Combined APOGEE-GALAH stellar catalogues using the Cannon

Govind Nandakumar [ID],[1,2]★ Michael R. Hayden,[2,3] Sanjib Sharma,[2,3] Sven Buder [ID],[1,2] Martin Asplund,[4] Joss Bland-Hawthorn [ID],[2,3] Gayandhi M. De Silva,[5,6] Valentina D'Orazi [ID],[7] Ken C. Freeman,[1] Janez Kos,[8] Geraint F. Lewis [ID],[3] Sarah L. Martell [ID],[2,9] Katharine J. Schlesinger,[1] Jane Lin,[1,2] Jeffrey D. Simpson [ID],[2,9] Daniel B. Zucker,[6,10] Tomaž Zwitter [ID],[8] Thomas Nordlander [ID],[1,2] Luca Casagrande [ID],[1,2] Karin Lind,[11] Klemen Côtar [ID],[8] Dennis Stello,[2,3,9] Robert A. Wittenmyer[12] and Thor Tepper-Garcia [ID][2,3,13]

[1]*Research School of Astronomy & Astrophysics, Australian National University, ACT 2611, Australia*
[2]*Centre of Excellence for Astrophysics in Three Dimensions (ASTRO-3D), Australia*
[3]*Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006, Australia*
[4]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany*
[5]*Australian Astronomical Optics, Faculty of Science and Engineering, Macquarie University, Macquarie Park, NSW 2113, Australia*
[6]*Macquarie University Research Centre for Astronomy, Astrophysics & Astrophotonics, Sydney, NSW 2109, Australia*
[7]*Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[8]*Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia*
[9]*School of Physics, UNSW, Sydney, NSW 2052, Australia*
[10]*Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia*
[11]*Department of Astronomy, Stockholm University, AlbaNova University Center, SE-106 91 Stockholm, Sweden*
[12]*University of Southern Queensland, Centre for Astrophysics, USQ Toowoomba, QLD 4350 Australia*
[13]*Centre for Integrated Sustainability Analysis, The University of Sydney, 2006, Australia*

ABSTRACT

APOGEE and GALAH are two high resolution multi-object spectroscopic surveys that provide fundamental stellar parameters and multiple elemental abundance estimates for about half a million stars in the Milky Way. Both surveys observe in different wavelength regimes and use different data reduction pipelines leading to significant offsets and trends in stellar parameters and abundances for the common stars observed in both surveys. Such systematic differences/offsets in stellar parameters and abundances make it difficult to effectively utilize them to investigate Galactic abundance trends in spite of the unique advantage provided by their complementary sky coverage and different Milky Way components they observe. Hence, we use the *Cannon* data-driven method selecting a training set of 4418 common stars observed by both surveys. This enables the construction of two catalogues, one with the APOGEE-scaled and the other with the GALAH-scaled stellar parameters. Using repeat observations in APOGEE and GALAH, we find high precision in metallicity ($\sim$0.02–0.4 dex) and alpha abundances ($\sim$0.02–0.03 dex) for spectra with good signal-to-noise ratio (SNR $>$ 80 for APOGEE and SNR $>$ 40 for GALAH). We use open and globular clusters to validate our parameter estimates and find small scatter in metallicity (0.06 dex) and alpha abundances (0.03 dex) in APOGEE-scaled case. The final catalogues have been cross-matched with the *Gaia* EDR3 catalogue to enable their use to carry out detailed chemo-dynamic studies of the Milky Way from perspectives of APOGEE and GALAH.

Key words: surveys – stars: abundances – Galaxy: disc – Galaxy: evolution – Galaxy: formation – Galaxy: structure.

1 INTRODUCTION

The field of Galactic archaeology (Freeman & Bland-Hawthorn 2002) deals with dissecting the Milky Way into its various components with the aim to unravel the processes that contributed to the formation and evolution of our Galaxy. Stellar spectroscopy plays a crucial role in this field by enabling accurate measurement of stellar parameters and detailed chemical compositions of stars in the Galaxy. Such measurements are crucial in tracing these stars back to their birth sites providing clues to the

physical processes that led them to their present position in the Galaxy.

There are a plethora of data available in the form of spectra, astrometric, and photometric information, as well as multiwavelength maps with the advent of large-scale spectroscopic (Apache Point Observatory Galactic Evolution Experiment/APOGEE: Eisenstein et al. 2011, RAdial Velocity Experiment/RAVE: Steinmetz et al. 2006, Gaia-ESO: Gilmore et al. 2012, Large Sky Area Multi-Object Fiber Spectroscopic Telescope/LAMOST: Cui et al. 2012, Galactic Archaeology with HERMES/GALAH: De Silva et al. 2015, Abundances and Radial velocity Galactic Origins Survey/ARGOS: Ness et al. 2012), astrometric (*Hipparcos*: Perryman et al. 1997, *Gaia*: Gaia Collaboration 2016), and photometric surveys (Two-Micron All Sky Survey/2MASS: Skrutskie et al. 2006, Sloan Digital

Sky Survey/SDSS: Stoughton et al. 2002, Vista Variables in the Vía Láctea/VVV: Minniti et al. 2010, the SkyMapper Southern Survey : Wolf et al. 2018). These surveys have enabled the chemo-dynamic characterization of stellar populations in the Milky way that constitute different Milky Way components like thin disc, thick disc, halo, bulge, etc. For example, star count observations in the solar neighbourhood (Yoshii 1982; Gilmore & Reid 1983) led to the discovery of the thick disc, followed by its characterization as the old α-enhanced population in the double sequence exhibited by the solar neighbourhood stars in the [α/Fe] versus [Fe/H] plane (Fuhrmann 1998; Bensby, Feltzing & Lundström 2003; Reddy, Lambert & Allende Prieto 2006; Adibekyan et al. 2012; Haywood et al. 2013). At present, data from large-scale spectroscopic surveys (Anders et al. 2014; Hayden et al. 2015; Weinberg et al. 2019) have led to the discovery of this trend at different galactocentric radius, R, and average height, |Z|, across the Galaxy shedding light on the disc formation and evolution scenarios. In addition, many age determination methods have been developed that uses these survey data to provide valuable information about the star formation histories and age metallicity relation of disc stellar populations (Casagrande et al. 2011; Bedell et al. 2018; Lin et al. 2020; Nissen et al. 2020). Secular processes such as radial migration (Sellwood & Binney 2002; Schönrich & Binney 2009; Minchev & Famaey 2010), which leads to the mixing of stars across the Galaxy, are also being explored using a combination of accurate phase space information from *Gaia* (Gaia Collaboration 2018) and chemistry and age information of stars from large-scale spectroscopic surveys (Buder et al. 2019). The discovery of streams and dynamically different stellar populations in the Milky Way halo, considered to be the result of past accretion/merger events (Belokurov et al. 2018; Helmi et al. 2018; Ibata, Malhan & Martin 2019; Myeong et al. 2019) using the *Gaia* data and their further exploration with chemistry from large-scale spectroscopic surveys (Buder et al. in preparation) is another example. Multiple components in the Bulge metallicity distribution function discovered by multiple individual and large-scale spectroscopic observations, are being studied in detail to understand the origin of the Bulge and its connection with the Milky Way bar and Galaxy evolution (Ness et al. 2013; Rojas-Arriagada et al. 2017, 2020). There are many upcoming surveys [4-metre Multi-Object Spectroscopic Telescope/4MOST: de Jong et al. (2019), Sloan Digital Sky Survey/SDSS-V: Kollmeier et al. (2017), WEAVE: Dalton et al. (2018)] that will further improve our understanding of the formation and evolution of the Milky Way and its components.

The above-mentioned large-scale spectroscopic surveys derive fundamental stellar parameters and elemental abundances from observed stellar spectra via dedicated pipelines using spectral fitting routines that fit observed spectra with synthetic spectra generated from stellar model atmospheres, model grids, and line lists, all of which are different/specific to the respective survey. Thus, even though there are overlaps in the observed stars between many spectroscopic surveys, there are significant systematic differences in their derived stellar parameters as well as abundances. This difference can also lead to misinterpretation of abundance trends estimated using the derived parameters from different surveys. Thus it is necessary to combine such complementary surveys with their parameters scaled with respect to either survey so that the resulting volume complete sample can be used to map and decipher the global chemo-dynamic trends of stellar populations in the Milky Way.

One step towards this direction of combining surveys (or scaling them with respect to each other) was made with the introduction of the data driven approach known as the *Cannon* (Ness et al. 2015).

Ho et al. (2017) used the *Cannon* to derive the stellar parameters for around 450 000 giant stars in LAMOST (low spectral resolution survey) by bringing them to the scale of APOGEE (high spectral resolution) survey. Recently, Wheeler et al. (2020) used the *Cannon* to estimate abundances representing five different nucleosynthetic channels, for 3.9 million stars in LAMOST, by training LAMOST spectra with GALAH DR2 stellar parameters and abundances (also referred to as 'labels') for stars in common to both surveys. The *Cannon* has also been used to propagate information from one survey to another, and to derive higher precision stellar parameters, abundances, mass and age information using survey pipeline estimate as the training set labels (Ness et al. 2016; Casey et al. 2017; Buder et al. 2018; Zhang et al. 2019; Hasselquist et al. 2020, etc.). The *Starnet* (Fabbro et al. 2018), a convolutional neural network model, was able to predict stellar parameters by training on APOGEE spectra with APOGEE Stellar Parameter and Chemical Abundance Pipeline (ASPCAP) labels (T$_{eff}$, log g, and [Fe/H]). When compared with the *Cannon* results trained on the same data, the *Starnet* showed similar behaviour, though the *Starnet* performed poorly on small training sets compared to the *Cannon*. A deep neural network designed by Leung & Bovy (2019) was used to determine stellar parameters from APOGEE spectra using the full wavelength range, while censored portions of the spectrum were used to derive individual element abundances. The *Payne* (Ting et al. 2019) is another tool that explicitly models spectra as a function of stellar parameters. Xiang et al. (2019) used the data driven *Payne* to train a model that predicts stellar parameters and abundances for 16 elements from LAMOST DR5 spectra using stars in common with APOGEE DR14 and GALAH DR2 as the training set. Thus, there are many tools and methods available to put different surveys on the same scale.

In this work, we use the data-driven approach, the *Cannon* 2 (Ness et al. 2015; Casey et al. 2016), to put the stellar parameters (T$_{eff}$, log g, and [Fe/H]) and general [α/Fe] abundance on the same scale for the surveys, APOGEE and GALAH. For this, we have to select a training set composed of common stars observed in both the surveys, with high fidelity stellar labels, as well as high quality spectra. Since both the surveys yield stellar parameters and abundances with dedicated pipelines from high-resolution spectra (though in different wavelength ranges), we cannot choose either survey to be the best. Hence, we carry out the exercise in both ways, i.e. (i) train *Cannon* model on APOGEE spectra with GALAH labels and (ii) train *Cannon* model on GALAH spectra with APOGEE labels and derive stellar parameters and [α/Fe] values for both cases. We also (iii) train *Cannon* model on GALAH spectra with GALAH labels and (iv) *Cannon* model on APOGEE spectra with APOGEE labels, so that we can combine (i) and (iii) to provide the GALAH-scaled stellar parameter catalogue, and (ii) and (iv) to provide the APOGEE-scaled stellar parameter catalogue.

We describe the data used in this paper in Section 2. In Section 3, we give a brief description of the *Cannon*, followed by the training set selection, cross-validation of the *Cannon* estimates compared to the input labels from the training set, testing, flagging, and error estimation. We carry out astrophysical validation of our *Cannon* estimates using open and globular cluster members in Section 4. Finally, we discuss the limitations and caveats, as well as notable improvements in *Cannon* estimates with respect to survey pipeline estimates in Section 5

2 DATA

We use the latest available data release of APOGEE (DR16) and GALAH (DR3). We make use of the fundamental stellar parameters:

$T_{\rm eff}$, log g, [Fe/H], and [α/Fe] from those catalogues, along with the stellar spectra for each star.

2.1 APOGEE

The Apache Point Observatory Galaxy Evolution Experiment (APOGEE; Majewski et al. 2017) observes in the near-infrared H-band (15 000–17 000 Å), split into three bands at high spectral resolution (R $\sim$ 22 500) and high signal-to-noise ratios (SNR), in both the Northern and Southern hemispheres. As a part of SDSS-IV (Blanton et al. 2017), the APOGEE-2 survey makes use of the APOGEE instrument (Wilson et al. 2012) on the Sloan 2.5 m Telescope (Gunn et al. 2006) at the Apache Point Observatory (APO) for Northern hemisphere observations, and the 2.5m du Pont telescope at Las Campanas Observatory (LCO; Bowen & Vaughan 1973) with the twin near-infrared (NIR) spectrograph (Wilson et al. 2019) for Southern hemisphere observations.

We make use of the data products from DR16 (Ahumada et al. 2020), which contains a total of 473 307 sources with derived atmospheric parameters and elemental abundances of up to 26 species from the APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP; García Pérez et al. 2016) using MARCS grid of atmospheric models (Jönsson et al. 2020). For our work, we use the calibrated 'PARAM' stellar parameters and abundances described in Jönsson et al. (2020) as they have shown that there are several systematic issues with the spectroscopic (FPARAM) measurements of stellar parameters [section 6.3 and fig. 10 in (Jönsson et al. 2020)]. In addition to the catalogue, we also make use of the 'apStar' and 'asStar' spectra[1], which are the combined spectra of multiple visits, all in a common rest frame and identical wavelength solution across all sources.

2.2 GALAH

Galactic Archaeology with HERMES (GALAH; De Silva et al. 2015) is a high-resolution spectroscopic survey of the Milky Way using the High Efficiency and Resolution Multi-Element Spectrograph (HER-MES; Barden 2010 ; Sheinis et al. 2015) on the Anglo-Australian Telescope. The HERMES spectrograph provides high-resolution (R $\sim$ 28 000) spectra from 4700–7900 Å in four wavelength bands.

We make use of data from the latest GALAH data release (GALAH DR3; Buder et al. 2021), which includes the K2-HERMES survey (Wittenmyer et al. 2018; Sharma et al. 2019) and the TESS-HERMES survey (Sharma et al. 2018), and provides up to 30 element abundances from various nucleosynthesis channels for 678 423 spectra of 588 571 stars. Here, we will refer to this data set collectively as the GALAH survey. Observations are reduced through a standardized pipeline developed for the GALAH survey as described in Kos et al. (2017). In this data release, all stellar parameters and elemental abundances have been estimated via the spectrum synthesis code Spectroscopy Made Easy (SME; Valenti & Piskunov 1996; Piskunov & Valenti 2017) with 1D MARCS stellar atmosphere models (Gustafsson et al. 2008). In addition to the catalogue, we also make use of the GALAH spectra,[2] all in a common rest frame. After removing 12 181 stars that are missing the spectra for some of their chips, we are left with 576 390 GALAH spectra.

---

[1] https://data.sdss.org/sas/dr16/apogee/spectro/redux/r12/stars/
[2] https://docs.datacentral.org.au/galah/dr3/spectra-data-access/

3 METHOD

3.1 The *Cannon*

The *Cannon* (Ness et al. 2015), in simple terms, is a data-driven method that generates a model for stellar spectra based on a training set of spectra, for which the stellar parameters and abundances are known with high fidelity (i.e, high SNR and/or accurate estimates). This model can then be used to infer the same labels for any set of continuum normalized spectra that have been sampled on to a common wavelength grid with uniform start and end wavelengths as that of the training set. The *Cannon* relies on the following assumptions: similar labels imply similar spectra and each spectrum is a smooth function of its labels such that changes in labels result in a smooth variation of spectra.

In the training step, the spectral model coefficients are fit at each wavelength pixel while keeping the labels fixed for all training set star spectra. The spectral model thus generated characterizes the flux at each wavelength pixel as a function of the given labels, with label coefficient values describing the influence of the corresponding label at a certain wavelength pixel. In the label inference step, the label coefficients are fixed, while likelihood optimization is carried out to predict the labels from the flux values at each wavelength pixel of each test spectrum.

We use the *Cannon* 2 described in Casey et al. (2016)[3] and use a quadratic model with $T_{\rm eff}$, log g, [Fe/H], [α/Fe], microturbulence ($v_{\rm micro}$), and line broadening ($v_{\rm broad}$) as the labels, resulting in a spectral model of the following form:

$$F_{n\lambda} = v(l_n).\theta_\lambda + noise \tag{1}$$

where $F_{n\lambda}$ is the flux at each wavelength pixel, λ, for each star, n, in the training set. θ_λ is the set of spectral model coefficients for multiple label combinations at each λ pixel, l_n represents the labels and $v(l_n)$ is the 'vectorizing function', which is in the form of a 2 degree quadratic polynomial and the labels have been normalized or scaled in the following manner:

$$\hat{l}_n = \frac{l_n - l_{n,50}}{l_{n,97.5} - l_{n,2.5}} \tag{2}$$

where $l_{n,2.5}$, $l_{n,2.5}$, and $l_{n,2.5}$ are respectively the 2.5th, 50th, and 97.5th percentile values of the label, l, in the training set.

Finally, noise is drawn from a Gaussian distribution with zero mean and variance defined as the sum of the variance in flux, $\sigma_{n\lambda}^2$, and excess variance at each wavelength pixel, s_λ^2. While variance in flux at each wavelength pixel is obtained from the observed spectra file of each star, excess variance is estimated along with θ_λ in the training step by optimizing the likelihood function:

$$\arg\min_{\theta,s}\left[\sum_{n=0}^{N_{\rm stars}-1}\frac{[F_{n\lambda}-v(l_n).\theta_\lambda]^2}{\sigma_{n\lambda}^2+s_\lambda^2}+\sum_{n=0}^{N_{\rm stars}-1}\ln\left(\sigma_{n\lambda}^2+s_\lambda^2\right)\right] \tag{3}$$

In the test step, hereby called the label inference step, labels corresponding to each test spectrum, m, are predicted by optimizing the likelihood function:

$$\arg\min_l\left[\sum_{\lambda=0}^{N_{\rm pix}}\frac{[F_{m\lambda}-v(l_m).\theta_\lambda]^2}{\sigma_{m\lambda}^2+s_\lambda^2}\right] \tag{4}$$

In the following section, we explain the quality cuts used to choose a high-quality training set in order to generate the required *Cannon* spectral model as described above.
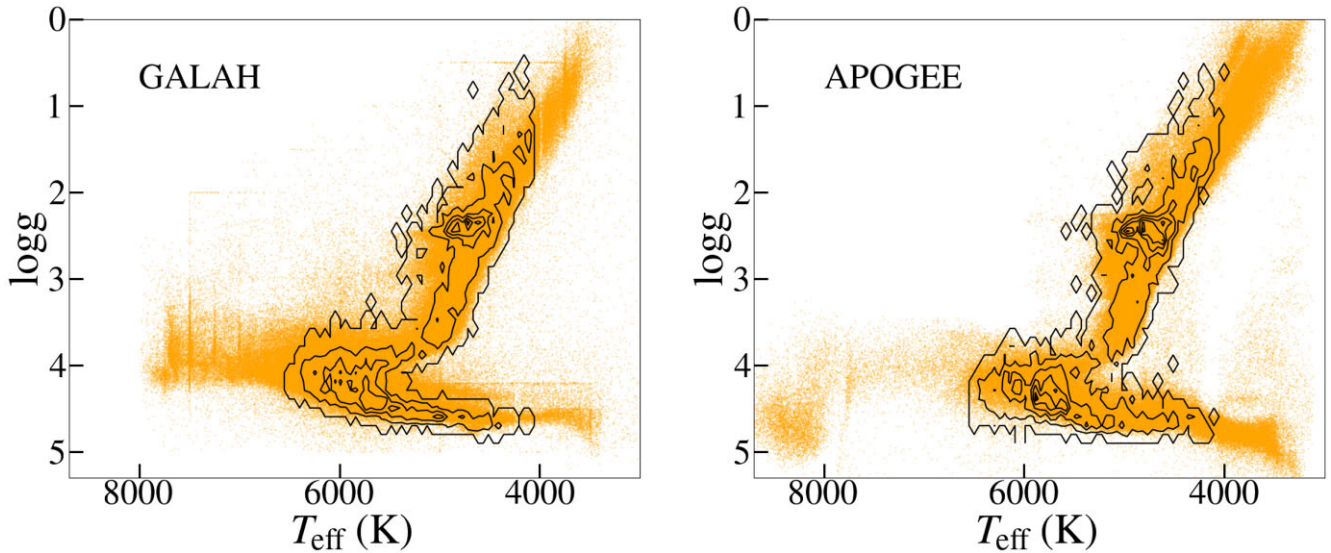
---

[3] https://github.com/andycasey/AnniesLasso

Figure 1. Kiel diagram (log g versus T_{eff}) with the respective pipeline estimates for all stars in GALAH (left-hand panel; orange points) and all stars in APOGEE (right-hand panel; orange points) overlaid with the black contours representing the training set which is chosen from among the common stars in both APOGEE and GALAH as described in Section 3.2.

3.2 Training set

We construct one training set that can be used to carry out four combinations of cross-survey *Cannon* training and modelling. Hence, we use the GALAH DR3 combined spectra catalogue with one entry per star for which spectra from multiple exposures have been stacked to estimate their stellar parameters and elemental abundances. In the case of APOGEE, we choose stars with stellar parameters derived from high SNR combined spectrum. Similarly, we choose combined spectra from both surveys[4] for constructing the training set and later in testing stage where predictions are made.

We find 14 406 stars based on cross-match using the *APOGEE_ID* and *star_id* columns in APOGEE DR16 and GALAH DR3 catalogues with *Topcat* (Taylor 2005, 2020). We remove stars with invalid GALAH and APOGEE stellar parameters and further select reliable and high-quality survey labels/stellar parameters by certain constraints for each survey.

For GALAH, we choose stars with SNR ratio of spectra in the green arm (*snr_c2_iraf*) > 25, chi-square value of stellar parameter fitting the following constraints (*chi2_sp*) < 4, and the flag that describes various GALAH reduction and analysis issues indicating the quality of spectra and estimated stellar parameters (*flag_sp*) is equal to zero.

For APOGEE, we choose stars with SNR > 80, in addition to removing stars for which selected bits (16: bad T_{eff}, 17: bad log g, 18: bad v_{micro}, 19: bad metals, 20: bad [α/Fe], and 23: bad overall for star) in the *ASPCAPFLAG*[5] have been set.

As for selecting good quality spectra, we neglect spectra with *STARFLAG*[6] bits set for selected few bits (0: bad pixels, 3: very bright neighbour, 4: low snr, 9: significant number of pixels in high persistence region, 10: significant number of pixels in medium

persistence region, 11: significant number of pixels in low persistence region, 12: obvious positive jump in blue chip, 13: obvious negative jump in blue chip, and 17: broad lines).

Fig. 1 shows the Kiel diagrams for all stars in GALAH (left-hand panel) and APOGEE (right-hand panel) overlaid with the training set contours with respective pipeline estimates in black. In Fig. 2, we show the systematic difference trends in T_{eff}, log g, [Fe/H], [α/Fe], v_{micro}, and v_{broad} for these stars, with $T_{eff, GALAH}$ on the x-axis and APOGEE-GALAH values on the y-axis. The two clumps, cool (4000 < T_{eff} < 5200 K) and hot (5200 < T_{eff} < 6200 K), seen for all parameters approximately represent difference trend for giants and dwarfs, respectively. Table 1 lists the mean and standard deviation (calculated as the mid-value of 84th−16th percentile to get rid of outliers) values for all stars, giants, and dwarfs. In this case, giants and dwarfs are selected based on log g cut of 3.5.

Among the six stellar parameters, we find significant differences/offsets in the case of [Fe/H], v_{micro} and v_{broad}. The differences for all parameters have a non-zero mean value and exhibits various trends with T_{eff}. APOGEE temperatures are higher and derive higher log g values for hot dwarfs compared to GALAH. Meanwhile, cool giants in APOGEE have lower [α/Fe] measurements compared to GALAH.

The giant and dwarf clumps in the Fig. 2c hints at different systematic trends of [Fe/H] with respect to T_{eff} for giants and dwarfs. There is a steep declining trend for giants, with higher metallicity values (∼0.4 dex relative to GALAH abundances) measured by APOGEE for cool giants (T_{eff} < 4500 K). This trend is similar to one of the possible caveats mentioned in GALAH DR3 [section 6.5 of Buder et al. 2021]. They have noticed a significant trend of underestimated [Fe/H] with increasing metallicity, when comparing with GALAH DR2, for the metal-rich ([Fe/H > 0) giants and red clump stars. As discussed in Buder et al. 2021, the reasons for this could be many fold, e.g, missing/unreliable molecular line data, the underestimation of blending and incorrect continuum normalization, over/under estimation of v_{micro} etc. For dwarfs, there is better agreement between APOGEE and GALAH metallicities with a slight increase (upto ∼0.15–0.2 dex) in APOGEE metallicity for hotter stars (T_{eff} > 6000 K).

---

[4]For APOGEE, combined spectra is provided as the HDU1 extension of 'apStar' fits files. For GALAH, combined spectra fits files are named after their 'sobject_id' from which the pseudo continuum normalized flux can be extracted.
[5]https://www.sdss.org/dr14/algorithms/bitmasks/
[6]https://www.sdss.org/dr14/algorithms/bitmasks/#APOGEE_STARFLAG

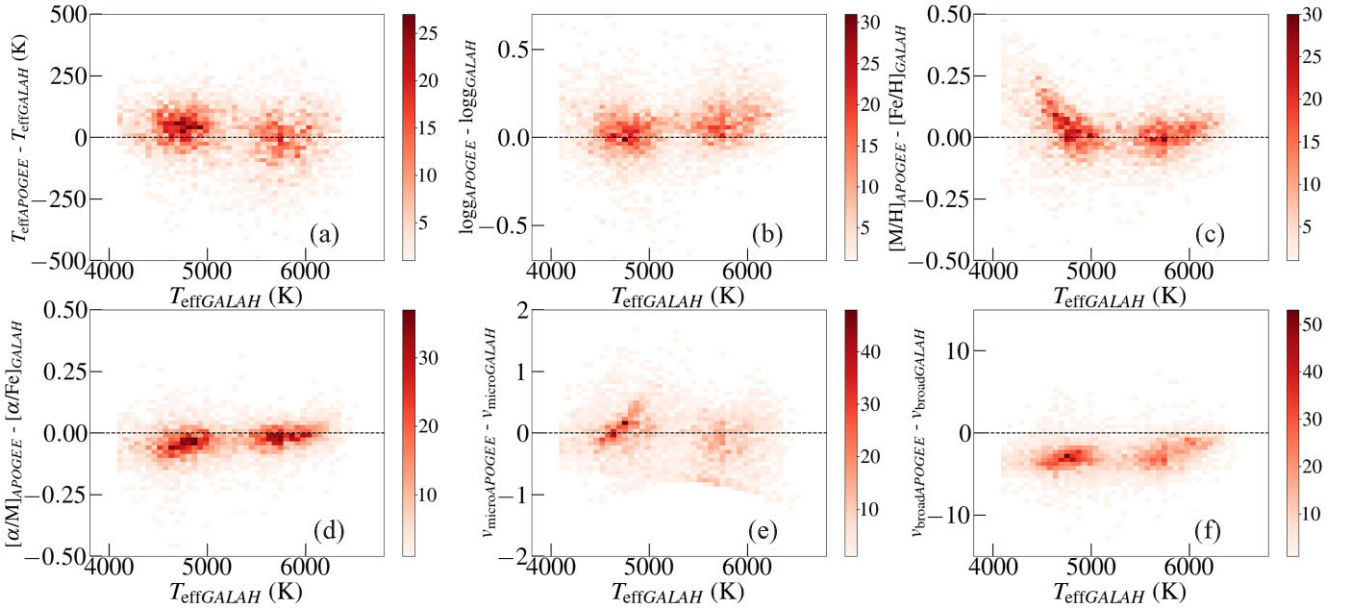APOGEE ASPCAP [DR16] - GALAH SME [DR3] (4418)

Figure 2. The difference in derived stellar parameters as a function of T_{eff} from GALAH SME values for 4418 stars (shown as 2D histogram colour coded by the number of stars in each bin) in the training set. The two clumps, cool ($4000 < T_{eff} < 5200$ K) and hot ($5200 < T_{eff} < 6200$ K), seen for all parameters approximately represent difference trend for giants and dwarfs, respectively.

All six parameters used as input labels in the *Cannon* are shown: (a) T_{eff}, (b) $\log g$, (c) [Fe/H], (d) [α/Fe], (e) v_{micro}, and (f) v_{broad}.

Table 1. Mean and scatter values of APOGEE-GALAH difference for six stellar labels in the training set.

APOGEE–GALAH	All		Giants		Dwarfs	
Parameter (unit)	μ	σ	μ	σ	μ	σ
T_{eff} (K)	14	84	35	68	0	94
$\log g$ (dex)	0.06	0.12	0.03	0.13	0.08	0.10
[*Fe/H*] (dex)	0.02	0.08	0.05	0.10	0.00	0.06
[α/*Fe*] (dex)	-0.03	0.05	-0.04	0.06	-0.02	0.05
v_{micro} (km s^{-1})	-0.05	0.37	0.07	0.30	-0.14	0.40
v_{broad} (km s^{-1})	-2.80	1.50	-3.08	-1.21	-2.59	1.62

In Fig. 2e, we find significant scatter ($> \sim 0.30$ km s^{-1}) in the v_{micro} difference between the two surveys, with mean differences of -0.05, 0.07, and -0.14 km s^{-1} for the whole sample, giants, and dwarfs, respectively. The difference in trends for giants and dwarfs are also evident from the two clumps. Such a significant difference may be attributed to the way in which v_{micro} is determined in GALAH and APOGEE. While empirical relations (equations 4 and 5 in Buder et al. 2021) are employed in the case of GALAH, APOGEE uses v_{micro} as a free parameter while determining synthetic spectra that best fits the observed spectra.

In Fig. 2f, a systematic difference in v_{broad} can be seen with the APOGEE values consistently lower than GALAH values. This may again be attributed to the difference in v_{broad} determination in both surveys. In GALAH, v_{broad} is determined using SME by setting v_{mac} (macroturbulent velocity) to 0 and only fitting for $v \sin i$ (rotational velocity). In APOGEE, v_{broad} for dwarfs are the $v \sin i$ values estimated in the same way as in the case of v_{micro} using a grid with seven steps of $v \sin i$ (1.5, 3.0, 6.0, 12.0, 24.0, 48.0, and 96.0 km s^{-1}), while for giants an empirical relation is used to estimate v_{mac} (Jönsson et al. 2020).

Even though there is reasonably a good agreement between the two surveys (especially for T_{eff} and $\log g$) in Fig. 2, there are systematic differences in [Fe/H], [α/Fe], v_{micro}, and v_{broad} for the same stars in both surveys. As mentioned in Section 2, both surveys observe stars in different wavelength regimes (optical for GALAH and NIR for APOGEE) and employ different methodology and pipelines to estimate these parameters which could result in such systematic differences. These systematic differences for same stars, as well as contrasting difference trends for giants and dwarfs, thus show the importance of placing these surveys on the same abundance scales if they are to be used in conjunction. This also emphasizes the need for observation of larger number of common stars between large-scale spectroscopic surveys, which will enable cross survey calibrations, as well as more consistent analysis pipelines in the future.

3.3 Training

Once the training set is finalized, we proceed to carry out the training and cross-validation. As mentioned in Section 1, the objective of this work is to provide two combined stellar parameter catalogues of APOGEE and GALAH, one scaled in terms of APOGEE and the other in terms of GALAH. Hence, we use spectra and labels from both surveys in four different combinations, starting from the training step. To avoid any confusion resulting from this, hereafter we introduce a naming convention to indicate each case in Table 2:

In the following section, we focus on APOGEE Cannon GALAH SME (ACGS) and GALAH Cannon APOGEE ASPCAP (GCAA), while similar exercises for APOGEE Cannon APOGEE ASPCAP (ACAA) and GALAH Cannon GALAH SME (GCGS) are explained in Appendix A.

We limit the labels that we train and infer to T_{eff}, $\log g$, [Fe/H], [α/Fe], v_{micro}, and v_{broad} where [Fe/H] and [α/Fe] refer to the general metallicity and alpha abundance labels in each survey. We do not go

Table 2. Naming convention used in this work to indicate four catalogues resulting from the use of two surveys and their spectra in different combinations.

Name	Spectra	Labels
APOGEE Cannon GALAH SME (ACGS)	APOGEE	GALAH SME
APOGEE Cannon APOGEE ASPCAP (ACAA)	APOGEE	APOGEE ASPCAP
GALAH Cannon APOGEE ASPCAP (GCAA)	GALAH	APOGEE ASPCAP
GALAH Cannon GALAH SME (GCGS)	GALAH	GALAH SME

beyond the general alpha abundance to include individual elements as in such cases, we need to know all abundances for each training set entry – which becomes an increasingly difficult task with an increasing number of elements. In particular, elements other than the alpha-elements, like Li or s- and r-process elements are impossible to measure always throughout the whole parameter space. Various challenges involved in determination of more elements using the *Cannon* is discussed in detail in Buder et al. (2018).

While we use the whole spectral wavelength range for training and determining label coefficients for T_{eff}, $\log g$, [Fe/H], v_{micro}, and v_{broad}, we make use of censoring/line masks in the case of [α/Fe]. This is similar to the use of line masks in SME to estimate abundances for each element in a spectrum. [α/Fe] for GALAH is determined from an error-weighted combination of selected lines of Si, Mg, Ti, and Ca, while APOGEE [α/M] is determined from a combination of O, Mg, Si, S, Ca, and Ti lines. We decided to use line masks for the common elements in both surveys (Si, Mg, Ti, and Ca) to be used as censors in the training step. This ensures that *Cannon* does not incorporate incorrect correlations from other lines in the respective spectra while predicting [α/Fe] values. In the case of GALAH, the line masks for Si, Mg, Ti, and Ca are available in the line list used to determine respective abundances with SME (table A2 in Buder et al. 2021). For APOGEE, Jönsson et al. (2020) provides the windows and weights used in the determination of stellar abundances in their table 3. We select the wavelength windows for Si, Mg, Ti, and Ca lines.

In the training step, four *Cannon* models corresponding to ACGS, GCGS, GCAA, and ACAA are created using spectra and labels of stars in the training set. Coefficient matrix, θ_λ, and scatter, s_λ, are obtained by optimizing equation (3.1), given the labels in each case.

3.4 Cross-validation

In order to verify how well the *Cannon* model is able to recover the training set labels, we carry out a 12-fold cross-validation. For this, we create 12 random subgroups from the training set, then use them individually as a test set while the remaining 11 are treated as the training data. We show the resulting one-to-one relation, i.e. predicted label values versus input label values for ACGS and GCAA in Figs 3 and 4, respectively.

When we train the *Cannon* model on APOGEE spectra using GALAH SME labels (ACGS; Fig. 3), all labels, except v_{broad} and [α/Fe], follow a tight one-to-one relation. T_{eff} relation is tight with mean difference of ~5 K and scatter of ~90 K. The *Cannon* estimates for $\log g$ and [Fe/H] values are similar to corresponding input GALAH labels, while larger dispersion in input GALAH [α/Fe] values is reflected in the *Cannon* output as well. The tightest one-to-one relation is seen in the case of v_{micro}, for which the input GALAH labels are estimated using empirical relations. As this relation is a function of GALAH T_{eff} and $\log g$, these values are well correlated with respective spectra in the training step and thus ensure that the

Cannon model is able to reproduce them from APOGEE spectra in the label inference step. The largest scatter is seen in the case of v_{broad}, meaning that the *Cannon* model is not able to correlate the input GALAH v_{broad} values with the features in APOGEE spectra. At the same time, we find that the *Cannon* model trained on GALAH spectra using GALAH labels (GCGS) are able to estimate similar v_{broad} values from GALAH spectra in the label inference step with less scatter (see Appendix, Fig. A1). Thus the reason for the large dispersion seen here in the case of v_{broad} is not clear.

When we train the *Cannon* model on GALAH spectra using APOGEE ASPCAP labels (Fig. 4), the *Cannon* output values for T_{eff}, $\log g$, [Fe/H], and [α/Fe] is as seen in ACGS, with tight one-to-one relations with the respective APOGEE ASPCAP values. There are significant deviations in the case of v_{micro} and v_{broad}. The *Cannon* outputs for input $v_{micro\ APOGEE} > 1$ km s^{-1} follow one-to-one relation, whereas the *Cannon* estimates are higher for lower $v_{micro\ APOGEE}$ values. This indicates that the *Cannon* model is unable to find significant correlation between $v_{micro\ APOGEE}$ values and corresponding GALAH spectra in this lower $v_{micro\ APOGEE}$ range. Much tighter one-to-one trend is seen in the case of v_{broad}, though there are significant deviations for lower $v_{broad\ APOGEE}$ values. When we use the *Cannon* model trained on APOGEE spectra using APOGEE labels (ACAA), the resulting *Cannon* v_{micro} and v_{broad} estimates also show significant deviations in similar v_{micro} and v_{broad} ranges (see Appendix A1). This may be because of the limited sensitivity at such low values of v_{micro} and v_{broad}.

In all cases and for all labels, there are deviations from the one-to-one line close to the high and low values of the respective labels, i.e. the label space edges. This most likely shows the inability of the *Cannon* to interpolate at the training set edges.

Overall, the 12-fold cross-validation shows the capability of the *Cannon* models trained on APOGEE (GALAH) spectra with GALAH (APOGEE) labels (with training set comprising good quality stars commonly observed by both surveys) to infer GALAH (APOGEE) scaled values of T_{eff}, $\log g$, [Fe/H], and [α/Fe] from APOGEE (GALAH) spectra.

3.5 Application to GALAH and APOGEE

We proceed to use the *Cannon* models after training and cross-validation to predict the GALAH-scaled labels for 437 445 APOGEE spectra and APOGEE-scaled labels for 576 390 GALAH spectra of unique stars. Using four *Cannon* spectral models corresponding to ACGS, GCGS, GCAA, and ACAA, all six labels are estimated for each case by optimizing the log likelihood function in equation (3.1), given the respective coefficient matrix, θ_λ, and scatter, s_λ, from the training step.

In the following sections, we demonstrate the effectiveness of our *Cannon* models by comparing model spectra with observed spectra (both GALAH and APOGEE) and plotting the systematic differences of all labels for common stars in APOGEE and GALAH.
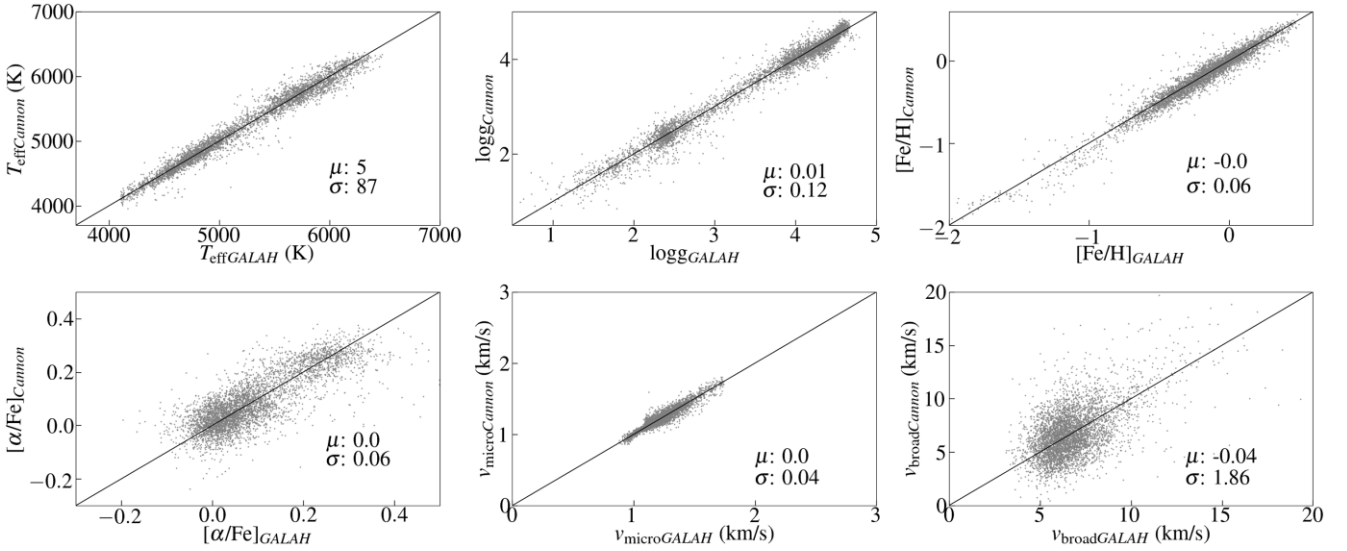
Figure 3. One-to-one relation of GALAH SME stellar labels (*x*-axis) in the training set versus corresponding values estimated by the *Cannon* from APOGEE spectra (*y*-axis) after carrying out a 12-fold cross-validation test. The mean and scatter (calculated as the mid value of 84th−16th percentile) in each label difference is indicated on the bottom right-hand side of each plot.
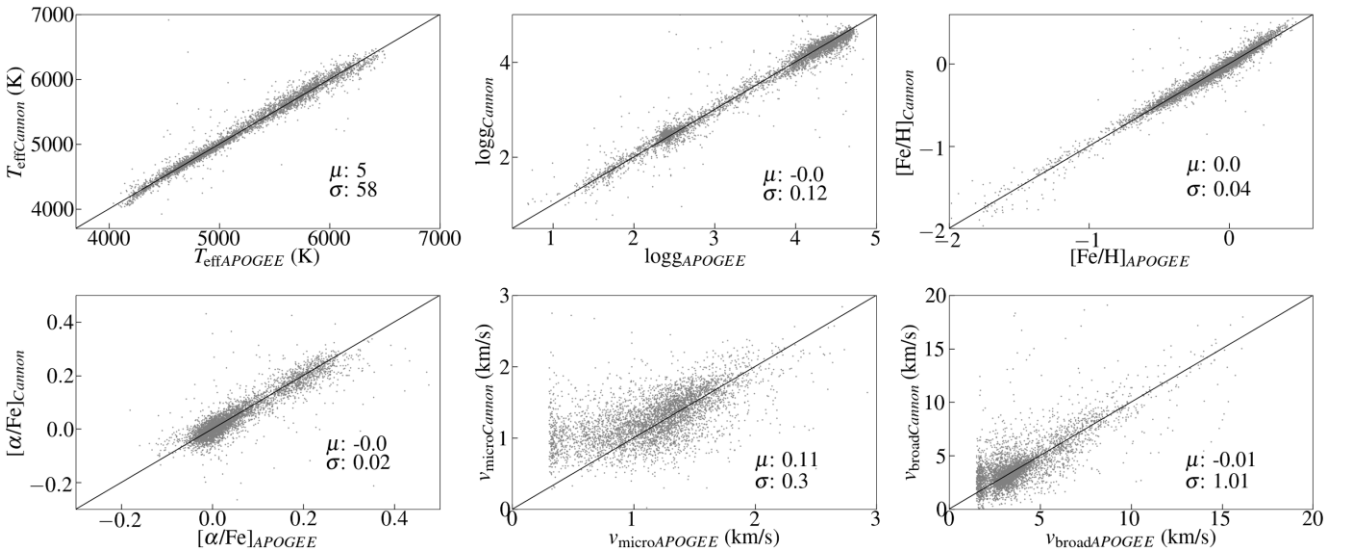

Figure 4. One-to-one relation of APOGEE ASPCAP stellar labels (*x*-axis) in the training set versus corresponding values estimated by the *Cannon* from GALAH spectra (*y*-axis) after carrying out a 12-fold cross-validation test. The mean and scatter (calculated as the mid value of 84th−16th percentile) in each label difference is indicated on the bottom right-hand side of each plot.

3.5.1 Cannon *model spectra*

Here, we compare observed survey spectra with the spectra generated by the *Cannon* models to demonstrate how well *Cannon* spectral models reproduce the observed spectra. For this, we first estimate the residual values (observed − model normalized flux) for all 4418 stars in the training set at each wavelength/pixel. We then plot the median residual value along with the respective 16th and 84th percentile values at each pixel as shown in Figs 5 and 6 for model spectra comparisons with GALAH spectra and APOGEE spectra, respectively. In both the figures, we show observed solar-type star spectrum from the respective survey in the top row to help the readers identify lines and the residual plots for APOGEE-scaled and GALAH-scaled cases, respectively, in the top and bottom rows with the full survey wavelength ranges plotted in the panels from left

to right. The solid scatter points represent the median residual value estimated at each pixel and the band around each point/pixel denote the 16th and 84th percentile values of residuals at the respective pixel.

Median values for majority of pixels are very close to 0 with the 16th and 84th percentile values lying within 0.01, except in the case of GALAH chip 1 which is found to be inherently noisy. In the case of GALAH (Fig. 5), slightly higher residual values are seen at either bad/noisy pixels or pixels that correspond to lines of elements (e.g. Cr at ∼5700 Å, Cu at ∼5782 Å, and strong K line at ∼7700 Å) that have not been modelled in our *Cannon* models. Comparing the pixel positions of higher residual values with lines in APOGEE solar-type star spectrum, it is evident that the bad/noisy spike pixels are the dominant reason for higher residuals in the case of APOGEE (Fig. 6).
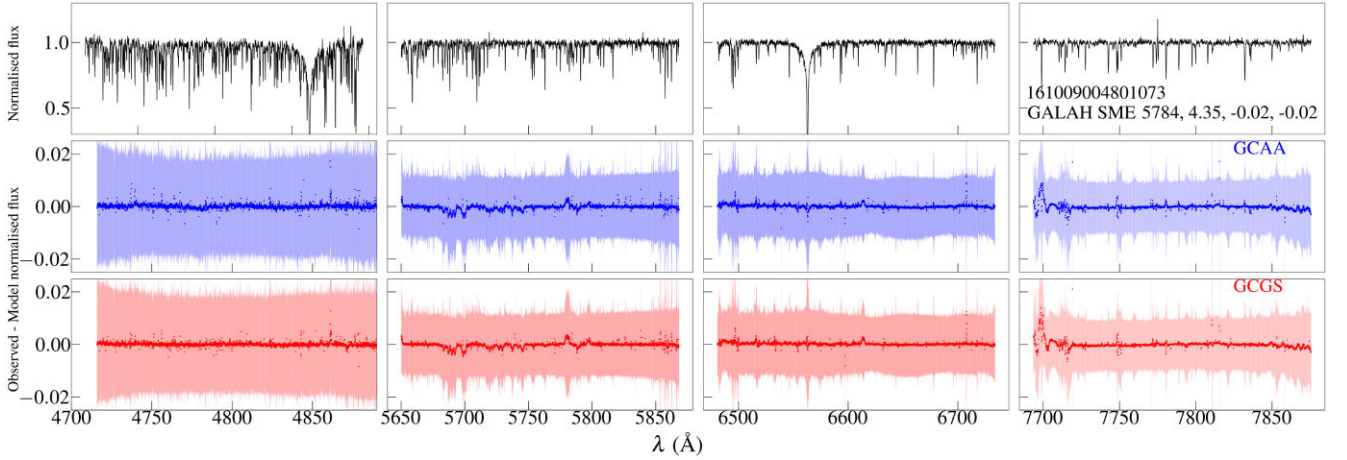
Figure 5. Pixel-by-pixel comparison of the *Cannon* model (GCAA and GCGS) spectra with the observed GALAH spectra for all 4418 stars in the training set. The top row shows a typical solar-type GALAH star spectrum with the GALAH SME stellar parameters (T_{eff}, $\log g$, [Fe/H], and [α/Fe]) listed in the right-most panel. The middle and bottom rows show the comparison for GCAA and GCGS, respectively, with the full GALAH wavelength ranges in the four chips plotted in four panels from left to right. The solid scatter points represent the median residual value (observed – model normalized flux) estimated at each pixel and the band around each point/pixel denote the 16th and 84th percentile values of residuals estimated at the respective pixel. Median residuals are very close to 0 for majority of pixels, while higher residuals are seen at either bad/noisy pixels or pixels that correspond to lines of elements (e.g. Cr at $\sim$5700 Å, Cu at $\sim$5782 Å, and strong K line at $\sim$7700 Å) that have not been modelled in our *Cannon* models.
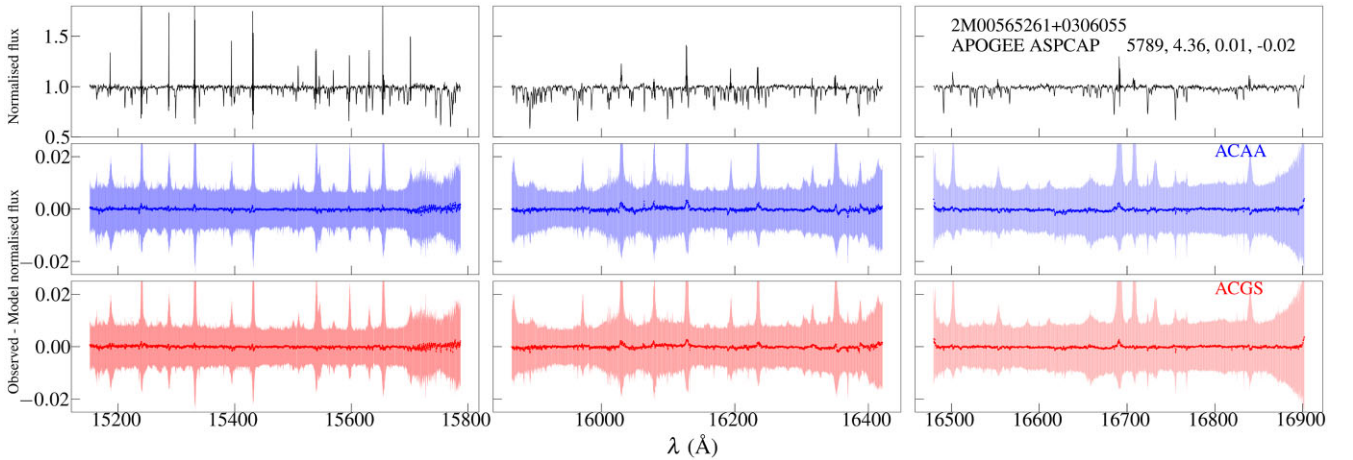

Figure 6. Pixel-by-pixel comparison of the *Cannon* model (ACAA and ACGS) spectra with the observed APOGEE spectra for all 4418 stars in the training set. The top row shows a typical solar-type APOGEE star spectrum with the APOGEE ASPCAP stellar parameters (T_{eff}, $\log g$, [Fe/H], and [α/Fe]) listed in the right-most panel. The middle and bottom rows show the comparison for ACAA and ACGS, respectively, with the full GALAH wavelength ranges in the three wavelength bands plotted in three panels from left to right. The solid scatter points represent the median residual value (observed – model normalized flux) estimated at each pixel and the band around each point/pixel denote the 16th and 84th percentile values of residuals estimated at the respective pixel. As in the case of GALAH, median residuals are very close to 0 for majority of pixels. Higher residuals are dominated by bad/noisy pixels that is evident on comparison with the solar-type star spectrum in the top panels.

This demonstrates the ability of the *Cannon* spectral models to reproduce the observed spectra for stars in the parameter space covered by our training set.

3.5.2 Systematic difference using common stars

In order to demonstrate the effectiveness of our method, we compare GALAH-scaled labels from APOGEE spectra (ACGS) and GALAH spectra (GCGS) for common stars which are in the training set. For the same stars, we compare APOGEE-scaled labels from

APOGEE spectra (ACAA) and GALAH spectra (GCAA). This is shown in Figs 7 and 8 where we plot the differences in all six labels as a function of T_{eff} for ACGS – GCGS and ACAA – GCAA, respectively. This is similar to Fig. 2 where we showed the differences between the pipeline values from APOGEE and GALAH. Table 3 lists the mean and scatter of the difference values for all labels in each case. On comparison with the mean and scatter values for APOGEE-GALAH listed in Table 1, we find reduced scatter and less trends for common stars in both surveys from GALAH-scaled and APOGEE-scaled catalogues. The improvement is especially evident in metallicity and alpha abundance with the
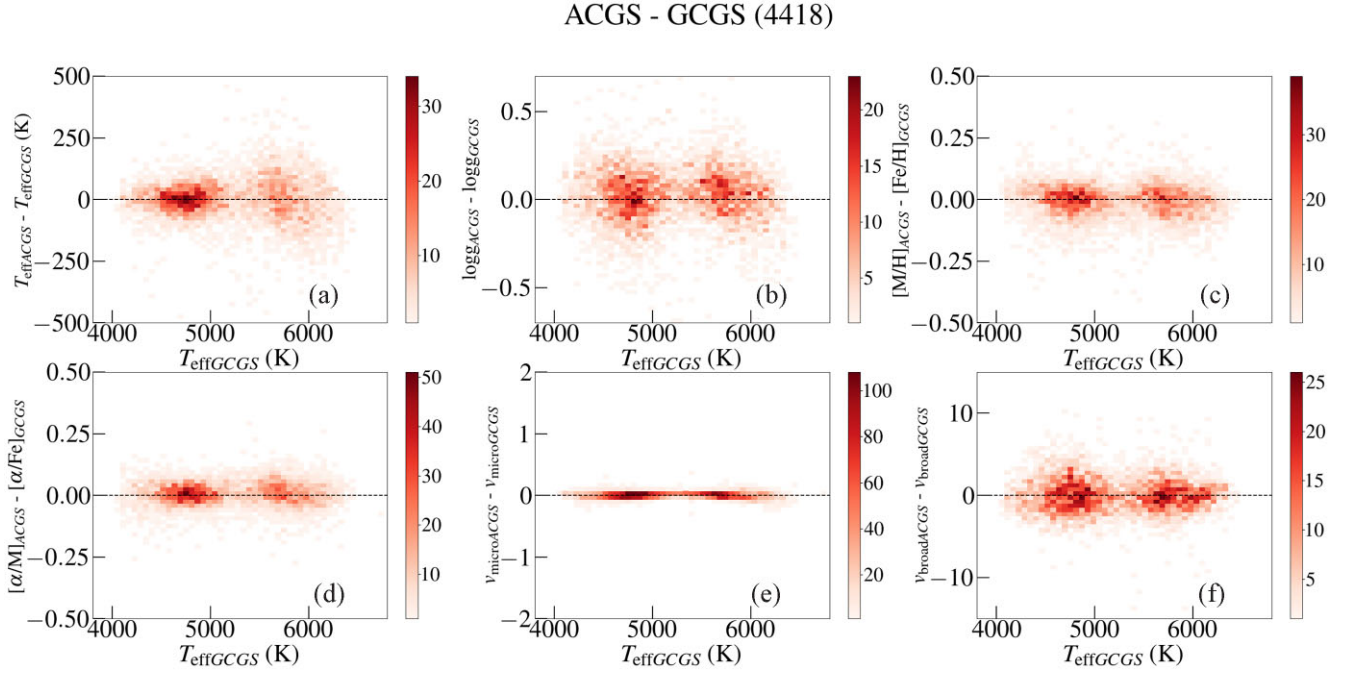
Figure 7. Same as Fig. 2 except for label differences between training set stars in ACGS and GCGS. Compared to differences seen in Fig. 2, metallicity and alpha abundance differences have decreased scatter and less trends, demonstrating the effectiveness of the *Cannon* models.
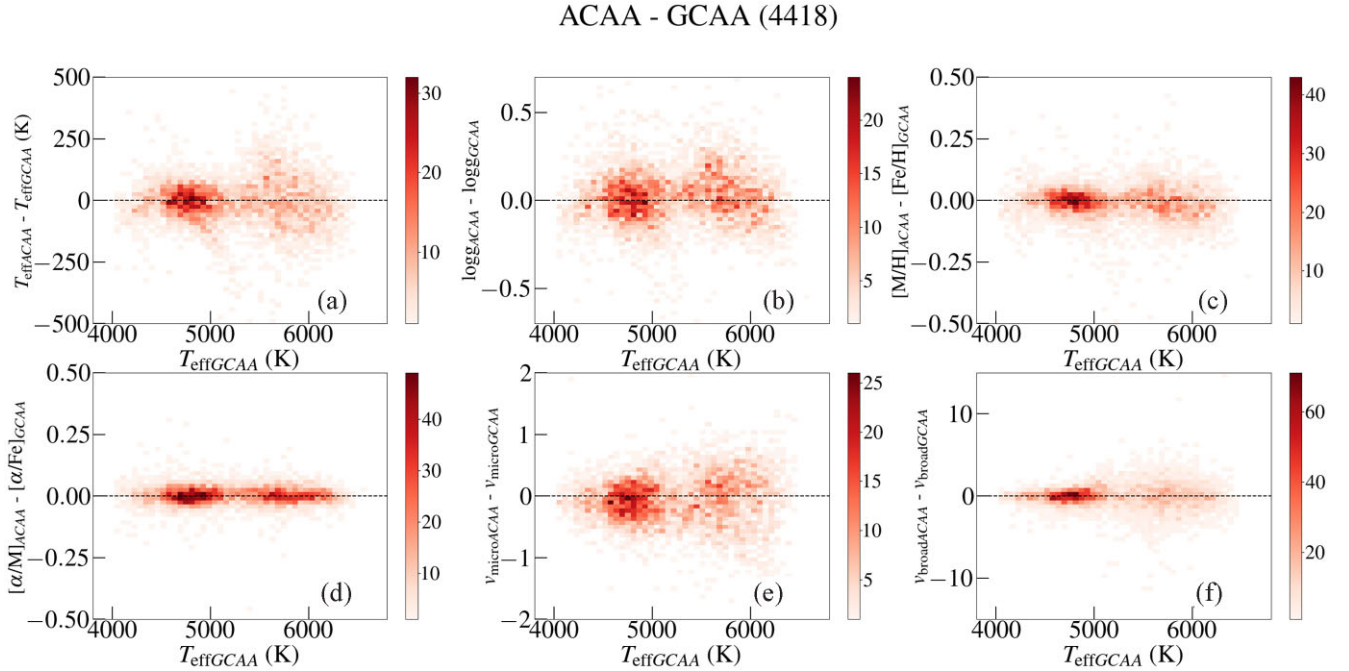

Figure 8. Same as Fig. 2 except for label differences between training set stars in ACAA and GCAA. Reduced scatter and not less trends are seen here, with the mean difference always lying close to 0. This demonstrates the effectiveness of our method and the *Cannon* models we used.

mean difference close to 0 and the absence of any significant trends.

3.6 Flagging

As a first step, we need to flag *Cannon* estimates that lie outside the bounds of the training set labels since the *Cannon* cannot reliably extrapolate to different regimes outside of the training set. For each test set spectra, we estimate the distance, D, of the test set *Cannon* estimate, l_{test}, to the training set labels, l_{nTs}, in similar way as described in Buder et al. (2018) and Ho et al. (2017) :

$$D = \sum_l \sum_{nTs} \frac{(l_{test} - l_{nTs})}{K_l^2} \tag{5}$$

Table 3. Mean and scatter values in ACGS – GCGS and ACAA – GCAA difference of six stellar labels for common stars in APOGEE and GALAH.

Parameter (unit)	ACGS – GCGS		ACAA – GCAA	
	μ	σ	μ	σ
$T_{\rm eff}$ (K)	2	66	−4	73
log g (dex)	0.03	0.12	0.02	0.12
[Fe/H] (dex)	0.00	0.05	0.00	0.05
[α/Fe] (dex)	0.01	0.04	0.00	0.02
$v_{\rm micro}$ (km s^{-1})	0.00	0.03	−0.06	0.34
$v_{\rm broad}$ (km s^{-1})	−0.01	1.78	−0.06	1.07

Table 4. Scaling factor for each label covariance error estimated from repeat observations using equation (6) for each catalogue.

Parameter (unit)	ACGS	GCGS	GCAA	ACAA
$T_{\rm eff}$ (K)	12200	7200	6925	11 325
log g (dex)	18	8.6	8.5	19
[Fe/H] (dex)	8.9	6.4	7.7	9
[α/Fe] (dex)	1.2	0.5	0.4	1.2
$v_{\rm micro}$ (km s^{-1})	3.8	2.1	5.3	12
$v_{\rm broad}$ (km s^{-1})	52.1	31.5	29	56

where K_l represents the uncertainties in each label for which we use RMS values obtained from cross-validation step in Section 3.4. Using $T_{\rm eff}$, log g, [Fe/H], [α/Fe] as the label space, l, we calculate the average distance to the closest 10 stars in the training set and flag those *Cannon* estimates that are farther than eight (2σ for four labels). We indicate this in the final catalogue by setting the *flag_Cannon_dist* to 1. In addition, we provide *flag_spectra* to indicate problematic spectra from APOGEE (STARFLAG bits as explained in Section 3.2) and GALAH, *flag_sp_aspcap* to indicate stars that have been flagged by respective surveys (flag_sp in GALAH and ASPCAP FLAG in APOGEE) and *flag_survey* to indicate the stars with invalid values (from ASPCAP and SME) for the labels we use. *flag_sp_aspcap* follows the same format as in the respective survey. *flag_spectra* for APOGEE stars follow the format in APOGEE DR16 catalogue, while value of 1 is used to indicate bad GALAH spectra. With these flags, we indicate those stars/spectra that have corresponding issues and emphasize that it is better to avoid these stars for exploring the Milky Way chemo-dynamic trends. Table A1 lists and describe all the above mentioned flags.

3.7 Error and precision

The *Cannon* provides covariance errors that are very small and likely to be underestimated (Casey et al. 2017). In the following sections, we explain precision and precision estimation using repeat observations in each survey and from the label difference of stars in the training set as a function of SNR.

3.7.1 Error

We use the *Cannon* results for repeat observations in each survey to determine the factor by which covariance errors for each label have to be scaled in order to get a reasonable error estimate. In APOGEE, there are 28 570 groups of repeated observations with the group size varying from 2 to 14. Meanwhile, there are 47 189 groups of repeated observations of sizes 2–15.

We calculate the pair-wise differences between the labels we derived from multiple visits over the quadrature sum of their formal covariance errors. For each label, l, we introduce a scaling factor, f, as shown below:

$$\delta l = \frac{l_1 - l_2}{\sqrt{\left(\sigma_{l1}^2 + \sigma_{l2}^2\right).f^2}} \tag{6}$$

Under the ideal assumption that the derived labels are unbiased and the *Cannon* covariance errors are correct, we expect δl (without scaling factor) to follow a normal distribution with zero mean and variance of unity. The resulting δl distribution is found to be significantly different from the expected normal distribution for all

labels except [α/Fe] for *Cannon* estimates in ACGS, GCGS, GCAA, and ACAA. This indicates that covariance errors obtained from *Cannon* are not correct. Hence, we proceed to vary f in equation (6) until variance of δl distribution tend to reach unity.

In Table 4, we list the scaling factor for each label estimated from repeat observations for each catalogue. In all cases, scaling factor for [α/Fe] are close to 1, which can be attributed to the fact that the *Cannon* covariance error for [α/Fe] is indicative of the actual error.

3.7.2 Precision

We further use repeat observations in APOGEE and GALAH to determine the label precision in the *Cannon* estimates as a function of SNR. For this, we calculate the differences between the *Cannon* estimates from the most high SNR spectrum and those from lower SNR spectra (from multiple visits) of the same star. These differences are then binned by SNR of the lower SNR spectra with the percentile difference (mid-value of 84th−16th percentile) in each bin denoting the precision achieved in the respective SNR range.

In Fig. 9, we show exponential fits to the percentile differences (shown as filled circles) as a function of SNR using the repeat observations in all four cases: ACGS, GCGS, GCAA and ACAA, as indicated by the colour of the lines in the plots. The precision for all labels except $v_{\rm mciro}$ tends to vary exponentially as a function of SNR depending on the survey to which the spectra of repeat observations belong. Precision estimated from APOGEE spectra (ACGS and ACAA) tends to flatten out at SNR > 80 while precision estimated from GALAH spectra follow the same trend around SNR > 40. Thus, when using combined catalogues (ACGS+GCGS or GCAA+ACAA), it is advisable to choose stars with SNR > 40 for GALAH spectra and SNR > 80 for APOGEE spectra to make sure that the parameters are of the same precision scale. In the case of $v_{\rm mciro}$, precision in APOGEE-scaled cases (GCAA and ACAA) follow similar trend independent of the spectra the label is inferred from. Also, higher precision is achieved for $v_{\rm mciro}$ in GALAH-scaled cases independent of the spectra. Thus, in the case of $v_{\rm mciro}$, we infer that the *Cannon* models are unable to find strong correlations for APOGEE values with either spectra. As mentioned in Section 3.2, this may be the result of different methods employed to determine $v_{\rm mciro}$ in APOGEE and GALAH (also see Sections 3.4 and A1).

We also list the precision at SNR > 80 (for ACGS and ACAA) and SNR > 40 (for GCAA and GCGS) in Fig. 9. We find higher precision for $T_{\rm eff}$ and log g when they are estimated from GALAH spectra and APOGEE spectra, respectively. Meanwhile, similar high precision (∼0.02–0.04 dex) is obtained in all cases for [Fe/H] and [α/Fe].

We find that the rescaled *Cannon* covariance errors and precision estimates closely follow each other as a function of SNR. Hence, we take the maximum value from among the rescaled *Cannon* covariance
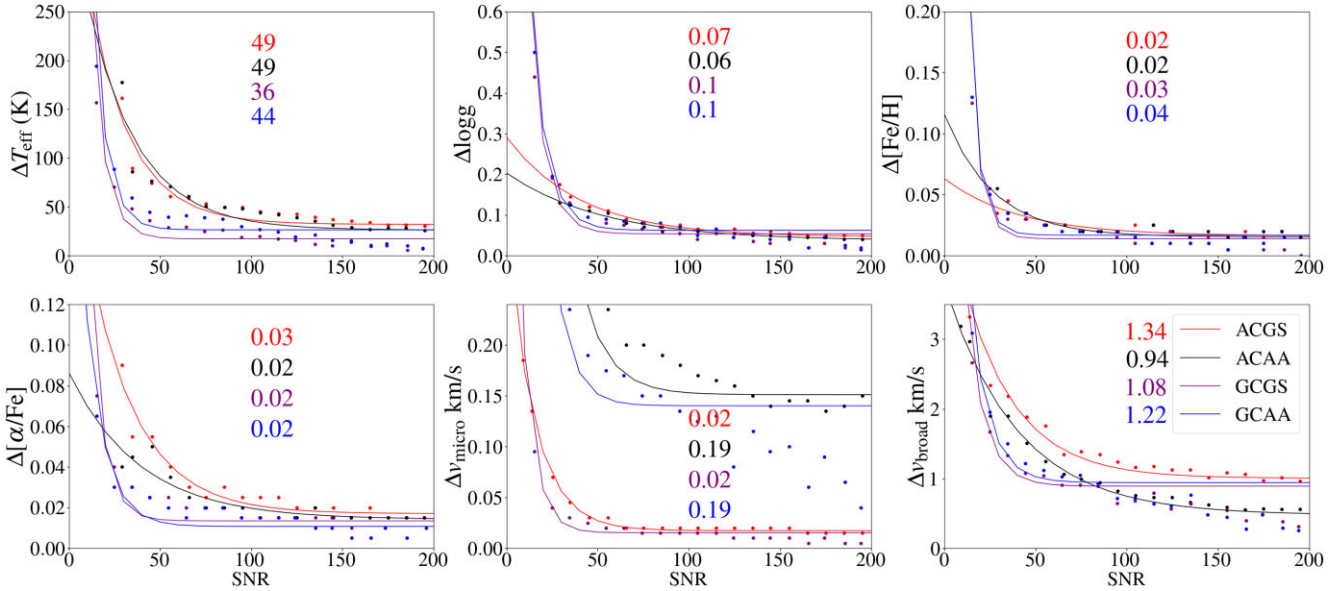
Figure 9. Exponential fits to the percentile differences (mid-value of 84th−16th percentile plotted as filled circles) of the *Cannon* estimates of labels between high-SNR and low-SNR spectra among multiple visits of same stars (repeat observations) in ACGS (red), ACAA (black), GCGS (purple), and GCAA (black) binned as a function of the SNR of the lower SNR spectra. The precision for all labels except $v_{\rm mciro}$ flattens out at approximately SNR > 40 when spectra belong to GALAH (GCGS and GCAA) and SNR > 80 when it is APOGEE spectra (ACGS and ACAA).

errors and precision estimate at the respective SNR as the final error for each label in all cases. This is published in the final catalogues.

4 VALIDATION

We combine *APOGEE Cannon GALAH SME* (ACGS) and *GALAH Cannon GALAH SME* (GCGS) to construct the GALAH-scaled catalogue. Similarly, we combine *GALAH Cannon APOGEE AS-PCAP* (GCAA) and *APOGEE Cannon APOGEE ASPCAP* (ACAA) to construct the APOGEE-scaled catalogue. Both the catalogues provide stellar parameters, metallicity, and global alpha abundances for 1 013 835 stars corresponding to the sum of number of unique stellar spectra in each survey (437 445 in APOGEE and 576 390 in GALAH). After applying the flags as mentioned in the Section 3.6, we are left with slightly less than 50 per cent of the total number of stars in the catalogues, but with good quality spectra and the survey parameters on a common scale in each catalogue.

In this section, we carry out external astrophysical validation of the *Cannon* estimates to investigate how well the *Cannon* is able to learn from GALAH and APOGEE labels, along with their respective spectra, to produce the GALAH- and APOGEE-scaled catalogues. To do this though, we have only a limited number of stars from other high-resolution spectroscopic studies that have been observed by APOGEE and/or GALAH (e.g. Reddy et al. 2003, 2006; Bensby, Feltzing & Oey 2014 etc). So we rely on open and globular clusters, which contain stars that are born from the same parental molecular cloud and are expected to have similar ages. Hence, we expect the stars in open and globular clusters to follow the same theoretical isochrone track in the Kiel diagram and to have similar metallicities and abundances.

4.1 Astrophysical validation

In this section, we investigate the astrophysical validity of the stellar parameters in our catalogues. We start by searching for members of

previously observed, well studied open and globular clusters in our combined GALAH-scaled and APOGEE-scaled catalogues. We then check the consistency of our *Cannon* estimated metallicities of these member stars with that in APOGEE and/or GALAH, as well as with that in the literature.

4.1.1 Open clusters

We cross-match our catalogues with the open cluster member catalogue provided by Spina et al. (2021) for 205 clusters observed by at least one of the survey, APOGEE (DR16) or GALAH (DR3). From our GALAH-scaled and APOGEE-scaled catalogues, we choose stars with *flag_Cannon_dist*, *flag_survey*, *flag_spectra* set to 0, SNR cuts as in the training set and remove stars flagged by both surveys (*flag_sp_aspcap*). From the open cluster catalogue, we choose member stars with membership probability, P_mem, > 0.5. We identify 113 open clusters with at least one member in our catalogues that satisfy these conditions. For our investigation, we choose 5 open clusters: Blanco 1, IC 4665, Melotte 22, NGC 2682, and Ruprecht 147.

We show Kiel diagrams for members of the above mentioned clusters colour coded by their metallicities in Fig. 10. The plots for clusters are arranged from left to right in increasing order of their ages from literature (Cantat-Gaudin et al. 2020), with the top row showing survey pipeline (SME/ASPCAP) estimates, middle and bottom row showing GALAH-scaled and APOGEE-scaled values, respectively, for same stars. We also overplot PARSEC isochrones (Bressan et al. 2012) in each panel, where the isochrone ages are from Cantat-Gaudin et al. (2020) and metallicities are adopted from Heiter et al. (2014). For each cluster, we separate out the spectra of stars from each survey (indicated by different symbols in the Fig. 10) and list median metallicities and standard deviations of GALAH-scaled and APOGEE-scaled cases for their respective spectra in Table 5. In addition, we also list the median and standard deviation values of the pipeline metallicity estimates, enabling us to compare and quantify
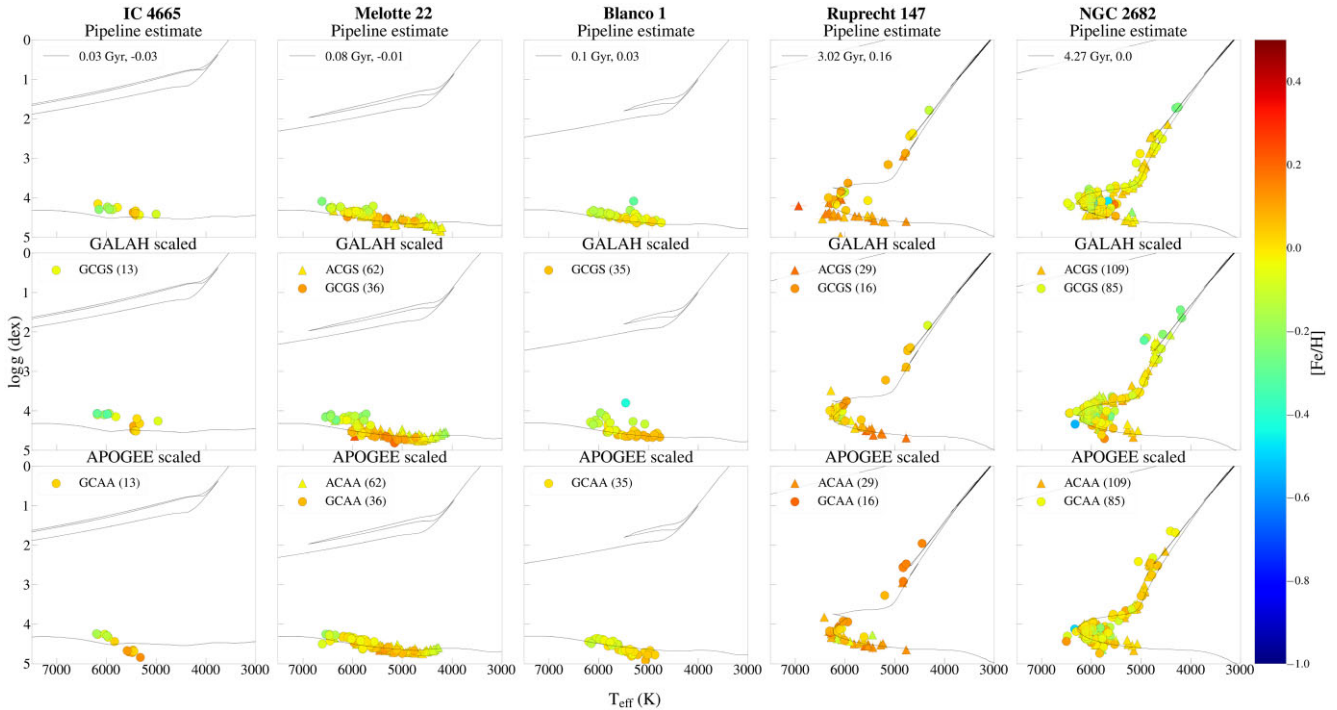
Figure 10. Kiel diagram for members of five open clusters from the catalogue of Spina et al. (2021) colour coded by their metallicities. Clusters are arranged from left to right in increasing order of their ages from literature. We show survey pipeline (GALAH SME and APOGEE ASPCAP) estimates in the top row, GALAH-scaled values (ACGS and GCGS) in the middle row, and APOGEE-scaled values (GCAA and ACAA) in the bottom row. PARSEC isochrones (Bressan et al. 2012) based on cluster ages and metallicities adopted from literature (age: Cantat-Gaudin et al. 2020, metallicity: Heiter et al. 2014) are over plotted (black line) in each panel. The values estimated from APOGEE spectra (ASPCAP as well as *Cannon* estimates) are indicated as triangle markers, while those from GALAH spectra (SME as well as *Cannon* estimates) are indicated as circle markers.

the similarity of *Cannon* estimates to the respective survey pipeline estimates.

For Blanco 1 and IC 4665, there are no APOGEE stars in our catalogues after quality cuts. Except for IC 4665, member stars from rest of the clusters consistently lie on the theoretical isochrone track defined by respective age and metallicity from literature, for both *Cannon* estimates, as well as individual survey pipeline estimates. In IC 4665, GALAH SME log g estimates also show deviation from the track at higher T_{eff} for a few stars in the GALAH-scaled case, which is more pronounced in GALAH-scaled case at $T_{eff} >$ 6000 K. This is not unexpected since there are fewer lines in the GALAH wavelength windows for these hotter stars. Similar deviations from tracks or scatter in log g are evident in the case of other two young clusters, Melotte 22 and Blanco 1, pointing out the problems in the GALAH SME estimates and spectra of hot dwarf stars.

Within the standard deviations, GALAH-scaled and APOGEE-scaled median metallicity values are consistent with that from the median survey pipeline estimates of stars in GALAH and APOGEE surveys, and with the literature values (see Table 5). At the same time, we see higher values of the standard deviations in GALAH-scaled metallicities compared to APOGEE-scaled cases. This is expected since GALAH SME estimates also show higher values of standard deviation compared to APOGEE ASPCAP values for all clusters. In the case of Ruprecht 147, GALAH-scaled median metallicities from APOGEE spectra (ACGS) are consistent with GALAH SME median metallicity. Similarly, APOGEE-scaled median metallicities estimated from GALAH spectra (GCAA) are consistent with the more metal rich APOGEE

ASPCAP median metallicity estimate. These are clear indications of the ability of the *Cannon* to carry out cross survey scaling.

4.1.2 Globular clusters

Unlike in the case of open clusters, there are no combined catalogues of globular clusters in APOGEE and GALAH. So we cross-match our catalogues with the latest *Gaia* EDR3 globular cluster catalogue (Vasiliev & Baumgardt 2021). As in the case of open clusters, we choose stars from our catalogues with *flag_Cannon_dist*, *flag_survey*, *flag_spectra* set to 0, SNR cuts as in the training set, globular cluster member stars with membership probability >0.5 and remove stars flagged by both surveys (*flag_sp_aspcap*). This results in nearly 28 globular cluster populations with only two of them having spectra in both APOGEE and GALAH. For our investigation, we choose five globular clusters: NGC 104 (47 Tuc), NGC 2808, NGC 288, NGC 6121, and NGC 6809.

We show Kiel diagrams for members of the above mentioned clusters colour coded by their metallicities in Fig. 11. The plots for clusters are arranged from left to right in decreasing order of their metallicities from literature (VandenBerg et al. 2013), with the top row showing survey pipeline (SME/ASPCAP) estimates, middle and bottom row showing GALAH-scaled and APOGEE-scaled values respectively for same stars. We also overplot PARSEC isochrones (Bressan et al. 2012) in all panels where the isochrone ages and metallicities of these clusters are from VandenBerg et al. (2013). In Table 6, we list the median metallicities and standard deviations

Table 5. Median metallicity and standard deviation estimated for members of five open clusters from the catalogue of Spina et al. (2021) cross-matched with the catalogues in this work. The names of clusters are listed in the first column, metallicity of the cluster from high resolution study (Heiter et al. 2014) in the second column, the abbreviations for the catalogues (see Section 3.5) with the number of stars from each catalogue in brackets in the third column, *Cannon* metallicity estimates from each catalogue in the fourth column, and survey pipeline metallicity estimates corresponding to the spectra mentioned (in the catalogue name) in the same row in the last column.

Cluster	$[Fe/H]_{literature}$ (median, std dev)	Catalogue (No.)	$[Fe/H]_{Cannon}$ (median, std dev)	$[Fe/H]_{Survey}$ (median, std dev)
IC 4665	−0.03, 0.04	GCGS (13)	−0.05, 0.13	−0.03, 0.10
		ACGS (0)	–	–
		GCAA (13)	0.03, 0.10	−0.03, 0.10
		ACAA (0)	–	–
Melotte 22	−0.01, 0.05	GCGS (36)	−0.05, 0.15	−0.03, 0.09
		ACGS (62)	0.00, 0.10	−0.01, 0.04
		GCAA (36)	0.00, 0.05	−0.03, 0.09
		ACAA (62)	0.00, 0.04	−0.01, 0.04
Blanco 1	0.03, 0.07	GCGS (35)	0.0, 0.09	−0.05, 0.07
		ACGS (0)	–	–
		GCAA (35)	−0.01, 0.06	−0.05, 0.07
		ACAA (0)	–	–
Ruprecht 147	0.16, 0.08	GCGS (16)	0.05, 0.06	0.03, 0.06
		ACGS (29)	0.07, 0.08	0.12, 0.03
		GCAA (16)	0.11, 0.07	0.03, 0.06
		ACAA (29)	0.10, 0.05	0.12, 0.03
NGC 2682	0.0, 0.06	GCGS (85)	−0.05, 0.09	−0.06, 0.07
		ACGS (109)	−0.06, 0.07	0.00, 0.03
		GCAA (85)	−0.02, 0.07	−0.06, 0.07
		ACAA (109)	0.00, 0.05	0.00, 0.03

of GALAH-scaled and APOGEE-scaled cases for their respective spectra, as well as from survey pipelines (SME/ASPCAP), in addition to respective cluster metallcities from VandenBerg et al. (2013).

From the plots, we find *Cannon* and survey pipeline estimates to follow theoretical isochrone tracks in the case of NGC 104. For NGC 2808, NGC 6121, and NGC 288, survey pipeline estimates are systematically offset (lower T_{eff}) from the respective theoretical isochrone tracks while the *Cannon* estimates are closer to them. This is a clear indication of the quality of *Cannon* estimates compared to the survey pipeline estimates and shows the ability of the *Cannon* models to correct the shift in pipeline estimates by learning it from field stars that are dominant in the training set. For the most metal poor cluster in our sample, NGC 6809, both the *Cannon* and survey pipeline estimates are offset from the stellar track. For this cluster, the median metallicity from the literature is ∼0.25 dex more metal poor than those from the *Cannon*, as well as survey pipeline estimates. This is most likely the reason for the shift in *Cannon* and pipeline estimates compared to the theoretical isochrone track. On the other hand, during the cross-validation of the training set, we have seen that the *Cannon* estimates for stars with metallicities lower than −1.5 dex may be wrong due to them being scarcely represented in the training set and values close to the training set boundary. But consistency of *Cannon* estimates with the survey pipeline estimates gives strength to the validity of the *Cannon* estimates in this case.

For all clusters except NGC 288 and NGC 104, median metallicities from the *Cannon* are consistent with the median metallicity values from survey pipelines. In NGC 288, median metallicity from GALAH SME (−1.07 dex) is more metal rich than GALAH-scaled (−1.12 dex) and APOGEE-scaled (−1.26 dex) values from GALAH spectra for same set of stars, while we find consistent

median metallicities from APOGEE ASPCAP and *Cannon* estimates from APOGEE spectra. This means that for NGC 288 members, metallicity information estimated from GALAH spectra with the *Cannon* model trained on GALAH spectra using APOGEE ASPCAP labels (GCAA) results in more accurate (and closer to literature estimate) metallicity values compared to GALAH SME values. Similarly, the *Cannon* model trained on APOGEE spectra using GALAH SME labels (ACGS) is able to derive consistent metallicity information from APOGEE spectra, while slight decrease in median metallicity compared to GALAH SME estimate is seen in the case of GCGS. This points out that the GALAH SME metallicity estimates could be wrong, which is confirmed in the next section where we find significant systematic differences in stellar parameters for these stars between APOGEE and GALAH.

In NGC 104, median metallicity estimated from *Cannon* models trained on GALAH spectra using GALAH SME (GCGS) and APOGEE ASPCAP (GCAA) labels are more metal rich compared to the pipeline estimates, as well as literature value. On further investigation by comparing GALAH spectra of NGC 104 members with similar T_{eff}, log *g*, and [Fe/H] from GALAH SME, we found differences in spectra of stars with more metal rich *Cannon* estimates. This difference is found to be due to enhanced N abundance in stars with more metal-rich *Cannon* estimates resulting in strong CN bands as shown in Figs A4 and A5. Thus, these pervasive CN lines may blend with atomic features, resulting in higher [Fe/H] estimates by the *Cannon* models for these stars. These stars are also found to belong to the second-generation population in NGC 104 based on their enhanced N, Na, Al, and low O, Mg abundances from GALAH SME. At the same time, majority of NGC 104 members in our APOGEE sample (ACGS and ACAA) belongs to the first-generation population
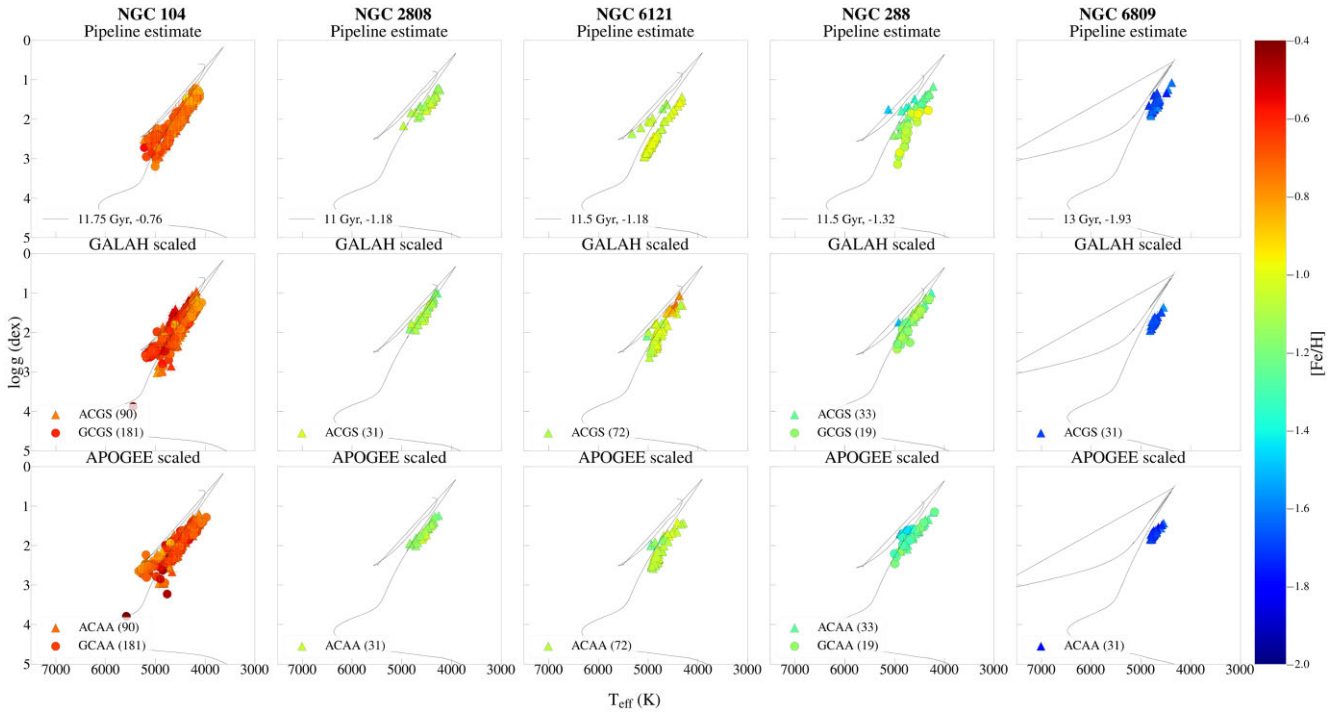
Figure 11. Kiel diagram for members of five globular clusters from the latest *Gaia* EDR3 globular cluster catalogue (Vasiliev & Baumgardt 2021) colour coded by their metallicities. Clusters are arranged from left to right in decreasing order of their median metallicities from literature. We show survey pipeline (GALAH SME and APOGEE ASPCAP) estimates in the top row, GALAH-scaled values (ACGS and GCGS) in the middle row and APOGEE-scaled values (GCAA and ACAA) in the bottom row. PARSEC isochrones (Bressan et al. 2012) based on cluster ages and metallicities adopted from literature (VandenBerg et al. 2013) are over plotted (black line) in each panel. The values estimated from APOGEE spectra (ASPCAP as well as *Cannon* estimates) are indicated as triangle markers, while those from GALAH spectra (SME as well as *Cannon* estimates) are indicated as circle markers.

based on their ASPCAP light element abundances. Hence, we do not see similar difference in [Fe/H] *Cannon* estimates compared to the pipeline estimates using APOGEE spectra (ACGS and ACAA).

For all clusters excluding young open clusters, we find a mean scatter of 0.06 dex in APOGEE-scaled metallicity estimates and slightly higher scatter of 0.065 dex in GALAH-scaled metallicity estimates. we find a smaller mean scatter of 0.03 dex in both APOGEE and GALAH-scaled alpha abundance estimates. Overall, we find the *Cannon* metallicity estimates to be consistent with that in literature and follow expected theoretical isochrone tracks in the Kiel diagram, showing the quality of our GALAH- and APOGEE-scaled catalogues.

4.1.3 Chemical trends in clusters

Though the median metallicities from the *Cannon* estimates show good consistency with that from respective survey pipelines, as well as those from the literature, it is important to check for any significant trends in chemistry (metallicity and alpha abundance) with temperature. For cluster members (both open and globular clusters), we expect such trends to be non-existent owing to the chemical homogeneity of their birth cloud. At the same time, atomic diffusion can lead to differences in surface abundance values for different evolutionary phases (i.e. main sequence turn-off stars and giants) of cluster members as well (Dotter et al. 2017). In addition to atomic diffusion, Spina et al. (2020) have shown that stellar parameters and abundances vary as a function of the stellar activity in young stars. Hence, we also

have to take this into account while investigating chemical trends in clusters.

In this work, where we have used the *Cannon* to estimate GALAH-scaled labels from APOGEE spectra (and GALAH spectra) and APOGEE-scaled labels from GALAH (and APOGEE spectra), we expect the *Cannon* estimates to follow the trends we see in the case of survey labels that have been used to train the model with. Thus, we expect GALAH-scaled cases (ACGS and GCGS) to exhibit similar trends as seen with GALAH SME values and APOGEE-scaled cases (GCAA and ACAA) that of APOGEE ASPCAP. All figures discussed below are in the same format as in previous section (Figs 10 and 11), with the top row showing the survey pipeline (SME/ASPCAP) estimates, middle and bottom rows showing the GALAH-scaled and APOGEE-scaled values, respectively, for the same stars. Estimates from GALAH spectra are indicated in blue circles and from APOGEE spectra in red triangles, with the median metallicity values in each case indicated with respectively coloured-dashed lines.

In Fig. 12, first five columns show metallicity trends of five open clusters with respect to T_{eff} and the next five columns show [α/Fe] trends of the same clusters with respect to T_{eff}. We have arranged them from left to right in increasing order of their ages, from which it is clear that chemical trends with T_{eff} are prominent for youngest clusters (IC 4665, Blanco 1, and Melotte 22). From GALAH SME estimates (first three columns in top row, blue circles), we see a slight negative trend for metallicity and positive trend for [α/Fe] with increase in T_{eff}. As mentioned above, this could be the effect of stellar activity in young cluster members that is

Table 6. Median metallicity and standard deviation estimated for members of five globular clusters from the latest *Gaia* EDR3 globular cluster catalogue (Vasiliev & Baumgardt 2021) cross-matched with the catalogues in this work. The names of clusters are listed in the first column, metallicity of the cluster from literature (VandenBerg et al. 2013) in the second column, the abbreviations for the catalogues (see Section 3.5) with the number of stars from each catalogue in brackets in the third column, *Cannon* metallicity estimates from each catalogue in the fourth column, and survey pipeline metallicity estimates corresponding to the spectra mentioned (in the catalogue name) in the same row in the last column.

Cluster	$[Fe/H]_{literature}$	Catalogue (No.)	$[Fe/H]_{Cannon}$ (median, std dev)	$[Fe/H]_{Survey}$ (median, std dev)
NGC 104	−0.76	GCGS (181)	− 0.66, 0.08	− 0.73, 0.07
		ACGS (90)	− 0.73, 0.09	− 0.71, 0.04
		GCAA (181)	− 0.67, 0.07	− 0.73, 0.07
		ACAA (90)	− 0.76, 0.04	− 0.71, 0.04
NGC 2808	−1.18	GCGS (0)	−	−
		ACGS (31)	− 1.10, 0.06	− 1.10, 0.05
		GCAA (0)	−	−
		ACAA (31)	− 1.12, 0.05	− 1.10, 0.05
NGC 6121	−1.18	GCGS (0)	−	−
		ACGS (72)	− 1.03, 0.04	− 1.01, 0.04
		GCAA (0)	−	−
		ACAA (72)	− 1.07, 0.03	− 1.01, 0.04
NGC 288	−1.32	GCGS (19)	− 1.12, 0.05	− 1.07, 0.04
		ACGS (33)	− 1.24, 0.06	− 1.24, 0.04
		GCAA (19)	− 1.26, 0.09	− 1.07, 0.04
		ACAA (33)	− 1.26, 0.07	− 1.24, 0.04
NGC 6809	−1.93	GCGS (0)	−	−
		ACGS (31)	− 1.67, 0.04	− 1.67, 0.07
		GCAA (0)	−	−
		ACAA (31)	− 1.71, 0.06	− 1.67, 0.07
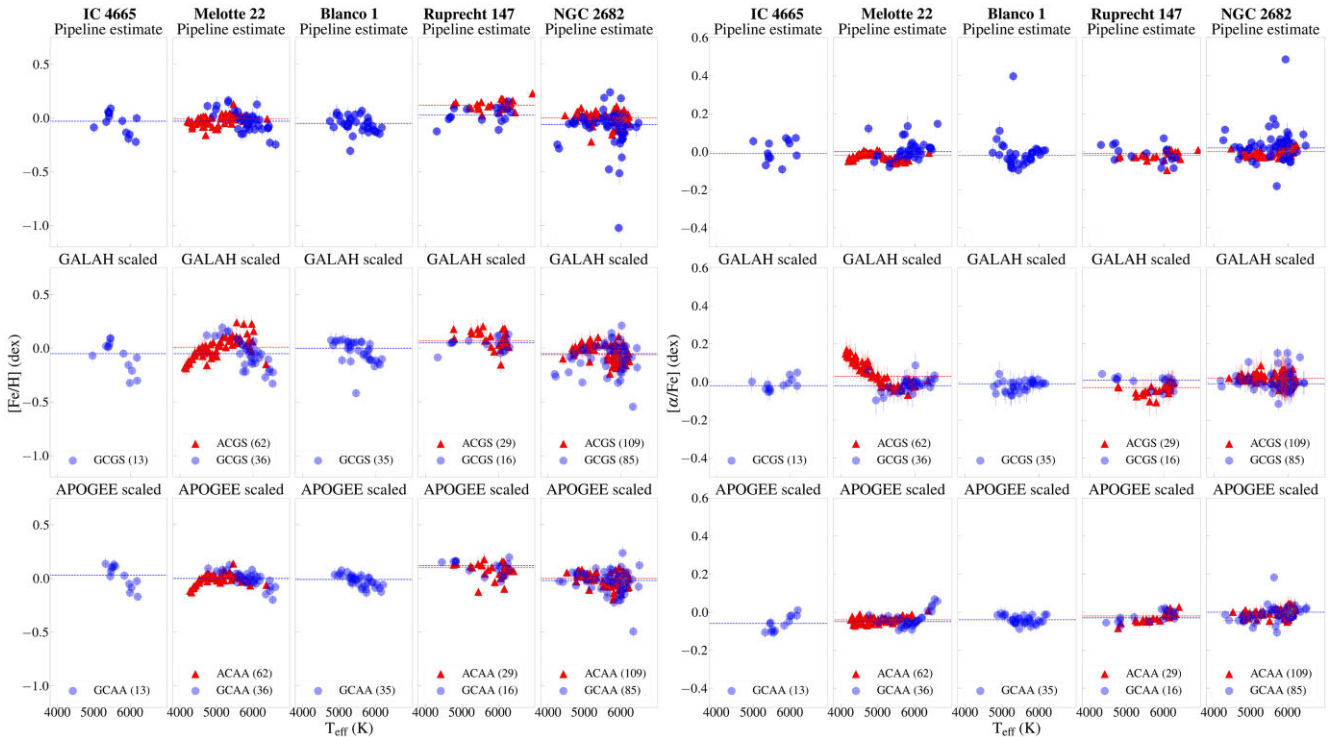
Figure 12. [Fe/H] versus T_{eff} (left five columns) and [α/Fe] versus T_{eff} (right five columns) plots for five open clusters. We show survey pipeline (GALAH SME and APOGEE ASPCAP) estimates in the top row, GALAH-scaled values (ACGS and GCGS) in the middle row and APOGEE-scaled values (GCAA and ACAA) in the bottom row. Estimates from GALAH spectra are indicated in blue circles and from APOGEE spectra in red triangles, with the median metallicity and [α/Fe] values in each case indicated with respectively coloured dashed lines.

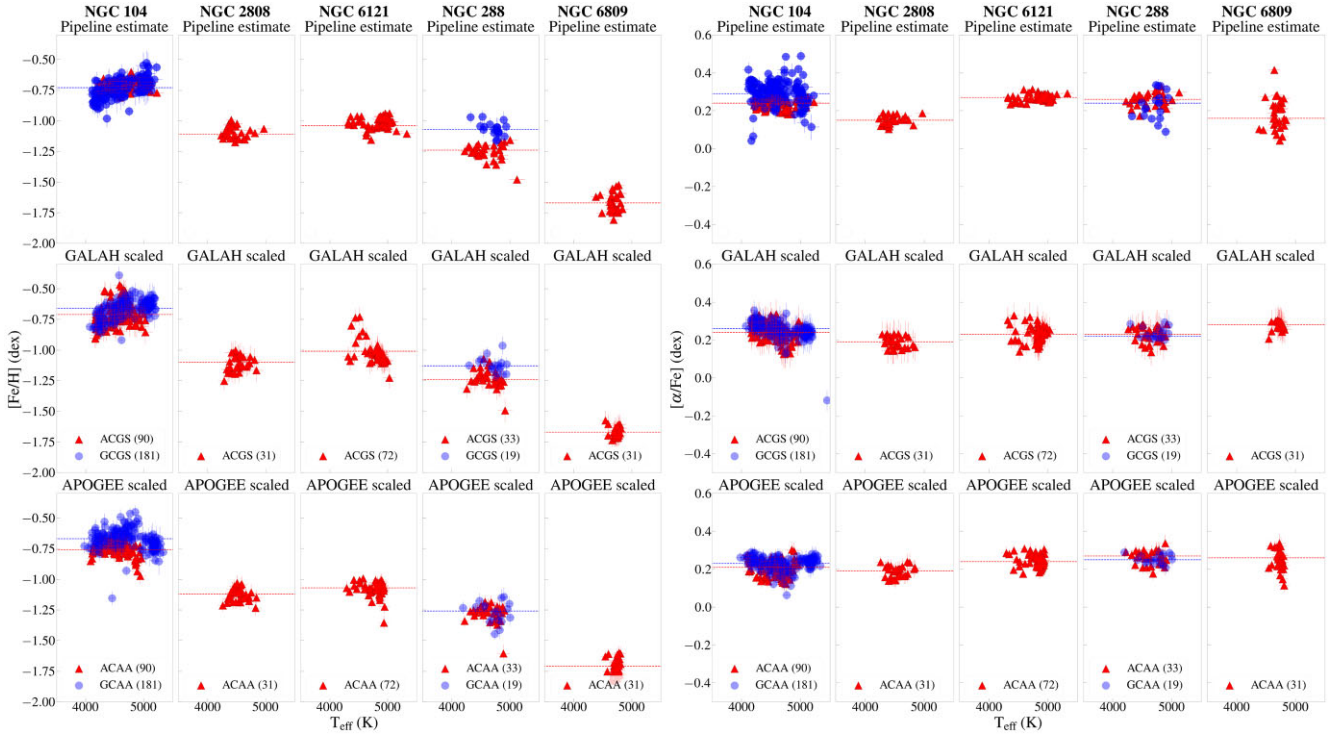**Figure 13.** [Fe/H] versus T_{eff} (left five columns) and [α/Fe] versus T_{eff} (right five columns) plots for five globular clusters in the same format as in Fig. 12.

expected to strengthen saturated lines, resulting in a complicated interplay between effects on T_{eff}, $\log g$ and chemical abundances (Spina et al. 2020). Among young clusters, only Melotte 22 has stars from both APOGEE and GALAH. For this cluster, GALAH-scaled metallicities and alpha abundances enhance the trends imprinted from the SME pipeline compared to APOGEE-scaled estimates. APOGEE ASPCAP [α/Fe] values in Melotte 22 shows a wave-like trend with T_{eff}, which completely disappears in the APOGEE-scaled values from APOGEE spectra. Meanwhile, GALAH-scaled [α/Fe] values are higher for the same stars, a trend that is seen in GALAH SME estimates for cool stars in Blanco 1. Meanwhile, this trend is removed in the case of APOGEE- and GALAH-scaled [α/Fe] values for Blanco 1. For older open clusters, Ruprecht 147 and NGC 2682, there are no significant trends seen in survey pipeline estimates and correspondingly in the *Cannon* estimates, with respect to T_{eff}. We also note that for NGC 2682, the median metallicity for the GALAH-scaled case and APOGEE-scaled case align with the respective survey pipeline estimates.

In Fig. 13, first five columns show metallicity trends of five globular clusters with respect to T_{eff} and the next five columns show [α/Fe] trends of the same clusters with respect to T_{eff}. We have arranged them from left to right in decreasing order of their metallicities. Unlike in open clusters, there are no obvious trends with T_{eff} for survey pipeline estimates and *Cannon* estimates. In all clusters, GALAH-scaled estimates have larger standard deviations or scatter as listed in the Table 6. As mentioned in the previous section, for NGC 104, differences in spectra due to strong CN bands in second-generation stars result in higher metallicity values estimated from GALAH spectra in both GALAH-scaled and APOGEE-scaled cases (see Figs A4 and A5). In addition to this, there are few GALAH stars for which APOGEE-scaled T_{eff} is higher (by $\sim$100–150 K) than GALAH SME estimates. On further investigation, these are found to be red clump stars. In the training set, we find similar difference

in T_{eff} between APOGEE ASCPAP and GALAH SME estimates for red clump stars in the same GALAH T_{eff} and $\log g$ ranges. Thus, this offset is likely real that is propagated from the systematic differences in training set.

In the case of NGC 288, GALAH SME and APOGEE ASPCAP metallicity values are different ($\sim$0.2 dex as shown in Table 6). GALAH-scaled metallicity estimates for GALAH and APOGEE stars also show difference which is smaller than that between survey pipeline estimates. Meanwhile, APOGEE-scaled metallicity estimates from both set of stars are consistent with the APOGEE ASPCAP estimate. At the same time, GALAH-scaled and APOGEE-scaled [α/Fe] estimates are consistent for both APOGEE and GALAH stars and with respective survey pipeline estimates. In Fig. 14, we show metallicity trends with $\log g$ for survey pipeline estimates (top panel), GALAH-scaled estimates (middle panel) and APOGEE-scaled estimates (bottom panel). There is a clear offset between the $\log g$ values of GALAH and APOGEE pipeline estimates (see Fig. 14), which resulted in the shift from theoretical isochrone tracks seen in Fig. 11. This offset is found to be the result of bad parallax measurements for NGC 288 members (located at a distance of $\sim$9 kpc) leading to wrong GALAH SME $\log g$ estimates (equation 1 in Buder et al. 2021). Interestingly, GALAH-scaled and APOGEE-scaled $\log g$ for GALAH stars are shifted to correct $\log g$ values (blue circles in middle and bottom panels in Fig. 14). This could also explain the improvement in APOGEE-scaled and GALAH-scaled metallicity estimates for GALAH stars. Thus, our *Cannon* models are able to improve incorrect pipeline estimates.

Overall, we find that in most cases, GALAH-scaled and APOGEE-scaled metallicities and [α/Fe] values are following the trends exhibited by the respective survey pipeline estimates. We also find GALAH-scaled estimates to have comparatively higher scatter compared to APOGEE-scaled values, which is expected from their respective survey pipeline values. The plots discussed above thus
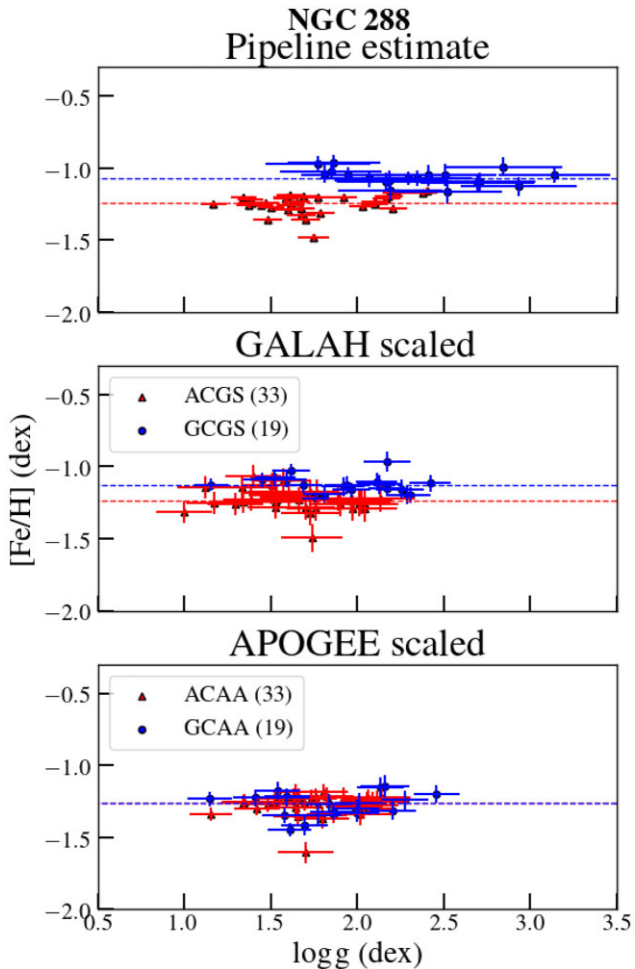
Figure 14. [Fe/H] versus log *g* for NGC 288 with survey pipeline (GALAH SME and APOGEE ASPCAP) estimates in the top row, GALAH-scaled values (ACGS and GCGS) in the middle row and APOGEE-scaled values (GCAA and ACAA) in the bottom row. Estimates from GALAH spectra are indicated in blue circles and from APOGEE spectra in red triangles, with the median metallicity in each case indicated with respectively coloured dashed lines.

clearly display both qualities, as well as limitations, of our catalogues and show the ability of the *Cannon* to carry out cross survey scaling of large data sets from high-resolution spectroscopic surveys.

5 DISCUSSION AND CONCLUSIONS

We used the data-driven approach, the *Cannon*, to produce two catalogues of stellar parameters and general alpha abundances, one scaled in terms of the APOGEE survey and the other in terms of the GALAH survey. We chose the training set from among ~20 000 stars commonly observed in both the surveys, which after quality cuts (removing stars with low SNR, bad spectra, bad pipeline estimates, etc) resulted in a final training set sample of 418 stars. We have shown the importance of cross survey scaling with plots showing the systematic differences in stellar parameters, metallicities ,and alpha abundances using common observed stars in the training set. While there is reasonably good agreement between the two surveys (especially for T_{eff} and log *g*), there are systematic differences in [Fe/H], [α/Fe], v_{micro} and v_{broad} for the same stars in both surveys (Fig. 2). Most notably, [Fe/H] for giants and hot dwarfs are underestimated by GALAH when compared to APOGEE, while GALAH [α/Fe]

measurements are overestimated for giants. Most of these offsets could be attributed to differences in the parameter estimation methods implemented in the APOGEE and GALAH surveys. We used our final training set to train four *Cannon* models in four combinations of spectra and labels, i.e. two models with spectra and labels from the different surveys respectively, and two models with spectra and labels from the same survey.

With the 12-fold cross-validation of stars in the training set (see Sections 3.4 and A1), we have shown that the *Cannon* models predict labels/parameters that are consistent with the input labels. We further demonstrate in Figs 5 and 6 that the model generated spectra with the *Cannon* estimates are very similar (16th and 84th percentile values of residuals at majority of pixels within −0.01 and 0.01) to the respective observed survey spectra. We have also shown that pixels with higher residuals are either bad/noisy pixels or lines of elements that have not been modelled by our *Cannon* models. We also demonstrate the effectiveness of our method in Figs 7 and 8, where we remade the Fig. 2 for common stars in APOGEE and GALAH but with Cannon estimates from GALAH-scaled and APOGEE scaled catalogues. We show that the mean systematic difference for metallicity and alpha abundance is very close to zero with lesser scatter and no significant trends in comparison with survey pipeline estimates in Fig. 2.

These models are then used to estimate labels from 437 445 APOGEE and 576 390 GALAH spectra, thus providing two combined catalogues with stellar parameters scaled to both the APOGEE and GALAH surveys. We carried out the validation of these catalogues with selected open and globular cluster members by comparing survey pipeline estimates and *Cannon* estimates in kiel diagrams and chemical trend plots as shown in Figs 10–13. In Kiel diagrams, both GALAH-scaled and APOGEE-scaled *Cannon* estimates follow respective theoretical isochrone tracks (based on literature ages and metallicities) even when survey pipeline estimates do not (NGC 2808, NGC 6121, and NGC 288). In addition, median metallicity estimates within standard deviations are consistent with literature values thus validating our *Cannon* estimates. In the case of T_{eff} trends for metallicities and alpha abundance in clusters, as expected, *Cannon* estimates follow spurious trends exhibited by respective pipeline estimates to which they are scaled. Meanwhile, there are a few cases where GALAH-scaled metallicities and alpha abundances enhance the trends imprinted from the SME pipeline, particularly for the very young clusters.

Significant offsets of ~0.5–1 dex between GALAH SME and APOGEE ASPCAP log *g* estimates for NGC 288 members (located at a distance of ~9 kpc) in Fig. 14 showed that incorrect parallaxes can have an effect on log *g* determination in GALAH. Interestingly, log *g* values of GALAH stars in both GALAH-scaled and APOGEE-scaled cases have been shifted to the correct range. This has also led to the correction of the overestimated GALAH SME metallicities in the APOGEE-scaled case and significant improvement in the GALAH-scaled case. This shows the ability of the *Cannon* models to correct survey pipeline estimates and illustrates the quality of our *Cannon* estimates.

We have thus demonstrated the effectiveness of our method to carry out cross survey scaling and showed the quality of the resulting *Cannon* estimates in both catalogues. However, there are some limitations and caveats in the method that we briefly discuss below.

5.1 Limitations in the training set

The training set is the starting point and one of the major factors that can affect parameter estimation using data-driven approaches like the *Cannon*. While the training set should be a representative of different
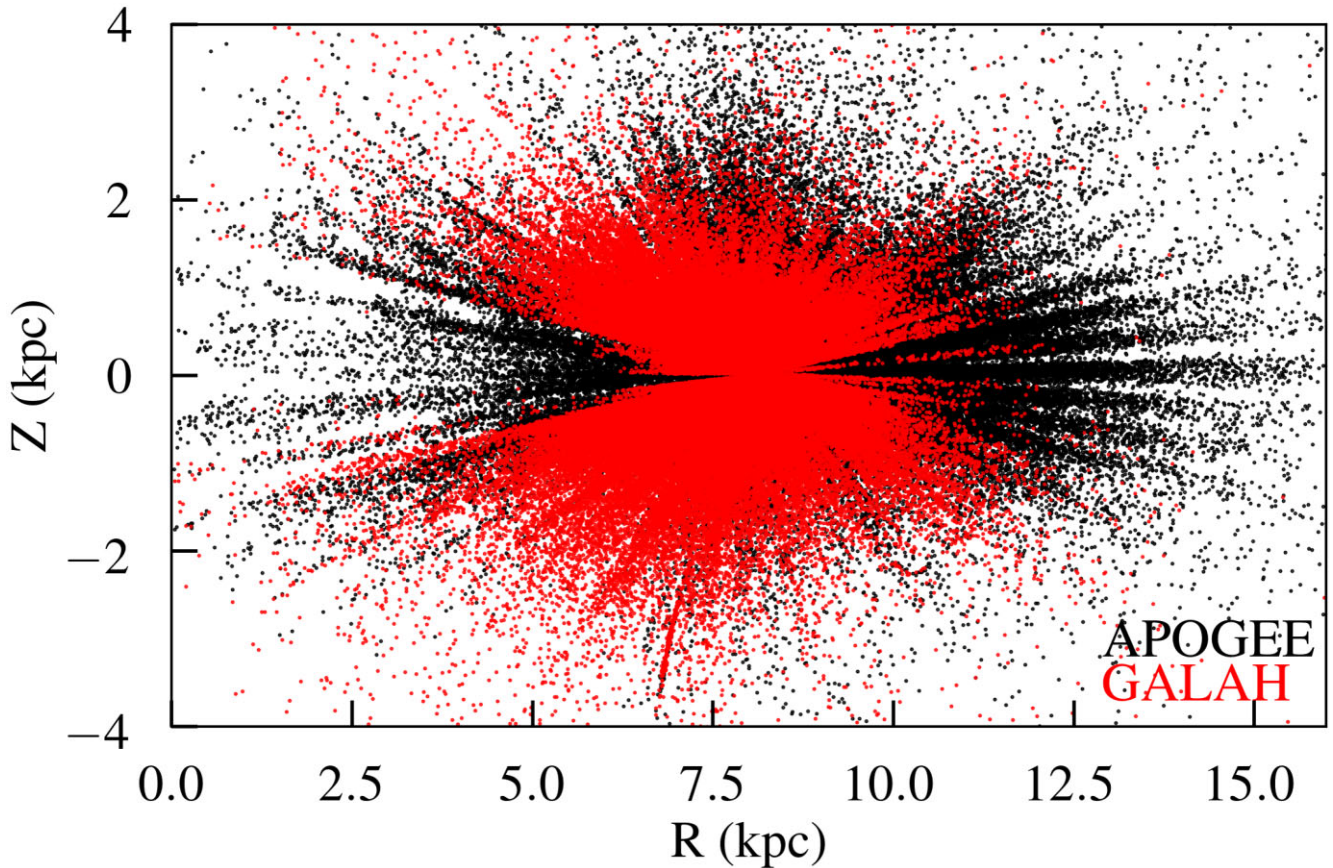
Figure 15. Z versus R distribution of stars with valid *Cannon* estimates based on flags set in both GALAH-scaled (black) and APOGEE-scaled (red) stellar parameter catalogues. R and Z are estimated based on the *Gaia* EDR3 distances from Bailer-Jones et al. (2021).

populations of stars that the survey (from which the input labels are taken) observed, the labels and spectra should also be reliable and of high quality. Since we are restricted to choosing the common stars observed in both the surveys for the training set in this work, we have limited options regarding the former criterion. We have carried out quality cuts as recommended by each survey for both the labels, as well as spectra, ensuring a good-quality training set. Ideally, in cases where the spectra and label are from the same survey (ACAA and GCGS), one can use a larger training set. Since our goal is to put the surveys on the same scale, we choose a single training set that passes all quality criteria and have sufficient number statistics. Thus our training set is limited by the fact that not all the best quality spectra or labels in the respective catalogue are included in it, rather we select the best labels from among the common stars observed in both the surveys. Still our final training set sample has reasonably broad coverage of the parameter space of both surveys (see Fig. 1), but one should exercise caution at the edges of the training set, in particular for metal-poor ([Fe/H]<-0.8) stars and for M-dwarfs, for which the *Cannon* labels deviate from the one-to-one relation. We have assigned dedicated flags that will help the user to identify such stars (see Section 3.6).

5.2 Propagation of parameter trends

By reproducing APOGEE-scaled labels from GALAH spectra and GALAH-scaled labels from APOGEE spectra, we are propagating the trends/issues in these parameters that arise from the input survey pipeline/analysis method. Such issues are evident in the labels when

we compare them in Fig. 2. These include the lower metallicities determined for metal rich giants in GALAH and large differences in microturbulence and broadening values for stars in GALAH as well as APOGEE (see Section 3.2). We show the inability of the *Cannon* models to reproduce APOGEE microturbulence values in Figs 4 and A2. We also see the effect of this on *Cannon* temperature and surface gravity estimates in the case of ACAA (see Fig. A2). When we adopted the empirical relations used in GALAH to redetermine microturbulence values for APOGEE, these trends are corrected (see Fig. A3). Thus, it is important to have accurate and reliable input labels in order to get better results from data-driven methods and machine learning tools and future surveys should aim to achieve this.

5.3 Limitations in the error determination

With repeat observations, we have been able to show that the errors for the labels from the *Cannon* are either incorrect or underestimated. We then estimate a scaling factor for each label and rescale their *Cannon* covariance errors. Using repeat observations, we also estimated precision for all labels as a function of SNR in the range of 36–50 K (T_{eff}), 0.06–0.1 dex (log g), 0.02–0.04 dex ([Fe/H]), and 0.02–0.03 dex ([α/Fe]) for SNR > 40 in GALAH and SNR > 80 in APOGEE. We choose the maximum value among rescaled covariance uncertainty and precision estimate at respective SNR to be the final error for each label. Thus, our errors are determined without taking into account any possible additional dependence on temperature and surface gravity as it is done in APOGEE (Jönsson

et al. 2020). Thus, there is a possibility that our errors are still underestimated.

Finally, we have ∼1 million stars with APOGEE-scaled and GALAH-scaled parameters in both catalogues, with a spatial coverage that spans inner and outer parts of the Milky Way, including thin disc, thick disc, and halo components.The mean metallicity and alpha abundance errors is 0.07 and 0.06 dex in the GALAH-scaled case, and 0.06 and 0.03 dex in the APOGEE-scaled cases. Once we implement flags in both catalogues to remove stars outside the training set boundary (*flag_Cannon_dist*), stars with bad spectra (*flag_spectra*), bad SME and ASPCAP labels (*flag_sp_aspcap* and *flag_survey*), and good snr (*flag_snr*), we have ∼280 000 GALAH stars and ∼170 000 stars from APOGEE, with a spatial coverage that spans the mid-plane, as well as halo regions of the Milky Way as shown in Fig. 15. Thus, we end up with less than 50 per cent of the total number of stars in the catalogues, but with good quality spectra and the parameters on a common scale in each catalogue. This is still an impressive number of stars with good spatial coverage to study the metallicity and alpha abundance trends of the Milky Way components.

Combining these catalogues with the latest *Gaia* EDR3 catalogue enables a comprehensive chemo-dynamic study of these different components that constitutes the Milky Way from the perspective of APOGEE and GALAH, separately. We have cross matched with *Gaia* EDR3 and included source ids from both DR2 and DR3, along with distances estimated by Bailer-Jones et al. (2021) in the final catalogues as listed in the table schema, Table A1.

5.4 Final conclusions

With this work, we have demonstrated the ability of the data-driven method, *Cannon*, to reliably estimate respective survey-scaled stellar parameters, metallicity, and alpha abundance. While having a few caveats, we find that the *Cannon* estimates are as good as or showed improvement over the respective survey pipeline estimates. These are encouraging signs to use currently available data-driven and machine learning tools to do cross-survey scaling with many more ongoing and upcoming spectroscopic surveys.

Among the limitations and caveats discussed above, the limited size and coverage of parameter space by the common stars observed in both surveys for the training set is foremost. This will hopefully improve in coming years, as both GALAH and APOGEE South continue to observe larger samples of stars going forward. Improvements in stellar model atmospheres and incorporation of 3D NLTE models in spectroscopic analysis will enable ongoing, as well as upcoming ground-based surveys, like SDSS-V (Kollmeier et al. 2017), 4MOST (de Jong et al. 2019), WEAVE (Dalton et al. 2018) etc., to achieve better accuracy and precision in stellar parameter and elemental abundances. Thus, we will have all the necessary ingredients to construct training sets that can be used to carry out cross-survey scaling, which will provide catalogues of stars with consistent stellar parameters and elemental abundances with a significant coverage of all components of the Galaxy.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data underlying this article are available in the Data Central at https://cloud.datacentral.org.au/teamdata/GALAH/public/GALAH_DR3/

REFERENCES

Adibekyan V. Z., Sousa S. G., Santos< N. C., Delgado Mena E., González Hernández J. I., Israelian G., Mayor M., Khachatryan G., 2012, A&A, 545, A32
Ahumada R. et al., 2020, APJS, 249, 3
Anders F. et al., 2014, A&A, 564, A115
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147
Barden S. C., 2010, Proceedings of the SPIE, 7735, 773509
Bedell M. 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., Lundström I., 2003, A&A, 410, 527
Bensby T., Feltzing S., Oey M. S., 2014, A&A, 562, A71
Blanton M. R. et al., 2017, AJ, 154, 28
Bowen I. S., Vaughan A. H. J., 1973, Appl. Opt., 12, 1430
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Buder S. et al., 2021, MNRAS, 506, 150
Buder S. et al., 2018, MNRAS, 478, 4513
Buder S. et al., 2019, A&A, 624, A19
Cantat-Gaudin T. et al., 2020, A&A, 640, 19
Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, A&A, 530, A138
Casey A. R. et al., 2017, ApJ, 840, 59
Casey A. R., Hogg D. W., Ness M., Rix H.-W., Ho A. Q. Y., Gilmore G., 2016, preprint (arXiv:1603.03040)
Cui X.-Q. et al., 2012, Res. Astron. Astrophys., 12, 1197
Dalton G. et al., 2018, SPIE, 10702, 107021B
de Jong R. S. et al., 2019, The Messenger, 175, 3
De Silva G. M. et al., 2015, MNRAS, 449, 2604
Dotter A., Conroy C., Cargile P., Asplund M., 2017, ApJ, 840, 99
Eisenstein D. J. et al., 2011, AJ, 142, 72
Fabbro S., Venn K. A., O'Briain T., Bialek S., Kielty C. L., Jahandar F., Monty S., 2018, MNRAS, 475, 2978

Freeman K., Bland-Hawthorn J., 2002, ARA&A, 40, 487
Fuhrmann K., 1998, A&A, 338, 161
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2018, A&A, 616, A1
García Pérez A. E. et al., 2016, AJ, 151, 144
Gilmore G. et al., 2012, The Messenger, 147, 25
Gilmore G., Reid N., 1983, MNRAS, 202, 1025
Gunn J. E. et al., 2006, AJ, 131, 2332
Gustafsson B., Edvardsson B., Eriksson K., Jørgensen U. G., Nordlund Å.,
    Plez B., 2008, A&A, 486, 951
Hasselquist S. et al., 2020, ApJ, 901, 19
Hayden M. R. et al., 2015, ApJ, 808, 132
Haywood M., Di Matteo P., Lehnert M. D., Katz D., Gómez A., 2013, A&A,
    560, A109
Heiter U., Soubiran C., Netopil M., Paunzen E., 2014, A&A, 561, A93
Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown
    A. G. A., 2018, Nature, 563, 85
Ho A. Y. Q. et al., 2017, ApJ, 836, 5
Ibata R. A., Malhan K., Martin N. F., 2019, ApJ, 872, 152
Jönsson H. et al., 2020, AJ, 160, 120
Kollmeier J. A. et al., 2017, preprint (arXiv:1711.03234)
Kos J. et al., 2017, MNRAS, 464, 1259
Leung H. W., Bovy J., 2019, MNRAS, 483, 3255
Lin J. et al., 2020, MNRAS, 491, 2043
Majewski S. R. et al., 2017, AJ, 154, 94
Minchev I., Famaey B., 2010, ApJ, 722, 112
Minniti D. et al., 2010, New Astron., 15, 433
Myeong G. C., Vasiliev E., Iorio G., Evans N. W., Belokurov V., 2019,
    MNRAS, 488, 1235
Ness M. et al., 2012, ApJ, 756, 22
Ness M. et al., 2013, MNRAS, 430, 836
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., Martig M., Pinsonneault M. H., Ho A. Y.
    Q., 2016, ApJ, 823, 114
Nissen P. E., Christensen-Dalsgaard J., Mosumgaard J. R., Silva Aguirre V.,
    Spitoni E., Verma K., 2020, A&A, 640, A81
Perryman M. A. C. et al., 1997, A&A, 500, 501
Piskunov N., Valenti J. A., 2017, A&A, 597, A16
Reddy B. E., Lambert D. L., Allende Prieto C., 2006, MNRAS, 367,
    1329
Reddy B. E., Tomkin J., Lambert D. L., Allende Prieto C., 2003, MNRAS,
    340, 304
Rojas-Arriagada A. et al., 2017, A&A, 601, A140
Rojas-Arriagada A. et al., 2020, MNRAS, 499, 1037
Schönrich R., Binney J., 2009, MNRAS, 399, 1145
Sellwood J. A., Binney J. J., 2002, MNRAS, 336, 785
Sharma S. et al., 2018, MNRAS, 473, 2004
Sharma S. et al., 2019, MNRAS, 490, 5335
Sheinis A. et al., 2015, J. Astron. Telesc. Instrum. Syst., 1, 035002
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Spina L. et al., 2020, ApJ, 895, 52
Spina L. et al., 2021, MNRAS, 503, 3279
Steinmetz M. et al., 2006, AJ, 132, 1645
Stoughton C. et al., 2002, AJ, 123, 485
Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf.
    Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV.
    Astron. Soc. Pac., San Francisco, p. 29
Taylor M. B., 2020, in Ballester P., Ibsen J., Solar M., Shortridge K., eds, ASP
    Conf. Ser. Vol. 522, Astronomical Data Analysis Software and Systems
    XXVII. Astron. Soc. Pac., San Francisco, p. 67
Ting Y.-S., Conroy C., Rix H.-W., Cargile P., 2019, ApJ, 879, 69
Valenti J. A., Piskunov N., 1996, A&AS, 118, 595
VandenBerg D. A., Brogaard K., Leaman R., Casagrande L., 2013, ApJ, 775,
    134
Vasiliev E., Baumgardt H., 2021, MNRAS, 505, 5978

Weinberg D. H. et al., 2019, ApJ, 874, 102
Wheeler A. et al., 2020, ApJ, 898, 58
Wilson J. C. et al., 2012, in McLean I. S., Ramsay S. K.,
    Takami H., eds, SPIE Conf. Ser. Vol. 8446, Ground-based and
    Airborne Instrumentation for Astronomy IV. SPIE, Bellingham,
    p. 84460H
Wilson J. C. et al., 2019, PASP, 131, 055001
Wittenmyer R. A. et al., 2018, AJ, 155, 84
Wolf C. et al., 2018, Publ. Astron. Soc. Aust., 35, e010
Xiang M. et al., 2019, ApJS, 245, 34
Yoshii Y., 1982, PASJ, 34, 365
Zhang X., Zhao G., Yang C. Q., Wang Q. X., Zuo W. B., 2019, PASP, 131,
    094202

APPENDIX A: *GALAH CANNON GALAH SME (GCGS)* & *APOGEE Cannon APOGEE ASPCAP (ACAA)*

We use the training set to generate *Cannon* model trained on GALAH (APOGEE) spectra using GALAH SME (APOGEE ASPCAP) labels, which is then used to predict stellar parameters for rest of GALAH (APOGEE) spectra. We carry out 12-fold cross-validation, error estimation from repeat observations and internal validation for both cases as shown below.

A1 Cross-validation

We follow the method described in Section 3.4 to carry out 12-fold cross-validation of the *Cannon* estimates for six labels using the training set. In the case of *GALAH Cannon GALAH SME*, we compare the *Cannon* outputs with input GALAH SME labels, while we compare the *Cannon* outputs with input APOGEE ASPCAP labels for *APOGEE Cannon APOGEE ASPCAP*. Figs A1 and A2 show the cross-validation results for *GALAH Cannon GALAH SME* and *APOGEE Cannon APOGEE ASPCAP*, respectively

In the case of *GALAH Cannon GALAH SME*, all labels follow a tight one-to-one relation as indicated by the mean and scatter values as well. Largest scatter (0.48 km s^{-1}) is seen in the case of v_{broad}, which is still much lower than in the case of *APOGEE Cannon GALAH SME*.

In the case of *APOGEE Cannon APOGEE ASPCAP*, all labels except v_{micro} and v_{broad} follow a tight one-to-one relation. Still, we find the *Cannon* to under estimate T_{eff} for cool stars ($T_{eff, APOGEE} < 4500$ K) and stars with $T_{eff, APOGEE}$ 5000–5400 K, while slightly over estimate T_{eff} for stars with $T_{eff, APOGEE}$ in between these two limits. Similarly, we find the *Cannon* to slightly over predict $\log g$ for stars with $\log g_{APOGEE} > 4$ dex.

There are significant deviations in the case of v_{micro} and v_{broad}, similar to that seen in the case of *GALAH Cannon APOGEE ASPCAP* (see Fig. 4). We find the *Cannon* outputs for input $v_{micro\ APOGEE} > 1$ km s^{-1} follow one-to-one relation, whereas the *Cannon* estimates have higher dispersion for lower $v_{micro\ APOGEE}$ values. This indicates that the *Cannon* model is unable to find significant correlation between $v_{micro\ APOGEE}$ values and corresponding APOGEE spectra. Much tighter one-to-one trend is seen in the case of v_{broad}, though there are significant deviations for lower $v_{broad\ APOGEE}$ values. As mentioned in Section 3.4, this again possibly indicates the inefficiency/problems associated with the way v_{micro} and v_{broad} have been determined for APOGEE stars.

Table A1. Table schema of APOGEE-scaled and GALAH-scaled stellar parameter catalogues.

Column name	Unit	Description
Unique ID		Unique survey star id
dr2_source_id		*Gaia* dr2 ID
dr3_source_id		*Gaia* dr3 ID
ra	deg	Right ascension
dec	deg	Declination
teffCann	K	*Cannon* estimate of effective temperature
e_teffCann	K	error for teffCann
loggCann	$\log(\mathrm{cm\,s^{-2}})$	*Cannon* estimate of surface gravity
e_loggCann	$\log(\mathrm{cm\,s^{-2}})$	error for loggCann
fehCann	dex	*Cannon* estimate of metallicity
e_fehCann	dex	error for fehCann
alphafeCann	dex	*Cannon* estimate of [α/Fe]
e_alphafeCann	dex	error for alphafeCann
vmicroCann	$\mathrm{km\,s^{-1}}$	*Cannon* estimate of microturbulence
e_vmicroCann	$\mathrm{km\,s^{-1}}$	error for vmicroCann
vbroadCann	$\mathrm{km\,s^{-1}}$	*Cannon* estimate of broadening velocity
e_vbroadCann	$\mathrm{km\,s^{-1}}$	error for vbroadCann
r_chi_sq		reduced chi-square from *Cannon*
snrCann		Signal-to-noise ratio from *Cannon*
Cannon_distance		Distance to the training set labels
flag_id		0 – GALAH spectra, 1 – APOGEE spectra
flag_Cannon_dist		flag for quality of *Cannon* estimates (0/1)
flag_spectra		flag for problematic spectra
flag_sp_aspcap		survey pipeline flag
flag_survey		invalid survey pipeline estimates (1)
flag_training		star belonging to the training set (0)
teff	K	Survey effective temperature
e_teff	K	Survey effective temperature error
logg	$\log(\mathrm{cm\,s^{-2}})$	Survey surface gravity
e_logg	$\log(\mathrm{cm\,s^{-2}})$	Survey surface gravity error
feh	dex	Survey metallicity
e_feh	dex	Survey metallicity error
alphafe	dex	Survey [α/Fe]
e_alphafe	dex	Survey [α/Fe] error
vmicro	$\mathrm{km\,s^{-1}}$	Survey microturbulence
e_vmicro	$\mathrm{km\,s^{-1}}$	Survey microturbulence error
vbroad	$\mathrm{km\,s^{-1}}$	Survey broadening velocity
e_vbroad	$\mathrm{km\,s^{-1}}$	Survey broadening velocity error
snr		Survey signal-to-noise ratio
r_est	pc	Estimated distance from Bailer-Jones+21
r_lo	pc	Lower bound from Bailer-Jones+21
r_hi	pc	Higher bound from Bailer-Jones+21
parallax	mas	Gaia EDR3 parallax
e_parallax	mas	Gaia EDR3 parallax error

To investigate this further, we carry out an exercise in which instead of the APOGEE v_{micro}, we make use of the empirical relation adopted in GALAH (equations 4 and 5 in Buder et al. 2021) to determine v_{micro} for APOGEE as well. We do this since the *Cannon* could correlate the v_{micro} from GALAH SME with APOGEE spectra in *APOGEE Cannon GALAH SME* (see Fig. 3). Including this new v_{micro}, we train *Cannon* model on APOGEE spectra using APOGEE labels. The cross-validation results from this exercise are shown in the Fig. A3. First of all, there is a tight one-to-one relation for v_{micro}. This is no surprise since the new empirical v_{micro} is a function of T_{eff} and $\log g$, and thus the *Cannon* could identify tight correlation with corresponding spectra as well. This is again reflected in the T_{eff} and $\log g$ estimated by the *Cannon*, which no longer show the issues that we found when APOGEE v_{micro} is used to train the *Cannon* model. This can be considered as an evidence of the inefficiency/problems associated with the way v_{micro} have been determined for APOGEE stars.
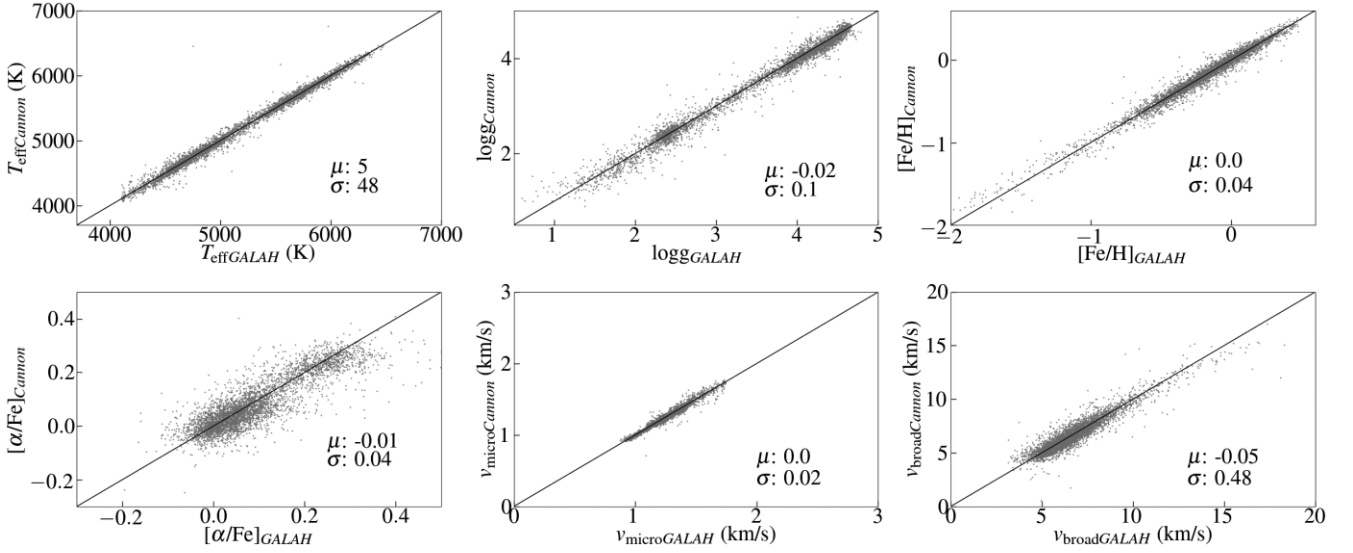
Figure A1. One-to-one relation of GALAH SME stellar labels (*x*-axis) in the training set versus corresponding values estimated by the *Cannon* from GALAH spectra (*y*-axis) after carrying out a 12-fold cross-validation test. The mean and scatter (calculated as the mid-value of 84th−16th percentile) in each label difference is indicated on the bottom right-hand side of each plot.
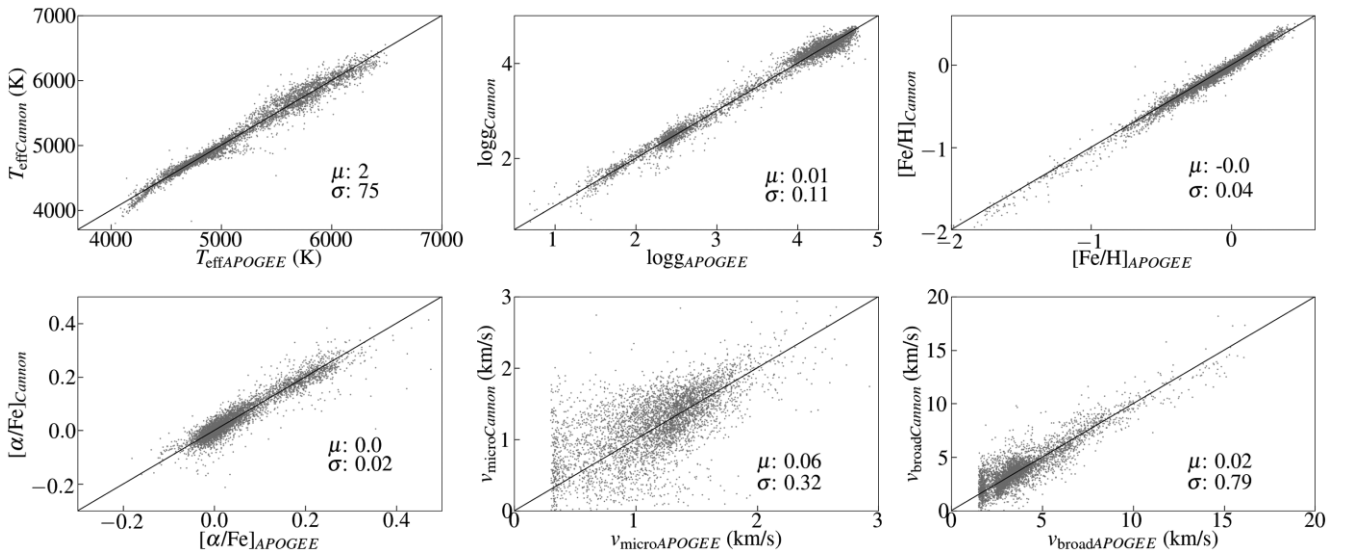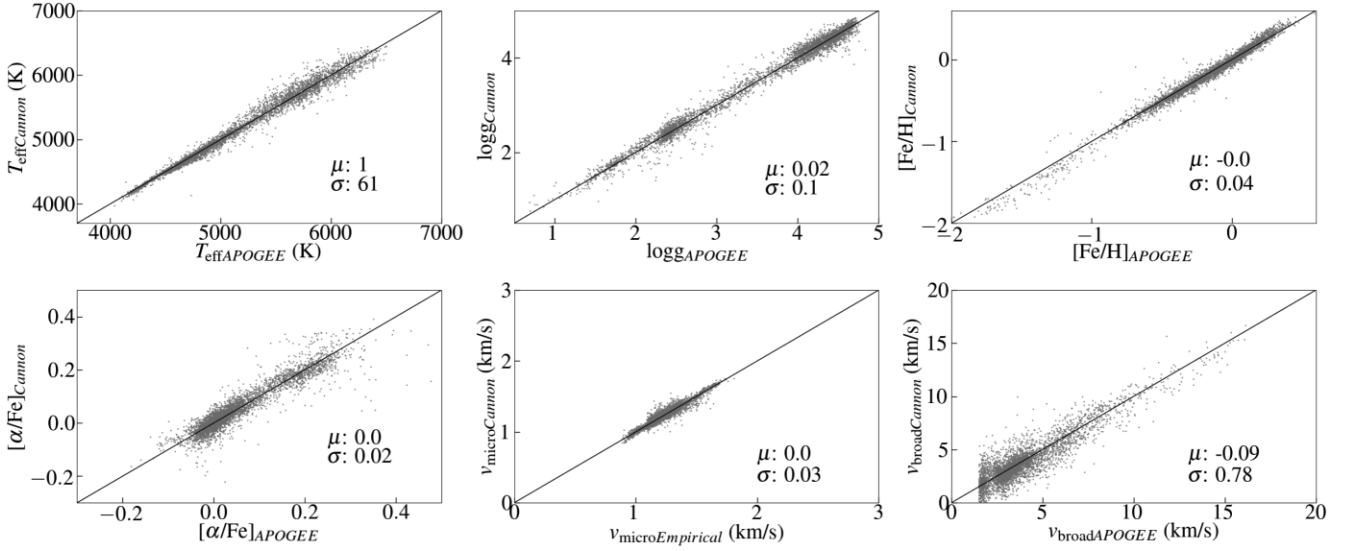
Figure A2. One-to-one relation of APOGEE ASPCAP stellar labels (x-axis) in the training set versus corresponding values estimated by the *Cannon* from APOGEE spectra (y-axis) after carrying out a 12-fold cross-validation test. The mean and scatter (calculated as the mid value of $84^{th}-16^{th}$ percentile) in each label is indicated on the bottom right hand side of each plot.

Figure A3. One-to-one relation of APOGEE ASPCAP stellar labels (*x*-axis) in the training set versus corresponding values estimated by the *Cannon* from APOGEE spectra (*y*-axis) after carrying out a 12-fold cross-validation test. The mean and scatter (calculated as the mid-value of 84th−16th percentile) in each label difference is indicated on the bottom right-hand side of each plot.
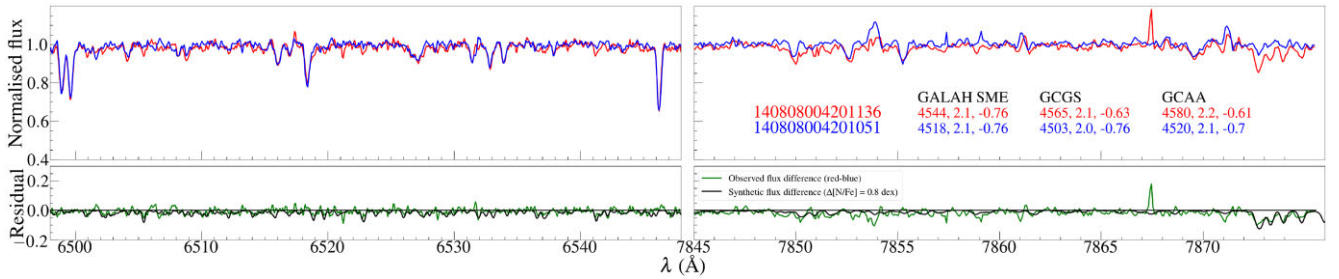

Figure A4. GALAH spectra for two NGC 104 members with similar stellar parameters from GALAH. In blue, we show star spectrum for which GALAH- and APOGEE-scaled stellar parameters from GALAH spectra (GCGC and GCAA) are consistent with GALAH SME. Star spectrum in red is an example where GCGS and GCAA metallicity estimates are more metal rich than GALAH SME value. This inconsistency can be attributed to the difference between the two spectra (red-blue), shown in green in the bottom panels (mean flux difference is indicated in these panels), arising due to pervasive CN lines that blend with atomic features in the red spectrum resulting in more metal-rich Cannon estimates.
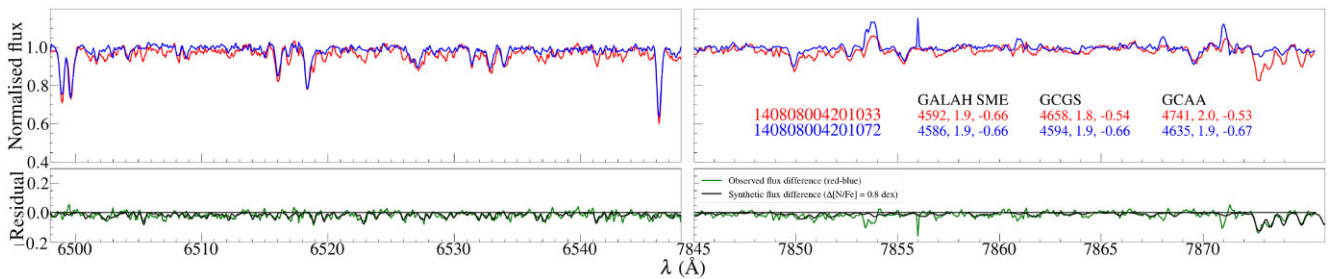

Figure A5. Same as Fig. A4 but for another pair of stars with similar stellar parameters in NGC 104.

APPENDIX B: METALLICITY DIFFERENCE IN NGC 104

As mentioned in Section 4.1.2, GCGS and GCAA metallicity estimates are found to be more metal-rich than both survey pipeline estimates, as well as *Cannon* estimates from APOGEE spectra. To understand the reason for this, we compare GALAH spectra of two stars in NGC 104 with similar stellar parameters, out of which GCGS and GCAA metallicity is the same as GALAH SME estimate for one star and not for the other. Two such examples in the red and infrared wavelength ranges are shown in Figs A4 and A5, where spectra in blue are stars with consistent *Cannon* estimate and red spectra represent the inconsistent case. Residuals resulting from difference between red and blue spectra is shown in green in

the bottom panels. We have computed illustrative synthetic spectra following Buder et al. (2021) for the relevant stellar parameters with a solar chemical composition, as well as with an enhancement in N of 0.8 dex, the difference between which is plotted in black in the same panel. The residual from observed spectra clearly follows that from the synthetic spectra, especially evident from the CN bands at $> \sim$ 7870 Å, suggesting that the stars with more metal-rich *Cannon* estimates (red spectra) have enhanced N abundances and thus stronger CN bands.

This paper has been typeset from a TEX/LATEX file prepared by the author.