

## RESEARCH PAPER

# The GALAH Survey: Data Release 4

S. Buder,<sup>1,2</sup> J. Kos,<sup>3</sup> E. X. Wang,<sup>1,2</sup> M. McKenzie,<sup>1,2</sup> M. Howell,<sup>4,2</sup> S. L. Martell,<sup>5,2,6</sup> M. R. Hayden,<sup>7,2</sup> D. B. Zucker,<sup>8,9,2</sup> T. Nordlander,<sup>1,2</sup> B. T. Montet,<sup>5,2,6</sup> G. Traven,<sup>3</sup> J. Bland-Hawthorn,<sup>7,2</sup> G. M. De Silva,<sup>9</sup> K. C. Freeman,<sup>1,2</sup> G. F. Lewis,<sup>7</sup> K. Lind,<sup>10</sup> S. Sharma,<sup>11</sup> J. D. Simpson,<sup>5,2</sup> D. Stello,<sup>5,7,2,12</sup> T. Zwitter,<sup>3</sup> A. M. Amarsi,<sup>13</sup> K. Banks,<sup>5,2</sup> K. Beeson,<sup>3</sup> B. Chen,<sup>1,2</sup> I. Ciucă,<sup>1,2</sup> G. S. Da Costa,<sup>1,2</sup> R. de Grijs,<sup>14,15,16</sup> B. Martin,<sup>1</sup> D. M. Nataf,<sup>17</sup> M. K. Ness,<sup>1,2</sup> A. Rains,<sup>18</sup> T. Scarr,<sup>1</sup> R. Vogrinčič,<sup>3</sup> Z. Wang,<sup>7,2,19</sup> Y. Xie,<sup>1,2</sup> and The GALAH Collaboration<sup>20</sup>

<sup>1</sup>Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

<sup>2</sup>ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

<sup>3</sup>Faculty of Mathematics & Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

<sup>4</sup>School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia

<sup>5</sup>School of Physics, UNSW, Sydney, NSW 2052, Australia

<sup>6</sup>UNSW Data Science Hub, University of New South Wales, Sydney, NSW 2052, Australia

<sup>7</sup>Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006, Australia

<sup>8</sup>School of Mathematical and Physical Sciences, Macquarie University, Sydney, NSW 2109, Australia

<sup>9</sup>Macquarie University Research Centre for Astronomy, Astrophysics and Astrophotonics, Sydney, NSW 2109, Australia

<sup>10</sup>Department of Astronomy, Stockholm University, AlbaNova University Centre, SE-106 91 Stockholm, Sweden

<sup>11</sup>Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD, 21218, USA

<sup>12</sup>Stellar Astrophysics Centre, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

<sup>13</sup>Theoretical Astrophysics, Department of Physics and Astronomy, Uppsala University, Box 516, 751 20 Uppsala, Sweden

<sup>14</sup>School of Mathematical and Physical Sciences, Macquarie University, Balaclava Road, Sydney, NSW 2109, Australia

<sup>15</sup>Astrophysics and Space Technologies Research Centre, Macquarie University, Balaclava Road, Sydney, NSW 2109, Australia

<sup>16</sup>International Space Science Institute-Beijing, 1 Nanertiao, Zhongguancun, Beijing 100190, China

<sup>17</sup>Department of Physics & Astronomy, University of Iowa, Iowa City, IA 52242, USA

<sup>18</sup>Department of Physics and Astronomy, Uppsala University, Box 516, SE-75120 Uppsala, Sweden

<sup>19</sup>Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA

<sup>20</sup><https://www.galah-survey.org>

**Author for correspondence:** S. Buder, Email: sven.buder@anu.edu.au.

## Abstract

The stars of the Milky Way carry the chemical history of our Galaxy within 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 31 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 (CNO) and more measurements of rare earth elements critical to modern life electronics, offering unparalleled insights into the chemical composition of the Milky Way.

For this data release, we employ neural networks to simultaneously fit stellar parameters and chemical abundances across the full spectrum, leveraging synthetic grids computed with Spectroscopy Made Easy. The grids include atomic line formation in non-local thermodynamic equilibrium (NLTE) for 14 elements. In a two-iteration process, we first fit stellar labels for all 1 085 520 spectra, then co-add repeat spectra and refine these labels using astrometric data from *Gaia* and 2MASS photometry to deliver improved accuracy and precision. Our validation thoroughly assesses the reliability of these measurements and highlights key caveats for catalogue users.

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

## 1. INTRODUCTION AND WORKFLOW

### 1.1 Motivation

The history of our Milky Way galaxy is written in starlight. By capturing and analysing the light of millions of stars that are now billions of years old, we can learn about the chemical composition of the material that was locked into their stellar atmospheres at birth. We can use stars as time capsules into the past evolution of the Milky Way. The light of stars can thus guide us to explore and map our environment, just as it

has guided Aboriginal and Torres Strait Islander people and their astronomers on Country 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), or RAVE (Steinmetz et al., 2020) 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 changes of chemical compositions. By combining stellar ages, orbits, and chemistry, we have made major advances in our 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 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. 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 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. By using neural networks, we can model the entire wavelength range, including broad molecular absorption features from C<sub>2</sub> and CN, rather than focusing on narrow atomic line windows. This approach allows us to model all stellar labels—global parameters and elemental abundances—simultaneously. Additionally, we can derive interstellar properties from the differences between observed and synthetic spectra, while also incorporating non-spectroscopic information during the optimization process.

We outline our workflow in the following section and provide detailed explanations throughout this manuscript, hoping that it will prove useful for upcoming surveys like 4MOST (de Jong et al., 2019) and WEAVE (Dalton et al., 2014).

## 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 Sec. 2, most notably the observation of HERMES spectra. We explain how we create synthetic stellar spectra to compare with the observed ones in Sec. 3. This comparison is done in two consecutive steps. In Sec. 4, we explain how we extract stellar labels from individual observations (without non-spectroscopic information folded into the optimization), while Sec. 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 Sec. 6. The data products of this data release are explained in Sec. 7. We describe identified caveats in Sec. 8 and make suggestions for minimising them in the future, before concluding this manuscript with thought in Sec. 9.

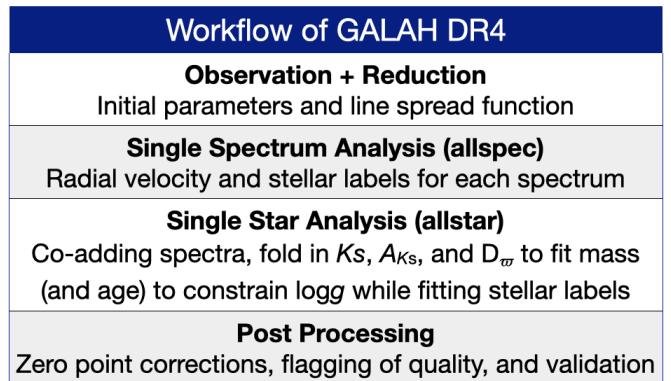


Figure 1. Workflow of GALAH DR4.

## 2. DATA

The GALAH Survey uses the 3.9-metre 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 entrances on one of two metal field plates, which can be tumbled to allow observing of 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 et al., 2010; Brzeski et al., 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  $\sim 1000 \text{ \AA}$  in the range of 4713 – 4903 (blue CCD or CCD1), 5648 – 5873 (green / CCD2), 6478 – 6737 (red / CCD3), and 7585 – 7887  $\text{\AA}$  (infrared IR / CCD4). The data used in this data release is primarily based on observations of stars with said 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 and observed (Sec. 2.1) with the 2dF-HERMES setup, including the first description of the second phase of GALAH observations (GALAH Phase 2) with a sharper focus on main-sequence turnoff stars to estimate more precise ages. In Sec. 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 Sec. 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 (Sharma et al., 2019) and TESS missions (Sharma et al., 2018), as well as hand-picked and random observations as part of other programs. 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 halo stars (PI S. Buder) in addition to their by-chance observation through the main surveys. The column `survey_name` in our catalogues denotes this 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, we used the 2MASS photometric survey (Skrutskie et al., 2006) with its  $J$  and  $K_s$  filters as a parent sample from which we selected stars within estimated visual magnitudes

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

For GALAH Phase 1, a tiling pattern (with unique `field_id` entries) with 2 deg fields of view below declination  $\delta < +10$  deg were created for regions with Galactic latitude  $|b| > 10$  deg to avoid crowding and strong extinction. For each tile, a selection of 400 stars within magnitudes  $9 < V_{JK} < 12$  for a bright magnitude cut and  $12 < V_{JK} < 14$  for the nominal magnitude cut is randomly selected from the complete parent sample of 2MASS. This typically selects 2/3 main sequence and turnoff 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 magnitude between  $2 < M_G < 4$ , where

$$M_G = \text{phot\_g\_mean\_mag} + 5 \cdot \log_{10} \left( \frac{\text{parallax}}{100 \text{ mas}} \right) \quad (2)$$

with  $G$  magnitude and parallax measurements from *Gaia* DR2 (Gaia Collaboration et al., 2018; Evans et al., 2018; Lindegren et al., 2018). Remaining fibres are again filled with targets as done in Phase 1.

#### 2.1.2 Observational setup

We list the observations under various sub-programs in Table 1. Except for 2935 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 minutes. Under sufficient conditions (no clouds and seeing below 2 arcsec), GALAH Phase 1 and TESS-HERMES observed  $3 \times 6$  minutes for bright targets ( $9 < V_{JK} < 12$ ) and  $3 \times 20$  minutes for the majority of targets ( $12 < V_{JK} < 14$ ).

**Table 1.** Overview of stars observed for the programs included in this data release. Numbers of open and globular cluster observations were estimated after observations as described in Sec. 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

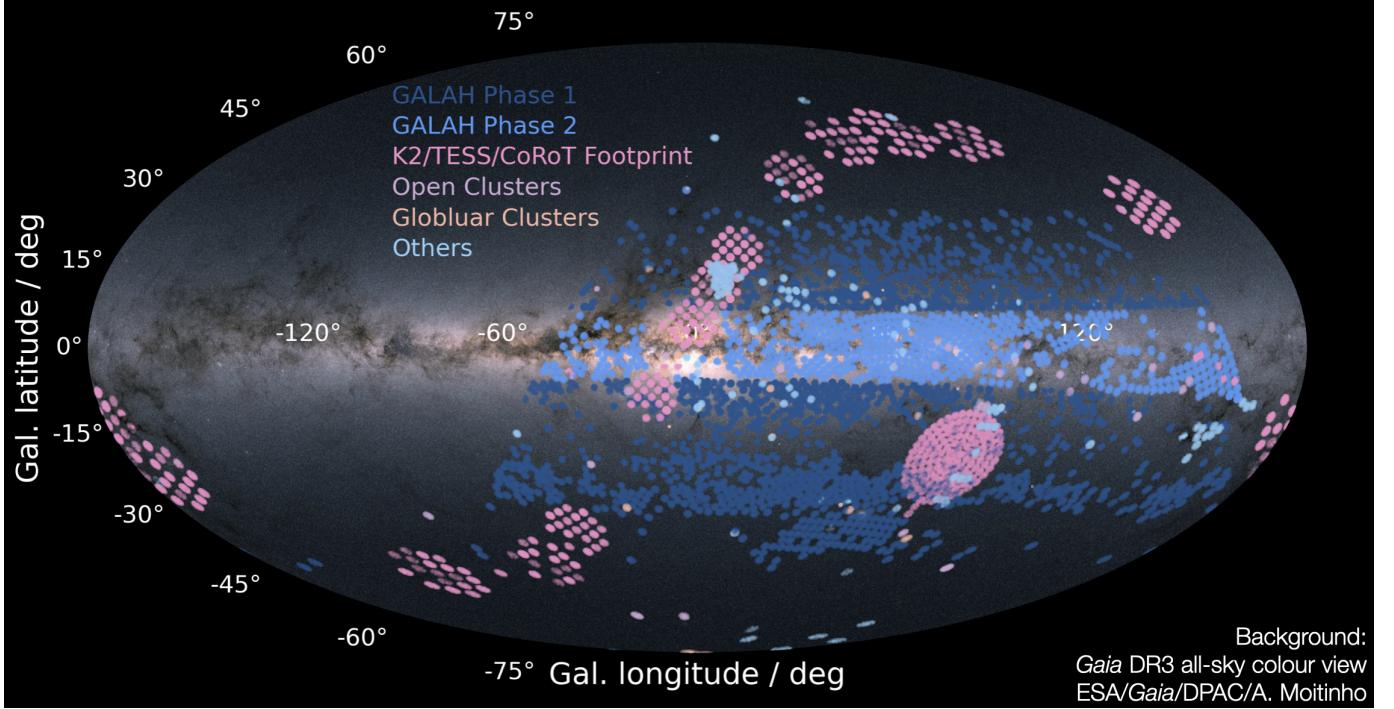
**Table 2.** Data product 1: FITS files of reduced spectra.

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
Ext. 7	Resolution profile / FWHM

GALAH Phase 2 extended these times to  $3 \times 10$  and  $3 \times 30$  minutes, respectively, and included repeat observations of GALAH Phase 1 main targets with another  $3 \times 15$  minutes. K2-HERMES observations targeted stars with  $13 < V_{JK} < 15$  or even  $13 < V_{JK} < 15.8$  to complement the K2 Galactic Archaeology Program (Stello et al., 2015). These fields were observed for 2 hours, similar to most globular and open cluster stars. Worse seeing conditions or thin clouds triggered between one ( $2 < \text{seeing} < 2.5$  arcsec) and 3 ( $2.5 < \text{seeing} < 3$  arcsec) additional exposures. In addition to the science frames, quartz fibre flat and ThXe arc lamp 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. A reduced example spectrum of the asteroid 4 Vesta (used as a reference for a Solar spectrum) is shown in Fig. 3. The reduction process to create FITS files of reduced spectra from two-dimensional images of the cameras employs an updated and publicly available version 6 of the



**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 sky.

already well-tested reduction pipeline (Kos et al., 2017). The file extensions are listed in Tab. 2 and created as follows.

Science frames are corrected by removing the bias, dividing out the different gain of the two readout amplifiers per CCD, flagging bad pixels, and dividing by master flatfield 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 ThXe arc frames and the spectra are interpolated onto a linearly increasing wavelength grid. The starting wavelength CRVAL1 and dispersion CDELT1 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 1 of the reduction pipeline FITS files and used for the subsequent analysis.

Fractional noise/uncertainties are saved in extension 2 and calculated from the square root of the sum of squared flux, 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 a boxiness parameter  $b$  and full width half maximum  $\text{fwhm}$  for each wavelength point in the spectrum, that is

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

The array  $x$  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  $\text{fwhm}$  are saved in extension 7 with  $b$  saved in the headers.

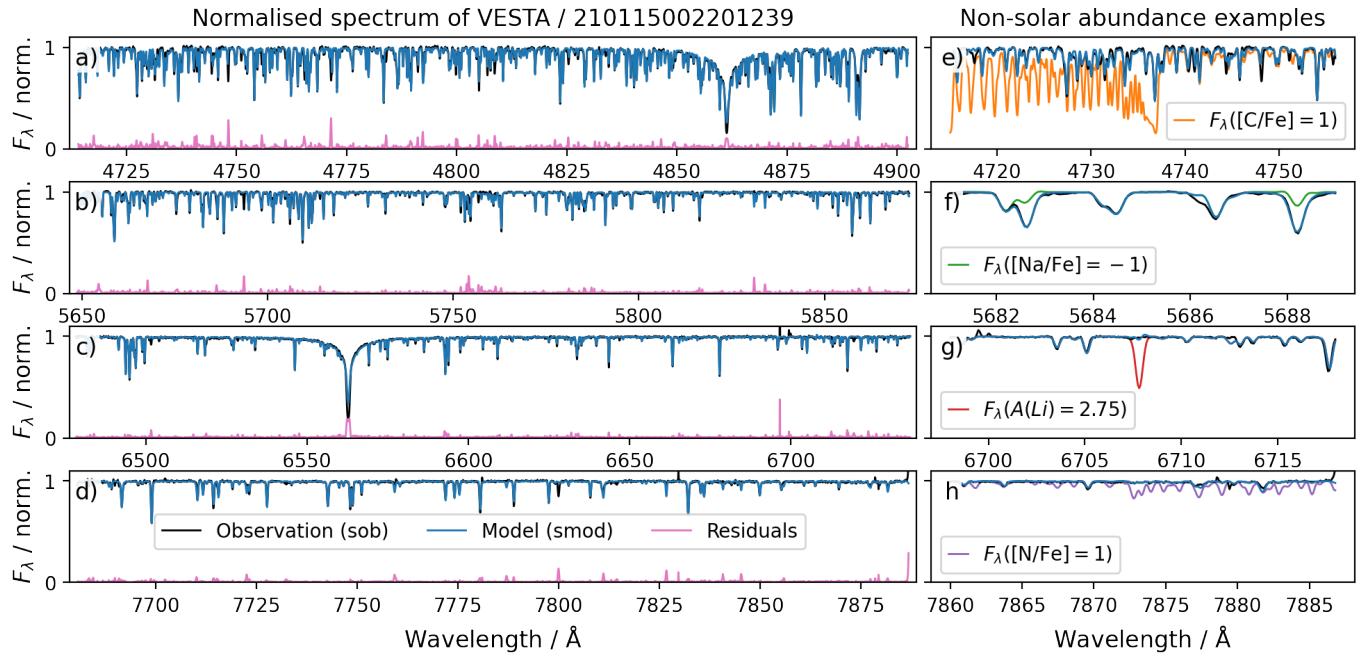
The achieved Signal-to-Noise Ratio (SNR) of the individual exposures depends on the spectral types, reddening, and observational conditions. Especially the repeat observations of previous pointings have increased the SNR for co-added spectra with respect to GALAH DR3. This can be appreciated from Fig. 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

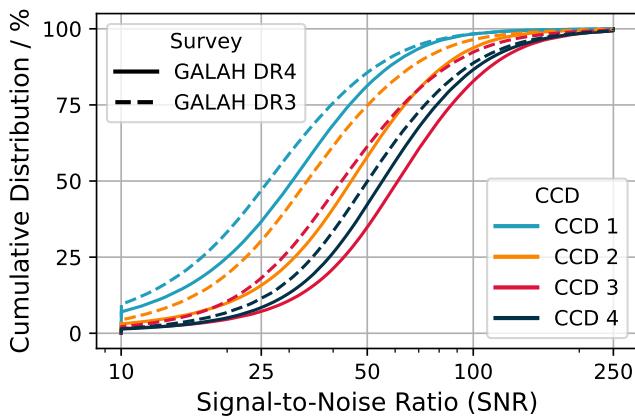
To support our spectroscopic analysis, we make use of astrometric and photometric information from the *Gaia* satellite (Gaia Collaboration et al., 2016) and 2MASS survey (Skrutskie et al., 2006), which is available for essentially all our targets. We further use the value-added catalogues, like distance 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*



**Figure 3.** Comparison of normalised observed (black) and synthetic spectra (blue) of the asteroid 4 Vesta with solar composition as well as examples of synthetic spectra with non-solar abundances. Panels a-d) show the observed and best-fit 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<sub>2</sub> 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 (purple).



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

DR3 (Torra et al., 2021). 911 754 (99.0 %) also have astrometric information (Lindegren 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 (Lindegren et al., 2021a) information. Where possible we prefer the photogeometric 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  $\sigma_{\omega}/\omega = 1.6^{+2.0}_{-0.9}\%$ . Only 2.3 % of GALAH stars have no parallax measurements 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 RJCE method (Majewski et al., 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 color extinction of  $B - V$ , that is,  $A_{K_S} \sim 0.36 \cdot E(B - V)$  (Cardelli et al., 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 aim of our spectroscopic analysis is to estimate the best set of stellar properties (labels) that influence a stellar spectrum by minimising the difference between observed stellar spectra and synthetic ones that were created with those best stellar labels. In our endeavour to push the limits even further, we are advancing our analysis to fit all 31 elemental abundances and stellar parameters across the full GALAH wavelength range simultaneously with the appropriate model spectrum.

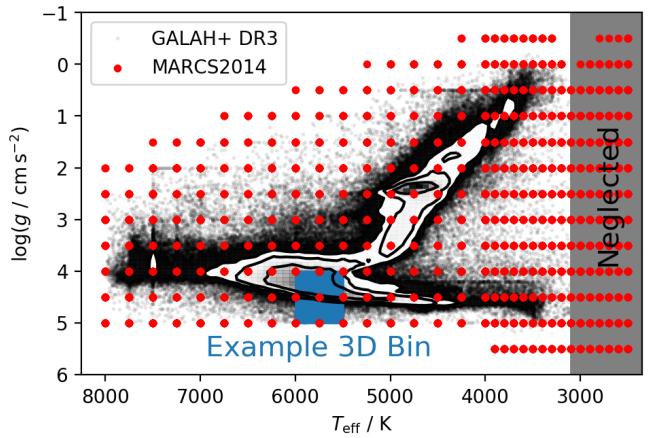
To make this computationally feasible, we follow an idea reported by Rix et al. (2016) and create flexible models for smaller regions of the parameter space from only a limited number of *ab initio* models (see also Ting et al., 2016). Our *ab initio* models are calculated with Spectroscopy Made Easy (SME Valenti & Piskunov, 1996; Piskunov & Valenti, 2017) for the whole wavelength range and all visible atomic and molecular lines for random selections of elemental abundances and stellar parameters within the range of MARCS atmospheres (Gustafsson et al., 2008) at much higher resolution and sampling than our HERMES spectra. We then select subsets of these spectra within a restricted space of the three main spectroscopic parameters  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . This idea is comparable to selecting Solar or stellar twin spectra when analysing the Sun (Nissen, 2015, see e.g.) or differential abundance analysis of globular cluster stars (Yong et al., 2013; Monty et al., 2023, e.g.). As they reduce the impact of several systematic issues of line data and atmospheric effect, these approaches have been highly successful (Nissen & Gustafsson, 2018). For each subset, we train a neural network that correlates stellar flux and labels (stellar parameters and abundances) similar to *The Payne* (Ting et al., 2019). With these models, we can then create model spectra with all lines over the whole wavelength range for any combination of element abundances within this restricted parameter space within less than a second (as compared to minutes or hours for physics-driven syntheses).

Another reason to create smaller training sets rather than a "one-fit-all" approach was the limited flexibility of both quadratic and non-linear interpolation routines within computationally reasonable model sizes and ability to test the model accuracy over the full parameter space. Surveys like GALAH, RAVE and APOGEE aim to fit basically all types of spectra at once. This means we demand one model to predict Sun-like stars, red clump stars, metal-poor stars with almost no absorption features, cool evolved stars with strong molecular features, and hot stars with shallow and broad lines – all at once and for up to 31 elemental abundances. As one can imagine, this is an impossible task and has led to numerous systematic trends in catalogues for the most extreme cases that use such model interpolations (Casey et al., 2016; Buder et al., 2018;

Ting et al., 2019). In this data release, we therefore purposefully limit the spectral complexity by creating smaller models. The hot star model therefore does not need to also predict the strong molecular absorption features of a cool star. We discuss the possible caveats and disadvantages of this approach and our particular implementation in Sec. 8.

In this section, we lay out how we create smaller bins in the parameter space from which we sample a training set (Sec. 3.2) rather than using one training set for all stars. We describe how we create the parent sample of high-resolution synthetic stellar spectra (Sec. 3.2) and train neural networks to quickly predict/interpolate new synthetic spectra (Sec. 3.3).

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



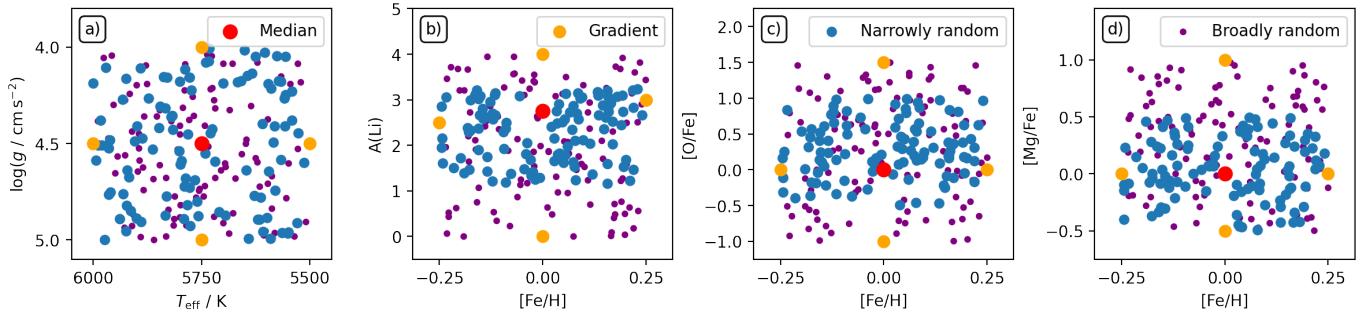
**Figure 5. Coverage in  $T_{\text{eff}}$  and  $\log g$  of the MARCS2014 grid (red) and GALAH DR3 (black, including density contours).** Shown is also an example of one of the 3D bins used to create models with *The Cannon*. MARCS grid points  $T_{\text{eff}} < 3100$  K or  $[\text{Fe}/\text{H}] < -3$  dex are neglected throughout GALAH DR4.

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  $\pm 100$  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  $T_{\text{eff}} = 5750 \pm 250$  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<sup>a</sup> synthetic spectra with no rotational broadening, which are later broadened with different rotational velocities  $v \sin i$  to create between 1680 and 2240 training set spectra for each bin. For clarity, we explain the parameter and abundance sampling for an example 3D bin centred on  $T_{\text{eff}} = 5750 \pm 250$  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 30 elements are randomly sampled within reasonable limits (see examples in Tab. 3) and fed into SME to create self-consistent synthetic spectra over the full HERMES wavelength range for MARCS atmospheres.

<sup>a</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 to test the gradient change of spectra with individual labels.

**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	$A(\text{Li})$	1.05..2.75..3.26	$A(\text{Li})$	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 Eq. 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 Eq. 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

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

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

### 3.2 Higher-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 over-sampled synthetic intensity spectra at ten times higher resolution than the typical GALAH resolution with SME for seven equal-area angles of a stellar surface (see Fig. 7).

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 by Heiter et al. (2021) and implements numerous updates to line data, such as updates or corrections of  $\log g/f$  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.

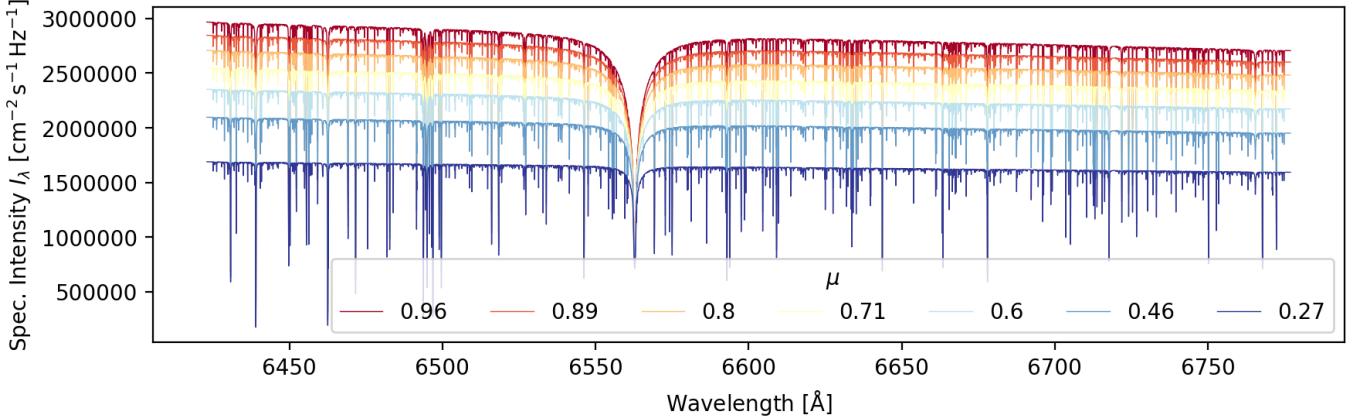
Spectra are computed at a resolution of  $R = 300,000$  on a fine wavelength grid with 60,819 pixels spread over the extended wavelengths 4675.1–4949.9, 5624.1–5900.9, 6424.1–6775.9, and 7549.1 – 7925.9 Å. We note that these extend significantly beyond the actual GALAH wavelength range.

We use 1D MARCS atmospheres from the MARCS grid (Gustafsson et al., 2008, version 2014) and interpolate them for combinations of  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . We use grids of non-LTE departure coefficients by Amarsi et al. (2020b) for atomic lines of H, Li, C, N, O, Na, Mg, Al, Si, K, Ca, Mn, and Ba. For all except C, we use grids that include background scattering.

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 polynomial interpolation 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.cmmod`) and the specific intensities (`sme.sint`) as a function of the equal-area midpoints of each equal-area annulus  $\mu$  (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, 2021) of SME (Piskunov & Valenti, 2017). Depending on the expected rotational velocities (increasing



**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 equal-area midpoints of each equal-area annulus  $\mu$ .

with temperature) we sample a range of

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

Note that  $\nu \sin i = 24 \text{ km s}^{-1}$  is only included for bins with  $T_{\text{eff}} \geq 5000 \text{ K}$  and  $\nu \sin i = 36 \text{ km s}^{-1}$  for those with  $T_{\text{eff}} \geq 6000 \text{ K}$ .

### 3.3 Interpolating synthetic spectra with neural networks

To allow a fast interpolation with new and different stellar labels, we use the method of training descriptive 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, 2016; Casey et al., 2016, 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, 2021). Because of the needed flexibility<sup>b</sup> to predict synthetic spectra with 36 stellar labels for a large parameter space, we are also choosing neural networks to interpolate between our synthetic spectra in this data release.

In this work, we utilize 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 stellar labels  $\ell$  with the flux  $f$  at each wavelength pixel  $\lambda$  via

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

which encapsulates the so-called layers of a neural network with  $i = 300$  neurons and where we use the default leaky

<sup>b</sup>For a more detailed discussion on the advantages of neural networks for predicting spectra see Ting et al. (2019).

Rectified Linear Unit (lReLU)

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

After optimizing the loss function for  $10^4$  steps, we consider the network trained with an optimized combination of three sets of weights and biases within the minimum and maximum ranges of each label. We discuss the caveats of this particular neural network architecture and training setup in Section 8. The trained networks can then be used with new input labels to quickly create synthetic spectra for the label optimization.

## 4. SINGLE SPECTRUM ANALYSIS (ALLSPEC)

As outlined in Sec. 1, the workflow of GALAH DR4 includes a first analysis step of all observed spectra without taking non-spectroscopic information for the optimization. This allows us to identify shifts in radial velocity between separate spectroscopic observations of the same star<sup>c</sup> and a more accurate co-adding of spectra for the allstar analysis (see Sec. 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 optimization of stellar labels is thus aiming to minimise the absolute difference between synthetic and observed spectra, weighted by their uncertainty. Starting from a set of initial labels (Sec. 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, we used a repeated combination of spectrum normalisation followed by stellar parameter optimization 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

<sup>c</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.

abundances. This allows a more accurate comparison of synthetic and observed spectra (Sec. 4.2) and thus a more accurate stellar label optimization (see Sec. 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 (Sec. 2.2). Because of the limited accuracy of the latter for cool stars with  $T_{\text{eff}} < 4000$  K as well as the hottest stars with  $T_{\text{eff}} > 6500$  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., 2022; Fouesneau et al., 2022) as initial values.

In selected cases, we further adjust the starting parameters toward reasonable limits, for example for hot stars which are likely to be young and close to Solar metallicity. Furthermore, we recalculate the initial  $v_{\text{mic}}$  based on Eq. 4 and limit rotational broadening values to  $3 \leq v \sin i \leq 10 \text{ km s}^{-1}$  for stars below  $T_{\text{eff}} = 5500$  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 and are depicted in Fig. 29.

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  $-0.4 \times [\text{Fe}/\text{H}]$  for  $-1 \leq [\text{Fe}/\text{H}] \leq 0$ . All other abundances are initialised at  $[\text{X}/\text{Fe}] = 0$ .

#### 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 Sec. 3.3 and the observational data at actually measured resolution and sampling (presented in Sec. 2.2).

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

Because dedicated line-spread-function measurements are available for every spectrum (see Sec. 2.2), we use this information to downgrade our synthetic spectrum to the measured resolution of each observation. We then interpolate the over-sampled 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 spectrum and re-normalise our observed spectra and their uncertainties via this normalisation function.

This specific approach is similar to the internal routine of SME (Piskunov & Valenti, 2017) 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 (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  $\sigma_{f,\lambda}$ .

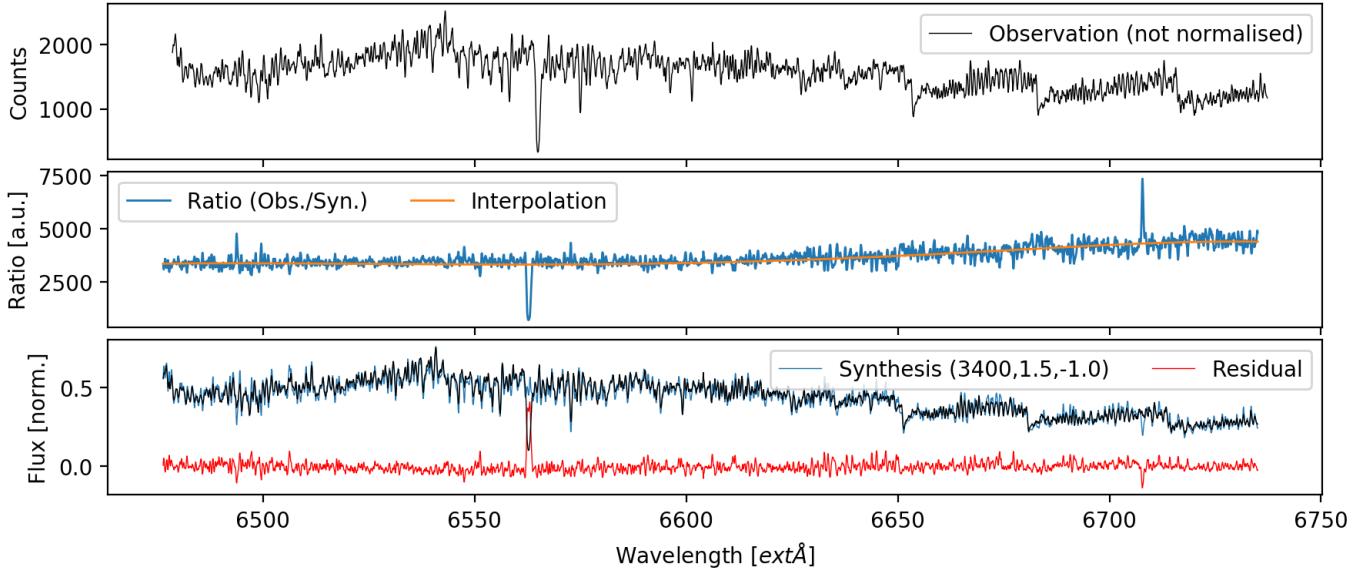
As we compare the observation to model spectra, we do not have to restrict ourselves to an *a priori* normalisation, but can take into account the residual information on the continuum in parts of the spectrum. For each model spectrum that we compare to, we therefore perform a normalisation by fitting a 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  $T_{\text{eff}}$  as well as  $v_{\text{rad}}$ .

#### 4.3 Stellar label optimization

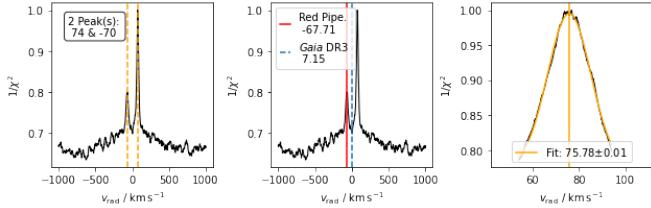
In up to four major loops, we optimize the radial velocities and all other stellar labels and report a) their values, b) their co-variances, c) the best fit synthetic and re-normalised spectra along with d) their uncertainties and e) masks that indicate which pixels were used in the final optimization.

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  $v_{\text{rad}} = -1000..(2)..1000 \text{ km s}^{-1}$  (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 optimization 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



**Figure 8.** Example of normalisation for GALAH DR4 for a model spectrum ( $T_{\text{eff}} = 3400$  K,  $\log g = 1.5$ ,  $[\text{Fe}/\text{H}] = -1.0$ ) that is selected during the label optimization. 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 J06084657-7815235, 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 or  $v_{\text{rad}}$ . Panel c) shows the narrow window of  $-20.00..(0.04)..20.00 \text{ km s}^{-1}$  around the highest peak and the Gaussian fit (yellow) to it.

the least squares optimization within less than  $10^4$  iterations and a desired relative error (`xtol`) below  $10^{-4}$ .

For each optimization loop, a new, best-fit 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 deleted or initialised with  $[\text{X}/\text{Fe}] = 0$ .

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

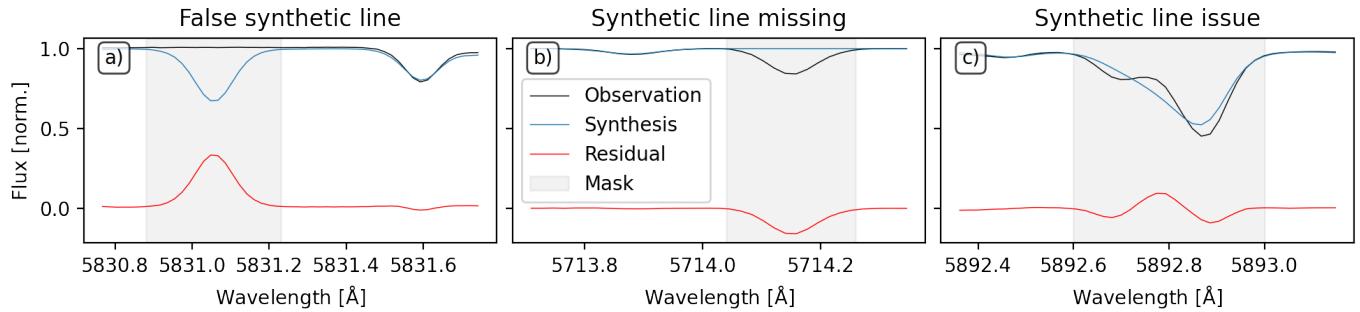
We measured the time taken for the individual steps in the `CURVE_FIT` function's execution to be approximately 80 ms. The total fitting process for stellar labels, including input/output overheads, was timed at  $89^{+77}_{-29}$  s for the `allspec` module, and

$125^{+81}_{-33}$  s for the more complex `allstar` module.

#### 4.3.1 Which stellar labels are optimized?

As part of the synthetic grid computations, we have perturbed each label of stellar parameters and elemental abundances individually to our chosen maximum and minimum ranges (see Sec. 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: Does the spectrum between minimum and maximum label value at any pixel change more than a certain threshold (0.07 of the normalised spectrum)? Does the spectrum between minimum and maximum label value change by more than 0.005 of the normalised spectrum for at least 25% of the spectrum? While the first case is constructed for atomic lines, such as  $\text{Li I } 6708 \text{ \AA}$ , the second case is addressing in particular molecular lines like the  $\text{C}_2$  and  $\text{CN}$  lines. For stars that are missing data from the infrared arm of HERMES (CCD4) we fix the abundances of N, O, L, and Rb to the scaled-Solar values.

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 (Sec. 4.3.2) at the beginning of each optimization loop. To minimise this effect, we therefore shift the interim abundance values towards the narrow label boundaries. In practise, we limit the initial and interim abundances to  $1.05..3.26$  for  $\text{A}(\text{Li})$ ,  $[\text{X}/\text{Fe}] = -0.5..1.0$  for C, N, O, Y, Ba, La, Ce, and Nd, and  $[\text{X}/\text{Fe}] = -0.5..0.5$  for all other elements before optimizing them again. For warm and hot stars ( $T_{\text{eff}} > 6000$  K), this effect was seen to affect multiple



**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.

abundances, such that we needed to implement a zeroing of all abundances except Li for stars above 6000 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 by Hinkle et al. (2000) with the flux that would be predicted through our pipeline for a star with Solar labels ( $T_{\text{eff}} = 5772$  K,  $\log g = 4.438$  dex,  $[\text{Fe}/\text{H}] = 0.00$  dex,  $\nu_{\text{mic}} = 1.06 \text{ km s}^{-1}$ ,  $\nu \sin i = 1.6 \text{ km s}^{-1}$ ,  $\nu_{\text{mac}} = 4.2 \text{ km s}^{-1}$  (Prša et al., 2016; Jofré et al., 2017), and  $[\text{X}/\text{Fe}] = 0.00$  dex for the default Solar abundance pattern for MARCS by Grevesse et al. (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>d</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 spectrum differ by more than  $5\sigma$  or a flux of 0.3 (0.4 before the initial optimization). To avoid the masking of lines that are vital for our spectroscopic analysis, we have created a list<sup>e</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 optimization 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.

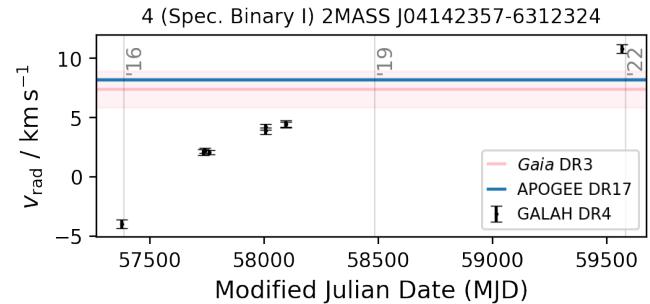
In addition to this default masking, we exclude pixels for

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). For presumably cool stars (with initial  $T_{\text{eff}} < 4100$  K), we therefore double the observational uncertainty of the blue arm.

## 5. SINGLE STAR ANALYSIS (ALLSTAR)

After the `allspec` module (Sec. 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. The optimization 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.



**Figure 11.** Example of radial velocity evolution over time for a single-lined spectroscopic binary (SBI).

To speed up computation, we use the mean results of the

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

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

`allspec` analyses as initial stellar labels for the `allstar` analysis. All other methodology of the comparison of synthetic spectra to observations (Sec. 4.2) and label optimization (Sec. 4.3) apply also to this module, with the exception of the optimization 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 Eq. 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, 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 by the distance modulus, bolometric correction, and extinction. Because essentially all stars in GALAH DR4 have outstanding 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

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

While the values for  $K_S$ ,  $D_{\infty}$ , and  $A(K_S)$  are readily available (see Sec. 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 by Casagrande & VandenBerg (2018).

Because of the appearance of  $\log g$  in Eq. 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 an earlier version of the ELLI code by Lin et al. (2018) for a likelihood-weighted estimate with default uncertainties of 100 K, 0.25 dex, 0.2 dex, and an average uncertainty of  $L_{\text{bol},\odot}$  from propagated uncertainties of Eq. 9.

We interpolate over the default tables of 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  $[\alpha/\text{Fe}]$  enhancement<sup>f</sup> via the correlation by 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 Secs. 4 and 5, respectively), a post-processing routine is used to estimate additional parameters from the residuals of the spectra (Sec. 6.1), estimate and validate accuracy and precision uncertainties (Sec. 6.2), and perform quality assurance tests on a global scale (`flag_sp`, see Sec. 6.3) as well as for the individual abundances of elements X (`flag_X_fe`, see Sec. 6.4).

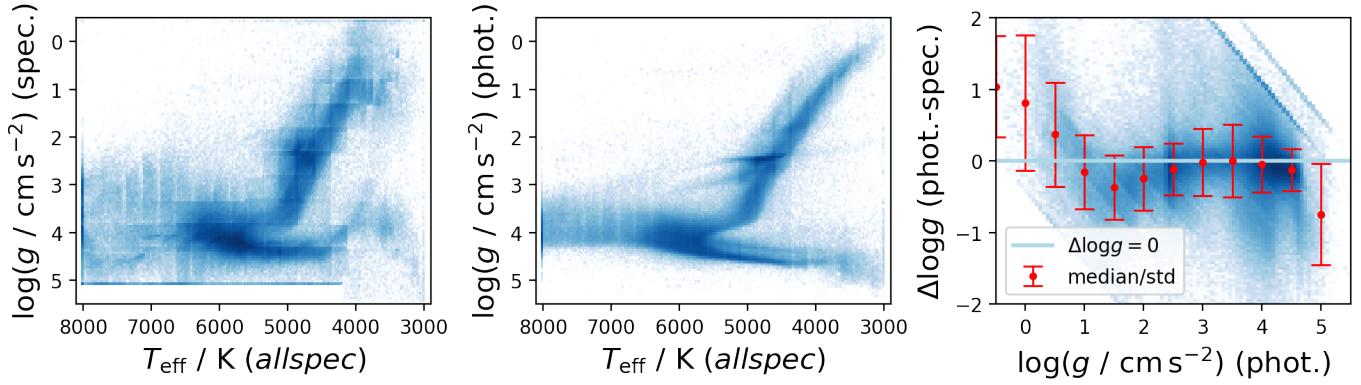
### 6.1 Analysis of spectral residuals

#### 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 ( $-70 \text{ km s}^{-1}$  in addition to  $74 \text{ km s}^{-1}$ ) 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 4890 and 4891 Å, Ni at 6644 Å) as well as those with the largest expected wavelength shift in the infrared detector (O triplet at 7772 – 7775 Å as well as Mg at 7692 Å). If we find several peaks with a reasonably similar radial velocity, the likely  $X \in 16, 50, 84^{\text{th}}$  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 our stars, we find a difference with respect to *Gaia* DR3 larger than  $10 \text{ km s}^{-1}$ . For these stars, as well as stars for which we estimate unrealistic  $v_{\text{mic}}$  and  $v \sin i$  below  $0 \text{ km s}^{-1}$  in `allspec`. We note that the `allspec` analysis was

<sup>f</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.

run without boundary conditions for global parameters and thus also resulted in negative velocities, which are later flagged. 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 Sec. 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 Figs. 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 Sec. 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 at least for the high quality spectra (where  $\log g$  precisions are below the single to binary system offset of  $\Delta \log g = 0.3 \text{ dex}$ ), as discussed in Sec. 6.3.

### 6.1.3 Interstellar absorption

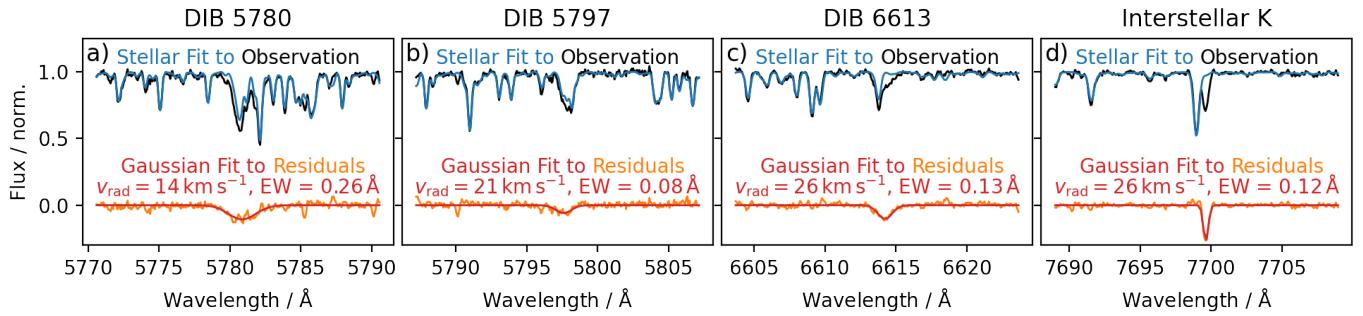
Because we can create synthetic stellar spectra for the full wavelength range, we can now also trace interstellar absorption in the residuals of observed spectra. By default, we try to calculate the equivalent width via Gaussian fits to the three diffuse interstellar bands ( $5780.59, 5797.19, 6613.66 \text{ \AA}$ ) with central wavelengths identified by Vogrinčič et al. (2023) as well as for interstellar K ( $7698.9643 \text{ \AA}$ ), see Fig. 13. We report the equivalent widths `eq_x`, standard deviations `sigma_x` and radial velocities `rv_x` for  $x$  in `k_is` for interstellar K and  $x$  in `DIB_5780`, `DIB_5797`, and `DIB_6613` for the DIBs.

### 6.1.4 Emission estimates for the Balmer lines

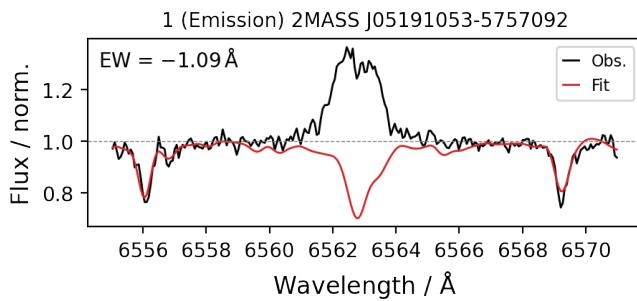
The difference between synthetic and observed Balmer line absorption holds valuable information on active stars, as well as the known inaccuracy of the synthetic Balmer lines, as well as masses for evolved stars (Bergemann et al., 2016) and possibly even information on unresolved binary systems (Sayeed et al., 2024). We therefore perform a trapezoidal integration around the Balmer lines at  $4861.3230$  and  $6562.7970 \text{ \AA}$  whose values we report in `ew_h_beta` and `ew_h_alpha`. By default we integrate in a window of  $\pm 0.75$  and  $1.25 \text{ \AA}$  for  $H_{\beta}$  and  $H_{\alpha}$ , respectively, and increase this window to  $5 \text{ \AA}$  if the average observed, normalised flux within  $\pm 0.5 \text{ \AA}$  of the Balmer line core exceeds 1. An example of such a star is shown in Fig. 14, for which we measure a residual EW of  $1.09 \text{ \AA}$ . Most emission line stars in the GALAH sample are found in the region of pre-main-sequence and hot stars (see Fig. 37a).

## 6.2 Uncertainty estimation and validation

The uncertainties that we report for our spectroscopic data analysis are based on the comparison to literature measure-



**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.** Example of a star with clear emission in the Balmer lines (here  $H_{\alpha}$ ). The bit indicating emission has been raised in the major quality flag `flag_sp` for this star.

ments to estimate accuracy uncertainties and a combined precision uncertainty estimate from adjusted covariance estimates of 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. In the following sections, we lay out how we estimate and validate accuracy and precision uncertainties in Secs. 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 yet. 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 and ignore the possibly large accuracy uncertainties for the elemental abundances. In all cases, we do estimate an overall bias with respect to literature values and

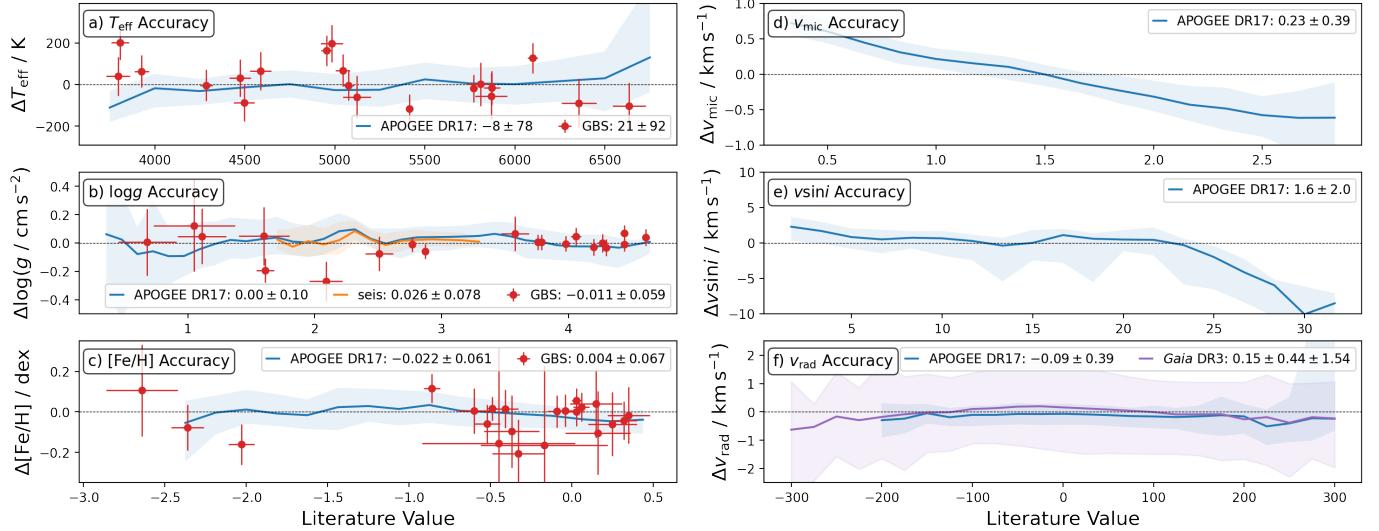
**Table 4.** List of accuracy and representative precision uncertainties for stellar parameters in GALAH DR4. Accuracy values are estimated from comparisons with literature references (see Sec. 6.2.1), whereas precision estimates are estimated from covariance uncertainties and repeat observations (Secs. 6.2.2). Here, we list the median precision uncertainties for stars with  $\text{SNR} = 50 \pm 10$  in CCD2 (see Fig. 19).

Parameter / Unit	Accuracy	Precision ( $\text{SNR} = 50$ )
$T_{\text{eff}} / \text{K}$	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	$0.17 \pm 0.02$

then combine these estimates to a globally applied zeropoint correction. Where not explicitly stated otherwise, we use 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 8. 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 remain the *Gaia* FGK benchmark stars (Jofré et al., 2014, 2015, 2018; Heiter et al., 2015). Additionally, we use estimates from asteroseismic estimates of 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,b; 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 figure Fig. 15.

$T_{\text{eff}}$ : The effective temperature estimates from GALAH DR4 show good agreement with the *Gaia* FGK benchmark stars (Fig. 15a). Specifically, we find a mean difference of  $\Delta T_{\text{eff}} = 21 \pm 92$  K, indicating no significant bias between



**Figure 15. 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). 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).

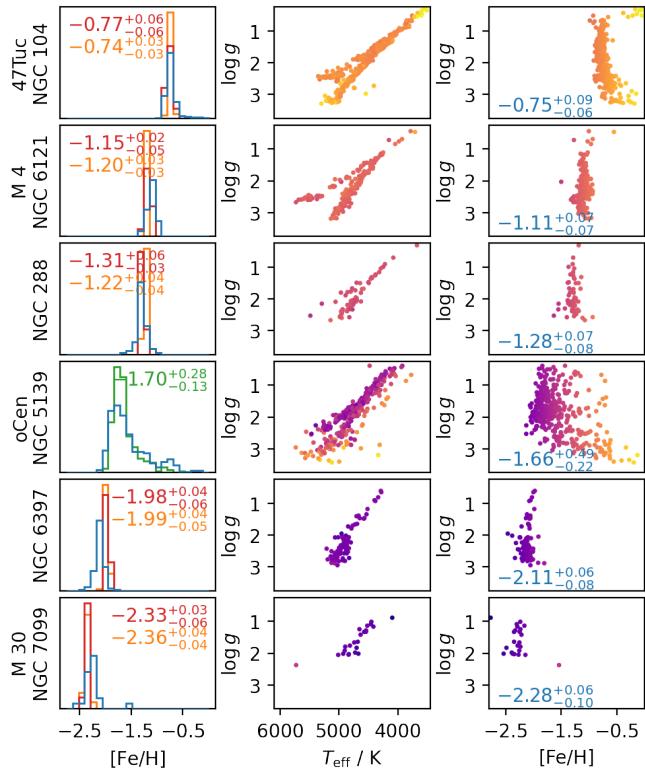
our temperatures and the benchmark values. Comparisons with APOGEE DR17 show an equally robust agreement, with  $\Delta T_{\text{eff}} = -8 \pm 78$  K. This small offset and uncertainty suggest that the GALAH DR4  $T_{\text{eff}}$  estimates are highly reliable across a wide range of at least G- and 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. 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 do 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 Sec. 8.4. Because this single group is driving the scatter of 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 on the 0.049 dex level. The application of a zeropoint correction (see Table 8) 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. 36. 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. (2009b,a) for NGCs 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. 16). While this is not necessarily confirming our accuracy, it is showing consistency within this uncertain parameter regime. We note however, a specific region in NGC 104, where the metallicity of the most luminous giants ( $T_{\text{eff}} < 3750$  K and  $\log g < 0.5$ ) is incorrectly estimated near the Solar value. We discuss this problem in detail as a caveat in Sec. 8.6, 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. A more detailed analysis of globular cluster abundances beyond [Fe/H] will be performed in a dedicated 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., in preparation).

$v_{\text{mic}}$ : 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 $^{-1}$ . However, the comparison reveals a linear mismatch: APOGEE DR17 tends to measure lower  $v_{\text{mic}}$  values for stars with low microturbulence and larger  $v_{\text{mic}}$  values for stars with higher microturbulence



**Figure 16. 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 colored 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.

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. Fitting  $v_{\text{mic}}$  returned a similar pattern as the empirical relation by Dutra-Ferreira et al. (2016) and show a significantly different behaviour of  $v_{\text{mic}}$  for the hottest, coolest, and red clump stars (see Fig. 29). 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 are starting 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 are fitting two Gaussian distributions to the overall difference of GALAH and *Gaia* radial velocities (see Fig. 17). 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.

**Elemental abundances [X/Fe]:** While there is no model-independent benchmark for abundance accuracy, we continue to perform a variety of comparisons with literature measurements to at least analyse the consistency with other estimates. In GALAH DR4, we analyse the overall zeropoints of abundances with up to five different estimates (see Fig. 32): (1) the spectroscopic analysis of a spectrum with solar composition of the asteroid 4 Vesta (210115002201239), (2) the abundance estimate from Solar twins for an Solar age of 4.5 Gyr (see Fig. 18), (3) the abundances estimates of *Gaia* FGK benchmark stars (Jofré et al., 2015, 2018), (4) the stars with Solar-like metallicity  $-0.1 < [\text{Fe}/\text{H}] < 0.1$  in the Solar neighbourhood of  $D_{\odot} < 500 \text{ pc}$  (a method introduced by Jönsson et al., 2020), and (5) the abundance difference of stars overlapping with the high-resolution large scale spectroscopic APOGEE DR17 (Abdurro'uf et al., 2022). We stress that our abundance corrections and as a consequence thereof our Solar abundances in Table 8, are estimated in 1D LTE or 1D NLTE and do not represent the most accurate Solar abundances, but merely reflect our rather best attempt to minimise the disagreement of abundances for a variety of cases. For several science cases, an adjustment of the abundance zeropoints might be advisable.

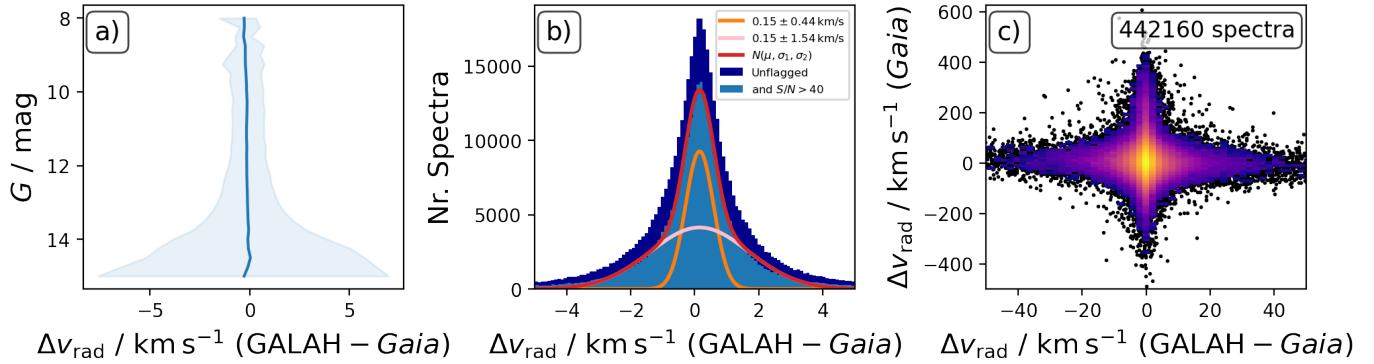
While we are not able to include all of our validation plots, we refer the interested reader to the publicly available code in our code repository<sup>g</sup>. Generally speaking, we have found that a large number of systematic trends of abundances with temperature and surface gravity has decreased with 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. 16.

### 6.2.2 Precision estimation and validation

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

We note that while we report the raw fitting covariance matrix for each spectrum and module (see Fig. 31 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

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



**Figure 17. 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, that is 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.

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 signal-to-noise of CCD2. Contrary to GALAH DR3, we have extended this rescaling function to be fitted in bins of signal-to-noise with both a constant, linear, and exponential term with `snr_px_ccd2` as the independent variable, that is,  $c_1 + c_2 \cdot S/N + c_3 \cdot \exp(c_4 \cdot S/N)$ . We report the fitted constants for both `allspec` and `allstar` in the online repository<sup>h</sup> for each stellar parameter and abundance.

In Figs. 19 and 33, 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 overpredict the precision uncertainty of stellar parameters, we do not expect a too strong overprediction for elemental 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. We note, however, that the precision of Eu seems to have decreased.

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

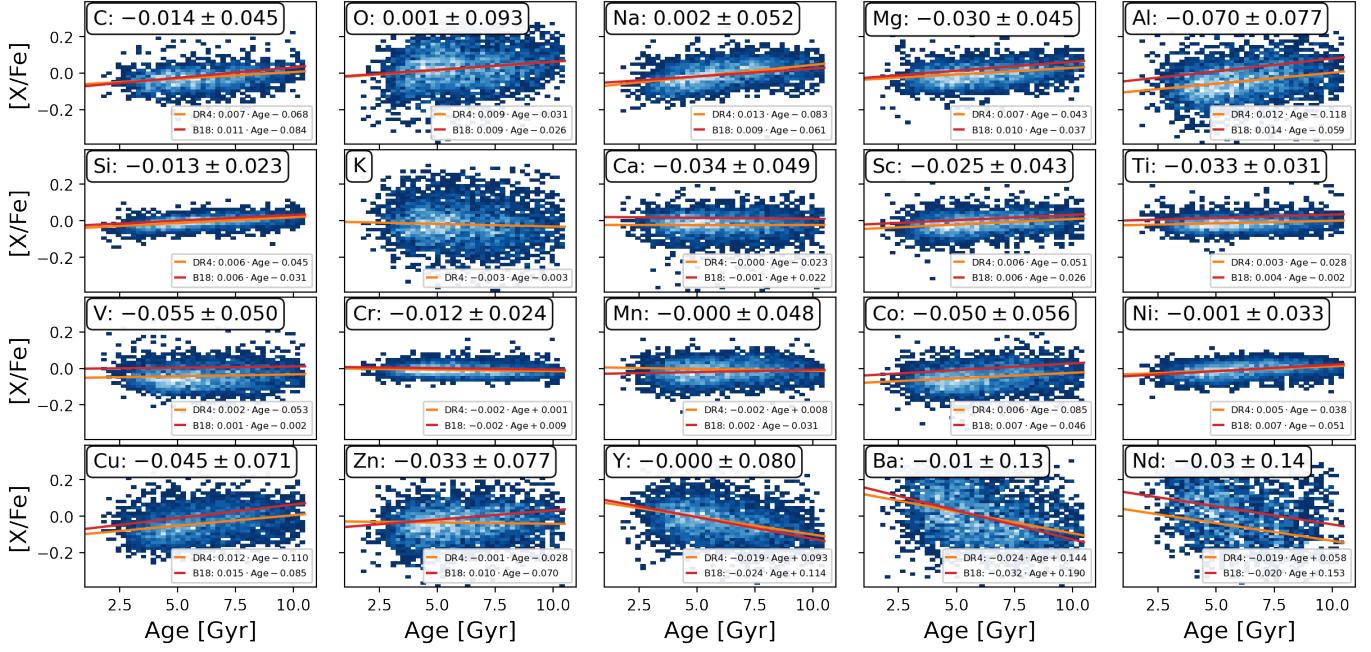
To get a better idea of the actual improvement of accuracy and improvement, we have performed more elaborate comparisons as those in Fig. 15 and only showcase a few in this manuscript with reference to the online repository. We have found the

comparison of GALAH DR4 with both GALAH DR3 and APOGEE DR17 highly informative.

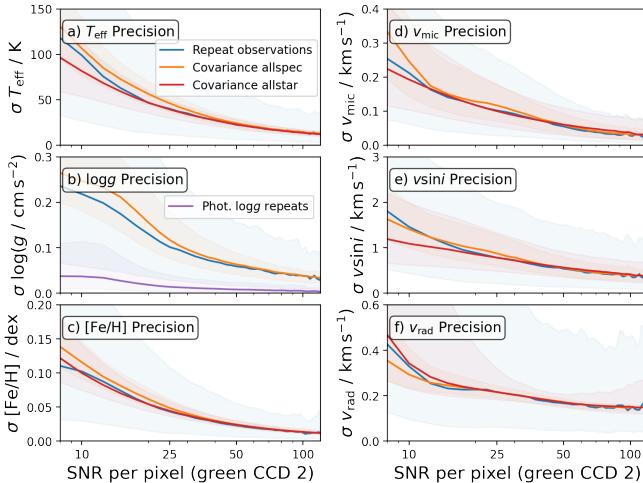
Because GALAH DR3 did not include N measurements yet and only a limited amount of C measurements, our first comparison concerns the abundances of C and N between GALAH DR4 and APOGEE DR17 in Fig. 20. We attach the comparisons for the other overlapping elements O, Na, Al, Si, K, Ca, Ti, V, Cr, Mn, Co, and Ce in Figs. 34 and 35. While we see a generally good agreement of the shapes, we notice both a bias of  $-0.03$  dex and  $0.10$  dex for C and N, respectively. These can be, however, explained with 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 for 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 et al., 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. 21. The  $\alpha$ -process element Mg has significant value Galactic studies because it is predominantly produced by core-collapse supernovae (Kobayashi et al., 2020). In GALAH DR3, only the Mg i5711 line was used, whereas we are now using a combination of several lines. This has led to a significant improvement of precision, as can be appreciated from the comparison of Figs. 21a and 21b. Even more positive, we see an improved agreement of the [Fe/H] vs. [Mg/Fe] measurements between GALAH DR4 (Fig. 21b) and APOGEE DR17 (Fig. 21c), 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 increment in used flux information 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 Figs. 21c-f).

<sup>h</sup>\*precision\_correction\_factors\* in [https://github.com/svenbuder/GALAH\\_DR4/tree/main/catalogs](https://github.com/svenbuder/GALAH_DR4/tree/main/catalogs).

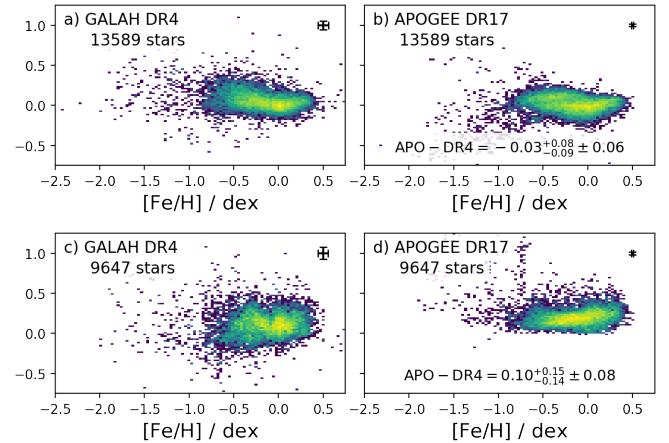


**Figure 18. Chemical abundances [X/Fe] of Solar twin stars as a function of ages that were estimated as part of the mass and age estimation of the allstar spectrum analysis.** We overplot linear fits to our age-abundance relations for Solar twins in orange as well as the literature values from Bedell et al. (2018) in red.



**Figure 19. Precision monitoring 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).

In addition to these instructive comparisons, we also want to return to the precision of Solar twins from Fig. 18. Here we specifically want to highlight the significant improvement of precision from GALAH DR3 to GALAH DR4 with respect to the linear estimate from 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

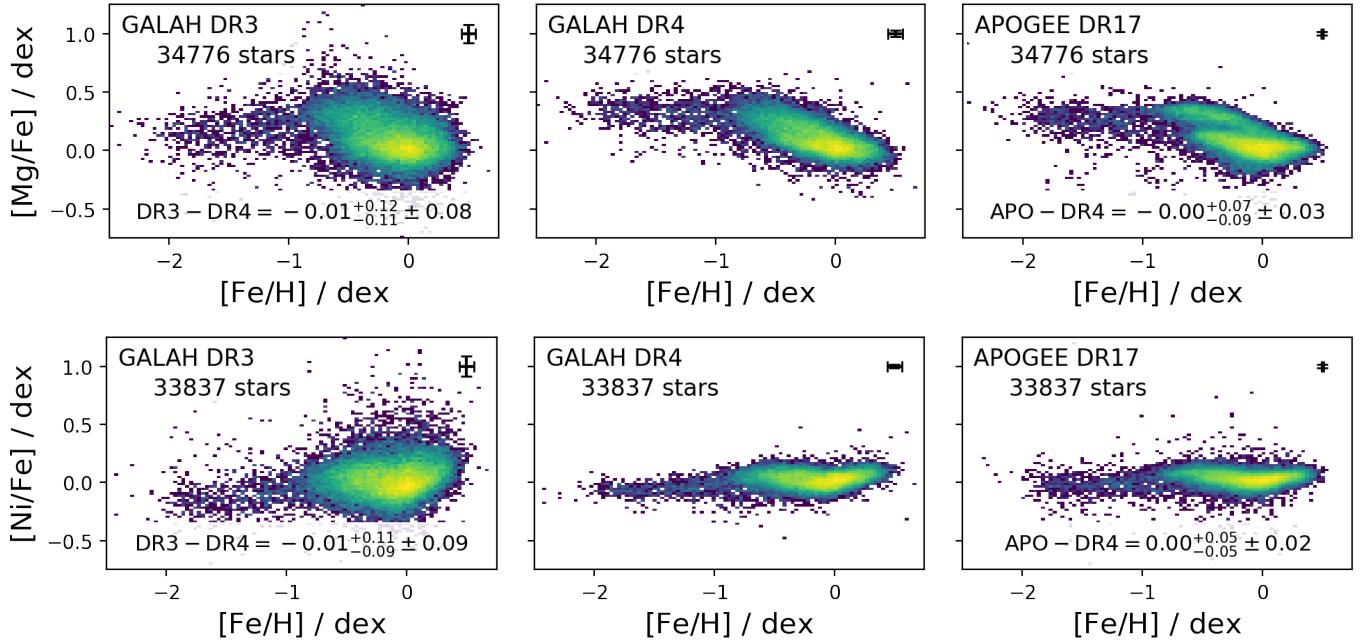


**Figure 20. Comparison of stars with available measurements in GALAH DR4 (left column) and APOGEE DR17 (right) for [C/Fe] (top row) and [N/Fe] (bottom row).**

that these values are representative of a precision improvement beyond the Solar twins, as shown for example for Ni in Fig. 21.

### 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 Tab. 5 and distributions in the



**Figure 21.** Comparison of stars with available measurements in GALAH DR3 (left column), GALAH DR4 (middle) as well as APOGEE DR17 (right) for [Mg/Fe] (top row) and [Ni/Fe] (bottom row).

Kiel diagram ( $T_{\text{eff}}$  and  $\log g$ ) are shown for each raised bit in Fig. 37 for the `allstar` catalogue. For examples of stars with raised flags, we refer back to the emission line star (`flag_sp` = 1) of Fig. 14 and the clearly double-lined binary of Fig. 11.

**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 the of 917 588 stars.

Raised Bit	Flag	Description	<code>allspec</code>	<code>allstar</code>
	0	No flag	700125	663075
0	1	Emission	9568	7646
1	2	CCD missing	70078	44344
2	4	Spectr. Binary 1	0	25538
3	8	Spectr. Binary 2	34833	32566
4	16	$\chi^2 > 3\sigma$	66859	20544
5	32	$v \sin i$ warning	138317	95990
6	64	$v_{\text{mic}}$ warning	99692	78686
7	128	Triple Binary warning	0	0
8	256	$T_{\text{eff}}$ warning	0	0
9	512	$\log g$ warning	19863	10900
10	1024	[Fe/H] warning	0	0
11	2048	S/N low	123736	71154
12	4096	Not converged	32986	0
13	8192	Model extrapolated	69613	5953
14	16384	No Results	7400	10899

While not intended to identify binaries, we believe that both the  $v \sin i$  and  $v_{\text{mic}}$  flags are informative for binaries below  $T_{\text{eff}} < 6000$  K (see their elevated position in Figs. 37f and 37g). We have trained the stars of this region with a lower maximum

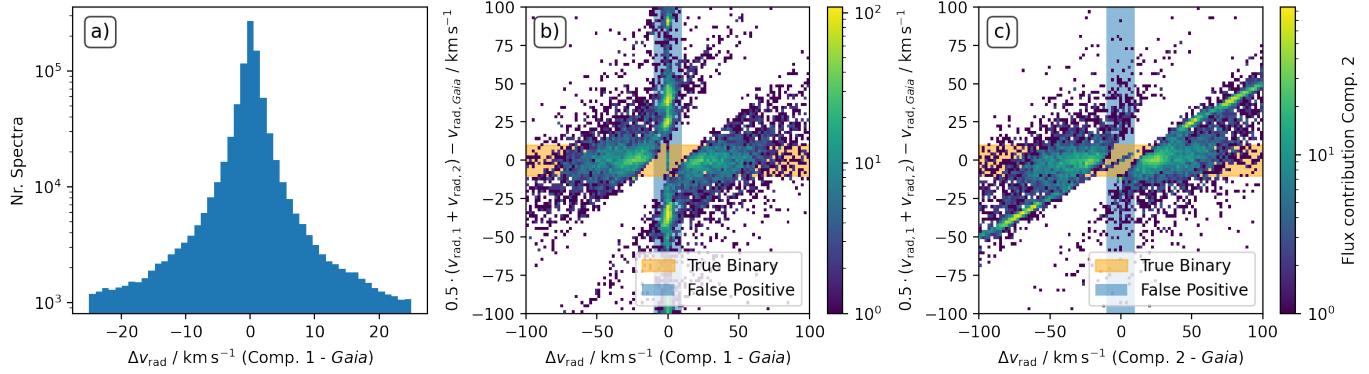
$v \sin i$  range that would be reached for a spectrum that is broadened due to binarity. This region certainly overlaps with the one of identified single-lined and double-lined binaries with `flag_sp` = 4 and 8, respectively (see Figs. 37c and 37d). For the latter, we notice that especially cool giants 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 sufficiently and thus showing up as residual signal. While these stars are possibly flagged false-positively, we do also find a remarkable amount of true binaries, 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. 22, 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 S/N, 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 S/N too low to allow for an estimate. The list of bits and flags for elemental abundances, `flag_X_fe`, is shown in Table 6.



**Figure 22. 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.

**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	measurement of Li, Ca, or Ba

## 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 fit 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-fit spectrum and 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. We note that our initial test of overall detectability (Sec. 4.3.1) allowed us to identify elements for which not even an upper limit was expected, raising flag `FLAG_X_FE` by 2.

## 7. DATA RELEASE PRODUCTS

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

The data products are provided directly on the AAO Data-Central website at <https://cloud.datacentral.org.au/teamdata/>

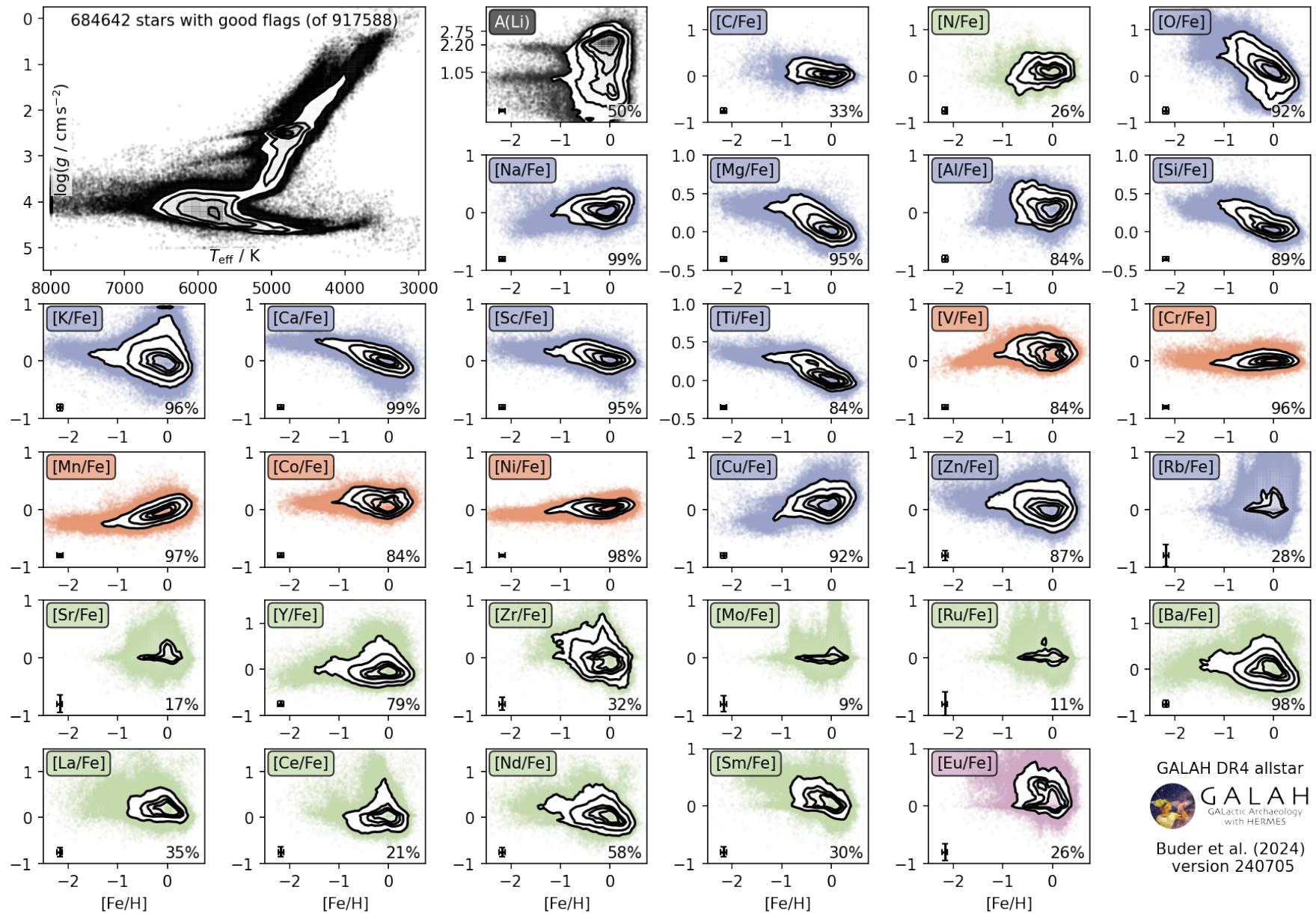
GALAH/public/GALAH\_DR4/. We further provide multiple ways to interact with the data release products, which are described in Sec. 7.4.

### 7.1 Main Data Release Catalogues

1. `galah_dr4_allspec_240705.fits`: analysis for each spectrum (incl. 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 7, 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. 24, 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) and Martig et al. (2016). Our measurement demonstrate the potential of [C/N] abundances to better separate the core-helium burning from the red giant phase (around the blue areas of Figs. 24b and 24c).



**Figure 23. 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 colored 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 color schema from Kobayashi et al. (2020).

## 7.2 Value-Added Catalogues (VAC)

We provide several value-added catalogues, namely a cross-match 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 dynamic 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>i</sup> with the *Gaia* DR3, 2MASS, and WISE catalogues as well as the distance catalogue by 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>j</sup> (Skrutskie et al., 2006) via its designation and linked this match to the *Gaia* DR3 catalogue via the best neighbour<sup>k</sup> and joined<sup>l</sup> catalogues of 2MASS to *Gaia* DR3 (Torra et al., 2021). When cross-matching between *Gaia* DR3 and 2MASS, some stars were associated with multiple possible matches. To ensure the best match, the data was sorted by from brightest to faintest G-band magnitude, and only the brightest match for each sobject\_id was retained.

The crossmatch to the WISE far-infrared photometry catalogue<sup>m</sup> (Cutri et al., 2014) was performed via the *Gaia* DR3's best neighbour catalogue<sup>n</sup> (Torra et al., 2021). The match to the distance catalogue<sup>o</sup> by Bailer-Jones et al. (2021) via the *Gaia* DR3 source\_id.

The catalogue also includes uncertainties in the *Gaia* DR3 photometric magnitudes (G, GBP, GRP) 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 zeropoint 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 ( $\nu_{\text{eff}}$ ) used in astrometry, pseudocolour, latitude, and the astrometric solution type. The parallax zeropoints and original parallaxes are reported as plx\_zpt\_corr and parallax\_raw.

In addition to 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 are overlapping

with the gaiadr3.vari\_classifier\_result catalogue. In particular, we find 17 256 SOLAR\_LIKE variables, 14 477 stars in the DSCT/GDOR/SXPHE category ( $\delta$ Scuti,  $\gamma$  Doradus, or SX-Phoenicis), 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>p</sup> includes the kinematic and dynamical properties for stars in the GALAH DR4 survey. The catalogue is created with a publicly available script<sup>q</sup> as part of GALAH DR4. We define the position of the Sun in our Galactic reference frame as  $R_{\text{GC}} = 8.21 \text{ kpc}$  (McMillan, 2017),  $\varphi_{\text{GC}} = 0 \text{ rad}$ , and  $z_{\text{GC}} = 25 \text{ pc}$  (Bland-Hawthorn & Gerhard, 2016). We then combine the total velocity in  $V$  of the Sun at  $R_{\text{GC}}$  based on the proper motion measurement of  $6.379 \pm 0.024 \text{ mas yr}^{-1}$  by (Reid & Brunthaler, 2004), that is,  $V_{\odot} = 248.27 \text{ km s}^{-1}$  with the circular velocity of  $V_{\text{circ}} = 233.10 \text{ km s}^{-1}$  by McMillan (2017) to estimate a peculiar velocity of the Sun with respect to the local standard of rest of  $15.17 \text{ km s}^{-1}$ . For the other two components, we use the estimate by Schönrich et al. (2010), leading to a peculiar velocity of the sun of  $(U, V, W) = (11.1, 15.17, 7.25) \text{ km s}^{-1}$ .

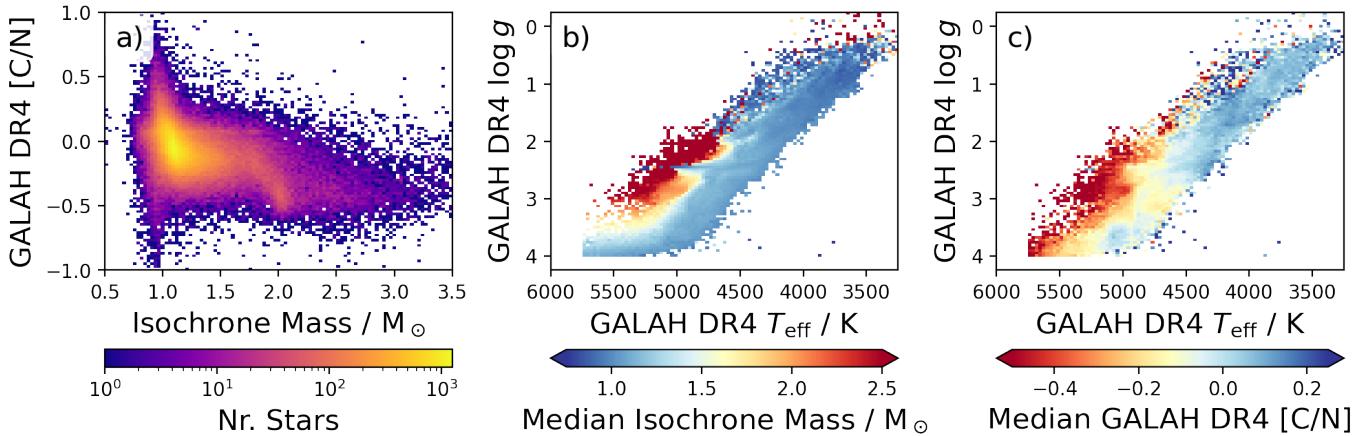
Starting from the crossmatch of GALAH DR4 with the *Gaia* DR3 (see Sec. 7.2.1), we use the GALPY.ORBIT module by Bovy (2015) to estimate heliocentric Cartesian coordinates ( $X, Y, Z$ ) and velocities ( $U, V, W$ ) as well as Galactocentric Cylindrical coordinates ( $R, \varphi, Z$ ) and velocities ( $\nu_R, \nu_\varphi, \nu_Z$ ). We approximate the orbit actions  $J_R, J_\varphi = L_Z, J_Z$  and frequencies  $\omega_i$  with the GALPY.ACTIONANGLE.ACTIONANGLESTAECHEL 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 Staeckel approximation (Binney, 2012) to calculate eccentricity, maximum orbit Galactocentric height, and apocentre/pericentre radii with GALPY's ECCZ-MAXRPERIRAP (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. 25) and angular momentum  $L_Z$  versus orbit energy  $E$  (Fig. 26) and visualises the potential of GALAH DR4 observations to complement Galactic dynamics studies and enable Galactic chemodynamic studies.

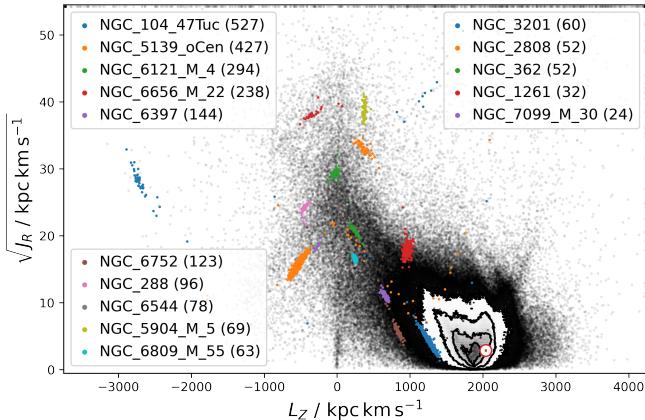
<sup>i</sup>galah\_dr4\_vac\_wise\_tmass\_gaiadr3  
<sup>j</sup>gaiadr1\_tmass\_original\_valid  
<sup>k</sup>gaiadr3\_tmass\_psc\_xsc\_best\_neighbour  
<sup>l</sup>gaiadr3\_tmass\_psc\_xsc\_join  
<sup>m</sup>gaiadr1\_allwise\_original\_valid  
<sup>n</sup>gaiadr3\_allwise\_best\_neighbour  
<sup>o</sup>external\_gaiaedr3\_distance

<sup>p</sup>galah\_dr4\_vac\_dynamics

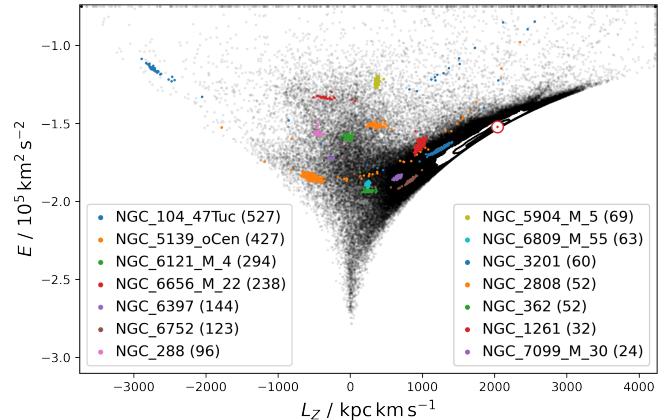
<sup>q</sup>Accessible in the GALAH DR4 repository here.



**Figure 24.** The ratio of [C/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 25. Distribution of the dynamic 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 26. Distribution of the dynamic properties of angular momentum  $L_Z$  and orbit 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.

### 7.2.3 VAC of 3D NLTE lithium abundances

In this value-added catalogue<sup>r</sup>, we use spectrum fitting to infer 3D non-local thermodynamic equilibrium (NLTE) lithium abundances. For each spectra, the Li line is modelled with a 3D NLTE BREIDABLIK line profile. 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 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. (2024) for a detailed description of the methodology.

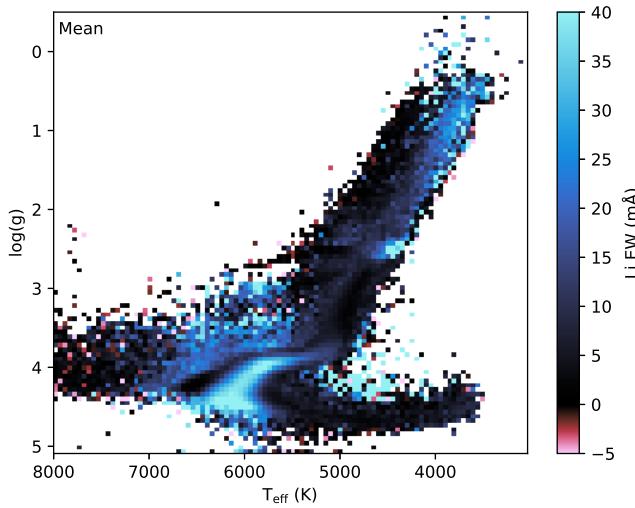
We update this methodology to provide better performance and fitting constraints. Wang et al. (2024) measured a local line width for the Li region which incorporated the thermal broadening, rotational broadening, and instrumental profile

of blending lines. For this work, we use the GALAH instrumental profile convolved with the GALAH rotational velocity for the line width. Wang et al. (2024) measured a separate Li line width which includes only the rotational broadening and instrumental profile as the 3D NLTE profiles are already thermally broadened. For this work, we set the width of the Li line to be the same as the blending lines. Whilst we still measure a local radial velocity due to imperfections of the wavelength solution in CCD3, 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 3000 K, but the STAGGER model atmospheres only reach 4000 K, therefore, we provide an additional column of 1D NLTE  $A(\text{Li})$  inferred through our EWs using BREIDABLIK on MARCS model atmospheres. Using the updated methodology it takes  $\sim 2$  minutes per star in comparison to 0.5–2 hours

<sup>r</sup>galah\_dr4\_vac\_3dnlte\_a\_li

per star reported in Wang et al. (2024), with the main speed up coming from the Li FWHM constraint.

EWs are measured for 892 223 stars (97% of GALAH DR4), with 3D NLTE A(Li) detections reported for 417 825 stars (46%) and upper limits for 474 398 stars (52%). Fig. 27 shows the mean EW over  $T_{\text{eff}}$  and  $\log g$ . The Li-dip can be seen on the main sequence turn-off at  $T_{\text{eff}} \approx 6500$  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. 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 followup papers. 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. (2024) for more details on the bitmask flag). We recommend `flag_ALi < 2` when using the 3D NLTE A(Li), and `flag_ALi < 4` when using Li EWs. A similar quality flag `flag_ALi_1D` is provided corresponding to the 1D NLTE A(Li) included in the VAC.



**Figure 27.** 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.

For convenience, we have already included the most important columns of this catalogue in the `allstar` catalogue (see Table 7), 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

TBD.

### 7.3 Data products for each spectrum and star

We provide individual data products in an orderly fashion that allows to create links to these products based solely on the `sobject_id`. To download data products for individual stars we recommend to create a url string and use `wget` or

similar commands. For bulk downloads of the advanced data products of this section, we recommend to contact the GALAH collaboration or use 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>s</sup> and sorted into directories with four spectra – one for each of the four CCDs. These spectra are produced by the reduction pipeline (see Sec. 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 `CDELT1` 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 Sec. 7.3.3 for co-added spectra). We note that not all files might be available for a given exposure due to the seldom failure of CCD readouts.

#### 7.3.2 Additional products of the `allspec` module

The `allspec` analysis product directory<sup>t</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 Fig. 9), the raw fitting results `*results.fits` and their covariance matrices `*covariances.npy` (similar to the entries used to produce Fig. 31). 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-fit model spectrum interpolated onto the same wavelength. Finally, we provide an `*comparison.pdf` (similar to Fig. 30) 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>u</sup> also include the radial velocity monitoring `*rv.png`, results files `*results.fits`, combined spectra `*spectrum.fits` and `*comparison.pdf` overview, similar to the ones described in Sec. 7.3.2. In addition, each directory also includes a `*sobject_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.

<sup>s</sup>`observations/YYMMDD/spectra/com/sobject_id*.fits`

<sup>t</sup>`analysis_products_single/YYMMDD/sobject_id/`

<sup>u</sup>`analysis_products_allstar/YYMMDD/sobject_id/`

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 along 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). While we have attempted to already flag possible caveats, we also lay out problematic flags in (Section 8.7). We summarise the most important caveats in Section 8.8.

### 8.1 Spectrum reduction

Although a significant amount of work was spent on improving the spectrum reduction, we want to lay out a few areas that could still suffer from imperfection.

#### 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 amount of problematic wavelength solutions towards the end of CCD3, with the only used absorption lines of Li and Eu, from an initially bad wavelength solution for 7.9% of the spectra.

#### 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 1-dimensional (1D) atmospheres with local thermodynamic equilibrium (LTE) towards 1D NLTE, that is, statistical (or non-local) thermodynamic equilibrium (Amarsi et al., 2020b). This includes the use of 1D NLTE synthesis for atomic lines H (Amarsi et al., 2018b), Li (Lind et al., 2009; Wang et al., 2020), 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, 2022), and Ba (Gallagher et al., 2020) for the MARCS model atmosphere grid. The work by Wang et al. (2024) 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 optical and infrared (compare e.g. Bensby et al., 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 C2 and CN have to be modelled in 1D LTE.

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 since. We also note that a Python-based version of SME, pySME (Wehrhahn, 2021), has become available. In addition, the spectrum synthesis package KORG (Wheeler et al., 2023, 2024) has been published in Julia with a Python interface and offers a faster alternative to SME once 1D NLTE synthesis has been implemented, which is vital for an application to many of the 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 use an imperfect atmosphere in the synthesis.

### 8.2.2 Mismatch of atmosphere and spectrum chemistry

For several of our synthetic spectra, the chosen chemical composition differs significantly from the scaled-Solar pattern of

the MARCS model atmospheres. We note that dedicated MARCS atmospheres have already been computed with adjusted alpha-process, oxygen, and nitrogen abundances to match the diverse chemical composition needed to model APOGEE infrared spectra (Abdurro'uf et al., 2022).

### 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 zeropoints are. In the case of several elements, like Na, Al, and Eu they are significant, on the order of 0.2 – 0.3 dex. In that case, stars with high abundances of 0.7 – 0.8 dex are not sufficiently covered, although expected, for example in old stars and especially globular clusters (see e.g. Carretta et al., 2009b).

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 randomized, uncorrelated parameter sampling, which can lead to unrealistic combinations of elemental abundances. Elements that share a similar nucleosynthesis channel often exhibit correlated behavior, 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; Budér et al., 2021). To better capture this behavior in the training set, the use of scaled linear functions or normalizing flows could be advantageous. These approaches would help minimize 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. 18 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 of spectral features and 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 (Budér 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 line was mainly measured only from the weak Eu6645 line, whereas the neural networks are not restricted to this region.

#### 8.3.3 Flexibility of neural networks in general

The choice of using a large number neural networks of restricted regions in the  $T_{\text{eff}}$ ,  $\log g$ , and [Fe/H] space was done based on the hope that they would require less flexibility from neural networks and thus a smaller network architecture. 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 have found the lowest root mean square errors for our chosen activation function. We have further tested a larger number of neurons, but found the root mean square errors to stabilize 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, MSTO 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, and would also decrease 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., 2020). 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 work of fitting Gaussian lines to multiple components in the wavelength range around 6708 Å is actually delivering appropriate measurements of  $\text{EW}(\text{Li})$ , 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 J22353100-6658174 (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  $[\text{Y}/\text{Fe}] = 1.2$ ,  $[\text{Ba}/\text{Fe}] = 1.5$ ,  $[\text{La}/\text{Fe}] = 1.5$ ,  $[\text{Ce}/\text{Fe}] = 1.1$ ,

$[\text{Nd}/\text{Fe}] = 1.9$ , and  $[\text{Sm}/\text{Fe}] = 1.2$ . A neural network that is not trained on such high abundances is likely to improperly extrapolate stellar spectra.

While we have tried to stars with diverse abundance, 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 between the different approaches. For future analyses, a possible solution could therefore be to follow a 2-step approach of first running one generic neural network for all spectra and then using optimized neural networks – or full spectrum synthesis – on smaller target samples of specific science cases.

## 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) or 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 the information of both lines. This information could thus be used to better constrain high masses and counter-act 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 large giant star masses are.

### 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., 2021b; 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. 28, where the module was applied to a spectroscopic binary type 2 and resulted in a significantly better fit than the single star analysis.

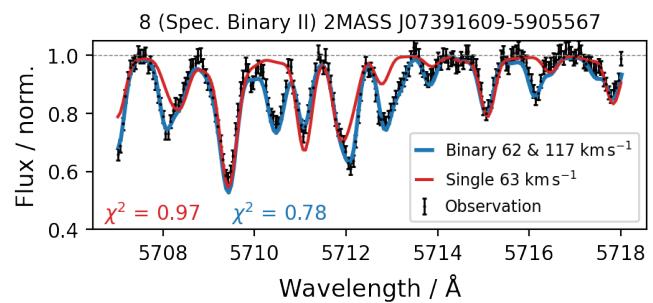


Figure 28. Example spectrum for a double-lined spectroscopic binary star (SBII) that is better fit with our binary fitting algorithm.

## 8.6 Fit optimization

As described in Sec. 4, we are using the `CURVE_FIT` function of `scipy.optimize` (Virtanen et al., 2020) to optimize the difference of observed and synthetic spectra, whose optimization 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 gotten stuck. In particular for some red clump stars as well as cool giant stars with  $T_{\text{eff}} < 3750 \text{ K}$  and  $\log g < 0.5$  (see Sec. 6.2.1), we have been able to recover a pattern of abundances being 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 `zeropoints` in Table 8).

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. Contrary to this, *The Cannon* version by (Casey et al., 2016) uses the `SCIPY.OPTIMIZE.LEASTSQ` fitting approach.

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 `DIFFERENTIAL_EVOLUTION` function of the `scipy.optimize` module could be used for slow but global optimization. Finally, multiple randomized initial starting guesses could be applied for `curve_fit`, but would multiply the computing costs linearly by the number of initial guesses.

Finally, the fitting optimization 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 another Bayesian inference engine 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 optimization to only a few major stellar labels (see e.g. Traven et al., 2020).

## 8.7 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.

### 8.7.1 Bug in `flag_fe_h`

The quality flag for iron abundance, `flag_fe_h`, was computed similar to the elemental abundances, that is, by comparing the best-fit spectrum with a spectrum with the lowest grid value. In the case of  $[\text{Fe}/\text{H}]$ , however, this is not the appropriate reference value. It had led to a comparison of a spectrum with  $[\text{Fe}/\text{H}] = -0.74 \text{ dex}$  to one with the lowest grid value of  $[\text{Fe}/\text{H}] = -0.75 \text{ dex}$ , which can easily be too similar within the SNR range. This has affected up to 34% of stars 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  $[\text{Fe}/\text{H}] = -4$ .

### 8.7.2 Fitting machinery stuck in local minimum

As laid out in Sec. 8.6, we have not been able to automatically flag all estimates for which our fitting machinery has gotten stuck in local minima, most notably the initial value.

### 8.7.3 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).

## 8.8 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 difficult effect might be that some elements could be fit 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 is causing increasingly inaccurate models and inferred stellar properties towards the coolest stars.
- The coolest giant stars ( $T_{\text{eff}} < 3750 \text{ K}$  and  $\log g < 0.5$ ) still are rather unreliable.

- Lower precision for Eu due to missing masking of neural networks.

These caveats are the unfortunate negative effects of our successful attempt to increase the accuracy and precision of most stellar parameters and elemental abundances while simultaneously pushing the number of stars for which we can report abundances.

## 9. CONCLUSIONS

The GALAH survey celebrates its 10th anniversary with the release of GALAH DR4, marking a decade of significant contributions to our understanding of the Milky Way and the elemental composition of its stars. Over the years, GALAH has been instrumental in measuring and cataloging 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 improved upon the precision and accuracy of stellar parameters and elemental abundances for over 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 further strengthened the reliability of stellar parameters, particularly for surface gravities, breaking long-standing degeneracies in spectroscopic data.

The unique value of GALAH lies in its detailed mapping of elements crucial to life as we know it. By tracking the abundances of carbon, nitrogen, and oxygen (CNO), as well as rare heavy elements used in modern electronics, GALAH has provided key insights into how the building blocks of life and technology were forged in the interiors of stars and distributed throughout the Milky Way over billions of years.

The next decade promises further improvements and discoveries as GALAH continues to push the boundaries of stellar spectroscopy. Upcoming surveys and future releases will likely address the caveats discussed in this work, refine the data products, and expand upon the wealth of knowledge already provided. The legacy of GALAH will undoubtedly continue to influence not only stellar and galactic archaeology but also our broader understanding of the cosmos and the elements that shape modern life.

planets, origin of elements, other galaxies, cosmology, stellar evolution, Galactic archaeology, other surveys.

## Acknowledgements

We acknowledge the traditional owners of the land on which the AAT and ANU stand, the Gamilaraay, the Ngunnawal and Ngambri people. 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 thank the staff at Siding Spring Observatory for their decade-long maintenance of 2dF-HERMES and support during observations, without whom this project would not have been possible: NAME1, NAME2, ...

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.

## 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](http://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 et al., 2000).

## 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, 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);

## References

- Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35  
 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., Grevesse, N., Grumer, J., et al. 2020a, A&A, 636, A120  
 Amarsi, A. M., Liljegegren, 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., Nordlander, T., Barklem, P. S., et al. 2018b, A&A, 615, A139  
 Amarsi, A. M., Lind, K., Osorio, Y., et al. 2020b, A&A, 642, A62  
 Andrae, R., Fouesneau, M., Sordo, R., et al. 2022, arXiv e-prints, arXiv:2206.06138  
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
- Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147
- Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58
- Barden, S. C., Jones, D. J., Barnes, S. I., et al. 2010, SPIE, 7735, 09
- Baumgardt, H., & Vasiliev, E. 2021, MNRAS, 505, 5957
- Bedell, M., Bean, J. L., Meléndez, J., et al. 2018, ApJ, 865, 68
- 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., Serenelli, A., Schönrich, R., et al. 2016, A&A, 594, A120
- Bergemann, M., Gallagher, A. J., Eitner, P., et al. 2019, A&A, 631, A80
- Binney, J. 2012, MNRAS, 426, 1324
- Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529
- Bovy, J. 2015, ApJS, 216, 29
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
- Brzeski, J., Case, S., & Gers, L. 2011, SPIE, 8125, 04
- Buchner, J. 2021, The Journal of Open Source Software, 6, 3001
- Buder, S., Asplund, M., Duong, L., et al. 2018, MNRAS, 478, 4513
- Buder, S., Sharma, S., Kos, J., et al. 2021, MNRAS, 506, 150
- Cantat-Gaudin, T., & Anders, F. 2020, A&A, 633, A99
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
- Carretta, E., Bragaglia, A., Gratton, R., & Lucatello, S. 2009a, A&A, 505, 139
- Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2009b, A&A, 505, 117
- Casagrande, L., & Vandenberg, D. A. 2018, MNRAS, 475, 5023
- Casey, A. R., Hogg, D. W., Ness, M., et al. 2016, ArXiv e-prints, arXiv:1603.03040
- Casey, A. R., Hawkins, K., Hogg, D. W., et al. 2017, ApJ, 840, 59
- Cutri, R. M., et al. 2014, VizieR Online Data Catalog, 2328
- Dalton, G., Trager, S., Abrams, D. C., et al. 2014, SPIE, 9147, 0L
- de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, The Messenger, 175, 3
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, MNRAS, 449, 2604
- Dutra-Ferreira, L., Pasquini, L., Smiljanic, R., Porto de Mello, G. F., & Steffen, M. 2016, A&A, 585, A75
- Evans, D. W., Riello, M., De Angeli, F., et al. 2018, ArXiv e-prints, arXiv:1804.09368
- Farrell, T. J., Birchall, M. N., Heald, R. W., et al. 2014, SPIE, 9152, 23
- Foreman-Mackey, D. 2016, The Journal of Open Source Software, 1, 24
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
- Fouesneau, M., Frémat, Y., Andrae, R., et al. 2022, arXiv e-prints, arXiv:2206.05992
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
- . 2021a, A&A, 649, A1
- . 2021b, A&A, 649, A1
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1
- Gallagher, A. J., Bergemann, M., Collet, R., et al. 2020, A&A, 634, A55
- Gent, M. R., Bergemann, M., Serenelli, A., et al. 2022, A&A, 658, A147
- Gilmore, G., Randich, S., Worley, C. C., et al. 2022, A&A, 666, A120
- Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98
- Grevesse, N., Asplund, M., & Sauval, A. J. 2007, Space Sci. Rev., 130, 105
- Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
- Heijmans, J., Asplund, M., Barden, S., et al. 2012, SPIE, 8446, 0W
- Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, A&A, 582, A49
- Heiter, U., Lind, K., Bergemann, M., et al. 2021, A&A, 645, A106
- Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Nature, 563, 85
- 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., Ness, M. K., Hogg, D. W., et al. 2017, ApJ, 836, 5
- Hon, M., Huber, D., Kuszlewicz, J. S., et al. 2021, ApJ, 919, 131
- Hunter, J. D. 2007, Comput Sci Eng, 9, 90
- Jofré, P., Heiter, U., Tucci Maia, M., et al. 2018, Research Notes of the American Astronomical Society, 2, 152
- Jofré, P., Heiter, U., Soubiran, C., et al. 2014, A&A, 564, A133
- . 2015, A&A, 582, A81
- Jofré, P., Heiter, U., Worley, C. C., et al. 2017, A&A, 601, A38
- Johnson, C. I., & Pilachowski, C. A. 2010, ApJ, 722, 1373
- Jönsson, H., Holtzman, J. A., Prieto, C. A., et al. 2020, AJ, 160, 120
- Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, A&A, 674, A5
- Kawka, A., Simpson, J. D., Vennes, S., et al. 2020, MNRAS, 495, L129
- Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, ApJ, 900, 179
- Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv e-prints, arXiv:1711.03234
- Kos, J., Lin, J., Zwitter, T., et al. 2017, MNRAS, 464, 1259
- Kos, J., Bland-Hawthorn, J., Betters, C. H., et al. 2018, MNRAS, 480, 5475
- Lewis, A. 2019, arXiv e-prints, arXiv:1910.13970
- Lewis, I. J., Cannon, R. D., Taylor, K., et al. 2002, MNRAS, 333, 279
- Lin, J., Dotter, A., Ting, Y.-S., & Asplund, M. 2018, MNRAS, doi:10.1093/mnras/sty709
- Lind, K., Asplund, M., & Barklem, P. S. 2009, A&A, 503, 541
- Lind, K., Asplund, M., Barklem, P. S., & Belyaev, A. K. 2011, A&A, 528, A103
- Lindgren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
- Lindgren, L., Bastian, U., Biermann, M., et al. 2021a, A&A, 649, A4
- Lindgren, L., Klioner, S. A., Hernández, J., et al. 2021b, A&A, 649, A2
- Mackereth, J. T., & Bovy, J. 2018, PASP, 130, 114501
- Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, ApJ, 739, 25
- Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
- Martell, S. L., Sharma, S., Buder, S., et al. 2017, MNRAS, 465, 3203
- Martig, M., Fouesneau, M., Rix, H.-W., et al. 2016, MNRAS, 456, 3655
- Masseron, T., & Gilmore, G. 2015, MNRAS, 453, 1855
- McMillan, P. J. 2017, MNRAS, 465, 76
- Monty, S., Yong, D., Marino, A. F., et al. 2023, MNRAS, 518, 965
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16
- Ness, M., Hogg, D. W., Rix, H.-W., et al. 2016, ApJ, 823, 114
- Ness, M. K., Johnston, K. V., Blancato, K., et al. 2019, ApJ, 883, 177
- Nissen, P. E. 2015, A&A, 579, A52
- Nissen, P. E., & Gustafsson, B. 2018, A&A Rev., 26, 6
- Nordlander, T., & Lind, K. 2017, A&A, 607, A75
- Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23
- Osorio, Y., Barklem, P. S., Lind, K., et al. 2015, A&A, 579, A53
- Osorio, Y., Lind, K., Barklem, P. S., Allende Prieto, C., & Zatsarinny, O. 2019, A&A, 623, A103
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J Mach Learn Res, 12, 2825
- Pérez, F., & Granger, B. E. 2007, Comput Sci Eng, 9, 21
- Piskunov, N., & Valenti, J. A. 2017, A&A, 597, A16
- Prša, A., Harmanec, P., Torres, G., et al. 2016, AJ, 152, 41
- Rains, A. D., Žerjal, M., Ireland, M. J., et al. 2021, MNRAS, 504, 5788
- Reggiani, H., Amarsi, A. M., Lind, K., et al. 2019, A&A, 627, A177
- Reid, M. J., & Brunthaler, A. 2004, ApJ, 616, 872
- Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
- Rimoldini, L., Holl, B., Gavras, P., et al. 2023, A&A, 674, A14
- Rix, H.-W., Ting, Y.-S., Conroy, C., & Hogg, D. W. 2016, ApJ, 826, L25
- Salaris, M., & Cassisi, S. 2006, Evolution of Stars and Stellar Populations (J. Wiley)
- Saydari, A. K., Uzsoy, A. S. M., Zucker, C., Peek, J. E. G., & Finkbeiner, D. P. 2023, ApJ, 954, 141

- Sayeed, M., Ness, M. K., Montet, B. T., 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., Stello, D., Buder, S., et al. 2018, MNRAS, 473, 2004
- Sharma, S., Stello, D., Bland-Hawthorn, J., et al. 2019, MNRAS, 490, 5335
- Sheinis, A., Anguiano, B., Asplund, M., et al. 2015, J. Astron. Telesc. Instrum. Syst., 1, 035002
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
- Spina, L., Meléndez, J., Karakas, A. I., et al. 2016, A&A, 593, A125
- Steinmetz, M., Matijević, G., Enke, H., et al. 2020, AJ, 160, 82
- Stello, D., Huber, D., Sharma, S., et al. 2015, ApJ, 809, L3
- Taylor, M. B. 2005, ASPC, 347, 29
- 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., Castañeda, J., Fabricius, C., et al. 2021, A&A, 649, A10
- Traven, G., Feltzing, S., Merle, T., et al. 2020, A&A, 638, A145
- Valenti, J. A., & Piskunov, N. 1996, A&AS, 118, 595
- van Leeuwen, F. 2007, A&A, 474, 653
- Vasiliev, E., & Baumgardt, H. 2021, MNRAS, 505, 5978
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
- Vogrinčič, R., Kos, J., Zwitter, T., et al. 2023, arXiv e-prints, arXiv:2303.14016
- Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, Comput Sci Eng, 13, 22
- Wang, E. X., Nordlander, T., Asplund, M., et al. 2020, MNRAS, arXiv:2010.15248
- Wang, E. X., Nordlander, T., Buder, S., et al. 2024, MNRAS, 528, 5394
- Wehrhahn, A. 2021, in The 20.5th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun (CS20.5), Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, 1
- 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
- Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, ApJS, 245, 34
- Xiang, M., Rix, H.-W., Ting, Y.-S., et al. 2021, arXiv e-prints, arXiv:2108.02878
- Yong, D., Meléndez, J., Grundahl, F., et al. 2013, MNRAS, 434, 3542
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, Research in Astronomy and Astrophysics, 12, 723
- Zinn, J. C., Stello, D., Elsworth, Y., et al. 2020, ApJS, 251, 23

## Appendix 1. Initial parameters

In this section, we append the overview of initial and final stellar parameters of GALAH DR4. We show the density distribution of  $\log g$ , [Fe/H],  $v_{\text{mic}}$ , and  $v \sin i$  in each row as a function of  $T_{\text{eff}}$ . The columns show the same stars and their final parameters, the initial value, the reduction pipeline suggestion, the *Gaia* DR3 gspphot estimates and the GALAH DR3 estimates, respectively.

## Appendix 2. Data Products

In this section, we append examples of data products of GALAH DR4 that were not already shown in the main manuscript. We present the shortened table schema of the `allstar` and `allspec` catalogues in Table 7. In Fig. 30, we show the automatically created fit comparison of the `allstar` module

for Vesta (210115002201239). In Fig. 31, we show examples of the covariance matrices for Vesta and Arcturus, as representative examples for main sequence and giant stars. We note that these covariance matrices are a direct output of `scipy.optimize.curve_fit` and note scaled up to the final label uncertainties.

## Appendix 3. Stellar parameter and abundance validation

### Appendix 3.1 Accuracy

We list the final stellar parameter and abundance zeropoints of the `allstar` module in Table 8. A complete table, including the zeropoints for the `allspec` module can be found as FITS file in the online repository. As discussed in Sec. 6.2.1, the choice of abundance zeropoints was made as a compromise between the different accuracy indicators shown in Fig. 32 for the `allstar` module.

### Appendix 3.2 Precision

Fig. 33

Although multiple 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. 36, 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.

### Appendix 3.3 Distribution of flagged stars

Fig. 37

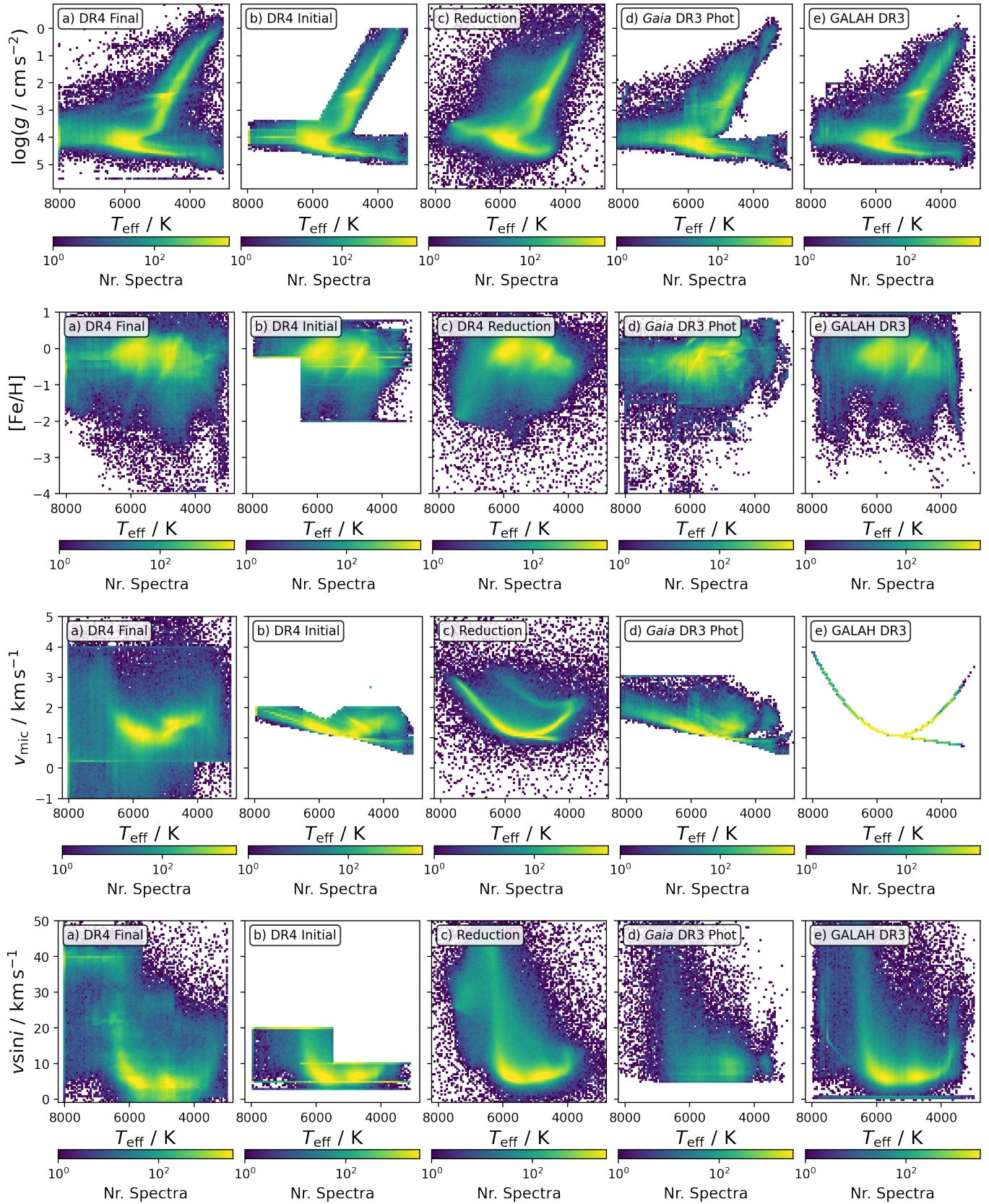
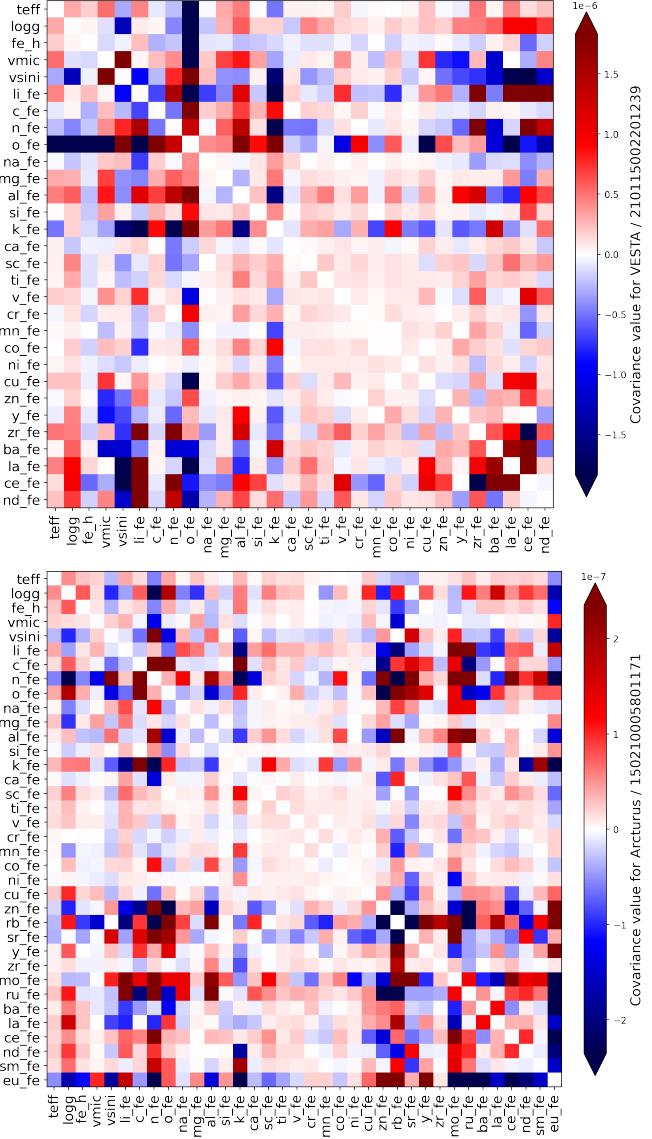


Figure 29. Comparison of final GALAH DR4 stellar parameters (first column) against the initial parameters used in the *allstar* analysis (second column), estimates from the GALAH DR4 reduction pipeline (third column), Gaia DR3 (fourth column with  $v_{\text{mic}}$  based on the adjusted formula from Dutra-Ferreira et al., 2016), and GALAH DR3 (fifth column).

**Table 7.** 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
<code>sobject_id</code>	GALAH identifier	<code>tmass_id</code>	2MASS identifier
<code>gaiadr3_source_id</code>	Gaia DR3 source_id	<code>survey_name</code>	2dF-HERMES Program
<code>field_id</code>	GALAH Field ID	<code>setup</code>	allspec/allstar
<code>mjd</code>	Modified Julian Date	<code>ra</code>	propagated from Gaia DR3
<code>dec</code>	propagated from Gaia DR3	<code>flag_sp</code>	Major spectroscopic quality bitmask flag
<code>flag_sp_fit</code>	Major fitting quality flag	<code>opt_loop</code>	Nr of optimisation loops used for fitting
<code>flag_red</code>	Quality bitmask flag of reduction pipeline	<code>snr_px_ccdN</code>	Average SNR for CCD N
<code>chi2_sp</code>	Chi2 value of spectroscopic fitting	<code>px_used_perc</code>	Percentage of spectrum used for fit
<code>model_name</code>	Used neural network for synthesis	<code>closest_model</code>	Closest neural network for synthesis
<code>comp_time</code>	Computation time spent on spectrum	<code>fit_global_rv</code>	RV fitted or fixed after co-adding?
<code>rv_comp_1</code>	Radial velocity of primary source	<code>e_rv_comp_1</code>	Uncertainty of <code>rv_comp_1</code>
<code>rv_comp_2</code>	Radial velocity of potential secondary source	<code>e_rv_comp_2</code>	Uncertainty of <code>rv_comp_1</code>
<code>rv_gaia_dr3</code>	Radial velocity in Gaia DR3	<code>e_rv_gaia_dr3</code>	Uncertainty of <code>rv_gaia_dr3</code>
<code>v_bary_eff</code>	Barycentric velocity correction	<code>teff</code>	Spectr. effective temperature
<code>e_teff</code>	Uncertainty teff	<code>logg</code>	Photometric surface gravity
<code>e_logg</code>	Uncertainty logg_plx	<code>fe_h</code>	Abundance of Fe as pseudo-metallicity
<code>e_fe_h</code>	Uncertainty fe_h	<code>flag_fe_h</code>	Quality flag fe_h
<code>vmic</code>	Microturbulence velocity (fitted)	<code>e_vmic</code>	Uncertainty vmic
<code>vsini</code>	Broadening velocity	<code>e_vsini</code>	Uncertainty of vsini
<code>nn_li_fe</code>	Elemental abundance for [Li/Fe]	<code>nn_e_li_fe</code>	Uncertainty Li_fe
<code>nn_flag_li_fe</code>	Quality bitmask flag of Li_fe	<code>x_fe</code>	Elemental abundance for [X/Fe]
<code>e_x_fe</code>	Uncertainty of elemental abundance [X/Fe]	<code>flag_x_fe</code>	Quality bitmask flag of [X/Fe]
<code>mass</code>	Mass used for calculating logg(plx)	<code>age</code>	Age estimated when calculating mass
<code>bc_ks</code>	Bolometric Correction of Ks band	<code>a_ks</code>	Attenuation in Ks-band A(Ks)
<code>lbol</code>	Bolometric Luminosity	<code>r_med</code>	Median Distance
<code>r_lo</code>	Lower Limit Distance	<code>r_hi</code>	Higher Limit Distance
<code>sb2_rv_N</code>	Nth perc. of RV residual signal	<code>ew_h_beta</code>	Equivalent Width of observed Hbeta core
<code>ew_h_alpha</code>	Equivalent Width of observed Halpha core	<code>res_h_beta</code>	Residual EW in Hbeta core
<code>res_h_alpha</code>	Residual EW in Halpha core	<code>ew_k_is</code>	EW of K7699 Interstellar Line
<code>sigma_k_is</code>	Gaussian sigma of K7699 Interstellar Line	<code>rv_k_is</code>	RV of K7699 Interstellar Line
<code>ew_dib5780</code>	Equivalent width of DIB NNNN	<code>sigma_dib5780</code>	Gaussian sigma of DIB NNNN
<code>rv_dib5780</code>	RV of DIB NNNN	<code>ebv</code>	Extinction E(B-V)
<code>phot_g_mean_mag</code>	Mean Gaia DR3 G-band apparent magnitude	<code>bp_rp</code>	Color of BP-RP bands
<code>j_m</code>	2MASS J-band magnitude	<code>j_msigcom</code>	Uncertainty of j_m
<code>h_m</code>	2MASS H-band magnitude	<code>h_msigcom</code>	Uncertainty of h_m
<code>ks_m</code>	2MASS Ks-band magnitude	<code>ks_msigcom</code>	Uncertainty of ks_m
<code>W2mag</code>	AllWISE W2-band magnitude	<code>e_W2mag</code>	uncertainty of W2mag
<code>ruwe</code>	RUWE reported by Gaia DR3	<code>parallax</code>	Astrometric parallax used for GALAH DR4
<code>parallax_error</code>	Uncertainty of astrometric parallax	<code>ew_li</code>	Eqiuvalent width of Lithium 6708 Ll line
<code>e_ew_li_low</code>	Lower uncertainty ew_li	<code>e_ew_li_upp</code>	Upper uncertainty ew_li
<code>a_li</code>	Absolute 3D NLTE Li abundance	<code>a_li_upp_lim</code>	Upper limit of absolute 3D NLTE A(Li)
<code>e_a_li_low</code>	Lower uncertainty of a_li	<code>e_a_li_upp</code>	Upper uncertainty of a_li
<code>e_a_li_teff</code>	Uncertainty of A(Li) due to temperature	<code>flag_a_li</code>	Flag for a_li measurement
<code>rv_comp_nr</code>	Nr RV cross-correlation function peaks	<code>rv_comp_1_p</code>	Prominence of <code>rv_comp_1</code> in CCF
<code>rv_comp_2_h</code>	Height of <code>rv_comp_1</code> in CCF	<code>rv_comp_2_p</code>	Prominence of <code>rv_comp_1</code> in CCF
<code>logg_spec</code>	Spectroscopic surface gravity estimate	<code>e_logg_spec</code>	Uncertainty logg_spec

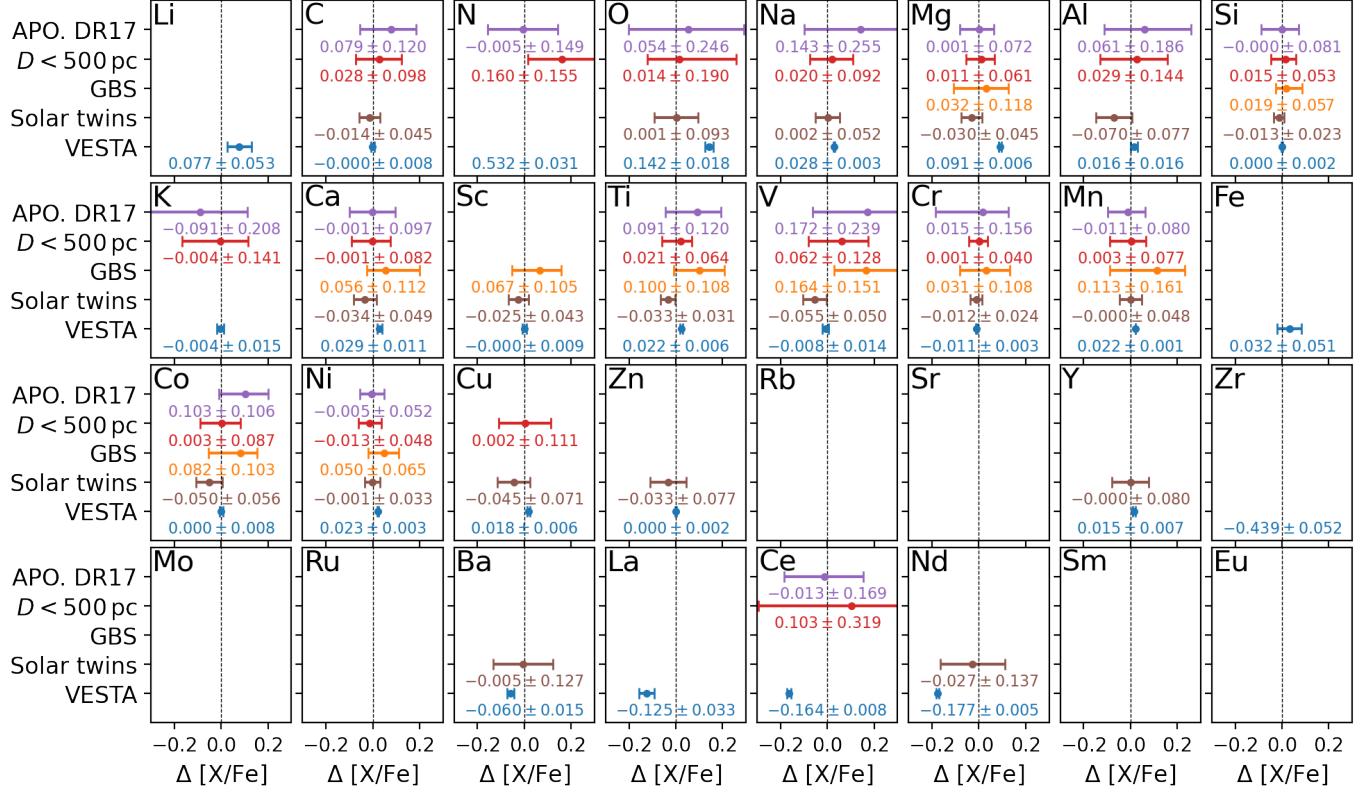




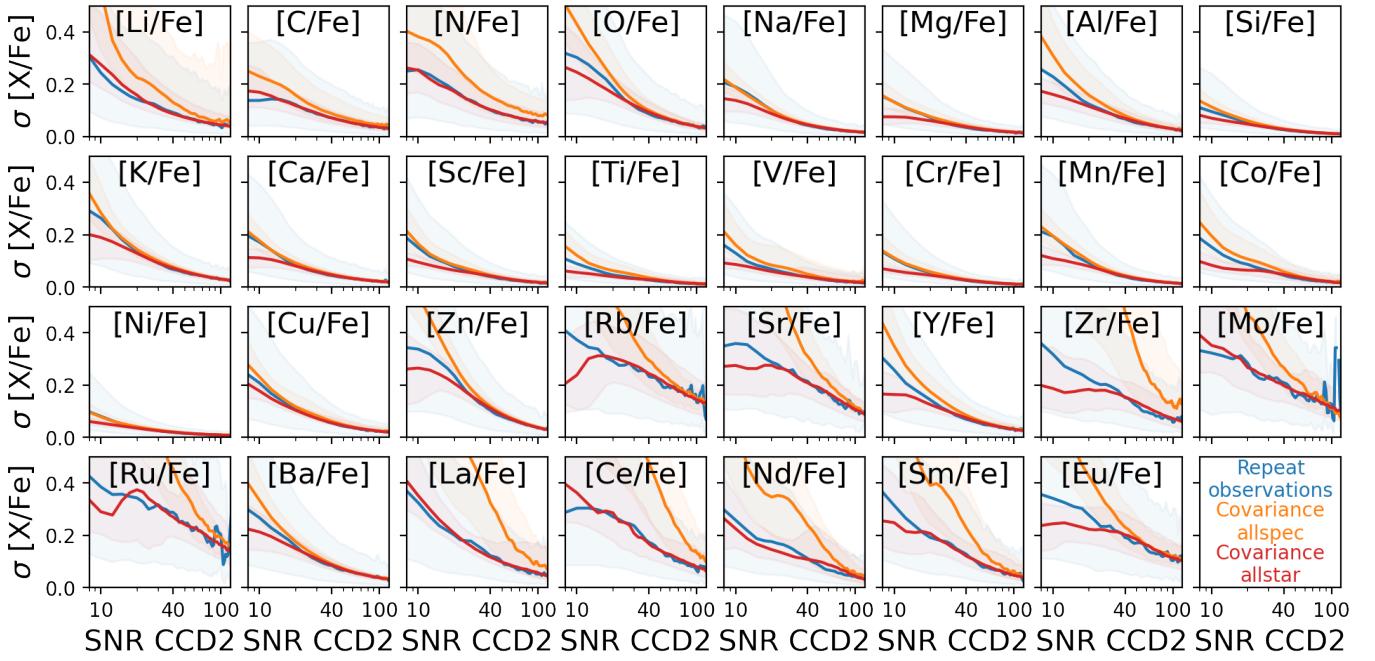
**Figure 31.** Covariance matrices for labels for Vesta (panel a) and Arcturus (panel b).

**Table 8.** Zero point corrections applied to the allstar measurements. We used Prša et al. (2016) as reference for Solar parameters and Grevesse et al. (2007) as reference for Solar abundances. For reference, we also show the combined rotational and macroturbulence as well as microturbulence velocities from Jofré et al. (2014).

Property	Reference	Zeropoint	Shift	Vesta	Vesta Shift
$T_{\text{eff}}$	5772.0	5772.0	0.0	5752.261	-19.739
$\log g$	4.438	4.438	0.0	4.429	-0.009
[Fe/H]	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 32. Zeropoint 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  $|[\text{Fe}/\text{H}]| \leq 0.1$  closer than  $D_\omega < 0.5$  kpc (red), as well as stars that were also observed by APOGEE DR17 (purple).



**Figure 33. Precision monitoring of elemental abundances as a function of SNR for the green CCD2 across for GALAH DR4..** Each panel shows the behavior 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.

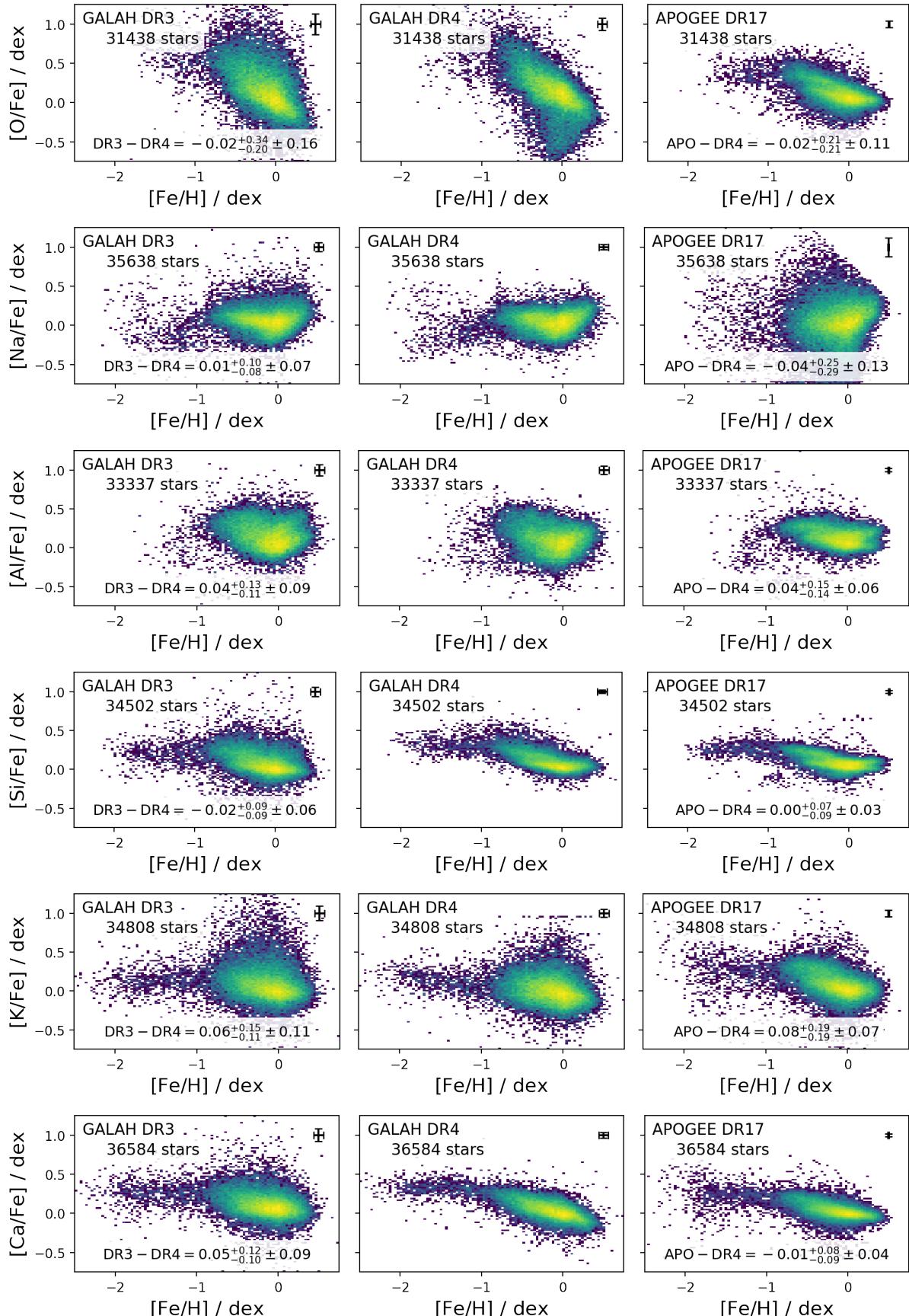


Figure 34. Comparison of stars with available measurements in GALAH DR3 (left column), GALAH DR4 (middle column) and APOGEE DR17 (right) for O, Na, Al, Si, K, and Ca.

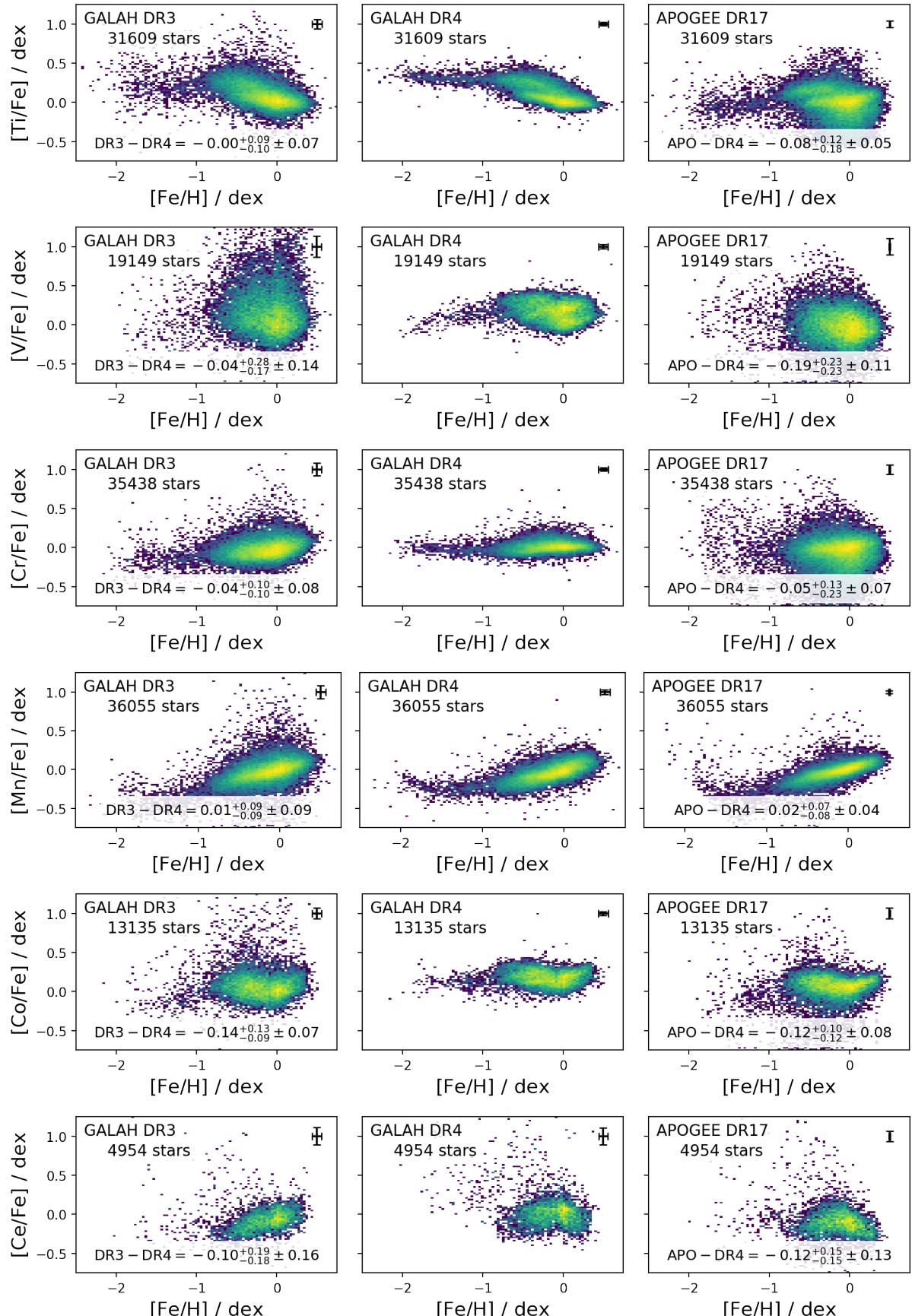
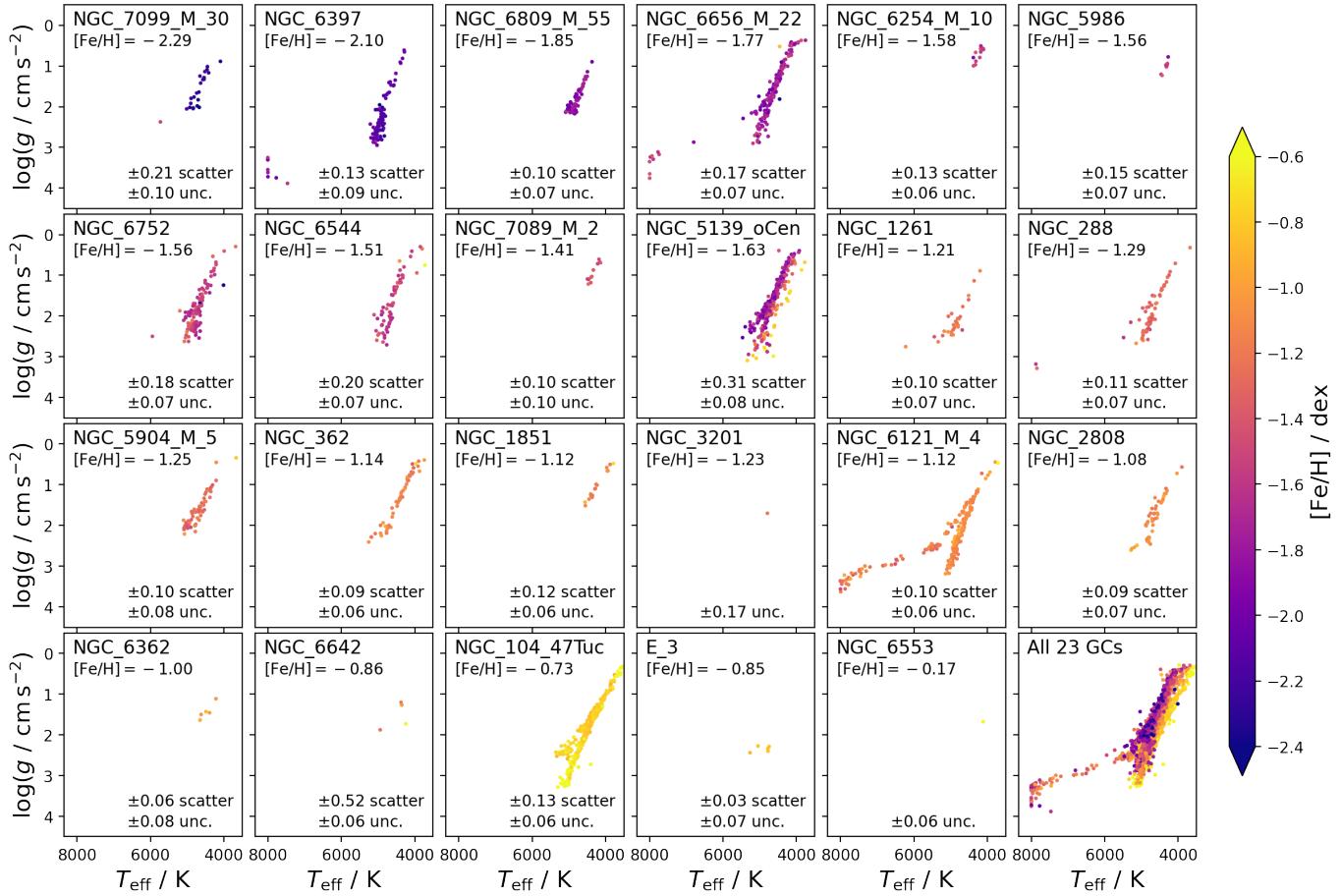
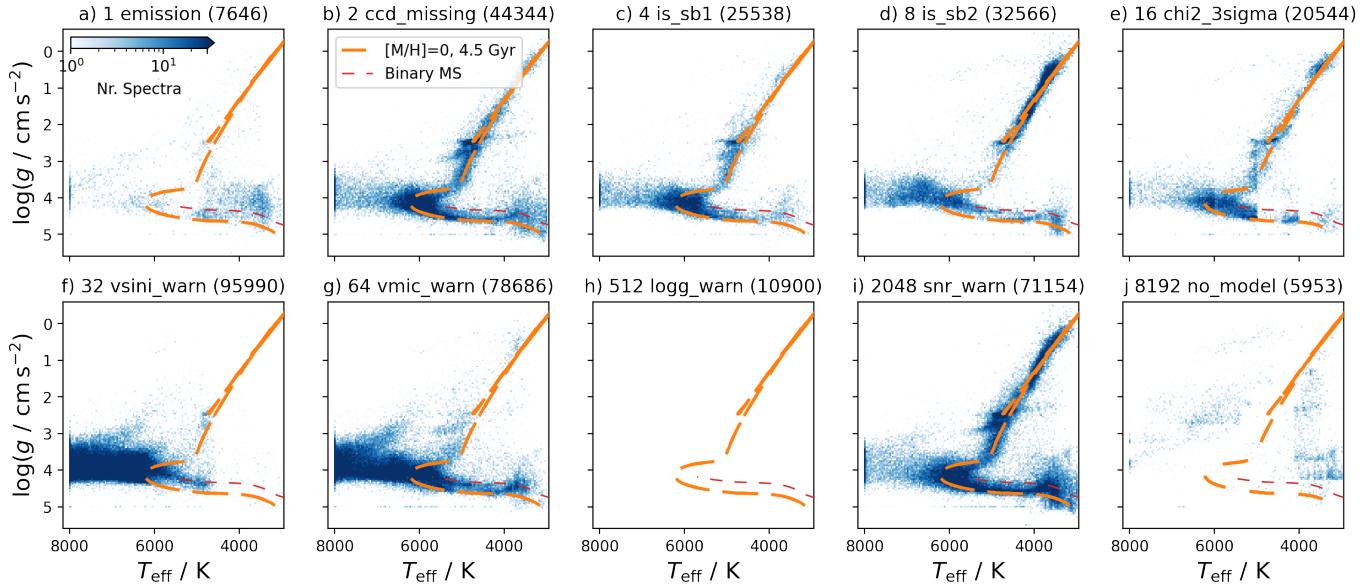


Figure 35. Continuation of Fig. 34 for Ti, V, Cr, Mn, Co, and Ce.



**Figure 36. Collage of globular clusters in the  $T_{\text{eff}}$ - $\log g$  space, coloured by stellar metallicity  $[\text{Fe}/\text{H}]$ . There are no strong systematic 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 37. Parameter overview of stars with raised major quality flag  $\text{flag}_{\text{sp}}$  for  $\text{allstar}$ .** Each panel shows the logarithmic density distribution of stars in the  $T_{\text{eff}}$  and  $\log g$  plane with blue colormaps. 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).