



Universiteit  
Leiden  
The Netherlands

## Refining open cluster parameters with Gaia XP metallicities

Nizovkina, M.; Larsen, S.S.; Brown, A.G.A.; Helmi, A.

### Citation

Nizovkina, M., Larsen, S. S., Brown, A. G. A., & Helmi, A. (2025). Refining open cluster parameters with Gaia XP metallicities. *Astronomy And Astrophysics*, 703.  
doi:10.1051/0004-6361/202555437

Version: Accepted Manuscript

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4288500>

**Note:** To cite this publication please use the final published version (if applicable).

# Refining open cluster parameters with Gaia XP metallicities

M. Nizovkina<sup>1</sup>, S. S. Larsen<sup>1</sup>, A. G. A. Brown<sup>2</sup>, and A. Helmi<sup>3</sup>

<sup>1</sup> Department of Astrophysics/IMAPP, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands  
e-mail: m.nizovkina@astro.ru.nl

<sup>2</sup> Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

<sup>3</sup> Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands

Received 9 May 2025 / Accepted 14 September 2025

## ABSTRACT

**Context.** Open clusters (OCs) are crucial objects for studying stellar evolution and Galactic dynamics due to the shared origin and composition of the stars within each cluster. The precision of cluster parameter determination has significantly improved with the availability of homogeneous photometric Gaia data; however, challenges such as age-metallicity degeneracy and lack of spectroscopic observations remain.

**Aims.** We investigate whether metallicities derived from low-resolution Gaia XP spectra can be effectively used to break degeneracies and improve the accuracy of OC parameter determinations.

**Methods.** We analysed 20 OCs using isochrone fitting methods on Gaia DR3 photometry and metallicity estimates from several Gaia XP-based catalogues. We used the synthetic cluster method, implemented through ASteCA, to derive age, distance modulus, and extinction using the approximate Bayesian computation (ABC).

**Results.** We compared the parameter estimates of these open clusters to the values obtained in other works through isochrone fitting with spectroscopically constrained metallicities or through neural network techniques applied only to the photometry. We found the systematic difference between Gaia XP derived metallicities and those obtained from high-resolution spectroscopy to be 0.1–0.15 dex. We found a systematic age difference of  $< 0.03 \pm 0.13$  dex compared to isochrone fitting using high-resolution spectroscopy, and  $< 0.08 \pm 0.21$  dex compared to neural network-based methods, and a median individual error,  $\sigma_{\log(\text{age})}$ , of  $\sim 0.065$  dex.

**Conclusions.** Metallicities from Gaia XP spectra prove highly effective in constraining the parameters of open clusters that do not have a populated red giant branch (RGB), despite the low spectral resolution. When used with stringent quality cuts and incorporated as priors, they make it possible to determine ages comparable in precision to those based on high-resolution spectroscopy. Moreover, compared to studies of open clusters that use neural networks without metallicity constraints, our method yields smaller uncertainties. These results highlight the potential of Gaia data for accurate cluster parameter analysis and detailed Galactic studies, even in the absence of traditional spectroscopic observations.

**Key words.** Methods: statistical - Galaxy: open clusters and associations:general - Galaxy: stellar content

## 1. Introduction

Open clusters (OC) are groups of stars that are located in the Galactic disc and have a common origin, therefore the same age and compositions, which makes them invaluable laboratories for studying stellar evolution and the dynamics of the galaxy. They used to be identified from the ground-based data (Dias et al. 2002; Kharchenko et al. 2005), which was limited to brighter, closer and more populated clusters.

The open cluster census was tremendously improved by data from the Gaia satellite (Gaia Collaboration et al. 2016), as it provided homogeneous astrometric and photometric data of unprecedentedly high precision for the whole sky. This led to a significant increase in the amount of identified OCs based on their precise positions, proper motions and parallaxes using clustering methods, starting from  $\sim 3000$  OCs known pre-Gaia and more than 7000 known afterwards (e.g. Sim et al. 2019; Liu & Pang 2019; Monteiro et al. 2019). A large portion of the newly discovered clusters was caused by the development of machine learning techniques that were applied to OCs to perform ‘blind searches’ (Castro-Ginard et al. 2019). Most notably, Cantat-Gaudin et al. (2018) compiled a homogeneous catalogue of 1229 OCs by applying an unsupervised membership determination algorithm to Gaia Data Release 2 (Gaia Collaboration

et al. 2018). Their method included using a clustering method DBSCAN on astrometric data and applying a neural network to the colour-magnitude diagram (CMD) of clusters to validate the membership. Hunt & Reffert (2021) took this work further first by comparing various clustering methods and establishing that instead HDBSCAN shows the most sensitivity to OCs in Gaia data. In their follow-up work (Hunt & Reffert 2023, hereafter HR23) they applied HDBSCAN together with the neural network to Gaia DR3 data, thus compiling a comprehensive catalogue of 7167 OCs, which is the largest homogeneous OC catalogue up to date.

The astrophysical parameters of OCs have also been determined on a larger scale and with higher precision in the Gaia era. While the ages of small samples of OCs have been measured with asteroseismology (Palakkatharappil & Creevey 2023) and gyrochronology (Angus et al. 2019), the isochrone fitting method enables the determination of ages for numerous clusters in large amounts, while relying solely on Gaia photometric data. Cantat-Gaudin et al. (2020) (hereafter CG20) and HR23 determined the ages using artificial neural networks (ANN) that were trained on synthetic clusters generated from PARSEC isochrones, while other studies applied isochrone fitting via least

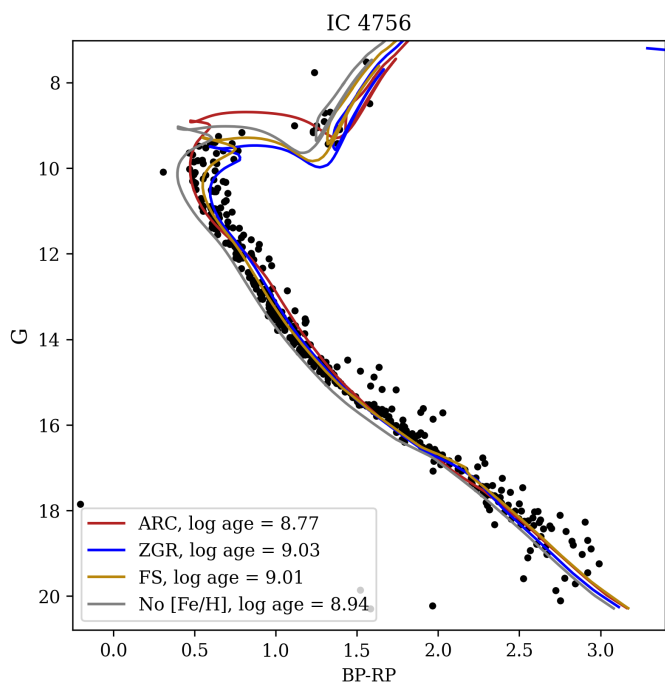


Fig. 1: CMD of the open cluster IC 4756 from Gaia DR3 photometry. The isochrones correspond to the derived parameters based on the metallicity prior from the respective Gaia XP-based stellar parameters estimates catalogue (see Section 2.2 for the description of the catalogues), namely ARC (red), ZGR (blue) and FS (yellow). The grey isochrone corresponds to the fit performed with metallicity as a free parameter.

squares fitting (Sim et al. 2019) or an MCMC-based method (Bossini et al. 2019) (hereafter B19).

Isochrone fitting is a multi-parametric process during which the age, metallicity, distance modulus and extinction of a cluster are determined simultaneously. An additional factor is the age-metallicity degeneracy, which introduces an additional uncertainty to the age determination. B19 calculated that a  $[\text{Fe}/\text{H}]$  uncertainty of 0.06 dex leads to a  $\log(\text{age})$  uncertainty of 0.01 dex, whereas a  $[\text{Fe}/\text{H}]$  uncertainty of 0.3 dex in turn results in the  $\log(\text{age})$  error of 0.05 dex. To limit the age uncertainty they fixed the  $[\text{Fe}/\text{H}]$  value obtained from high-resolution spectra of the clusters, if the spectra were available. However, not all known OCs have this luxury.

The expansion of Gaia data continued further, as Gaia DR3 (Gaia Collaboration et al. 2023) has introduced the XP spectra for 220 million objects. XP spectra are low-resolution spectra obtained in blue and red Gaia passbands (BP, RP). Despite the low resolution ( $R \sim 20\text{--}70$ ) of these spectra, a growing number of recent studies were able to provide estimates for stellar parameters using machine learning algorithms. Zhang et al. (2023) used a forward model to provide estimates for effective temperature ( $T_{\text{eff}}$ ), surface gravity ( $\log g$ ), and metallicity  $[\text{Fe}/\text{H}]$  for the whole XP dataset, whereas Andrae et al. (2023b) estimated  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{M}/\text{H}]$  for 220 million objects using the XGboost machine learning algorithm. Fallows & Sanders (2024) were able to determine not only these stellar parameters, but also  $[\text{C}/\text{Fe}]$ ,  $[\text{N}/\text{Fe}]$  and  $[\alpha/\text{Fe}]$ . Lucey et al. (2023) also applied XGBoost to the XP spectra to search for carbon-enhanced stars and provided probabilities for each object being carbon-enhanced. Moreover, Kane et al. (2025) identified  $\sim 900$  candidates of escapees from

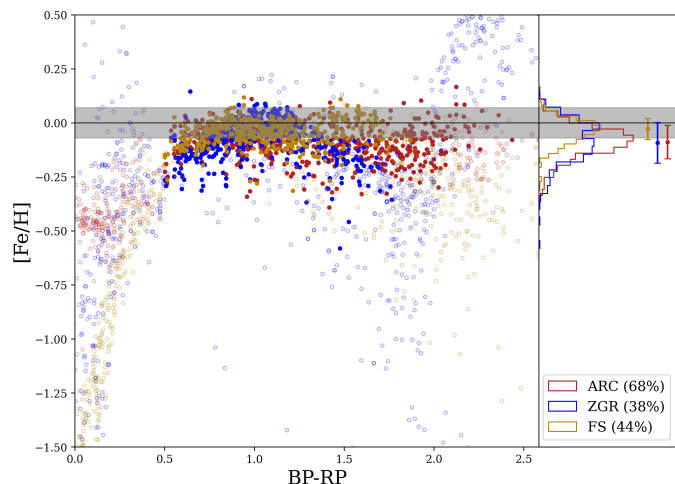


Fig. 2: Metallicities determined from the catalogues ARC, ZGR and FS of the open cluster NGC 3532 as a function of Gaia  $G_{\text{BP}} - G_{\text{RP}}$  colour. Empty circles show the estimates that were discarded after the respective cuts, whereas filled circles show the values that satisfied the cuts' criteria. The black line and the shaded area show the  $[\text{Fe}/\text{H}]$  and  $\sigma[\text{Fe}/\text{H}]$  determined with high-resolution spectroscopy by Netopil et al. (2016) and used by B19, whereas the points with error bars show the weighted mean  $[\text{Fe}/\text{H}]$  and its dispersion of the respective catalogue. The percentage indicates the fraction of the stars that were accepted and used for  $[\text{Fe}/\text{H}]$  determination after the respective cuts.

globular clusters by determining  $[\text{Al}/\text{Fe}]$  and  $[\text{N}/\text{O}]$  from Gaia XP spectra using a trained neural network. Other examples of XP spectra exploitation with machine learning methods include studies by Guiglion et al. (2024); Li et al. (2024); Yao et al. (2024); Khalatyan et al. (2024); Ye et al. (2025); Hattori (2025), which show the growing interest in extracting chemical information from Gaia data.

In this paper we aim to apply metallicities derived from Gaia XP spectra to the parameter derivation of Milky Way open clusters and test the possibilities of studying OCs with Gaia data alone. Having an estimate of metallicity will help lift the age-metallicity degeneracy, therefore providing more accurate age estimates. We use a selection of clusters previously studied with high-resolution spectroscopy and revisit them with the context of new Gaia data to compare how these age estimates change.

In Section 2 we describe the selection of clusters and the determination of metallicities, and in Section 3 we describe the process of the clusters parameters determination. Section 4 presents the obtained results of dynamical properties for the selected clusters. We discuss the reliability of the input data and the outcome results in Section 5, and we present the overview of our work in Section 6.

## 2. Data selection

### 2.1. Cluster membership

We started our sample selection from a subset of 37 clusters from B19 with high-resolution spectroscopy (HRS) measurements of metallicity. Their sample was defined to include clusters that have extinction less than  $A_V < 2.5$  mag and are older than 10 Myr ( $\log(\text{age}) > 7$  dex) based on the values from Dias et al. (2002) and Kharchenko et al. (2013). The latter avoids the potential uncertainty in the age determination that comes

with fitting isochrones to young clusters without a clearly defined main-sequence turnoff (MSTO). The [Fe/H] measurements were adopted from Netopil et al. (2016) that includes measurements for 88 OCs derived from high-quality, high-resolution ( $R \geq 25000$ ), and high signal-to-noise ratio ( $S/N \geq 50$ ) spectra. These spectra were obtained from various literature sources and the results were homogenised. This selection allows us to compare the effect of using low-resolution spectroscopy measurements for age determination.

For isochrone fitting to the cluster CMDs, we used Gaia DR3 photometry due to its all-sky coverage and exceptional precision at the mmag level for the brighter stars. It is available in three bands: broad  $G$  band (330–1050 nm),  $G_{BP}$  band (330–680 nm), and  $G_{RP}$  band (630–1050 nm).

To select cluster members, we used the cluster membership probabilities (MP) derived by Hunt & Reffert (2023). Their catalogue includes stars up to  $G = 20$  mag, in contrast to Cantat-Gaudin et al. (2018) which is limited to  $G = 18$  mag. As a result, the Hunt & Reffert (2023) catalogue contains almost two times as many cluster member stars. Additionally, this catalogue was compiled using not only a clustering algorithm (HDBSCAN) but also a Bayesian convolutional neural network trained on cluster CMDs to verify cluster memberships. In our analysis, we only included cluster members with  $MP > 0.5$ . An example CMD is shown in Figure 1.

After initial testing, we noticed that in clusters with a heavily populated bottom of the main sequence, the fit was being skewed, thus diminishing the goodness of fit to the stars in the area of the MSTO and the red clump. To accommodate for that, we imposed a magnitude cut on several clusters:  $G < 15$  mag for NGC 2632,  $G < 16$  mag for NGC 6475, and  $G < 17$  mag for NGC 2516 and NGC 3532.

## 2.2. Metallicity

To calculate the metallicities of the open clusters, we used measurements based on Gaia XP spectra from several recently published catalogues. Using Gaia XP spectra instead of the higher resolution Gaia RVS spectra allowed us to increase the number of measurements per cluster, since the number of stars with metallicities derived from RVS spectra is 5.6 million in contrast to 220 million of XP spectra. Due to the very low resolution of the latter, it becomes problematic to determine the stellar parameters using standard methods such as spectral fitting or equivalent width measurements. Therefore, studies utilise machine learning algorithms that are trained on spectroscopic libraries of high-resolution surveys in order to extract the stellar parameters information from the spectra.

We used the results from the following catalogues:

*ZGR* (Zhang, Green, & Rix 2023) contains approximately 220 million stars. The empirical forward model was trained on six million LAMOST spectra ( $R \sim 1800$ ) and augmented with 2MASS and WISE photometry. Deliverables of the catalogue include temperature ( $T_{\text{eff}}$ ), surface gravity ( $\log g$ ), metallicity ([Fe/H]), parallax ( $\varpi$ ), and extinction ( $E$ ). They report a typical error of [Fe/H] measurement to be 0.15 dex. We applied the recommended quality cut that denotes that all three parameters ( $T_{\text{eff}}$ ,  $\log g$ , and [Fe/H]) are reliable: `quality_flags < 8`. An example of this cut is seen in Figure 2.

*ARC* (Andrae, Rix, & Chandra 2023b) provides metallicity [M/H], as well as  $T_{\text{eff}}$  and  $\log g$ , for 175 million stellar objects. This is an improvement upon the previous work of Rix et al. (2022) that did this work for two million bright giant stars. ARC utilised the XGBoost algorithm that was trained on 600,000

APOGEE DR17 (Abdurro'uf et al. 2022) spectra ( $S/N > 50$ ) and supplemented with CatWISE photometry (Marocco et al. 2021). The model covers the temperature range from 3107 K to 6867 K, and the metallicity range from -4.37 to +0.5 dex. The training sample includes main-sequence dwarfs and red giant stars, but there is limited coverage for  $[M/H] < -2$  dex and  $T_{\text{eff}} < 4000$  K, as well as very few training examples for metal-poor dwarfs with  $\log g > 3.5$  dex and  $[M/H] < -1$  dex. They achieved a precision of  $\sigma([M/H]) = 0.1$  dex, making it outstanding quality data for its size. Following the prescriptions of the study, we applied several selection cuts:

- $0.5 < G_{BP} - G_{RP} < 2.5$ ,
- $G < 17$  mag,
- filtering out the OBA stars based on the Gaia Golden OBA sample (Gaia Collaboration et al. 2023).

*FS* (Fallows & Sanders 2024) present the predictions of a machine learning model for stellar atmospheric parameters of using Gaia BP/RP spectra and broadband photometric colours from Gaia, 2MASS, and WISE. The model predicts six stellar parameters:  $T_{\text{eff}}$ ,  $\log g$ , iron [Fe/H], carbon [C/Fe], nitrogen [N/Fe], and alpha [ $\alpha/M$ ], along with uncertainties for these parameters. It was trained on a dataset of APOGEE spectra, which includes over 650,000 stars. The model's predictions were validated against APOGEE and LAMOST spectroscopic data, and Gaia GSP-Phot parameters. The quality of the deliverables, particularly for metallicity ([Fe/H]), is reported with a median formal uncertainty of 0.068 dex, indicating high precision in the metallicity predictions. While there are no explicitly recommended cuts for this catalogue, we applied the following limitations:

- $G_{BP} - G_{RP} > 0.5$ ,
- $\sigma [Fe/H] < 0.1$ .

The first cut was dictated by a sharp decline in [Fe/H] for the hottest stars of the MS observed in almost every cluster in the sample (for example, see Figure 2) and is caused by the model being undertrained with OB stars. The second cut was based on the reported median [Fe/H] uncertainty of the catalogue.

Other catalogues of metallicities based on Gaia XP spectra were considered:

*GSP-Phot* (Andrae et al. 2023a) is a module in the Gaia processing pipeline that aims to determine stellar parameters using only Gaia photometry, parallaxes, and BP/RP spectra. However, the [M/H] values exhibit strong systematics and are not recommended for use by the Gaia collaboration. Moreover, Bianchini & Mastrobuono-Battisti (2024) showed the influence of these systematics in the globular cluster NGC 5904, which was previously reported to host populations of two metallicities based on the GSP-Phot data. Comparisons with high-quality ARC metallicity values revealed that the cluster is, in fact, monometallic.

*Pristine* (Martin et al. 2024) presented synthetic CaHK-based metallicities for 220 million stars by training a neural network on CaHK metallicities derived from the Pristine survey. We do not use this otherwise extensive catalogue since it was calibrated for metal-poor stars ( $[Fe/H] < -0.3$  dex), which is lower than the metallicity values of the clusters in our sample.

For the analysis of the HRS sample, we only considered open clusters with more than 100 measurements in both the Zhang et al. (2023), Andrae et al. (2023b) and Fallows & Sanders (2024) catalogues in order to ensure the statistical significance of the derived metallicities. After applying these cuts, our final sample consisted of 20 open clusters.

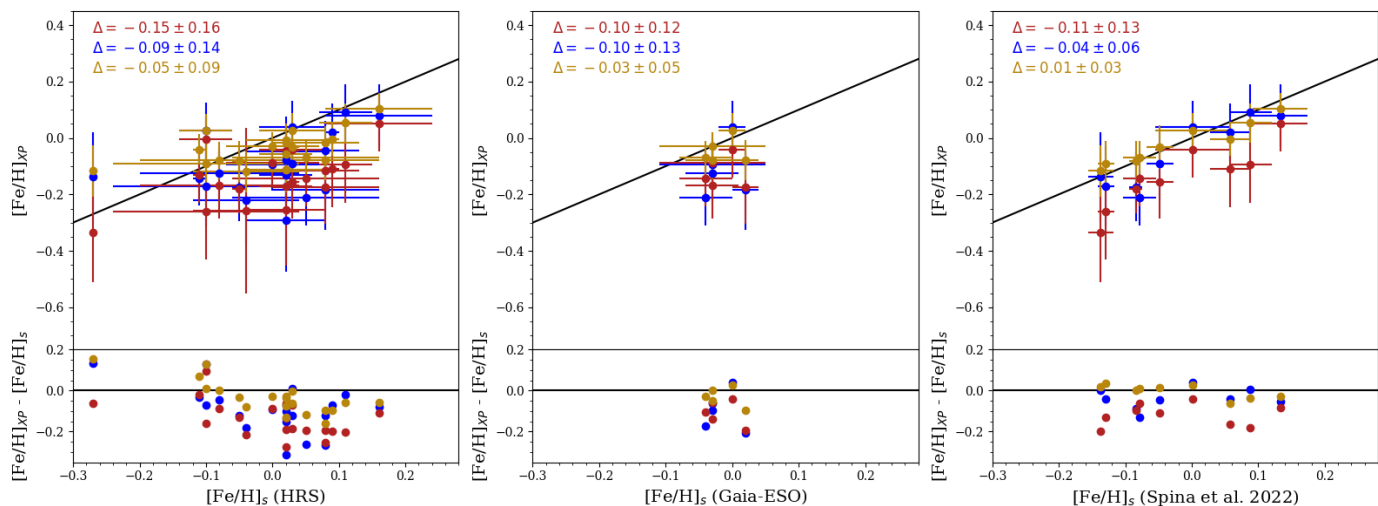


Fig. 3: Top panels: Metallicity of individual OCs of the sample derived from three Gaia XP-based catalogues (on the Y axis) presented against that from three high-resolution spectroscopic surveys (on the X axes of the left, middle and right plots). Bottom: Difference between the two values for each cluster. The same colour-coding scheme applies as in Fig. 1. The 1:1 relation is shown by the diagonal black line in the top plot and, respectively, by the horizontal line indicating zero difference in the bottom plot. The number in the top corner indicates the median difference and the root mean square difference (RMSD) of the respective catalogue.

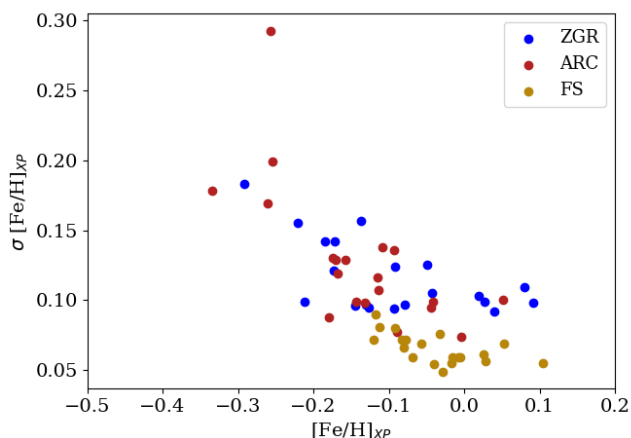


Fig. 4: Iron abundance dispersion  $\sigma[\text{Fe}/\text{H}]$  against metallicity  $[\text{Fe}/\text{H}]$  for individual open clusters. The same colour-coding scheme applies as in Fig. 1.

### 3. Cluster parameters

We used the ASteCA (Automated Stellar Cluster Analysis) package (Perren et al. 2015) to analyse the cluster properties. It is based on the ‘synthetic CMD method’, in which a synthetic cluster is generated based on a provided isochrone and cluster parameters, and then compared to the observational data. The best model in this process is determined with Bayesian statistics and is described by four parameters:  $[\text{Fe}/\text{H}]$ , age, distance modulus and extinction  $A_V$ . We used the 0.5.8 version of the code, which supports the use of approximate Bayesian computation (ABC) via the implementation of pyABC (Schälte et al. 2022).

As input for the code, we used the  $G$  and  $G_{\text{BP}} - G_{\text{RP}}$  magnitudes, their corresponding errors, and a set of PARSEC

1.2S<sup>1</sup> isochrones in the Gaia EDR3 photometric system. For the isochrones, we used a grid of  $7.2 \text{ dex} < \log(\text{age}) < 10.1 \text{ dex}$  with a step of 0.01 dex, and  $-2.1 < [\text{Fe}/\text{H}] < 0.5$  with a step of 0.05 dex.

Although ASteCA includes a decontamination algorithm which then assigns a membership probability to each source, we used the MPs provided with Hunt & Reffert (2023) catalogue. To populate the isochrone of a synthetic cluster we used the initial mass function (IMF) derived by Kroupa (2002). In contrast to the BASE-9 code (von Hippel et al. 2006), which was used by B19, ASteCA allows one to include binary systems in the synthetic cluster. To accommodate for the binary distribution we used the standard values of  $\alpha = 0.5$  and  $\beta = 0.15$ , where the probability of a star being in a binary system is determined by the formula:

$$P = \alpha + \beta \log_{10}(M), \quad (1)$$

and where  $M$  is the mass of the system expressed in  $M_{\odot}$ .

We provided the priors for the parameters since it is a necessary component of the Bayesian method. The prior of the distance modulus was provided as a Gaussian function, the standard deviation  $\omega$  of which was 0.3 mag, and the mean was specified by utilising the median parallaxes of the cluster available with Gaia via the following equation:

$$(m - M) = -5 \log(\bar{\varpi}) - 5, \quad (2)$$

where  $(m - M)_0$  is the distance modulus and  $\bar{\varpi}$  is the median value of the parallax of the stars included as cluster members. The prior of the extinction  $A_V$  was flat and limited to a maximum of 1.4 mag, which is stricter than the 2.5 mag limit used by B19. This choice was made to narrow the parameter space while still safely covering the full range of  $A_V$  values observed in our selected subsample, based on B19 measurements. The prior for age was not limited and the age could vary across the grid values. Lastly, for the prior of metallicity we used one of the following: 1) the fixed  $[\text{Fe}/\text{H}]$  value used by B19, 2) the derived  $[\text{Fe}/\text{H}]$  values of clusters and their respective

<sup>1</sup> <http://stev.oapd.inaf.it/cgi-bin/cmd>

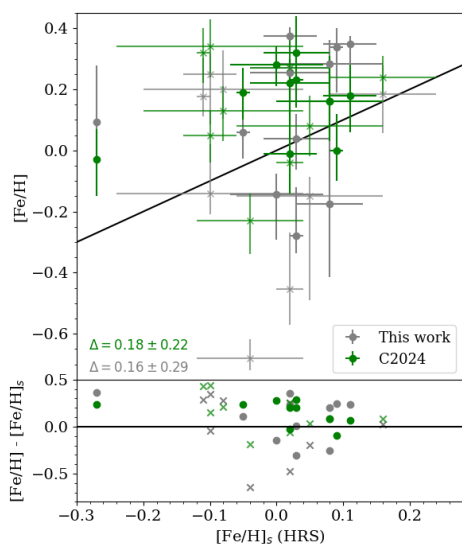


Fig. 5: Metallicity of individual OCs of the HRS sample determined as a free parameter in the procedure. The grey colour denotes this work; the green colour shows the predictions of the ANN of Cavallo et al. (2024). Figure layout is the same as in Fig. 4: X axis shows the  $[\text{Fe}/\text{H}]$  from high-resolution spectroscopy, the bottom plot shows the difference between the values, and the black line indicates 1:1 relation on the top plot and difference of zero in the bottom plot. Clusters with three or fewer red giant branch (RGB) stars are marked with an X symbol. The number in the bottom left corner indicates the median difference and RMSD.

$\sigma_{[\text{Fe}/\text{H}]}$  from the ZGR, ARC, or FS survey,

3) no prior provided and the metallicity was a free parameter.

We performed each run using the pyABC algorithm with a population size of 100 particles per iteration and a maximum of 20 iterations allowed. We established the acceptance threshold at  $\epsilon = 0.01$ , which determines the maximum allowed distance between the simulated and observed datasets for a sample to be accepted.

In order to correctly accommodate for extinction in Gaia photometric bands during the isochrone fitting process, we utilised the conversion of coefficients  $a_{,m}$  from Gaia (E)DR3 auxiliary data<sup>2</sup> available in ASteCA:

$$\frac{A_m}{A_V} = a_{1m} + a_{2m}X + a_{3m}X^2 + a_{4m}X^3 + a_{5m}A_V + a_{6m}A_V^2 + a_{7m}A_V^3 + a_{8m}A_VX + a_{9m}A_VX^2 + a_{10m}XA_V^2, \quad (3)$$

where  $m$  is  $G$ ,  $G_{BP}$  or  $G_{RP}$ , and  $X$  is  $G_{BP} - G_{RP}$ .

## 4. Results

### 4.1. Comparison of $[\text{Fe}/\text{H}]$ with spectroscopic surveys

First, we investigated the distribution of  $[\text{Fe}/\text{H}]$  measurements from each catalogue in each of the clusters of the sample. The median rate of stars in each cluster that have measurements in the ARC, ZGR, and FS catalogues is 72%, 84%, and 75%, respectively. After the cuts applied to these catalogues, the final number of stars for which the  $[\text{Fe}/\text{H}]$  measurements are accepted

<sup>2</sup> <https://www.cosmos.esa.int/web/gaia/edr3-extinction-law>

drops to a median of 50%, 49%, and 38%, respectively. The difference of  $\sim 10\%$  between the FS catalogue and the other two already demonstrates the resulting strictness of the cuts applied to the values of FS.

For every cluster in the sample, we calculated the weighted mean metallicity value from the ZGR and FS catalogues, using the provided errors. For the ARC catalogue, as it does not contain information about errors, we only calculated the mean metallicity. While these surveys provide estimates of either  $[\text{M}/\text{H}]$  (ARC) or  $[\text{Fe}/\text{H}]$  (ZGR, FS), in the further analysis we refer to both of them as  $[\text{Fe}/\text{H}]$ , since the standard deviation of these values exceeds the typical alpha-enhancement.

To assess the reliability and potential biases of these measurements, we compared them with those obtained from high-resolution spectroscopic surveys and used in the HRS sample of B19 (see left panel of Figure 3). According to these estimates, the  $[\text{Fe}/\text{H}]$  of the cluster sample derived from Gaia XP-based catalogues is systematically lower than HRS values. This difference varies between the three catalogues that were used. The median difference between HRS and ARC  $[\text{Fe}/\text{H}]$  estimates equals to  $-0.15 \pm 0.16$  dex, comparable to  $-0.09 \pm 0.14$  dex difference between HRS and ZGR  $[\text{Fe}/\text{H}]$  values. The FS catalogue, in turn, shows a far better agreement with the HRS values. The median difference between HRS and FS is  $-0.05 \pm 0.09$  dex.

While the HRS sample of B19 is based on high-resolution spectroscopic measurements from Netopil et al. (2016), the number of measured stars per cluster is quite limited. Namely, 15 out of 20 clusters in our sample have  $<5$  measurements in Netopil et al. (2016). This is in stark contrast with the average number of 350 stars from the ARC and ZGR catalogues and 250 stars from the FS catalogue, after applying the cuts. To ensure the robustness and consistency in our results, we compared them to the more recent and extensive studies of the metallicity of open clusters with high-resolution spectroscopy: Gaia-ESO and Spina et al. (2022).

The Gaia-ESO survey (GES, Magrini et al. (2023); Randich et al. (2022)) intended to complement the astrometry of Gaia with high-precision radial velocities and abundances using both GIRAFFE ( $R \sim 20000$ ) and UVES ( $R \sim 47000$ ) spectrographs. A part of the survey was specifically dedicated to open clusters and comprised 62 clusters accompanied by ESO archival spectra for another 18 clusters, in total covering an extensive range of ages (0.001–10 Gyr). The main sample has been observed with both spectrographs, whereas the clusters in the ESO archive only have UVES spectroscopy available. We found only five open clusters in our sample that overlap with the GES OC survey: NGC 2516, NGC 3532, NGC 6633, and two archival clusters, NGC 5822 and NGC 2682. While the number of measurements for non-archival data is greater than 700 per cluster, the archival clusters have 4 and 131 measurements, respectively. In total, the average error  $\sigma([\text{Fe}/\text{H}])$  of this sample is 0.04 dex, which is indicative of the high quality of the spectra and benefits from the large number of measurements per cluster. However, this level of precision is available for only a small number of clusters, limiting its broader applicability.

Despite covering a narrower range of  $[\text{Fe}/\text{H}]$ , the results shown in middle panel of Figure 3 are similar to the previous comparison with B19 both in terms of accuracy and precision:  $[\text{Fe}/\text{H}]$  derived from the FS catalogue persistently shows excellent agreement and an RMSD of 0.05 dex. In contrast, the other two catalogues differ from Gaia-ESO measurements by approximately 0.1 dex.

Given the limited overlap of our cluster sample with GES, we also compared the derived metallicities with the catalogue of

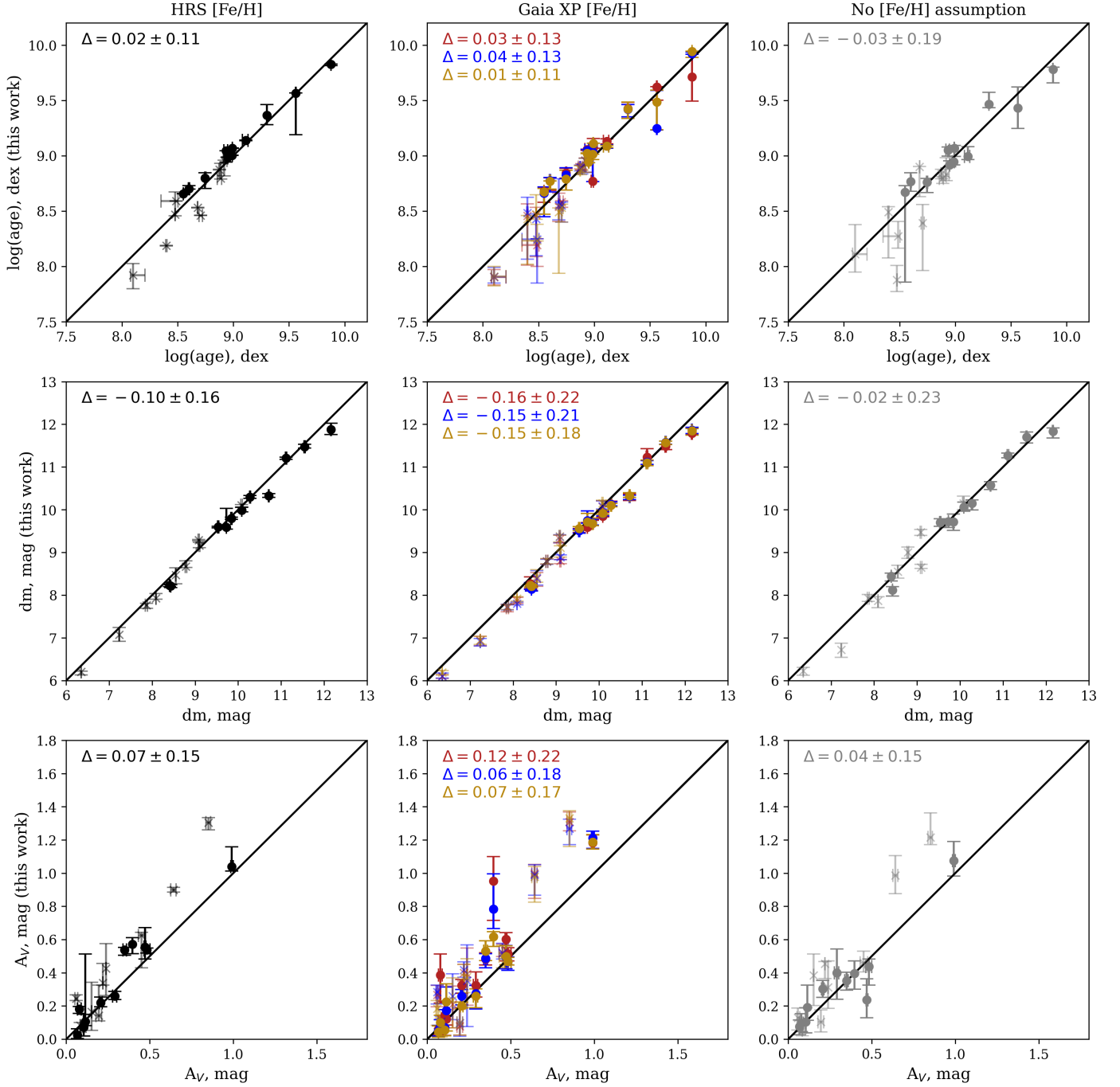


Fig. 6: Parameter estimates for age, distance modulus and extinction (rows, Y axis) of the OCs determined with different [Fe/H] prior information (columns) compared to those from B19 on the X axis. The red, blue, and yellow colours correspond to runs with metallicity prior based on ARC, ZGR and FS catalogues, respectively, whereas black and grey symbols, respectively, depict the runs with the fixed HRS [Fe/H] prior and no prior at all. The legend in the upper corner of the figures shows the median absolute difference and the RMSD of the respective parameter. The OCs with three or fewer RGB stars are marked with an X symbol.

the metallicity of 251 OCs by Spina et al. (2022) (see right panel of Figure 3). They traced the Galactic metallicity gradient by using homogenised [Fe/H] from high-resolution spectroscopic surveys: APOGEE, GES, GALAH, OCCASO, and SPA. We find that 9 out of 20 clusters from our sample have metallicities in Spina et al. (2022). From this selection, the results are consistent with the previous comparisons, as ARC shows a noticeable systematic difference of approximately  $-0.1$  dex, and FS shows a small RMSD of less than 0.05 dex.

The standard error  $\sigma([\text{Fe}/\text{H}])$  exhibits a clear [Fe/H] dependence (Fig.4) and is more pronounced in ZGR and ARC catalogues (see Section 5). The median  $\sigma([\text{Fe}/\text{H}])$  derived from these catalogues for an individual cluster are 0.12 dex and 0.13 dex, respectively, which are comparable to the reported precision of 0.15 dex and 0.10 dex for the whole catalogues. FS shows a smaller median  $\sigma([\text{Fe}/\text{H}])$  of 0.07 dex, which once more reflects the quality of the data combined with the imposed maximum of

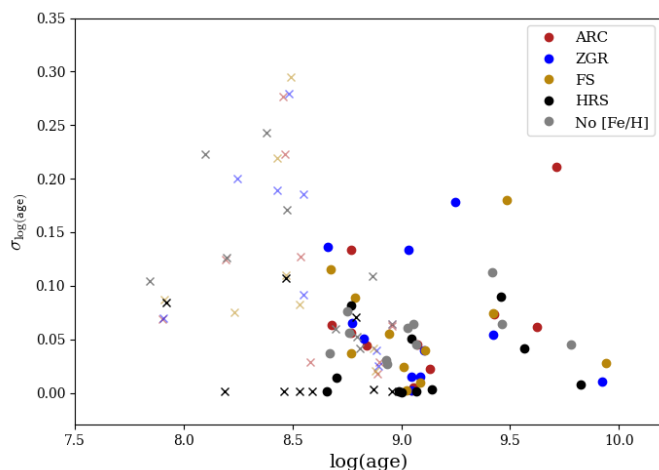


Fig. 7: Error of the fitted age parameter, calculated from the 84th and 16th percentiles of the posterior distribution,  $\sigma_{\log(\text{age})}$  against the  $\log(\text{age})$  (50th percentile) for individual open clusters of the sample. The same colour-coding scheme applies as in Fig. 6.

0.1 dex for uncertainties for [Fe/H] measurements of individual cluster members.

#### 4.2. [Fe/H] as a free parameter

One of the three regimes in which we determined the clusters' parameters is by not providing a [Fe/H] prior at all, leaving it as a free parameter, and allowing it to range across the whole grid. This test was done in order to evaluate how, if at all, necessary the [Fe/H] prior is and how its accuracy in turn affects other parameters. While we do discuss these parameters in the next section, this regime does leave the [Fe/H] estimates found from the best fit as corollary.

As is clearly seen in Figure 5, the [Fe/H] obtained as a parameter in the CMD fitting shows considerably less accurate results than when it is derived from Gaia XP-based catalogues. When compared to the [Fe/H] from the HRS sample (shown on the x axis), these values have a significantly larger RMSD of 0.29 dex, which is nearly double the RMSD achieved using XP-based [Fe/H], which is detailed in Section 4.1. This demonstrates that leaving metallicity unconstrained in the fit leads to degraded performance.

We also compared these results to a neural network-based (NN) study, where [Fe/H] was also determined as a free parameter. Recent study of Cavallo et al. (2024) has shown how to determine the parameters of 5400 open clusters, along with determining the metallicity. They used an approach in which the information about a cluster is extracted in the form of features from the photometry using the QuadTree algorithm, and these features are in turn passed to the ANN trained on a grid of synthetic clusters. We refer to the original paper for a more detailed explanation of the technique.

While the method of Cavallo et al. (2024) allows for the retrieval of some [Fe/H] information from the photometry of open clusters without using the spectra, it has shown an RMSD of 0.35 dex when verified against the study of Dias et al. (2021) and was reported to show 'mild correlation' with the values of Spina et al. (2022). For our sample, we extracted the [Fe/H] values from Cavallo et al. (2024) and compared them to the HRS data: the resulting RMSD is 0.22 dex, which is similar to the free-fit

approach but also shows a large systematic offset of  $\sim 0.2$  dex, with [Fe/H] of some clusters overestimated by as much as 0.5 dex. This result highlights that XP-derived metallicities, despite the low resolution of the spectra, provide far better constraints on [Fe/H] than fitting it as a free parameter or relying solely on photometric machine-learning estimates.

#### 4.3. Comparison of cluster parameters with B19

The values of  $\log(\text{age})$ , distance modulus (dm) and extinction  $A_V$  that were determined for the clusters in the sample are the median values of the respective posterior distribution, whereas the reported uncertainties are the 16th and 84th quantiles. We show these estimates as rows in Figure 6, where the columns show different [Fe/H] prior assumptions. The median absolute differences and RMSD of these parameters with B19 are presented in Table 1.

##### 4.3.1. Age

Since our sample is based on the HRS sample of B19, we first analysed the cluster parameters that were derived with the [Fe/H] being fixed at the same values as in their analysis (top left panel of Figure 6). Before analysing the effect of various metallicity assumptions on the determination of ages, we evaluated possible discrepancies that arise due to differences between the methodologies of our work and their study.

The ages demonstrate a reasonable agreement with the ages of B19, with a median difference of 0.02 dex and an RMSD of 0.11 dex. The clusters that show the biggest difference in the ages are those without a significant red giant population, which in turn tend to be on the younger side of the covered age range ( $\log \text{age} < 8.5$  dex). This result is similar in accuracy to how the B19 HRS sample compares to the well-established catalogues of MWSC and DAML, with differences of  $\Delta \log \text{age} = 0.04 \pm 0.10$  dex and  $0.05 \pm 0.11$  dex, respectively (see Table 4 in their work).

While in this run the [Fe/H] prior is the same as in their work, as well as the PARSEC isochrones and the [Fe/H] and age steps of the isochrone grid, the discrepancies between the results can be attributed to the algorithms used for isochrone fitting. We use pyABC as implemented by ASteCA, whereas B19 fit the parameters with MCMC via BASE-9. Moreover, while we allow  $A_V$  to vary across the whole range of values, they use prior  $A_V$  information from DAML and MWSC catalogues and limit the prior within  $\sigma_{A_V} = \frac{1}{3}A_V$ .

We then considered how a [Fe/H] based on Gaia XP, instead of high-resolution spectroscopy, alters the age determination of the clusters (top middle panel of Figure 6). The results show reasonable agreement and significant dispersion, with the median differences between the ages not exceeding 0.05 dex. The ages derived using the [Fe/H] from the FS catalogue show better agreement with the HRS sample, with an RMSD of 0.11 dex, than those derived with ARC or ZGR values (0.13 dex). The accuracy of the ages using the FS catalogue is equal to the ages derived with metallicities from high-resolution spectroscopy, demonstrating the potential of Gaia XP-based metallicities.

Lastly, we inspected the ages derived from the isochrone fits with no [Fe/H] prior, shown in the top right panel of Figure 6. Since the [Fe/H] derived from these fits has a substantial RMSD (see Section 4.2), the age estimates also show a considerably larger RMSD of  $\sim 0.2$  dex in contrast to the fits with [Fe/H]

Table 1: Median absolute difference and the RMSD of the parameter estimates determined with different metallicity priors with respect to B19 values.

		HRS [Fe/H]	Gaia XP [Fe/H]			No [Fe/H] assumption
			ARC	ZGR	FS	
$\Delta \log(age)$	All	$0.02 \pm 0.11$	$0.03 \pm 0.13$	$0.04 \pm 0.13$	$0.01 \pm 0.11$	$-0.03 \pm 0.19$
	RGB	$0.05 \pm 0.07$	$0.10 \pm 0.13$	$0.10 \pm 0.14$	$0.07 \pm 0.09$	$0.02 \pm 0.11$
	No RGB	$-0.09 \pm 0.14$	$-0.01 \pm 0.14$	$-0.05 \pm 0.13$	$0.00 \pm 0.14$	$-0.09 \pm 0.25$
$\Delta dm$	All	$-0.10 \pm 0.16$	$-0.16 \pm 0.22$	$-0.15 \pm 0.21$	$-0.15 \pm 0.18$	$-0.02 \pm 0.23$
	RGB	$-0.10 \pm 0.18$	$-0.14 \pm 0.22$	$-0.15 \pm 0.20$	$-0.17 \pm 0.19$	$-0.03 \pm 0.17$
	No RGB	$-0.10 \pm 0.13$	$-0.17 \pm 0.21$	$-0.16 \pm 0.21$	$-0.12 \pm 0.17$	$0.00 \pm 0.29$
$\Delta A_V$	All	$0.07 \pm 0.15$	$0.12 \pm 0.22$	$0.06 \pm 0.18$	$0.07 \pm 0.17$	$0.04 \pm 0.15$
	RGB	$0.05 \pm 0.09$	$0.12 \pm 0.21$	$0.02 \pm 0.14$	$0.02 \pm 0.11$	$0.01 \pm 0.09$
	No RGB	$0.17 \pm 0.21$	$0.16 \pm 0.23$	$0.11 \pm 0.22$	$0.12 \pm 0.22$	$0.07 \pm 0.21$
$\Delta [Fe/H]$	All	—	$-0.15 \pm 0.16$	$-0.09 \pm 0.14$	$-0.05 \pm 0.09$	$0.16 \pm 0.29$
	RGB	—	$-0.19 \pm 0.18$	$-0.12 \pm 0.16$	$-0.06 \pm 0.09$	$0.20 \pm 0.25$
	No RGB	—	$-0.11 \pm 0.13$	$-0.07 \pm 0.13$	$-0.03 \pm 0.07$	$0.03 \pm 0.34$

**Notes.** The results for all clusters are also shown in Fig. 6, and the results of the Gaia XP metallicities are shown in Fig. 3. They are also separated into two groups, based on the presence of RGB stars ('RGB': >3 stars, 'No RGB': ≤3 stars).

prior information. The main contributors to this dispersion are young clusters with no defined post-main sequence (MS) stars, as expected, whereas older clusters show a small difference in age. This is seen in Table 1, where the 'No RGB' subgroup exhibits the largest RMSD of 0.25 dex in  $\log(age)$  when no metallicity prior is used while clusters with RGB stars remain largely unaffected by using different metallicity priors. These findings suggest that the inclusion of a metallicity prior is particularly beneficial for clusters lacking evolved stars, as it significantly enhances the accuracy of their age estimates.

The comparisons outlined above establish the accuracy of the ages that we derived. In order to assess their precision, we calculated the median  $\sigma_{\log(age)}$  from the individual age estimates (see Fig. 7). The median  $\sigma_{\log(age)}$  of the ARC, ZGR, and FS catalogues is 0.062 dex, 0.065 dex, and 0.069 dex, respectively. If we only take into consideration the clusters with >3 post-MS stars, the respective average  $\sigma_{\log(age)}$  decreases to 0.056 dex, 0.050 dex, and 0.040 dex. This is comparable to the median  $\sigma_{\log(age)}$  for estimates obtained without a [Fe/H] prior (0.056 dex). The precision of the ages derived with the HRS [Fe/H] is 0.008 dex, which is an order of magnitude lower and is dictated by the fixed value of [Fe/H], which in turn reduces the degeneracy between the fit parameters.

#### 4.3.2. $dm$ and $A_V$

We also compared the other two parameters of the fitting procedure: distance modulus and extinction  $A_V$  (middle and bottom rows of Figure 6 respectively). When derived from fits with different [Fe/H] assumptions, distance modulus demonstrates the same trend of dispersion in comparison to the values of B19, as the fits with Gaia XP-based [Fe/H] show bigger dispersion than with fixed HRS [Fe/H] (0.16 mag) but smaller than with no [Fe/H] prior at all (0.23 mag). Fits that used the FS catalogue have a smaller RMSD between the three catalogues, in contrast to the fits that used [Fe/H] of the ARC catalogue. Similar to  $\log(age)$  estimates, the  $dm$  estimates for clusters without RGB stars (see Table 1) show the largest improvement from including Gaia XP-based priors, as RMSD drops from 0.29 mag without the prior to 0.17 for the FS catalogue. Clusters with RGB stars, however, show minimal improvement of the  $dm$  estimate.

As for the extinction  $A_V$ , it is consistently overestimated in all three types of fits for two clusters with  $A_V > 0.6$  mag in B19 and an unpopulated RGB. All three types of fits show a positive offset, with that of the ARC catalogue being the largest. Notably, the RMSD of the  $A_V$  of clusters with HRS [Fe/H] and no [Fe/H] prior is the same. The presence of RGB stars has a clear influence on the accuracy of the extinction estimate, as clusters with RGB stars have a significantly lower RMSD (0.09 mag) compared to those without (0.21 mag). However, Gaia XP-based priors do not appear to improve the precision of  $A_V$  in either group, in contrast to  $\log(age)$  and  $dm$  parameters. For clusters with RGB stars, the use of Gaia XP [Fe/H] actually increases the dispersion in extinction estimates. We examine this issue in the next section.

#### 4.4. Correlation analysis

In the previous section we evaluated the accuracy of the parameters independently of each other. However, since isochrone fitting is a multi-parameter fit, we analysed the correlations between the offsets in all fitted parameters.

For the solutions without a metallicity prior (Fig. 8a), we find a strong negative correlation between  $\Delta \log(age)$  and  $\Delta dm$  (Pearson correlation coefficient  $r = -0.92$ ,  $p < 0.001$ ). This is to be expected, since increasing the age shifts the MSTO lower and decreasing the distance modulus shifts the isochrone upwards in the CMD, thus compensating each other. This degeneracy has also been noted in other isochrone-fitting works (e.g. Ying et al. (2023)). Two clusters with no RGB (NGC 1912 and NGC 6475, marked as green) stand out from this clear trend by having a large  $\Delta \log(age)$  and were not included in the correlation calculations. These clusters also show significant offsets in the other two parameters, and, in case of NGC 6475, also  $\Delta [Fe/H]_{fit}$ . Although their isochrones appear consistent with the photometric data upon visual inspection (Figure B.1), the absence of an RGB substantially reduces the reliability of the fits.

The metallicity offset  $\Delta [Fe/H]_{fit}$  with respect to the B19 values also shows noticeable correlations with all other parameters: with  $\log(age)$  ( $r = -0.66$ ,  $p = 0.003$ ),  $dm$  ( $r = +0.71$ ,  $p = 0.001$ ), and  $A_V$  ( $r = -0.73$ ,  $p < 0.001$ ). An isochrone with a higher [Fe/H] will have a redder main sequence and RGB, which can be compensated by increasing  $A_V$ , lowering the turn-off age, or shifting the  $dm$ . As mentioned in Section 4.3.2,  $A_V$  values in

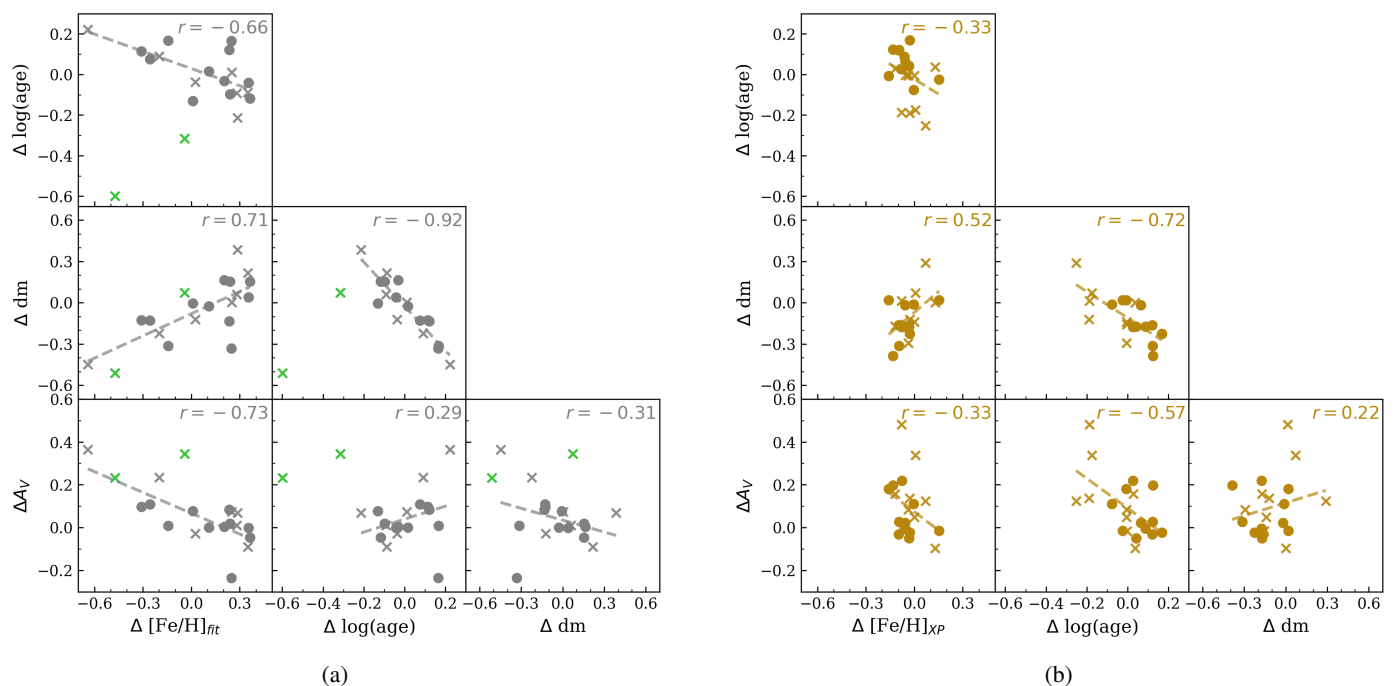


Fig. 8: Corner plot of parameter offsets relative to B19 for the solutions performed (a) without a  $[\text{Fe}/\text{H}]$  prior and (b) with a  $[\text{Fe}/\text{H}]$  prior calculated from the FS catalogue. Two clusters with exceptionally large  $\Delta \log(\text{age})$  are marked with green markers and are excluded from the correlation calculation in the plot (a). The dashed lines show linear fits to the data, and the Pearson correlation coefficient  $r$  is shown for each pair of parameters. Clusters with unpopulated RGB are shown with an X marker.

this regime exhibit less scatter compared to the fits with Gaia XP prior. In the absence of a  $[\text{Fe}/\text{H}]$  constraint, the fit is free to adjust the metallicity jointly with  $A_V$ , so the  $[\text{Fe}/\text{H}]$  estimate can absorb the uncertainties. Whereas when  $[\text{Fe}/\text{H}]$  prior is provided, the degeneracy forces this tension into  $A_V$ , increasing its dispersion.

For the solutions with a  $[\text{Fe}/\text{H}]$  prior from the FS catalogue (Fig. 8b), the  $\Delta \log(\text{age})$ – $\Delta \text{dm}$  correlation persists, but with a weaker slope ( $r = -0.72$ ,  $p < 0.001$ ), while other strong correlations are weakened. The extinction offsets are also reduced and remain close to those of the high-resolution spectroscopy solutions, reflecting the small systematic offset in  $[\text{Fe}/\text{H}]$  with respect to B19.

The ARC catalogue (Figure B.2a) shows a correlation of  $r = -0.55$  ( $p = 0.11$ ) between  $\Delta[\text{Fe}/\text{H}]_{\text{XP}}$  and  $\Delta A_V$ , in contrast to  $r = -0.33$  ( $p = 0.16$ ) for the FS catalogue. This is consistent with ARC’s larger median metallicity offset ( $-0.1$  dex relative to FS, see Table 1), which in turn propagates into larger extinction offsets. The ZGR catalogue shows similar behaviour but with slightly weaker correlations ( $r = -0.50$ ,  $p = 0.24$  between  $\Delta[\text{Fe}/\text{H}]_{\text{XP}}$  and  $\Delta A_V$ ), see Figure B.2b.

The fits that used the HRS  $[\text{Fe}/\text{H}]$  prior (Figure B.3) do not show a correlation between  $\Delta \log(\text{age})$  and  $\Delta \text{dm}$ , but instead some correlation between  $\Delta \log(\text{age})$  and  $\Delta A_V$  ( $r = -0.60$ ,  $p = 0.005$ ). However, the clusters driving these offsets are those lacking a well-populated RGB, whereas for other clusters these offsets are not correlated.

These correlations demonstrate the presence of  $\log(\text{age})$ – $\text{dm}$  degeneracy in isochrone fitting, as well as highlight that  $[\text{Fe}/\text{H}]$  errors are a primary source of extinction biases. The strength of these correlations varies between the adopted  $[\text{Fe}/\text{H}]$  priors: the ARC prior showed the strongest metallicity–extinction coupling, whereas the FS was closest to the benchmark values of B19. Adopting these priors helps diminishing the strong corre-

lations between all four parameters seen in the regime with no  $[\text{Fe}/\text{H}]$  prior at all.

#### 4.5. Comparison of cluster parameters with NN studies

To further evaluate the impact of incorporating a metallicity prior ( $[\text{Fe}/\text{H}]$ ) in our parameter estimates, we compared our results with three recent studies that derive cluster parameters using trained neural networks: CG20, HR23, and Cavallo et al. (2024). These studies did not rely on spectroscopic information for individual clusters but instead obtained parameters directly from their machine-learning approaches, bypassing the need for standard isochrone fitting procedures. We demonstrate these results in Figure 9, in a similar fashion: rows show the three parameters obtained in the fit, whereas columns denote the study to which the comparison was made.

CG20 (left column) employed a neural network trained on 347 reference clusters, 269 of which were also part of the B19 sample. They do not derive or fix  $[\text{Fe}/\text{H}]$  as one of the input parameters, instead their network was trained on a grid that covered a large range of metallicities. The agreement in the cluster ages and our work is good, with a comparable level of accuracy to the agreement we observed with B19. This is expected due to the overlap in training data used by their network.

However, the distance modulus (dm) values derived by CG20 show a significant negative systematic offset and dispersion compared to our results. The extinction values also exhibit a large offset and RMSD, indicating discrepancies in their estimation for clusters of not just high, but also low extinction.

HR23 (middle column) included metallicity information in their analysis in the same implicit way as CG20. The cluster ages derived in their work show poorer agreement with ours with a dispersion of  $\sim 0.2$  dex. This highlights the limitations of their approach for precise age determinations.

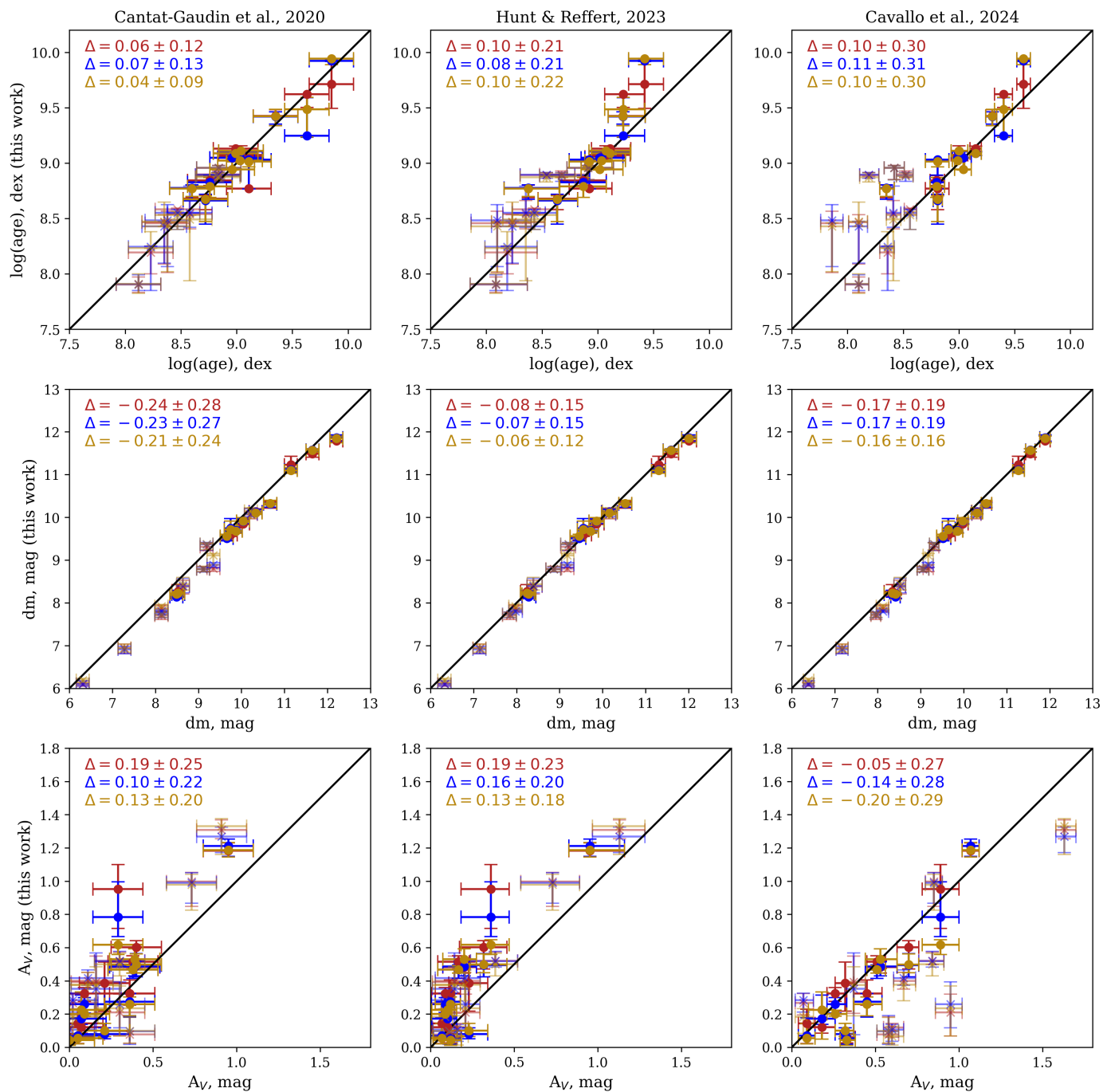


Fig. 9: Parameter estimates for age, distance modulus and extinction (rows, Y axis) of the OCs determined with a  $[\text{Fe}/\text{H}]$  prior based on Gaia XP metallicity compared to those estimates reported by the studies of CG20, HR23, and Cavallo et al. (2024) on the X axis. The same colour-coding scheme applies as in Fig. 6. The legend in the upper corner of the figures shows the median absolute difference and the RMSD of the respective parameter. The OCs with three or fewer RGB stars are marked with an X symbol.

In contrast, the dm values from HR23 align well with our results and even show better agreement compared to the B19 sample. However, similar to CG20, the extinction values  $A_V$  follow a similar pattern of significant offsets and large dispersion.

As previously discussed, Cavallo et al. (2024) (left column) utilised neural networks to estimate cluster parameters. Their age estimates, particularly for young clusters and those lacking a significant RGB population, show large discrepancies and a high level of scatter. This issue is consistent with the challenges highlighted in the CG20 and HR23 analyses. The overestimation of

ages was also noted by Cavallo et al. (2024) during the scientific verification of their whole dataset.

The distance modulus results from Cavallo et al. (2024) are similar in accuracy to those derived from B19. The extinction values  $A_V$  show considerable scatter, especially for young clusters with no RGB population, reinforcing the limitations of their methodology in handling such cases.

In summary, these comparisons underscore the importance of incorporating reliable metallicity priors, such as those derived from high-resolution spectroscopy or Gaia XP, to enhance the

accuracy and consistency of cluster parameter estimates. While neural network-based approaches offer promising alternatives considering the demands of large datasets and an ever-growing number of clusters, they face significant challenges, particularly in estimating ages, metallicity and extinction values for specific types of clusters. As a potential solution, one can use the [Fe/H] values based on Gaia XP spectra as priors in the NN.

## 5. Discussion

In this study we relied exclusively on Gaia DR3 data, including Gaia photometry, parallax, and studies that use Gaia XP spectra, for deriving cluster metallicities and use them in age determination through isochrone fitting. While Gaia data provides extensive coverage and unprecedented astrometric precision, the low resolution of the XP spectra introduces several challenges in accurately estimating stellar parameters, particularly in metallicities. Machine learning techniques allow us to extract metallicity information from these XP spectra, but they are in turn limited by the parameter range of the training data.

The [M/H] that we obtained from the ARC catalogue shows a far larger negative offset compared to the values derived from high-resolution spectroscopy in contrast with the ZGR and FS catalogues. The algorithm of ARC was undertrained for hot stars, since the reliable  $T_{\text{eff}}$  range of the training data is 3107 K to 6867 K, which corresponds to the unreddened  $BP - RP$  colour indices between 0.5 and 2.3. The imposed cut of  $BP - RP > 0.5$  mag does eliminate most of these hot stars that show lower metallicities than the rest; however, that is not the case for several clusters in the sample. In the clusters with higher  $A_V$ , the observed  $BP - RP$  value is affected by the extinction, which in turn makes the hot stars appear redder and pass the cut, thereby lowering the derived [Fe/H] and increasing the corresponding error (see IC 2714 in Fig. A.1).

We tested a stricter colour range of  $0.8 < BP - RP < 1.2$  to see if it would improve the metallicity estimates. However, the median improvement for ARC was  $\sim 0.04$  dex, which is negligible within the catalogue uncertainty. Moreover, this cut drastically decreased the amount of usable members, which in turn cut in half our sample of clusters with  $>100$  members in each catalogue.

We performed another test in which we first corrected the  $BP - RP$  values using the  $A_V$  values determined in B19 and the conversion (3), then we repeated the process of metallicity determination. The  $A_V$  correction did only improve both the median difference and RMSD of ARC [Fe/H] by  $\sim 0.02$  dex, which is negligible and does not subsequently lead to significant differences in cluster ages, as seen in the results obtained with three XP-based metallicity catalogues. This demonstrates that our initial approach allows us to derive reliable cluster parameters using only Gaia data, in contrast to B19, which based the  $A_V$  prior on DAML and MWSC catalogues.

Another reason for the random error of [Fe/H] measurements is the main population of the open clusters in the sample, which is main sequence stars. Andrae et al. (2023b) report an error of  $\sim 0.33$  dex at  $G \sim 17$ , which is caused by the noise in the XP spectra that have a lower S/N ratio with increasing magnitude.

The FS catalogue was also trained on the APOGEE data, much like ARC, yet it consistently shows better agreement with the results of B19 both in derived metallicities and cluster parameters. While ARC and ZGR predict [M/H], FS's methodology predicts [Fe/H] along with [C/Fe], [N/Fe] and [ $\alpha$ /M] and with a better precision, as it uses uncertainty estimation and additionally uses 2MASS photometry to break temperature degeneracies.

In this study we equated [M/H] to [Fe/H] for these catalogues, which led to noticeable discrepancies in the results.

Between the derived cluster parameters, the largest discrepancies with literature values were seen in  $A_V$ , which could be caused by a combination of factors. First, in our fitting procedure with ASteCA we include unresolved binaries, which can mimic the effects of extinction by making stars appear redder, if not accounted for. Meanwhile, the BASE-9 code used by B19 did not include binary systems. Second, the majority of clusters in our sample have relatively low extinction values ( $A_V < 0.5$ ), where even small differences in photometry or assumptions can lead to large relative deviations from other studies. Third, the largest discrepancies between our  $A_V$  estimates and those of B19 are found at  $A_V > 0.5$  and persist across all solutions, independent of whether they used a Gaia XP prior, an HRS prior, or no metallicity prior. These extinction estimates lie outside the uncertainty threshold of  $\sigma_{A_V} = \frac{1}{3}A_V$  adopted by B19, which was not applied in our analysis. Introducing an extinction prior from an independent source and marginalising over it would help to reduce these discrepancies. Finally, some clusters lack a red giant branch or red clump, which typically provide strong constraints on extinction in CMD fitting. In their absence, extinction estimates become more sensitive to the placement and shape of the main sequence,

Lastly, we imposed a floor of 50% for the membership probability of the cluster members to be included in the analysis. This allowed us to include the most amount of stars for [Fe/H] determination. However, this cut in turn included individual stars or groups of stars that did not resemble cluster members upon a visual inspection of the CMD (see NGC 3532 in Fig. B.1), but we did not exclude them to preserve the consistency of the methodology. Since ASteCA does not include assigning weights to the stars depending on their MP, their presence could potentially affect the goodness of the fit and the fit parameters.

## 6. Conclusion

We conducted an analysis of 20 Milky Way open clusters, focusing on their metallicity and cluster parameter determination using data from the Gaia mission alone. We used three existing catalogues that determined stellar parameters from all available 220 million Gaia XP spectra, from which we derived the [Fe/H] values of open clusters, observing differences of 0.1 dex for the ZGR catalogue, 0.15 dex for the ARC catalogue and 0.07 dex for the FS catalogue when matched against high-resolution spectroscopy values. The median errors in [Fe/H] measurements per cluster using these catalogues were found to be 0.12 dex, 0.13 dex, and 0.07 dex, respectively. These findings indicate that while the Gaia XP spectra yield reliable metallicity estimates, they exhibit systematic differences from traditional methods, particularly at lower metallicities. Moreover, the better agreement of FS metallicities was achieved by imposing a stricter cut on the accepted metallicity values in the clusters of the study.

By utilising the ASteCA code, we estimated the parameters of open clusters by comparing synthetic clusters generated with PARSEC isochrones. Our results show a good agreement with the values of B19. This method allowed us to determine cluster ages with the comparable accuracy to their study that required costly and time-consuming spectroscopic observations to lift the age-metallicity degeneracy. Our approach effectively reduced uncertainties and degeneracies between the parameters by incorporating [Fe/H] as a prior, enhancing the accuracy of our age determinations primarily for open clusters without a populated RGB. The results indicate an improvement in age determi-

nation accuracy compared to the ages determined with a neural network by CG20, HR23 and Cavallo et al. (2024). However, the estimates of extinction and distance modulus do not demonstrate a significant uncertainty reduction.

Our findings demonstrate the potential of utilising only Gaia data for studying the parameters of Milky Way open clusters, after applying sufficient cuts to the Gaia XP metallicity estimates, thereby enhancing our understanding of stellar evolution within these groupings. Looking forwards, this method can be expanded to the whole sky to encompass all clusters. Moreover, the estimations of  $[\alpha/\text{Fe}]$  from Gaia XP spectra (Guiglion et al. 2024) can help in further constraints of the chemical composition of the open clusters and its influence of the cluster parameters and the evolution of the galaxy. While these methods remain model-dependent, the integration of asteroseismology and chemical clocks to determine the ages will further refine the methodologies of cluster age determination.

## 7. Data availability

Tables B.1 and B.2 with the derived cluster parameters are only available in electronic form at the CDS via anonymous ftp to <https://cdsarc.cds.unistra.fr/> (130.79.128.5) or via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/>.

*Acknowledgements.* The authors thank the anonymous referee for their suggestions, which helped improve the quality of this manuscript. The authors also thank G. Perren for his extensive advice regarding the applications of ASteCA. This work used the following open source software: TOPCAT (Taylor 2005), ezpadova. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This research has made use of the tool provided by Gaia DPAC (<https://www.cosmos.esa.int/web/gaia/dr3-software-tools>) to reproduce (E)DR3 Gaia photometric uncertainties described in the GAIA-C5-TN-UB-JMC-031 technical note using data in Riello et al. (2021). This research or product makes use of public auxiliary data provided by ESA/Gaia/DPAC/CU5 and prepared by Carine Babusiaux. This research has made use of NASA's Astrophysics Data System Bibliographic Services. This work was supported by NOVA (the Netherlands Research School for Astronomy).

## References

Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35  
 Andrae, R., Fouesneau, M., Sordo, R., et al. 2023a, A&A, 674, A27  
 Andrae, R., Rix, H.-W., & Chandra, V. 2023b, ApJS, 267, 8  
 Angus, R., Morton, T. D., Foreman-Mackey, D., et al. 2019, AJ, 158, 173  
 Bianchini, P. & Mastrobuono-Battisti, A. 2024, MNRAS, 527, L32  
 Bossini, D., Vallenari, A., Bragaglia, A., et al. 2019, A&A, 623, A108  
 Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1  
 Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93  
 Castro-Ginard, A., Jordi, C., Luri, X., et al. 2019, A&A, 627, A35  
 Cavallo, L., Spina, L., Carraro, G., et al. 2024, AJ, 167, 12  
 Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871  
 Dias, W. S., Monteiro, H., Moitinho, A., et al. 2021, MNRAS, 504, 356  
 Fallows, C. P. & Sanders, J. L. 2024, MNRAS, 531, 2126  
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1  
 Gaia Collaboration, Creevey, O. L., Sarro, L. M., et al. 2023, A&A, 674, A39  
 Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1  
 Guiglion, G., Nepal, S., Chiappini, C., et al. 2024, A&A, 682, A9  
 Hattori, K. 2025, ApJ, 980, 90  
 Hunt, E. L. & Reffert, S. 2021, A&A, 646, A104  
 Hunt, E. L. & Reffert, S. 2023, A&A, 673, A114  
 Kane, S. G., Belokurov, V., Cranmer, M., et al. 2025, MNRAS, 536, 2507  
 Khalatyan, A., Anders, F., Chiappini, C., et al. 2024, A&A, 691, A98  
 Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R.-D. 2005, A&A, 438, 1163  
 Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R. D. 2013, A&A, 558, A53

Kroupa, P. 2002, Science, 295, 82  
 Li, J., Wong, K. W. K., Hogg, D. W., Rix, H.-W., & Chandra, V. 2024, ApJS, 272, 2  
 Liu, L. & Pang, X. 2019, ApJS, 245, 32  
 Lucey, M., Al Kharusi, N., Hawkins, K., et al. 2023, MNRAS, 523, 4049  
 Magrini, L., Viscasillas Vázquez, C., Spina, L., et al. 2023, A&A, 669, A119  
 Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., et al. 2021, ApJS, 253, 8  
 Martin, N. F., Starkeburg, E., Yuan, Z., et al. 2024, A&A, 692, A115  
 Monteiro, H., Dias, W. S., Moitinho, A., & Lépine, J. R. D. 2019, MNRAS, 487, 2385  
 Netopil, M., Paunzen, E., Heiter, U., & Soubiran, C. 2016, A&A, 585, A150  
 Palakkatharappil, D. B. & Creevey, O. L. 2023, A&A, 674, A146  
 Perren, G. I., Vázquez, R. A., & Piatti, A. E. 2015, A&A, 576, A6  
 Randich, S., Gilmore, G., Magrini, L., et al. 2022, A&A, 666, A121  
 Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3  
 Rix, H.-W., Chandra, V., Andrae, R., et al. 2022, ApJ, 941, 45  
 Schälte, Y., Klinger, E., Alamoudi, E., & Hasenauer, J. 2022, JOSS, 7, 4304  
 Sim, G., Lee, S. H., Ann, H. B., & Kim, S. 2019, JKoAS, 52, 145  
 Spina, L., Magrini, L., & Cunha, K. 2022, Universe, 8, 87  
 Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29  
 von Hippel, T., Jefferys, W. H., Scott, J., et al. 2006, ApJ, 645, 1436  
 Yao, Y., Ji, A. P., Koposov, S. E., & Limberg, G. 2024, MNRAS, 527, 10937  
 Ye, X., Wu, W., Allende Prieto, C., et al. 2025, A&A, 695, A75  
 Ying, J. M., Chaboyer, B., Boudreux, E. M., et al. 2023, AJ, 166, 18  
 Zhang, X., Green, G. M., & Rix, H.-W. 2023, MNRAS, 524, 1855

## Appendix A: [Fe/H] estimates

Figures A.1 and A.2 show the [Fe/H] estimates from the catalogues ARC, ZGR, and FS for individual stars of each cluster in our sample.

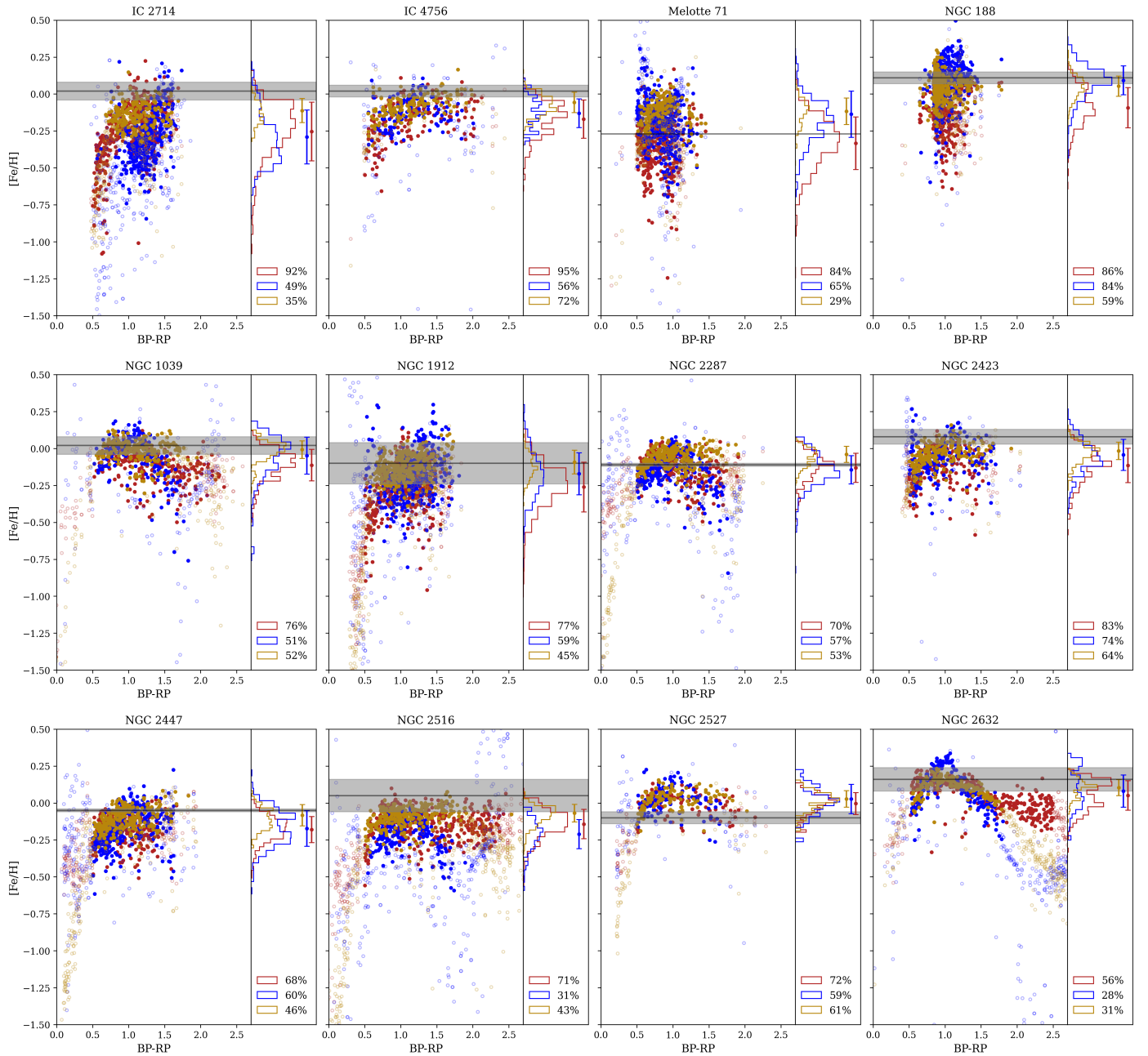


Fig. A.1: [Fe/H] estimates from three catalogues shown as a function of Gaia  $G_{BP} - G_{RP}$  colour for each cluster of the sample. The same colour-coding scheme and legend description apply as in Fig. 2. The black line and the shaded area represent the HRS [Fe/H] and  $\sigma$ [Fe/H] determined by [Netopil et al. \(2016\)](#) and used by B19. The points with error bars show the weighted mean [Fe/H] and its dispersion of the respective catalogue, these values are listed in Table B.1. Melotte 71 does not depict the error of the HRS [Fe/H] measurement since [Netopil et al. \(2016\)](#) obtained it from one star.

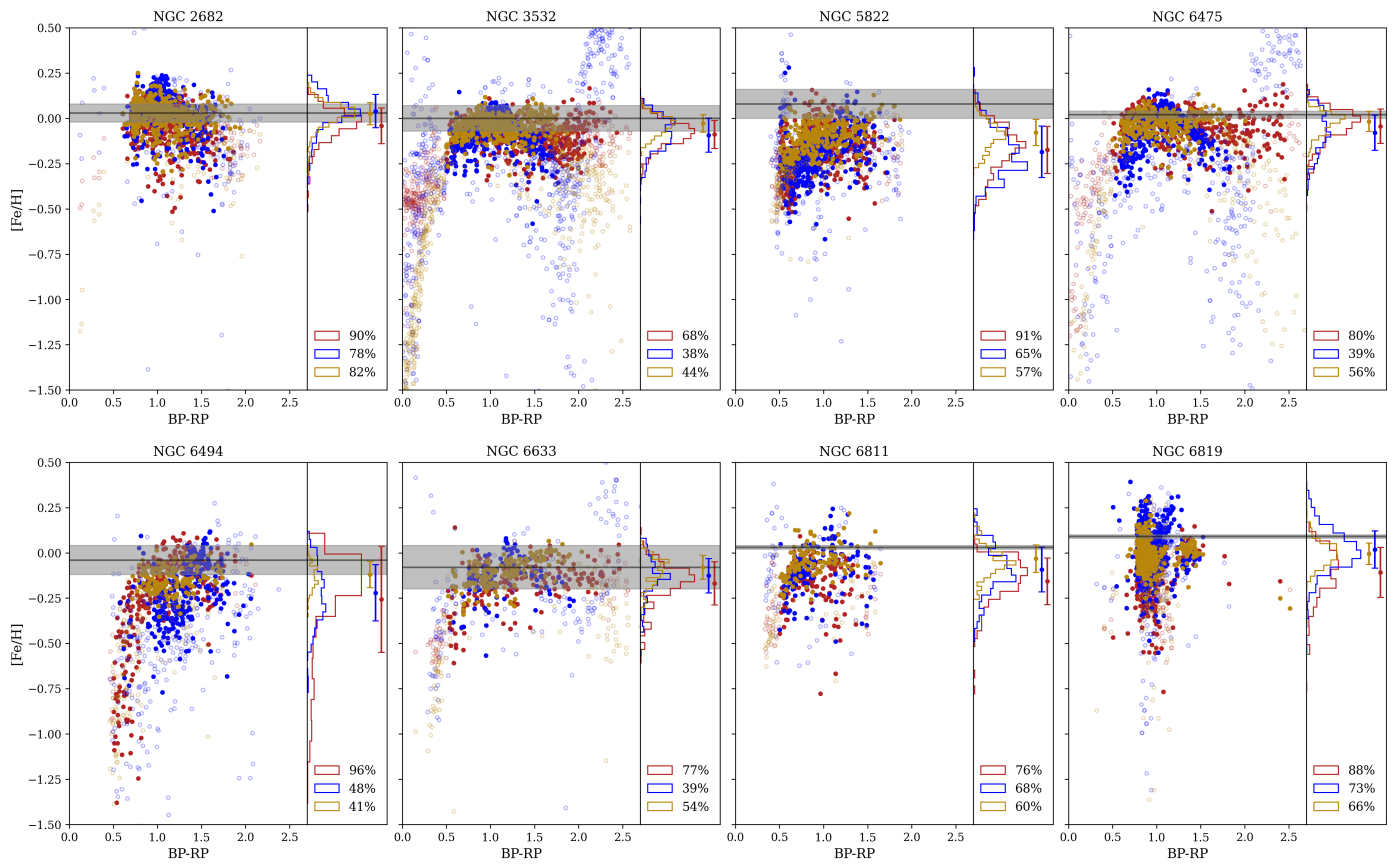


Fig. A.2:  $[Fe/H]$  estimates from three catalogues shown as a function of Gaia  $G_{BP} - G_{RP}$  colour for each cluster of the sample. The same colour-coding scheme and legend description apply as in Fig. A.1.

**Appendix B: Cluster parameters**

Parameters of 20 OCs derived with our analysis and the isochrones corresponding to the respective sets of parameters.

Table B.1: List of clusters and their [Fe/H] estimates subsequently used as prior assumptions.

Cluster	RA (deg)	Dec (deg)	[Fe/H]*	[Fe/H] (dex)	$\sigma$ [Fe/H] (dex)
IC 2714	169.369	-62.734	ARC	-0.254	0.199
			ZGR	-0.291	0.183
			FS	-0.112	0.081
			None	0.256	-
			HRS	0.02	-
IC 4756	279.627	5.437	ARC	-0.17	0.129
			ZGR	-0.13	0.097
			FS	-0.057	0.069
			None	0.374	-
			HRS	0	-
Melotte 71	114.377	-12.058	ARC	-0.334	0.178
			ZGR	-0.137	0.157
			FS	-0.117	0.09
			None	0.093	-
			HRS	-0.27	-
...	...	...	...	...	...

**Notes.** [Fe/H]\* indicates if the metallicity was derived from a Gaia XP-based catalogue (ARC, ZGR, FS), or adopted as a HRS value from [Bossini et al. \(2019\)](#). ‘None’ indicates that [Fe/H] was a free fit parameter, and the 50th percentile of its PDF is listed. The entire table is available at the CDS.

Table B.2: Cluster parameters  $A_V$ ,  $dm$  and  $\log(age)$  derived using the specified [Fe/H]\* prior assumption.

Cluster	[Fe/H]*	$A_V$ (mag)	$A_{V16}$ (mag)	$A_{V84}$ (mag)	$dm$ (mag)	$dm_{16}$ (mag)	$dm_{84}$ (mag)	$\log age$ (dex)	$\log age_{16}$ (dex)	$\log age_{84}$ (dex)
IC 2714	ARC	1.182	1.146	1.23	10.31	10.218	10.368	8.678	8.581	8.707
	ZGR	1.211	1.148	1.253	10.307	10.238	10.341	8.661	8.448	8.720
	FS	1.188	1.145	1.232	10.324	10.286	10.396	8.674	8.473	8.704
	None	1.075	0.983	1.191	10.577	10.475	10.656	8.671	7.859	8.729
	HRS	1.029	1.008	1.15	10.318	10.298	10.375	8.657	8.656	8.659
IC 4756	ARC	0.953	0.716	1.1	8.267	8.152	8.43	8.770	8.766	9.033
	ZGR	0.784	0.667	0.996	8.2	8.135	8.295	9.031	8.767	9.034
	FS	0.617	0.559	0.648	8.226	8.194	8.254	9.013	8.966	9.014
	None	0.396	0.302	0.472	8.441	8.332	8.511	8.946	8.917	8.994
	HRS	0.566	0.539	0.611	8.231	8.209	8.249	9.003	9.001	9.004
Melotte 71	ARC	0.514	0.471	0.55	11.486	11.404	11.529	9.131	9.107	9.152
	ZGR	0.468	0.415	0.496	11.576	11.533	11.609	9.088	9.069	9.099
	FS	0.468	0.448	0.493	11.568	11.528	11.604	9.088	9.075	9.094
	None	0.437	0.326	0.484	11.702	11.56	11.82	8.994	8.967	9.083
	HRS	0.537	0.523	0.547	11.472	11.45	11.539	9.138	9.136	9.141
...	...	...	...	...	...	...	...	...	...	...

**Notes.** The [Fe/H]\* notation is defined as in Table B.1. The 50th, 16th and 84th percentile of the PDF for each cluster parameter is reported. The entire table is available at the CDS.

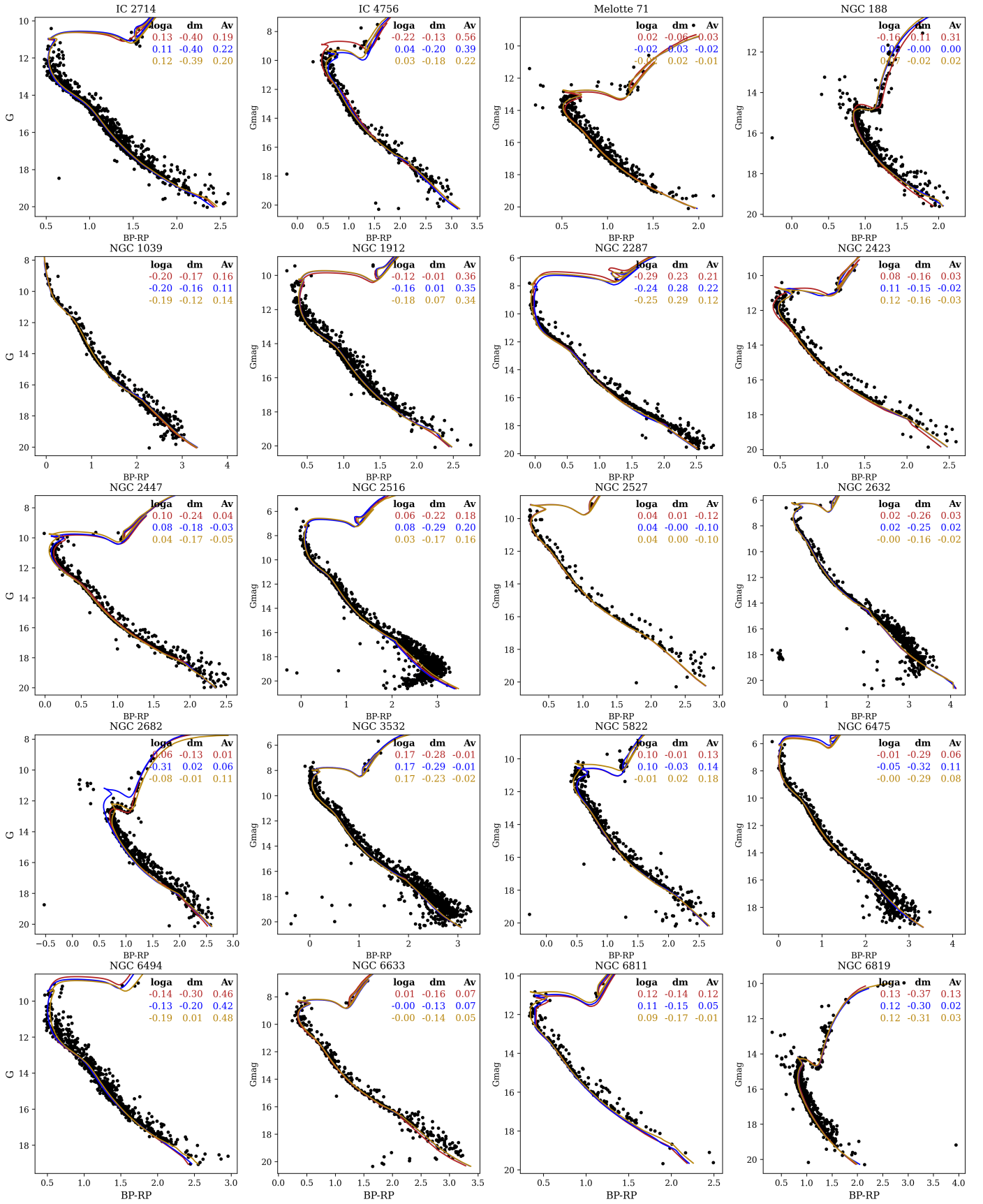


Fig. B.1: CMDs of the sample of OCs with photometric data from Gaia. The isochrones match the cluster parameters derived by using  $[\text{Fe}/\text{H}]$  of ZGR (blue), ARC (red), and FS (yellow) catalogues; the depicted values show the difference between the parameters (columns) estimated by using the respective prior and B19 values of the same parameter.

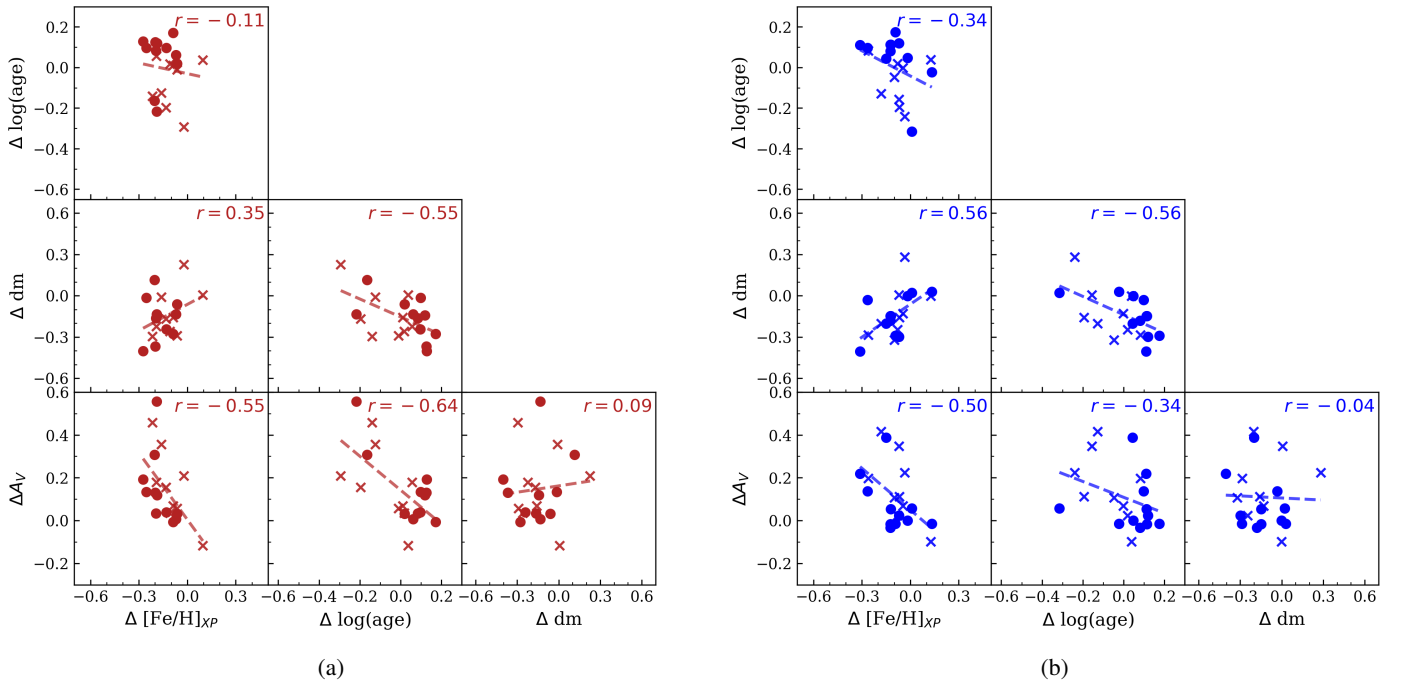


Fig. B.2: Same as Figure 8 but for (a) ARC catalogue and (b) ZGR catalogue.

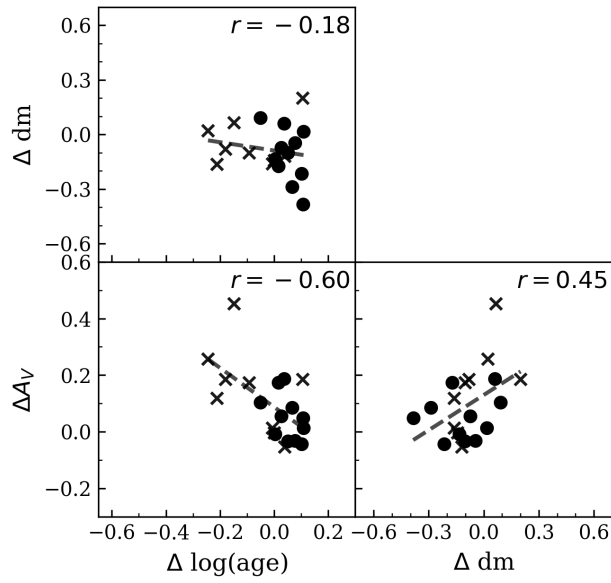


Fig. B.3: Same as Figure 8 but for solutions with  $[\text{Fe}/\text{H}]$  fixed to the HRS values from B19. The  $[\text{Fe}/\text{H}]$  offset is not displayed, as it is zero.