

RESEARCH PAPER

The GALAH Survey: Data Release 4

S. Buder,^{1,2} J. Kos,³ E. X. Wang,^{1,2} M. McKenzie,^{1,2} M. Howell,^{4,2} S. L. Martell,^{5,2} M. R. Hayden,^{6,2} D. B. Zucker,^{7,8,2} T. Nordlander,^{1,2} G. Traven,³ J. Bland-Hawthorn,^{6,2} G. M. De Silva,⁸ K. C. Freeman,^{1,2} G. F. Lewis,⁶ K. Lind,⁹ S. Sharma,¹⁰ J. D. Simpson,^{5,2} D. Stello,^{5,6,2,11} T. Zwitter,³ K. Banks,^{5,2} K. Beeson,³ I. Ciucă,^{1,2} M. K. Ness,^{1,2} A. Rains,²² R. Vogrinčič,³ and The GALAH Collaboration³³

¹Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

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

³Faculty of Mathematics & Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

⁴School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia

⁵School of Physics, UNSW, Sydney, NSW 2052, Australia

⁶Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006, Australia

⁷School of Mathematical and Physical Sciences, Macquarie University, Sydney, NSW 2109, Australia

⁸Macquarie University Research Centre for Astronomy, Astrophysics and Astrophotonics, Sydney, NSW 2109, Australia

⁹Department of Astronomy, Stockholm University, AlbaNova University Centre, SE-106 91 Stockholm, Sweden

¹⁰Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD, 21218, USA

¹¹Stellar Astrophysics Centre, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

²²Department of Physics and Astronomy, Uppsala University, Box 516, SE-75120 Uppsala, Sweden

³³<https://www.galah-survey.org>

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

(Received dd Mmm 2024; revised dd Mmm 2024; accepted dd Mmm 2024; first published online dd Mmm 2024)

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 bar codes, we can extract the chemical fingerprint of stars from high-resolution spectroscopy. The fourth data release (DR4) of the Galactic Archaeology with HERMES (GALAH) Survey provides the chemical abundances of up to 31 elements for 917 588 stars that also have exquisite astrometric information from the *Gaia* satellite.

For this data release, we utilise neural networks to simultaneously fit all stellar labels (stellar parameters and elemental abundances) based on the full wavelength range. The neural networks are trained on grids of synthetic spectra that are computed with Spectroscopy Made Easy with atomic line formation calculated in non-local thermodynamic equilibrium for 14 elements. In a first iteration, we fit the best set of labels and in particular radial velocities for all 1 085 520 spectra, including repeat observations. These are used to properly co-add spectra (for example of spectroscopic binaries without line-splitting) for the second major iteration. This loop uses the distances inferred from astrometric data of the *Gaia* satellite and 2MASS photometry to self-consistently infer more accurate and precise stellar labels.

Notable improvements of this data release are the newly available abundances of carbon and nitrogen from molecular features as well as the significantly improved precision and number of elemental abundances, in particular for neutron-capture elements.

Some note that higher precision and more daring extraction of abundances through multiple neural networks comes at a risk, as we line out in caveats section.

The combination of chemical compositions across multiple nucleosynthetic paths, orbit information, and age estimates for almost a million stars from GALAH DR4 provides a rich data set for Galactic exploration.

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 C₂ 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 observations 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 guide line 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 of our data in Sec. 6 and validate our measurements in Sec. 7. The data products of this data release are explained in Sec. 8. We describe identified caveats in Sec. 9 and make suggestions for minimising them in the future, before concluding this manuscript with thought in Sec. 10.

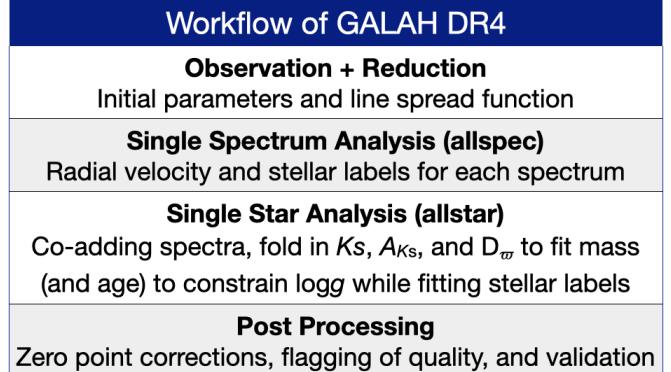


Figure 1. Workflow of GALAH DR4

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 \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.

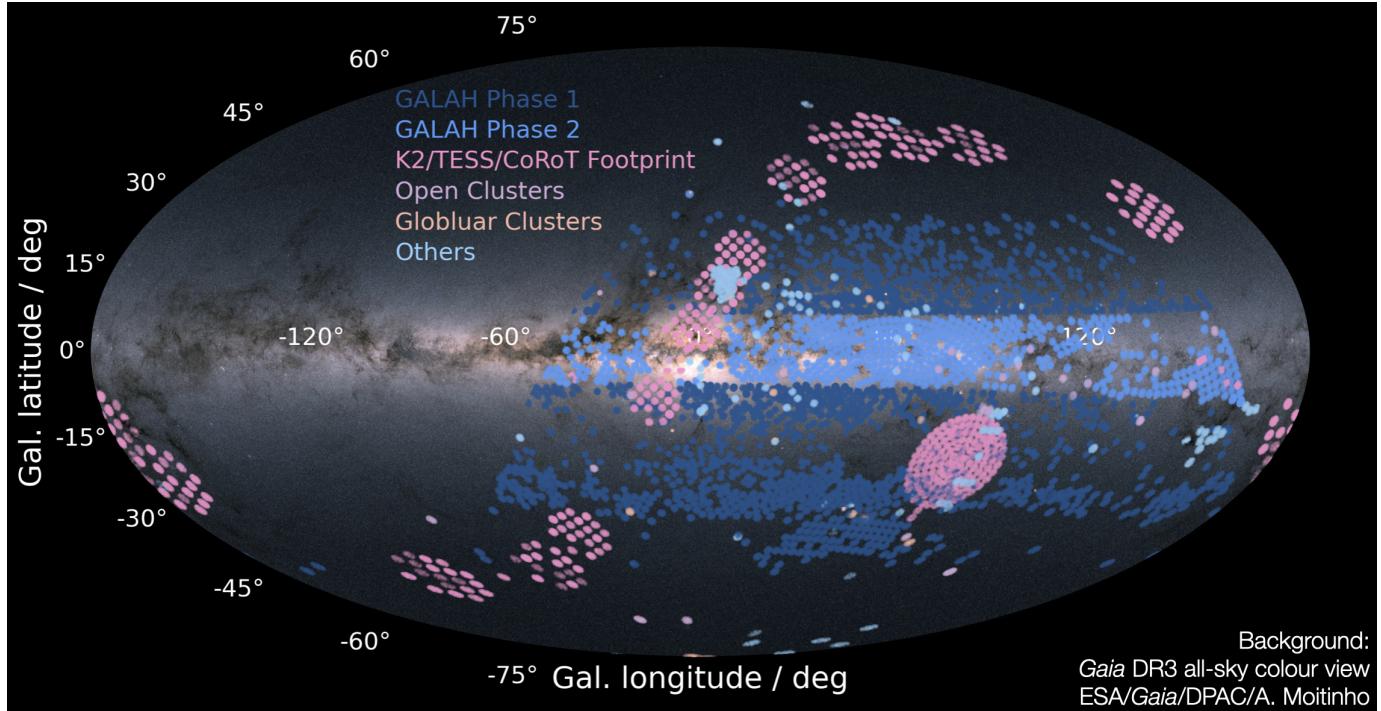


Figure 2. Overview of the distribution of stars included in this fourth GALAH data release in Galactic coordinates with the centre of the Galaxy at the origin and the *Gaia* DR3 all-sky colour view (Gaia Collaboration et al., 2023) as background. Shown are the targets of GALAH Phase 1 (dark blue) and 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.

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

GALAH Phase 2 extended these times to 3×10 and 3×30 minutes, respectively, and included repeat observations of

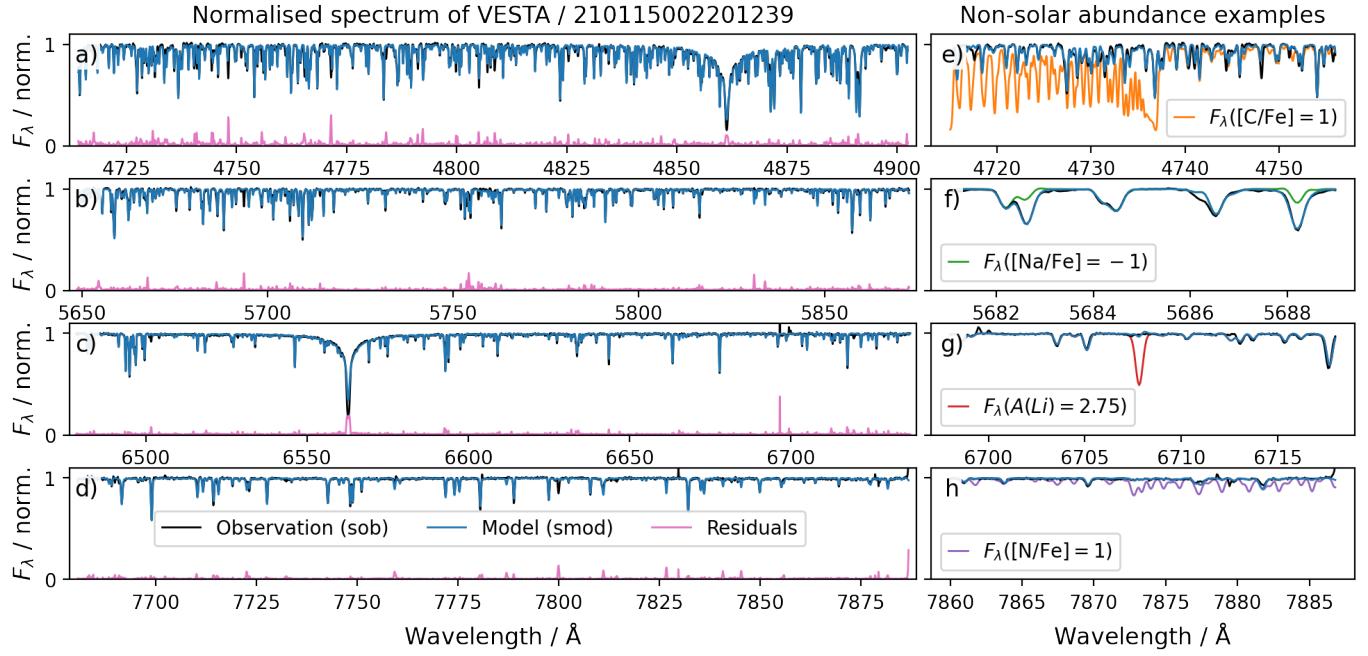


Figure 3. Comparison of normalised observed (black) and synthetic spectra (blue) of the asteroid 4 Vesta with solar composition as well as examples of synthetic spectra with non-solar abundances. Panels a-d) show the observed and best-fit synthetic spectrum as well as their absolute residual (pink) for 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 C₂ 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 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 in purple.

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 ≥ 5 stars) and 361 open clusters (109 with ≥ 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

GALAH Phase 1 main targets with another 3×15 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 ($2 < \text{seeing} < 2.5 \text{ arcsec}$) and 3 ($2.5 < \text{seeing} < 3 \text{ arcsec}$) 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 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 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

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

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 $fwhm$ for each wavelength point in the spectrum, that is

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

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 $fwhm$ are saved in extension 7 with b saved in the headers.

To achieve signal-to-noise ratio of the individual exposures depends on the spectral types, reddening, and observational conditions. Especially the repeat observations of previous pointings has increased the signal-to-noise 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.

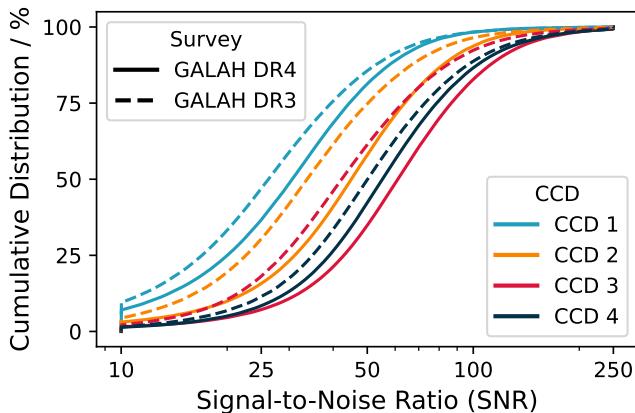


Figure 4. Cumulative Distribution Function (CDF) of the logarithmic Signal-to-Noise Ratio (SNR) for the 4 CCDs of the HERMES spectrograph comparing GALAH DR4 (solid lines) and GALAH DR3 (dashed lines).

2.3 Auxiliary data from Gaia, 2MASS, and literature

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

2.3.1 Gaia DR3

We crossmatch our observations to the *Gaia* DR3 catalogue (Gaia Collaboration et al., 2021a, 2023) using the 2MASS ID, via the nearest neighbour crossmatches provided as part of *Gaia* DR3 (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., 2022). 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 $\sigma_{\varpi}/\varpi = 1.6^{+2.6}_{-0.9}\%$. 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_{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 cancel out 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. 9.

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 lay out 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

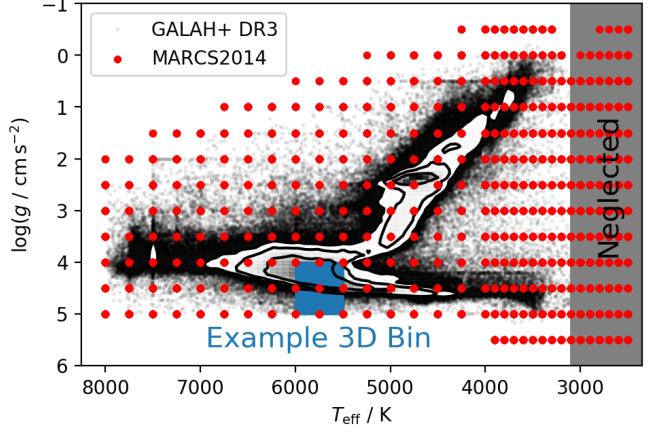


Figure 5. Coverage in T_{eff} and $\log g$ of the MARCS2014 grid (red) and GALAH DR3 (black, including density contours). Shown is also an example of one of the 3D bins used to create models with *The Cannon*. MARCS grid points $T_{\text{eff}} < 3100 \text{ K}$ or $[\text{Fe}/\text{H}] < -3 \text{ dex}$ 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_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$ within ± 1 grid points in T_{eff} (with either ± 250 or $\pm 100 \text{ K}$), $\log g$ ($\pm 0.5 \text{ dex}$), and $[\text{Fe}/\text{H}]$ (± 0.5 or $\pm 0.25 \text{ dex}$). An example box is shown for Solar siblings as a blue box in Fig. 5, which is centred on $T_{\text{eff}} = 5750 \pm 250 \text{ K}$, $\log g = 4.5 \pm 0.5 \text{ dex}$ and $[\text{Fe}/\text{H}] = 0.0 \pm 0.25 \text{ dex}$.

Within these bins we sample 280^a 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 $T_{\text{eff}} = 5750 \pm 250 \text{ K}$, $\log g = 4.5 \pm 0.5 \text{ dex}$ and $[\text{Fe}/\text{H}] = 0.0 \pm 0.25 \text{ dex}$ (see blue box in Fig. 5).

Stellar parameters (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, v_{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.

v_{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 $\log g' = \log g - 4.0$

^aThis number is chosen to match the 28 CPUs of our computing nodes.

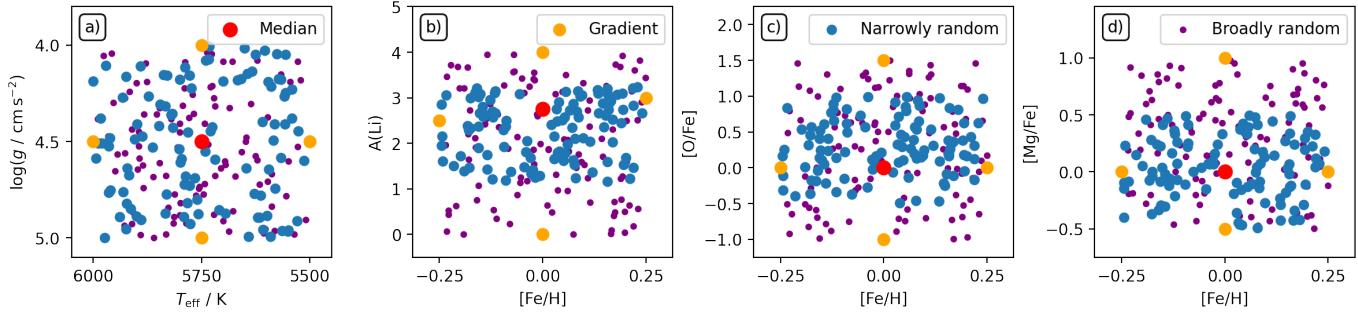


Figure 6. Coverage of stellar parameters and abundances for one of the 3D bins. Shown is the example of the Solar 3D bin ($T_{\text{eff}} / \text{K} = 5750$, $\log g / \text{dex} = 4.5$, $[\text{Fe}/\text{H}] / \text{dex} = 0.0$). **Panel a):** T_{eff} and $\log g$. **Panel b):** $[\text{Fe}/\text{H}]$ 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_{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

to return:

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

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).

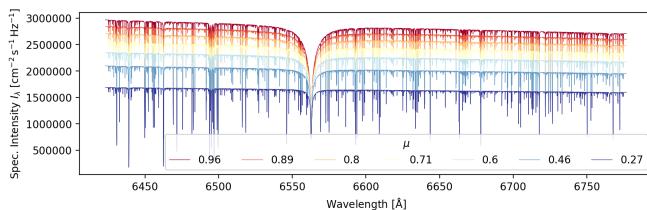


Figure 7. Example output of SME for a solar spectrum in HERMES CCD3 (around the Balmer H_{α} line). Shown are the specific intensities (sme.sint) as a function of the equal-area midpoints of each equal-area annulus μ .

For each spectrum, we first run a test on all available lines in the GALAH linelist, which is adapted from Heiter et al. (2021) and includes small changes to correct wrong $\log g / \text{dex}$ values for a few lines within the GALAH wavelength range. We keep all atomic lines for the final synthesis and restrict the molecular lines to those with SME.depth above 0.001.

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_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$. We use grids of non-LTE departure coefficients by Amarsi et al. (2020b) for 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 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 ($\nu_{\text{mac}} = v \sin i = 0 \text{ km s}^{-1}$), 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 μ (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 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^b 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 particular, we use the neural network architecture and training algorithms similar to *The Payne* (Ting et al., 2019) for our data. We describe the connection of stellar labels ℓ and the flux f at each wavelength pixel λ via

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

which encapsulates the so-called layers of a neural network with $i = 300$ neurons and where we use the default leaky Rectified Linear Unit (lReLU)

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

After optimising the loss function for 10^4 steps, we consider the network trained with an optimised combination of three sets of weights and biases within the minimum and maximum ranges of each label. We discuss this caveats of this particular neural network architecture and training setup in Section 9. The trained networks can then be used with new input labels to quickly create synthetic spectra for the label optimisation.

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 optimisation. This allows us to identify shifts in radial velocity between separate spectroscopic observations of the same star^c and a more accurate

^bFor a more detailed discussion on the advantages of neural networks for predicting spectra see Ting et al. (2019).

^cWhile 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.

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 optimisation 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. In an important improvement of our analysis, we perform an on-the-fly re-normalisation of the observed spectrum, which allows a more accurate comparison of synthetic and observed spectra (Sec. 4.2) and thus a more accurate stellar label optimisation (see Sec. 4.3).

4.1 Initial stellar labels

Initial values of all stellar labels are needed for creating a first synthetic spectrum. For ν_{rad} , T_{eff} , $\log g$, and $\nu \sin i$ 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 $T_{\text{eff}} < 4000 \text{ K}$ as well as the hottest stars with $T_{\text{eff}} > 6500 \text{ K}$, 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 ν_{mic} based on Eq. 4 and limit rotational broadening values to $3 \leq \nu \sin i \leq 10 \text{ km s}^{-1}$ for stars below $T_{\text{eff}} = 5500 \text{ K}$ and $3 \leq \nu \sin i \leq 20 \text{ km s}^{-1}$ for hotter stars. The explicit choices of starting values for T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, ν_{mic} , and $\nu \sin i$ are described in our online repository and are depicted in Fig. 32.

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

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 observations. 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_{λ} and $\sigma_{f,\lambda}$.

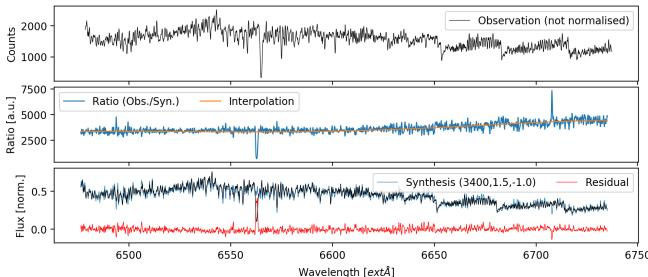


Figure 8. Example of normalisation for GALAH DR4 for a model spectrum that is selected during the label optimisation. Panel (a): Observed spectrum (counts). Panel (b): Ratio (blue) of observed spectrum and model spectrum as well as Chebyshev polynomial fit (orange). Panel (c): Normalised observed spectrum (black) compared to the model spectrum (blue). Residuals (red) can then be used as input for the likelihood function.

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_{eff} as well as v_{rad} .

4.3 Stellar label optimisation

In up to four major loops, we optimise the radial velocities and all other stellar labels and report a) their values, b) their 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 optimisation.

Starting from the initial values, a first synthetic spectrum is computed and compared with the observation in order to assess the initial radial velocity. This is done by applying the SCIPY.SIGNAL.FIND_PEAKS algorithm on the normalised inverse residuals of non-shifted observed and synthetic spectra, when the latter is shifted by $v_{\text{rad}} = -1000..(2)..1000 \text{ km s}^{-1}$ (see Fig. 9a). If no peak is found, the initial v_{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_{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.

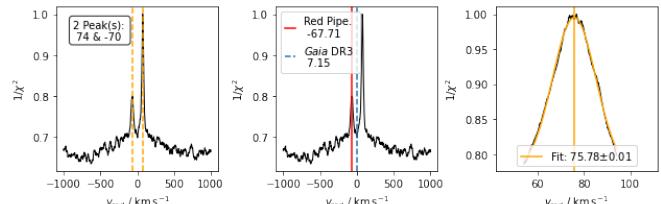


Figure 9. Output of the radial velocity fitting module. Panel a) shows the initial broad search on a v_{rad} array of $-1000..(2)..1000 \text{ km s}^{-1}$. In the case of 2MASS J06084657-7815235, two peaks (yellow, dashed) are visible for this double-lined spectroscopic binary. Panel b) shows the same plot, but overlaid with the GALAH DR4 reduction pipeline (red) and Gaia DR3 (blue, dashed) estimates of v_{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 centerpiece of our optimisation is the SCIPY.OPTIMISE module’s CURVE_FIT function (Virtanen et al., 2020), which we call with counts and uncertainties (our absolute sigmas) as input for a placeholder function that self-consistently re-normalises the observed spectrum. We estimate the labels via the least squares optimisation within less than 10^4 iterations and a desired relative error ($xtol$) below 10^{-4} .

For each optimisation loop, a new, best-fit 3D bin and neural network is identified via a grid search in the T_{eff} , $\log g$, and [Fe/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 $[X/\text{Fe}] = 0$.

While the optimisation has not converged (the final parameters T_{eff} , $\log g$, and [Fe/H] are not within the current 3D

bin), the optimisation is repeated, starting with the previous best-fit parameters as starting guesses.

4.3.1 Which labels are optimised?

As part of the synthetic grid computations, we have perturbed each label 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 Li I 6708, the second case is addressing in particular molecular lines like the C₂ and 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 optimisation 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 A(Li), [X/Fe] = -0.5..1.0 for C, N, O, Y, Ba, La, Ce, and Nd, and [X/Fe] = -0.5..0.5 for all other elements before optimising them again. For warm and hot stars ($T_{\text{eff}} > 6000$ K), this effect was seen to affect multiple 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, $v_{\text{mic}} = 1.06 \text{ km s}^{-1}$, $v \sin i = 1.6 \text{ km s}^{-1}$, $v_{\text{mac}} = 4.2 \text{ km s}^{-1}$ (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^d 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 5σ or a flux of 0.3 (0.4 before the initial optimisation). To avoid the masking of lines that are vital for our spectroscopic analysis, we have created a list^e with segments of such lines that is mainly based on the previous element masks from GALAH DR3 (Buder et al., 2021). The final mask of pixels to use for the optimisation then includes all vital line regions, as well as those wavelengths that do not show a too strong disagreement between observation and synthesis and are not deemed unreliable in their synthesis.

In addition to this default masking, we exclude pixels for each major iteration, for which the flux of observation and synthesis differ by more than 5σ 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. This approach 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 2σ . 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.

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

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

While we can use our spectroscopically determined T_{eff} in Eq. 8, the other values have to be estimated through models

^dExample masks can be found in the GALAH DR4 repository here.

^eThe list is available in the GALAH DR4 repository here.

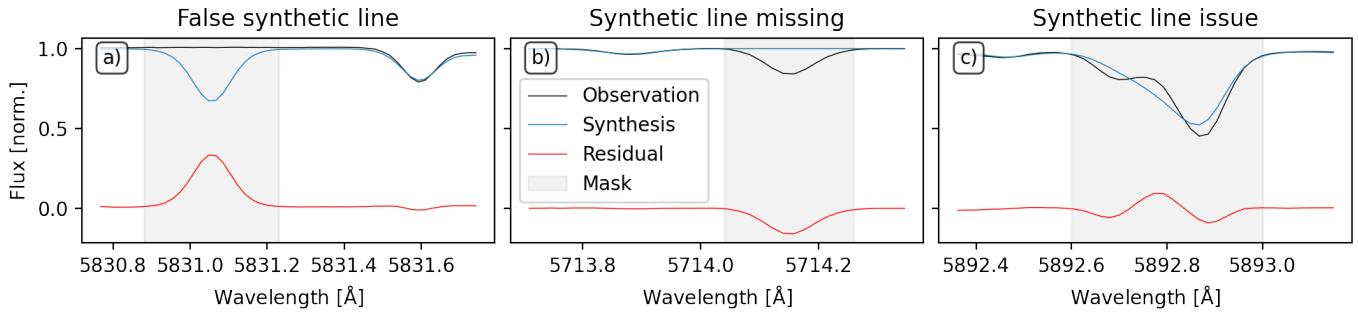


Figure 10. Examples of masks applied to unreliable pixels for the spectrum fitting, which is done by the minimisation of residuals (red) between observation (black) and synthesis (blue). Panel a) shows a strong synthetic line, where no line is observed in the data. Panel b) shows an observed line without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

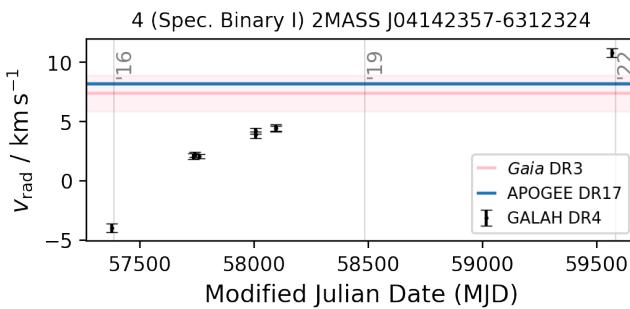


Figure 11. Example of radial velocity evolution over time for a single-lined spectroscopic binary (SBI).

or non-spectroscopic information. The logarithmic bolometric luminosity, L_{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_∞ , 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,

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

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^f 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).

^fWe 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 $-1 \leq [\text{Fe}/\text{H}] \leq 0$.

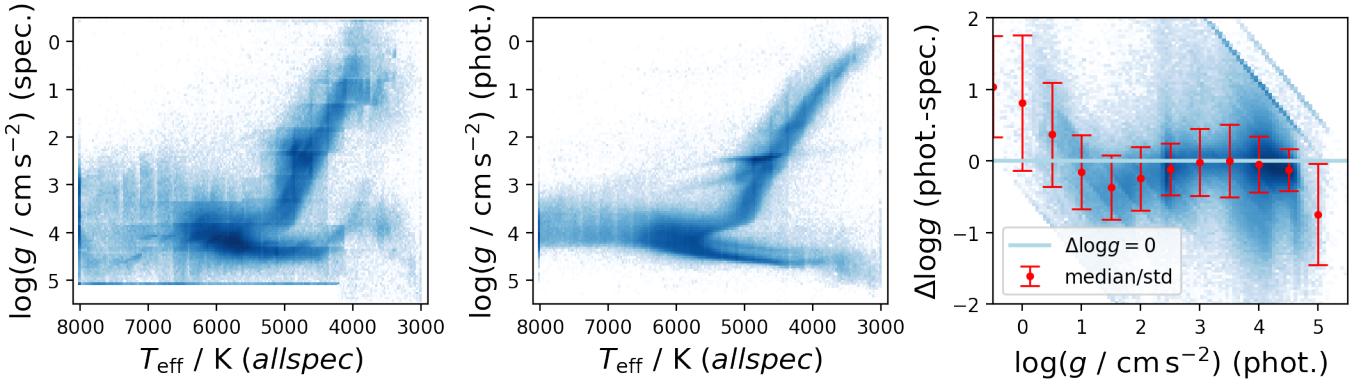


Figure 12. Comparison of spectroscopic and photometric $\log g$ estimates in the *allspec* analysis **Panel a)** shows the distribution of spectroscopic $\log g$ and T_{eff} from the *allspec* module. **Panel b)** shows the distribution of the same T_{eff} and photometric $\log g$. **Panel c)** shows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

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 χ^2 distribution as a function of radial velocity shifts (see Fig. 9). While a single star would only show one peak (saved as `rv_comp_1`), a binary system like 2MASS J06084657-7815235 shows a second peak (-70 km s^{-1} in addition to 74 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., 2022) 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.^g For 6% (50 577) of our stars, we find a difference with respect to *Gaia* DR3 larger than 10 km s^{-1} . For these stars, as well as stars for which we estimate unrealistic v_{mic} and $v \sin i$ below 0 km s^{-1} in *allspec*. We note that the *allspec* analysis was run without boundary conditions for global parameters and thus also resulted in negative velocities, which are later flagged. *allstar*, however, was run with v_{mic} and $v \sin i$ forced to be above 0 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 signal-to-noise. 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_{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. 9. 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.

^gKL: Expand on RV – this is important science in itself to many. Is there a plot of agreement?

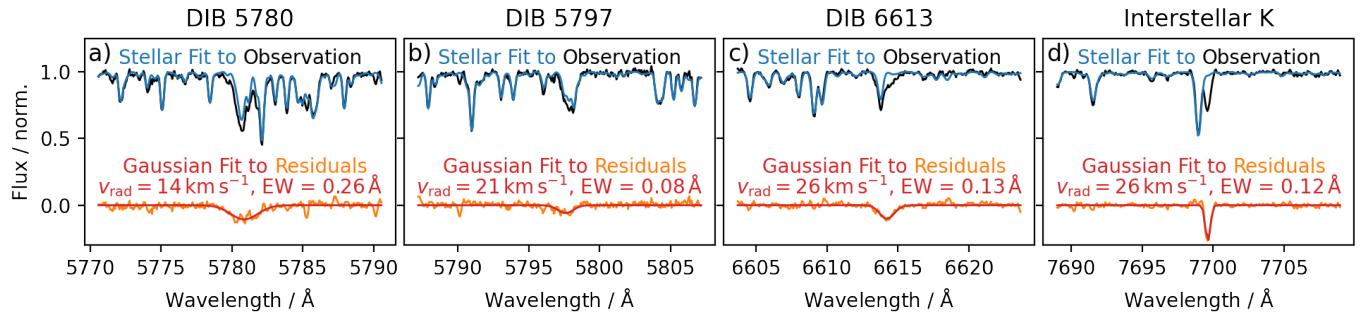


Figure 13. Example of three diffuse interstellar bands (DIBs) and interstellar K absorption for 2MASS J06453479-0102137 with an $E(B-V) = 0.84$ mag value from Schlegel et al. (1998). Shown are the observation (black) and stellar fit (blue) as well as a Gaussian fit (red) to the residual (orange), resulting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

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 Å) with central wavelengths identified by Vogrinčič et al. (2023) as well as for interstellar K (7698.9643 Å), 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 Å whose values we report in `ew_h_beta` and `ew_h_alpha`. By default we integrate in a window of ± 0.75 and 1.25 Å for H_β and H_α , respectively, and increase this window to 5 Å if the average observed, normalised flux within ± 0.5 Å 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 Å.

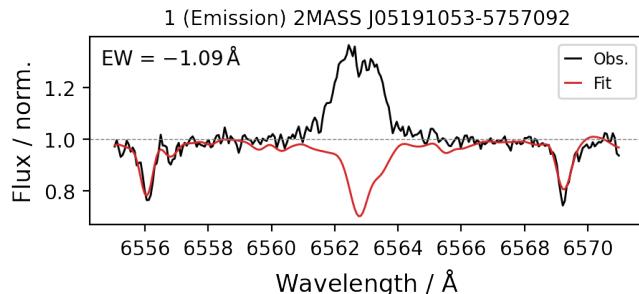


Figure 14. Example of a star with clear emission in the Balmer lines (here H_α). The bit indicating emission has been raised in the major quality flag `flag_sp` for this star.

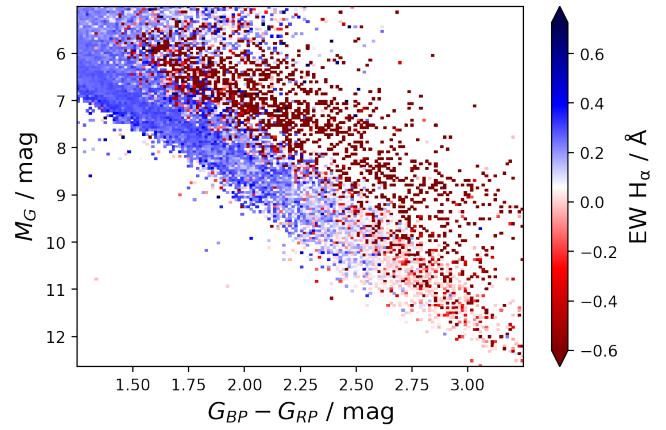


Figure 15. Distribution of residual equivalent width between observation and synthesis in the *Gaia* DR3 color-absolute magnitude diagram ($G_{\text{BP}} - G_{\text{RP}}$ vs. M_{G}) for cool dwarfs. Balmer lines in emission result in negative equivalent width values, as is the case for most young stars that are more luminous than the cool main sequence. Should we point out the interesting sequence of binary MS around $M_{\text{G}} \sim 7$ mag? KL: over-density of equal mass binaries?

6.2 Uncertainty estimation and validation

The uncertainties that we report for our spectroscopic data analysis are based on the covariance matrix of the fitting process. While we report the covariance matrix for each spectrum (see Fig. 23 for the covariance matrices of Vesta and Arcturus), we restrict ourselves to the standard deviations of each feature (square root values of the diagonal variance entries).

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. Subsequently, we describe the numerous comparisons that we have performed for both stellar parameters (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, ν_{mic} , $\nu \sin i$, and ν_{rad}) as well as the elemental abundance measurements. Due to the limited coverage of benchmark literature, we continue to apply a single accuracy estimate for each stellar parameter and no accuracy estimate for the elemental abundances (as for GALAH DR3 Buder et al., 2021).

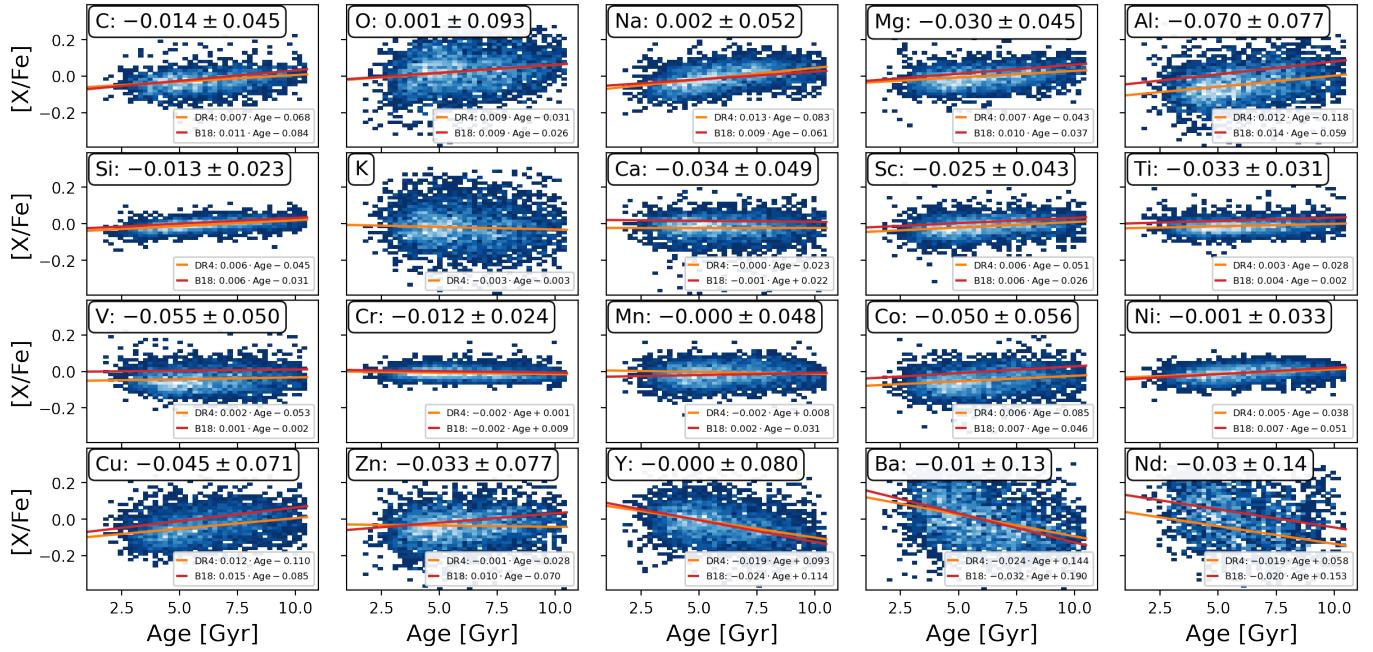


Figure 16. Chemical abundances [X/Fe] of Solar twin stars as a function of ages that were estimated as part of the mass and age estimation of the allstar spectrum analysis. We overplot linear fits to our age-abundance relations for Solar twins in orange as well as the literature values from Bedell et al. (2018) in red.

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 Fig. 17). Precision values are extracted from the median covariance uncertainties for stars with $\text{SNR} = 50 \pm 10$ in CCD2 (see Fig. 21).

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

Fig. 17 for stellar parameters and Fig. 19 for abundances.
Zeropoints in Table 5

- Gaia FGK Benchmark stars (Jofré et al., 2018) Fig. 17a–c
- IRFM (Casagrande et al., 2020)
- K2/TESS Zinn et al. (2020) Fig. 17b
- APOGEE DR17 Fig. 17a-f
- Gaia DR3 for ν_{rad} Fig. 17f
- Globular Clusters for accuracy at low [Fe/H] accuracy Fig. 20

- 47 Tuc Carretta et al. (2009c,a)
- NGC 6362
- M4
- NGC 362 Monty et al. (2023)?
- NGC 288 Monty et al. (2023)?
- NGC 1851
- M2 NGC 7089, with 3 populations in [Fe/H] accord-

ing to Yong et al. (2014). 21331790-0048198 was even observed by Yong et al. (2014) as NR 79 / HI-104 with $[\text{Fe}/\text{H}] = -1.69$. We get -1.37 ± 0.26 , but at SNR around 10.

- Omega Cen Johnson & Pilachowski (2010)
- NGC 6397 Carretta et al. (2009c,a)
- M30 Carretta et al. (2009c,a)
- Abundance Zeropoints
 - Vesta
 - APOGEE DR17
 - Solar twins
 - M 67, Ruprecht 147 and a few others cover stars across several evolutionary stages. Check them roughly (could also be problematic because of atomic diffusion etc.)

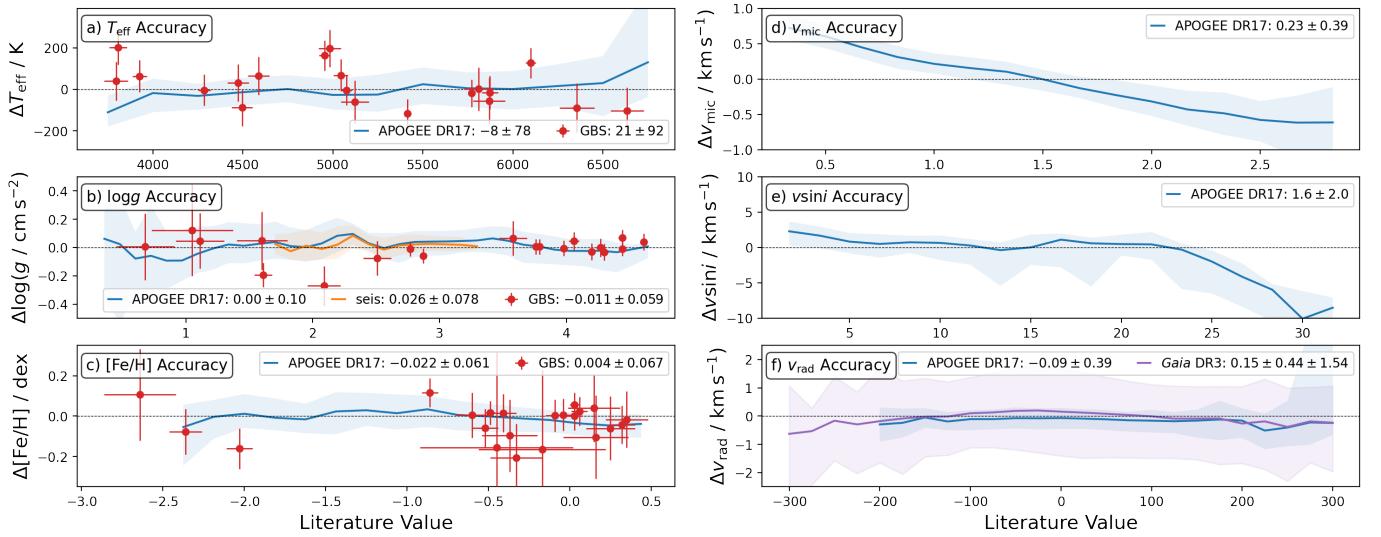


Figure 17. Accuracy of the main stellar parameters T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, v_{mic} , $v \sin i$, and v_{rad} for GALAH DR4. Each panel shows the comparison to literature (DR4 - literature). Comparisons are performed for the *Gaia* FGK Benchmark stars (red), APOGEE DR17 (blue), $\log g$ inferred from asteroseismic measurements (orange) and *Gaia* DR3 radial velocities (purple).

Table 5. 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.

Property	Reference	Zeropoint	Shift	Vesta	Vesta Shift
T_{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_{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	-	-

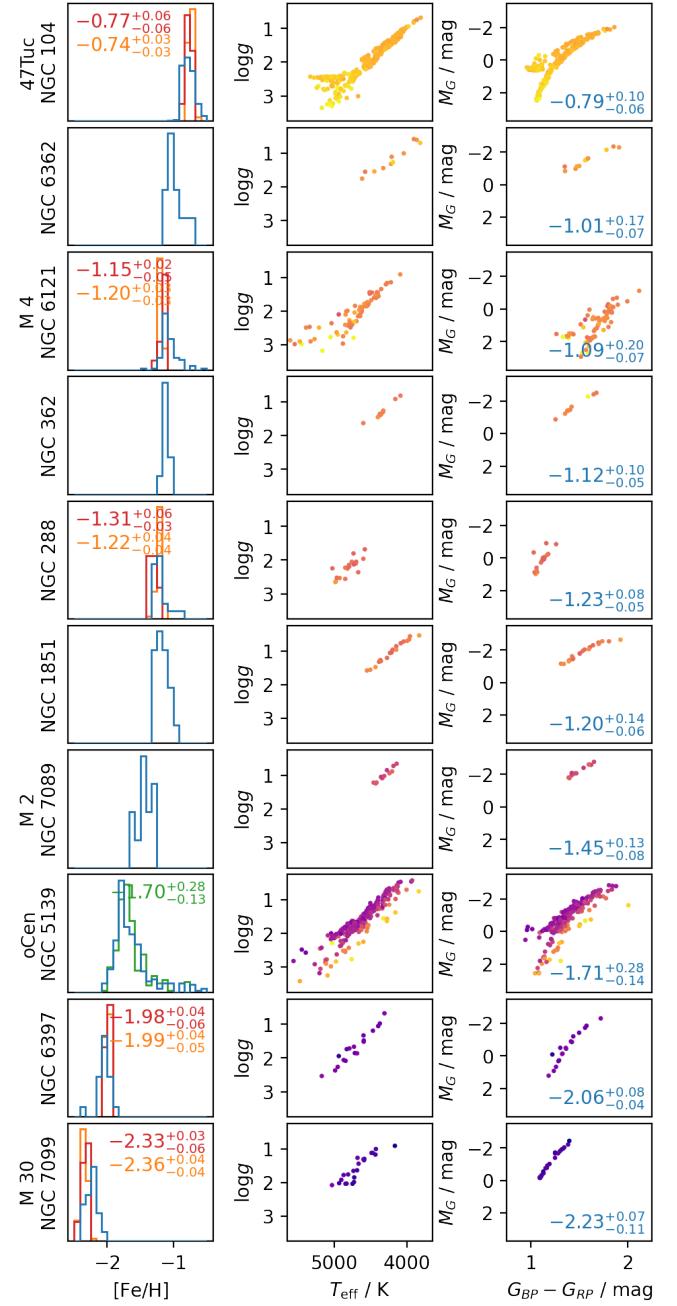


Figure 18. Comparison of iron abundances (16th, 50th and 84th percentiles) and overview of spectroscopic and photometric properties of globular cluster stars in GALAH DR4. Left panels show histograms of iron abundances from GALAH DR4 (blue) as well as literature estimates for the globular clusters from Giraffe (orange) and UVES (red) observations by (Carretta et al., 2009c,b) as well as observations from Johnson & Pilachowski (2010). Middle panels show the spectroscopic $T_{\text{eff}}-\log g$ diagrams colored by iron abundance [Fe/H]. Right panels show the Gaia DR3 color absolute magnitude diagrams.

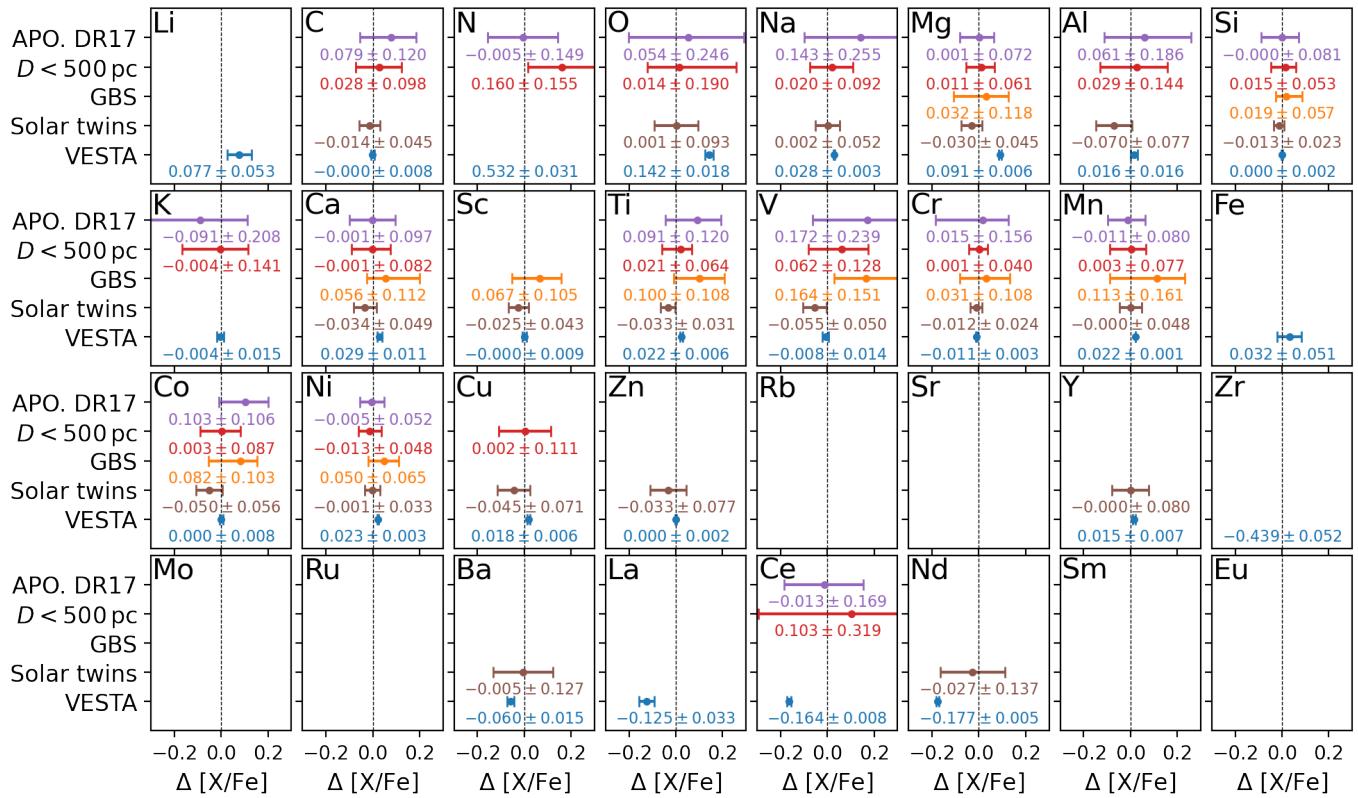


Figure 19. Zeropoint estimates of elemental abundances for GALAH DR4.. Each panel shows the comparison to literature (DR4 - literature) for Vesta (blue), Gaia FKG Benchmark Stars (orange), Stars with $|[\text{Fe}/\text{H}]| \leq 0.1$ closer than $D_{\odot} < 0.5$ kpc (red), as well as stars that were also observed by APOGEE DR17 (purple).

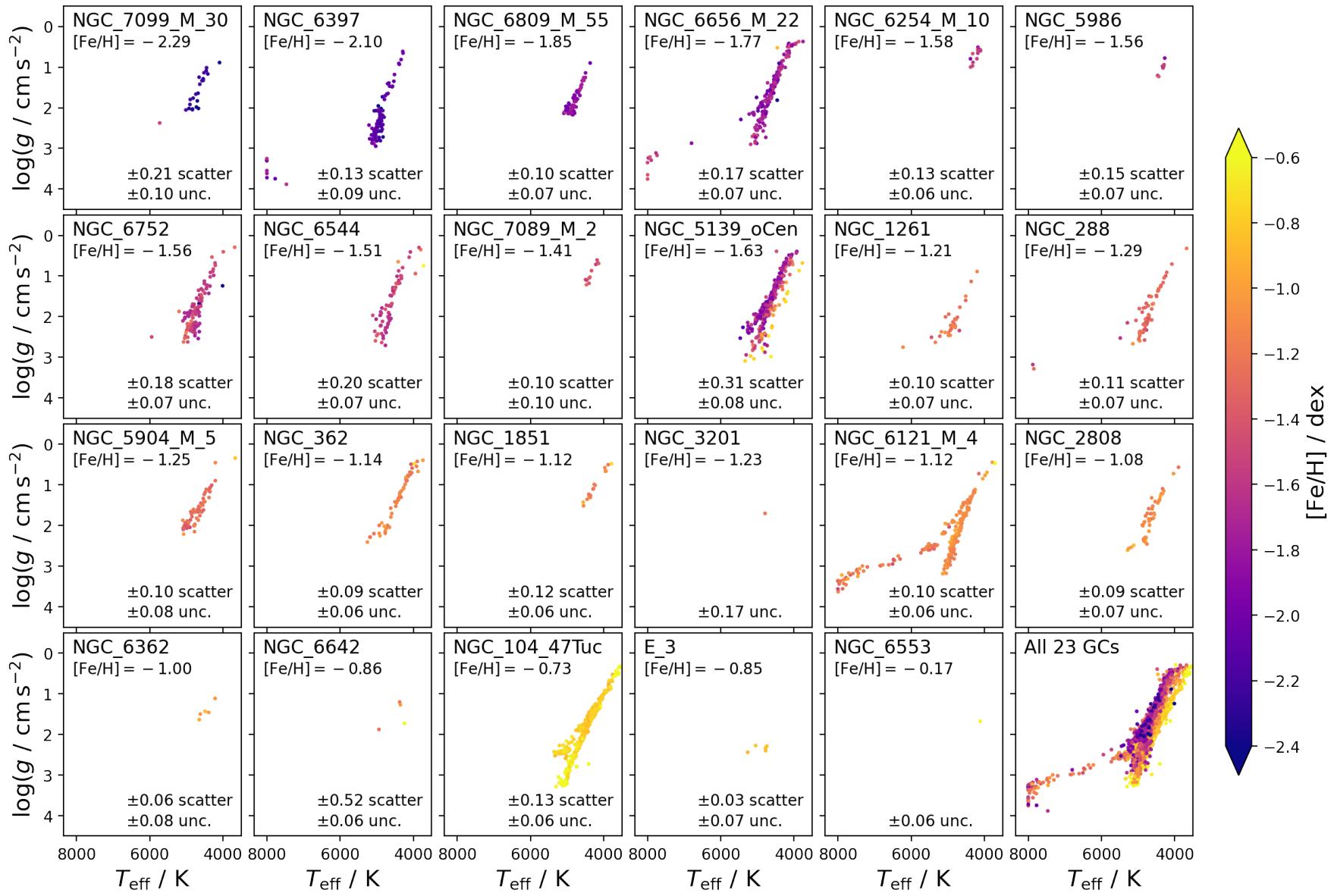


Figure 20. Collage of globular clusters in the T_{eff} - $\log g$ space, coloured by stellar metallicity $[Fe/H]$. There are no strong systematic trends between $[Fe/H]$ and T_{eff} , even for the horizontal branch stars in NGC 288, NGC 6656 (M22), and NGC 6121 (M4). NGC 5139 (!Cen) shows a significant range in $[Fe/H]$. RMS scatter and median metallicity uncertainties for each cluster are given in the lower right of each panel. Most other clusters are consistent with no metallicity spread beyond the measurement uncertainties.

6.2.2 Precision estimation and validation

- Fig. 21 and 22
- Create Table with overall scatter values for individual elements based on:
- Covariance uncertainties
- Repeat observations
- Scatter in open clusters
- Wide binary scatter

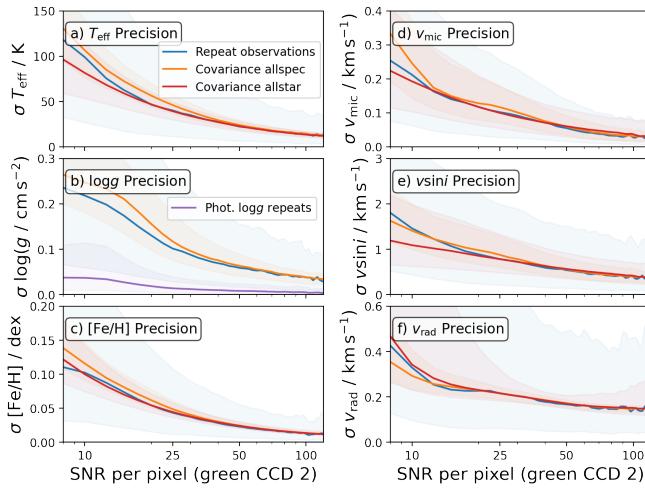


Figure 21. Precision monitoring of stellar parameters as a function of SNR for the green CCD2 across GALAH DR4.. Each panel shows the behaviour for bins of width 10 for the scatter of repeat observations of the allspec runs (blue), covariance uncertainties of allspec (orange) and allstar (red) setups as well as scatter of photometric $\log g$ from repeat observations (purple).

6.2.3 Covariances

We also calculate and report the covariances of all fitted labels, see Fig. 23 for Vesta and Arcturus.

6.3 Stellar parameter flags flag_sp

We have implemented a number of post-processing routines for quality assurance. Here, we describe the major quality flags of global parameters and overall fitting performance. If a spectrum does not pass one of these tests, a unique bit flag will be raised and added to `flag_sp`.

The description of the implemented bits/flags for `flag_sp` and how often they were raised is listed in Tab. 6 and distributions in the Kiel diagram (T_{eff} and $\log g$) are shown for each raised bit in Fig. 24 for the `allstar` catalogue.

Describe how the $vsini$ and v_{mic} flags are quite informative about binarity. Look at the binary main sequence for that!

SB2 warning currently often activated for luminous RGB stars. Is that maybe caused by detectable extinction? Inspect more examples and figure out what is going on!

Why are the chi2 stars mainly on the RGB? We use a strict cutoff value rather than an SNR dependent one. Maybe implement an SNR dependent on (GBS stars are also all flagged because of this issue)?

Fig. 14 and Fig. 11

In python, the check for a bitmask flag (e.g. 3rd bit with flag value $2^3 = 8$) is best performed via

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

Table 6. 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	allstar	allspec
	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	$\nu \sin i$ warning	138317	95990
6	64	v_{mic} warning	99692	78686
7	128	Triple Binary warning	0	0
8	256	T_{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

6.4 Elemental abundance flags flag_X_fe

Table 7. 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 element X, we produce synthetic spectra with the same stellar labels, but decrease the abundance [X/Fe] to the lower limit of the neural network (or decrease the [Fe/H] value by -0.5 for Fe). The residuals in units of sigma between the best-fit spectrum and spectrum with 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

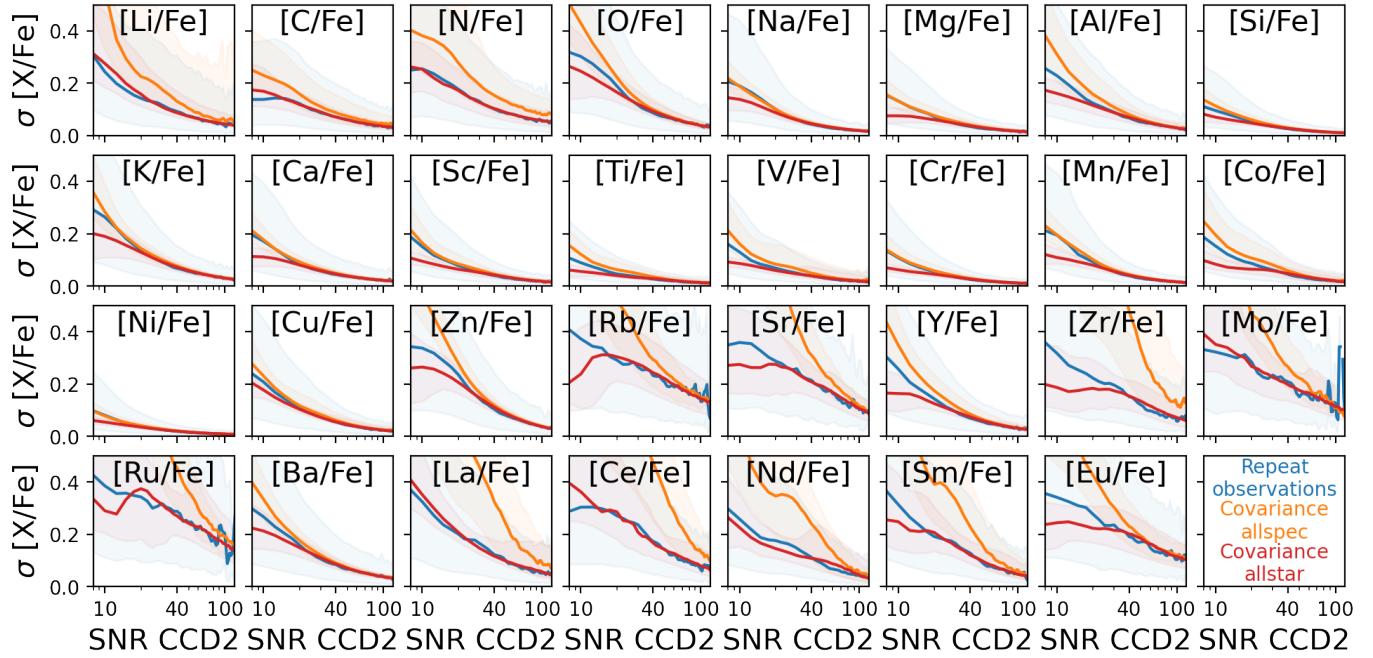


Figure 22. Precision monitoring of elemental abundances as a function of SNR for the green CCD2 across for GALAH DR4.. Each panel shows the behavior for bins of width 10 for the scatter of repeat observations of the allspec runs (blue) as well as covariance uncertainties of allspec (orange) and allstar (red) setups. Missing legend

initial test of overall detectability (Sec. 4.3.1) allowed us identify elements for which not even an upper limit was expected, raising flag FLAG_X_FE by 2.

7. VALIDATION

TESS CVZ asteroseismic info (Hon et al., 2022) and tables from Hon_2022_AJ_164_135.fits

7.1 Automatic Post-Processing Performance

In this section, we try to estimate how reliable the automatic flags in GALAH DR4 are.

Spectra have been classified by eye

- Spectroscopic Binary Type 1 (SB1) classification, Fig. 11
- Spectroscopic Binary Type 2 (SB2) classification, Fig. 31
- Equivalent width estimates of three diffuse interstellar bands (5780.59, 5797.19, 6613.66 Å) with central wavelengths identified by Vogrinčič et al. (2023) as well as for interstellar K (7698.9643 Å), see Fig. 13
- Emission classification and measurement, Fig. 14
- Upper Limit estimation

7.2 Comparison with GALAH DR3 and APOGEE DR17

Fig. 25

7.3 False positive rates for spectroscopic binaries

Binaries from Bashi et al. : 2304.00043 (Bashi et al., 2023)

Thanks to the radial velocity estimates provided for bright stars from the *Gaia* radial velocity spectrometer (Katz et al.,

2022), we have radial velocity estimates for 94% (774 914) of the stars observed for GALAH DR4.

Fig. 27

Binarity. How accurate is the flagging. Compare ratios of true/false, false-positive detections aided by Traven et al. (2020). Get in contact with Alex Wallace and Andy Casey regarding their binarity identification from BP/RP spectra.

7.4 Variable stars in GALAH DR4

We have 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 (δ Scuti, γ Doradus, or SXPhoenicis), 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).

Put list of vari catalogues and their respective publications in here, e.g. *Gaia* DR3 binary star parameters from *gaiadr3.nss_non_linear_spectro.fits* (Halbwachs et al., 2023).

7.5 Minimal Science Verification

Fig. 3 from Spoo et al. (2022): <https://arxiv.org/abs/2203.05463>: logAge vs. [C/N]

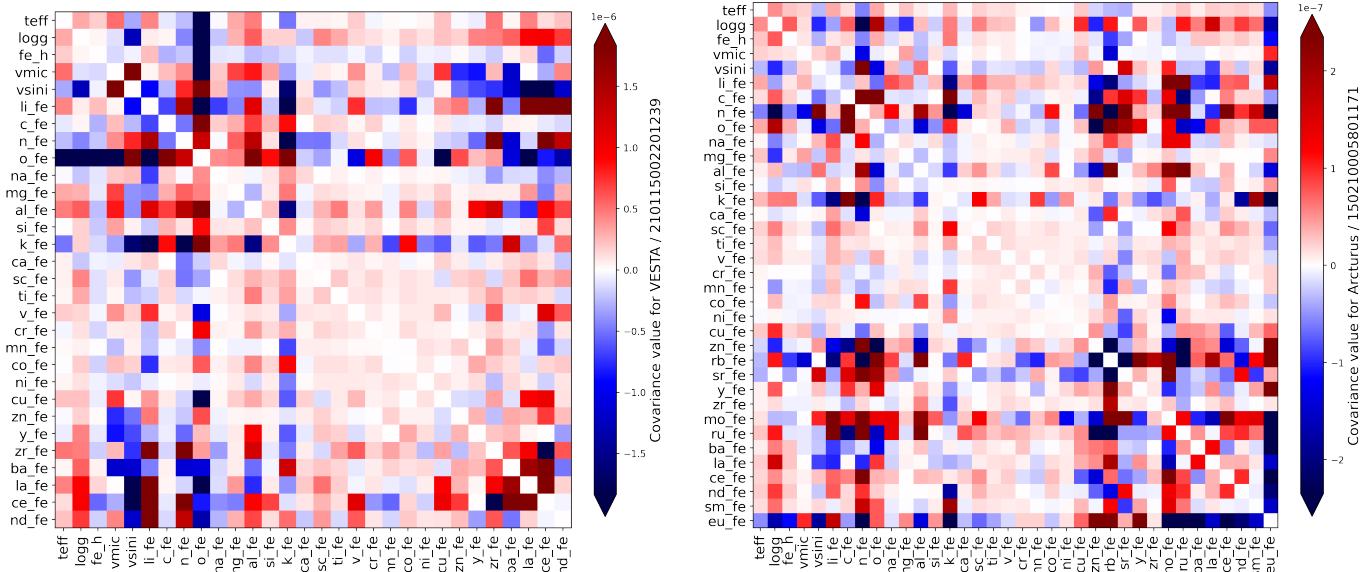


Figure 23. Covariance matrices for labels for Vesta (panel a) and Arcturus (panel b).

8. DATA RELEASE PRODUCTS

8.1 Main Data Release Catalogues

1. galah_dr4_allspec.fits: analysis for each spectrum (incl. RV) assuming single star
2. galah_dr4_allstar.fits: analysis for each star based on co-added spectra for each star and using non-spectroscopic information to constrain $\log g$

Table 8. Table schema of the GALAH DR4 main catalogues.

Keyword	Type	Description	Section
sobject_id	int64	GALAH identifier	
tmass_id	str	2MASS identifier	
gaiadr3_source_id	int64	Gaia DR3 source_id	
survey_name	str	HERMES-2dF Survey/Program/Pointing (other if not available)	
field_id	int32	GALAH Field ID (-1 if not available)	
setup	str	Analysis setup: single/binary/coadds	
mjd	float32	Modified Julian Date	
ra	float64	propagated from Gaia DR3	
dec	float64	propagated from Gaia DR3	
best_spec4star	bool	Best spectrum for given star (identifiable via 2MASS ID)	
flag_sp	int32	Major spectroscopic quality bitmask flag	
flag_sp_fit	int32	Major spectroscopic fitting quality bitmask flag	
opt_loop	int32	Nr of optimisation loops used for fitting	
flag_red	int32	Quality bitmask flag of reduction pipeline	
snr_px_ccdx	float32	Average signal-to-noise ratio (per pixel) of CCDX	
chi2_sp	float32	Chi2 value of spectroscopic fitting	
px_used_perc	int64	Percentage of spectrum used for spectroscopic fit	
model_name	str	Neural network model used for creating synthetic spectra	
closest_model	str	Neural network model closest for $T_{\text{eff}}/\log g/[{\text{Fe}/\text{H}}]$ combination	
comp_time	float32	Computation time spent on spectrum	
flux_contr	float32	Flux contribution of main source	
e_flux_contr	float32	Uncertainty flux_contr	
rv_comp_nr	int64	Number of peaks in RV cross-correlation function (CCF)	
rv_comp_x	float32	Radial velocity of primary or secondary sources	
e_rv_comp_x	float32	Uncertainty of rv_comp_x	
rv_comp_x_h	float32	Height of rv_comp_x in CCF	
rv_comp_x_p	float32	Prominence of rv_comp_x in CCF	
rv_gaia_dr3	float32	Radial velocity in Gaia DR3	
e_rv_gaia_dr3	float32	Uncertainty of rv_gaia_dr3	
v_bary_eff	float64	Barycentric velocity applied to reduced spectra	
teff	float32	Spectroscopic effective temperature (used for fitting)	
e_teff	float32	Uncertainty teff	
logg	float64	Surface gravity adjusted via parallax information	
e_logg	float32	Uncertainty logg_plx	
fe_h	float32	Abundance of Fe and all other elements not fitted in GALAH (Fe: 1D-NLTE)	
e_fe_h	float32	Uncertainty fe_h	
flag_fe_h	int32	Quality flag fe_h	
vmic	float32	Microturbulence velocity (fitted)	
e_vmic	float32	Uncertainty vmic	
vsini	float32	Broadening velocity (fitted sme.vsini with sme.vmac=0)	
e_vsini	float32	Uncertainty of vsini	
x_comp_2	float32	Parameter value x of potential secondary source	
e_x_comp_2	float32	Uncertainty of x_comp_2	
x_fe	float32	Elemental abundance for [X/Fe]	
e_x_fe	float32	Uncertainty x_fe	
flag_x_fe	int32	Quality bitmask flag of x_fe	

Table 9. Table schema of the GALAH DR4 main catalogues. (Continuation of Tab. 8).

Keyword	Type	Description	Section
sb2_rv_16	float32	16th perc. radial velocity of fit to syn-obs residuals	
sb2_rv_50	float32	50th perc. radial velocity of fit to syn-obs residuals	
sb2_rv_84	float32	84th perc. radial velocity of fit to syn-obs residuals	
ew_h_beta	float32	Equivalent Width of fit for syn-obs residuals at Hbeta core	
ew_h_alpha	float32	Equivalent Width of fit for syn-obs residuals at Halpha core	
ew_k_is	float32	Equivalent Width of fit for K7699 Interstellar Line	
sigma_k_is	float32	Sigma auf Gaussian fit for K7699 Interstellar Line	
rv_k_is	float32	Radial velocity of fit to syn-obs residuals around K7699 line	
ew_dibX	float32	Equivalent Width of fit for X Diffuse Interstellar Band	
sigma_dibX	float32	Sigma auf Gaussian fit for X DIB	
rv_dibX	float32	Radial velocity of fit to syn-obs residuals around X DIB	
logg_spec	float32	Spectroscopic surface gravity (used for fitting)	
e_logg_spec	float32	Uncertainty logg_spec	
phot_g_mean_mag	float32	Mean Gaia DR3 G-band apparent magnitude	
phot_bp_mean_mag	float32	Mean Gaia DR3 BP-band apparent magnitude	
bp_rp	float32	Color of BP-RP bands	
h_m	float32	2MASS H-band magnitude	
h_msigcom	float32	Uncertainty of h_m	
ks_m	float32	2MASS Ks-band magnitude	
ks_msigcom	float32	Uncertainty of ks_m	
W2mag	float32	AllWISE W2-band magnitude	
e_W2mag	float32	uncertainty of W2mag	
ebv	float32	Extinction E(B-V)	
parallax	float32	Astrometric parallax used for GALAH DR4	
e_parallax	float32	Uncertainty of astrometric parallax used for GALAH DR4	
parallax_gaia_edr3	float32	Parallax reported with corrections by Gaia EDR3	
e_parallax_gaia_edr3	float32	Uncertainty of parallax reported with corrections by Gaia EDR3	
ruwe_gaia_dr3	float32	RUWE reported by Gaia DR3	
r_med	float32	Median Distance used for calculating logg(plx)	
r_lo	float32	Lower Limit Distance used for calculating logg(plx)	
r_hi	float32	Higher Limit Distance used for calculating logg(plx)	
a_ks	float32	Attenuation in Ks-band A(Ks) used for calculating logg(plx)	
mass	float32	Mass used for calculating logg(plx)	
age	float32	Age estimated when calculating mass	
bc_ks	float32	Bolometric Correction of Ks, BC(Ks), used for calculating logg(plx)	
lbol	float32	Bolometric Luminosity used for calculating logg(plx)	

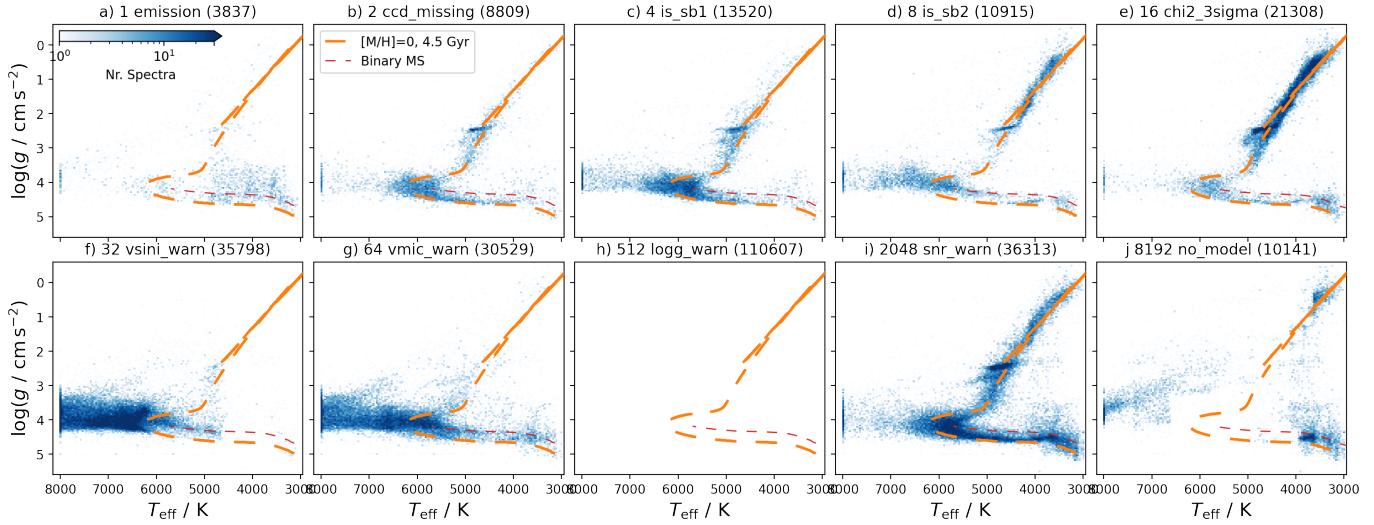


Figure 24. Parameter overview of stars with raised major quality flag `flag_sp` for `allstar`. Each panel shows the logarithmic density distribution of stars in the T_{eff} and $\log g$ plane with blue colormaps. A PARSEC isochrone with $[M/H] = 0$ and $\tau = 4.5$ Gyr is overplotted in orange and the 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).

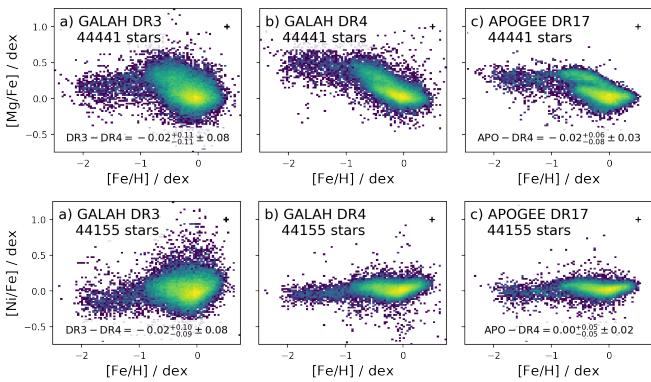


Figure 25. Comparison of stars with measurements for all of GALAH DR3 (left panel), GALAH DR4 (middle panel) as well as APOGEE DR17 (right panel) for $[\text{Mg}/\text{Fe}]$ (top row), $[\text{Ni}/\text{Fe}]$ (middle row).

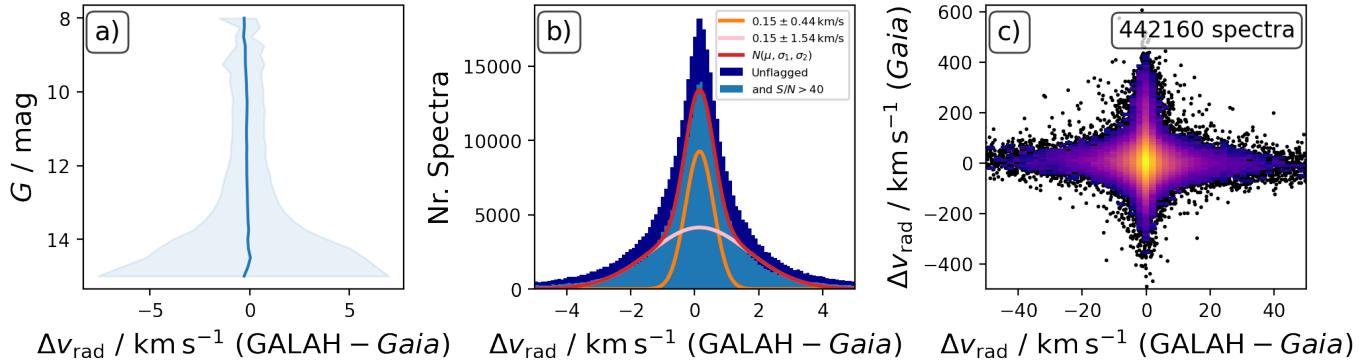


Figure 26. Comparison of radial velocities between GALAH DR4 and Gaia DR3. Panel a) shows the difference of radial velocities as function of *Gaia* *G* magnitude. Panel b) shows a histogram of the difference with two Gaussian distributions (with same mean) fitted to them to estimate a more robust, that 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.

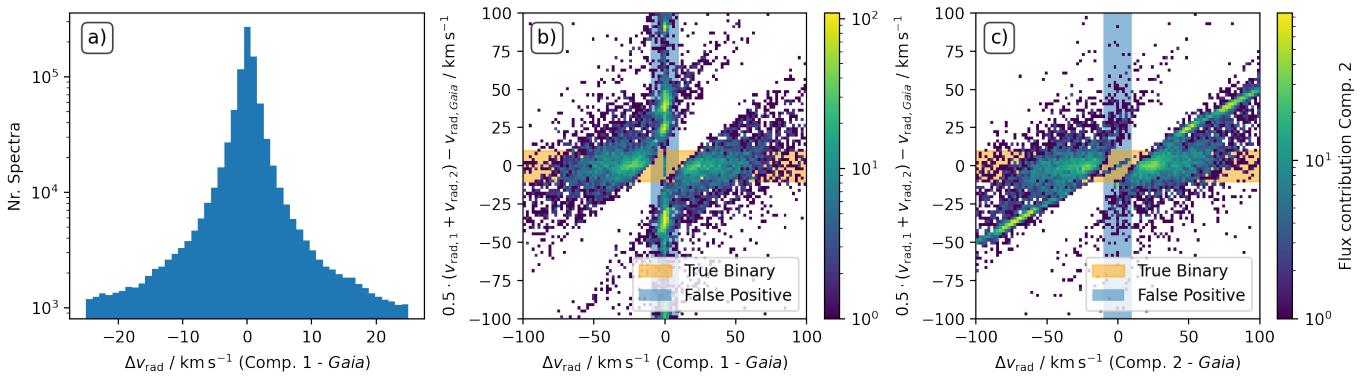


Figure 27. Comparison of radial velocity estimates of GALAH DR4 and Gaia DR3. Panel a) shows the difference of GALAH's primary component radial velocity with the mean *Gaia* DR3. Panels b) and c) show stars for which two components were detected in GALAH DR3 and shows the difference between each component 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.

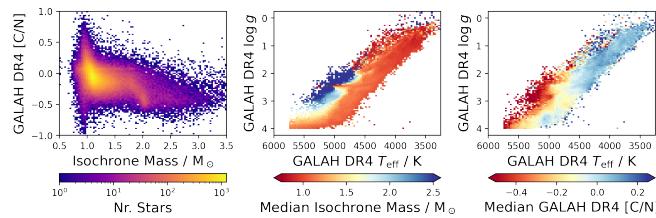


Figure 28. The ratio of [C/N] and isochrone masses in comparison (panel a), and as a function of T_{eff} and $\log g$ in panels b) and c), respectively

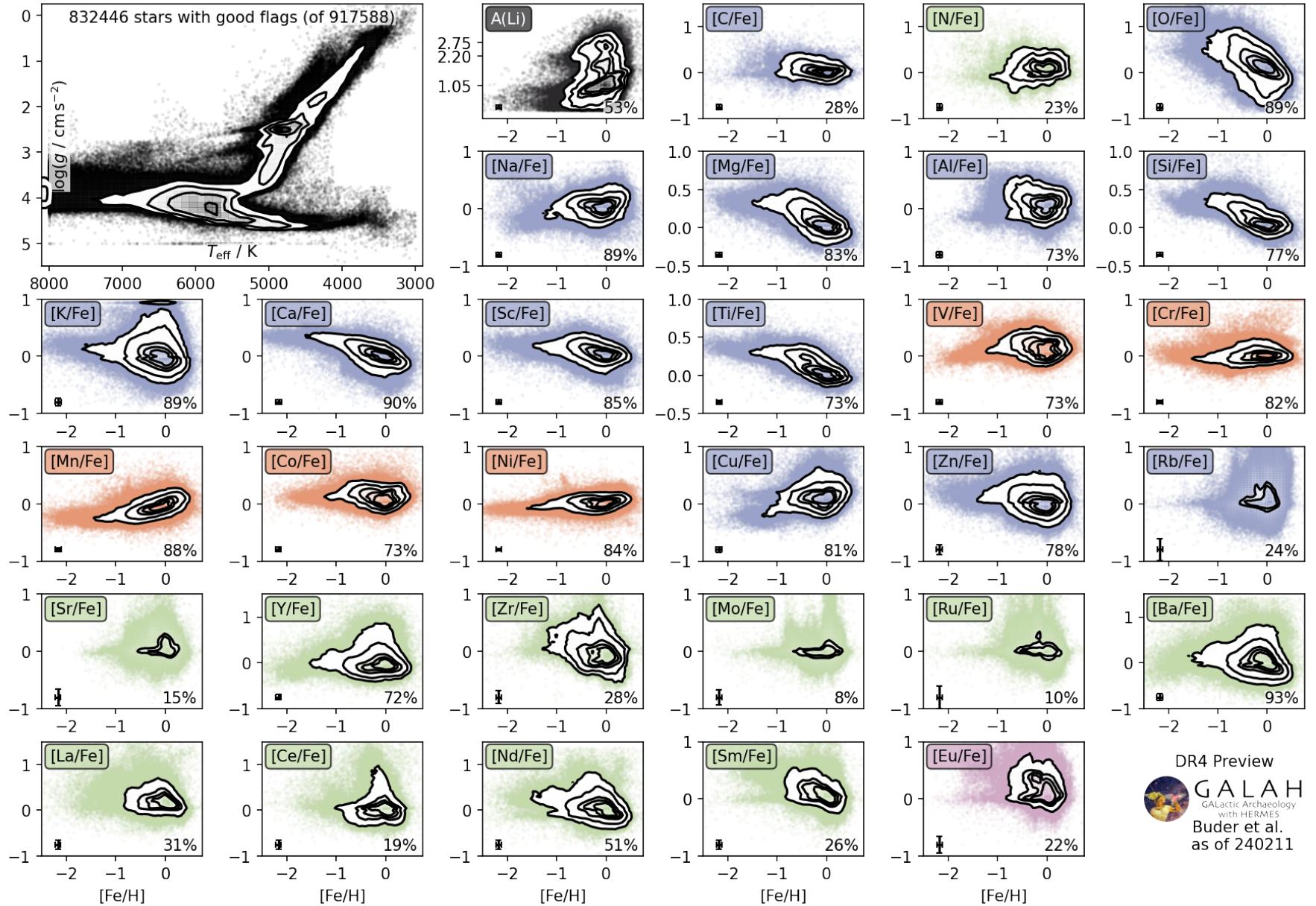


Figure 29. Overview of stellar parameters and elemental abundances for the *allstar* estimates of GALAH DR4. The top left panel shows the density distribution of stars in the Kiel diagram of T_{eff} and $\log g$. All other panels show the logarithmic elemental abundances (for elements indicated in the 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).

8.2 Data products for each spectrum

1. 210115002201239_single_fit_comparison.pdf (see Fig. 33)
Maybe actually use OmegaCen star 140314005201392: cool, metal-poor, strong CNO features and a good spectrum to explain why continuum points may not always work for a pipeline.
2. 210115002201239_single_fit_covariances.npz
3. 210115002201239_single_fit_results.fits
4. 210115002201239_single_fit_rv.png (see Fig. 9)
5. 210115002201239_single_fit_spectrum.fits

For easier downloading, we also provide these files in tarfiles per night. Furthermore, we have interpolated all normalised allstar spectra onto a common wavelength that includes the largest possible wavelength coverage of 99.9% of radial velocity corrected spectra (neglecting pixels below 7680Å for the infrared channel).

8.3 Value-Added Catalogues (VAC)

8.3.1 VAC of crossmatches with Gaia DR3, 2MASS and WISE

Description of the crossmatch with *Gaia* DR3, 2MASS, and WISE

8.3.2 VAC of stellar dynamics

Fig. 30

8.3.3 VAC of 3D NLTE lithium abundances

Ella: Can you add your description here?

8.3.4 Ages

9. 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 9.1), before extensively discussing the spectrum synthesis (Section 9.2) and spectrum interpolation (Section 9.3). We discuss possible problems arising from the use of photometric information (Section 9.4), in particular for stars that could be binaries (Section 9.5). While we have attempted to already flag possible caveats, we also lay out problematic flags in (Section 9.6). We summarise the most important caveats in Section 9.7.

9.1 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.

9.1.1 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.

9.1.2 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.

9.2 Imperfect spectrum synthesis

9.2.1 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 H (Amarsi et al., 2018b), Li (Lind et al., 2009; Wang et al., 2020), C (Amarsi et al., 2019), 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 C₁, 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,

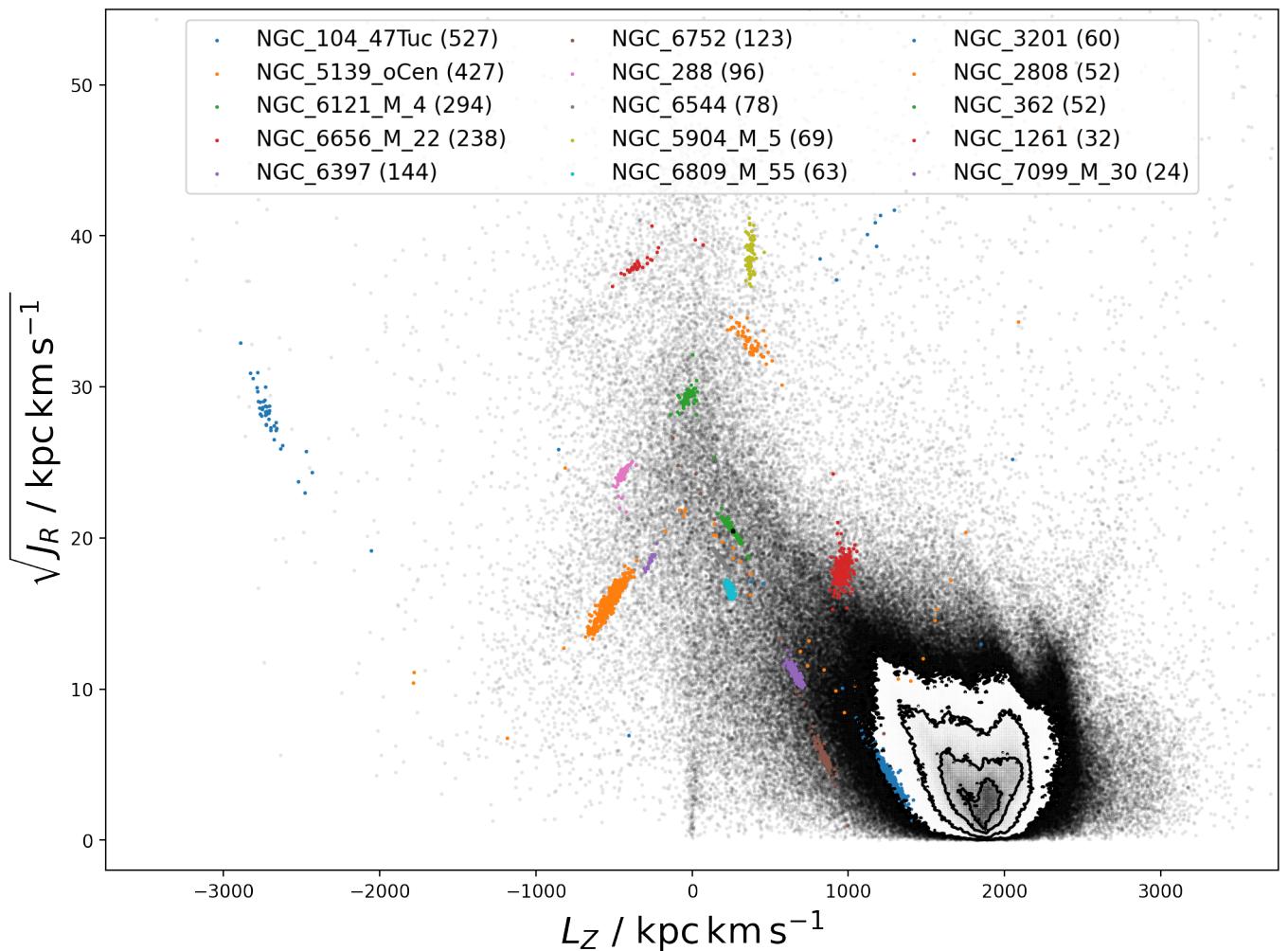


Figure 30. Overview of the dynamical content of GALAH DR4 (black) as well as selected globular clusters (colored) in action space of angular momentum L_Z versus radial action J_R . Cluster members were selected as those with more than 70 percent membership probability according to Vasiliev & Baumgardt (2021).

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.

9.2.2 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).

9.3 Spectrum interpolation with neural networks

9.3.1 Training set selection should be done after zeropoint corrections

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., 2009c).

9.3.2 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_{eff} and [Fe/H] for cool dwarfs) or two stellar properties tend to act in lockstep in actual stars (e.g. Mg, Si, and Ti as alpha-process elements). In GALAH DR2 (Buder et al., 2018), we attempted to overcome these issues by specifically masking the coefficients of spectrum interpolation, that is, effectively restricting the interpolation to only change smaller parts of the spectrum for a given stellar property.

In GALAH DR4, we have relaxed this restriction again, since we have trained on random abundance combinations in the hope of being able to break correlation degeneracies. We note, however, that too little information in spectra can again cause by-chance correlations (e.g. if neutron-capture lines are always very weak and the training set is not sufficiently large). We believe that this is the cause of the decrease in precision for Eu measurements from GALAH DR3 to GALAH DR4. The Eu line was mainly measured only from the weak Eu6645 line, whereas the neural networks are not restricted to this region.

9.3.3 Flexibility of neural networks in general

The choice of using a large number neural networks of restricted regions in the T_{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. It has to be acknowledged that due to the limit of human power

to properly train and test the neural networks, the chosen network architecture of a 2-layer perceptron with 300 neurons and a fixed learning rate is certainly imperfect at best. 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.

9.3.4 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 EW(Li), which are then used to infer 3D-NLTE based $A(\text{Li})$ abundances. This inference is preferable to our 1D-NLTE based neural network estimates, as it is independent of the network flexibility and superior to our less accurate spectrum synthesis in 1D.

While several studies have identified that the abundances of stars in the Galactic disk are often very similar (e.g. Ness et al., 2019), the Galactic halo offers a more diverse picture. An example is 2MASS 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 $[Y/\text{Fe}] = 1.2$, $[Ba/\text{Fe}] = 1.5$, $[La/\text{Fe}] = 1.5$, $[Ce/\text{Fe}] = 1.1$, $[Nd/\text{Fe}] = 1.9$, and $[Sm/\text{Fe}] = 1.2$. 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.

9.3.5 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.

9.4 Mismatch of spectroscopic and photometric information

9.4.1 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 ~ 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 [C/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 [C/N] information could at least serve as an indicator of how trustworthy large giant star masses are.

9.4.2 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.

9.5 Binaries

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

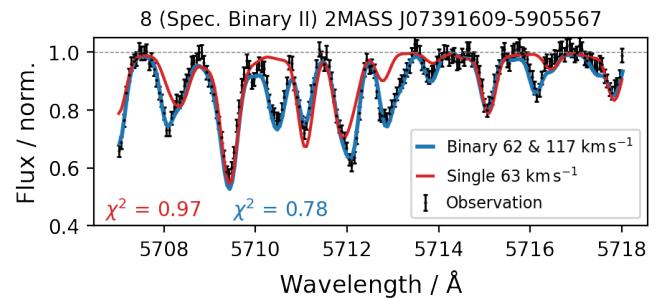


Figure 31. Example spectrum for a double-lined spectroscopic binary star (SBII) that is better fit with our binary fitting algorithm.

9.6 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.

9.6.1 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 [Fe/H], however, this is not the appropriate reference value. It had led to a comparison of a spectrum with $[Fe/H] = -0.74$ dex to one with the lowest grid value of $[Fe/H] = -0.75$ dex, which can easily be too similar within the signal-to-noise 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 $[\text{Fe}/\text{H}] = -4$.

9.6.2 Fitting machinery stuck in local minimum

We have tried to automatically identify regions of the parameter space where the `SCIPY.OPTIMIZE.CURVE_FIT` function has gotten stuck at a local minimum. In particular for cool giants and some red clump stars, we have been able to recover a pattern of abundances being stuck around their initial value. 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 5).

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. Secondly, the `SCIPY.OPTIMIZE.DIFFERENTIAL_EVOLUTION` 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.

9.6.3 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 6).

9.7 Summary of caveats

In summary, the most important caveats are:

- Noding in T_{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_{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.
- 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.

10. CONCLUSIONS

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. 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) 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., & Zatsariny, O. 2018a, A&A, 616, A89
- Amarsi, A. M., Barklem, P. S., Collet, R., Grevesse, N., & Asplund, M. 2019, A&A, 624, A111
- Amarsi, A. M., Grevesse, N., Grumer, J., et al. 2020a, A&A, 636, A120
- Amarsi, A. M., Liljegren, S., & Nissen, P. E. 2022, A&A, 668, A68
- 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
- Bashi, D., Mazeh, T., & Faigler, S. 2023, MNRAS, 522, 1184
- 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
- Bovy, J. 2015, ApJS, 216, 29
- Bressan, A., Mariño, P., Girardi, L., et al. 2012, MNRAS, 427, 127
- Brzeski, J., Case, S., & Gers, L. 2011, SPIE, 8125, 04
- 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., D’Orazi, V., & Lucatello, S. 2009a, A&A, 508, 695
- Carretta, E., Bragaglia, A., Gratton, R., & Lucatello, S. 2009b, A&A, 505, 139
- Carretta, E., Bragaglia, A., Gratton, R. G., et al. 2009c, A&A, 505, 117
- Casagrande, L., & Vandenberg, D. A. 2018, MNRAS, 475, 5023
- Casagrande, L., Lin, J., Rains, A. D., et al. 2020, arXiv e-prints, arXiv:2011.02517
- 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
- Fouesneau, M., Frémant, 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
- 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
- Halbwachs, J.-L., Pourbaix, D., Arenou, F., et al. 2023, A&A, 674, A9
- Heijmans, J., Asplund, M., Barden, S., et al. 2012, SPIE, 8446, 0W
- 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., Kuszlewicz, J. S., Huber, D., Stello, D., & Reyes, C. 2022, AJ, 164, 135
- 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., Worley, C. C., et al. 2017, A&A, 601, A38
- Johnson, C. I., & Pilachowski, C. A. 2010, ApJ, 722, 1373
- Katz, D., Sartoretti, P., Guerrier, A., et al. 2022, arXiv e-prints, arXiv:2206.05902
- 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, 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
- Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, ApJ, 739, 25
- Mariño, 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
- 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., & Zatsariny, 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
- 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)
- Saydjari, 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
- 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
- Spoo, T., Tayar, J., Frinchaboy, P. M., et al. 2022, AJ, 163, 229
- 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
- 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
- Yong, D., Roederer, I. U., Grundahl, F., et al. 2014, MNRAS, 441, 3396
- 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

Appendix 2. Example Fit Comparison

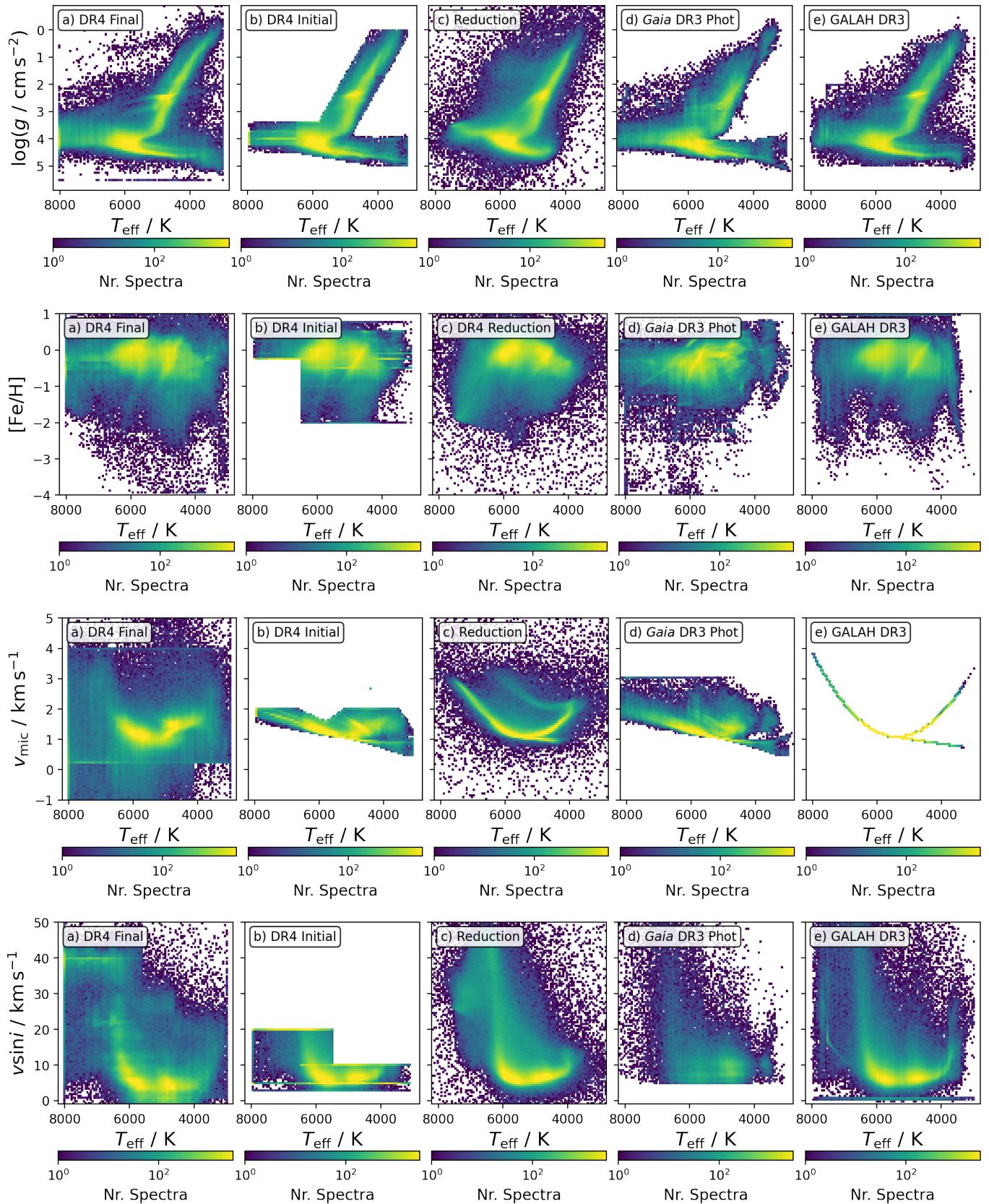
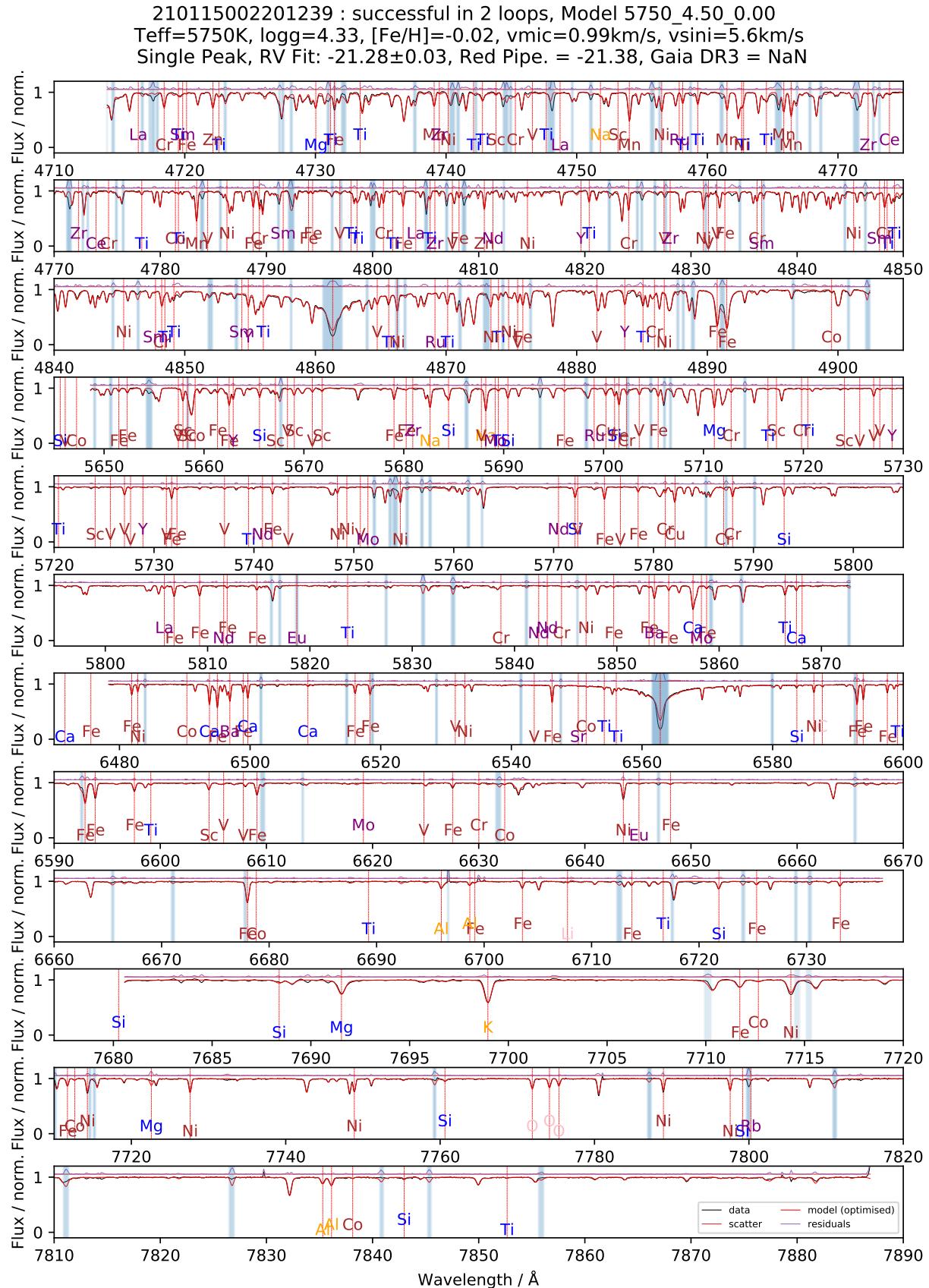


Figure 32. Comparison of final GALAH DR4 stellar parameters (first column) against the initial parameters used in the *allsta* analysis (second column), estimates from the GALAH DR4 reduction pipeline (third column), Gaia DR3 (fourth column with v_{mic} based on the adjusted formula from Dutra-Ferreira et al. (2016)), and GALAH DR3 (fifth column).

Figure 33. Example output of the *allspec* fitting routine for Vesta / 210115002201239.