ti (0.05 to 0.031), V (0.13 to 0.050), Cr (0.06 to 0.024), Ni (0.07 to 0.033), and Y (0.12 to 0.080). This improvement of sometimes a factor of 2 is remarkable and most of our comparisons indicate that these values are representative of a precision improvement beyond the Solar twins, as shown for example for Ni in Fig. 22.

### 6.3. Stellar parameter flags `flag_sp`

We have implemented a series of post-processing routines to assess the quality of the stellar parameter determinations. These routines check for a variety of potential issues with the spectra and stellar label fitting, with each flag corresponding to a specific quality check. If any of these checks are not passed, the respective bit in the quality flag `flag_sp` is raised. The description of the implemented bits/flags for `flag_sp` and how often they were raised is listed in Table 5 and distributions in the Kiel diagram ( $T_{\text{eff}}$  and  $\log g$ ) are shown for each raised bit in Fig. AC6 for the *allstar* catalogue. For examples of stars with raised flags, we refer back to the emission line star (`flag_sp` = 1) of Fig. 15 (Section 6.1.4) and the clearly double-lined binary of Fig. 11.

Because quality cuts should be applied based on the specific science case at hand, we do not make a strict recommendation for which upper limit of `flag_sp` should be applied. We note, however, that we have tried to implement flags that increase in concert. The first 8 bit masks (with values up to  $2^9 - 1 = 511$ ) are therefore less problematic than those of 9 or higher (`flag_sp`  $\leq 512$ ).

While not intended to identify binaries, we believe that both the  $v \sin i$  and  $v_{\text{mic}}$  flags are informative for binaries below  $T_{\text{eff}} < 6\,000$  K (see their elevated position in Fig. AC6f and g). We have trained the stars of this region with a lower maximum  $v \sin i$  range that would be reached for a spectrum that is broadened due to binarity. This region certainly overlaps with the one of identified single-lined and double-lined binaries with `flag_sp` = 4 and 8, respectively (see Fig. AC6c and d). For the latter, we notice that especially cool giants are picked up by the automatic algorithm

as well. This might be either due to strong extinction biasing our analysis or due to lines in the spectrum not being modelled properly and thus showing up as residual signal. While these stars are possibly flagged false-positively, we also find a remarkable amount of true binaries ( $> 41\%$  in orange area of Fig. 23b), for which the *Gaia* DR3 radial velocity is likely the systemic radial velocity, as it is close to the mean radial velocity of both components identified in GALAH DR4. In Fig. 23, we visualise how one could use the radial velocities from GALAH DR4 and *Gaia* DR3 to further assess the reliability of this flag. To check if a particular bitmask flag (e.g.  $2^3 = 8$ ) is raised, one can perform the check in PYTHON via

```
flag_8_raised = (dr4['flag_sp'] & 8) != 0
```

### 6.4. Elemental abundance flags `flag_X_fe`

The quality of elemental abundance measurements is also captured through flags. When an element is reliably detected in the spectrum, no flag is raised. However, if the abundance of an element is estimated as an upper limit, often due to weak spectral lines or low SNR, an upper limit flag is triggered. If no measurement of the element is possible, a flag is raised to indicate that the relevant spectral features were too weak or the SNR too low to allow for an estimate. The list of bits and flags for elemental abundances, `flag_X_fe`, is shown in Table 6.

By default, we recommend to only use significant detections (`flag_x_fe` = 0) for an element. Because of a bug in the flagging of the [Fe/H] detection (see discussion in Section 8.8), we do not recommend to consider `flag_fe_h` for quality cuts.

### 6.5. Abundance detection or upper limit

To assess whether the abundance estimates are a true detection or an upper limit for each element X, we compare a synthetic spectrum with the best-fitting parameters to a synthetic spectrum with the same parameters, except for element X, for which we use the lower limit abundance of the neural network. The residuals in units of  $\sigma$  between the best-fitting spectrum and the spectrum with the lowest possible [X/Fe] or lowered [Fe/H] then allow us to identify a detection (with maximum differences beyond  $3\sigma$ ) or upper limits, for which we raise the flag `flag_x_fe` by 1. Our initial test of overall detectability (Section 4.3.1) allowed us to raise the flag `flag_x_fe` by 2 for elements for which not even an upper limit was expected.

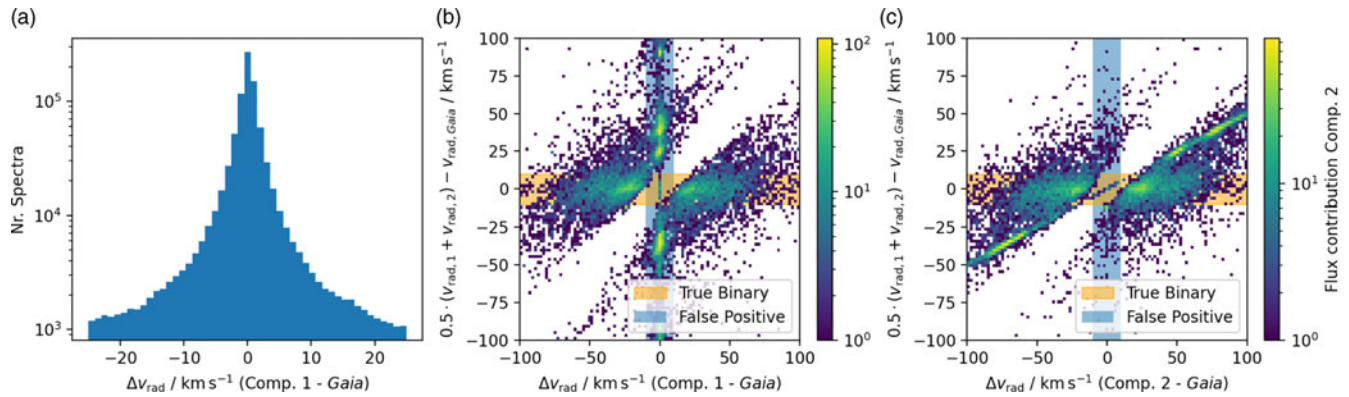
We further raise a flag for *allspec* abundances, if the element was fit above (3) or below (4) the neural network training set range. For CNO, we have identified specific regions, in particular dwarfs, for which could not verify abundances and therefore caution their use (flag 5). We have further tried to identify abundances, for which the optimisation may have failed and flagged these with flag 6 (see Section 8.7).

## 7. Data release products

GALAH DR4 encompasses a diverse range of data products. We describe the most important main catalogues in Section 7.1 and value-added catalogues in Section 7.2. We further explain the data products for each spectrum and star, that is, the reduced spectra (Section 7.3.1), *allspec* products (Section 7.3.2), and *allstar* products (Section 7.3.3).

The data products are provided directly on the AAO DataCentral website at [https://cloud.datacentral.org.au/teamdata/GALAH/public/GALAH\\_DR4/](https://cloud.datacentral.org.au/teamdata/GALAH/public/GALAH_DR4/). We further provide multiple





**Figure 23.** Comparison of radial velocity estimates of GALAH DR4 and *Gaia* DR3. Panel (a) shows the difference of GALAH's primary component radial velocity with the mean *Gaia* DR3. Panels (b) and (c) show stars for which two components were detected in GALAH DR4 and shows the difference between each component and *Gaia* DR3 against the difference of mean (roughly systemic) radial velocities. The panels also include regions where actual binaries and false positive detections are expected.

ways to interact with the data release products, which are described in Section 7.4.

### 7.1. Main data release catalogues

1. `galah_dr4_allspec_240705.fits`: analysis for each spectrum (including radial velocity estimation for each spectrum) based on a single spectrum.
2. `galah_dr4_allstar_240705.fits`: analysis for each star based on co-added spectra of each star and using non-spectroscopic information to constrain log *g*.

We present the main catalogue table schema in Table A1 (see also Fig. 24), but refer the reader to the FITS headers of each catalogue for more detailed information.

One of our greatest achievements as part of this data release is the extraction of C and N abundances for giant stars from molecular absorption features. In Fig. 25, we show how stellar mass and [C/N] ratios are correlated in GALAH DR4, as is expected based on the pioneering work by Masseron & Gilmore (2015), Martig et al. (2016), and Ness et al. (2016). Our measurements demonstrate the potential of [C/N] abundances to better separate the core-helium burning from the red giant phase (around the blue areas of Fig. 25b and c) or at least better constrain stellar masses.

### 7.2. Value-Added Catalogues (VAC)

We provide several value-added catalogues, namely a crossmatch catalogue to all entries of the *Gaia* DR3 main source catalogue and the most important entries from the 2MASS and WISE catalogues, a catalogue of stellar dynamics properties, a catalogue of 3D NLTE measurements of Li, and a catalogue with ages inferred via isochrone interpolation in a Bayesian framework.

#### 7.2.1. VAC of crossmatches with *Gaia* DR3, 2MASS and WISE

The value-added catalogue of the crossmatch<sup>n</sup> with the *Gaia* DR3, 2MASS, and WISE catalogues as well as the distance catalogue of Bailer-Jones et al. (2021) was calculated by performing an OUTER JOIN ADQL-query in the *Gaia* archive.

The query first performed an INNER JOIN with the 2MASS near-infrared photometry catalogue<sup>o</sup> via its designation and linked this match to the *Gaia* DR3 catalogue via the best neighbour<sup>p</sup> and joined<sup>q</sup> catalogues of 2MASS to *Gaia* DR3 (Torra et al. 2021). When cross-matching between *Gaia* DR3 and 2MASS, less than 1% of stars were associated with multiple possible matches. To ensure the best match, the data were sorted from brightest to faintest *G*-band magnitude, and only the brightest match for each `subject_id` was retained.

The crossmatch to the WISE far-infrared photometry catalogue<sup>r</sup> (Cutri et al. 2014) was performed via the *Gaia* DR3's best neighbour catalogue<sup>s</sup> (Torra et al. 2021). The match to the distance catalogue<sup>t</sup> of Bailer-Jones et al. (2021) via the *Gaia* DR3 `source_id`.

The catalogue also includes uncertainties in the *Gaia* DR3 photometric magnitudes (*G*, *G*<sub>BP</sub>, *G*<sub>RP</sub>) that were recalculated following the recommendations from the *Gaia* Early Data Release 3 (EDR3) documentation (Riello et al. 2021). The total uncertainties were computed by combining the photon flux error with an additional systematic term.

We further corrected the *Gaia* DR3 parallaxes for systematic zero-point errors by applying the correction model provided by Lindegren et al. (2021a). This correction depends on several factors, including the *G*-band magnitude, effective wavenumber (*v*<sub>eff</sub>) used in astrometry, pseudocolour, latitude, and the astrometric solution type. The parallax zero-points and original parallaxes are reported as `plx_zpt_corr` and `parallax_raw`, respectively.

Beyond the crossmatch with the *Gaia* DR3 `gaia_source` catalogue, multiple other crossmatches can easily be performed via the `gaiadr3_source_id` column. We have for example crossmatched the sources from GALAH DR4 with those from *Gaia* DR3's variability catalogues (Rimoldini et al. 2023). We find 47 493 stars in GALAH DR4 that overlap with the `gaiadr3.vari_classifier_result` catalogue. In particular, we find 17 256 SOLAR\_LIKE variables, 14 477 stars in the δScuti, γ Doradus,

<sup>o</sup>`gaiadr1.tmass_original_valid`.

<sup>p</sup>`gaiadr3.tmass_psc_xsc_best_neighbour`.

<sup>q</sup>`gaiadr3.tmass_psc_xsc_join`.

<sup>r</sup>`gaiadr1.allwise_original_valid`.

<sup>s</sup>`gaiadr3.allwise_best_neighbour`.

<sup>t</sup>`external.gaiadr3_distance`.

<sup>n</sup>`galah_dr4_vac_wise_tmass_gaiadr3`.

**Table 6.** List of elemental abundance quality flags `flag_fe_h` for [Fe/H] or `flag_X_fe` for element X.

| Raised Bit | Flag | Description                  |
|------------|------|------------------------------|
|            | 0    | Detection                    |
| 0          | 1    | Upper limit                  |
| 1          | 2    | No measurement available     |
| 2          | 4    | No convergence               |
| 3          | 8    | Measurement above limit      |
| 4          | 16   | Measurement below limit      |
| 5          | 32   | Measurement issue of CNO     |
| 6          | 64   | Optimisation may have failed |

or SXPhoenicis category (DSCT/GDOR/SXPHE), 6 247 LPV (long-period variables), 4 074 ECL (eclipsing binaries), 3 355 RS (RS Canum Venaticorum variables), 1 096 YSO (young stellar objects), 401 RR (RR Lyrae types), and a large variety of other variables, including the white dwarf 2MASS J05005185-0930549 that was already found in GALAH data by Kawka et al. (2020).

### 7.2.2. VAC of stellar dynamics

The value-added catalogue for stellar dynamics<sup>u</sup> includes the kinematic and dynamical properties for stars in the GALAH DR4 survey. The catalogue is created with a publicly available script<sup>v</sup> as part of GALAH DR4. We define the position of the Sun in our Galactic reference frame as  $R_{GC} = 8.21$  kpc (McMillan 2017),  $\varphi_{GC} = 0$  rad, and  $z_{GC} = 25$  pc (Bland-Hawthorn & Gerhard 2016). We then combine the total velocity in  $V$  of the Sun at  $R_{GC}$  based on the proper motion measurement of  $6.379 \pm 0.024$  mas yr<sup>-1</sup> by (Reid & Brunthaler 2004), that is,  $V_{\odot} = 248.27$  km s<sup>-1</sup> with the circular velocity of  $V_{circ} = 233.10$  km s<sup>-1</sup> from McMillan (2017) to estimate a peculiar velocity of the Sun with respect to the local standard of rest of 15.17 km s<sup>-1</sup>. For the other two components, we use the estimate by Schönrich, Binney, & Dehnen (2010), leading to a peculiar velocity of the Sun of  $(U, V, W) = (11.1, 15.17, 7.25)$  km s<sup>-1</sup>.

Starting from the crossmatch of GALAH DR4 with the *Gaia* DR3 (see Section 7.2.1), we use the GALPY.ORB module by Bovy (2015) to estimate heliocentric Cartesian coordinates  $(X, Y, Z)$  and velocities  $(U, V, W)$  as well as Galactocentric cylindrical coordinates  $(R, \varphi, Z)$  and velocities  $(v_R, v_{\varphi}, v_Z)$ . We approximate the orbit actions  $J_R, J_{\varphi} = L_Z, J_Z$  and frequencies  $\omega_i$  with the GALPY.ACTIONANGLE.ACTIONANGLESTAECKEL function with a focal length of the confocal coordinate system  $\delta = 0.45$  in the Milky Way potential by McMillan (2017). We further use the Staekel approximation (Binney 2012) to calculate eccentricity, maximum orbit Galactocentric height, and apocentre/pericentre radii with GALPY's ECCZMAXRPERIRAP (Mackereth & Bovy 2018). Our assumption of a time-invariant, axisymmetric potential further allows us to extract the orbit energy via GALPY. ORBIT.E.

In particular the dedicated observing programs of GALAH towards low angular momentum stars (PI S. Buder) and globular clusters (PI M. McKenzie and PI M. Howell) have increased the number of spectroscopic observations for stars on halo-like orbits.

This is showcased by both the action-action diagram of angular momentum  $L_Z$  versus radial action  $\sqrt{J_R}$  (Fig. 26) and angular momentum  $L_Z$  versus orbit energy  $E$  (Fig. 27) and visualises the potential of GALAH DR4 observations to complement Galactic dynamics studies and enable Galactic chemodynamic studies.

### 7.2.3. VAC of 3D NLTE lithium abundances

In this value-added catalogue,<sup>w</sup> we use spectrum fitting to infer 3D non-local thermodynamic equilibrium (NLTE) lithium abundances. For each spectrum, the Li line is modelled with a 3D NLTE BREIDABLIK line profile (Wang et al. 2021). In cases where the Li line is blended with nearby lines such as Fe and CN, we model blending lines as Gaussian absorption profiles. From this model, we measure the equivalent width (EW) and errors in EW of the Li line using ULTRANEST (Buchner 2021), a Monte Carlo nested sampling algorithm. The Li abundance,  $A(\text{Li})$ , is then inferred from the measured EW using BREIDABLIK and the stellar parameters from GALAH DR4's allstar. See Wang et al. (2024a) for a detailed description of the methodology.

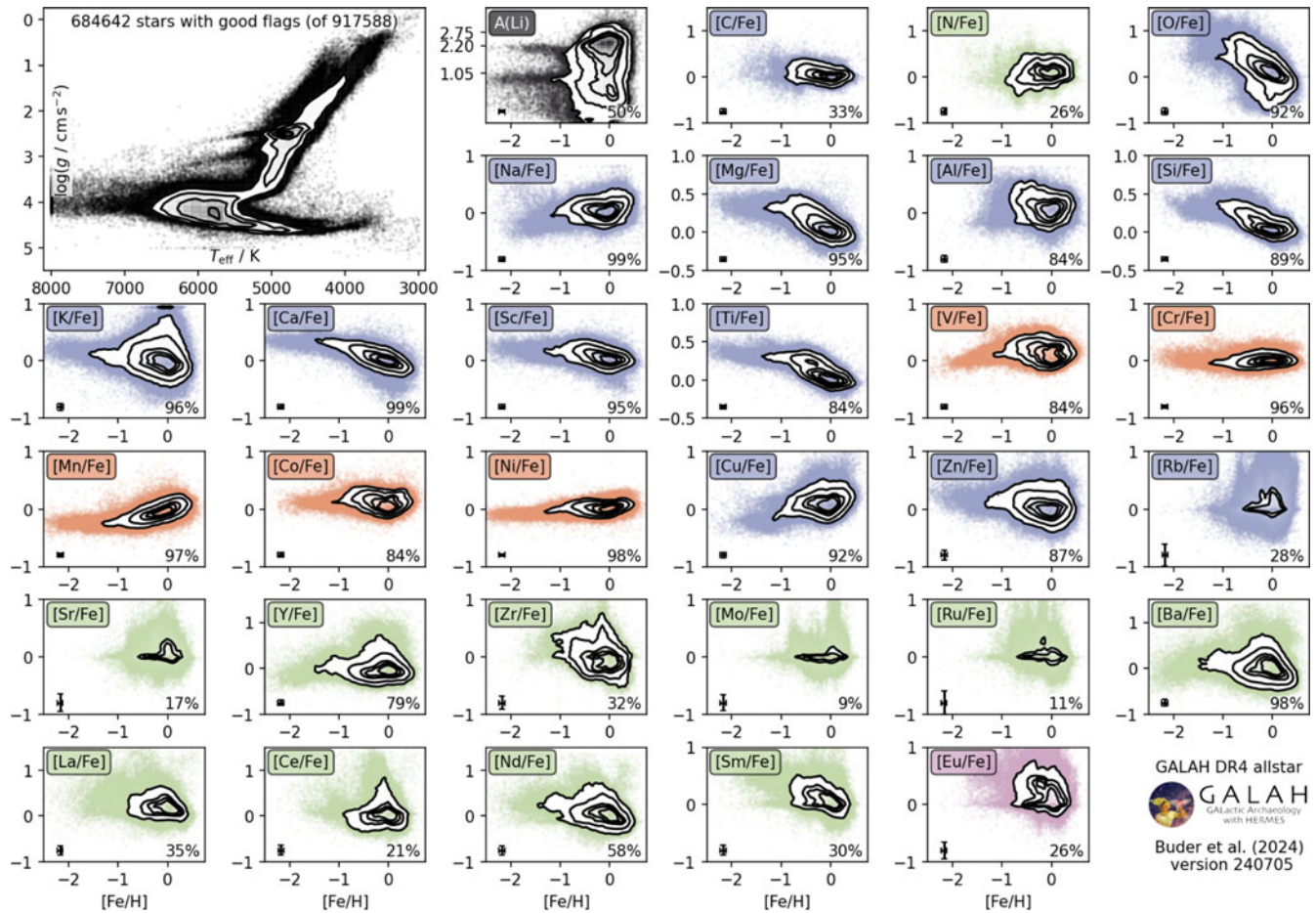
Wang et al. (2024a) measured a local line width for the Li region and fit the width of the Li line separately from other lines. For this work, we use the GALAH instrumental profile convolved with the rotational velocity as the width of our blending lines and set the width of the Li line based on this convolved kernel, better constraining the line widths. Whilst we still measure a local radial velocity due to a lack of ThXe arc lines for CCD3 (see Section 8.1), we apply the GALAH radial velocity for poorly constrained Li depleted stars where we cannot measure the local radial velocity. In addition, the sampled EW posterior is now modelled using a first order boundary corrected kernel density estimator from Lewis (2019), which has better convergence than histograms. Lastly, GALAH DR4 analyses stars down to 3 000 K, but the STAGGER model atmospheres only reach 4 000 K, therefore, we provide an additional column of 1D NLTE<sup>x</sup>  $A(\text{Li})$  inferred through our measured EWs using a new interpolator. Similar to the existing EW interpolators in BREIDABLIK (Wang et al. 2021), we train a feed-forward neural network on NLTE Li abundances synthesised using the 1D MARCS model atmospheres. We use a 2-layer architecture with the ReLU activation function and find best hyperparameters:  $i = 900$  neurons, and  $\alpha = 0.1$  L2 penalty. This model is included in the BREIDABLIK package. Using the updated methodology it takes  $\sim 2$  min per star in comparison to 0.5–2 h per star reported in Wang et al. (2024a), with the main speed up coming from fixing the Li line width.

EWs are measured for 892 223 stars (97% of GALAH DR4), with 3D NLTE  $A(\text{Li})$  detections reported for 417 825 stars (46%) and upper limits for 474 398 stars (52%). Fig. 28 shows the mean EW over  $T_{\text{eff}}$  and  $\log g$ . The Li-dip can be seen on the main-sequence turn-off at  $T_{\text{eff}} \approx 6 500$  K and  $\log g \approx 4.2$  and extends up the subgiant branch. There is a Li enhanced population of stars at  $\log g \approx 2.5$  in the red clump whilst the horizontal branch is depleted in Li. Although the secondary red clump appears to be depleted in Li, these stars have an overestimated  $\log g$  driven by incorrectly inferred masses (see Section 8.4), and should be primary red clump stars. The increase of mean Li EW up the giant branch is due to a large proportion of stars with  $\text{EW} \approx 100$  mÅ. Features of this figure will be studied in follow-up papers.

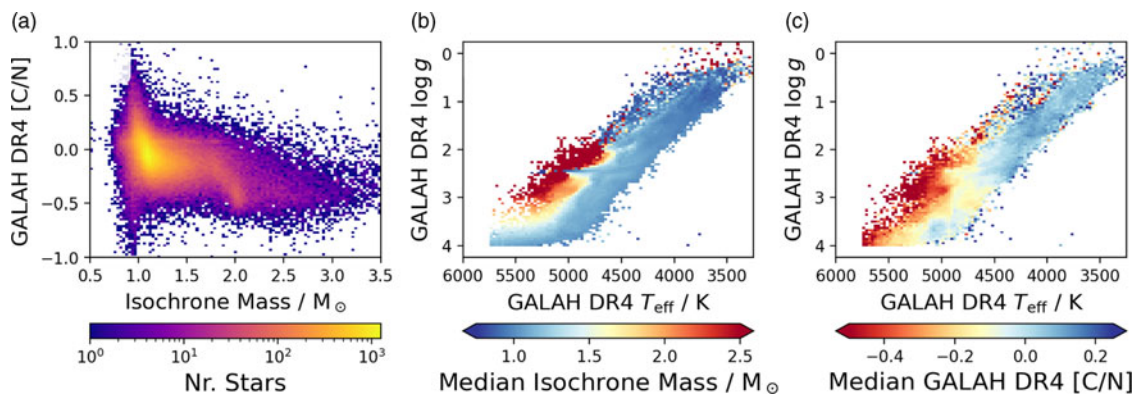
<sup>u</sup>galah\_dr4\_vac\_dynamics.

<sup>v</sup> Accessible in the GALAH DR4 repository here.

<sup>w</sup>galah\_dr4\_vac\_3dnlte\_a\_li.

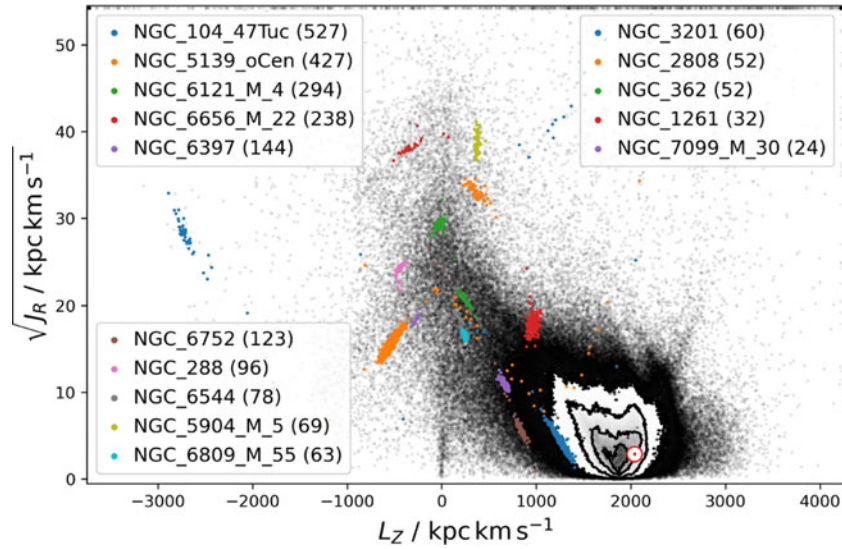


**Figure 24.** Overview of stellar parameters and elemental abundances for the allstar estimates of GALAH DR4. The top left panel shows the density distribution of stars in the Kiel diagram of  $T_{\text{eff}}$  and  $\log g$ . All other panels show the logarithmic elemental abundances (for elements indicated in the top left of the panel) as a function of the logarithmic iron abundances  $[\text{Fe}/\text{H}]$ . Elements are coloured by different nucleosynthetic channels (black for big bang nucleosynthesis, blue for core-collapse supernovae, red for supernovae Type Ia, green for asymptotic giant branch star contributions and pink for the rapid neutron capture process with contributions from merging neutron stars) following the colour schema from Kobayashi et al. (2020). Percentages indicate the fraction of detections of stars for each element.

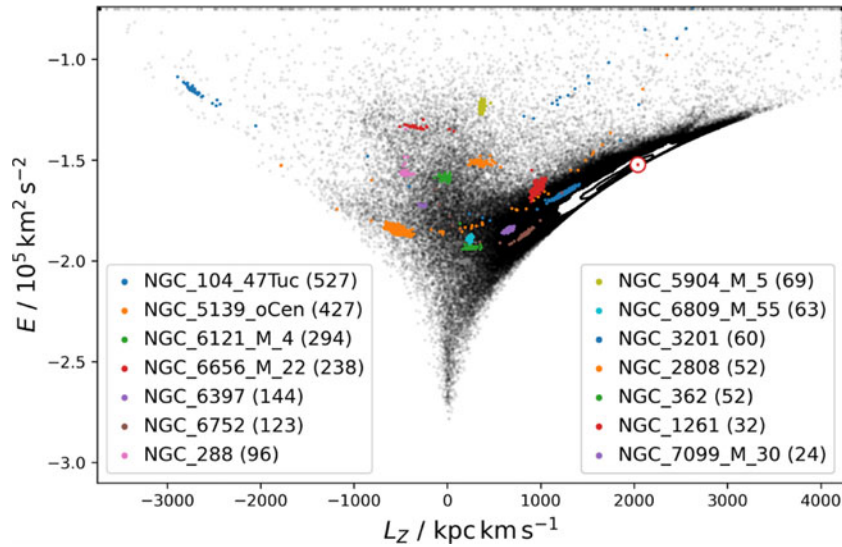


**Figure 25.** The ratio of  $[\text{C}/\text{N}]$  and isochrone masses in comparison panel (a), and as a function of  $T_{\text{eff}}$  and  $\log g$  in panels (b) and (c), respectively.





**Figure 26.** Distribution of the dynamical properties of angular momentum  $L_z$  and radial action  $J_R$  of stars in GALAH DR4 (black), with globular cluster members highlighted in colour. Cluster members were selected as those with more than 70 percent membership probability according to Vasiliev & Baumgardt (2021). The Sun is indicated with a red  $\odot$  symbol.



**Figure 27.** Distribution of the dynamical properties of angular momentum  $L_z$  and orbital energy  $E$  of stars in GALAH DR4 (black), with globular cluster members highlighted in colour. Cluster members were selected as those with more than 70 percent membership probability according to Vasiliev & Baumgardt (2021). The Sun is indicated with a red  $\odot$  symbol.

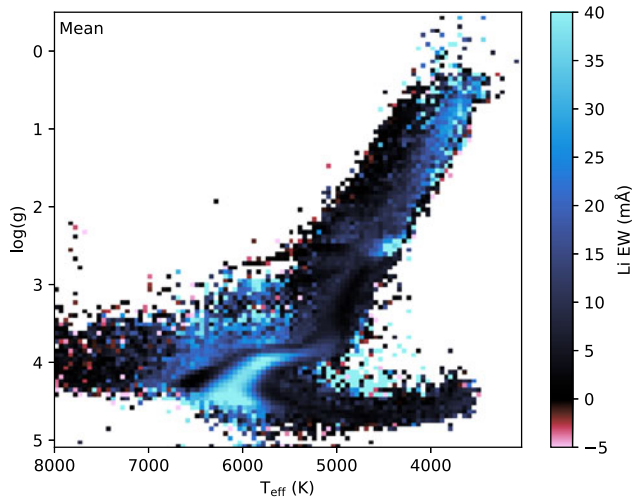
A quality flag (flag\_ALi) is raised by 1 for upper limits, 2 or more to indicate other quality issues, such as stellar parameters falling outside of the model atmosphere grid (see Wang et al. 2024a) for more details on the bitmask flag). We recommend  $\text{flag\_ALi} < 2$  when using the 3D NLTE A(Li), and  $\text{flag\_ALi} < 4$  when using Li EWs. A similar quality flag  $\text{flag\_ALi\_1D}$  is provided corresponding to the 1D NLTE A(Li) included in the VAC.

<sup>3</sup>Note that these 1D NLTE Li abundances are different from the 1D NLTE Li abundances published in *allstar*.

For convenience, we have included the most important columns of this catalogue in the *allstar* catalogue (see Table A1), as we recommend to use them instead of the less accurate 1D NLTE abundances estimated with an imperfect neural network interpolation, which we indicate with *nn\_li\**.

#### 7.2.4. Ages

A value-added-catalogue for stellar ages and masses from BSTEP (Sharma et al. 2018) is currently in preparation. In the meantime, users can rely on the on-the-fly age and mass estimates



**Figure 28.** Mean EW binned in  $T_{\text{eff}}$  and  $\log g$ . The Li-dip can be seen at  $T_{\text{eff}} \approx 6500$  K and  $\log g \approx 4.2$ . At  $\log g \approx 2.5$ , red clump stars have a higher mean Li EW whilst horizontal branch stars have a lower mean Li EW compared to surrounding stars. The mean Li EW increases going up the red giant branch.

already provided in the *allstar* and *allspec* catalogues from the pipeline.

### 7.3. Data products for each spectrum and star

We provide individual data products in an orderly fashion that allow users to create links to these products based solely on the `subject_id`. To download data products for individual stars we recommend creating a url string and using `WGET` or similar commands. For bulk downloads of the advanced data products of this section, we recommend contacting the GALAH collaboration or using the bulk download interfaces of AAO DataCentral.

#### 7.3.1. Reduced spectra

The reduced spectra of each night are provided in the observations directory<sup>y</sup> and sorted into directories with four spectra – one for each of the four CCDs. These spectra are produced by the reduction pipeline (see Section 2.2) and include several extensions as outlined in Table 2, with wavelength information stored in the fits headers with starting wavelength `CRVAL1` in Å and linear pixel scale `CDEL1` in Å/px, and the number of pixels `NAXIS1`. The reduced spectra are only provided per exposure and not in a co-added manner, since the co-adding was performed as part of the *allstar* module (see Section 7.3.3 for co-added spectra). We note that not all files might be available for a given exposure due to the rare failure of CCD readouts.

#### 7.3.2. Additional products of the *allspec* module

The *allspec* analysis product directory<sup>z</sup> provides the files that were produced by the *allspec* module. These include the on-the-fly assessment of the radial velocity fit `*rv.png` (similar to Fig. 9), the raw fitting results `*results.fits` and their covariance matrices `*covariances.npz` (similar to the entries used to produce Fig. AB2). We also provide a combined `*spectrum.fits`

file (concatenated over the four bands) that includes the wavelength, flux, and flux uncertainty of the velocity-corrected and re-normalised observed spectrum as well as the best-fitting model spectrum interpolated onto the same wavelength. Finally, we provide a `*comparison.pdf` (similar to Fig. AB1) which displays the fit results, comparison of observed and model spectrum, masked wavelength regions, and wavelengths of the most important element lines. If the module did not run to completion, for example because the SNR of the spectra was below the threshold of  $\text{SNR} = 10$  for any CCD to even attempt a fit, not all products are available for a spectrum.

#### 7.3.3. Additional products of the *allstar* module

The *allstar* analysis product directory<sup>aa</sup> also includes the radial velocity monitoring `*rv.png`, results files `*results.fits`, combined spectra `*spectrum.fits` and `*comparison.pdf` overview, similar to the ones described in Section 7.3.2. In addition, each directory also includes a `*subject_ids.txt` file that lists all individual spectra that were co-added to create the observed spectrum and its uncertainty in `*spectrum.fits`.

### 7.4. Interactive access via AAO DataCentral

In collaboration with the AAO Data Central, a number of interactive ways are provided to explore the data of this release. Data Central provides both Simple Spectral Access and Single Object Viewer services. In addition, we recommend to download files or easily crossmatch user catalogues with the TAP server <https://datacentral.org.au/vo/tap> in TOPCAT (Taylor 2005). <https://apps.datacentral.org.au/galah/spectra> also provides an interactive plotting application to show normalised or un-normalised spectra of different repeat observations. As these tools are under active development, we refer to the latest documentation on both the DataCentral and the main Survey website <https://www.galah-survey.org>.

## 8. Caveats and future improvements

In this section, we attempt a detailed discussion of caveats at different steps of our analysis, while also giving suggestions for future improvements – both for GALAH and other surveys. We first discuss caveats of the spectrum reduction (Section 8.1), before extensively discussing the spectrum synthesis (Section 8.2) and spectrum interpolation (Section 8.3). We discuss possible problems arising from the use of photometric information (Section 8.4), in particular for stars that could be binaries (Section 8.5). We elaborate on caveats regarding globular clusters in Section 8.6 and the fitting itself in Section 8.7. Finally, we point out caveats regarding the flags in Section 8.8 as well as a bug and its correction in the reported radial velocity of interstellar K in Section 8.9. We summarise the most important caveats in Section 8.10.

### 8.1. Spectrum reduction

Although a significant amount of work was spent on improving the spectrum reduction, several persistent issues remain, which are summarised below.

<sup>y</sup>observations/YYYYMMDD/spectra/com/subject\_id\*.fits.

<sup>z</sup>analysis\_products\_single/YYYYMMDD/subject\_id/.

<sup>aa</sup>analysis\_products\_allstar/YYYYMMDD/subject\_id/.

### 8.1.1. Wavelength solutions

For each CCD, the reduction pipeline estimates the most suitable wavelength solution, linking pixels with actual wavelengths based on the ThXe arc lines. In GALAH DR3 (Buder et al. 2021) we identified several issues for spectra where not enough ThXe lines could be used to constrain the wavelength solution. Improvements have been made for the new reduction version to improve the number of useful ThXe lines and restrict the flexibility of wavelength solutions to move them closer to previous results. This has helped us to decrease the number of problematic wavelength solutions towards the red end of CCD3 which includes the used absorption features of Li and Eu. We have decreased bad wavelength solutions for this CCD from initially 7.9% of the spectra to roughly 1% bad solutions, that is, similar to the other CCDs.

### 8.1.2. Holistic spectrum extraction

Although much work has been spent on improving telluric and sky lines in the reduction step, most reduction steps are currently run sequentially rather than in parallel. Using the information of stellar spectra when modelling the wavelength solution would certainly help to overcome the limited information in ThXe calibration spectra in the absence of laser combs (Kos et al. 2018). Multiple steps in this direction have been taken (Saydjari et al. 2023) and should be rolled out in future spectrum analysis. This would especially help to mitigate imperfect telluric and sky line removal while simultaneously improving the wavelength solution – among many other effects.

## 8.2. Imperfect spectrum synthesis

### 8.2.1. Spectrum synthesis

The GALAH survey's success relies heavily on the ability to accurately model stellar spectra to infer accurate stellar properties. The survey has seen significant improvements in moving from the approximation of 1D LTE towards 1D NLTE (Amarsi et al. 2020b). This includes the use of 1D NLTE synthesis for atomic lines using the 3D NLTE code BALDER (Amarsi et al. 2018b), a custom version of Multi3D (Botnen & Carlsson 1999; Leenaarts & Carlsson 2009). The code employs model atoms for H (Amarsi et al. 2018b), Li (Lind, Asplund, & Barklem 2009a; Wang et al. 2021), C (Amarsi et al. 2019a), N (Amarsi et al. 2020a), O (Amarsi et al. 2018a), Na (Lind et al. 2011), Mg (Osorio et al. 2015), Al (Nordlander & Lind 2017), Si (Amarsi & Asplund 2017), K (Reggiani et al. 2019), Ca (Osorio et al. 2019), Mn (Bergemann et al. 2019), Fe (Amarsi et al. 2018b; Amarsi et al. 2022), and Ba (Gallagher et al. 2020) over the MARCS model atmosphere grid. The work by Wang et al. (2024a) also enables us to present measurements of Li in 3D NLTE as part of this release.

All of these advances contrast with the lack of a proper way of modelling molecular features appropriately. This could explain the significant mismatch of oxygen abundances between the optical and infrared (compare e.g. Selsby, Feltzing, & Oey 2014; Abdurro'uf et al. 2022). It can, however, also lead to mismatches in the GALAH wavelength range, where atomic features, such as C I, can be modelled in 1D NLTE, whereas much stronger molecular features of C<sub>2</sub> and CN have to be modelled in 1D LTE and line lists of molecules, such as TiO, might be incomplete (Hoeijmakers et al. 2015; McKemmish et al. 2019).

For our synthesis, we have employed version 580 of the IDL-based code Spectroscopy Made Easy (Valenti & Piskunov 1996;

Piskunov & Valenti 2017). As part of the continuing improvement of this code, several bugs have been identified and fixed. We also note that a Python-based version of SME, `PySME` (Wehrhahn et al. 2023), has become available. In addition, the spectrum synthesis code KORG (Wheeler et al. 2023; Wheeler, Casey, & Abruzzo 2024) has been published in Julia with a Python interface. It offers a faster alternative to SME once 1D NLTE synthesis is implemented, which is essential for applying to many NLTE-sensitive lines, such as O and K, in the GALAH wavelength range. KORG already internally adjusts the metallicity that is used to interpolate atmospheres based on the overall chemical abundances, whereas this would need to be adjusted in SME by hand, since atmospheres are interpolated with the `SME.FEH` entry that is independent of the chemical composition `SME.ABUND`. Because we have not performed said adjustment, we note that the spectrum synthesis for chemical compositions far from scaled-Solar may have used an mismatched atmosphere in the synthesis in SME for GALAH DR4.

### 8.2.2. Mismatch of atmosphere and spectrum chemistry

For several of our synthetic spectra, the chosen chemical composition deviates significantly from the scaled-Solar pattern of the MARCS model atmospheres, particularly for  $\alpha$ -process elements such as O and Mg, as well as C and N. These elements can substantially affect opacity and energy transport, and therefore, their abundances must be adjusted to match observed spectra more accurately. For instance,  $\alpha$ -enhancements in stars with non-Solar abundance patterns can shift line strengths and depths significantly (Asplund 2005; VandenBerg et al. 2012). Likewise, variations in C and N abundances, particularly in cooler stars, can impact molecular equilibrium, altering CO and CN molecular line strengths significantly (Tsuji 1976; Smith et al. 2013). Dedicated MARCS atmospheres with modified  $\alpha$  and C abundances (Mészáros et al. 2012; Jönsson et al. 2020), such as those used in modelling APOGEE spectra (Abdurro'uf et al. 2022), or more flexible interpolation schemes by Westendorp Plaza, Asensio Ramos, & Allende Prieto (2023), address this mismatch. However, the NLTE grids would also need to be expanded to cover all grid points of the extended MARCS models to ensure consistency.

## 8.3. Spectrum interpolation with neural networks

### 8.3.1. Training set selection

Before the neural networks are computed, it should actually be tested what the abundance zero-points are. In the case of several elements like Na and Al they are significant, on the order of 0.2 dex. When this occurs, stars with actual high abundances of 0.7 – 0.8 dex, for example in old stars and especially in globular clusters (see e.g. Carretta et al. 2009b), are not sufficiently covered.

One of the primary challenges in creating an optimal training set for spectrum interpolation lies in the choice of parameter sampling. A common caveat is the use of randomised, uncorrelated parameter sampling, which can lead to unrealistic combinations of elemental abundances. Elements that share a similar nucleosynthesis channel often exhibit correlated behaviour, for instance, stars with high abundances of Mg are typically also enhanced in Si, Ca, and Ti, while Na and Al tend to be elevated together. Similarly, neutron-capture elements like Y and Ba often follow similar trends (e.g. Ting et al. 2012; Kobayashi et al. 2020; Buder et al. 2021). To better capture this behaviour in the training set, the use of scaled linear functions or normalising flows could be advantageous. These approaches would help minimise the occurrence of



unlikely parameter combinations and yield a more representative sample.

For stars of the thin disk population, one could for example consider sampling from a noisy age-[X/Fe] relation to model chemical evolution (see Fig. 19, Nissen 2015; Spina et al. 2016; Bedell et al. 2018). This approach becomes more complicated when considering the thick disk, halo, and peculiar stars, where distinct nucleosynthesis histories introduce greater variability in elemental abundance trends.

### 8.3.2. Masking of spectra

Because the correlation between spectral features, stellar parameters, and abundances is often complex, degeneracies can arise when two stellar properties influence similar pixels of a spectrum (e.g. C and N for CN, or  $T_{\text{eff}}$  and [Fe/H] for cool dwarfs) or two stellar properties tend to act in lockstep in actual stars (e.g. Mg, Si, and Ti as  $\alpha$ -process elements). In GALAH DR2 (Buder et al. 2018), we attempted to overcome these issues by specifically masking the coefficients of spectrum interpolation, that is, effectively restricting the interpolation to only change smaller parts of the spectrum for a given stellar property.

In GALAH DR4, we have relaxed this restriction again, since we have trained on random abundance combinations in the hope of being able to break correlation degeneracies. We note, however, that too little information in spectra can again cause by-chance correlations (e.g. if neutron-capture lines are always very weak and the training set is not sufficiently large). We believe that this is the cause of the decrease in precision for Eu measurements from GALAH DR3 to GALAH DR4. The Eu abundance was mainly measured only from the weak Eu 6 645 Å line in DR3, whereas the neural networks of DR4 are not restricted to this region.

### 8.3.3. Flexibility of neural networks in general

The decision to use a large set of neural networks, each covering a restricted region in the  $T_{\text{eff}}$ , log  $g$ , and [Fe/H] space, was motivated by the goal of reducing the complexity required of any single model. By dividing the parameter space into smaller subsets, each neural network can be specialised and therefore less flexible, which allows for more precise modelling within its specific region. This approach avoids the trade-off faced by a single, monolithic neural network, which would either lack sufficient flexibility across the entire parameter space or be computationally more expensive to train and evaluate. For this data release, we have fixed the chosen network architecture of a 2-layer perceptron with 300 neurons and specific learning rate. While we have tested other activation functions than leaky rectified linear units, namely sigmoid, tanh and exponential linear unit functions, we found the lowest root mean square errors for our chosen activation function. Given the found issues with model fluxes above 1, we also would recommend to test a sigmoid as last activation layer of the neural network to ensure that the neural network always predicts fluxes between 0 and 1, as is expected from modelled stellar spectra. We have further tested a larger number of neurons, but found the root mean square errors to stabilise around 300 neurons for our test cases. It has to be acknowledged that due to the limit of human power to properly train and test the neural networks, we have not been able to properly test all neural networks and explore more flexible architectures. For this data release, we have decided not to rerun these steps, but make the current results available to the community. In the future, the restriction to one or only a few network models is recommended. The latter could cover regions of cool

dwarfs, main-sequence turn-off stars, hot stars, and giant stars with individual models – and possibly explore the split in metal-poor and solar-like regimes. This would also decrease overhead, in particular for training and loading different models as well as possible nodding effects between different models.

### 8.3.4. Flexibility of neural networks for extreme abundances

While this approach has proved to be powerful for all elements across their abundance ranges, we have noticed sinusoidal shapes for weak Li lines (see also Wang et al. 2021). This is likely caused by the large dynamical range of  $0 < A(\text{Li}) < 4$  that has to be covered by the neural network. For Li, the more sophisticated approach is to fit Gaussian lines to multiple components in the wavelength range around 6 708 Å, measure EW(Li), which are then used to infer 3D-NLTE based  $A(\text{Li})$  abundances. This inference is preferable to our 1D-NLTE based neural network estimates, as it is independent of the network flexibility and superior to our less accurate spectrum synthesis in 1D.

While several studies have identified that the abundances of stars in the Galactic disk are often very similar (e.g. Ness et al. 2019), the Galactic halo offers a more diverse picture. An example is 2MASS J2253100-6658174 (140707003601047), a turn-off star with extremely high s-process abundances and actually visible lines of La and Nd in addition to the usually visible Y and Ba. In this case, the fits to the La and Nd lines are significantly weaker than the observations. GALAH DR3 actually produced reasonable fits to this star with high abundances in [Y/Fe]=1.2, [Ba/Fe]=1.5, [La/Fe]=1.5, [Ce/Fe]=1.1, [Nd/Fe]=1.9, and [Sm/Fe]=1.2. A neural network that is not trained on such high abundances is likely to improperly extrapolate stellar spectra.

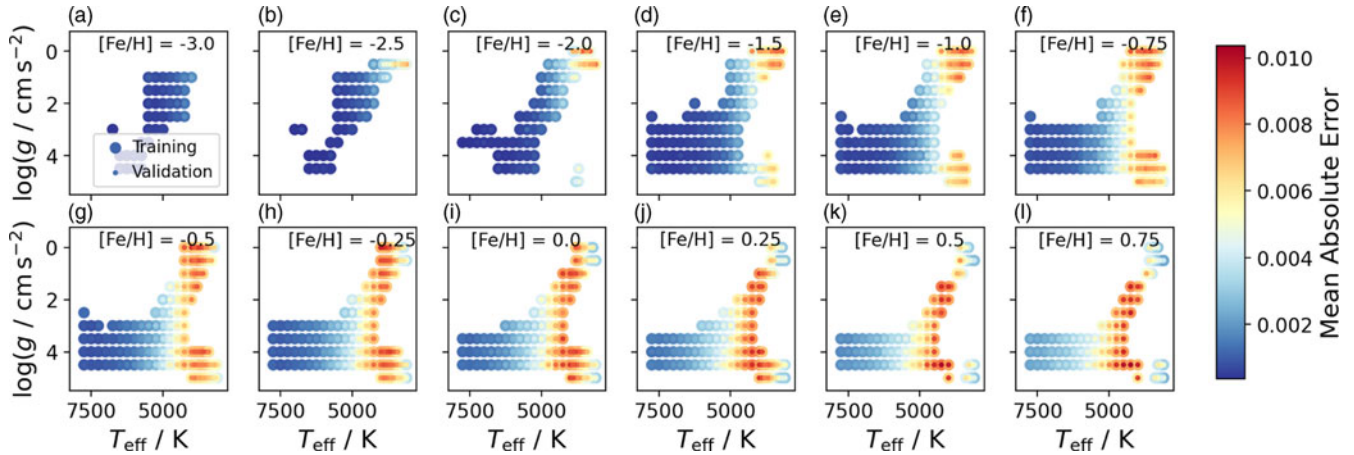
While we have tried to extract abundances of chemically peculiar stars, such as carbon-enhanced metal-poor stars, the significant effect of their molecular features onto the whole stellar spectrum is not to be underestimated and can in-itself pose a problem to the flexibility of neural networks.

### 8.3.5. Over- and underdensities at neural network edges

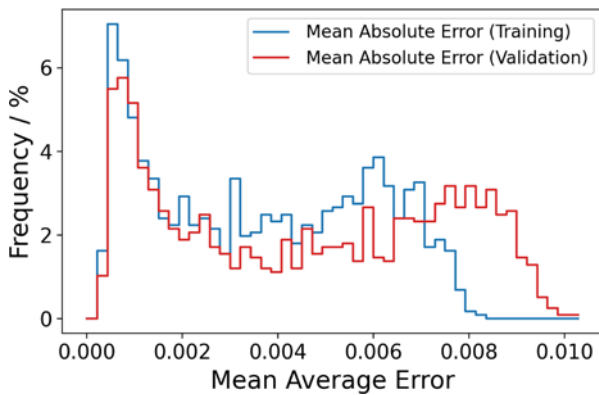
While the use of one neural network to interpolate the high-dimensional spectrum space is preferable, in practice, different science cases may drive the decision to use several networks. If the science case is to reach maximum precision, one neural network that is trained on the typical spectrum could be used at the expense of properly modelling peculiar spectra. If the science case is to reach maximum accuracy, only the regions with reliable line data and spectrum synthesis might be preferable. If the science case is to find peculiar stars, a larger coverage is needed to avoid the inaccurate extrapolation of stars with extreme abundances. In practice, large collaborations likely unite all of these goals, and a compromise has to be struck among the different approaches. For future analyses, a possible solution could therefore be to follow a two-step approach of first running one generic neural network for all spectra and then using optimised neural networks – or full spectrum synthesis – on smaller target samples of specific science cases.

### 8.3.6. Quantitative performance of neural networks

Throughout the training of our neural networks, we optimised model parameters using a mean absolute error (MAE) loss function across the spectrum pixels. The MAE remains consistently below 0.01 for all neural networks, indicating high accuracy in



**Figure 29.** Neural network performance shown as a function of  $T_{\text{eff}}$  vs.  $\log g$  with each panel showing a different range of  $[\text{Fe}/\text{H}]$ . Colours indicate the mean absolute errors of the training (large circles) and validation (small circles) for the neural networks.



**Figure 30.** Histogram of the mean absolute errors for the neural networks. These were used as loss function during the training (blue) and validation (red) on seen and unseen spectra, respectively.

model predictions, particularly for turn-off and most metal-poor stars where errors are typically below 0.001 (see Figs. 29 and 30).

Despite these low average error rates, the performance of neural networks can vary significantly across different spectral regions. Errors are minimal in continuum areas but tend to increase around strong or strongly changing absorption features, such as those of lithium, which are discussed in Section 8.3.4. The neural network architecture does not track uncertainty for each weight and bias, limiting our ability to generate perturbed models for assessing the impact of interpolation uncertainties on derived parameters and abundances. Additionally, retraining networks with varied initial conditions to evaluate prediction stability is computationally intensive.

To gauge the practical impact of these uncertainties, we compared the MAE against the noise levels in the GALAH spectra. Errors significantly lower than the noise levels for stars above  $T_{\text{eff}} > 5000$  K suggest that the interpolation inaccuracies minimally impact our analysis. However, for cooler stars with MAE around 0.01, interpolation inaccuracies could potentially influence precise chemical abundance studies more substantially.

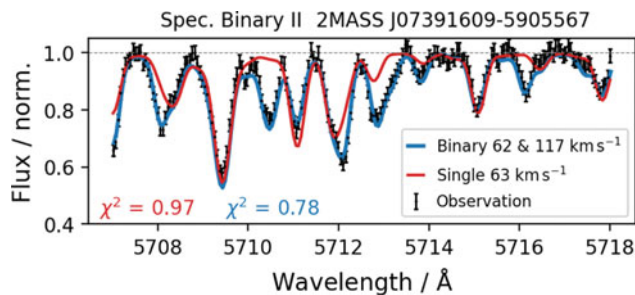
Despite the high degree of accuracy achieved, the limitations outlined necessitate careful interpretation of derived parameters, especially in regions with significant absorption features (see Figs. 29 and 30).

#### 8.4. Mismatch of spectroscopic and photometric information

##### 8.4.1. Incorrect masses driving incorrect stellar parameters

We estimate masses and ages through isochrone matching, where stellar parameters (validated against photometric estimates) are known for not being fully consistent with spectroscopic values. We believe this leads to significant mismatches especially for stars close to the red clump. In this region, a small change in spectroscopic and photometric information can imply a significant change in inferred mass (e.g. from primary to secondary red clump, with the latter being 2 or more solar masses and thus significantly more than the usual  $\sim 1$  solar mass). This issue has only become noticeable after the production runs and we have therefore decided not to rerun this particular region of the parameter space for this data release. We have extensively tested the possible reasons and identified the mismatch of isochrones and actual stellar spectroscopic parameters as the cause. We have not been able to fully resolve this issue by either including a prior based on the initial mass function to weigh against massive stars (see e.g. Sharma et al. 2018) and move from likelihood-weighted to posterior mass estimates. Similarly, we have not been able to resolve these effects by artificially upscaling the spectroscopic uncertainties when calculating the likelihood-weighted masses. More work needs to be done to mitigate the current inconsistencies of theoretical isochrones and spectroscopic estimates.

Another solution for this particular region of the parameter space could be the use of chemical stellar evolution through the correlation of core and thus total mass with the ratio of  $[\text{C}/\text{N}]$  after the first dredge-up (Masseron & Gilmore 2015; Martig et al. 2016), given that GALAH spectra also contain information on both elements. This could thus be used to better constrain high masses and counteract the information from isochrone-inferred masses. For this data release, the  $[\text{C}/\text{N}]$  information could at least serve as an indicator of how trustworthy high masses for giant stars are.



**Figure 31.** Example spectrum for a double-lined spectroscopic binary star (SB2) that is better fitted with our binary fitting algorithm.

#### 8.4.2. To use or not to use non-spectroscopic information?

The implementation of non-spectroscopic information, as done in our *allstar* module, has the advantage of overcoming spectroscopic degeneracies (as proven for the limited information on  $\log g$  in the HERMES wavelength range) as well as improving accuracy and precision also for the lowest quality spectra (because  $\log g$  is no longer solely dependent on the spectrum information).

However, this approach is only useful if the non-spectroscopic information is not biased (as it would be for astrometric and photometric information in the case of unresolved binarity). While the astrometric information for almost all GALAH targets is exquisite, this may not be the case for other surveys. The significant improvement from GALAH DR3 to GALAH DR4 has most definitely benefited from the improved astrometric information of *Gaia* EDR3 (Gaia Collaboration et al. 2021a; Lindegren et al. 2021b) and *Gaia* DR3 (Gaia Collaboration et al. 2023) with respect to *Gaia* DR2 (Gaia Collaboration et al. 2018; Lindegren et al. 2018). Further improvement could be expected when also taking *Gaia*'s photometric information into account, in addition to our use of 2MASS photometry.

### 8.5. Binaries

Although not part of this release, we have created an analysis module for spectroscopic binaries. The module will be presented in a separate work (Lach et al., in preparation) with a catalogue becoming a value-added catalogue of this release. The module is motivated by the extensive study of GALAH binary star spectra by Traven et al. (2020) and our ability to model the full spectrum via neural networks. We show a first analysis result of the module in Fig. 31, where the module was applied to a spectroscopic binary type 2 and resulted in a significantly better fit than the single star analysis.

### 8.6. Globular clusters

Globular clusters are well known for their light element anti-correlations (i.e. the Na-O or Mg-Al anti-correlations), though the underlying cause remains a subject of debate (for recent reviews see Bastian & Lardo 2018; Gratton et al. 2019; Milone & Marino 2022). It is widely accepted that one population is enhanced in elements including He, N, Na and Al, and depleted in O and C. Previous GALAH data releases have encountered issues in removing trends between abundances and stellar parameters (as discussed in Section 6.2.1), and DR4 represents a marked decrease of scatter within the Kiel diagrams of the globular clusters (see Appendix C). Despite these improvements, light element

abundance anti-correlations are still not well reproduced for DR4. We attribute this to two key factors: abundance zero points (Section 6.2.1) and the masking of spectra (Section 8.3.2).

Table C1 illustrates that both Na and Al have some of the largest zero-point shifts ( $-0.171$  and  $-0.185$ ), meaning that for some clusters the full extent of the anti-correlations is not realised (particularly for the light element enhanced populations). Secondly, when inspecting the optimal synthesis for particular lines (e.g. the Na lines at 5 682.6 and 5 688.2 Å, or the O triplet around 7 770 Å), the fits are poorly constrained, leading to a more significant scatter in these critical elements than what has previously been reported in the literature. We expect this is related to the relaxed restrictions on the neural network. Based on the abundances in their current form, we do not recommend using these light element abundances to distinguish between the multiple populations in globular clusters.

However, the 3D NLTE Li abundances discussed in Section 7.2.3 have effectively mitigated the above issues by adopting the GALAH stellar parameters and focusing exclusively on fitting the Li line. When analysing this Li data for the globular clusters, we can effectively reproduce the Li depletion patterns reported by Lind et al. (2009b). The large sample of clusters allows for a homogeneous study of Li depletion around the RGB bump, which will be detailed in McKenzie et al. (in preparation).

Dedicated observing programs have increased the average SNR for some clusters and expanded the sample to include additional clusters, such as M 22 (PI: M. McKenzie) and M 4 (PI: M. Howell). As discussed in Section 6.2.1, M 4 will be used to spectroscopically confirm whether the stars with lower asteroseismic masses belong to the light-element-enhanced population. This confirmation will be achieved through the re-analysis of Na, O, Mg, and Al lines, following the approach used for Li, since we advise against relying on these current light element abundances for globular cluster stars.

The cluster M 22, renowned for its bimodal *s*-process population (Marino et al. 2011; McKenzie et al. 2022, 2024), was observed as a crucial test case to evaluate the GALAH pipeline's ability to detect *s*-process abundance variations. While the pipeline successfully recovers the bimodal distribution, the scatter is larger than reported in previous studies. Additionally, as noted in Section 6.2.2, the precision of Eu measurements appears to have decreased between DR3 and DR4, particularly within globular cluster populations. Therefore, we recommend against using Eu from DR4 in future globular cluster publications.

Due to their low metallicity, globular clusters are particularly susceptible to the bug in the `flag_fe_h` discussed in Section 8.8.1. If this condition is relaxed, we recommend that all spectra and corresponding fits be manually inspected for quality before being included in any publications. Again, we reiterate that a boutique, custom reanalysis aiming to address these caveats in the globular cluster data will be the focus of upcoming work from McKenzie et al. (in preparation).

### 8.7. Fit optimisation

As described in Section 4, we are using the `CURVE_FIT` function of `scipy.optimize` (Virtanen et al. 2020) to fit synthetic spectra to observed spectra, whose optimisation can get stuck in local minima. We have tried to automatically identify regions of the parameter space where the `SCIPY.OPTIMIZE.CURVE_FIT` function has become stuck. In particular for some red clump stars as well as cool giant stars with  $T_{\text{eff}} < 3\,750$  K and  $\log g < 0.5$  (see Section 6.2.1),



we have been able to recover a pattern of abundances that are stuck around their initial value. However, this pattern is not consistent enough to flag stars without a significant amount of false-positives. Because of the zero point corrections, these are shifted away from the usual initial guess of 0 dex depending on the element (see zero-points in Table C1).

Such a fitting failure would also be expected when applying *The Payne* (Ting et al. 2019) with its similar default setup that adopts parameter bounds for the fitting parameters and thus employs the `curve_fit` function with the *trust region reflective* (`trf`) method. Given the common use of `curve_fit`, future pipelines should test a range of approaches to avoid this issue. Firstly, instead of using `trf`, the *Dogbox* (`dogbox`) method, could be used. The method is potentially slower but more reliable for complex parameter spaces. It could be used to randomly check the convergence of the `trf` method or be applied only to regions where multiple local minima are expected.

Moving away from the `curve_fit` function, the `LEASTSQ`,<sup>ab</sup> `MINIMIZE` or the `DIFFERENTIAL_EVOLUTION` function of `scipy`'s `optimize` module could be used to test options of a more expensive but more extensive optimisation. Finally, multiple randomised initial starting guesses could be applied for `curve_fit`, but would multiply the computing costs linearly by the number of initial guesses.

The fitting optimisation and uncertainty estimation should be performed in a more sophisticated Bayesian framework that folds in photometric, astrometric, and asteroseismic information and their uncertainties. We have indeed implemented such a framework with a likelihood estimate from spectroscopic information and prior information based on photometric, astrometric, and asteroseismic estimates for test purposes. When implementing the resulting posterior into the Markov-Chain Monte-Carlo machinery of `EMCEE` (Foreman-Mackey et al. 2013), we have not been able to limit the computational time (when fitting all labels) to a competitive level with `curve_fit` and thus not implemented this approach for the analysis of a million spectra. We note, however, that a future analysis should implement this approach – either with `EMCEE` or Monte Carlo nested sampling algorithms like `ULTRANEST` (Buchner 2021). Furthermore, we suggest to either separate the likelihood and posterior estimation steps (see e.g. Gent et al. 2022) or limit the optimisation to only a few major stellar labels (see e.g. Traven et al. 2020).

## 8.8. Reliability of flags

We have tried to develop a quality assurance pipeline that automatically flags results and stars that may not be adequately analysed with our assumptions.

### 8.8.1. Bug in flag\_fe\_h">Bug in flag\_fe\_h

The quality flag for iron abundance, `flag_fe_h`, was computed similarly to the elemental abundances, that is, by comparing the best-fitting spectrum with a spectrum with the lowest grid value of the neural network subgrids. In the case of  $[\text{Fe}/\text{H}]$ , however, this is not the appropriate reference value. For example, for a star with  $[\text{Fe}/\text{H}] = -0.74$  dex, the spectrum will be compared to a reference with  $[\text{Fe}/\text{H}] = -0.75$  dex, which will appear essentially identical within the spectrum uncertainties, and the code concludes there are no spectral features that are significantly different. This has affected up to 34% of stars – most with detectable iron lines – and

we therefore do not recommend the use of this flag at all. In the future, such a test should be performed with respect to an actually low (undetectable) amount of iron, such as  $[\text{Fe}/\text{H}] = -4$  dex.

### 8.8.2. Fitting machinery stuck in local minimum

As laid out in Section 8.7, we have not been able to automatically flag all estimates for which our fitting machinery has become stuck in local minima, most notably at the initial value.

### 8.8.3. Binary or fast rotating star?

With the increasing number of turn-off stars as part of ongoing GALAH observations, we have tried to implement a more sensitive approach to identify binaries in this region. This may, however, mean that we have also introduced more false-positive detections of stars that are only fast rotating with higher  $v \sin i$ , rather than being a binary system. We therefore suggest carefully considering using or neglecting the accompanying flag in GALAH DR4 (see Table 5).

## 8.9. Bug of interstellar K velocity

As mentioned in Section 6.1.3, `rv_k_is` in v240705 is reported relative to the stellar radial velocity. To compute the barycentric radial velocity of the measured interstellar K, `rv_comp_1` has to be added to `rv_k_is`.

## 8.10. Summary of caveats

In summary, the most important caveats are:

- Noding in  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  around edges between neural networks: Our tests when switching between neural networks indicate that this effect for  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  should stay within the precision uncertainties. A more problematic effect might be that some elements could be fitted as part of one neural network based on the detectability tests that were performed at the grid centres of each neural network.
- Mismatches of photometry and spectroscopy: Both imperfect isochrone and spectrum models can drive a mismatch in the estimation of spectroscopic parameters. This is most notable around the secondary red clump region and also expected for highly extincted regions.
- Imperfect synthesis leading to trends in cool stars: The unreliable line data in cool stars causes increasingly inaccurate models and inferred stellar properties towards the coolest stars (see Kos et al. 2025). The coolest giant stars ( $T_{\text{eff}} < 3\,750$  K and  $\log g < 0.5$ ) still have unreliable parameters.
- Lower precision for Eu due to missing masking of neural networks.
- The radial velocity of interstellar K has to be corrected from the stellar to barycentric frame by adding `rv_comp_1`.

These caveats are a by-product of our ambitious goal to enhance the accuracy and precision of stellar parameters and elemental abundances, while vastly expanding the number of stars for which we report measurements. Each region of the Hertzsprung–Russell diagram brings its own set of challenges, whether in the complex physics of evolved stars or the fine-tuned data analysis

<sup>ab</sup> `LEASTSQ` is used for example by *The Cannon* version by Casey et al. (2016).

required for main-sequence stars. Yet, these efforts have culminated in the remarkable success of GALAH DR4, providing an incredibly rich and robust dataset for researchers. As we continue to explore the Galaxy, this data enables new discoveries and insights, but it is essential to take measurements and peculiar findings with a grain of salt—they may sometimes reflect the complexity of data analysis rather than intrinsic stellar properties. Despite these challenges, GALAH DR4 marks a significant leap forward, opening up exciting opportunities for the community to unravel the mysteries of our Galaxy.

## 9. Conclusions

The GALAH survey celebrates its 10th anniversary with the release of GALAH DR4, marking a decade of transformative contributions to our understanding of the Milky Way and the elemental composition of its stars. Over the years, GALAH has been pivotal in measuring and cataloguing the chemical fingerprints of stars, which serve as cosmic barcodes that reveal their formation histories, migration patterns, and the evolutionary processes that shaped our Galaxy.

With GALAH DR4, we have achieved notable advancements in the precision and accuracy of stellar parameters and elemental abundances for nearly a million stars. This release benefits from a decade of continuous development in spectroscopic techniques, calibration processes, and the adoption of cutting-edge models like 1D NLTE and even 3D NLTE synthesis for lithium abundances. The inclusion of photometric and astrometric information from *Gaia* DR3 has enhanced the reliability of stellar parameters, particularly for surface gravities, helping to resolve degeneracies of spectroscopic data. The unique value of GALAH lies in its detailed mapping of elements crucial to the studies of exoplanets and life as we know it. By tracking the abundances of carbon, nitrogen, and oxygen (CNO), rock-forming elements (e.g. Mg, Si, and Fe), as well as rare heavy elements used in modern electronics (e.g. Ce, La, and Nd), GALAH has provided key insights into how the building blocks of planets, life, and technology were forged in the interiors of stars and distributed throughout the Milky Way over billions of years.

In the last decade, 321 research outputs (176 of them refereed) have mentioned GALAH in their abstract.<sup>ac</sup> GALAH DR3 (Buder et al. 2021), the predecessor of this data release, was by far the most cited paper of the Monthly Notices of the Royal Astronomical Society in 2021 at the time when this manuscript was published. GALAH DR3 has made significant contributions across several major research fields. In stellar physics and evolution, GALAH has expanded our understanding of stellar structures, nucleosynthesis (Sanders, Belokurov, & Man 2021; Griffith et al. 2022), and lithium enrichment (Martell et al. 2021; Simpson et al. 2021; Bouma et al. 2021; Sayeed et al. 2024; Wang et al. 2024a). In galactic astronomy and archaeology, GALAH has mapped the Milky Way's chemical and kinematic properties (e.g. Bland-Hawthorn et al. 2019; Sharma et al. 2021; Sharma et al. 2022), shedding light on its formation, dynamics, and past mergers (Buder et al. 2022). The survey has also influenced planetary formation by examining the chemical environments of exoplanet host stars (Clark et al. 2021; Soares-Furtado et al. 2021; Spaargaren et al. 2023; Wang et al. 2024b), while deepening our knowledge of the Galaxy's chemical evolution and complexity (Kos et al. 2021), especially

regarding neutron-capture and r-process elements (Matsuno et al. 2021; Aguado et al. 2021; Horta et al. 2022; Manea et al. 2024). Additionally, GALAH has provided insights into open clusters and star formation across the Galactic disc (e.g. Spina et al. 2021), with broader applications in extragalactic astronomy through a refined understanding of surviving structures of galaxy mergers and streams (Myeong et al. 2022; Buder et al. 2022; Manea, Hawkins, & Maas 2022) through its innovative chemical tagging techniques (Buder et al. 2022; Buder, Mijnders, & Buck 2024). In addition to its scientific discoveries, GALAH's influence has always been mutually beneficial with both photometric (Huang et al. 2021), asteroseismic (Zinn et al. 2022), and spectroscopic analyses (Nandakumar et al. 2022; Tsantaki et al. 2022; Soubiran, Brouillet, & Casamiquela 2022). GALAH information has aided the calibration and validation of surveys (Casagrande et al. 2021; Katz et al. 2023; Frémat et al. 2023) as well as the improvement of stellar ages by exploring chemical abundances (Hayden et al. 2022; Ratcliffe et al. 2024) and combining spectroscopic and other data (Hayden et al. 2022; Sahlholdt, Feltzing, & Feuillet 2022; Queiroz et al. 2023). GALAH's extensive observations also covered a range of rare or peculiar objects, such as variable stars (Jayasinghe et al. 2021) or metal-poor stars (Da Costa et al. 2023).

The next decade holds tremendous potential for further breakthroughs as GALAH continues its mission to observe and analyse stars across the Milky Way. With a clear goal of surpassing the 1 million star milestone, GALAH not only refines its data reduction and spectral analysis techniques but also paves the way for other ambitious surveys, such as SDSS-V (Kollmeier et al. 2017), 4MOST (de Jong et al. 2019), and WEAVE (Dalton et al. 2014) at similar spectral resolution or MSE (The MSE Science Team et al. 2019) and HRMOS (Magrini et al. 2023) at higher spectral resolution. As a trailblazer in the field of stellar spectroscopy, GALAH's approach has set the standard for these upcoming surveys, and its legacy will be cemented by the release of one final data set that will address the caveats and challenges discussed in this fourth data release. GALAH will undoubtedly continue to influence not only planetary, stellar and galactic astronomy but also broaden our understanding of the cosmos and the elements that shape modern life.

**Acknowledgements.** We acknowledge the traditional owners of the land on which the AAT and ANU stand, the Gamilaraay, the Ngannawal and the Ngambri peoples. We pay our respects to Elders, past and present, and are proud to continue their tradition of surveying the night sky in the Southern hemisphere.

We extend our heartfelt thanks to the entire staff at Siding Spring Observatory, both past and present, for their dedicated maintenance of 2dF-HERMES and invaluable support during the decade of observations. This project would not have been possible without the collective efforts of the many individuals who have contributed their expertise, time, and hard work, including Ashley Anderson, Paula Boubel, Rob Brookfield, James Cameron, Steve Chapman, Tony Farrell, Kristin Feigert, Andy Green, Doug Grey, Dionne Haynes, Steve Lee, Chris Lidman, Angel Lopez-Sanchez, Chris McCowage, Quentin Parker, Rob Patterson, Susan Patterson, Michael Andre Phillips, Chris Ramage, Murray Riding, Mike Sharrott, Andy Sheinis, Lee Spitler, Darren Stafford, Lew Waller, Fred Watson, Duncan Wright, as well as Tayyaba Zafar, and all those who played a part in making GALAH a success.

The operation of the AAT is funded by the AAT Consortium, which includes The Australian National University (operator), The University of New South Wales, The University of Sydney, Macquarie University, Western Sydney University, The University of Melbourne, Swinburne University, Monash University, The University of Queensland, The University of Southern Queensland and The University of Tasmania, with Astronomy Australia Limited (AAL) as Consortium Manager.

<sup>ac</sup>A total of 1 539 astronomical research outputs mentions (1 193 refereed) mentioned GALAH throughout their manuscript.

This work was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. SB acknowledges support from the Australian Research Council under grant number DE240100150. SLM, DBZ and GFL acknowledge support from the Australian Research Council through Discovery Program grant DP220102254. SLM, BTM and KB acknowledge support from the UNSW Scientia Program. JK, GT and TZ acknowledge financial support of the Slovenian Research Agency (research core funding No. P1-0188) and the European Space Agency (PRODEX Experiment Arrangements No. 4000142234 and No. 4000143450). AMA acknowledges support from the Swedish Research Council (VR 2020-03940) and from the Crafoord Foundation via the Royal Swedish Academy of Sciences (CR 2024-0015).

**Facilities.** AAT with 2dF-HERMES at Siding Spring Observatory: AAT observations for this data release were performed under programs 2013B/13, 2014A/25, 2015A/3, 2015A/19, 2015B/1, 2015B/19, 2016A/22, 2016B/10, 2016B/12, 2017A/14, 2017A/18, 2017B/16, 2018A/18, 2018B/15, 2019A/1, 2019A/15, 2020B/14, 2020B/23, 2022B/02, 2022B/05, 2023A/04, 2023A/08, 2023A/09, 2023B/04, and 2023B/05.

**AAO Data Central:** This paper includes data that has been provided by AAO Data Central (datacentral.org.au) and makes use of services and code that have been provided by AAO Data Central.

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

**Other facilities:** This publication makes use of data products from the Two Micron All Sky Survey (Skrutskie et al. 2006) and the CDS VizieR catalogue access tool (Ochsenbein, Bauer, & Marcout 2000). This research was supported by computational resources provided by the Australian Government through the National Computational Infrastructure (NCI) under the National Computational Merit Allocation Scheme and the ANU Merit Allocation Scheme (project y89) and HPCAI Talent Programme Scholarship (project hl99).

**Software.** The research for this publication was coded in PYTHON (version 3.7.4) and included its packages ASTROPY (v. 3.2.2; Astropy Collaboration et al. 2013; Astropy Collaboration et al. 2018), ASTROQUERY (v. 0.4; Ginsburg et al. 2019), CORNER (v. 2.0.1; Foreman-Mackey 2016), GALPY (version 1.6.0; Bovy 2015), IPYTHON (v. 7.8.0; Pérez & Granger 2007), MATPLOTLIB (v. 3.1.3; Hunter 2007), NUMPY (v. 1.17.2; Walt et al. 2011), SCIPY (version 1.3.1; Virtanen et al. 2020), SKLEARN (v. 0.21.3; Pedregosa et al. 2011). We further made use of TOPCAT (version 4.7; Taylor 2005);

**Linelist.** Our linelist, as described in Section 3.2, makes use of the following work: References: 1982ApJ...260.395C: Cardon et al. (1982), 1983MNRAS.204.883B|1989A&A...208.157G: Blackwell, Menon, & Petford (1983), Grevesse et al. (1989), 1990JQSRT.43.207C: Chang & Tang (1990), 1992A&A...255.457D: Davidson et al. (1992), 1993A&AS...99.179H: Hibbert et al. (1993),

1993PhyS...48.297N: Nahar (1993), 1998PhRvA.57.1652Y: Yan, Tambasco, & Drake (1998), 1999ApJS.122.557N: Nitz et al. (1999), 2008JPCRD.37.709K: Kelleher & Podobedova (2008),

2009A&A...497.611M: Meléndez & Barbay (2009),

2014ApJS.211...20W: Wood et al. (2014), 2014ApJS.215...20L: Lawler et al. (2014), 2014ApJS.215...23D: Den Hartog et al. (2014a), 2014MNRAS.441.3127R: Ruffoni et al. (2014),

2015ApJS.220...13L: Lawler, Sneden, & Cowan (2015),

2015ApJS.220...13L\_1982ApJ...260.395C: Lawler et al. (2015); Cardon et al. (1982), 2017MNRAS.471.532P: Palmeri et al. (2017), 2017PhRvA.95e2507T: Trubko et al. (2017), BGHL: Biemont et al. (1981), BIPS: Blackwell et al. (1979), BK: Bard & Kock (1994), BK+BWL: Bard & Kock (1994); O'Brian et al. (1991), BK+GESB82d+BWL: Bard & Kock (1994), Blackwell et al. (1982b), O'Brian et al. (1991), BKK: Bard, Kock, & Kock (1991), BKK+GESB82c+BWL: Bard et al. (1991); Blackwell et al. (1982a);

O'Brian et al. (1991), BLNP: Blackwell-Whitehead et al. (2006), BWL: O'Brian et al. (1991),

BWL+2014MNRAS.441.3127R: O'Brian et al. (1991); Ruffoni et al. (2014), BWL+GESHRL14: O'Brian et al. (1991); Den Hartog et al. (2014a), CB: Corliss & Bozman (1962), DLSSC: Den Hartog et al. (2011), FMW: Fuhr, Martin, & Wiese (1988), GARZ|BL: Garz (1973); O'Brian & Lawler (1991), GESB82c+BWL: Blackwell et al. (1982a); O'Brian et al. (1991), GESB86: Blackwell et al. (1986), GESB86+BWL: Blackwell et al. (1986); O'Brian et al. (1991), GESMCHF: Froese Fischer, Tachiev, & Irimia (2006), Grevesse2015: Grevesse et al. (2015), HLSC: Den Hartog et al. (2003), K06: Kurucz (2006), K07: Kurucz (2007), K08: Kurucz (2008), K09: Kurucz (2009), K10: Kurucz (2010), K13: Kurucz (2013), K14: Kurucz (2014), KL-astro: astrophysical, KR|1989ZPhyD.11.287C: Kock & Richter (1968), Carlsson et al. (1989), LBS: Lawler et al. (2001a), LD: Lawler & Dakin (1989), LD-HS: Lawler et al. (2006), LGWSC: Lawler et al. (2013), LSCI: Lawler et al. (2009), LWSH: Lawler et al. (2001b), MA-astro: astrophysical, MC: Meggers, Corliss, & Scribner (1975), MFW: Martin, Fuhr, & Wiese (1988), MRW: May, Richter, & Wichelmann (1974), NIST: Ralchenko et al. (2010), NWL: Nitz, Wickliffe, & Lawler (1998), PQWB: Palmeri et al. (2000), RU: Raassen & Uylings (1998), S: Smith (1988), SLS: Sobeck, Lawler, & Sneden (2007), SR: Smith & Raggett (1981), VGH: Vaeck, Godefroid, & Hansen (1988), WLSC: Wood et al. (2013), WSL: Wickliffe, Salih, & Lawler (1994).

**Data availability.** See Section 7.

## References

- Abdurro'uf, ACCETTA, K., Aerts, C., et al. 2022, *ApJS*, **259**, 35  
Aguado, D. S., et al. 2021, *ApJ*, **908**, L8  
Amarsi, A. M., & Asplund, M. 2017, *MNRAS*, **464**, 264  
Amarsi, A. M., Barklem, P. S., Asplund, M., Collet, R., & Zatsarinny, O. 2018a, *A&A*, **616**, A89  
Amarsi, A. M., Barklem, P. S., Collet, R., Grevesse, N., & Asplund, M. 2019a, *A&A*, **624**, A111  
Amarsi, A. M., et al. 2020a, *A&A*, **636**, A120  
Amarsi, A. M., Liljegren, S., & Nissen, P. E. 2022, *A&A*, **668**, A68  
Amarsi, A. M., Nissen, P. E., & Skúladóttir, Á. 2019b, *A&A*, **630**, A104  
Amarsi, A. M., et al. 2018b, *A&A*, **615**, A139  
Amarsi, A. M., et al. 2020b, *A&A*, **642**, A62  
Andrae, R., et al. 2023, *A&A*, **674**, A27  
Asplund, M. 2005, *ARA&A*, **43**, 481  
Astropy Collaboration, et al. 2013, *A&A*, **558**, A33  
Astropy Collaboration, et al. 2018, *AJ*, **156**, 123  
Bailer-Jones, C. A. L., Rybizki, J., Fouseneau, M., Demleitner, M., & Andrae, R. 2021, *AJ*, **161**, 147  
Bailer-Jones, C. A. L., Rybizki, J., Fouseneau, M., Mantelet, G., & Andrae, R. 2018, *AJ*, **156**, 58  
Bard, A., Kock, A., & Kock, M. 1991, *A&A*, **248**, 315, (BKK)  
Bard, A., & Kock, M. 1994, *A&A*, **282**, 1014, (BK)  
Barden, S. C., et al. 2010, *SPIE*, **7735**, 09  
Bastian, N., & Lardo, C. 2018, *ARA&A*, **56**, 83  
Baumgardt, H., & Vasiliev, E. 2021, *MNRAS*, **505**, 5957  
Bedell, M., et al. 2018, *ApJ*, **865**, 68  
Beeson, K. L., et al. 2024, *MNRAS*, **529**, 2483  
Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, *MNRAS*, **478**, 611  
Bensby, T., Feltzing, S., & Oey, M. S. 2014, *A&A*, **562**, A71  
Bergemann, M., et al. 2016, *A&A*, **594**, A120  
Bergemann, M., et al. 2019, *A&A*, **631**, A80  
Biemont, E., Grevesse, N., Hannaford, P., & Lowe, R. M. 1981, *ApJ*, **248**, 867, (BGHL)  
Binney, J. 2012, *MNRAS*, **426**, 1324  
Blackwell, D. E., Booth, A. J., Menon, S. L. R., & Petford, A. D. 1986, *MNRAS*, **220**, 289  
Blackwell, D. E., Ibbetson, P. A., Petford, A. D., & Shallis, M. J. 1979, *MNRAS*, **186**, 633, (BIPS)  
Blackwell, D. E., Menon, S. L. R., & Petford, A. D. 1983, *MNRAS*, **204**, 883



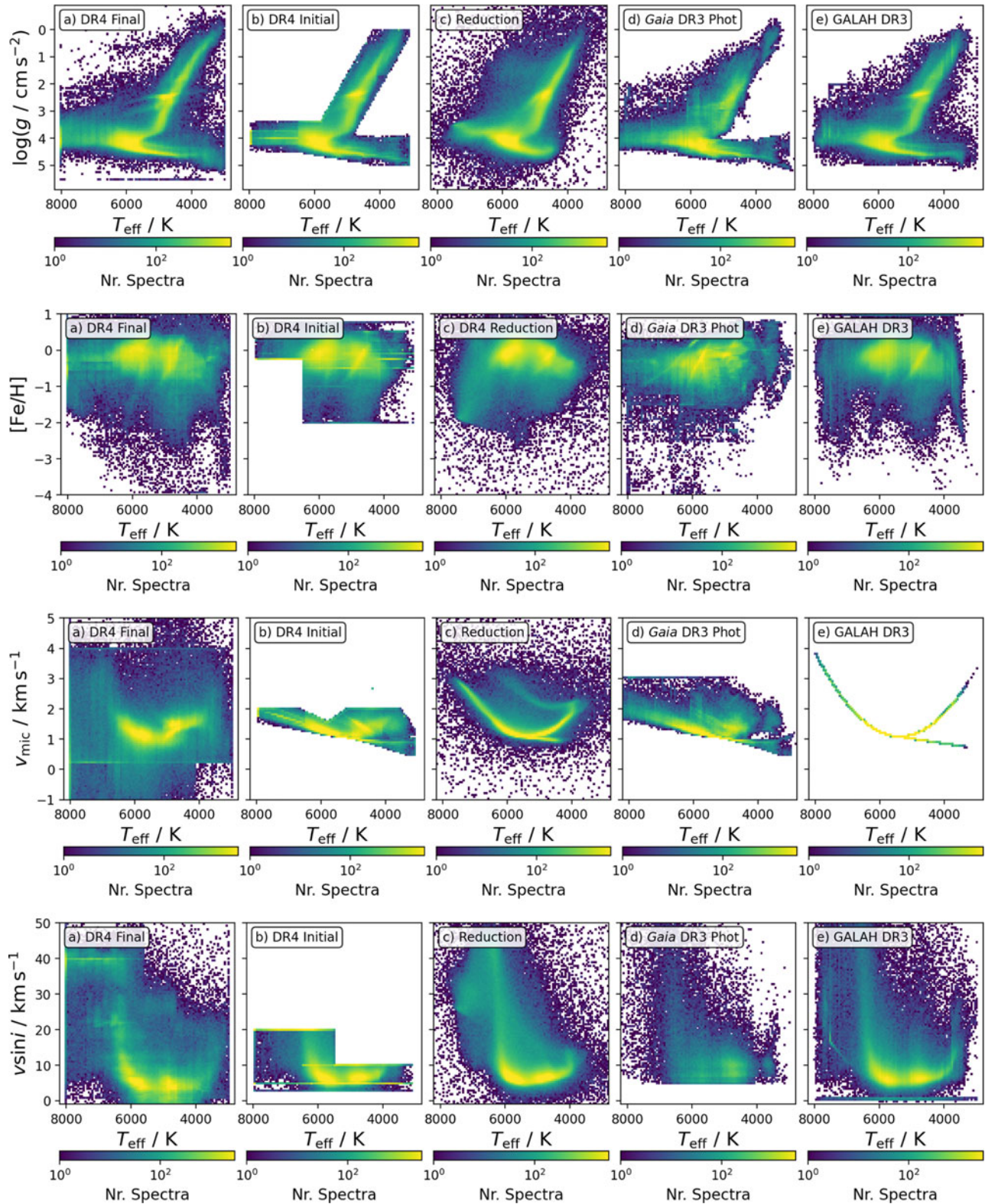
- Blackwell, D. E., Petford, A. D., Shallis, M. J., & Simmons, G. J. 1982a, *MNRAS*, **199**, 43
- Blackwell, D. E., Petford, A. D., & Simmons, G. J. 1982b, *MNRAS*, **201**, 595
- Blackwell-Whitehead, R. J., et al. 2006, *MNRAS*, **373**, 1603, (BLNP)
- Bland-Hawthorn, J., & Gerhard, O. 2016, *ARA&A*, **54**, 529
- Bland-Hawthorn, J., et al. 2019, *MNRAS*, **486**, 1167
- Botnen, A., & Carlsson, M. 1999, in *Astrophysics and Space Science Library*, Vol. 240, Numerical Astrophysics, ed. S. M. Miyama, K. Tomisaka, & T. Hanawa, 379
- Bouma, L. G., Curtis, J. L., Hartman, J. D., Winn, J. N., & Bakos, G. Á. 2021, *AJ*, **162**, 197
- Bovy, J. 2015, *ApJS*, **216**, 29
- Bressan, A., et al. 2012, *MNRAS*, **427**, 127
- Brzeski, J., Case, S., & Gers, L. 2011, *SPIE*, 8125, 04
- Buchner, J. 2021, *JOSS*, **6**, 3001
- Buder, S., Mijnders, L., & Buck, T. 2024, *MNRAS*, **532**, 1010
- Buder, S., et al. 2021, *MNRAS*, **506**, 150
- Buder, S., et al. 2022, *MNRAS*, **510**, 2407
- Cantat-Gaudin, T., & Anders, F. 2020, *A&A*, **633**, A99
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, **345**, 245
- Cardon, B. L., Smith, P. L., Scalo, J. M., Testerman, L., & Whaling, W. 1982, *ApJ*, **260**, 395
- Carlsson, J., Stenroos, L., & Svanberg, S. 1989, *ZPhAMC*, **11**, 287
- Carretta, E., Bragaglia, A., Gratton, R., & Lucatello, S. 2009a, *A&A*, **505**, 139
- Carretta, E., et al. 2009b, *A&A*, **505**, 117
- Casagrande, L., & VandenBerg, D. A. 2018, *MNRAS*, **475**, 5023
- Casagrande, L., et al. 2021, *MNRAS*, **507**, 2684
- Casey, A. R., et al. 2016, *ArXiv e-prints*, [arXiv:1603.03040](https://arxiv.org/abs/1603.03040)
- Chang, T. N., & Tang, X. 1990, *J. Quant. Spec. Radiat. Transf.*, **43**, 207
- Clark, J. T., et al. 2021, *MNRAS*, **504**, 4968
- Corliss, C. H., & Bozman, W. R. 1962, *NBS Monograph*, Vol. 53, Experimental Transition Probabilities for Spectral Lines of Seventy Elements (US Government Printing Office), (CB)
- Cutri, R. M., et al. 2014, *VizieR Online Data Catalog*, **2328**
- Da Costa, G. S., et al. 2023, *MNRAS*, **520**, 917
- Dalton, G., et al. 2014, *SPIE*, **9147**, 0L
- Davidson, M. D., Snoek, L. C., Volten, H., & Doenszelmann, A. 1992, *A&A*, **255**, 457
- de Jong, R. S., et al. 2019, *Msngr*, **175**, 3
- Den Hartog, E. A., Lawler, J. E., Sneden, C., & Cowan, J. J. 2003, *ApJS*, **148**, 543, (HLSC)
- Den Hartog, E. A., Lawler, J. E., Sobek, J. S., Sneden, C., & Cowan, J. J. 2011, *ApJS*, **194**, 35, (DLSSC)
- Den Hartog, E. A., et al. 2014a, *ApJS*, **215**, 23
- Dutra-Ferreira, L., Pasquini, L., Smiljanic, R., Porto de Mello, G. F., & Steffen, M. 2016, *A&A*, **585**, A75
- Evans, D. W., et al. 2018, *A&A*, **616**, A4
- Farrell, T. J., et al. 2014, *SPIE*, **9152**, 23
- Foreman-Mackey, D. 2016, *JOSS*, **1**, 24
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Fouesneau, M., et al. 2023, *A&A*, **674**, A28
- Frémat, Y., et al. 2023, *A&A*, **674**, A8
- Froese Fischer, C., Tachiev, G., & Irimia, A. 2006, *ADNDT*, **92**, 607
- Fuhr, J. R., Martin, G. A., & Wiese, W. L. 1988, *JPCRD*, **17**, (FMW)
- Gaia Collaboration, et al. 2021a, *A&A*, **649**, A1
- Gaia Collaboration, et al. 2023, *A&A*, **674**, A1
- Gallagher, A. J., et al. 2020, *A&A*, **634**, A55
- Garz, T. 1973, *A&A*, **26**, 471, (GARZ)
- Gent, M. R., et al. 2022, *A&A*, **658**, A147
- Gilmore, G., et al. 2022, *A&A*, **666**, A120
- Ginsburg, A., et al. 2019, *AJ*, **157**, 98
- Gratton, R., et al. 2019, *A&A Rev.*, **27**, 8
- Grevesse, N., Asplund, M., & Sauval, A. J. 2007, *Space Sci. Rev.*, **130**, 105
- Grevesse, N., Blackwell, D. E., & Petford, A. D. 1989, *A&A*, **208**, 157
- Grevesse, N., Scott, P., Asplund, M., & Sauval, A. J. 2015, *A&A*, **573**, A27
- Griffith, E. J., et al. 2022, *ApJ*, **931**, 23
- Gustafsson, B., et al. 2008, *A&A*, **486**, 951
- Hayden, M. R., et al. 2022, *MNRAS*, **517**, 5325
- Heijmans, J., et al. 2012, *SPIE*, **8446**, 0W
- Heiter, U., et al. 2015, *A&A*, **582**, A49
- Heiter, U., et al. 2021, *A&A*, **645**, A106
- Helmi, A., et al. 2018, *Nature*, **563**, 85
- Hibbert, A., Biemont, E., Godefroid, M., & Vaeck, N. 1993, *A&AS*, **99**, 179
- Hinkle, K., Wallace, L., Valenti, J., & Harmer, D., eds. 2000, *Visible and Near Infrared Atlas of the Arcturus Spectrum, 3727-9300 Å* (Astron. Soc. Pac.)
- Ho, A. Y. Q., et al. 2017, *ApJ*, **836**, 5
- Hoeijmakers, H. J., et al. 2015, *A&A*, **575**, A20
- Hon, M., et al. 2021, *ApJ*, **919**, 131
- Horta, D., Ness, M. K., Rybizki, J., Schiavon, R. P., & Buder, S. 2022, *MNRAS*, **513**, 5477
- Hourihane, A., et al. 2023, *A&A*, **676**, A129
- Howell, M., Campbell, S. W., Stello, D., & De Silva, G. M. 2022, *MNRAS*, **515**, 3184
- Howell, S. B., et al. 2014, *PASP*, **126**, 398
- Huang, Y., et al. 2021, *ApJ*, **907**, 68
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Jayasinghe, T., et al. 2021, *MNRAS*, **503**, 200
- Jofré, P., et al. 2018, *RNAAS*, **2**, 152
- Jofré, P., et al. 2014, *A&A*, **564**, A133
- Jofré, P., et al. 2015, *A&A*, **582**, A81
- Jofré, P., et al. 2017, *A&A*, **601**, A38
- Johnson, C. I., & Pilachowski, C. A. 2010, *ApJ*, **722**, 1373
- Jönsson, H., et al. 2020, *AJ*, **160**, 120
- Katz, D., et al. 2023, *A&A*, **674**, A5
- Kawka, A., et al. 2020, *MNRAS*, **495**, L129
- Kelleher, D. E., & Podobedova, L. I. 2008, *JPCRD*, **37**, 709
- Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, *ApJ*, **900**, 179
- Kock, M., & Richter, J. 1968, *ZAp*, **69**, 180, (KR)
- Kollmeier, J. A., et al. 2017, *arXiv e-prints*, [arXiv:1711.03234](https://arxiv.org/abs/1711.03234)
- Kos, J., et al. 2018, *MNRAS*, **480**, 5475
- Kos, J., et al. 2021, *MNRAS*, **506**, 4232
- Kos, J., et al. 2025, *arXiv e-prints*, [arXiv:2501.06140](https://arxiv.org/abs/2501.06140)
- Kurucz, R. L. 2006, Database of observed and predicted atomic transitions
- Kurucz, R. L. 2007, Database of observed and predicted atomic transitions
- Kurucz, R. L. 2008, Database of observed and predicted atomic transitions
- Kurucz, R. L. 2009, Database of observed and predicted atomic transitions
- Kurucz, R. L. 2010, Database of observed and predicted atomic transitions
- Kurucz, R. L. 2013, Database of observed and predicted atomic transitions
- Kurucz, R. L. 2014, Database of observed and predicted atomic transitions
- Lawler, J. E., Bonvallet, G., & Sneden, C. 2001a, *Astrophys. J.*, **556**, 452, (LBS)
- Lawler, J. E., & Dakin, J. T. 1989, *JOSAB*, **6**, 1457, (LD)
- Lawler, J. E., Den Hartog, E. A., Sneden, C., & Cowan, J. J. 2006, *ApJS*, **162**, 227, (LD-HS)
- Lawler, J. E., Guzman, A., Wood, M. P., Sneden, C., & Cowan, J. J. 2013, *ApJS*, **205**, 11
- Lawler, J. E., Sneden, C., & Cowan, J. J. 2015, *ApJS*, **220**, 13
- Lawler, J. E., Sneden, C., Cowan, J. J., Ivans, I. I., & Den Hartog, E. A. 2009, *ApJS*, **182**, 51, (LSCI)
- Lawler, J. E., Wickliffe, M. E., den Hartog, E. A., & Sneden, C. 2001b, *ApJ*, **563**, 1075, (LWHS)
- Lawler, J. E., et al. 2014, *ApJS*, **215**, 20
- Leenaarts, J., & Carlsson, M. 2009, *ASPC*, **415**, 87
- Lewis, A. 2019, *arXiv e-prints*, [arXiv:1910.13970](https://arxiv.org/abs/1910.13970)
- Lewis, I. J., et al. 2002, *MNRAS*, **333**, 279
- Lind, K., Asplund, M., & Barklem, P. S. 2009a, *A&A*, **503**, 541
- Lind, K., Asplund, M., Barklem, P. S., & Belyaev, A. K. 2011, *A&A*, **528**, A103
- Lind, K., Primas, F., Charbonnel, C., Grundahl, F., & Asplund, M. 2009b, *A&A*, **503**, 545
- Lindgren, L., et al. 2018, *A&A*, **616**, A2
- Lindgren, L., et al. 2021a, *A&A*, **649**, A4
- Lindgren, L., et al. 2021b, *A&A*, **649**, A2
- Mackereth, J. T., & Bovy, J. 2018, *PASP*, **130**, 114501
- Magrini, L., et al. 2023, *arXiv e-prints*, [arXiv:2312.08270](https://arxiv.org/abs/2312.08270)
- Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, *ApJ*, **739**, 25
- Manea, C., Hawkins, K., & Maas, Z. G. 2022, *MNRAS*, **511**, 2829

- Manea, C., et al. 2024, *ApJ*, **972**, 69
- Mann, A. W., Gaidos, E., Lépine, S., & Hilton, E. J. 2012, *ApJ*, **753**, 90
- Marigo, P., et al. 2017, *ApJ*, **835**, 77
- Marino, A. F., et al. 2011, *A&A*, **532**, A8
- Martell, S. L., et al. 2017, *MNRAS*, **465**, 3203
- Martell, S. L., et al. 2021, *MNRAS*, **505**, 5340
- Martig, M., et al. 2016, *MNRAS*, **456**, 3655
- Martin, G., Fuhr, J., & Wiese, W. 1988, *JPhChRDS*, 17
- Masseron, T., & Gilmore, G. 2015, *MNRAS*, **453**, 1855
- Matsuno, T., et al. 2021, *A&A*, **650**, A110
- May, M., Richter, J., & Wichelmann, J. 1974, *A&AS*, **18**, 405, (MRW)
- McKemmish, L. K., et al. 2019, *MNRAS*, **488**, 2836
- McKenzie, M., et al. 2022, *MNRAS*, **516**, 3515
- McKenzie, M., et al. 2024, *MNRAS*, **527**, 7940
- McMillan, P. J. 2017, *MNRAS*, **465**, 76
- Meggers, W. F., Corliss, C. H., & Scribner, B. F. 1975, Tables of spectral-line intensities. Part I, II - arranged by elements. (NBS), (MC)
- Meléndez, J., & Barbuy, B. 2009, *A&A*, **497**, 611
- Mészáros, S., Allende Prieto, C., Edvardsson, B., et al. 2012, *AJ*, **144**, 120
- Milone, A. P., & Marino, A. F. 2022, *Universe*, **8**, 359
- Miszalski, B., Shortridge, K., Saunders, W., Parker, Q. A., & Croom, S. M. 2006, *MNRAS*, **371**, 1537
- Monty, S., et al. 2023, *MNRAS*, **518**, 965
- Myeong, G. C., et al. 2022, *ApJ*, **938**, 21
- Nahar, S. N. 1993, *PhS*, **48**, 297
- Nandakumar, G., et al. 2022, *MNRAS*, **513**, 232
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, **808**, 16
- Ness, M., et al. 2016, *ApJ*, **823**, 114
- Ness, M. K., et al. 2019, *ApJ*, **883**, 177
- Nissen, P. E. 2015, *A&A*, **579**, A52
- Nissen, P. E., et al. 2020, *A&A*, **640**, A81
- Nissen, P. E., & Gustafsson, B. 2018, *A&A Rev.*, **26**, 6
- Nitz, D. E., Kunau, A. E., Wilson, K. L., & Lentz, L. R. 1999, *ApJS*, **122**, 557
- Nitz, D. E., Wickliffe, M. E., & Lawler, J. E. 1998, *ApJS*, **117**, 313, (NWL)
- Nordlander, T., & Lind, K. 2017, *A&A*, **607**, A75
- O'brian, T. R., & Lawler, J. E. 1991, *PhRvA*, **44**, 7134, (BL)
- O'Brian, T. R., Wickliffe, M. E., Lawler, J. E., Whaling, W., & Brault, J. W. 1991, *JOSAB*, **8**, 1185, (BWL)
- Ochsenbein, F., Bauer, P., & Marcout, J. 2000, *A&AS*, **143**, 23
- Osorio, Y., et al. 2015, *A&A*, **579**, A53
- Osorio, Y., Lind, K., Barklem, P. S., Allende Prieto, C., & Zatsarinny, O. 2019, *A&A*, **623**, A103
- Palmeri, P., et al. 2017, *MNRAS*, **471**, 532
- Palmeri, P., Quinet, P., Wyart, J., & Biémont, E. 2000, *PhS*, **61**, 323, (PQWB)
- Pedregosa, F., et al. 2011, *J Mach Learn Res*, **12**, 2825
- Pérez, F., & Granger, B. E. 2007, *CSE*, **9**, 21
- Piskunov, N., & Valenti, J. A. 2017, *A&A*, **597**, A16
- Prša, A., et al. 2016, *AJ*, **152**, 41
- Queiroz, A. B. A., et al. 2023, *A&A*, **673**, A155
- Raassen, A. J. J., & Uylings, P. H. M. 1998, *A&A*, **340**, 300, (RU)
- Rains, A. D., et al. 2021, *MNRAS*, **504**, 5788
- Rains, A. D., et al. 2024, *MNRAS*, **529**, 3171
- Ralchenko, Y., Kramida, A., Reader, J., & NIST ASD Team. 2010, NIST Atomic Spectra Database (ver. 4.0.0), [Online].
- Ratcliffe, B., et al. 2024, *MNRAS*, **528**, 3464
- Recio-Blanco, A., et al. 2023, *A&A*, **674**, A29
- Reggiani, H., et al. 2019, *A&A*, **627**, A177
- Reid, M. J., & Brunthaler, A. 2004, *ApJ*, **616**, 872
- Ricker, G. R., et al. 2015, *JATIS*, **1**, 014003
- Riello, M., et al. 2021, *A&A*, **649**, A3
- Rimoldini, L., et al. 2023, *A&A*, **674**, A14
- Rix, H.-W., Ting, Y.-S., Conroy, C., & Hogg, D. W. 2016, *ApJ*, **826**, L25
- Ruffoni, M. P., et al. 2014, *MNRAS*, **441**, 3127
- Sahlholdt, C. L., Feltzing, S., & Feuillet, D. K. 2022, *MNRAS*, **510**, 4669
- Salaris, M., & Cassisi, S. 2006, Evolution of Stars and Stellar Populations (J. Wiley)
- Sanders, J. L., Belokurov, V., & Man, K. T. F. 2021, *MNRAS*, **506**, 4321
- Saydjari, A. K., Uzsoy, A. S. M., Zucker, C., Peek, J. E. G., & Finkbeiner, D. P. 2023, *ApJ*, **954**, 141
- Sayeed, M., et al. 2024, *ApJ*, **964**, 42
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, **403**, 1829
- Sharma, S., et al. 2018, *MNRAS*, **473**, 2004
- Sharma, S., et al. 2019, *MNRAS*, **490**, 5335
- Sharma, S., et al. 2021, *MNRAS*, **506**, 1761
- Sharma, S., et al. 2022, *MNRAS*, **510**, 734
- Sheinis, A., et al. 2015, *JATIS*, **1**, 035002
- Simpson, J. D., et al. 2021, *MNRAS*, **507**, 43
- Skrutskie, M. F., et al. 2006, *AJ*, **131**, 1163
- Smith, G. 1988, *JPhB*, **21**, 2827, (S)
- Smith, G., & Raggett, D. S. J. 1981, *JPhB*, **14**, 4015, (SR)
- Smith, V. V., et al. 2013, *ApJ*, **765**, 16
- Soares-Furtado, M., Cantiello, M., MacLeod, M., & Ness, M. K. 2021, *AJ*, **162**, 273
- Sobeck, J. S., Lawler, J. E., & Sneden, C. 2007, *ApJ*, **667**, 1267, (SLS)
- Soubiran, C., Brouillet, N., & Casamiquela, L. 2022, *A&A*, **663**, A4
- Spaargaren, R. J., Wang, H. S., Mojszsis, S. J., Ballmer, M. D., & Tackley, P. J. 2023, *ApJ*, **948**, 53
- Spina, L., et al. 2016, *A&A*, **593**, A125
- Spina, L., et al. 2021, *MNRAS*, **503**, 3279
- Steinmetz, M., et al. 2020, *AJ*, **160**, 82
- Stello, D., et al. 2015, *ApJ*, **809**, L3
- Taylor, M. B. 2005, *ASPC*, **347**, 29
- The MSE Science Team, et al. 2019, arXiv e-prints, arXiv:1904.04907
- Ting, Y.-S., Conroy, C., & Rix, H.-W. 2016, *ApJ*, **826**, 83
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Asplund, M. 2018, *ApJ*, **860**, 159
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, *ApJ*, **879**, 69
- Ting, Y.-S., Freeman, K. C., Kobayashi, C., De Silva, G. M., & Bland-Hawthorn, J. 2012, *MNRAS*, **421**, 1231
- Torra, F., et al. 2021, *A&A*, **649**, A10
- Traven, G., et al. 2020, *A&A*, **638**, A145
- Trubko, R., Gregoire, M. D., Holmgren, W. F., & Cronin, A. D. 2017, *Phys. Rev. A*, **95**, 052507
- Tsantaki, M., et al. 2022, *A&A*, **659**, A95
- Tsuji, T. 1976, *PASJ*, **28**, 543
- Vaeck, N., Godefroid, M., & Hansen, J. E. 1988, *Phys. Rev. A*, **38**, 2830, (VGH)
- Valenti, J. A., & Piskunov, N. 1996, *A&AS*, **118**, 595
- van Leeuwen, F. 2007, *A&A*, **474**, 653
- VandenBerg, D. A., et al. 2012, *ApJ*, **755**, 15
- Vasiliev, E., & Baumgardt, H. 2021, *MNRAS*, **505**, 5978
- Virtanen, P., et al. 2020, *NM*, **17**, 261
- Vogrincić, R., et al. 2023, *MNRAS*, **521**, 3727
- Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, **13**, 22
- Wang, E. X., et al. 2021, *MNRAS*, **500**, 2159
- Wang, E. X., et al. 2024a, *MNRAS*, **528**, 5394
- Wang, H. S., Quanz, S. P., Mahadevan, S., & Deal, M. 2024b, *A&A*, **688**, A225
- Wehrhahn, A., Piskunov, N., & Ryabchikova, T. 2023, *A&A*, **671**, A171
- Westendorp Plaza, C., Asensio Ramos, A., & Allende Prieto, C. 2023, *A&A*, **675**, A191
- Wheeler, A. J., Abruzzo, M. W., Casey, A. R., & Ness, M. K. 2023, *AJ*, **165**, 68
- Wheeler, A. J., Casey, A. R., & Abruzzo, M. W. 2024, *AJ*, **167**, 83
- Wickliffe, M. E., Salih, S., & Lawler, J. E. 1994, *J. Quant. Spec. Radiat. Transf.*, **51**, 545, (WSL)
- Wood, M. P., Lawler, J. E., Sneden, C., & Cowan, J. J. 2013, *ApJS*, **208**, 27
- Wood, M. P., Lawler, J. E., Sneden, C., & Cowan, J. J. 2014, *ApJS*, **211**, 20
- Xiang, M., et al. 2019, *ApJS*, **245**, 34
- Xiang, M., et al. 2022, *A&A*, **662**, A66
- Yan, Z.-C., Tambasco, M., & Drake, G. W. F. 1998, *Phys. Rev. A*, **57**, 1652
- Yong, D., et al. 2013, *MNRAS*, **434**, 3542
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *RAA*, **12**, 723
- Zinn, J. C., et al. 2020, *ApJS*, **251**, 23
- Zinn, J. C., et al. 2022, *ApJ*, **926**, 191



## Appendix A. Initial Parameters

We append the overview of the initial and final stellar parameters of GALAH DR4 in Fig. A1. We show the density distribution of  $\log g$ ,  $[\text{Fe}/\text{H}]$ ,  $v_{\text{mic}}$ , and  $v \sin i$  in each row as a function of  $T_{\text{eff}}$ .



**Figure A1.** Comparison of final GALAH DR4 stellar parameters (first column) against the initial parameters used in the *allstar* analysis (second column), estimates from the GALAH DR4 reduction pipeline (third column), *Gaia* DR3 (fourth column with  $v_{\text{mic}}$  based on the adjusted formula from Dutra-Ferreira et al. 2016), and GALAH DR3 (fifth column).



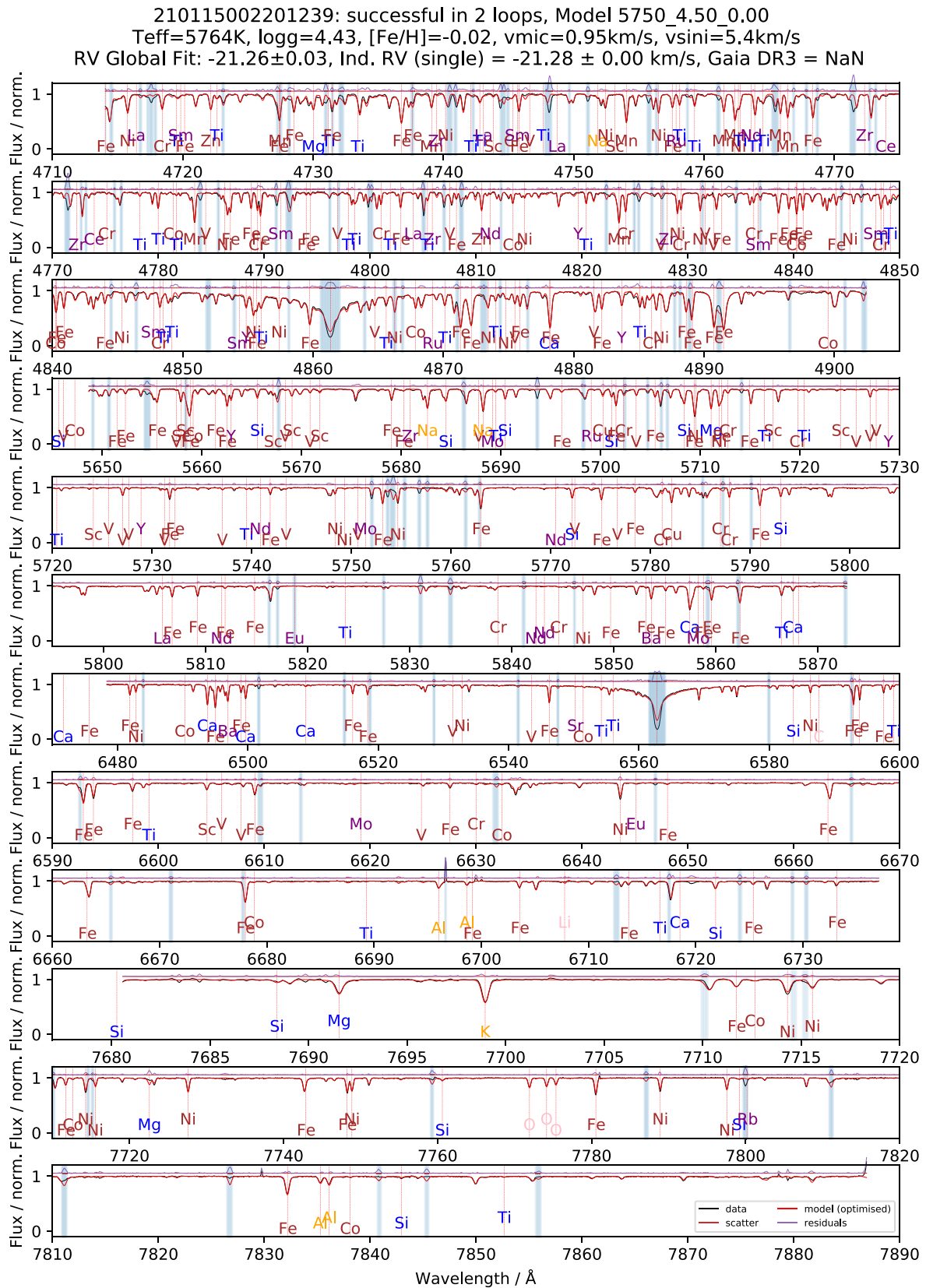
## Appendix B. Data Products

We append examples of data products of GALAH DR4 that were not already shown in the main manuscript. [Table B1](#) shows a shortened table schema of the `allstar` and `allspec` catalogues.

[Fig. B1](#) shows the automatically created fit comparison of the `allstar` module for Vesta (210115002201239). [Fig. B2](#) shows examples of the covariance matrices for Vesta and Arcturus, as representative examples for main-sequence and giant stars.

**Table B1.** Table schema of the GALAH DR4 main catalogues. Columns that are part of `allspec`, but not `allstar` are listed below the middle line. For compactness, we have combined repetitive columns (for example with integers  $N$ ). Detailed table schemas are available in the FITS headers of each catalogue file.

| Column            | Description                                    | Column        | Description                               |
|-------------------|--|---------------|---|
| subject_id        | GALAH identifier                               | tmass_id      | 2MASS identifier                          |
| gaiadr3_source_id | <i>Gaia</i> DR3 source_id                      | survey_name   | 2dF-HERMES Program                        |
| field_id          | GALAH Field ID                                 | setup         | allspec/allstar                           |
| mjd               | Modified Julian Date                           | ra            | propagated from <i>Gaia</i> DR3           |
| dec               | propagated from <i>Gaia</i> DR3                | flag_sp       | Major spectroscopic quality bitmask flag  |
| flag_sp_fit       | Major fitting quality flag                     | opt_loop      | Nr of optimisation loops used for fitting |
| flag_red          | Quality bitmask flag of reduction pipeline     | snr_px_ccdN   | Average SNR for CCD N                     |
| chi2_sp           | Chi2 value of spectroscopic fitting            | px_used_perc  | Percentage of spectrum used for fit       |
| model_name        | Used neural network for synthesis              | closest_model | Closest neural network for synthesis      |
| comp_time         | Computation time spent on spectrum             | fit_global_rv | RV fitted or fixed after co-adding?       |
| rv_comp_1         | Radial velocity of primary source              | e_rv_comp_1   | Uncertainty of rv_comp_1                  |
| rv_comp_2         | Radial velocity of potential secondary source  | e_rv_comp_2   | Uncertainty of rv_comp_1                  |
| rv_gaia_dr3       | Radial velocity in <i>Gaia</i> DR3             | e_rv_gaia_dr3 | Uncertainty of rv_gaia_dr3                |
| v_bary_eff        | Barycentric velocity correction                | teff          | Spectr. effective temperature             |
| e_teff            | Uncertainty teff                               | logg          | Photometric surface gravity               |
| e_logg            | Uncertainty logg_plx                           | fe_h          | Abundance of Fe as pseudo-metallicity     |
| e_fe_h            | Uncertainty fe_h                               | flag_fe_h     | Quality flag fe_h                         |
| vmic              | Microturbulence velocity (fitted)              | e_vmic        | Uncertainty vmic                          |
| vsini             | Broadening velocity                            | e_vsini       | Uncertainty of vsini                      |
| nn_li_fe          | Elemental abundance for [Li/Fe]                | nn_e_li_fe    | Uncertainty nn_li_fe                      |
| nn_flag_li_fe     | Quality bitmask flag of Li_fe                  | x_fe          | Elemental abundance for [X/Fe]            |
| e_x_fe            | Uncertainty of elemental abundance [X/Fe]      | flag_x_fe     | Quality bitmask flag of [X/Fe]            |
| mass              | Mass used for calculating log $g$ (plx)        | age           | Age estimated when calculating mass       |
| bc_ks             | Bolometric Correction of $K_S$ band            | a_ks          | Attenuation in $K_S$ -band $A(K_S)$       |
| lbol              | Bolometric Luminosity                          | r_med         | Median Distance                           |
| r_lo              | Lower Limit Distance                           | r_hi          | Higher Limit Distance                     |
| sb2_rv_N          | Nth perc. of RV residual signal                | ew_h_beta     | Equivalent Width of observed Hbeta core   |
| ew_h_alpha        | Equivalent Width of observed Halpha core       | res_h_beta    | Residual EW in Hbeta core                 |
| res_h_alpha       | Residual EW in Halpha core                     | ew_k_is       | EW of K7699 Interstellar Line             |
| sigma_k_is        | Gaussian sigma of K7699 Interstellar Line      | rv_k_is       | RV of K7699 Interstellar Line             |
| ew_dib5780        | Equivalent width of DIB NNNN                   | sigma_dib5780 | Gaussian sigma of DIB NNNN                |
| rv_dib5780        | RV of DIB NNNN                                 | ebv           | Extinction $E(B - V)$                     |
| phot_g_mean_mag   | Mean <i>Gaia</i> DR3 G-band apparent magnitude | bp_rp         | Color of $BP - RP$ bands                  |
| j_m               | 2MASS $J$ -band magnitude                      | j_msigcom     | Uncertainty of j_m                        |
| h_m               | 2MASS $H$ -band magnitude                      | h_msigcom     | Uncertainty of h_m                        |
| ks_m              | 2MASS $K_S$ -band magnitude                    | ks_msigcom    | Uncertainty of ks_m                       |
| W2mag             | ALLWISE W2-band magnitude                      | e_W2mag       | uncertainty of W2mag                      |
| ruwe              | RUWE reported by <i>Gaia</i> DR3               | parallax      | Astrometric parallax used for GALAH DR4   |
| parallax_error    | Uncertainty of astrometric parallax            | ew_li         | Equivalent width of Lithium 6708 LiI line |
| e_ew_li_low       | Lower uncertainty ew_li                        | e_ew_li_upp   | Upper uncertainty ew_li                   |
| a_li              | Absolute 3D NLTE Li abundance                  | a_li_upp_lim  | Upper limit of absolute 3D NLTE A(Li)     |
| e_a_li_low        | Lower uncertainty of a_li                      | e_a_li_upp    | Upper uncertainty of a_li                 |
| e_a_li_teff       | Uncertainty of A(Li) due to temperature        | flag_a_li     | Flag for a_li measurement                 |
| rv_comp_nr        | Nr RV cross-correlation function peaks         | rv_comp_1_p   | Prominence of rv_comp_1 in CCF            |
| rv_comp_2_h       | Height of rv_comp_1 in CCF                     | rv_comp_2_p   | Prominence of rv_comp_1 in CCF            |
| logg_spec         | Spectroscopic surface gravity estimate         | e_logg_spec   | Uncertainty logg_spec                     |



**Figure B1.** Example output of the *a11star* analysis for Vesta (210115002201239). The observed flux (black) is compared with the fitted model flux (red), and the residuals (purple) show the difference between the observed and modelled spectra. Important spectral lines are annotated with their corresponding elements, with element groups colour-coded for clarity. Blue-shaded regions represent the 5% of the spectrum that was masked and excluded from the fit to avoid contamination from outliers or poorly modelled lines.

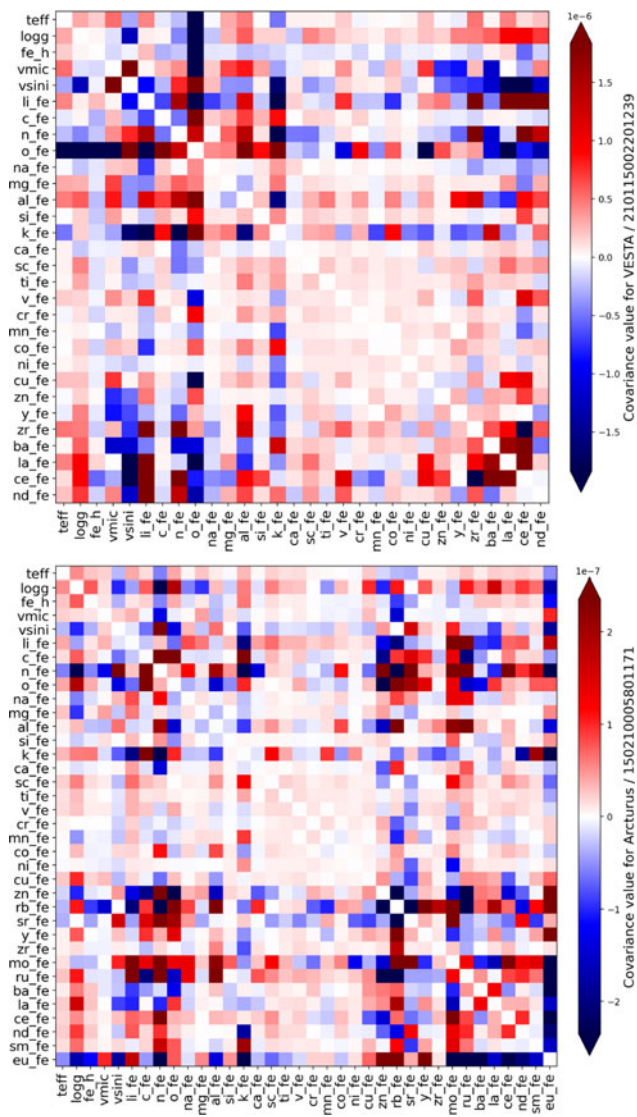


Figure B2. Covariance matrices for labels for Vesta (panel a) and Arcturus (panel b).

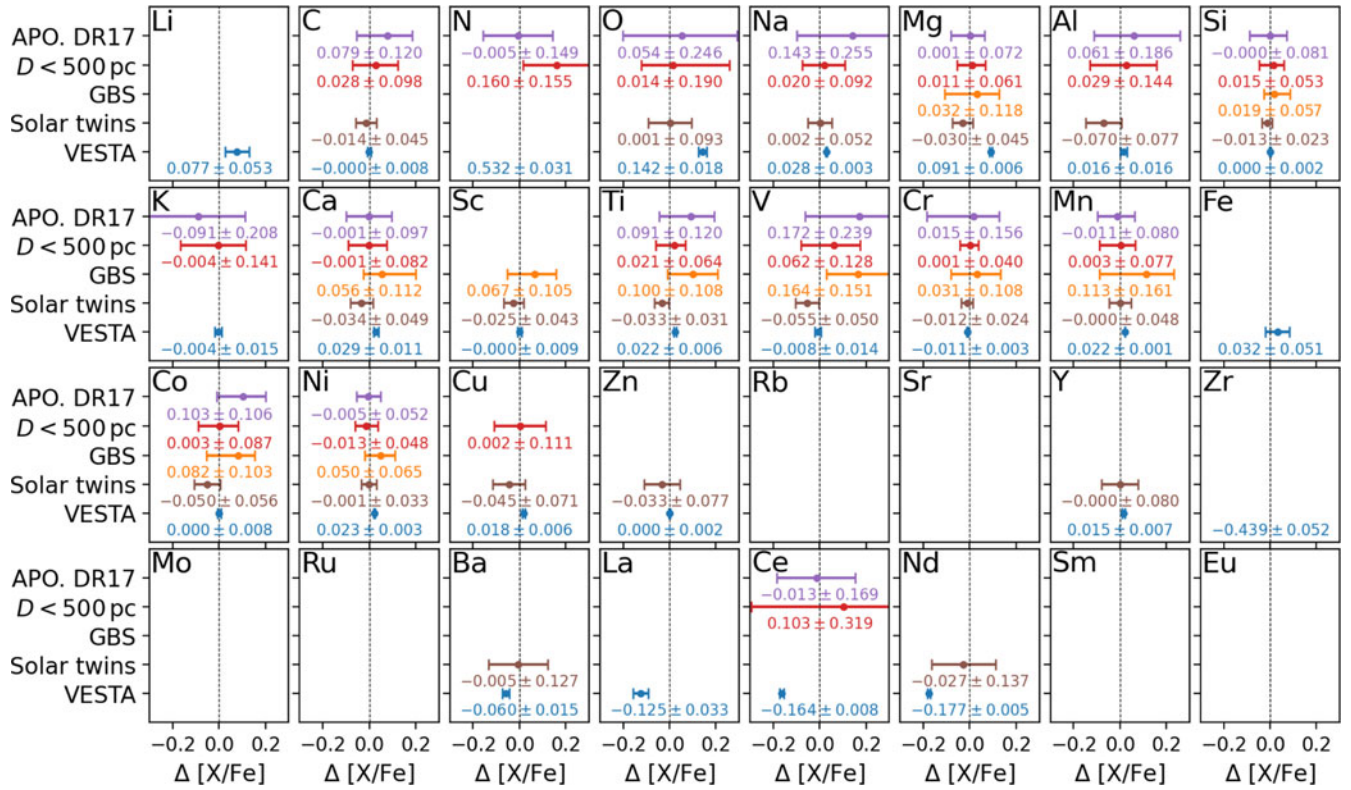
Appendix C. Stellar Parameter and Abundance Validation

Stellar parameter and abundance zero-points of the `allstar` module are listed in Table C1. A complete table, including the zero-points for the `allspec` module can be found as FITS file in the online repository. A compromise between the different accuracy abundance indicators is shown in Fig. C1 for the `allstar` module. Fig. C2 shows the precision of individual abundances. Figs. C3 and C4 show the remaining comparisons with APOGEE DR17 in addition to Figs. 21 and 22. Although many of the observed globular clusters are expected to show an abundance spread, including for iron, we show a collage of globular clusters with ascending iron abundance in Fig. C5, with each panel indicating the median iron abundance per cluster as well as the spread (scatter) of the iron abundance distribution and the average measurement uncertainty. A more comprehensive analysis of the globular clusters will be presented in upcoming work (McKenzie et al., in preparation). Finally, Fig. C6 shows the distribution of flagged stars in the Kiel diagram.

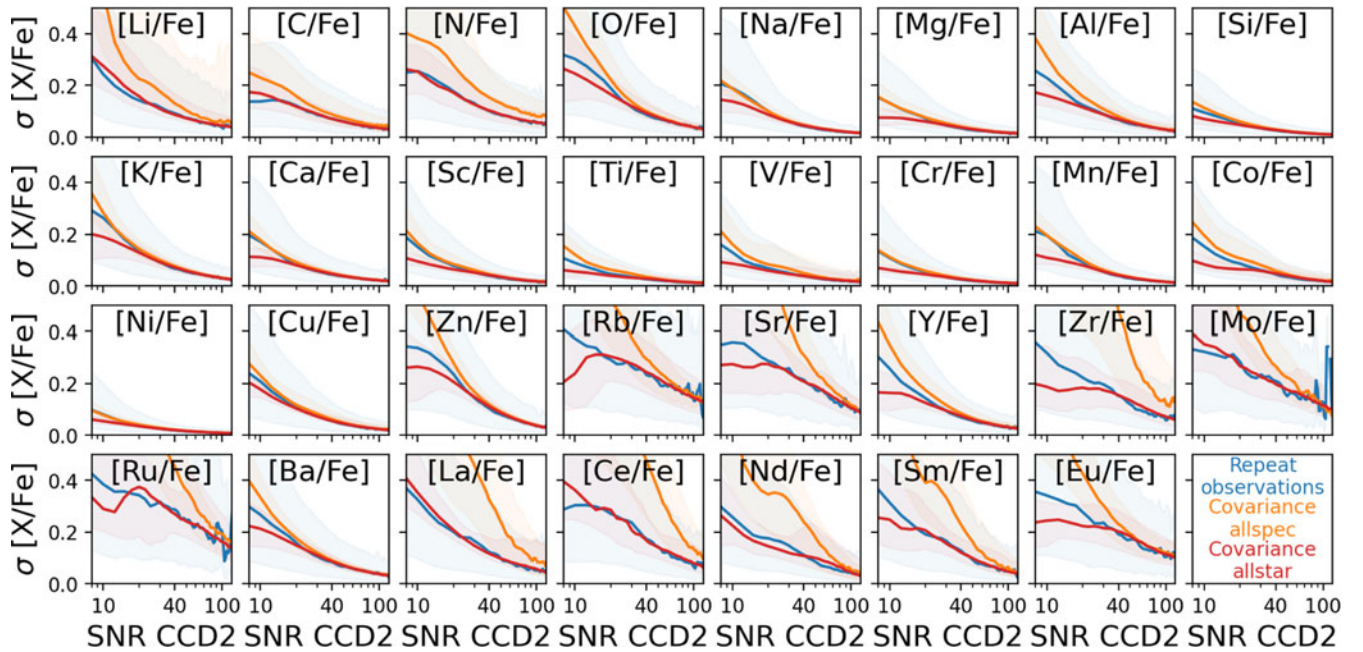
**Table C1.** Zero point estimates and corrections applied to the `allstar` measurements. We used Prša et al. (2016) as reference for Solar parameters and Grevesse et al. (2007), consistent with the MARCS model atmosphere composition (Gustafsson et al. 2008), as reference for Solar abundances. For reference, we also show the combined rotational and macroturbulence as well as micro-turbulence velocities from Jofré et al. (2014). Values for Vesta indicate our uncorrected measurements for the Vesta spectrum.

| Property         | Reference | Zeropoint | Shift  | Vesta     | $\Delta$ Vesta |
|------------------|-----------|-----------|--------|-----------|----------------|
| $T_{\text{eff}}$ | 5 772.0   | 5 772.0   | 0.0    | 5 752.261 | −19.739        |
| $\log g$         | 4.438     | 4.438     | 0.0    | 4.429     | −0.009         |
|                  | 0.0       | 0.049     | 0.049  | −0.019    | −0.068         |
| A(Fe)            | 7.45      | 7.499     | 0.049  | 7.431     | −0.068         |
| $v_{\text{mic}}$ | 1.06      | 1.06      | 0.0    | 1.0       | −0.06          |
| $v \sin i$       | 4.5       | 4.5       | 0.0    | 5.552     | 1.052          |
| A(Li)            | 1.05      | 1.05      | 0.0    | 1.108     | 0.058          |
| A(C)             | 8.39      | 8.393     | 0.003  | 8.348     | −0.045         |
| A(N)             | 7.78      | 7.705     | −0.075 | 8.368     | 0.663          |
| A(O)             | 8.66      | 8.659     | −0.001 | 8.784     | 0.125          |
| A(Na)            | 6.17      | 5.999     | −0.171 | 6.35      | 0.351          |
| A(Mg)            | 7.53      | 7.445     | −0.085 | 7.687     | 0.242          |
| A(Al)            | 6.37      | 6.185     | −0.185 | 6.552     | 0.367          |
| A(Si)            | 7.51      | 7.486     | −0.024 | 7.515     | 0.029          |
| A(K)             | 5.08      | 5.029     | −0.051 | 5.108     | 0.079          |
| A(Ca)            | 6.31      | 6.287     | −0.023 | 6.361     | 0.074          |
| A(Sc)            | 3.17      | 3.167     | −0.003 | 3.12      | −0.047         |
| A(Ti)            | 4.9       | 4.876     | −0.024 | 4.882     | 0.006          |
| A(V)             | 4.0       | 4.124     | 0.124  | 3.849     | −0.275         |
| A(Cr)            | 5.64      | 5.64      | 0.0    | 5.61      | −0.03          |
| A(Mn)            | 5.39      | 5.289     | −0.101 | 5.494     | 0.205          |
| A(Co)            | 4.92      | 5.05      | 0.13   | 4.771     | −0.279         |
| A(Ni)            | 6.23      | 6.228     | −0.002 | 6.236     | 0.008          |
| A(Cu)            | 4.21      | 4.418     | 0.208  | 4.002     | −0.416         |
| A(Zn)            | 4.6       | 4.651     | 0.051  | 4.53      | −0.121         |
| A(Rb)            | 2.6       | 2.6       | 0.0    | –         | –              |
| A(Sr)            | 2.92      | 2.92      | 0.0    | –         | –              |
| A(Y)             | 2.21      | 2.204     | −0.006 | 2.152     | −0.052         |
| A(Zr)            | 2.58      | 2.58      | 0.0    | 2.122     | −0.458         |
| A(Mo)            | 1.92      | 1.92      | 0.0    | –         | –              |
| A(Ru)            | 1.84      | 1.84      | 0.0    | –         | –              |
| A(Ba)            | 2.17      | 2.108     | −0.062 | 2.113     | 0.005          |
| A(La)            | 1.13      | 1.19      | 0.06   | 0.986     | −0.204         |
| A(Ce)            | 1.7       | 1.77      | 0.07   | 1.447     | −0.323         |
| A(Nd)            | 1.45      | 1.328     | −0.122 | 1.276     | −0.052         |
| A(Sm)            | 1.0       | 1.0       | 0.0    | –         | –              |
| A(Eu)            | 0.52      | 0.52      | 0.0    | –         | –              |

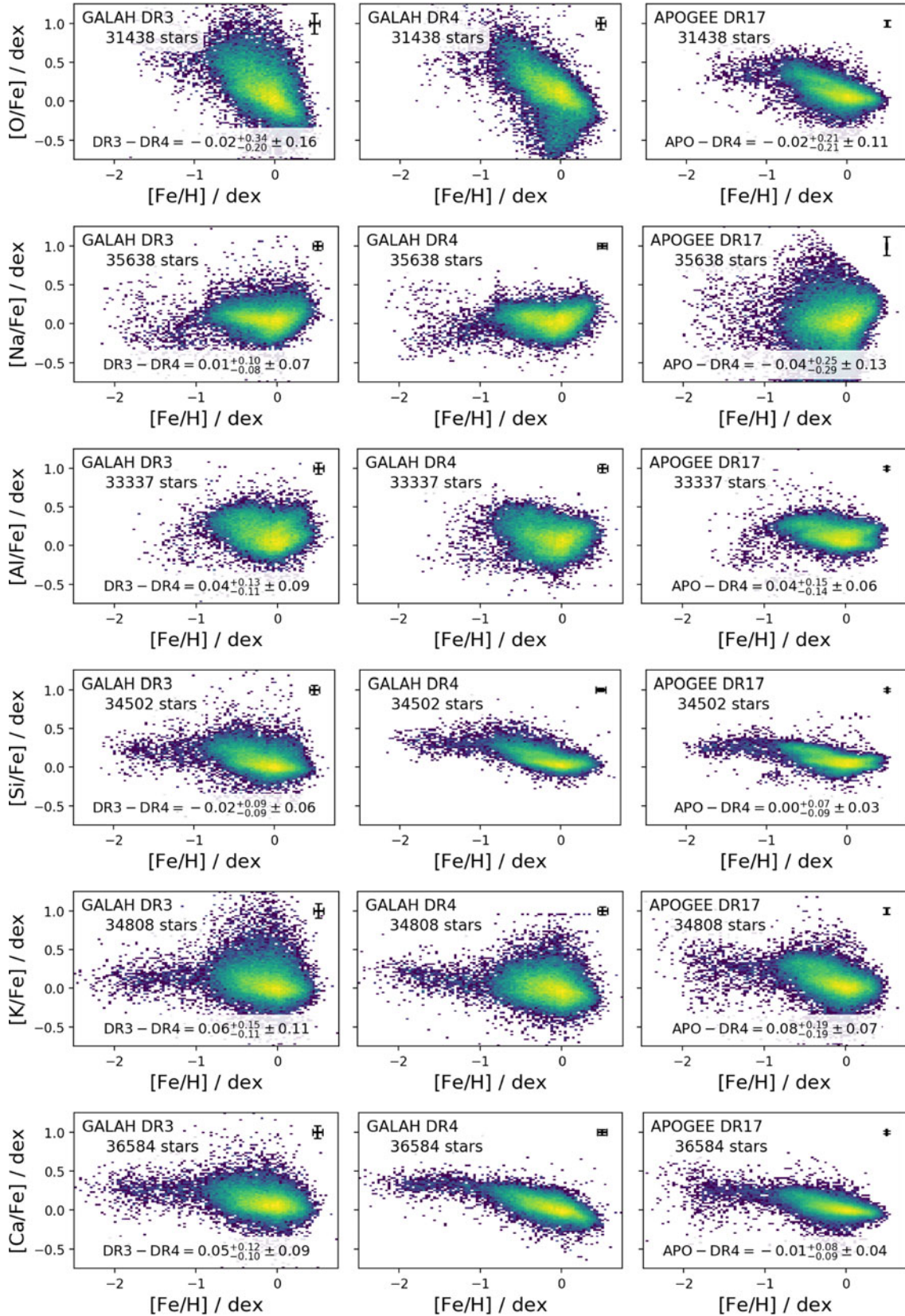




**Figure C1.** zero-point estimates of elemental abundances for GALAH DR4. Each panel shows the comparison to literature (DR4 – literature) for Vesta (blue), *Gaia* FKG Benchmark Stars (orange), Stars with  $|\text{[Fe/H]}| \leq 0.1$  closer than  $D_{\text{tr}} < 0.5$  kpc (red), as well as stars that were also observed by APOGEE DR17 (purple).



**Figure C2.** Precision monitoring (with a median line and standard deviation shading) of elemental abundances as a function of SNR for the green CCD2 across for GALAH DR4. Each panel shows the behaviour for bins of width 10 for the scatter of repeat observations of the allspec runs (blue) as well as covariance uncertainties of allspec (orange) and allstar (red) setups.



**Figure C3.** Comparison of stars with available measurements in GALAH DR3 (left column), GALAH DR4 (middle column) and APOGEE DR17 (right) for O, Na, Al, Si, K, and Ca.



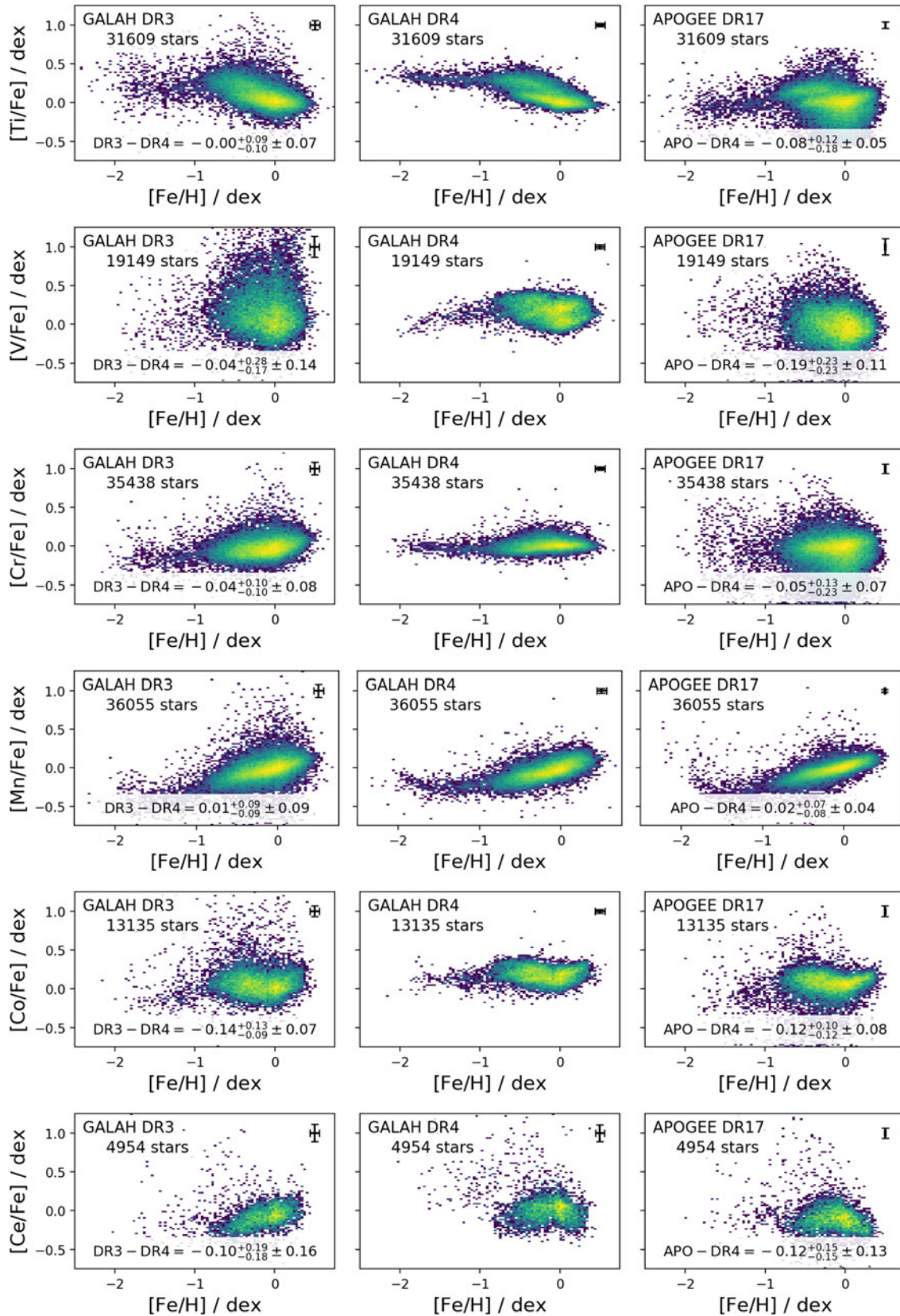
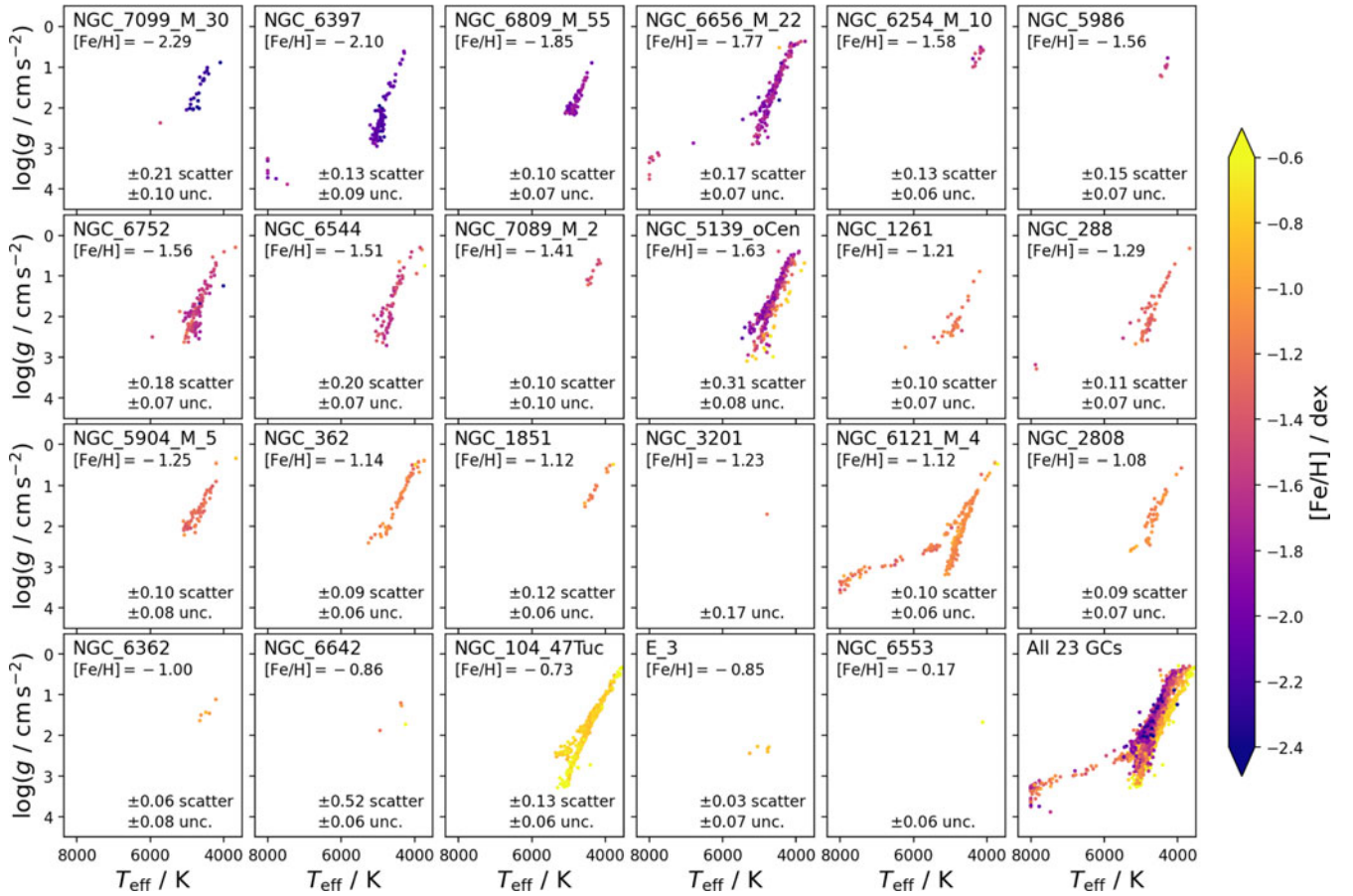
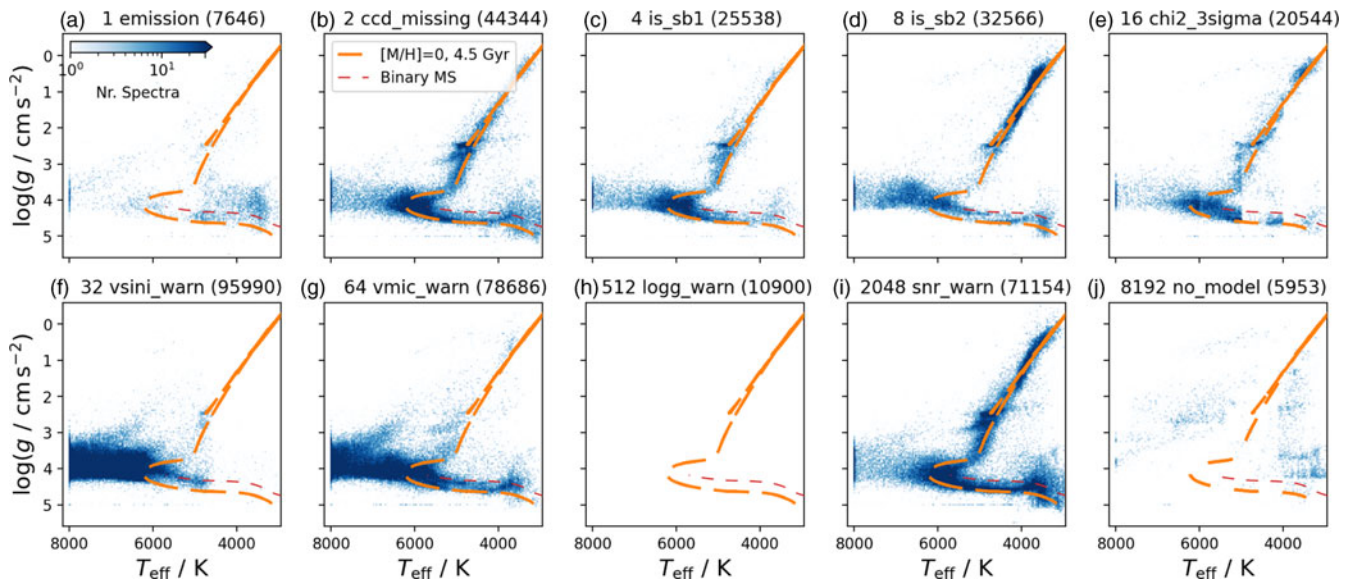


Figure C4. Continuation of Fig. C3 for Ti, V, Cr, Mn, Co, and Ce.





**Figure C5.** Collage of globular clusters in the  $T_{\text{eff}}\text{-}\log g$  space, coloured by stellar metallicity  $[\text{Fe}/\text{H}]$ . There are only minor trends between  $[\text{Fe}/\text{H}]$  and  $T_{\text{eff}}$ , even for the horizontal branch stars in NGC 288, NGC 6656 (M22), and NGC 6121 (M4). NGC 5139 ( $\omega$ Cen) shows a significant range in  $[\text{Fe}/\text{H}]$ . RMS scatter and median metallicity uncertainties for each cluster are given in the lower right of each panel.



**Figure C6.** Parameter overview of stars with raised major quality flag `flag_sp` for `allstar`. Each panel shows the logarithmic density distribution of stars in the  $T_{\text{eff}}$  and  $\log g$  plane with blue colourmaps. A PARSEC isochrone with  $[\text{M}/\text{H}] = 0$  and  $\tau = 4.5$  Gyr is overplotted in orange and the same mass binary main-sequence (shifted from the single star one by  $\Delta \log g = -0.3$  dex) is shown in red. Panel heads denote the bit mask and its description as well as how many times the flag was raised. We neglect distributions with no flag (0), for flags which have not been raised (8,9,11), and for which no results were available (15).