A&A, 709, A51 (2026)
https://doi.org/10.1051/0004-6361/202659713

Astronomy & Astrophysics

A critical analysis of main-sequence fitting in open clusters to derive the helium-to-metal enrichment ratio $\Delta Y/\Delta Z$

G. Valle[1,2,⋆], N. Ricci[1], M. Dell'Omodarme[1], P. G. Prada Moroni[1,2], S. Degl'Innocenti[1,2], and S. Cassisi[2,3]

[1] Dipartimento di Fisica "Enrico Fermi", Università di Pisa, Largo Pontecorvo 3, I-56127 Pisa, Italy
[2] INFN, Sezione di Pisa, Largo Pontecorvo 3, I-56127 Pisa, Italy
[3] INAF-Osservatorio Astronomico d'Abruzzo, Via Mentore Maggini s.n.c., 64100 Teramo, Italy

Received 5 March 2026 / Accepted 23 March 2026

ABSTRACT

Aims. We aim to investigate the feasibility of accurately determining the helium-to-metal enrichment ratio $\Delta Y/\Delta Z$ for open clusters using Gaia DR3 photometry.
Methods. To test the reliability of this calibration, we performed a theoretical investigation using mock open clusters. We generated synthetic photometric data from isochrones calculated by five different stellar evolution codes (FRANEC, PARSEC 1.2s, PARSEC 2.0, BASTI, and MIST), for which the true $\Delta Y/\Delta Z$ is known. We then fitted these mock clusters with two sets of isochrones calculated with the FRANEC code, differing only in the implementation of bolometric corrections (BCs). The analysis focused on the G-band absolute magnitude range (4.3–6.5 mag) to minimise the impact of poorly constrained physics. Synthetic clusters were generated at [Fe/H] values from 0.0 to 0.15 dex, for different numbers of populating stars and different levels of photometric uncertainties.
Results. The Monte Carlo experiments revealed significant and code-dependent biases. Unbiased results were achieved only when the stellar models used for synthetic-cluster generation and fitting were identical. Using identical FRANEC stellar models but different BCs introduced a significant bias of up to 0.6. Furthermore, using different stellar models for synthetic cluster generations resulted in even larger biases: $\Delta Y/\Delta Z$ was underestimated by up to 0.8 for PARSEC target isochrones, while it was overestimated for BASTI and MIST isochrones by up to 0.6 and 1.5, respectively.
Conclusions. The magnitude and the inconsistency of these biases strongly suggest that the photometric calibration of $\Delta Y/\Delta Z$ using open clusters is not reliably robust.

Key words. methods: statistical – stars: abundances – stars: evolution – stars: fundamental parameters – stars: interiors

1. Introduction

The effort to calibrate stellar models has achieved significant results in improving the accuracy of predictions compared with observations. Nevertheless, several physical phenomena – such as convective transport, microscopic diffusion, and competing processes – are still affected by notable uncertainties (see e.g. Viallet et al. 2015; Moedas et al. 2022). Moreover, the assumed chemical composition plays a non-negligible role in the overall uncertainty of stellar model predictions. While absorption lines in stellar spectra allow for a robust determination of the surface metallicity, direct measurements of helium abundance are generally infeasible for stars cooler than approximately 15 000 K due to the lack of observable helium spectral lines. Therefore, stellar models must rely on an assumed initial helium abundance. A common approach is to employ a linear relationship between initial helium abundance, Y, and metallicity, Z:

$$Y = Y_p + \frac{\Delta Y}{\Delta Z} Z, \tag{1}$$

where Y_p is the primordial helium abundance produced in the Big Bang nucleosynthesis and $\Delta Y/\Delta Z$ is the helium-to-metal enrichment ratio.

The precise value of the helium-to-metal enrichment ratio has been extensively studied. A variety of techniques have been deployed to constrain this parameter, including comparing theoretical isochrones with observational data in the Hertzsprung–Russell (HR) diagram (e.g. Pagel & Portinari 1998; Casagrande et al. 2007; Gennaro et al. 2010; Tognelli et al. 2021); fitting evolutionary tracks to a census of nearby field stars (e.g. Jimenez et al. 2003; Valcarce et al. 2013; Ricci et al. 2025); utilising asteroseismic data (e.g. Silva Aguirre et al. 2017; Verma et al. 2019; Nsamba et al. 2021); calibrating from detached, double-lined eclipsing binaries (Ribas et al. 2000; Fernandes et al. 2012; Valle et al. 2024); developing a standard solar model that accurately replicates the Sun's current luminosity, radius, and surface Z/X ratio (e.g Bahcall et al. 1995; Serenelli 2010; Valcarce et al. 2012; Vinyoles et al. 2017; Magg et al. 2022; Buldgen et al. 2025); calibrating stellar models against evolved stars, specifically horizontal branch and red giant stars (Renzini 1994; Marino et al. 2014; Valcarce et al. 2016); and leveraging the properties of galactic and extragalactic H II regions (e.g. Peimbert & Torres-Peimbert 1974; Pagel et al. 1992; Chiappini & Maciel 1994; Peimbert et al. 2000; Fukugita & Kawasaki 2006; Méndez-Delgado et al. 2020; Kurichin et al. 2021) or planetary nebulae (D'Odorico et al. 1976; Chiappini & Maciel 1994; Peimbert & Serrano 1980; Maciel 2001). The results derived from these different methodologies exhibit significant scatter and are susceptible to systematic biases. As an illustration, standard solar models commonly yield $\Delta Y/\Delta Z$ values near to 1, while analyses focusing on evolved stars suggest a range from 2 to 3. Conversely,

⋆ Corresponding author: valle@df.unipi.it

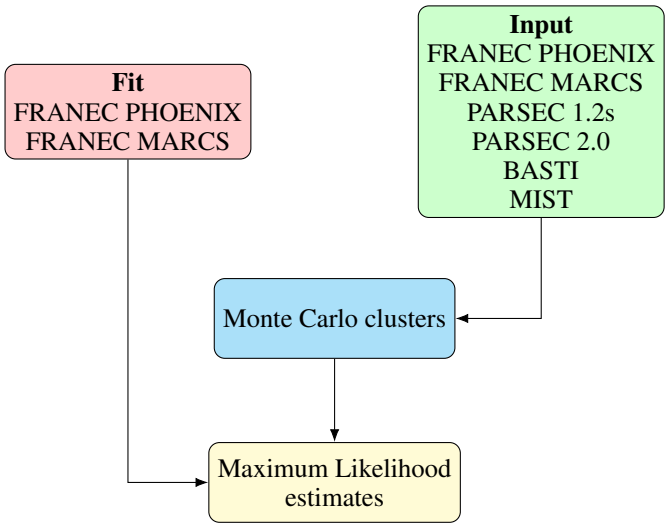

Fig. 1. Block diagram of the adopted framework.

comparisons involving main-sequence (MS) binary stars and the HR diagram generally point towards values spanning 1–3, while fitting MS field stars suggests plausible values from 2 to 5.

We recently investigated the robustness and reliability of calibration using detached eclipsing binaries and MS field stars (Valle et al. 2024; Ricci et al. 2025). The results indicated the presence of significant systematic errors that undermine the use of these objects for $\Delta Y/\Delta Z$ calibration. In this paper, we aim to investigate the reliability of estimates attainable from fitting open-cluster MS using a set of isochrones, specifically adopting Gaia Data Release 3 (DR3; Gaia Collaboration 2021) photometry. We particularly focused on results from young (100–800 Myr) and metal-rich ([Fe/H] $>$ 0) clusters, as this age and metallicity range contain near, well-populated open clusters such as the Pleiades, Hyades, Alpha Persei, and Praesepe (e.g. Dahm 2015; Brandt & Huang 2015; Martín et al. 2018; Gossage et al. 2018). The major challenge in this calibration is the well-known degeneracy in the photometric space among global metallicity Z, initial helium abundance Y, and age (see among many Chaboyer et al. 1992; Pagel & Portinari 1998; Castellani et al. 1999; Lebreton et al. 2001; Pinsonneault et al. 2003; Jimenez et al. 2004; An et al. 2007; Casagrande et al. 2007).

This degeneracy makes the effect on the isochrones of a variation of $\Delta Y/\Delta Z$ almost indistinguishable from the one due to the metallicity ([Fe/H]) change, within the limits allowed by observational uncertainty. A primary objective of our investigation is to precisely determine if and how well the exquisite precision of Gaia DR3 photometry and astrometry allows us to lift this degeneracy.

To explore the feasibility of this type of calibration, we chose an approach that offers the best sensitivity. Instead of working directly with observational data, we performed a theoretical investigation of the maximum achievable performance using mock data. This approach offers several advantages, mainly allowing for the firm identification of possible hidden biases and systematic sources of uncertainty. Unavoidable discrepancies exist between theoretical isochrones and actual cluster data. These arise from two main sources: modelling imperfections and observational contamination. Modelling imperfections arise from missing or incomplete implementation of physical processes in stellar model computations. A related factor is the non-negligible variability that persists in stellar model computations due to the freedom that stellar modelers have in adopting different input physics within their allowed uncertainties (see e.g. Valle et al. 2013; Stancliffe et al. 2016). Observational contamination occurs because selecting a sample of MS single stars from observed clusters is a process plagued by the presence of unresolved binaries, peculiar objects, and field stars that blur the cluster MS passing all selection steps. These problems become more relevant for high-mass MS stars and near the turn-off. In these regions, physical phenomena not yet robustly accounted for in stellar models, such as convective core overshooting or rotation, strongly affect the isochrone morphology. Moreover, these zones are significantly less populated than the low MS. This makes the identification of peculiar objects and unresolved binaries more difficult because the effectiveness of techniques such as iterative sigma clipping – used to identify bona fide single stars (see e.g. de Marchi & Pulone 2007; Valle et al. 2021; Brandner et al. 2023a, 2024) – strongly depends on having a well-populated colour-magnitude diagram throughout all its parts.

2. Methods

2.1. Framework

To explore the feasibility of reconstructing the $\Delta Y/\Delta Z$ value from mock data, we implemented a structured pipeline. This framework comprises several key components: a grid of fitting models used to estimate the best $\Delta Y/\Delta Z$ value from the synthetic cluster data; a set of isochrones generated from different stellar evolutionary codes; a Monte Carlo procedure designed to generate the synthetic clusters from these isochrones; and, finally, a fitting procedure to obtain the maximum-likelihood estimates of [Fe/H] and $\Delta Y/\Delta Z$. A block diagram illustrating this entire framework is provided in Fig. 1. The adoption of target isochrones from various stellar evolutionary codes allowed us to investigate the existence of possible systematic effects in the $\Delta Y/\Delta Z$ estimation within a controlled environment. Discrepancies between stellar models and observations are, in fact, to be expected. Therefore, exploring the impact of these differences – especially when they originate simply from variations in the morphology of theoretical isochrones – is of fundamental importance.

Since the various evolutionary sequences along the isochrone are affected by distinct sources of systematic uncertainty, the first step is to identify a portion of the MS suitable for our investigation. We aimed for a region whose morphology is highly consistent among different stellar evolutionary codes and minimally influenced by age changes. Following Tognelli et al. (2021), we selected a portion of the isochrones that is most sensitive to changes in $\Delta Y/\Delta Z$. Figure 2 illustrates this by showing a set of isochrones from our fitting grids (FRANEC isochrones) at [Fe/H] = 0.1 dex, $\Delta Y/\Delta Z$ = 2.0, and three ages representative of the range considered in this paper. For comparison, the corresponding PARSEC 1.2s isochrones (Bressan et al. 2012) are also shown. The figure demonstrates the impact of age in the high MS region, which begins to manifest itself clearly for $G \lesssim 3.5$ mag. A direct comparison of isochrones at 750 Myr and 100 Myr reveals that, at $G = 4.3$ mag, the difference in the $BP-RP$ colour due to age is 0.007 mag for both the FRANEC and PARSEC sets of isochrones. Moving to lower magnitudes causes the impact of age to decrease to a minimum value of $\Delta(BP-RP) = 0.004$ mag at $G = 5.0$ mag and then increase again, reaching 0.011 mag at $G = 6.5$ mag. Similar trends are observed at different metallicities. Assuming the expected observational uncertainties from

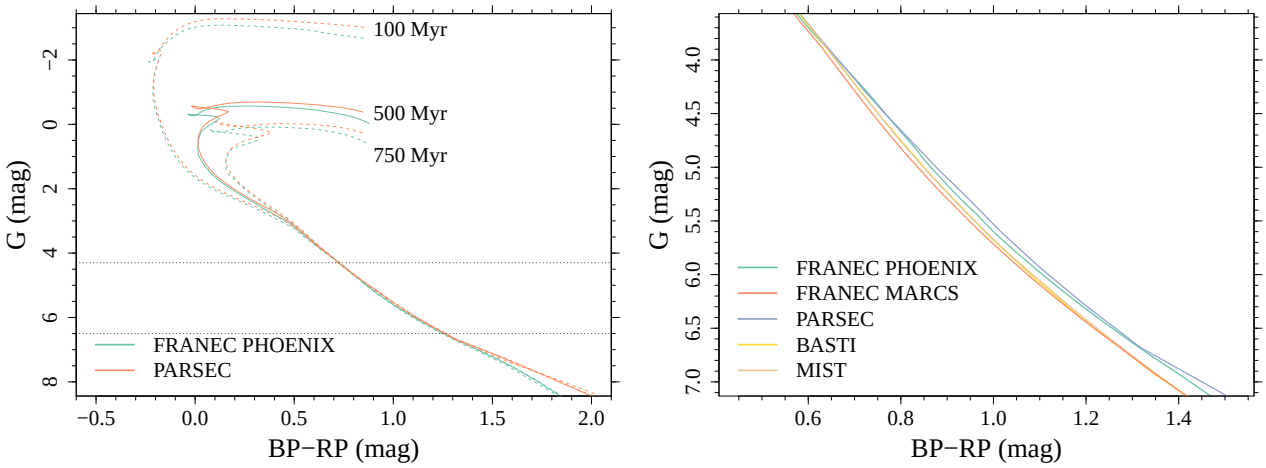

Fig. 2. Comparison of isochrones from different stellar evolutionary codes. *Left*: PARSEC and FRANEC PHOENIX isochrones at [Fe/H] = 0.1 dex and different ages in the Gaia DR3 colour-magnitude diagram. The dotted horizontal lines mark the edges of the selected zone. FRANEC isochrones are computed with $\Delta Y/\Delta Z = 2.0$. *Right*: Comparison of isochrones from the FRANEC, PARSEC, BASTI, and MIST codes, at [Fe/H] = 0.1 dex, in the region selected for the analysis.

Gaia photometry range from 0.005 mag to 0.02 mag (depending on data quality, proximity, and extinction), we safely adopted the range of $G = (4.3,6.5)$ mag for the following analyses. Enlarging this range is not advisable for theoretical reasons. Specifically, there are well-known discrepancies between isochrones and data at magnitudes below our lower edge (Brandner et al. 2023a,b; Wang et al. 2025; Ricci et al. 2025). Conversely, stars brighter than the selected upper edge are potentially affected by physical phenomena, such as rotation and convective-core overshooting, whose implementation in stellar evolutionary codes is still an area of active research. Figure 2 shows the differences in the near-turn-off region, where the isochrones from the two stellar evolutionary codes diverge, mainly due to differences in the implementation and efficiency of the convective core overshooting mechanism.

The right panel of Fig. 2 presents a comparison of the corresponding isochrones derived from the different stellar evolutionary codes selected for this analysis; that is, PARSEC 1.2s, PARSEC 2.0, BASTI, and MIST (Bressan et al. 2012; Nguyen et al. 2025; Hidalgo et al. 2018; Choi et al. 2016), as discussed in Sect. 2.3. The non-negligible spread among the isochrones is immediately apparent, highlighting the relevant impact of adopting different atmosphere models in the computation of the bolometric corrections (BCs). In fact, the difference observed between the two FRANEC isochrones shown here is solely attributable to the choice of alternative BC tables (see Sect. 2.2).

2.2. Stellar model grid for fitting

The grid of stellar evolutionary models was calculated for the 0.50 to 4.00 $M_\odot$ mass range with a step of 0.025 $M_\odot$, spanning the evolutionary stages from the pre-MS to the onset of the red giant branch (RGB). The initial metallicity [Fe/H] ranged from −0.5 dex to 0.35 dex with a step of 0.01 dex. We adopted the solar heavy-element mixture by Asplund et al. (2009). For each metallicity, we considered a range of initial helium abundances based on the commonly used linear relation in Eq. (1), with the primordial helium abundance, $Y_p = 0.2471 \pm 0.001$, from Planck Collaboration VI (2020). The helium-to-metal enrichment ratio, $\Delta Y/\Delta Z$, was varied from 0.4 to 3.2 with a step of 0.2.

The models were computed with the FRANEC code in the same configuration previously adopted to compute the Pisa Stellar Evolution Data Base[1] for low-mass stars (Dell'Omodarme et al. 2012). The outer boundary conditions were established by the solar semi-empirical $T(\tau)$ of Vernazza et al. (1981), which aptly approximates the results obtained using the hydro-calibrated $T(\tau)$ (Salaris & Cassisi 2015; Salaris et al. 2018). The models were computed assuming a mixing-length parameter of $\alpha_{\rm ml} = 2.02$, which was calibrated by computing the solar standard model. Atomic diffusion was included by adopting the coefficients given by Thoul et al. (1994) for gravitational settling and thermal diffusion. To prevent extreme variations in surface chemical abundances, the diffusion velocities were multiplied by a suppression parabolic factor that is equivalent to one for 99% of the mass of the structure and zero at the base of the atmosphere (Chaboyer et al. 2001).

The raw stellar evolutionary tracks were reduced to a set of isochrones spanning an age range from 100 Myr to 750 Myr, which is the estimated age range of nearby open clusters. BCs used to derive the Gaia magnitudes were obtained from both the PHOENIX2011 grid (Allard et al. 2011) and the MARCS grid (Gustafsson et al. 2008).

2.3. Stellar models for the synthetic clusters construction

To explore the accuracy and precision attainable in a controlled framework, we used several models to build synthetic clusters. We selected isochrones at a fixed age of 500 Myr and metallicities of [Fe/H] = 0.00, 0.05, 0.10, and 0.15 from PARSEC 1.2s (Bressan et al. 2012), PARSEC 2.0 (Nguyen et al. 2025), BASTI (Hidalgo et al. 2018), and MIST (Choi et al. 2016). Isochrones at the desired ages and metallicities were obtained using the respective web interpolator tools. For PARSEC 2.0 and MIST, we selected models without rotation ($\omega/\omega_c = 0$). For BASTI, we used models computed considering convective core overshooting, but without microscopic diffusion. The target pool also includes the FRANEC MARCS and PHOENIX models, which were sampled at $\Delta Y/\Delta Z = 1.8$.

[1] http://astro.df.unipi.it/stellar-models/

Table 1. Relevant input parameters for the stellar evolution codes adopted in this analysis.

Case	Models	Mixture	Y_p	$\Delta Y/\Delta Z$	BC
A	FRANEC	Photospheric Asplund 2009	0.247	1.8	MARCS[a]
B	FRANEC	Photospheric Asplund 2009	0.247	1.8	PHOENIX[b]
C,D	PARSEC 1.2s	Caffau 2011	0.248	1.78	ATLAS9 ODFNEW, PHOENIX BT-Settl[c]
E	PARSEC 2.0	Caffau 2011	0.248	1.78	ATLAS9 ODFNEW, PHOENIX BT-Settl[c]
F	BASTI	Caffau 2011	0.247	1.31	ATLAS9 ODFNEW, PHOENIX BT-Settl[c]
G	MIST 1.2	Proto-solar Asplund 2009	0.249	1.5	C3K[d]

Notes. (a) Gustafsson et al. (2008); (b) Allard et al. (2011); (c) Castelli & Kurucz (2003), Allard et al. (2012); (d) Conroy et al. (2018).

These target isochrones differ in some key aspects relevant to our investigation. First, they adopt different solar reference mixtures: MIST uses the proto-solar mixture from Asplund et al. (2009), while both the PARSEC and BASTI models employ the mixture from Caffau et al. (2011). Consequently, the assumed $(Z/X)_\odot$ value varies across the different sets of isochrones. Second, the reference values of $\Delta Y/\Delta Z$ used to compute the models vary significantly. While PARSEC 1.2s and 2.0 adopt $\Delta Y/\Delta Z$ = 1.78, BASTI uses $\Delta Y/\Delta Z$ = 1.31 and MIST uses $\Delta Y/\Delta Z$ = 1.5. The differences in these inputs affect the Z-to-[Fe/H] relations. For example, [Fe/H] = 0.0 corresponds to Z values of 0.01471, 0.01492, and 0.01429 for PARSEC, BASTI, and MIST, respectively. For reference, the Z range corresponding to [Fe/H] = 0.0 of the fitting grid spans the range (0.01266, 0.01329), depending on the value of $\Delta Y/\Delta Z$. Finally, the primordial helium abundance values assumed by the different set of isochrones are almost identical (see Table 1); this key point is further discussed in Sect. 2.5.

Since the comparison is performed in the Gaia DR3 magnitude space, another relevant difference relates to the BCs adopted to compute the synthetic photometry. This input plays a key role in isochrone fitting, as shown by Valle et al. (2021), causing major systematic differences among the calculated magnitudes. The right panel in Fig. 2 clearly shows the relevance of this input, which causes a large shift between identical FRANEC models. Regarding the significance of this parameter, the displacement in $BP - RP$, resulting from the use of PHOENIX BCs rather than MARCS, roughly corresponds to a difference of 0.15 dex in [Fe/H]. Given this critical importance of the BCs, in the following analysis, we preferred not to anchor the fitting in the target isochrone's [Fe/H] value. Instead, we allowed [Fe/H] to vary as a supplementary fitting parameter. Table 1 summarises the key inputs adopted by the selected stellar models.

2.4. Monte Carlo synthetic clusters

For all the target isochrones discussed above, we generated synthetic clusters by means of Monte Carlo simulations. The simulations were performed assuming three different numbers of stars (N = 50, 100, and 150) – representing the populations expected within the adopted magnitude range of nearby open clusters – and three different levels of photometric uncertainty (σ = 0.005, 0.01, and 0.02 mag). Each experiment was repeated ten times, and the estimated parameters were obtained by the mean of the resulting samples. Therefore, the pipeline steps are as follows: (1) selection of a target isochrone at a given [Fe/H]; (2) random generation of N synthetic stars in the chosen range in the Gaia DR3 magnitudes from this isochrone; (3) random perturbation of the synthetic data adopting Gaussian random errors with a standard deviation of σ. This intrinsically assumes a synthetic cluster where all the stars to be fitted are effectively single stars, thus implying a perfect rejection of unresolved binaries, field stars, and peculiar objects; (4) fitting of the data using the reference set of isochrones; (5) repetition of steps 1–4 ten times to reduce random errors[2]. The repeated experiments were summarised by computing the mean values of the fitted $\Delta Y/\Delta Z$ and [Fe/H].

Overall, we considered the cases listed below.

- Case A: Both sampling and reconstruction adopt the same FRANEC MARCS grid. This case serves as our reference and provides the maximum achievable performance in the absence of any systematic discrepancies between synthetic data and models.
- Case B: Sampling from the FRANEC PHOENIX grid and reconstruction over the FRANEC MARCS grid. This scenario tests the importance of a change in BCs only, since the underlying stellar structure computations are identical.
- Case C: Sampling from PARSEC 1.2s isochrones and reconstruction over the FRANEC MARCS grid.
- Case D: Sampling from PARSEC 1.2s isochrones and reconstruction over the FRANEC PHOENIX grid. Comparing Cases C and D allowed us to further investigate the importance of BCs when a systematic discrepancy exists between the synthetic data and the fitting isochrones.
- Case E: Sampling from PARSEC 2.0 isochrones and reconstruction over the FRANEC MARCS grid.
- Case F: Sampling from BASTI isochrones and reconstruction over the FRANEC MARCS grid.
- Case G: Sampling from MIST isochrones and reconstruction over the FRANEC MARCS grid.

2.5. Fitting technique

To perform the stellar fit, we utilised a grid of isochrones computed over a grid of given a enrichment ratio, $\Delta Y/\Delta Z$, and [Fe/H]. The approach is mathematically equivalent to the standard observational approach of independently fitting Y and Z. In the classical framework, the enrichment ratio is typically derived a posteriori by performing a linear regression on the best-fit (Z, Y) pairs obtained for a sample of clusters. By contrast, our method explores the parameter space treating $\Delta Y/\Delta Z$ as a fundamental input of the grid. However, the validity of this equivalence strictly depends on the assumption that the primordial helium abundance, Y_p, is consistent across the different stellar evolution libraries. In fact any offset in the adopted Y_p between the synthetic cluster and the reference grid would introduce a systematic bias in the inferred $\Delta Y/\Delta Z$. Given that the variation of Y_p among the adopted isochrone libraries is only 0.002 (Table 1), the hypothesis of constant initial helium abundance is robust for the purposes of this analysis.

[2] We verified that adopting more repetitions did not modify the results.

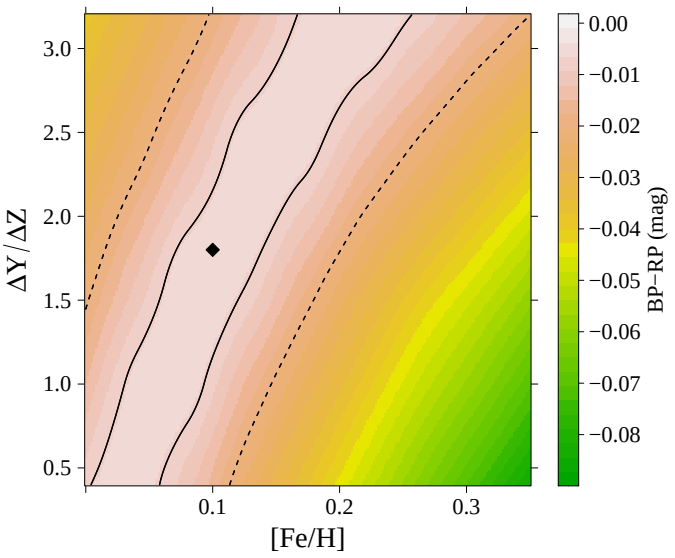

Fig. 3. Filled contour plot of the mean absolute differences in the colour $BP - RP$ between the reference isochrone with [Fe/H] = 0.1 and $\Delta Y/\Delta Z = 1.8$, and all other isochrones in the grid (for the FRANEC MARCS set). The reference model is identified by the black diamond. Solid and dashed black lines indicate regions where the mean absolute difference is below 0.007 mag and 0.02 mag, respectively.

We adopted the fitting technique described in Valle et al. (2021), which relies on the computation of the Mahalanobis distances between observed data and isochrones. Briefly, we consider a data set consisting of N observations and their associated observational uncertainties. For every j-th ($j = 1, \ldots, J$) isochrone in the fitting set, the sum χ^2_j of the squared Mahalanobis distances, d, between data and isochrones is obtained:

$$\chi^2_j = \sum_{i=1}^{N} d_i^2. \tag{2}$$

The likelihood of the j-th isochrone is then given by

$$\mathcal{L}_j = \exp(-\chi^2_j/2). \tag{3}$$

Best-fit values of the metallicity [Fe/H] and $\Delta Y/\Delta Z$ are then computed by a weighted average. Let us define θ_j = (age, [Fe/H], $\Delta Y/\Delta Z$) as the vector of meta-parameters characterising the isochrone. The best-fitting values are then

$$\tilde{\theta} = \frac{\sum_{j=1}^{J} \theta_j \mathcal{L}_j}{\sum_{j=1}^{J} \mathcal{L}_j}. \tag{4}$$

The goodness of fit was assessed by means of an χ^2 test, as described in Valle et al. (2021). For each of the seven considered cases, a Bonferroni correction (Abdi 2007) was adopted to protect against Type I errors. Globally, 12 out of 2520 Monte Carlo simulations ended with no isochrones passing the goodness-of-fit test at level $\sigma = 0.05$; one passed for case D; two passed for cases B, E, F, and G; and three passed for case C. These simulations were removed from the pool when computing the best-fit values.

3. Estimated $\Delta Y/\Delta Z$

3.1. Sensitivity of the fitting process

As already discussed, the calibration of $\Delta Y/\Delta Z$ is intrinsically a difficult exercise due to the counteracting effect of metallicity and initial helium abundance on the isochrone. This creates a well-known degeneracy. To quantify the relevance of this degeneracy for our purpose, we performed an evaluation over the grid of the differences among the isochrones. We adopted the FRANEC MARCS isochrone with an age of 500 Myr, [Fe/H] = 0.1, and $\Delta Y/\Delta Z = 1.8$ as our references. We then evaluated the mean absolute difference in the colour $BP - RP$ between this reference and all other isochrones in the grid. This comparison was performed over the magnitude range selected for the analysis, spanning from $G = 6.5$ mag to $G = 4.3$ mag. The result of this exercise is shown in Fig. 3 as a function of [Fe/H] and $\Delta Y/\Delta Z$. The figure highlights some important points relevant for interpreting the results of this section. First, there is a large region in the metallicity–helium plane, where the difference among isochrones is lower than or comparable to the expected observational uncertainty of 0.007 mag. Second, the boundary values chosen for $\Delta Y/\Delta Z$ in the grid-building process have a fundamental and direct impact on the results. In fact, the figure shows that the zone with low difference extends to both the upper and the lower edges of $\Delta Y/\Delta Z$. Third, even a quite precise determination of [Fe/H], say at 0.05 dex, does not significantly restrict the range of allowed $\Delta Y/\Delta Z$ values.

3.2. Estimated $\Delta Y/\Delta Z$ and dependence on the target [Fe/H]

The results of the Monte Carlo simulations confirm that the calibration of $\Delta Y/\Delta Z$ is prone to severe and uncontrolled biases. Figure 4 shows the results of the experiments, grouped according to the metallicity of the isochrone used for the synthetic-cluster generation. Case A (FRANEC MARCS isochrones for both target and fit) shows that the technique is in principle unbiased in the absence of systematic discrepancies between the data and the models adopted in the fit. The recovered $\Delta Y/\Delta Z$ values are consistent with the value of $\Delta Y/\Delta Z = 1.8$ adopted for the target isochrone, with a remarkable precision of about ±0.2, independent of the target's metallicity. For all other cases, the results show a substantial bias. Even in Case B, where the difference between data and models is limited to the BC adopted (PHOENIX for the target and MARCS for the models), the inferred $\Delta Y/\Delta Z$ value is biased towards lower values, and the effect increases with the metallicity. At [Fe/H] = 0.15, the $\Delta Y/\Delta Z$ value is underestimated by as much as 0.6.

The bias is even more severe when the models underlying the synthetic data come from a different stellar evolutionary code. Cases C and D show that the $\Delta Y/\Delta Z$ value estimated when using PARSEC 1.2s isochrones to construct synthetic clusters is largely underestimated when FRANEC models are used for the fit, by about 0.8 for Case C. In agreement with the comparison of Cases A and B, Cases C and D show that the bias is reduced – by about 0.15 – when FRANEC PHOENIX models are adopted in the fit. Results from Case E show that the difference between PARSEC 1.2s and 2.0 is quite small up to [Fe/H] = 0.05 dex, but it then substantially increases: at [Fe/H] = 0.15, the $\Delta Y/\Delta Z$ value is underestimated by about 1.2.

The use of BASTI isochrones as targets (Case F) shows an opposite behaviour, as FRANEC MARCS models substantially overestimated the $\Delta Y/\Delta Z$ target value. The overestimation is almost constant, with a positive bias of about 1.3 across the metallicity range and over the target value of 1.3. Finally, Case G, with synthetic clusters built from MIST isochrones, shows a tendency toward overestimation again, even if it is less severe than Case F. In this case, the estimated value of $\Delta Y/\Delta Z$ increases from 1.9 to 2.2 with increasing metallicity, resulting in a mean overestimation of about 0.6 over the target value of 1.5.

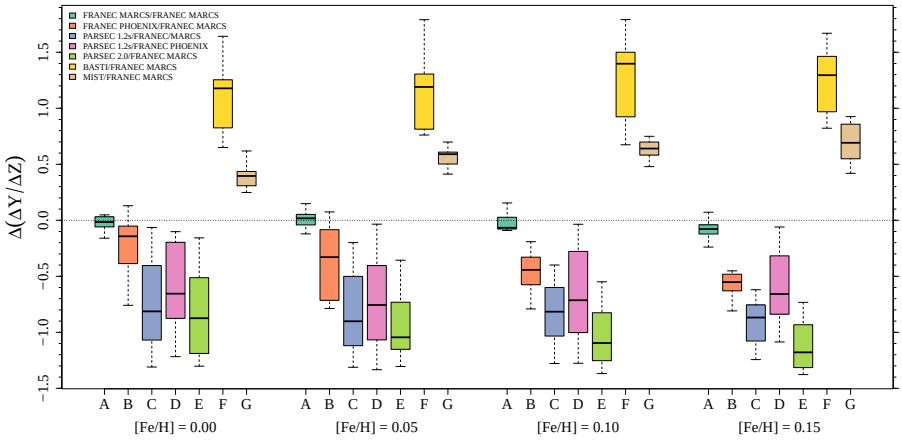

Fig. 4. Differences in the estimated values of $\Delta Y/\Delta Z$ for the seven different considered cases with respect to the values adopted by the generating $\Delta Y/\Delta Z$, according to different [Fe/H] values of the target isochrone.

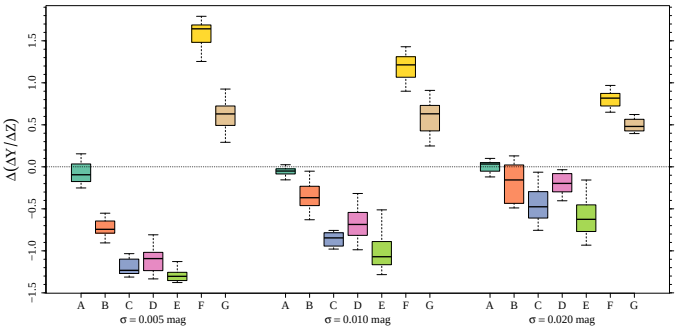

Fig. 5. Same as in Fig. 4, but according to the different adopted photometric errors.

3.3. Dependence on the other simulation parameters

The dependence on the assumed photometric errors is shown in Fig. 5. It reveals an interesting behaviour that warrants further discussion. As shown in the figure, increasing the photometric errors adopted in both the synthetic cluster generation and the fitting procedure reduces the bias between the target and the estimated $\Delta Y/\Delta Z$. However, this reduction occurs by chance and is simply the effect of a regression towards the mean value available in the grid. It is likely that the larger photometric errors typical of past observations acted as a statistical smoothing factor, effectively masking the systematic biases that we identified here. Figure 6 helps us to understand what is happening. This figure is analogous to Fig. 3, which is discussed in Sect. 3.1, but the target isochrone comes from the BASTI data set. The differences were computed with respect to the FRANEC MARCS grid. In this case, the lowest discrepancy region is in the upper part of the allowed $\Delta Y/\Delta Z$ range; therefore, the tendency of Case F fits, discussed above, to prefer $\Delta Y/\Delta Z$ values larger than 2.0 is understandable. When the photometric errors are at the 0.007 mag level, isochrones in this upper region of the [Fe/H] – $\Delta Y/\Delta Z$ plane would provide an acceptable fit. However, when larger photometric errors are allowed, say 0.02 mag, a greater portion of the parameter space must be considered in the fit. As a result, lower values of $\Delta Y/\Delta Z$ become compatible with the data, and the estimated values regress towards the mean $\Delta Y/\Delta Z$ value represented in the fitting grid, which is $\Delta Y/\Delta Z = 1.8$. This behaviour, with increasing observational uncertainty, is not surprising. A similar regression towards the mean value in the grid was previously reported by Valle et al. (2024) in their investigation into the possibility of constraining the $\Delta Y/\Delta Z$ value from MS binary stars.

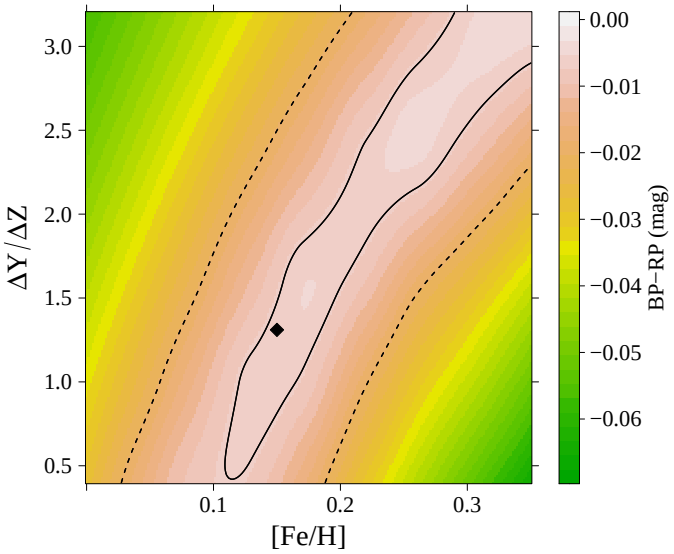

Fig. 6. Same as in Fig. 3, but adopting the BASTI isochrone with [Fe/H] = 0.15 dex as the target.

Regarding the dependence on the sample size, the results align with expectations. The systematic biases generally increase with sample size, with a magnitude that depends on the specific case considered. In fact, increasing the sample size of the synthetic clusters from 50 to 150 objects limits the impact of random fluctuations, enhances the possibility of detecting the differences in the morphology of the underlying isochrones, and restricts the pool of good fitting isochrones.

The increase in the bias ranges from 0.25 to 0.40 for PARSEC isochrones (Cases C–E), and it is about 0.35 for BASTI (Case F). The effect is more modest (about 0.1) for the comparison between FRANEC models with different BSs and the comparison with MIST models.

3.4. Restricting to the target [Fe/H]

When fitting real stellar clusters, the metallicity [Fe/H] is routinely adopted among the constraints. However, we conducted our simulations lifting this specific constraint, as discussed in

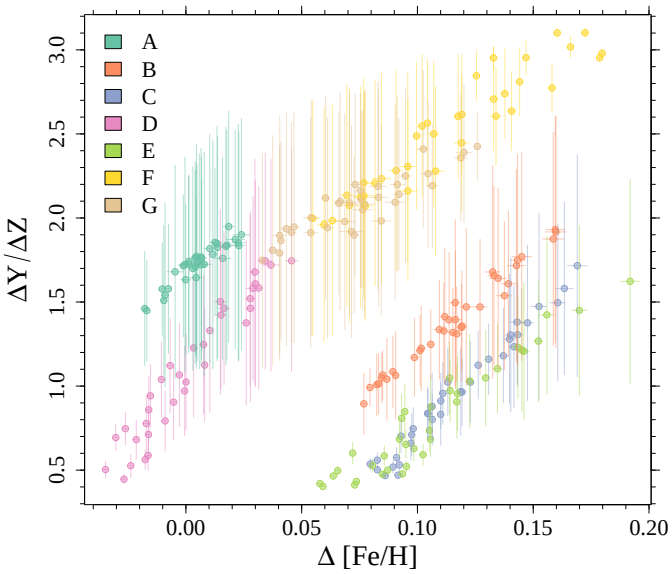

Fig. 7. Scatter plot of the fitted [Fe/H] and $\Delta Y/\Delta Z$ values. The plotted [Fe/H] are the differences between fitted and target values.

Sect. 2.3. In this section, we explore the impact of imposing a prior in the metallicity.

For this purpose we computed the value Δ[Fe/H] as the difference between the estimated and target metallicities' [Fe/H] for all the discussed simulations. Figure 7 shows the results of our fits in the $\Delta Y/\Delta Z$ versus Δ[Fe/H] plane, grouped by the different explored cases. It is apparent that only Cases A and D provide a metallicity consistent with the assumed value. The result of Case A is expected because the fitting and target models coincide. In this case, the dispersion of the metallicities is modest, and all the estimated values lie within 0.02 dex of the target. Case D – fitting target models based on PARSEC 1.2s and reconstruction with FRANEC PHOENIX models – is also unbiased in [Fe/H], although the dispersion around the true values is substantially larger than in Case A, being about 0.04 dex.

All other cases show a noticeable bias. Interesting comparisons exist between Cases A and B and between Cases C and D. The former comparison shows a shift of about 0.12 dex in the estimated [Fe/H] for equivalent stellar models that rely upon different BCs. This implies that, as directly verified by means of a dedicated simulation, restricting the fit within 0.05 dex of the value characterising the target isochrones produces no valid output. Increasing the allowed range to 0.10 dex – which is quite large considering the quoted values for several open clusters – biases the results toward the low end of the allowed $\Delta Y/\Delta Z$ values. Although the argument is not rigorous, this behaviour can be easily understood by considering the Case B models in Fig. 7: only a few models are compatible with a shift of 0.10 dex, and these correspond to the low $\Delta Y/\Delta Z$ tail of the distribution. A similar behaviour exists between Cases C and D, which are characterised by a difference in the BCs of the fitting isochrone pool.

4. Conclusions

We explored the feasibility of constraining the helium-to-metal enrichment ratio, $\Delta Y/\Delta Z$, using the MS of young open clusters. To this aim, we focused on the Gaia DR3 photometry (Gaia Collaboration 2021), which provides data of exquisite precision. This method has already been adopted in works relying upon other photometric bands (e.g. Pagel & Portinari 1998; Casagrande et al. 2007; Gennaro et al. 2010; Tognelli et al. 2021), yielding quite different results depending on the models and photometric bands adopted in the calibrations.

To test the reliability of the results that can be obtained with such a calibration, we performed a theoretical investigation. First, following Tognelli et al. (2021), we identified a region of the cluster MS that is minimally affected by changes in age and is not influenced by still poorly understood input physics, such as stellar rotation or convective core overshooting. This selection restricted the range of absolute G magnitudes to (4.3, 6.5) mag, a range where the differences among various stellar evolutionary codes are lower than in more advanced evolutionary phases.

We investigated the relevance of the morphological differences between data and isochrones on the $\Delta Y/\Delta Z$ calibration by working on mock clusters' data generated from isochrones from a set of different stellar evolutionary codes: PARSEC 1.2s, PARSEC 2.0, BASTI, and MIST (Bressan et al. 2012; Nguyen et al. 2025; Hidalgo et al. 2018; Choi et al. 2016). This setup allowed us to establish the presence of biases in a controlled environment because the target values of $\Delta Y/\Delta Z$ were precisely known. We adopted two different sets of fitting isochrones based on identical stellar models computed with the FRANEC code, but implementing different BCs – namely PHOENIX and MARCS grids (Allard et al. 2011; Gustafsson et al. 2008) – to compute synthetic photometry. Synthetic clusters from target isochrones were generated at [Fe/H] values from 0.0 to 0.15 dex for different numbers of populating stars from 50 to 150, and different levels of photometric uncertainties from 0.005 mag to 0.02 mag.

The results of the Monte Carlo experiments evidenced noticeable biases. Only when the synthetic cluster generation and fitting were performed by adopting the same stellar models were the recovered $\Delta Y/\Delta Z$ values unbiased. On the other hand, even adopting underlying identical FRANEC stellar models but different BCs to obtain Gaia magnitudes resulted in biases as high as 0.6 at [Fe/H] = 0.15 with respect to the target value of 1.8. The biases were even more important when the underlying stellar models were different. While for target PARSEC isochrones the values of $\Delta Y/\Delta Z$ were underestimated by up to 0.8 from the target value of 1.78, opposite behaviour was noted for both BASTI and MIST isochrones. In these cases, we found an overestimation by about 1.3 over the target 1.3 for BASTI and 0.6 over 1.5 for MIST.

An interesting dependence on the magnitude of the assumed photometric error was evidenced. The biases in the $\Delta Y/\Delta Z$ estimation were found to decrease as the photometric error increased. This spurious phenomenon is due to a regression towards the mean value in the fitting grid, which becomes more and more relevant for increasing simulated errors. This may explain why, for several decades, it was widely believed that $\Delta Y/\Delta Z$ could be reliably constrained; the lower precision of earlier data sets unintentionally concealed the intrinsic degeneracies of the fits, leading to an overestimation of the method robustness. In the current era of high-precision photometry and dense theoretical grids, the situation has changed. Our results suggest that as observational uncertainties shrink, the underlying systematic biases become the dominant source of error, revealing the fundamental limitations of this estimation technique. The transition from sparse data to the highly accurate observations and massive computational power available today has transformed $\Delta Y/\Delta Z$ from a seemingly accessible parameter into a significant methodological challenge, necessitating a critical re-evaluation

of long-standing assumptions in the field. Actually, when the adopted uncertainty in the simulated data is low, the possibility of discrimination between different isochrone morphology is high.

The occurrence of biases that vary both in direction and magnitude across different tests strongly suggests that the calibration of $\Delta Y/\Delta Z$ using the open cluster MS is fundamentally non-robust. Importantly, this limitation is not a by-product of the specific stellar models adopted here; rather, it reflects an intrinsic sensitivity to the underlying assumptions in the input physics and BCs adopted in the stellar model computations. The fact that these biases scatter significantly when different sets of stellar models are employed – even when one restricts the comparison to state-of-the art evolutionary codes based on up-to-date and entirely reasonable physical assumptions – suggests that any alternative analysis conducted with a different code would face the same fundamental limitations. If a different set of isochrones were used as the fitting pool, the specific values of the biases might shift, but the overall lack of convergence in the results would remain. The inadequacy lies in the method itself: the MS position does not uniquely decouple $\Delta Y/\Delta Z$ from other model uncertainties. Consequently, any calibration of the helium-to-metal enrichment ratio derived from open cluster MS photometry must be viewed with scepticism, as the result is essentially determined by the specific physical input of the chosen model grid and cannot be generalised.

A fundamental limitation identified in our analysis is the significant impact of BC selection on the determination of $\Delta Y/\Delta Z$. Direct evaluation of the mean $BP - RP$ colour difference among the adopted isochrones – within the G magnitude range selected for this study – reveals a spread of approximately 0.05 mag. This dispersion significantly exceeds the precision of Gaia photometry for nearby clusters. However, these discrepancies are substantially mitigated in the G versus $T_{\rm eff}$ plane, where the mean spread across the full dataset is reduced to only 20 K. Consequently, performing the $\Delta Y/\Delta Z$ estimation in the G versus $T_{\rm eff}$ plane represents a superior alternative (see e.g. Lebreton et al. 1999 and Casagrande et al. 2007 for applications to different photometric systems), provided that high-precision effective temperatures are available.

Acknowledgements. G.V., P.G.P.M., S.D., and S.C. acknowledge INFN (Iniziativa specifica TAsP) and support from PRIN MIUR2022 Progetto "CHRONOS" (PI: S. Cassisi) finanziato dall'Unione Europea – Next Generation EU. S.C. acknowledges also the support from INAF – Theory grant "Lasting".

References

Abdi, H. 2007, in Encyclopedia of Measurement and Statistics, ed. C. S. N. Salkind (Thousand Oaks, CA: Sage), 103
Allard, F., Homeier, D., & Freytag, B. 2011, ASP Conf. Ser., 448, 91
Allard, F., Homeier, D., Freytag, B., & Sharp, C. M. 2012, EAS Publ. Ser., 57, 3
An, D., Terndrup, D. M., Pinsonneault, M. H., et al. 2007, ApJ, 655, 233
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Bahcall, J. N., Pinsonneault, M. H., & Wasserburg, G. J. 1995, Rev. Mod. Phys., 67, 781
Brandner, W., Calissendorff, P., & Kopytova, T. 2023a, AJ, 165, 108
Brandner, W., Calissendorff, P., & Kopytova, T. 2023b, A&A, 677, A162
Brandner, W., Sorg, A., Röser, S., & Schilbach, E. 2024, AJ, 168, 282
Brandt, T. D., & Huang, C. X. 2015, ApJ, 807, 24
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Buldgen, G., Noels, A., Amarsi, A. M., et al. 2025, A&A, 694, A285
Caffau, E., Ludwig, H.-G., Steffen, M., Freytag, B., & Bonifacio, P. 2011, Sol. Phys., 268, 255
Casagrande, L., Flynn, C., Portinari, L., Girardi, L., & Jimenez, R. 2007, MNRAS, 382, 1516
Castellani, V., degl'Innocenti, S., & Marconi, M. 1999, A&A, 349, 834
Castelli, F., & Kurucz, R. L. 2003, IAU Symp., 210, A20
Chaboyer, B., Sarajedini, A., & Demarque, P. 1992, ApJ, 394, 515
Chaboyer, B., Fenton, W. H., Nelan, J. E., Patnaude, D. J., & Simon, F. E. 2001, ApJ, 562, 521
Chiappini, C., & Maciel, W. J. 1994, A&A, 288, 921
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Conroy, C., Villaume, A., van Dokkum, P. G., & Lind, K. 2018, ApJ, 854, 139
Dahm, S. E. 2015, ApJ, 813, 108
de Marchi, G., & Pulone, L. 2007, A&A, 467, 107
Dell'Omodarme, M., Valle, G., Degl'Innocenti, S., & Prada Moroni, P. G. 2012, A&A, 540, A26
D'Odorico, S., Peimbert, M., & Sabbadin, F. 1976, A&A, 47, 341
Fernandes, J., Vaz, A. I. F., & Vicente, L. N. 2012, MNRAS, 425, 3104
Fukugita, M., & Kawasaki, M. 2006, ApJ, 646, 691
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1
Gennaro, M., Prada Moroni, P. G., & Degl'Innocenti, S. 2010, A&A, 518, A13
Gossage, S., Conroy, C., Dotter, A., et al. 2018, ApJ, 863, 67
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Hidalgo, S. L., Pietrinferni, A., Cassisi, S., et al. 2018, ApJ, 856, 125
Jimenez, R., Flynn, C., MacDonald, J., & Gibson, B. K. 2003, Science, 299, 1552
Jimenez, R., MacDonald, J., Dunlop, J. S., Padoan, P., & Peacock, J. A. 2004, MNRAS, 349, 240
Kurichin, O. A., Kislitsyn, P. A., & Ivanchik, A. V. 2021, Astron. Lett., 47, 674
Lebreton, Y., Perrin, M.-N., Cayrel, R., Baglin, A., & Fernandes, J. 1999, A&A, 350, 587
Lebreton, Y., Fernandes, J., & Lejeune, T. 2001, A&A, 374, 540
Maciel, W. J. 2001, Ap&SS, 277, 545
Magg, E., Bergemann, M., Serenelli, A., et al. 2022, A&A, 661, A140
Marino, A. F., Milone, A. P., Przybilla, N., et al. 2014, MNRAS, 437, 1609
Martín, E. L., Lodieu, N., Pavlenko, Y., & Béjar, V. J. S. 2018, ApJ, 856, 40
Méndez-Delgado, J. E., Esteban, C., García-Rojas, J., Arellano-Córdova, K. Z., & Valerdi, M. 2020, MNRAS, 496, 2726
Moedas, N., Deal, M., Bossini, D., & Campilho, B. 2022, A&A, 666, A43
Nguyen, C. T., Costa, G., Bressan, A., et al. 2025, A&A, 701, A258
Nsamba, B., Moedas, N., Campante, T. L., et al. 2021, MNRAS, 500, 54
Pagel, B. E. J., & Portinari, L. 1998, MNRAS, 298, 747
Pagel, B. E. J., Simonson, E. A., Terlevich, R. J., & Edmunds, M. G. 1992, MNRAS, 255, 325
Peimbert, M., & Serrano, A. 1980, Rev. Mex. Astron. Astrofis., 5, 9
Peimbert, M., & Torres-Peimbert, S. 1974, ApJ, 193, 327
Peimbert, M., Peimbert, A., & Ruiz, M. T. 2000, ApJ, 541, 688
Pinsonneault, M. H., Terndrup, D. M., Hanson, R. B., & Stauffer, J. R. 2003, ApJ, 598, 588
Planck Collaboration VI. 2020, A&A, 641, A6
Renzini, A. 1994, A&A, 285, L5
Ribas, I., Jordi, C., & Giménez, Á. 2000, MNRAS, 318, L55
Ricci, N., Valle, G., Dell'Omodarme, M., Prada Moroni, P. G., & Degl'Innocenti, S. 2025, A&A, 702, A194
Salaris, M., & Cassisi, S. 2015, A&A, 577, A60
Salaris, M., Cassisi, S., Schiavon, R. P., & Pietrinferni, A. 2018, A&A, 612, A68
Serenelli, A. M. 2010, Ap&SS, 328, 13
Silva Aguirre, V., Lund, M. N., Antia, H. M., et al. 2017, ApJ, 835, 173
Stancliffe, R. J., Fossati, L., Passy, J.-C., & Schneider, F. R. N. 2016, A&A, 586, A119
Thoul, A. A., Bahcall, J. N., & Loeb, A. 1994, ApJ, 421, 828
Tognelli, E., Dell'Omodarme, M., Valle, G., Prada Moroni, P. G., & Degl'Innocenti, S. 2021, MNRAS, 501, 383
Valcarce, A. A. R., Catelan, M., & Sweigart, A. V. 2012, A&A, 547, A5
Valcarce, A. A. R., Catelan, M., & De Medeiros, J. R. 2013, A&A, 553, A62
Valcarce, A. A. R., Catelan, M., Alonso-García, J., Contreras Ramos, R., & Alves, S. 2016, A&A, 589, A126
Valle, G., Dell'Omodarme, M., Prada Moroni, P. G., & Degl'Innocenti, S. 2013, A&A, 549, A50
Valle, G., Dell'Omodarme, M., & Tognelli, E. 2021, A&A, 649, A127
Valle, G., Dell'Omodarme, M., Prada Moroni, P. G., & Degl'Innocenti, S. 2024, A&A, 687, A294
Verma, K., Raodeo, K., Basu, S., et al. 2019, MNRAS, 483, 4678
Vernazza, J. E., Avrett, E. H., & Loeser, R. 1981, ApJS, 45, 635
Viallet, M., Meakin, C., Prat, V., & Arnett, D. 2015, A&A, 580, A61
Vinyoles, N., Serenelli, A. M., Villante, F. L., et al. 2017, ApJ, 835, 202
Wang, F., Fang, M., Fu, X., et al. 2025, ApJ, 979, 92