



HAL
open science

Assessment of $[\text{Fe}/\text{H}]$ determinations for FGK stars in spectroscopic surveys

C. Soubiran, N. Brouillet, L. Casamiquela

► **To cite this version:**

C. Soubiran, N. Brouillet, L. Casamiquela. Assessment of $[\text{Fe}/\text{H}]$ determinations for FGK stars in spectroscopic surveys. *Astronomy and Astrophysics - A&A*, 2022, 663, pp.A4. 10.1051/0004-6361/202142409 . hal-03711970

HAL Id: hal-03711970

<https://hal.science/hal-03711970>

Submitted on 1 Jul 2022

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

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

Assessment of [Fe/H] determinations for FGK stars in spectroscopic surveys[★]

C. Soubiran[✉], N. Brouillet[✉], and L. Casamiquela

Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
e-mail: caroline.soubiran@u-bordeaux.fr

Received 11 October 2021 / Accepted 13 December 2021

ABSTRACT

Context. The iron abundance [Fe/H] in the atmosphere of FGK-type stars is crucial in stellar and galactic physics. The number of stars with a measured value of [Fe/H] is increasing considerably thanks to spectroscopic surveys. However, different methodologies, inputs, and assumptions used in spectral analyses lead to different precisions in [Fe/H] and possibly to systematic differences, which need to be evaluated. It is essential to understand the characteristics of each survey to fully exploit their potential, in particular if the surveys are combined to probe a larger galactic volume and to improve statistics.

Aims. The purpose of this study is to compare [Fe/H] determinations from the largest spectroscopic surveys to other catalogues taken as reference. Offsets and dispersions of the residuals are examined, as are their trends with other parameters. The investigated surveys are the latest public releases of APOGEE, GALAH, RAVE, LAMOST, SEGUE, and the *Gaia*-ESO Survey.

Methods. We use reference samples that provide independent determinations of [Fe/H], which are compared to those from the surveys for common stars. The distribution of the residuals is assessed through simple statistics that measure the offset between two catalogues and the dispersion representative of the precision of both catalogues. When relevant, linear fits are performed. A large sample of FGK-type stars with [Fe/H] based on high-resolution, high-signal-to-noise spectroscopy was built from the PASTEL catalogue to provide a reference sample. We also use FGK members in open and globular clusters to assess the internal consistency of [Fe/H] of each survey. The agreement of median [Fe/H] values for clusters observed by different surveys is discussed.

Results. All the surveys overestimate the low metallicities, and some of them also underestimate the high metallicities. They perform well in the most populated intermediate metallicity range whatever the resolution. In most cases, the typical precision that we deduce from the comparisons is in good agreement with the uncertainties quoted in the catalogues. Some exceptions to this general behaviour are discussed.

Key words. stars: abundances – surveys

1. Introduction

We are in the middle of a new era where stellar atmospheric parameters (APs) and abundances are produced on massive scales by spectroscopic surveys. The iron abundance [Fe/H] is an essential stellar property that has to be known for the determination of other parameters through stellar models, such as the mass and the age. Iron abundances are also needed in galactic archeology in order to understand how the different stellar populations have formed and evolved.

The pioneer of spectroscopic surveys, the Radial Velocity Experiment (RAVE), produced its first data release (DR) 15 years ago (Steinmetz et al. 2006) and its final one, DR6, last year (Steinmetz et al. 2020a,b). In the meantime, other surveys have been operated, and survey designers have developed new methodologies, learning progressively from their own and one another's experience on how to reduce biases and uncertainties in the automated determination of APs and abundances from massive datasets of stellar spectra with various resolutions and spectral coverages. Besides RAVE, other surveys have published successive DRs that are available for public use. At the time of writing, there is open access to the DR16 of the Apache Point Observatory Galactic Evolution Experiment (APOGEE;

Jönsson et al. 2020), the DR3 of the Galactic Archaeology with HERMES project (GALAH; Buder et al. 2021), the DR6 of RAVE (Steinmetz et al. 2020a,b), the DR5 of the Large sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST; Luo et al. 2015, 2019), Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009), and the *Gaia*-ESO Survey (DR3; Gilmore et al. 2012; Randich & Gilmore 2013). Additional versions of APs, based on different methods, are also provided for RAVE DR6 (Guiglion et al. 2020) and for LAMOST DR5 (Xiang et al. 2019). The next generation of optical and near-infrared spectrographs, wide-field and massively multiplexed, is in preparation and will soon provide even larger catalogues of APs and abundances, such as the WHT Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2012), the Multi-Object Optical and Near-infrared Spectrograph (MOONS) on ESO's Very Large Telescope (Taylor et al. 2018), the 4-metre Multi-Object Spectroscopic Telescope (4MOST; de Jong et al. 2019), the Prime Focus Spectrograph (PFS; Takada et al. 2014), and the Maunakea Spectroscopic Explorer (MSE; McConnachie et al. 2016). In terms of numbers of stars, the most revolutionary survey will certainly be that of the *Gaia* space mission (Gaia Collaboration 2016), which will deliver in its DR3 in 2022¹ estimates of the physical properties, including metallicities, for millions of stars obtained with

[★] The catalogue of atmospheric parameters of FGK stars is only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](ftp://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/663/A4>

¹ The *Gaia* DR scenario and the DR3 content can be found at <https://www.cosmos.esa.int/web/gaia>

various methods through an astrophysical parameter inference system (Bailer-Jones et al. 2013).

Each survey has its own strategy for the calibration and validation of APs and abundances. The term calibration usually invokes standard stars with true APs. While effective temperatures and surface gravities, T_{eff} and $\log g$, can be determined independently of atmospheric models thanks to fundamental relations (Heiter et al. 2015), this is not the case for the metallicity [Fe/H], which has therefore no absolute zero point. Abundances are expressed relative to the Sun, the chemical composition of which is still subject to debate (Asplund et al. 2009, 2021). It is thus impossible to measure the typical accuracy of a given catalogue of metallicities; however, the zero-point agreement between two catalogues can be assessed instead. It is also possible to evaluate the relative precisions of different catalogues by comparing them to another independent source. In classical spectroscopy, the assessment of uncertainties usually evaluates the random errors due to the characteristics of the input spectra and to the line selection, as well as the systematic errors due to the adopted assumptions, for example local thermodynamic equilibrium (LTE), or to the method itself, for example equivalent width versus synthetic spectrum fitting. Comparisons to independent reference datasets and inter-comparisons of surveys are mandatory for tracking systematic differences, although these comparisons are limited by the number of stars in common and their range of properties. The strategies adopted by the ongoing surveys for calibrations and validations are reviewed in Jofré et al. (2019).

The validation of the *Gaia*'s APs is challenging due to the size of the dataset, the large magnitude range, and the observing mode. *Gaia* collects all the objects down to a limiting magnitude, including stars with properties that prevent a reliable determination of APs via automated methods (e.g., rotation, emission, binarity). All the information from ground-based surveys and catalogues of APs is being used to assess the accuracy and precision of *Gaia*'s APs. In this context, an important task is to deepen our knowledge and understanding of the AP uncertainties of ground-based surveys. Any systematic difference between large surveys has potentially important implications for the study of stellar populations in the Milky Way and the galactic chemical evolution. It is also important to make these comparisons in the perspective of combining different surveys to probe a larger galactic volume and to improve the statistics, as attempted by Nandakumar et al. (2020) for instance.

In this study we focus on the [Fe/H] of FGK-type stars in the effective temperature range 4000–6500 K. The upper limit avoids hot stars, the metallicity of which can be affected by rotation and chemical peculiarities. The lower limit avoids cool giants and dwarfs, the metallicity of which is reputedly difficult to measure because of many blended lines in the spectra. FGK stars span the full age range of the Galaxy, with their chemical composition reflecting the chemical composition of the interstellar matter from which they formed, from very low to very high metallicity. FGK members in open clusters (OCs) and globular clusters (GCs) are supposed to share the same iron abundance. Indeed, high-precision differential studies have shown that the chemical homogeneity is at the level of 0.02 dex in OCs (Liu et al. 2016; Casamiquela et al. 2020) and of 0.03 in most GCs (e.g., Yong et al. 2013). This property offers the possibility to assess the precision of a given catalogue by measuring the typical dispersion of [Fe/H] among members of a given cluster. The consistency of the metallicity can be tested all over the temperature range of FGK stars and among giants and dwarfs. Many OCs and GCs have been observed for decades in spectroscopy,

so their chemical composition is reasonably known. Clusters are ideal for multi-object spectroscopy, and a significant observing time is dedicated to them by spectroscopic surveys. *Gaia* DR2 (Gaia Collaboration 2018b) and Early Data Release 3 (EDR3; Gaia Collaboration 2021) have considerably enlarged the number of stars and clusters for which membership probabilities are available (Gaia Collaboration 2018a; Cantat-Gaudin et al. 2018; Cantat-Gaudin & Anders 2020; Vasiliev & Baumgardt 2021). This offers an opportunity to revise the metallicity of clusters.

In order to evaluate the precision and zero-point agreement of [Fe/H] determinations in surveys, we constructed three reference samples based on: (1) the PASTEL catalogue (Soubiran et al. 2016), (2) OC members, and (3) GC members. PASTEL, updated in 2020, compiles [Fe/H] determinations based on high-resolution, high-S/N spectroscopy, with a significant fraction of metal-poor stars. Together with GCs, it provides a means to test the poorly constrained low metallicity part of the AP space. Our procedure measures the dispersion of the residuals resulting from the [Fe/H] comparison between the different surveys and the reference catalogues and looks for trends with magnitude and APs.

In this paper we first present our compilation of cluster members and the PASTEL catalogue. We then briefly present the investigated surveys and the selections applied on T_{eff} , [Fe/H] errors, flags, or other criteria, depending on the survey, to retain the best quality APs. We cross-match the resulting samples to the reference catalogues to evaluate the [Fe/H] residuals and their dependence on other parameters. We discuss the results in terms of the typical precision of the surveys in the metal-rich and metal-poor range as well as in the giant and dwarf subspaces. We also compare the surveys to APOGEE. In the whole paper we use median values, denoted MED, to measure offsets. We evaluate the dispersion of the various distributions measured through their median absolute deviation, denoted MAD. When relevant, we fit a line to highlight a trend.

2. Reference catalogues

2.1. Cluster members

For the purpose of testing the metallicity precision of the surveys, we selected clusters with reasonably known metallicity and their members of highest probability.

For OCs we adopted the mean metallicities per cluster compiled by Netopil et al. (2016) for 172 OCs. In that paper, the metallicities are derived from spectroscopy at high resolution and high S/N (88 clusters) or lower resolution (12 clusters) or from photometry (72 clusters). The spectroscopic metallicities are updated from Heiter et al. (2014). We completed this compilation with the high-precision and homogeneous mean [Fe/H] determined by Casamiquela et al. (2021) for 47 OCs based on clump giants only. This adds 18 clusters. We adopted the metallicity from Casamiquela et al. (2021) over that of Netopil et al. (2016) for three clusters (NGC 7245, NGC 6940, and King 1) because of a photometric determination and for two other clusters (NGC 2266 and NGC 2639) because of a spectroscopic metallicity that relies on one star only. The resulting compilation of mean [Fe/H] per cluster is not of homogeneous quality, but our purpose is to use clusters to test the precision of surveys, not their accuracy, so an absolute reference value is not mandatory. The most important thing is to have a large number of reliable members per cluster, giving a significant intersection with the surveys. In the following we analyse the distribution of

the [Fe/H] residuals in terms of dispersion per cluster since an offset seen for a given OC could be due to an erroneous mean metallicity from the literature. Despite the different precisions of the reference metallicities of OCs, trends can still be observed.

We adopted the list of members with a probability higher than 70% of belonging to their parent cluster that Cantat-Gaudin et al. (2020) used to determine the physical properties of ~ 2000 OCs based on *Gaia* DR2 data. All the stars have their *Gaia* magnitude $G_{\text{mag}} < 18$. Three clusters from Netopil et al. (2016) were not found in Cantat-Gaudin et al. (2020) (Saurer 1, Loden 807, and Collinder 173). This gives 77 899 stars in 187 OCs, spanning metallicities from -0.50 to $+0.43$. The number of members per cluster ranges from 11 to nearly 3000.

For GCs, we adopted the catalogue of Harris (2010), which provides the metallicity of 152 GCs in the Milky Way, and the membership probabilities recently computed by Vasiliev & Baumgardt (2021) using *Gaia* EDR3 for 170 galactic GCs. We selected the most reliable members that have $G_{\text{mag}} < 18$ and a membership probability higher than 70%, and we kept only the GCs that have at least ten members remaining after these cuts. It is well established that most GCs have a scatter in metallicity lower than 0.05 dex (Carretta et al. 2009); there are a few exceptions that show a larger dispersion, which is possibly related to multiple populations (Gratton et al. 2012). These objects, NGC 5139 (Omega Cen), NGC 6715 (M 54), NGC 6656 (M 22), Terzan 5, NGC 1851, and NGC 2419, have been removed from our reference sample. This gives 146 147 stars in 134 GCs, spanning metallicities from -2.37 to 0 dex.

2.2. The PASTEL catalogue

The compilation of APs started in the 1980s with the so-called [Fe/H] catalogue (Cayrel de Strobel et al. 1980, 1981, 1985, 1992, 1997, 2001) and continued in 2010 with the PASTEL catalogue, which was regularly updated until 2020 (Soubiran et al. 2010, 2016). Only [Fe/H] determinations based on high-resolution ($R \geq 25\,000$), high-S/N ($S/N \geq 50$) spectra are recorded in PASTEL, with a few exceptions as explained in Soubiran et al. (2016). PASTEL also provides effective temperatures and surface gravities determined from various methods. PASTEL does not include AP determinations from spectroscopic surveys that have a resolution and S/N fitting the criteria (e.g., GALAH and the UVES part of the *Gaia*-ESO Survey).

As of January 2020, PASTEL has 81 362 records, including 42 932 determinations of the three APs (T_{eff} , $\log g$, [Fe/H]) for 18 119 different stars. The [Fe/H] determinations range from -4.80 to $+2.40$ dex. The solar metallicity is by far the most frequent value. More than 80% of the [Fe/H] determinations are between -0.50 dex and $+0.50$ dex. In the literature monitoring, a particular effort was put on trying to be as complete as possible for metal-poor stars. Such stars are rare in the solar neighbourhood, so observers have to consider targets at larger distances, which are fainter and challenging for spectroscopic observations at high resolution and high S/N. Nevertheless, PASTEL includes a significant number of AP determinations for metal-poor stars: 5544 values with $[\text{Fe}/\text{H}] \leq -1.0$ dex (~ 2000 different stars), 1973 values with $[\text{Fe}/\text{H}] \leq -2.0$ dex (~ 850 different stars), and 418 values with $[\text{Fe}/\text{H}] \leq -3.0$ dex (~ 240 different stars). The top five most studied metal-poor stars are HD140283 (weighted mean $[\text{Fe}/\text{H}] = -2.47 \pm 0.03$ dex, for $N = 58$ determinations), HD 103095 ($[\text{Fe}/\text{H}] = -1.34 \pm 0.02$ dex, $N = 44$), HD19445 ($[\text{Fe}/\text{H}] = -1.98 \pm 0.03$ dex, $N = 43$), Tau Cet ($[\text{Fe}/\text{H}] = -0.51 \pm 0.01$ dex, $N = 43$), and HD 122563 ($[\text{Fe}/\text{H}] =$

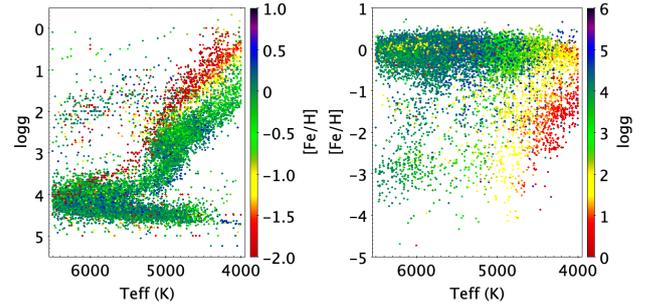


Fig. 1. AP distribution for the mean PASTEL catalogue limited to the FGK regime. *Left:* Kiel diagram coloured according to [Fe/H]. *Right:* [Fe/H] versus T_{eff} coloured according to $\log g$.

-2.67 ± 0.02 dex, $N = 42$). PASTEL includes 18 extremely metal-poor stars with $[\text{Fe}/\text{H}] \leq -4.0$ dex. Most have been studied only once at high resolution. The most studied one is CD-38 245 ($[\text{Fe}/\text{H}] = -4.10 \pm 0.08$ dex, $N = 12$).

For a practical use of the PASTEL catalogue in the comparison to other catalogues, it is necessary to have a single value of AP per star. There are different strategies for doing this, and we adopted the weighted average of the parameters, although we are aware that systematics make this procedure theoretically improper. However, PASTEL compiles more than 1200 papers with very different numbers of targets, making any kind of homogenisation impossible.

The mean APs and uncertainties were computed for each star with a weighting scheme based on the uncertainty of the individual measurements, following the method described in Soubiran et al. (2013). Errors on (T_{eff} , $\log g$, [Fe/H]) listed in PASTEL have median values of 50 K, 0.1 dex, and 0.06 dex, respectively, for FGK stars with $4000 \leq T_{\text{eff}} \leq 6500$ K, but not all the individual AP determinations are given with an error. Therefore, these median values are adopted as default errors for each AP determination, doubled when $[\text{Fe}/\text{H}] < -1.0$ dex and doubled again if the year of publication is before 1990. The error adopted for the weighting scheme is the maximum between this default value and that provided in PASTEL, when available. Flagged values of [Fe/H], corresponding to global metallicities [M/H] or to non-LTE abundances or based on ionised iron lines, are not considered in the average.

The resulting mean PASTEL catalogue, available in Vizier, provides the three parameters (T_{eff} , $\log g$, [Fe/H]) for 14 181 FGK stars, of which 13 506 are in *Gaia* DR2. For the 7400 stars that have at least two [Fe/H] determinations, the median uncertainty of the mean is 0.04 dex (0.06 dex for metal-poor stars with $[\text{Fe}/\text{H}] < -0.50$ dex). The Kiel diagram of the mean PASTEL APs with $4000 \text{ K} \leq T_{\text{eff}} \leq 6500 \text{ K}$ is shown in Fig. 1 together with the metallicity distribution as a function of T_{eff} .

2.3. Clusters in PASTEL

The cross-match between the mean PASTEL catalogue and the list of reference OC members gives 590 common FGK stars in 87 clusters. Figure 2 (upper panel) shows the metallicity difference between individual stars and the literature mean value as a function of the G magnitude, of T_{eff} and $\log g$ from PASTEL, and of the literature [Fe/H] of the cluster. The distribution of residuals is flat; there is no trend. The median difference is null, and the dispersion is $\text{MAD} = 0.04$, in perfect agreement with the typical uncertainty of the mean PASTEL [Fe/H]. The [Fe/H] value of individual stars agrees well in general with the mean

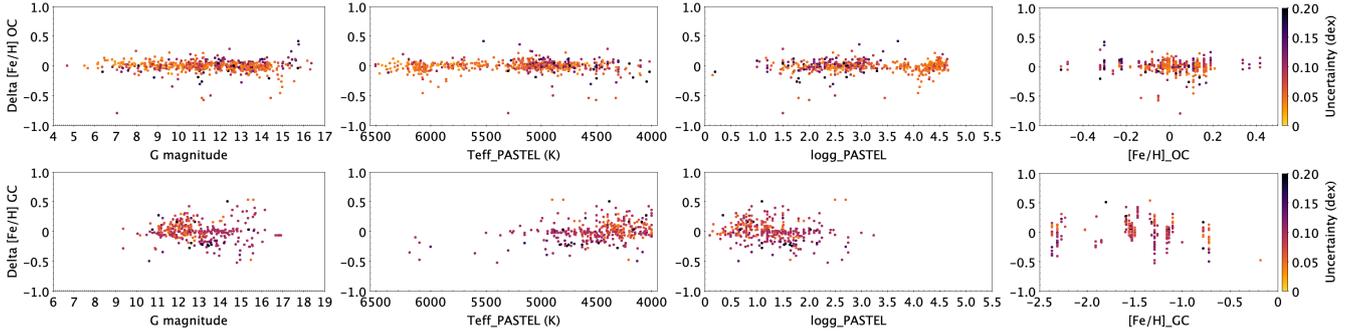


Fig. 2. Difference between mean PASTEL determinations of $[\text{Fe}/\text{H}]$ for individual members and the mean value per cluster from [Netopil et al. \(2016\)](#) for OCs (*upper panel*) and from [Harris \(2010\)](#) for GCs (*bottom panel*) versus G magnitude, T_{eff} , $\log g$ from PASTEL, and the mean $[\text{Fe}/\text{H}]$ per cluster. The colour is related to the $[\text{Fe}/\text{H}]$ uncertainty from PASTEL.

OC metallicity from the literature, which is not surprising since the previous version of PASTEL was used by [Heiter et al. \(2014\)](#) and [Netopil et al. \(2016\)](#) to build the compilation of OC metallicities. The two most represented OCs in PASTEL are M67 and the Hyades with, respectively, 91 and 55 FGK stars giving a metallicity of 0.0 ± 0.04 dex and $+0.14 \pm 0.03$ dex (MED \pm MAD).

The cross-match between the mean PASTEL catalogue and the list of reference GC members gives 350 common FGK stars in 29 clusters. The metallicity residuals are shown in Fig. 2 (bottom panel). The median difference is -0.01 dex, and the dispersion is MAD 0.08 dex, slightly larger than for the metal-poor field stars mentioned in the previous section. Four GCs show a remarkably low dispersion in metallicity, with a MAD lower or equal to 0.01 dex: NGC 2808 (MED $[\text{Fe}/\text{H}] = -1.12$ dex for 23 stars), NGC 4833 (MED $[\text{Fe}/\text{H}] = -2.02$ dex for 12 stars), NGC 7078 (MED $[\text{Fe}/\text{H}] = -2.37$ dex for 38 stars), and NGC 7078 (MED $[\text{Fe}/\text{H}] = -2.34$ dex for 9 stars). The other GCs with at least five members in PASTEL also have low dispersions, with their MAD ranging from 0.01 to 0.08 dex. The only exception is NGC 5904 (M5), which exhibits a large dispersion of 0.19 dex (MED $[\text{Fe}/\text{H}] = -1.31$ dex for 33 stars), clearly visible in Fig. 2. The chemical composition of this cluster has been extensively studied in the past. The dispersion that we observe in PASTEL reflects the fact that the authors of different analyses do not agree on the metallicity of this cluster. [Snedden et al. \(1992\)](#) and [Carretta & Gratton \(1997\)](#) report metallicities of individual members, giving on average $[\text{Fe}/\text{H}] = -1.17 \pm 0.01$ dex and -1.11 ± 0.03 dex for the cluster, respectively, while [Lai et al. \(2011\)](#) determine metallicities ranging from -1.82 to -1.33 dex.

In the next sections we cross-match the mean PASTEL catalogue with the most recent versions of spectroscopic surveys in order to compare $[\text{Fe}/\text{H}]$ determinations. Although the results of each comparison encompass the systematic errors and uncertainties of both PASTEL and the compared survey, they provide relevant information on the strengths and weaknesses of the datasets. In addition, we apply to each survey the same tests on stellar clusters that reveal typical dispersions in different ranges of magnitude, T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$, independently of any other catalogue.

3. Surveys versus reference catalogues

3.1. APOGEE

APOGEE DR16 ([Jönsson et al. 2020](#)) includes about 430 000 stars with APs. APOGEE spectra have a resolution of $\sim 22\,500$ and cover the near-infrared range from $15\,140 \text{ \AA}$ to $16\,940 \text{ \AA}$.

The APOGEE Stellar Parameters and Abundances Pipeline (ASCAP; [García Pérez et al. 2016](#)) compares APOGEE observations to a large library of synthetic spectra (from MARCS models for the FGK stars considered here) and determines the best matching synthetic spectrum using the code FERRE ([Allende Prieto et al. 2006](#)), which additionally allows for interpolation within the library. The spectroscopic T_{eff} and $\log g$ are then calibrated. The $[\text{Fe}/\text{H}]$ is measured in a second step by tuning the fit around iron lines. The uncertainty on $[\text{Fe}/\text{H}]$ is set through a function of the T_{eff} , global metallicity, and S/N of the spectrum, the coefficients of which are deduced from repeat observations. We used here the cleaned and calibrated $[\text{Fe}/\text{H}]$, as recommended, which verify `FE_H_FLAG = 0`. Together with the FGK selection $4000 \leq T_{\text{eff}} \leq 6500 \text{ K}$, this gives 236 966 stars with a median uncertainty of 0.01 dex.

There are 2155 stars in common between PASTEL and this APOGEE sample, 844 FGK cluster members in 43 OCs and 1958 FGK cluster members in 48 GCs. The residuals are shown in Fig. 3 as a function of magnitude and of (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$).

For PASTEL the residuals exhibit a rather low dispersion (MAD = 0.05), very similar to that computed by [Jönsson et al. \(2018\)](#) when comparing previous APOGEE DRs to independent studies. The few outliers correspond to stars with larger uncertainties. There is a trend in $[\text{Fe}/\text{H}]$ showing that APOGEE systematically overestimates metallicities of metal-poor stars ($[\text{Fe}/\text{H}] < -0.50$ dex) and underestimates those of the most metal-rich stars ($[\text{Fe}/\text{H}] > 0$ dex) compared to PASTEL. The offset is $+0.06$ dex in the metal-poor regime, while there is a decreasing trend with $[\text{Fe}/\text{H}]$ in the metal-rich regime (a linear fit gives a slope of -0.18 dex^{-1}). These metallicity trends seem to also be present in the residuals of clusters. [Nidever et al. \(2020\)](#) reported an offset of $+0.08$ dex for GC metallicities from APOGEE DR16 compared to high quality determinations from [Carretta et al. \(2009\)](#), which they considered consistent with uncertainties. Here, with a larger sample of GCs and with PASTEL, we confirm this positive offset of the metallicity scale in the metal-poor regime. The internal scatter (MAD) for clusters ranges from 0.007 to 0.05 dex for individual OCs and from 0.02 to 0.1 dex for GCs.

3.2. GALAH

The current public version is GALAH DR3 ([Buder et al. 2021](#)), which contains 588 571 stars. The data for GALAH consist of spectra at a resolution of $R \sim 28\,000$ in four wavelength ranges that cover $4713\text{--}4903$, $5648\text{--}5873$, $6478\text{--}6737$, and $7585\text{--}7887 \text{ \AA}$. The determination of APs and abundances is performed with the code Spectroscopy Made Easy (SME; [Valenti & Piskunov 2012](#))

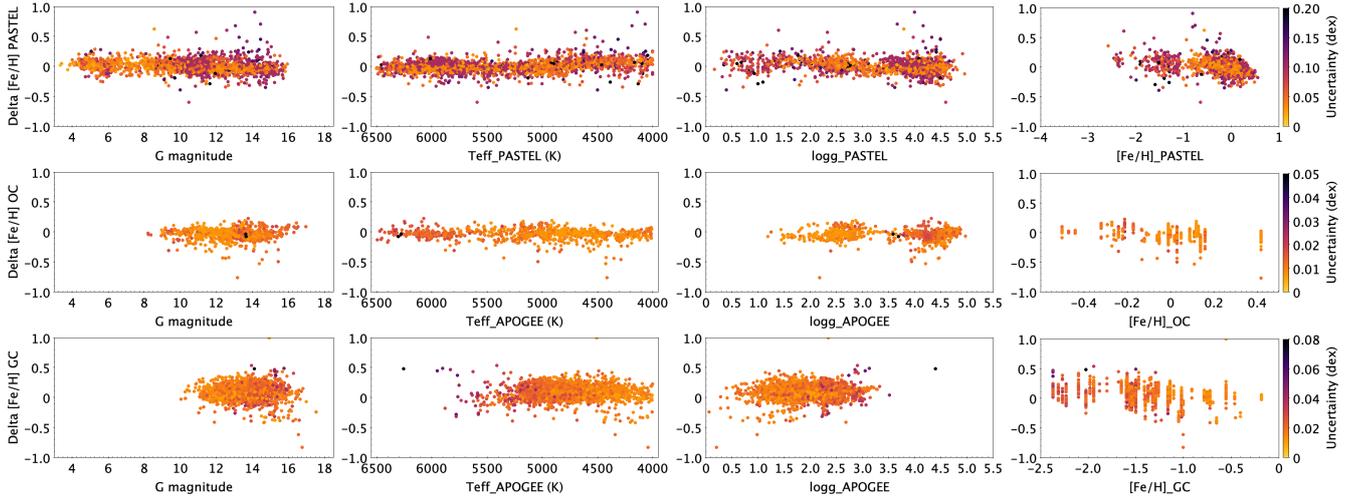


Fig. 3. Difference (Delta) between APOGEE determinations of [Fe/H] and those from the reference catalogues versus magnitude, T_{eff} , $\log g$, and [Fe/H]. The *upper panel* is for the literature mean value from PASTEL, the middle panel for OC members with cluster mean metallicities from Netopil et al. (2016), and the *bottom panel* for GC members with cluster mean metallicities from Harris (2010). The colour code reflects the metallicity uncertainty, quadratically combined for APOGEE versus PASTEL, from APOGEE only for the clusters (note the different scales).

through spectrum synthesis with MARCS models. Non-LTE corrections for Fe lines from Amarsi et al. (2016) are applied. The gravity $\log g$ is constrained from *Gaia* astrometry and 2MASS photometry. The precision of [Fe/H] is estimated with both the internal SME covariance errors and the standard deviation of repeat observations of the same star, resulting in an exponential function of S/N. The validation of iron abundances is made via comparison to the *Gaia* benchmark stars (Heiter et al. 2015; Jofré et al. 2015) and to clusters. We considered only the FGK stars in version 2 of the catalogue (GALAH_DR3_main_allstar_v2.fits) for which `flag_fe_h = 0` and `flag_sp = 0`, as recommended. The median [Fe/H] uncertainty for this selection of 407 276 stars is 0.08 dex.

There are 232 stars in common between PASTEL and this GALAH sample, 682 FGK cluster members in 25 OCs and 363 FGK cluster members in 11 GCs. The residuals are shown in Fig. 4 as a function of magnitude and of (T_{eff} , $\log g$, [Fe/H]).

In the comparison to PASTEL we note a larger dispersion for giants ($\log g < 3.8$) than for dwarfs ($\log g \geq 3.8$), with $\text{MAD} = 0.12$ dex and $\text{MAD} = 0.05$ dex, respectively. With the same MAD values, the dispersion for metal-poor stars ([Fe/H] < -0.5 dex) is larger than that for metal-rich stars ([Fe/H] > 0 dex), although there are some outliers. In addition, there is an offset of +0.14 dex for the metal-poor stars. In the comparison to OCs, there are trends that give a negative offset at the two extrema of the T_{eff} range and a pronounced oscillation along the $\log g$ axis. For dwarfs there seems to be a positive slope with $\log g$. Some outliers are visible, mainly on the negative side of the residuals. The largest uncertainties correspond to the faintest stars, but they do not exhibit large deviations. For the GCs, there are only giants, and there is no obvious trend.

3.3. *Gaia*-ESO Survey

The current public version of *Gaia*-ESO Survey (GES) is the DR3, available in the ESO science archive since 2016. It includes 25 533 stars observed with the FLAMES instrument on the Very Large Telescope, either at medium resolution ($R \sim 20\,000$) with GIRAFFE setups or at high resolution ($R \sim 47\,000$) with UVES. The APs have been determined by different groups using a

variety of parametrisation methodologies with common inputs (synthetic spectra, line list, solar abundance, LTE regime) from which recommended parameters and their errors were provided. The parameter homogenisation was performed with a weighting scheme that takes the performances of each group, after outlier rejection, into account. The parameters derived by the different groups were put on the same scale based on the calibrators analysed by all. Our selection of FGK stars with valid (T_{eff} , $\log g$, [Fe/H]) in the T_{eff} range 4000–6500 K gives 11 638 stars with a median [Fe/H] uncertainty of 0.09 dex.

We found 162 stars in common between PASTEL and GES, 239 FGK cluster members in 18 OCs and 513 FGK cluster members in 11 GCs. The residuals are shown in Fig. 5 as a function of magnitude and of (T_{eff} , $\log g$, [Fe/H]).

The comparison to PASTEL gives a good agreement over the full range of parameters, with a low dispersion that does not reflect the large quoted uncertainties. The comparison to OCs is very clean, with a remarkably low scatter among the members. Several OCs have a dispersion of $\text{MAD} = 0.02$ dex or lower: NGC 2243 (17 stars), NGC 2264 (7 stars), NGC 2682 (16 stars), NGC 6633 (6 stars), and NGC 6802 (9 stars). The most observed OC is NGC 2516, which has 55 stars ($\text{MAD} = 0.06$ dex). For the GCs with at least five members, the scatter ranges from $\text{MAD} = 0.03$ dex to $\text{MAD} = 0.06$ dex. NGC 104 has 111 observed members ($\text{MAD} = 0.03$ dex).

3.4. RAVE

The latest and final version of RAVE, DR6, contains 451 783 unique stars (Steinmetz et al. 2020a,b). RAVE spectra have an average resolution of $R = 7500$ and cover the infrared Ca triplet region at 8410–8795 Å. The MADERA pipeline derives APs by fitting the spectra to a grid of synthetic spectra built from MARCS models. The best model is found from a combination of a decision-tree algorithm and a projection method (see Kordopatis et al. 2011, for further details). Atmospheric parameters are then calibrated on a sample of reference stars. Next, the [Fe/H] is determined by the GAUGUIN pipeline, which fits individual lines on a pre-computed grid of synthetic spectra, interpolated to the MADERA APs. Errors on [Fe/H] are

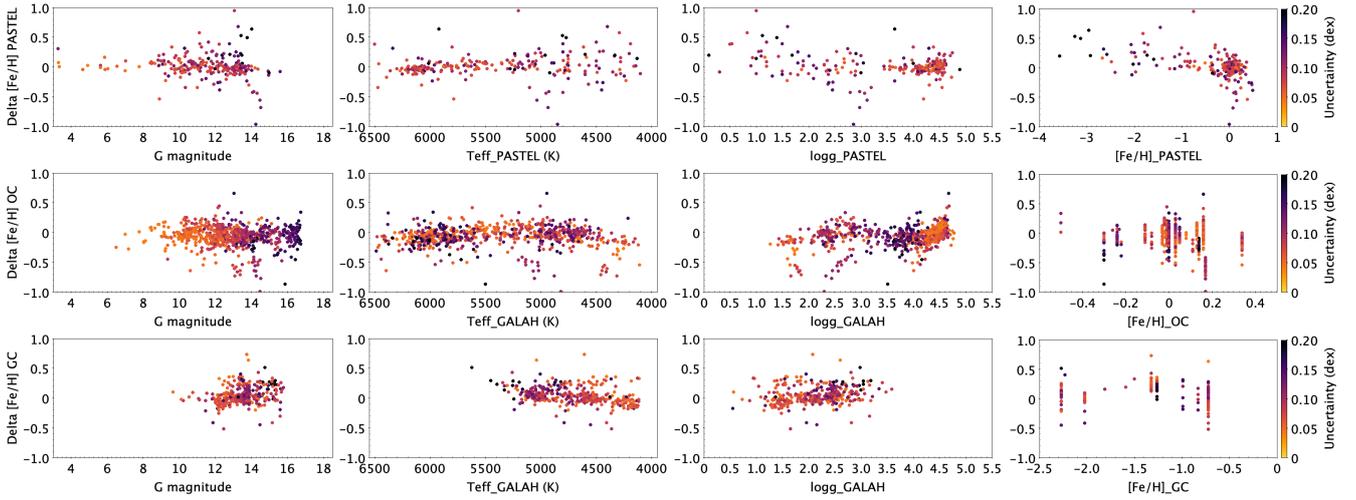


Fig. 4. Same as Fig. 3 but for GALAH.

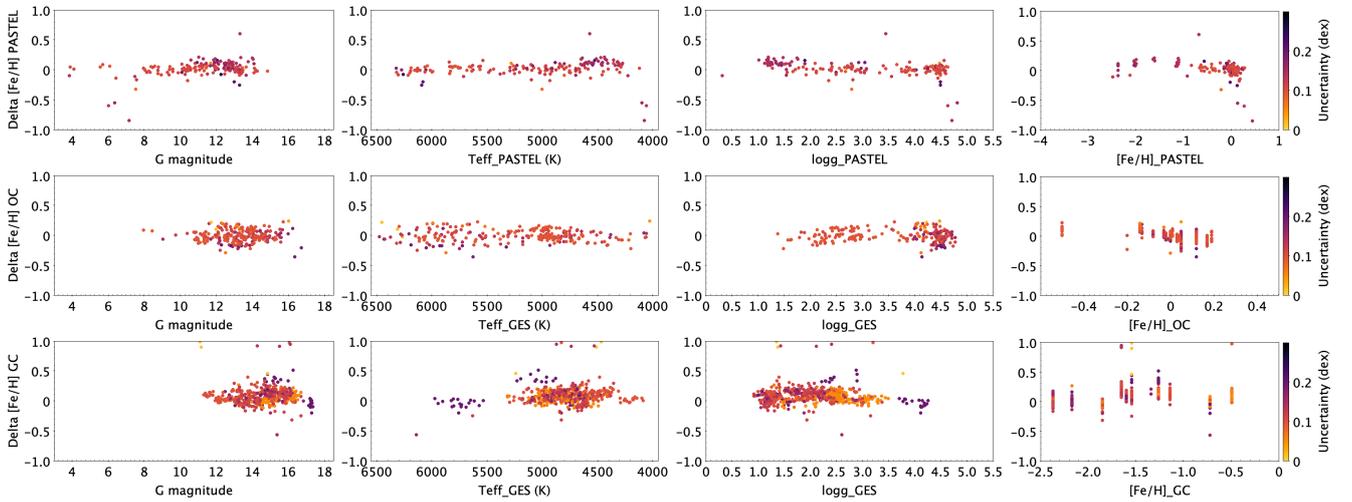


Fig. 5. Same as Fig. 3 but for GES.

computed by combining in a quadratic sum the propagation of errors of the stellar APs and the internal error of GAUGUIN due to noise. This internal error is the standard deviation, at a given S/N and for a given spectral line, of 500 measurements of [Fe/H] from noisy synthetic spectra of Sun-like and Arcturus-like stars. We note that the individual uncertainties quoted in the survey represent the precision of the metallicities. The accuracy has been tested using synthetic spectra with noise with the conclusion that GAUGUIN-derived values intrinsically do not suffer from large systematics. We selected the most reliable results for FGK stars with `flag_algo_conv_madera=0` as well as with `fe_h_chisq_gauguin < 2.5` and `fe_h_error_gauguin < 0.3`. The median [Fe/H] uncertainty for the corresponding 196 448 stars is 0.15 dex.

There are 427 stars in common between PASTEL and this RAVE sample, 119 FGK cluster members in 16 OCs and 35 FGK cluster members in 8 GCs. The residuals are shown in Fig. 6 as a function of magnitude and of (T_{eff} , $\log g$, [Fe/H]).

In the comparison to PASTEL, we note a positive offset of about +0.1 dex or larger for the coolest stars ($T_{\text{eff}} < 5000$ K), the most evolved giants ($\log g < 2$), and the most metal-poor stars ([Fe/H] < -1.5). The OC residuals show that the metallicity of giants is systematically underestimated compared to that

of dwarfs. The dispersion among OC members with at least five members ranges from 0.04 to 0.07 dex for the six OCs with at least five members. There are only three GCs with at least five members that show dispersions from 0.045 to 0.08 dex.

3.5. RAVE-CNN

Guiglion et al. (2020) provide another version of RAVE APs based on convolutional neural networks (CNNs) trained with a set of 3904 stars with high quality APs from APOGEE DR16. RAVE data are complemented by 2MASS and ALL_WISE photometry and by *Gaia* DR2 photometry and parallaxes. The parameters are averaged over 80 CNN runs, and the errors correspond to the dispersion of the runs. Repeat observations show that these internal errors are realistic. We considered only FGK stars with `fe_h_flag_cnn = 0` and `fe_h_error_cnn <= 0.3`. The median [Fe/H] uncertainty for the corresponding 381 681 stars is 0.045 dex.

There are 666 stars in common between PASTEL and this RAVE-CNN sample, 216 FGK cluster members in 25 OCs and 85 FGK cluster members in 10 GCs. The residuals are shown in Fig. 7 as a function of magnitude and of (T_{eff} , $\log g$, [Fe/H]).

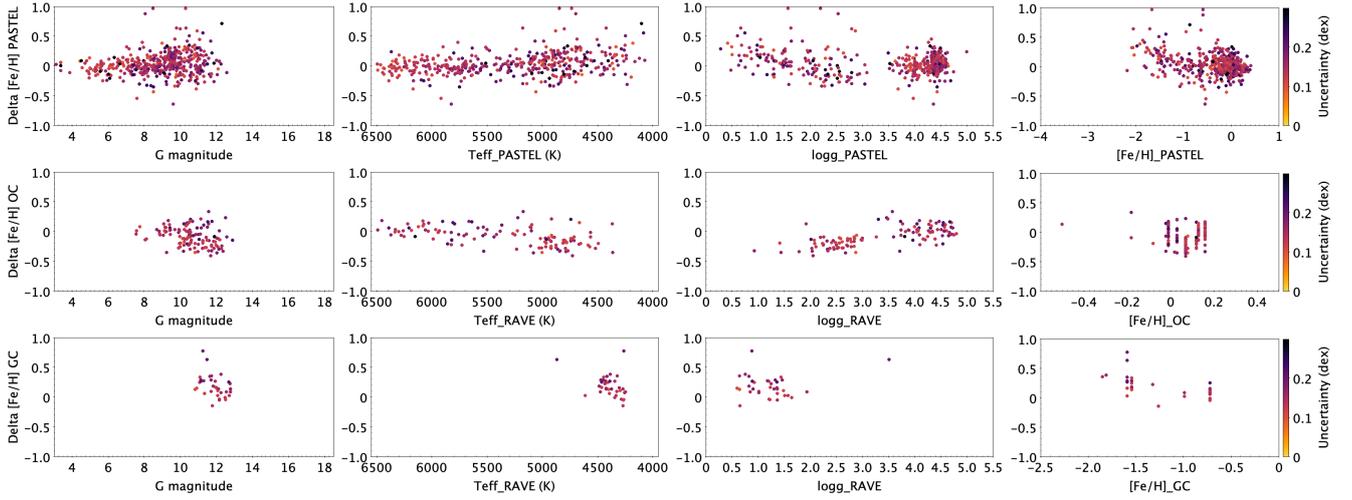


Fig. 6. Same as Fig. 3 but for RAVE.

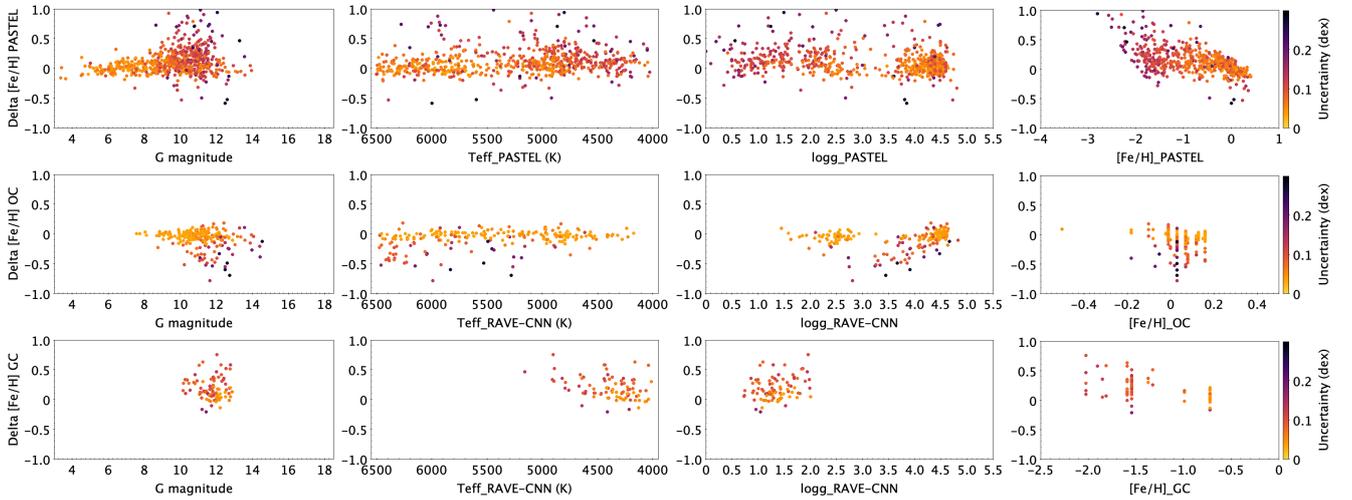


Fig. 7. Same as Fig. 3 but for RAVE-CNN.

In the comparison to PASTEL we note a change in the residual distribution at $G_{\text{mag}} = 9$. For fainter stars, the dispersion is higher, with a positive offset. The most deviating stars also have the largest uncertainties. The offset reaches $+0.18$ dex for the metal-poor stars ($[\text{Fe}/\text{H}] < -0.5$ dex), while there is a decreasing trend in the residuals versus $[\text{Fe}/\text{H}]$ for the metal-rich stars, with a slope of -0.32 dex^{-1} . The dispersion among OC members ranges from 0.02 to 0.06 dex, with only one cluster showing a large dispersion, of $\text{MAD} = 0.12$ dex (NGC 2423, 12 stars). The dwarfs and giants do not behave similarly: the residuals of giants are smaller, and there seems to be a trend with $\log g$ for the dwarfs. Outliers have large uncertainties in general, and they are more frequent among dwarfs. A significant positive offset for the most metal-poor GCs is visible, in agreement with the trend also observed for the metal-poor stars in PASTEL. The dispersion for four GCs with at least five members can be as low as 0.05 dex (NGC 104, 26 stars) and as high as 0.18 dex for the most metal-poor cluster (NGC 6397, 5 stars). Guiglion et al. (2020) state that the performance would improve a lot with a larger training sample. This pilot study was limited by the overlap with APOGEE DR16. Indeed, we note that the training sample has very few stars with $[\text{Fe}/\text{H}] < -0.5$ dex and almost none with $[\text{Fe}/\text{H}] < -1.0$ dex, which results in a poor parametrisation of halo stars.

3.6. LAMOST

The currently public version available in VizieR is LAMOST DR5 (Luo et al. 2015, 2019). LAMOST spectra have a resolution of $R \sim 1800$ and cover the optical range 3690–9100 Å. The LAMOST stellar pipeline (LASP; Wu et al. 2011; Xiang et al. 2015) derives $(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}])$ by matching the flux-calibrated spectra to empirical templates from the MILES library (Sánchez-Blázquez et al. 2006). Two algorithms are used successively, first a weighted mean of parameters of best-matching templates, then a χ^2 minimisation to further improve the parameters. Errors of the final parameters are estimated by combining the random and systematic errors, and are functions of S/N and of $(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}])$. Random errors are estimated from repeat observations, while the systematic errors are derived by applying the pipeline to the MILES templates. We considered only FGK stars with $e_{\text{Fe}_H} \leq 0.3$. The median $[\text{Fe}/\text{H}]$ uncertainty for the corresponding 4 539 240 stars is 0.08 dex.

The cross-match of LAMOST with the reference catalogues was performed with the equatorial J2000 coordinates and a radius of $3''$. There are 1767 stars in common between PASTEL and LAMOST, 1175 FGK cluster members in 51 OCs and 87 FGK cluster members in 13 GCs. The residuals are shown in Fig. 8 as a function of magnitude and of $(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}])$.

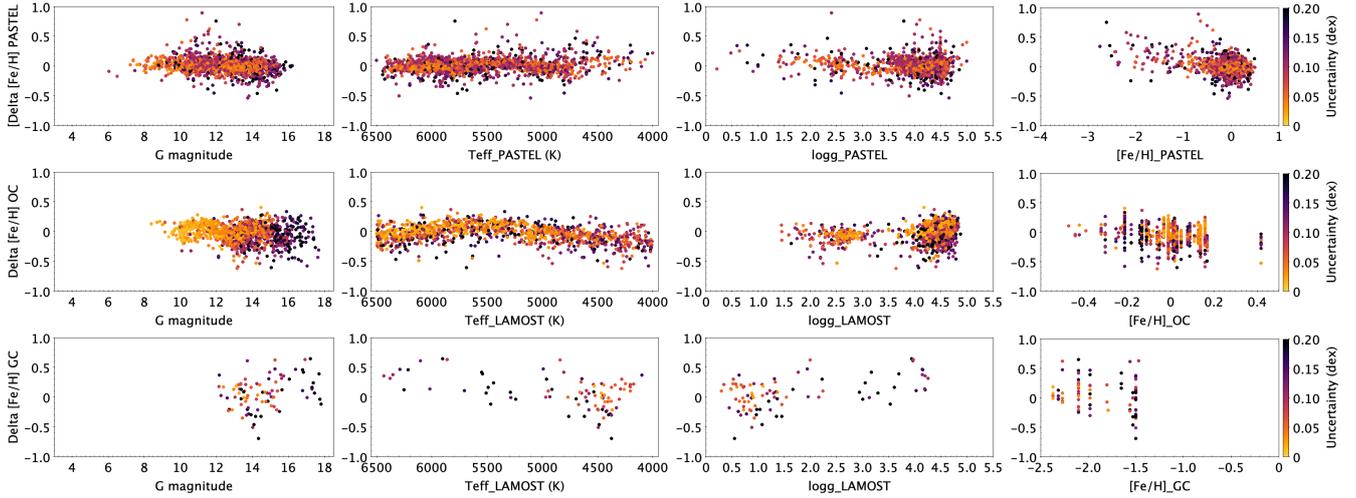


Fig. 8. Same as Fig. 3 but for LAMOST.

There is a good overlap between LAMOST and PASTEL, mostly dwarfs with $[\text{Fe}/\text{H}] > -0.5$ dex. The few metal-poor stars show mainly positive residuals, as for the other surveys, indicating that LAMOST tends to overestimate $[\text{Fe}/\text{H}]$ compared to high-resolution analyses. The overall dispersion is low ($\text{MAD} = 0.055$ dex), and the most deviating stars also have large uncertainties in general. There is also a good intersection with OCs, the majority having more than 15 members. The dispersion ranges from 0.01 to 0.14 dex with a median value of 0.05 dex. There is a marked oscillation of the residuals with T_{eff} in OCs, which gives negative residuals at the extrema, similar to what we observed with GALAH. Here the negative offset is more pronounced on the cool side. There are only four GCs with at least five members, and only one of them has a dispersion lower than 0.1 dex (NGC 5053, $\text{MAD} = 0.06$ dex, seven stars). A positive offset is seen for the faintest stars, the hottest stars, and dwarfs, these three types also having the largest uncertainties.

3.7. LAMOST-Payne

Another version of LAMOST DR5 has been released (Xiang et al. 2019) in which APs and abundances have been determined with the method *Data-Driven Payne*, a hybrid approach that combines constraints from theoretical spectral models (ATLAS12) and training on 4557 stars from GALAH DR2 and on 15 000 stars from APOGEE DR14. The internal precision of $[\text{Fe}/\text{H}]$ is deduced from the standard deviation of repeat observations at different S/N. We adopted the recommended parameters, selecting the FGK stars with $\text{FEH_FLAG} = 1$ (reliable). We applied in addition $\text{FEH_ERR} \leq 0.3$. The median $[\text{Fe}/\text{H}]$ uncertainty for the corresponding 5 967 849 stars is 0.06 dex.

As for LAMOST, the cross-match with the reference catalogues was performed with the equatorial J2000 coordinates and a radius of $3''$. We note that the intersection with the reference catalogues is larger than for LAMOST. There are 2243 stars in common between PASTEL and LAMOST-Payne, 1470 FGK cluster members in 53 OCs and 123 FGK cluster members in 14 GCs. The residuals are shown in Fig. 9 as a function of magnitude and of $(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}])$.

LAMOST-Payne behaves globally like LAMOST though with more outliers and larger systematics, which are inherited from the training sets. The largest deviations correspond quite well to the largest uncertainties. There are more metal-poor stars in

common with PASTEL but with a trend that gives a large positive offset of $+0.31$ dex at $[\text{Fe}/\text{H}] < -2$ dex, a large dispersion that also corresponds to large uncertainties. The OC residuals are significantly negative, with a dispersion larger than with LAMOST. Interestingly, the dwarfs and giants behave differently, with distributions of the residuals looking similar to those from RAVE-CNN. Residuals of dwarfs are more dispersed and show a positive slope with $\log g$. The dispersion among GC members is slightly lower than for LAMOST, from 0.02 to 0.10 dex. There is a pronounced positive offset reaching $+0.5$ dex at the lowest metallicities. The source of these strong biases is not clear since LAMOST-Payne uses two different training sets as well as constraints from synthetic spectra, but we presume that the performances could be improved with a larger homogeneous training set that covers the full parameter space better.

3.8. SEGUE

The SEGUE survey (Yanny et al. 2009) provides stellar spectra at $R \sim 2000$ over the wavelength range 3800–9200 Å for about 500 000 stars. The SEGUE Stellar Parameter Pipeline (SSPP; Lee et al. 2008) uses multiple techniques to estimate $(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}])$, up to 12 methods for $[\text{Fe}/\text{H}]$, with a procedure that gives at the end a recommended value and its error, which we adopted here. The SSPP was improved by Smolinski et al. (2011), who validated the results with OCs and GCs. They obtained a typical internal uncertainty of 0.05 dex on $[\text{Fe}/\text{H}]$ and a dispersion of 0.11 dex when the results are compared to high-resolution values.

The cross-match of SEGUE with other catalogues based on equatorial J2000 coordinates was performed with a radius of $5''$. We found 181 stars in common between PASTEL and SEGUE, 266 FGK cluster members in 10 OCs and 509 FGK cluster members in 9 GCs. The residuals are shown in Fig. 10 as a function of magnitude and of $(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}])$.

In the comparison to PASTEL, the faintest stars ($G \geq 13.5$), which seem to also be the most metal-poor ($[\text{Fe}/\text{H}] < -2$ dex), show a positive offset of $+0.18$ dex. For the other common stars there is no offset and the dispersion is $\text{MAD} = 0.085$ dex. For the six OCs with at least five members, the dispersion is remarkably low despite the resolution of the survey, from 0.02 dex (NGC 2682, 67 stars) to 0.06 dex (NGC 7789, 49 stars). Dwarfs, which seem to correspond to faint stars, tend to have lower metallicities than giants. The typical dispersion for GCs

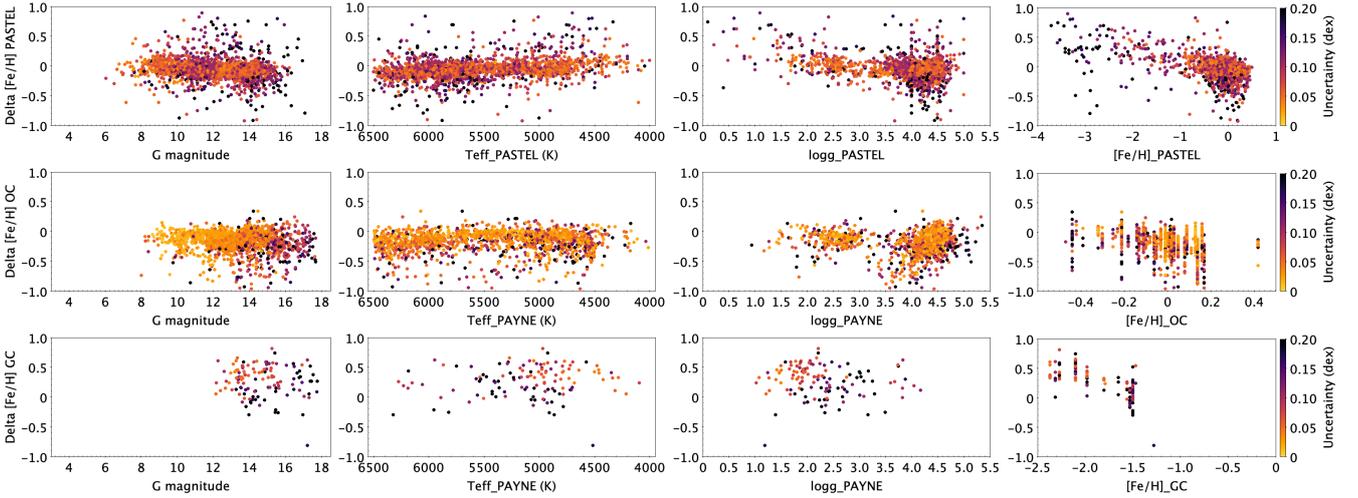


Fig. 9. Same as Fig. 3 but for LAMOST-Payne.

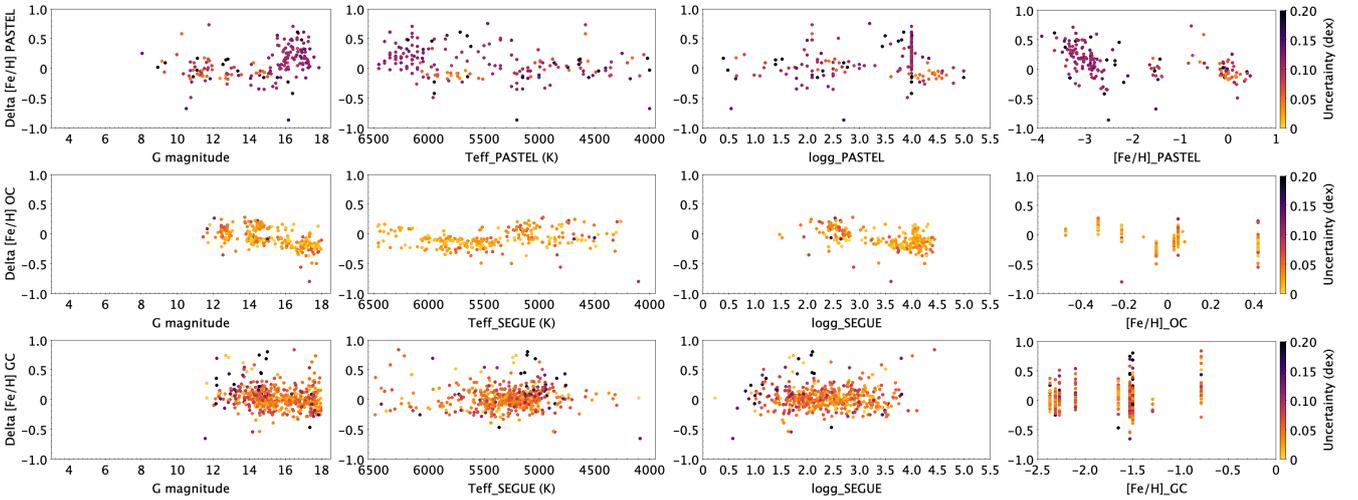


Fig. 10. Same as Fig. 3 but for SEGUE. The multiple stars at $\log g = 4.0$ in the PASTEL plot come from one bibliographical reference, a high-resolution spectroscopic follow-up of extremely metal-poor stars from SEGUE (Aoki et al. 2013), which assumed the surface gravity for all turn-off stars to be that value.

is around 0.08 dex. NGC 6205 is the most populated cluster for SEGUE with 186 members and has $\text{MAD} = 0.08$ dex. NGC 5024 shows a poorer performance, with $\text{MAD} = 0.16$ dex (17 members). There is no trend of the residuals with G magnitude, T_{eff} , or $\log g$.

4. Surveys versus PASTEL

All the comparisons to PASTEL are summarised in Table 1, which gives the median offsets of the [Fe/H] residuals and their MAD. When relevant, trends have been evaluated through a simple linear fit. In general, there is a good agreement for the metal-rich regime ($[\text{Fe}/\text{H}] \geq -0.5$ dex), with negligible offsets and typical dispersions of 0.04–0.06 dex for the higher-resolution surveys (APOGEE, GALAH, and GES) and up to 0.10 dex for the other surveys. In this metallicity range, there is, however, a significant correlation between the residuals and [Fe/H] for APOGEE, which is reproduced in RAVE-CNN and LAMOST-Payne, the surveys that use APOGEE as a training set for their parametrisation methods. We note that these data-driven methods give more outliers and larger offsets, dispersions, and trends than the more classical methods used for RAVE and

LAMOST. It is worth noting that LAMOST and SEGUE show a remarkable precision, better than 0.1 dex, despite their low resolution.

For all the surveys, the metal-poor stars ($[\text{Fe}/\text{H}] < -0.5$ dex) have their metallicity overestimated compared to the high-resolution, high-S/N determinations listed in PASTEL (also visible with GCs). Offsets range from +0.06 dex for APOGEE to +0.18 dex for RAVE-CNN. This result needs to be investigated further owing to the important implication that it has for galactic studies. If this bias is confirmed in massive spectroscopy, it implies in particular that metallicity gradients in the Milky Way cannot be reliably estimated from surveys. It is unlikely that PASTEL is the source of this bias, owing to the number of different papers (more than 1200) that have been considered when averaging APs and the fact that the cross-match between PASTEL and the surveys involves different samples of stars. This bias more likely results from the analysis pipelines of the surveys, which are poorly constrained in the metal-poor range due to a lack of reference stars fitting the specific observing requirements. This highlights the need for surveys to observe metal-poor stars for calibration purposes.

Table 1. Summary of the [Fe/H] differences between the surveys and PASTEL.

Survey	Constraint	N	MED	MAD	σ_{Survey}	σ_{PASTEL}	Slope
APOGEE		2155	-0.01	0.05	0.01	0.05	
APOGEE	[Fe/H] < -0.5	308	0.06	0.07	0.01	0.07	
APOGEE	[Fe/H] \geq -0.5	1847	-0.02	0.05	0.01	0.05	-0.18
GALAH		232	0.00	0.07	0.055	0.05	
GALAH	[Fe/H] < -0.5	48	0.14	0.12	0.07	0.085	
GALAH	[Fe/H] \geq -0.5	184	-0.02	0.06	0.05	0.05	
GALAH	$\log g < 3.8$	91	0.00	0.125	0.06	0.07	
GALAH	$\log g \geq 3.8$	141	0.00	0.05	0.05	0.05	
GES		162	0.02	0.05	0.10	0.05	
GES	[Fe/H] < -0.5	44	0.09	0.04	0.10	0.10	
GES	[Fe/H] \geq -0.5	118	0.00	0.04	0.10	0.05	
RAVE		427	0.01	0.10	0.135	0.05	
RAVE	$T_{\text{eff}} < 5000$ K	177	0.09	0.13	0.14	0.07	
RAVE-CNN		666	0.08	0.11	0.055	0.05	
RAVE-CNN	$G_{\text{mag}} > 9$	150	0.13	0.19	0.07	0.06	
RAVE-CNN	[Fe/H] < -0.5	336	0.18	0.14	0.07	0.07	
RAVE-CNN	[Fe/H] \geq -0.5	330	0.01	0.07	0.04	0.05	-0.32
LAMOST		1767	0.00	0.055	0.03	0.05	
LAMOST	[Fe/H] < -0.5	136	0.08	0.095	0.025	0.07	
LAMOST-Payne		2243	-0.07	0.07	0.03	0.05	-0.17
LAMOST-Payne	[Fe/H] < -0.5	246	0.15	0.16	0.05	0.08	
SEGUE		181	0.07	0.16	0.04	0.10	
SEGUE	[Fe/H] < -2	111	0.17	0.16	0.05	0.10	
SEGUE	[Fe/H] \geq -2	70	-0.045	0.085	0.03	0.07	

Notes. Differences are measured with the median value, MED, and the dispersion, MAD, for the N common stars with various constraints. The typical (median) [Fe/H] uncertainty listed in each catalogue for the corresponding selected stars is also given. When relevant, the slope of the linear fit on the distribution of the residuals versus [Fe/H] is provided.

Several surveys (APOGEE, GALAH, and SEGUE) seem to provide underestimated metallicities for stars with [Fe/H] > 0 dex. RAVE-CNN and LAMOST-Payne also show this behaviour, inherited from APOGEE. This trend is difficult to quantify due to the small extension of the metallicity range on the positive side.

Table 1 also provides the typical [Fe/H] uncertainties (median values) as given in each survey and in PASTEL for the corresponding selection of stars. This allowed us to verify that their combination through a quadratic sum is consistent with the dispersion of the residuals. In most cases, the MAD of the residuals is lower than the total uncertainty of the catalogues, which indicates that their precision is possibly better than expected. The most remarkable case is GES, which exhibits a small dispersion of MAD = 0.04 dex for [Fe/H] \geq -0.5 dex, although the quoted uncertainties have a median value of 0.10 dex. On the contrary, RAVE-CNN quotes small errors for the faint stars, which are not consistent with the large dispersion of the residuals. A similar disagreement is seen for LAMOST-Payne in the metal-poor regime.

5. Surveys versus open clusters

The [Fe/H] residuals for OC members identified by each survey are shown in the middle panels of Figs. 3–5 and discussed in the previous sections. The [Fe/H] reference value for each cluster was adopted from Netopil et al. (2016) and Casamiquela et al. (2021). The offsets that are seen for some clusters and some surveys may be due to a systematic error in the survey or in the literature, or both. They are therefore difficult to interpret. More relevant is the agreement of the median [Fe/H] value obtained

by different surveys for a given cluster. Figure 11 represents the median [Fe/H] computed for the 76 OCs that have at least five members in one of the nine surveys, including PASTEL. For the majority of clusters observed by several surveys, there is in general a good agreement, although there are a few clusters where the agreement is poor, with variations reaching more than 0.2 dex (e.g., Trumpler 20). The most metal-poor OCs (e.g., Trumpler 5, NGC 2243, and Berkeley 32) have their low metallicity confirmed by several surveys. Similarly, the high metallicity of NGC 6791 and Berkeley 81 is found in good agreement by several surveys. It is very clear in Fig. 11 that the median [Fe/H] from LAMOST-Payne systematically lies below the others, indicating that this survey has a more metal-poor zero point at the solar metallicity compared to all the other surveys.

Remarkably, NGC 2682 (M 67) is part of the nine surveys, and the Pleiades (Melotte 22) and the Hyades (Melotte 25) appear in seven surveys. The results for these well-observed clusters are presented in Table 2. For M 67, the median [Fe/H] values vary from -0.10 to 0.0 dex with low dispersions, from 0.02 to 0.06 dex. Interestingly, the survey with the lowest spectroscopic resolution, SEGUE, exhibits the lowest scatter. We cannot exclude that the good performances in all the surveys are related to the fact that M 67 is a common reference, used by most surveys for calibrations and validations. For the Pleiades, all the surveys except LAMOST-Payne agree well, with MED ranging from -0.02 to 0.04 dex and MAD ranging from 0.03 to 0.09 dex. For the Hyades, there are larger differences for the median [Fe/H] due to lower values from LAMOST and LAMOST-Payne. The Hyades is known to be a metal-rich OC (e.g., Casamiquela et al. 2020), which is confirmed by PASTEL, APOGEE, GALAH, RAVE, and RAVE-CNN, with

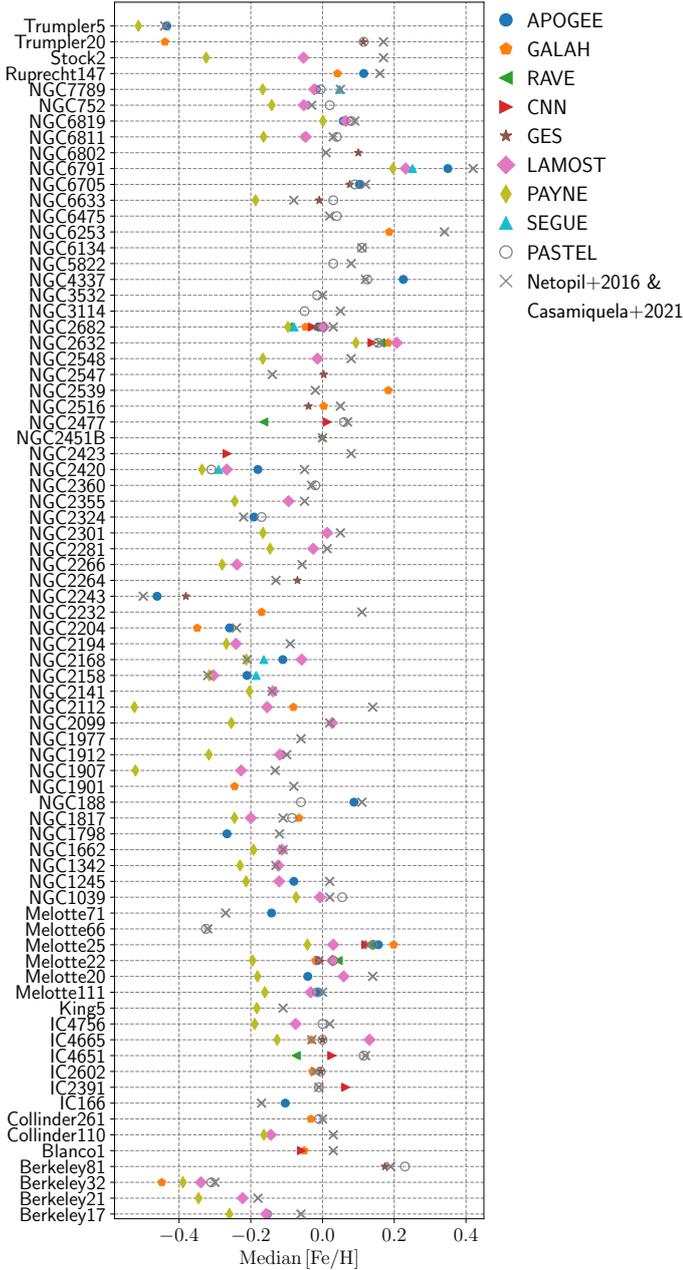


Fig. 11. Median [Fe/H] obtained by the different surveys and catalogues for OCs with at least five FGK members.

MED between +0.12 and +0.20 dex and MAD between 0.02 to 0.06 dex.

It is also informative to compare, for each catalogue, the dispersion among members of a given cluster and the typical uncertainties quoted in that catalogue, which we expect to be consistent. This is verified in most cases, with a few notable exceptions. We note that GES and RAVE provide [Fe/H] precisions that look pessimistic owing to the low dispersion obtained within OCs. On the contrary, LAMOST-Payne gives small [Fe/H] errors for the Pleiades and Hyades members, which do not correspond to the significant dispersion observed for these two clusters.

Figure 12 shows the histogram of the MAD [Fe/H] for each survey. APOGEE is clearly the survey that has the highest consistency among FGK members of OCs. This cannot directly be attributed to the ASCAP pipeline, which, according to Jönsson et al. (2020), does not use OCs for calibrations, con-

trary to previous APOGEE releases. APOGEE observed 25 OCs in common with our reference sample and with at least five members, and they have all a dispersion (MAD) lower than 0.05. The best performance is reached for Trumpler 5 (ten members, MED = -0.43 dex, MAD = 0.007 dex), then NGC 2324 (six members, MED = -0.19 dex, MAD = 0.008 dex) and NGC 1798 (nine members, MED = -0.27 dex, MAD = 0.008 dex). It is worth noting that APOGEE has observed 128 OCs in total (reported in DR16), which Donor et al. (2020) used to measure the radial metallicity gradient in the galactic disc.

6. Surveys versus globular clusters

The bottom panels of Figs. 3–5 show the [Fe/H] residuals for GC members identified by each survey. The [Fe/H] reference values are adopted from Harris (2010) and do not represent the most up-to-date metallicities for those clusters. Nevertheless, for the 42 GCs with at least five members, the median [Fe/H] from the surveys shows an excellent agreement with the Harris metallicities, as shown in Fig. 13, although they are slightly more metal-rich in general. When plotted all together, the [Fe/H] residuals are centred on ~ 0.1 dex (Fig. 14). The tendency for the surveys to overestimate the [Fe/H] of the metal-poor stars has already been noticed in the comparison to PASTEL and is confirmed here in a different way with GCs. As for OCs, LAMOST-Payne stands apart from the other surveys. In the previous section, LAMOST-Payne was systematically underestimating the median metallicity of clusters, while here in the metal-poor regime it overestimates it.

The mildly metal-poor cluster NGC 104 has been observed by six surveys, and NGC 5272 and NGC 7078 appear in five of them. The corresponding median [Fe/H] values are detailed in Table 3. For NGC 104 they vary from -0.63 dex with RAVE to -0.83 dex with PASTEL, a difference still compatible at the level of uncertainties of the two catalogues, 0.15 and 0.06 dex, respectively. For this cluster there is a high level of consistency within each survey, with dispersions ranging from 0.03 to 0.06 dex. The agreement between surveys is also very good for NGC 5272, with the median [Fe/H] ranging from -1.50 to -1.40 dex, although the dispersions are larger, in particular for LAMOST (MAD = 0.21 dex). For NGC 7078 the median [Fe/H] from LAMOST-Payne lies well above the others despite the smallest internal dispersion (MAD = 0.02 dex for 7 stars). Like for OCs, we note that the dispersion of the residuals is lower than the quoted uncertainties for RAVE and GES, while they are in better agreement for the other surveys.

Figure 12 shows the histogram of the MAD [Fe/H] for each survey. APOGEE has observed the largest number of clusters: 35 GCs with at least five members. They have dispersions (MAD) ranging from 0.02 to 0.10 dex (median 0.04 dex). The lowest scatter (MAD = 0.02 dex) is reached for NGC 6441 (six members, MED = -0.47 dex), NGC 6553 (seven members, MED = -0.19 dex), and NGC 6723 (seven members, MED = -1.00 dex).

7. Surveys versus APOGEE

In the previous sections we have shown that APOGEE performs very well, providing in particular a low [Fe/H] dispersion among members of OCs and GCs. It is thus relevant to use it as a reference catalogue to test the other surveys against it in order to strengthen the statistics with more common stars. Here we therefore compare the [Fe/H] determinations from various surveys to those from APOGEE without considering PASTEL. The

Table 2. Metallicity of M 67, the Pleiades, and the Hyades in the different surveys.

Survey	M67			Pleiades			Hyades					
	N	MED	MAD	σ	N	MED	MAD	σ	N	MED	MAD	σ
PASTEL	91	0.0	0.04	0.05	14	0.03	0.055	0.05	55	0.14	0.03	0.04
APOGEE	178	-0.01	0.03	0.01	180	-0.015	0.03	0.01	7	0.155	0.04	0.01
GALAH	130	-0.05	0.05	0.06	42	-0.02	0.07	0.06	23	0.20	0.06	0.04
GES	16	-0.015	0.02	0.10								
RAVE	8	-0.09	0.04	0.16	15	0.04	0.07	0.18	16	0.13	0.06	0.13
CNN	34	-0.03	0.06	0.08	25	-0.01	0.05	0.04	20	0.12	0.02	0.04
LAMOST	93	0.0	0.04	0.03	148	0.03	0.09	0.03	26	0.03	0.07	0.03
Payne	99	-0.10	0.05	0.03	184	-0.195	0.14	0.03	38	-0.04	0.18	0.02
SEGUE	67	-0.08	0.02	0.02								

Notes. The number of members (N), median [Fe/H] (MED), and median absolute deviation are provided. The σ column gives the typical (median) uncertainty of the N stars as provided in the corresponding catalogue.

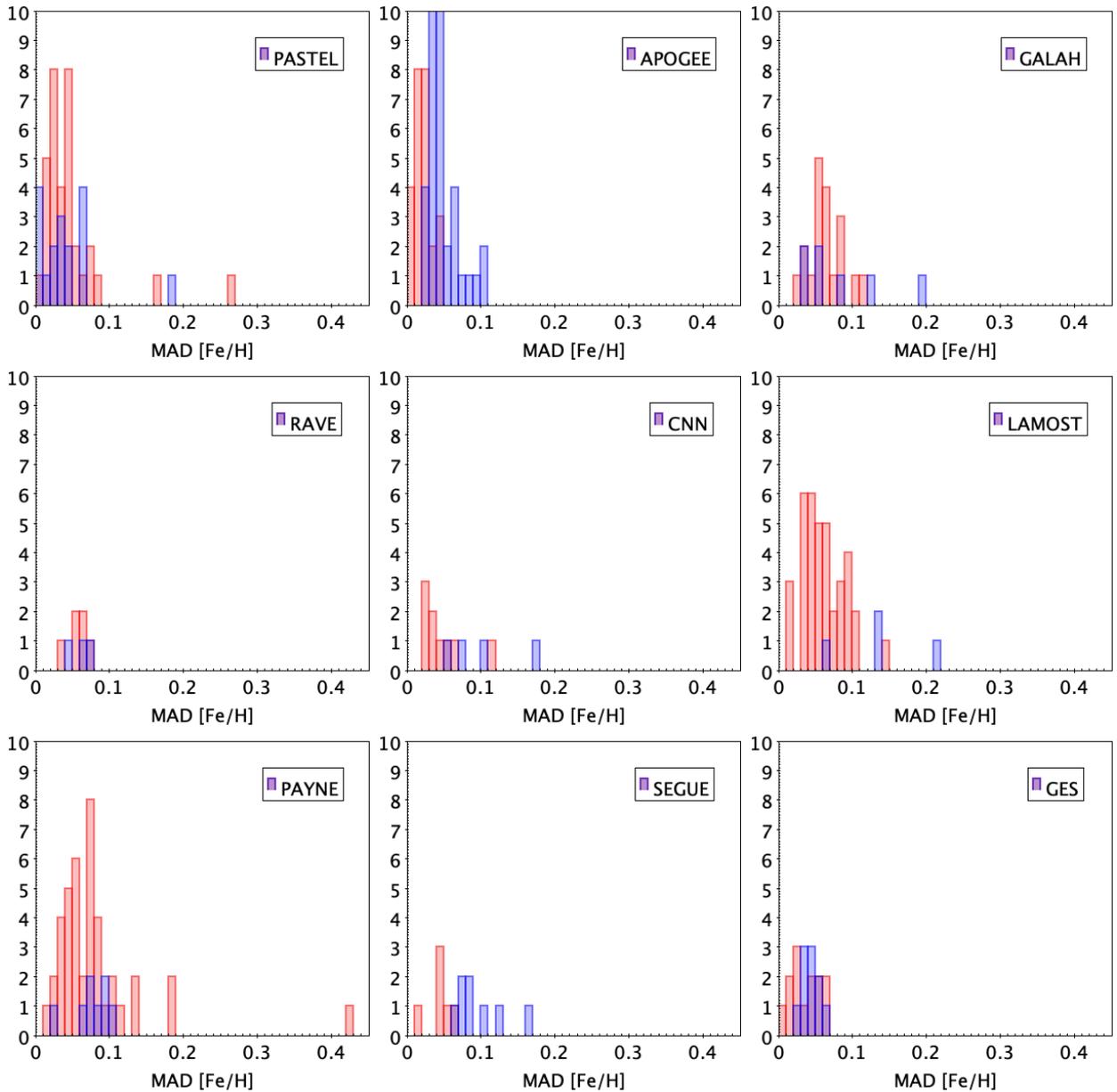


Fig. 12. Histogram of the dispersion (MAD) of [Fe/H] obtained by the different surveys for OCs (in red) and GCs (in blue) with at least five FGK members.

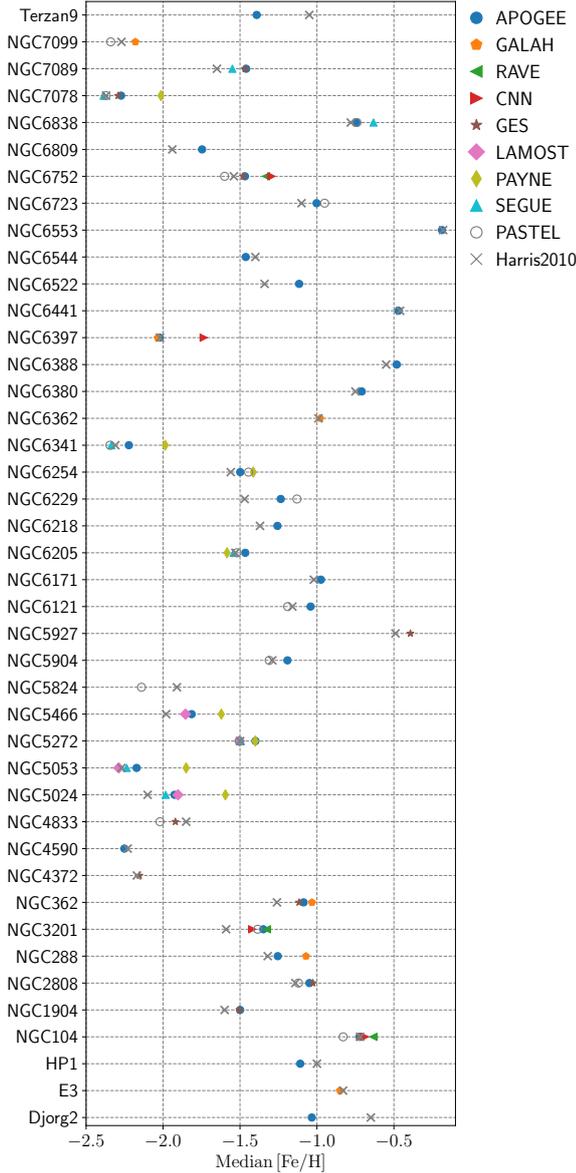


Fig. 13. Median $[\text{Fe}/\text{H}]$ obtained by the different surveys and catalogues for GCs with at least five FGK members.

residuals are shown in Fig. 15, and the MED and MAD are presented in Table 4, together with the uncertainties of the catalogues for the common stars. The intersection between the surveys and APOGEE varies from a few hundred to nearly 96 000 for LAMOST-Payne. In general, there is a good agreement of the $[\text{Fe}/\text{H}]$ determinations with flat distributions of the residuals, well centred on zero. The lowest scatter (0.04 dex) is reached by LAMOST and GES, which implies that these surveys have precisions at this level. This confirms the findings of the previous sections. For LAMOST this precision is a remarkable owing to its low resolution. For GES it implies that the quoted uncertainties are too pessimistic. The largest offset and dispersion are for RAVE (MED = -0.08 dex, MAD = 0.10 dex). No trend is visible with the G magnitude. There is a slight trend in T_{eff} for GALAH. The largest effects depend on $\log g$, but they occur essentially at the edges of the $\log g$ range, where the density of stars is lower. Compared to APOGEE, RAVE clearly underestimates the metallicity of giants. We note that there is no star more metal-poor than $[\text{Fe}/\text{H}] = -2.5$ dex in common between

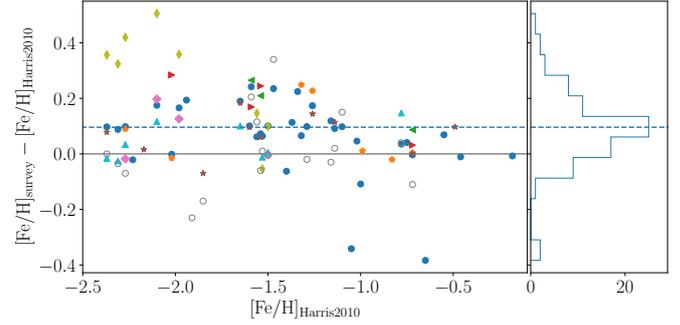


Fig. 14. $[\text{Fe}/\text{H}]$ residuals for 42 GCs with at least five FGK members. The dot colours are the same as in Fig. 13. The histogram, represented without the PASTEL values, shows a median offset of 0.096 dex for the surveys with respect to Harris (2010).

APOGEE and the other surveys. A striking feature is the positive offset for GALAH for metal-poor stars, while LAMOST, SEGUE, and GES have similar zero points along the metallicity axis.

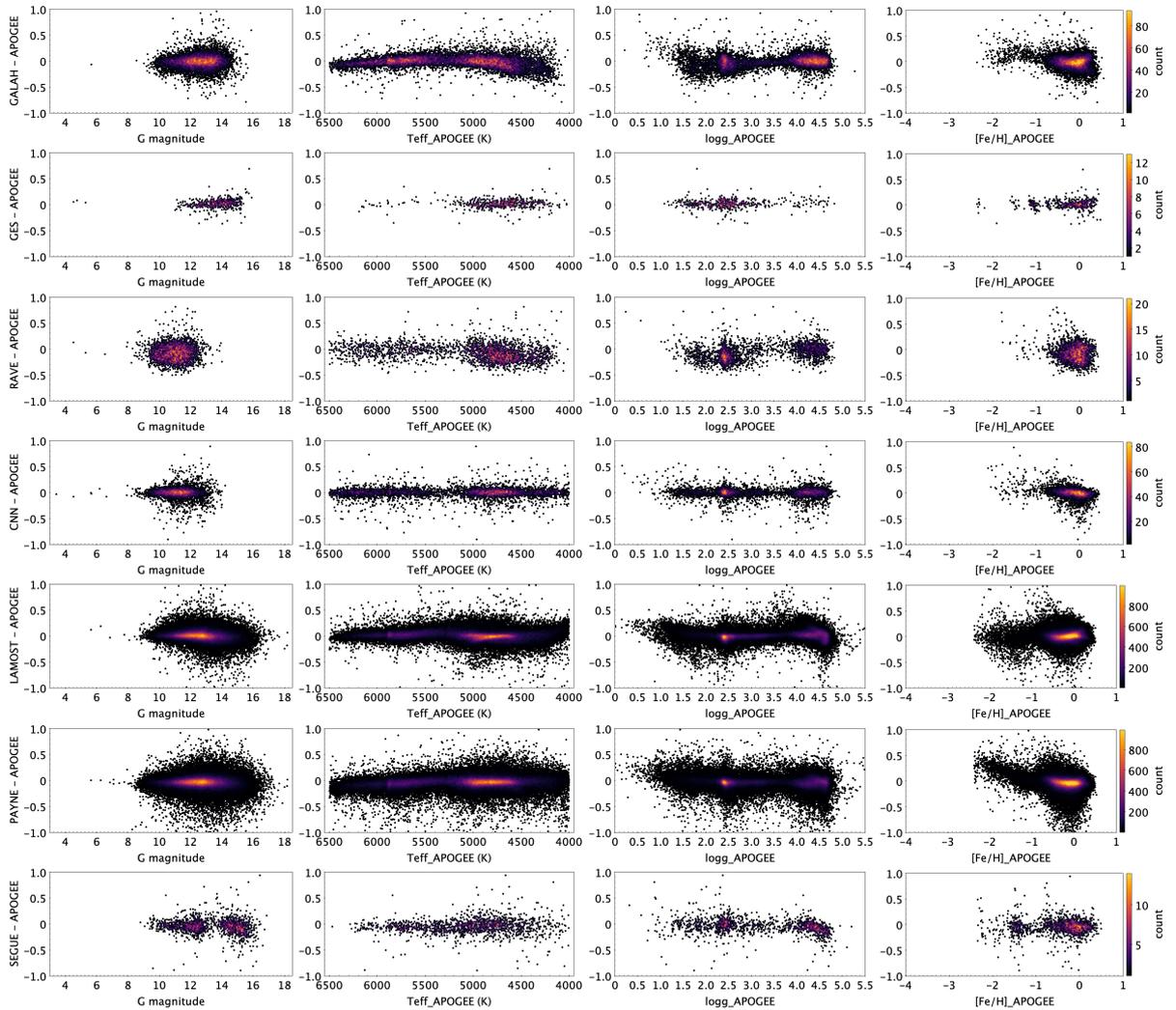
The APOGEE comparisons to RAVE-CNN and LAMOST-Payne are worth discussing since these two surveys use APOGEE for the training of their pipeline. The bulk of metallicities from RAVE-CNN are in good agreement with those from APOGEE, with no offset and a low dispersion of 0.05 (this does not apply, however, to the small fraction of metal-poor stars that exhibit a positive offset). The systematics and precision look better in comparison to RAVE. This good performance is well explained by the fact that nearly all the stars in common between RAVE-CNN and APOGEE were used for the training of the CNN pipeline. What we see here is essentially the trained $[\text{Fe}/\text{H}]$ versus the input $[\text{Fe}/\text{H}]$ from RAVE-CNN, already analysed by Guiglion et al. (2020). For LAMOST-Payne the distribution of the $[\text{Fe}/\text{H}]$ residuals is very similar to that obtained with LAMOST, albeit with a very pronounced linear trend for the metal-poor stars. This suggests that the training set for LAMOST-Payne (15 000 stars from APOGEE DR14) was possibly too sparse in the metal-poor regime, resulting in these systematics.

The next step of this work will be to evaluate the agreement of abundance ratios from the different surveys. APOGEE DR16 provides abundances of 26 species (C, C I, N, O, Na, Mg, Al, Si, P, S, K, Ca, Ti, Ti II, V, Cr, Mn, Fe, Co, Ni, Cu, Ge, Rb, Ce, Nd, and Yb), GALAH DR3 the abundances of 30 species (Li, C, O, Na, Mg, Al, Si, K, Ca, Sc, Ti, V, Cr, Mn, Co, Ni, Cu, Zn, Rb, Sr, Y, Zr, Mo, Ru, Ba, La, Ce, Nd, Sm, and Eu), and the current public GES version the abundances of 24 species (C, Li, N, O, Na, Mg, Al, S, Ca, Sc, Ti, Ti 2, V, Cr, Co, Ni, Zn, Y, Zr, Ba, La, Ce, Nd, and Eu), offering interesting perspectives for comparisons. This is particularly relevant with the advent of *Gaia* DR3, which will include chemical abundances (N, Mg, Si, S, Ca, Ti, Cr, Fe, FeII, Ni, Zr, Ce, and Nd) for several million FGK stars (Recio-Blanco et al. 2016).

Another task for the future would be to combine the different surveys into a single homogenised catalogue. The agreement between metallicities from the different surveys is reasonable enough in the metal-rich regime ($[\text{Fe}/\text{H}] \geq -0.5$) to attempt such a combination, which would increase the sample size and sky coverage for galactic archeology studies. This implies, however, the development of a proper procedure that takes the differences between surveys into account to calibrate the metallicities onto a common scale. Nandakumar et al. (2020) used the

Table 3. Similar to Table 2 but for the GCs NGC 104, NGC 5272, and NGC 7078.

Survey	NGC 104				NGC 5272				NGC 7078			
	N	MED	MAD	σ	N	MED	MAD	σ	N	MED	MAD	σ
PASTEL	18	-0.83	0.03	0.06	8	-1.505	0.04	0.07	38	-2.37	0.01	0.10
APOGEE	186	-0.72	0.03	0.01	152	-1.40	0.07	0.02	34	-2.27	0.05	0.025
GALAH	253	-0.72	0.05	0.07								
GES	111	-0.72	0.03	0.09					39	-2.29	0.05	0.13
RAVE	14	-0.63	0.045	0.15								
CNN	26	-0.69	0.05	0.05								
LAMOST					23	-1.50	0.21	0.13				
Payne					34	-1.40	0.09	0.145	7	-2.01	0.02	0.05
SEGUE					126	-1.495	0.10	0.05	39	-2.39	0.06	0.04

**Fig. 15.** [Fe/H] difference between the different surveys and APOGEE (surveys minus APOGEE) versus magnitude, T_{eff} , $\log g$, and [Fe/H] from APOGEE. From top to the bottom: GALAH, GES, RAVE, RAVE-CNN, LAMOST, LAMOST-Payne, and SEGUE. The colour is scaled on the counts.

data-driven approach *The Cannon* (Ness et al. 2015; Casey et al. 2016) to combine metallicities and alpha abundances from APOGEE DR16 and GALAH DR3 in order to explore the radial and vertical gradients and abundance trends in the Galactic disc. The stellar parameters of one survey were put on the scale of the other and vice versa, resulting in two catalogues that show some

differences. These discrepancies reflect the difficulty of dealing with an imperfect training set and with complex selection functions. Another ongoing project is the Survey of Surveys, which has already managed to homogenise radial velocities from different surveys (Tsantaki et al. 2022) and plans to make a similar analysis with APs.

Table 4. Summary of the [Fe/H] residuals of the surveys versus APOGEE.

Survey	N	MED	MAD	σ_{Survey}	σ_{APOGEE}
GALAH	9331	-0.01	0.06	0.08	0.01
GES	491	0.01	0.04	0.08	0.01
RAVE	2269	-0.08	0.10	0.15	0.008
CNN	4498	0.0	0.04	0.04	0.009
LAMOST	85 726	0.0	0.04	0.04	0.01
Payne	95 924	-0.04	0.05	0.04	0.01
SEGUE	1427	-0.05	0.07	0.03	0.01

8. Conclusion

We have assessed the [Fe/H] determinations of FGK stars in eight spectroscopic surveys by comparing them to independent sources built from the PASTEL catalogue and clusters. We have tested the latest public versions of APOGEE, GALAH, RAVE, LAMOST, SEGUE, and GES, as well as the data-driven CNN and Payne versions of RAVE and LAMOST.

PASTEL being a bibliographical catalogue, we first adopted a weighted mean of the APs for each star. We then selected FGK stars with $4000 \leq T_{\text{eff}} \leq 6500$ K, which have a typical [Fe/H] uncertainty of 0.06 dex. We obtain a scatter (MAD) of 0.04 dex when OC members are considered and 0.08 dex for GC members, with the reference metallicity for clusters adopted from the literature. PASTEL includes a number of metal-poor stars, which allowed us to probe the metal-poor regime in surveys.

To test the agreement between two sources of metallicity, we used the median value of the residuals to measure an eventual offset between them and the median absolute deviation of the residuals, which reflects the precision of both sources. We looked for trends in the distribution of residuals versus G magnitude and (T_{eff} , $\log g$, [Fe/H]). We also checked whether the scatter of the residuals was consistent with the combined uncertainties quoted in the considered sources.

Our main conclusions are as follows.

- In general, all the surveys perform well in the metal-rich regime ([Fe/H] ≥ -0.5 dex), with negligible offsets and dispersions lower than 0.10 dex whatever the resolution. This is verified with both PASTEL and the OCs.
- All the surveys overestimate [Fe/H] in the metal-poor regime ([Fe/H] < -0.5 dex), with offsets ranging from +0.06 to +0.18 dex on average. This is verified with both PASTEL and the GCs.
- The metallicities based on data-driven methodologies show offsets, dispersions, and trends that are significantly larger than those obtained with classical methods on the same spectra. The biases of the training set are amplified. In addition, the quoted uncertainties look too optimistic.
- In most cases, the uncertainties of the surveys are consistent with the scatter observed in clusters. A notable exception is GES, which has overly pessimistic uncertainties.
- APOGEE has a typical precision better than 0.05 dex over the full metallicity range but systematically overestimates low metallicities and underestimates high metallicities with a linear trend.
- LAMOST performs as well as surveys of higher resolution.

Our investigation has highlighted biases at the extrema of the metallicity range, where atmospheric models and automated pipelines are less constrained. The differences between surveys seriously hamper any attempt to simply combine them,

for instance in order to improve statistics. The combination of surveys requires an elaborate procedure of homogenisation of the iron abundances. We would like to encourage the builders of spectroscopic surveys to include common reference stars in their observing plans for calibration and validation purposes. This would enlarge the intersection between surveys and facilitate their homogenisation into a common scale. Such stars should have APs measured from high-resolution, high-S/N spectroscopy or belong to well-studied clusters. The PASTEL catalogue is a useful resource for searching for well-studied stars over the whole metallicity range, in particular at its extrema. Having calibration targets in common between surveys is very useful for tracking systematic differences. For this purpose, the OC M 67 and the GC NGC 104 are excellent targets. It is useful to observe the Hyades, Berkeley 81, and NGC 6791 to constrain the high metallicity regime, while the most metal-poor OCs (Berkeley 32, NGC 2243, Trumpler 5) can constrain metallicities around [Fe/H] = -0.4 dex. Many GCs are available to anchor the lowest metallicities. The recent revision of memberships in clusters based on *Gaia* data provides reliable lists of targets. We want to mention the sample of *Gaia* FGK benchmark stars (Heiter et al. 2015; Jofré et al. 2015, 2018), which was built for the calibration of the *Gaia* APs that will be delivered in DR3 and to serve as a common reference between *Gaia* and the ground-based surveys. A new extended version is in preparation.

Acknowledgements. We thank the anonymous referee for the useful comments and suggestions that helped clarify the paper. C.S. and N.B. warmly thank Philippe Prugniel who made a significant contribution in the 2020 version of PASTEL. The preparation of this work has made extensive use of Topcat (Taylor 2011), of the Simbad and VizieR databases at CDS, Strasbourg, France, and of NASA's Astrophysics Data System Bibliographic Services.

References

- Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, *ApJ*, **636**, 804
- Amarsi, A. M., Lind, K., Asplund, M., Barklem, P. S., & Collet, R. 2016, *MNRAS*, **463**, 1518
- Aoki, W., Beers, T. C., Lee, Y. S., et al. 2013, *AJ*, **145**, 13
- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
- Asplund, M., Amarsi, A. M., & Grevesse, N. 2021, *A&A*, **653**, A141
- Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, *A&A*, **559**, A74
- Buder, S., Sharma, S., Kos, J., et al. 2021, *MNRAS*, **506**, 150
- Cantat-Gaudin, T., & Anders, F. 2020, *A&A*, **633**, A99
- Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *A&A*, **618**, A93
- Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, *A&A*, **640**, A1
- Carretta, E., & Gratton, R. G. 1997, *A&AS*, **121**, 95
- Carretta, E., Bragaglia, A., Gratton, R., D'Orazi, V., & Lucatello, S. 2009, *A&A*, **508**, 695
- Casamiquela, L., Tarricq, Y., Soubiran, C., et al. 2020, *A&A*, **635**, A8
- Casamiquela, L., Soubiran, C., Jofré, P., et al. 2021, *A&A*, **652**, A25
- Casey, A. R., Hogg, D. W., Ness, M., et al. 2016, ArXiv e-prints [arXiv:1603.03040]
- Cayrel de Strobel, G., Bentolila, C., Hauck, B., & Curchod, A. 1980, *A&AS*, **41**, 405
- Cayrel de Strobel, G., Bentolila, C., Hauck, B., & Lovy, D. 1981, *A&AS*, **45**, 97
- Cayrel de Strobel, G., Bentolila, C., Hauck, B., & Duquennoy, A. 1985, *A&AS*, **59**, 145
- Cayrel de Strobel, G., Hauck, B., Francois, P., et al. 1992, *A&AS*, **95**, 273
- Cayrel de Strobel, G., Soubiran, C., Friel, E. D., Ralite, N., & Francois, P. 1997, *A&AS*, **124**, 299
- Cayrel de Strobel, G., Soubiran, C., & Ralite, N. 2001, *A&A*, **373**, 159
- Dalton, G., Trager, S. C., Abrams, D. C., et al. 2012, *SPIE Conf. Ser.*, **8446**, 84460P
- de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, *The Messenger*, **175**, 3
- Donor, J., Frinchaboy, P. M., Cunha, K., et al. 2020, *AJ*, **159**, 199
- Gaia* Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Gaia* Collaboration (Brown, A. G. A., et al.) 2018a, *A&A*, **616**, A1
- Gaia* Collaboration (Babusiaux, C., et al.) 2018b, *A&A*, **616**, A10

- Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, 649, A1
- García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, 151, 144
- Gilmore, G., Randich, S., Asplund, M., et al. 2012, *The Messenger*, 147, 25
- Gratton, R. G., Carretta, E., & Bragaglia, A. 2012, *A&ARv*, 20, 50
- Guiglion, G., Matijević, G., Queiroz, A. B. A., et al. 2020, *A&A*, 644, A168
- Harris, W. E. 2010, ArXiv e-prints [arXiv:1012.3224]
- Heiter, U., Soubiran, C., Netopil, M., & Paunzen, E. 2014, *A&A*, 561, A93
- Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, *A&A*, 582, A49
- Jofré, P., Heiter, U., Soubiran, C., et al. 2015, *A&A*, 582, A81
- Jofré, P., Heiter, U., Tucci Maia, M., et al. 2018, *Res. Notes Am. Astron. Soc.*, 2, 152
- Jofré, P., Heiter, U., & Soubiran, C. 2019, *ARA&A*, 57, 571
- Jönsson, H., Allende Prieto, C., Holtzman, J. A., et al. 2018, *AJ*, 156, 126
- Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, *AJ*, 160, 120
- Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011, *A&A*, 535, A106
- Lai, D. K., Smith, G. H., Bolte, M., et al. 2011, *AJ*, 141, 62
- Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, *AJ*, 136, 2022
- Liu, F., Yong, D., Asplund, M., Ramírez, I., & Meléndez, J. 2016, *MNRAS*, 457, 3934
- Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, *Res. Astron. Astrophys.*, 15, 1095
- Luo, A. L., Zhao, Y. H., Zhao, G., et al. 2019, VizieR Online Data Catalog: V/164
- McConnachie, A., Babusiaux, C., Balogh, M., et al. 2016, ArXiv e-prints [arXiv:1606.00043]
- Nandakumar, G., Hayden, M. R., Sharma, S., et al. 2020, *MNRAS*, 513, 232
- Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, 808, 16
- Netopil, M., Paunzen, E., Heiter, U., & Soubiran, C. 2016, *A&A*, 585, A150
- Nidever, D. L., Hasselquist, S., Hayes, C. R., et al. 2020, *ApJ*, 895, 88
- Randich, S., Gilmore, G., & Gaia-ESO Consortium 2013, *The Messenger*, 154, 47
- Recio-Blanco, A., de Laverny, P., Allende Prieto, C., et al. 2016, *A&A*, 585, A93
- Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, *MNRAS*, 371, 703
- Smolinski, J. P., Lee, Y. S., Beers, T. C., et al. 2011, *AJ*, 141, 89
- Snedden, C., Kraft, R. P., Prosser, C. F., & Langer, G. E. 1992, *AJ*, 104, 2121
- Soubiran, C., Le Campion, J. F., Cayrel de Strobel, G., & Caillo, A. 2010, *A&A*, 515, A111
- Soubiran, C., Jasniewicz, G., Chemin, L., et al. 2013, *A&A*, 552, A64
- Soubiran, C., Le Campion, J.-F., Brouillet, N., & Chemin, L. 2016, *A&A*, 591, A118
- Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, *AJ*, 132, 1645
- Steinmetz, M., Guiglion, G., McMillan, P. J., et al. 2020a, *AJ*, 160, 83
- Steinmetz, M., Matijević, G., Enke, H., et al. 2020b, *AJ*, 160, 82
- Takada, M., Ellis, R. S., Chiba, M., et al. 2014, *PASJ*, 66, R1
- Taylor, M. 2011, *TOPCAT: Tool for OPERations on Catalogues And Tables*
- Taylor, W., Cirasuolo, M., Afonso, J., et al. 2018, in *Proc. SPIE*, SPIE Conf. Ser., 10702, 107021G
- Tsantaki, M., Pancino, E., Marrese, P., et al. 2022, *A&A*, 659, A95
- Valenti, J. A., & Piskunov, N. 2012, *SME: Spectroscopy Made Easy Astrophysics Source Code Library*, record ascl:1202.013
- Vasiliev, E., & Baumgardt, H. 2021, *MNRAS*, 505, 5978
- Wu, Y., Luo, A. L., Li, H.-N., et al. 2011, *Res. Astron. Astrophys.*, 11, 924
- Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, *MNRAS*, 448, 822
- Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, *ApJS*, 245, 34
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, 137, 4377
- Yong, D., Meléndez, J., Grundahl, F., et al. 2013, *MNRAS*, 434, 3542