A method for identifying metal-poor stars with Gaia BP/RP spectra*

T. Xylakis-Dornbusch1,2, N. Christlieb1, K. Lind3, and T. Nordlander3,4

1 Zentrum für Astronomie der Universität Heidelberg, Landessternwarte, Königstuhl 12, 69117 Heidelberg, Germany. E-mail: ttxylaki@lsw.uni-heidelberg.de
2 Department of Astronomy, Stockholm University, AlbaNova University Center, 106 91 Stockholm, Sweden
3 Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
4 ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

Received 19 April 2022 / Accepted 12 July 2022

Abstract

Context. The study of the oldest and most metal-poor stars in our Galaxy promotes our understanding of the Galactic chemical evolution and the beginning of Galaxy and star formation. However, they are notoriously difficult to find, with only seven stars at [Fe/H] < −5.0 having been detected to date. Thus, the spectrophotometric data of 219 million sources which became available in the third Gaia Data Release comprise a very promising dataset for the identification of metal-poor stars.

Aims. We want to use the low-resolution Gaia Blue Photometer/Red Photometer (BP/RP) spectra to identify metal-poor stars. Our primary aspiration is to help populate the poorly constrained tail of the metallicity distribution function of the stellar halo of the Galaxy.

Methods. We developed a metal-poor candidate selection method based on flux ratios from the BP/RP Gaia spectra, using simulated synthetic spectra.

Results. We found a relation between the relative iron abundance and the flux ratio of the Ca H & K region to that of the Hγ line. This relation is temperature and surface gravity dependent, and it holds for stars with 4800 K ≤ Teff ≤ 6300 K. We applied it to noisy simulated synthetic spectra and inferred [Fe/H] with an uncertainty of σ[Fe/H] ≤ 0.65 dex for −3 ≤ [Fe/H] ≤ 0.5 and G = 15−17 mag, which is sufficient to identify stars at [Fe/H] < −2.0 reliably. We predict that by selecting stars with inferred [Fe/H] ≤ −2.5 dex, we can retrieve 80% of the stars with [Fe/H] ≤ −3 and have a success rate of about 50%, that is one in two stars we select would have [Fe/H] ≤ −3. We do not take into account the effect of reddening, so our method should only be applied to stars which are located in regions of low extinction.

Key words. stars: Population III – surveys – stars: Population II – stars: abundances – stars: late-type

1. Introduction

The Gaia survey (Gaia Collaboration 2016) poses an unprecedented opportunity to understand the structure, history, and evolution of our Galaxy better. Aside from the astrometry and photometry for over a billion stars, Gaia has also delivered spectrophotometric data in its third Data Release (Gaia DR3), which was made public on June 13, 2022. This could serve as a tool to map out the Galaxy in terms of the relative iron abundance [Fe/H] – which is a proxy for the total amount of metals in a star – and consequently find new metal-poor stars. Stars deficient in elements heavier than helium, called metal-poor stars, are usually very old, with ages comparable to that of the Universe (Frebel & Norris 2013). The basic idea is that stellar atmospheres preserve, to a large extent, the makeup of their birth cloud, hence a metal deficient star should have formed along with the first few generations of stars after the Big Bang. That is why these objects are very interesting, and they can facilitate our understanding of the past, even though they can be found close by. In particular, they can be used as tracers of the evolution of the elements and of the galaxy, help us understand how the distribution of the first population of stars (Pop III) looked, and assist in constraining Big Bang nucleosynthesis further (Frebel 2010; Beers & Christlieb 2005; Sneden et al. 2008). The Gaia Collaboration itself, through the Data and Analysis Consortium (DPAC), has already, and will also in the future, deliver astrophysical parameters using the Gaia data. With the astrometric and photometric data of Gaia DR2, Andrae et al. (2018) delivered temperatures, extinctions and reddening. DPAC also used the Gaia DR3 Blue Photometer / Red Photometer (BP/RP) spectra to estimate metallicities of stellar objects (Creevey et al. 2022; Fouesneau et al. 2022; Andrae et al. 2022). Liu et al. (2012) predicted that Aeneas – which is a Bayesian method employed by DPAC for the stellar parameter estimation – would estimate the metallicity with an accuracy of 0.1–0.2 dex, given that the extinction is A0 < 2 mag and that the magnitude is G = 15 mag. Now with Gaia DR3 available, Andrae et al. (2022) estimated [M/H] values for several million sources, and provide information on how to use them.

The search for metal-poor stars is as pertinent as ever now. The ongoing photometric SkyMapper Southern Sky (SMSS) survey is actively searching for extremely metal-poor stars ([Fe/H] < −3 dex according to the classification of Beers & Christlieb 2005). With its DR1.1, the SkyMapper team found that 40% of their candidates had [Fe/H] < −2.75 dex and only

* Full Table A1 and a table with all the coefficients of the method are only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/666/A58
** Member of the International Max Planck Research School for Astronomy & Cosmics Physics at the University of Heidelberg (IMPRS-HD).
approx. 7% had [Fe/H] > -2 dex, that is an efficacy of approx. 93% in finding stars with [Fe/H] < -2 dex (Da Costa et al. 2019). SkyMapper also led to the discovery of the star with the lowest detected Fe abundance (Nordlander et al. 2019). The Pristine survey, which is being carried out in the northern hemisphere, was tailored to find metal-poor stars with a narrow-band filter centered on the metallicity-sensitive Ca II H & K lines (Starkenburg et al. 2017). Pristine has a 56% reported efficacy of finding stars with [Fe/H] < -2.5 dex and 23% for stars with [Fe/H] < -3 dex (Aguado et al. 2019). The objective of this work is to develop a recipe that can efficiently identify metal-poor stars, and especially stars within and below the extremely metal-poor regime ([Fe/H] < -3 dex). In Sect. 2, we present the reasoning of our approach for this endeavor, as well as the tools we used. We also describe the parameter space where our method is applicable, along with a detailed description of the development of our technique. Further, in Sect. 3 we present our results from applying our method on noise-free and noisy spectra. Therein, we also investigate the dependence of our procedure on the different astrophysical parameters, and we additionally study the effect of extinction. Lastly, we explore the expected efficiency of our technique.

### 2. Methods

For this work, we used the Ulysses Simulator (Astraatmadja 2015) and synthetic spectra (Nordlander et al. 2019; see Sects. 2.1.2 and 2.1.1, respectively) in order to simulate the spectrophotometric data of BP and RP, respectively, on board Gaia. The very low resolution of the expected BP/RP spectra (see Fig. 1) inclined us to use integrated fluxes of different parts of the spectra, such that a relation varying with metallicity could be found. Those spectral areas had to be sensitive to the change in metallicity and also to at least one other quantity that can be known a priori or that can be directly derived from the spectra themselves.

#### 2.1. Data and simulations

Choosing the temperature and surface gravity space for our dataset was the first critical step for our method. The desired relation has to hold in that parameter space, which would then allow us to use those parameters as priors when applying the method described in this paper. Additionally, the dataset has to cover a wide metallicity range, so that both metal-poor and metal-rich stars can be covered, and we can ultimately distinguish between them. Since low metallicities are involved in the process, the last important factor that we have to pay attention to is the carbon enhancement that has been observed in many metal-poor stars (Beers & Christlieb 2005; Frebel & Norris 2015; Arentsen et al. 2021). The initial dataset parameters are shown in Table 1 and Fig. 2.

### Table 1. Initial parameter space of the synthetic spectra.

| Parameter      | Range         | Step |
|----------------|---------------|------|
| \( T_{\text{eff}} \) (K) | \( 3500 \leq T_{\text{eff}} \leq 7800 \) | 100  |
| \( \log g \) (dex) | \( 0.5 \leq \log g \leq 5.0 \) | 0.5  |
| [Fe/H] (dex) | \( -7.0 \leq [\text{Fe/H}] \leq 0.5 \) | 0.5  |
| [C/Fe] (dex) | \( -1.5 \leq [\text{C/Fe}] \leq 6.0 \) | 0.5  |

**Notes.** The metallicity and carbon enhancement combinations are restricted to physically meaningful combinations (see Fig. 2).
2.1.1. Synthetic spectra

For this work, we used synthetic spectra from a one dimensional (1D) grid where local thermodynamic equilibrium (LTE) is assumed (Nordlander et al. 2019), that was computed with the Turbospectrum code (Alvarez & Plez 1998; Plez 2012) and MARCS model atmospheres (Gustafsson et al. 2008). Nordlander et al. (2019) adopted the Solar chemical composition from Asplund et al. (2009); for \([\text{Fe/H}] < -1.0\), \(\alpha/\text{Fe} = +0.4\) was adopted; for models with \(-1.0 < \text{[Fe/H]} < 0.0\), a relationship of \(\alpha/\text{Fe} = +0.4 \cdot \text{[Fe/H]}\) was used. The models take continuum scattering into account and use \(v_{\text{mic}} = 1\,\text{km}\,\text{s}^{-1}\) for models with \(\log g \geq 4.0\). Models with \(\log g \leq 3.5\) take spherical symmetry into account and assume \(v_{\text{mic}} = 2\,\text{km}\,\text{s}^{-1}\). The line data are a combination of atomic data from VALD3 (Ryabchikova et al. 2015) and molecular data from a variety of sources; here the primary molecular line list is the one for CH from Masseron et al. (2014). This grid also contains spectra with varying carbon abundances, which we used for this work.

2.1.2. Ulysses

Ulysses (Astraatmadja 2015) is a simulator which takes spectra as input and delivers the equivalent BP/RP spectrum as it would be observed from Gaia. The final DR3 BP/RP spectra will be a combination of many different epoch observations, spanning up to the entire time of the missions’ operation. That is why the input spectrum is being convolved with an averaged line spread function (LSF). For each of the two spectrophotometers (BP and RP), an averaged LSF exists, which is the product of all the LSFs of the different CCDs and telescopes. The parameters of Ulysses that can be tuned, among others, are the number of transits – which are the number of times an object was observed – the amount of interstellar extinction \(A_0\), the apparent magnitude of the source, as well as the spectrum oversampling. The spectrum oversampling is essentially the number of spectra that will be combined into the final BP/RP spectrum. Each spectrum is sampled over 60 pixels, but for different observations, different parts of the spectrum are being sampled, so that there will be more than 60 pixels of data in the end. Carrasco et al. (2021) present a model for the internal calibration of the BP/RP spectra, in particular how to produce a mean spectrum from all the epoch spectra of the same source. The extinction curve models that Ulysses implements are those of Cardelli et al. (1989) and Fitzpatrick (1999). We selected the Fitzpatrick (1999) models for our simulations, after determining both models produce very similar results.

The core products of Ulysses are noise-free BP/RP spectra, noise-free Gaia photometry, and extinctions. Other outputs are also possible, for instance the end-of-mission noisy spectra, which are being generated with the addition of noise to the noise-free spectra. The noise model employed by Ulysses is the one from Jordi et al. (2010). An overview of the configuration we used to simulate our data is shown in Table 2.

2.1.3. Parameter space

A \(\chi^2\) test of the simulated spectra showed that they carry enough information in the region of the Ca II H & K lines to distinguish between metal-rich stars and stars with \(\text{[Fe/H]} < -2.0\). However, this was observed only for stars with \(T_{\text{eff}} \geq 4800\) and also depending on their surface gravity. A 12 Gyr and \(Z = 0.001\) isochrone (Bressan et al. 2012) was used to choose realistic \(T_{\text{eff}} - \log g\) pairs for the preliminary dataset (see Fig. 2).

2.2. Flux ratios

Differential simulated spectra were used to find out how the flux changes with changing metallicity. For this exercise, we used the spectra with \([\text{Fe/H}] = -2.5\) dex as a reference, and we subtracted the spectra of lower and higher metallicity that had all the other astrophysical parameters identical to the reference spectrum (Fig. 3). The regions we expected to be \([\text{Fe/H}]\) sensitive were the Ca II H & K and G-band wavelength ranges, which was confirmed: the lower the metallicity of the object, the higher the flux. The regions of the spectra that cover the Ca II near-infrared triplet as well as the H\(\beta\) absorption line

---

### Table 2. Configuration of the Ulysses simulator which we used in order to construct our method.

| Parameter               | Value |
|-------------------------|-------|
| \(G\) magnitude (mag)   | 15    |
| spectrum oversampling   | 4     |
| number of transits      | 75    |
| Extinction \(A_0\)      | 0.0   |

**Notes.** The method is described in Sect. 2.1.2. The same configuration was used for the test and applications described in Sect. 3, apart from the magnitude and extinction which are explicitly mentioned when they have different values than in the table above.

---

**Fig. 3.** Differential BP (top panel) and RP (bottom panel) spectra of Solar \([\text{C}/\text{Fe}]\)-scaled stars, with different \([\text{Fe/H}]\). Each differential spectrum results from reference spectrum – spectrum, where \([\text{Fe/H}]\)_ref = \(-2.5\). The shaded areas correspond to the regions we used for our flux ratios, i.e., the Ca II H & K region (purple), the G-band (magenta), the H\(\beta\) (yellow), and the Ca II NIR (green) regions.
showed a greater sensitivity to the change in all the other stellar parameters compared to [Fe/H]. Using the ratios of the aforementioned spectral regions, that is the ratio of the integrated Ca H & K flux to that of the Hβ region ($f_{\text{CaHK/H}\beta}$), and that of the G band integrated flux to the Ca NIR triplet ($f_{\text{G/CaNIR}}$), we can see a relation with metallicity (Fig. 4). Figure 4 was created from our entire dataset, that is temperature, surface gravity (Fig. 2), and all physically meaningful [C/Fe]–[Fe/H] combinations, in other words [C/Fe] + [Fe/H] ≤ −1 dex for [Fe/H] ≤ 0.5 and −1.5 ≥ [C/Fe] ≤ 0 for [Fe/H] ≥ −0.5.

2.3. Metallicity-temperature-$f_{\text{CaHK/H}\beta}$ relation

The relation between $f_{\text{CaHK/H}\beta}$, $f_{\text{G/CaNIR}}$, and [Fe/H] that is emerging in Fig. 4 encouraged us to try to find a relation of the form $F(f_{\text{CaHK/H}\beta}, f_{\text{G/CaNIR}}) = [\text{Fe/H}]$. In this venture, we found that $f_{\text{G/CaNIR}}$ is – by itself – an indicator for the effective temperature (see Fig. 5), but it does not carry any information about the metallicity. On the other hand, for constant $f_{\text{G/CaNIR}}$, the metallicity is changing smoothly with the change of $f_{\text{CaHK/H}\beta}$ (Fig. 6), hence $f_{\text{CaHK/H}\beta}$ carries information concerning the iron abundance. There is, however, a scatter that accounts for variations in $T_{\text{eff}}$, log $g$, and [C/Fe]. So, since $f_{\text{G/CaNIR}}$ is not metallicity-sensitive, as mentioned above, but rather a temperature indicator, we thought it best to proceed with the objective of developing a relation which is comprised of $f_{\text{CaHK/H}\beta}$ and some relevant stellar parameters. The usage of stellar parameters makes the range of applicability more straightforward to implement, and subsequently more user-friendly. By inspecting Fig. 6, we assessed that the therein $f_{\text{CaHK/H}\beta}$–[Fe/H] correlation can be described with an exponential relation (Figs. 4 and 6), which was confirmed by the residuals of the fit. This exponential relation can then be expressed with respect to the temperature and surface gravity of the model stars (Fig. A.1), enabling us to use them as priors, for example from Gaia itself (Andrae et al. 2018). Finally, we can use the iso-$f_{\text{CaHK/H}\beta}$ lines (see Fig. A.1), as well as the effective temperatures and surface gravities, in order to infer the metallicity:

$$F(f_{\text{CaHK/H}\beta}, T_{\text{eff}}, \log g) = [\text{Fe/H}]$$

(1)

$$F = -(T_{\text{eff}} \cdot \log g) \cdot e^{b \cdot f_{\text{CaHK/H}\beta} + c} + d,$$

(2)

where $b, c, d$ are $T_{\text{eff}}$ and log $g$ dependent coefficients.

The coefficients $b, c, d$ are the result of fitting Eq. (2) to $f_{\text{CaHK/H}\beta}$-metallicity pairs of a roughly constant temperature $- T_{\text{eff}} \equiv 30 \text{K}$ – and constant surface gravity (see Fig. A.2) for metallicities greater than $-3.5 \text{dex}$. The aforementioned $f_{\text{CaHK/H}\beta}$–[Fe/H] pairs result from the iso-$f_{\text{CaHK/H}\beta}$ lines.

3. Results

We first tested our method on the very same data we used to construct it, and then we examined how it is influenced by different stellar parameters. Finally, we applied our method to noisy spectra.
Specifically, for red giant branch stars shows that our method tends to overestimate the metallicity for other stars, we can identify stars with ties that are in and below the metal-poor range are most probably true ones. The benefit of this result is that the inferred metallicities to those spectra. Furthermore, we investigated the dependence on different relative carbon to iron abundances, for which we already have some information as stated above. To test that, we separated our data into surface gravity and [C/Fe] bins, and calculated the error of the inferred metallicities (Fig. A.4). There is an obvious difference in the performance pattern for stars below and above [C/Fe] = 0 dex. At and below that threshold, σ_{[Fe/H]} is almost independent of [C/Fe]. Above it, however, the performance declines as the relative carbon abundance is rising. This effect though is attenuated as temperature rises.

Another factor we considered is the extinction A_0. We used synthetic spectra with T_{eff} = 5500 K, log g = 3.5 dex, [Fe/H] = -2 dex, and all possible [C/Fe] combinations – for which [Fe/H] + [C/Fe] ≥ -1 is true – and passed them through Ulysses generating spectra of two kinds for each model: a noiseless spectrum and 20 noisy ones. We expected that the performance would decline with greater extinction, which is what we observed (see Fig. 9). We validated this result by performing the above exercise for a collection of spectra with varying astrophysical parameters. As extinction rises, f_{CaHK/450} and f_{CaNIR} decrease, in other words the model star appears to be colder and the Ca H & K features are less distinguishable. When we use Fig. 4 as a reference, a model star with increasing extinction shifts left and downwards in the plot. However, according to Christlieb et al. (2008), about 90% of stars at high galactic latitudes (i.e., at |b| ≥ 20°) have a reddening that is smaller than 0.06 (E(B − V) ≤ 0.06), which corresponds to A_0 < 0.19. Thus, the issue of extinction is outside the scope of this paper, but we consider addressing its effect in our follow-up work.

Lastly, we wanted to find out for which magnitude our method starts to break down. Hence, we generated noisy spectra (one noisy spectrum for each set of stellar parameter combinations) of magnitudes G = 16, 17, 18 mag. Up to 17 mag, the performance of our method is roughly on the same levels of precision (Fig. 8), and at 18 mag, σ_{[Fe/H]} < 0.85 for [Fe/H] ≥ -3. Our method does not seem to break down, but it rather gets less precise as the magnitude rises. Even so, at 18 mag, we can infer metallicities down to -2 dex with an uncertainty of σ_{[Fe/H]} ≤ 0.6 dex.

### 3.2. Application on noisy spectra and the dependence on stellar parameters

We subsequently applied our method to noisy spectra. For each of our model spectra (G = 15 mag), we generated 20 noisy ones: we inferred the metallicity with σ_{[Fe/H]} ≤ 0.6 dex for [Fe/H] ≥ -3 dex (Fig. 8). Again, the uncertainty in inferring metallicity for metal-poor stars ([Fe/H] = -1) and above is very low with σ_{[Fe/H]} ≤ 0.12 dex. Further we sought to determine how our method performs in the different temperature bins. For this purpose, we computed σ_{[Fe/H]} for all spectra of the same temperature, surface gravity, and relative carbon abundance. We found that the dependence on temperature is linked with [C/Fe]. When [C/Fe] ≤ 0.0 dex, our method performs slightly better for lower effective temperatures of the same log g when [Fe/H] < -4. In contrast, when the iron abundance is greater than -4, performance becomes independent of T_{eff}. When [C/Fe] > 0, performance has a similar behavior for [Fe/H] < -4, that is to say it is better for lower temperatures. However, above that threshold, the effect is reversed: the performance is much better for higher effective temperatures. That phenomenon is intensified as [C/Fe] grows. We expected that we could determine the metallicity for lower temperatures more accurately since the Ca H & K lines are stronger in cooler stars compared to hotter ones, even down to very low metallicities. This, as stated above, is the case for [C/Fe] ≤ 0, but not above that threshold. The reason for this is probably due to the many carbon lines that arise as a result of the high carbon enhancements, which in turn are also stronger at lower effective temperatures and, consequently, our method assigns higher metallicities to those spectra.

![Fig. 7. σ and bias of the inferred metallicities of our noise-free spectra of G = 15 mag.](image-url)
and bias of the inferred metallicities of noisy spectra of different magnitudes: $G = 15, 16, 17, 18$ mag.

Fig. 9. Testing the dependence of metallicity inference on different extinction coefficients has the following expected outcome: the higher the extinction, the higher the uncertainty. We used one set of stellar parameters and metallicity, i.e., $T_{\text{eff}} = 5500$ K, $\log g = 3.5$ dex, and $[\text{Fe/H}] = -2$ dex, but various relative carbon abundances. On the top are the $\sigma_{\text{[Fe/H]}}$ and bias in bins of extinctions, and in the bottom the results are also in bins of different $[\text{C/Fe}]$.

Fig. 10. Metallicity and carbonicity of the simulated dataset we used to predict the success rate of our method. The color bar designates the number of models in each $[\text{Fe/H}]$–$[\text{C/Fe}]$ bin.

of halo stars, and from two carbonicity distribution functions, respectively. Specifically, we used the MDF from Youakim et al. (2020) when the initially drawn metallicity was $>-2$ dex, and the MDF from Placco et al. (2014) otherwise. Concerning the carbonicity, we used the respective carbon-enhancement distribution for the Placco et al. (2014) MDF (Placco et al. 2014), and we used the $[\text{C/Fe}]$ distribution from Yoon et al. (2016) for $[\text{Fe/H}] > -2$.

We found that when we selected all stars with an inferred $[\text{Fe/H}]_{\text{inf}} \leq -2.5$, we recovered 80% of stars with $[\text{Fe/H}] \leq -3$, and we had a contamination of about 2% of stars with metallicities above $-2.5$. It should be noted, however, that these “contamination” stars all have $[\text{Fe/H}] < -2$. Furthermore, about 55% of the stars with $[\text{Fe/H}]_{\text{inf}} \leq -2.5$ have a reference $[\text{Fe/H}] \leq -3$, which means that one in two of the selected stars would at least be extremely metal-poor ($[\text{Fe/H}] \leq -3$). If we were to select the metallicity threshold at $[\text{Fe/H}]_{\text{inf}} \leq -3$, the hit rate for stars below $-3$ would increase to 9 out of 10. We would, however, fail to detect about 75% of stars with $[\text{Fe/H}] \leq -3$ (see Fig. 11). The overall results of this test are detailed in Fig. 11. Specifically, we show how the success rate, the completeness,
and the contamination change by selecting a different threshold, that is for [Fe/H]_{inf} = −3.5, −3.0, −2.5, −2.0. The contamination was calculated with reference to the threshold synthetic spectra, whereas the completeness and the success rate were calculated with reference to model spectra with [Fe/H] ≤ −3.

4. Conclusions

We developed a method using flux ratios of metallicity-sensitive regions from the Gaia BP/RP low resolution spectra in order to find new metal-poor stars. This method is applicable when stars have 4800 K ≤ T_{eff} ≤ 6300 K. We took into account the fact that a large fraction of metal-poor stars are carbon enhanced, and thus used a grid of synthetic spectra with varying [Fe/H] and [C/Fe]. We found an exponential relation between the metallicity and the $f_{\text{CaIIK,H}}$ flux ratio, which is temperature and surface gravity dependent. Therefore, our method requires both of those stellar parameters as priors. We first applied our method to the very same noiseless data (G = 15 mag) we used to construct it, and [Fe/H] was inferred with an uncertainty of $\sigma$ ≤ 0.6 dex at [Fe/H] ≤ −3.0 dex. Our method’s performance was approximately on the same level for noisy spectra of the same and greater magnitudes, that is $\sigma$ ≤ 0.65 for [Fe/H] ≤ −3.0 dex and G = 15, 16, 17 mag. We found that the performance depends on temperature, but in conjunction with the relative carbon abundance: for Solar [C/Fe] and below, performance is slightly better for lower temperatures of the same surface gravity when determining [Fe/H] ≤ −3.5. For [Fe/H] ≥ −3, the performance is practically independent of T_{eff}. When [C/Fe] > 0, $\sigma_{\text{[Fe/H]}}$ ≤ −3.5 is as in the Solar case, that is it is lower for lower T_{eff}. However, the determination of the metallicity above −3.5 dex presents a lower uncertainty at higher temperatures of the same log g. A dependence of the performance from the relative carbon-to-iron abundance is observed when [C/Fe] > 0. Further, even for spectra of 18th magnitude, we can determine metallicities down to −2 with an uncertainty of $\sigma_{\text{[Fe/H]}}$ < 0.6 dex, which is sufficient for identifying metal-poor stars. Nevertheless, we find that the overall performance of our method deteriorates with rising extinction A_0. Lastly, we predict that by selecting stars with [Fe/H]_{inf} ≤ −2.5 dex, we achieve a completeness of 80% of the stars with [Fe/H] ≤ −3, and that one in two stars we select is extremely metal-poor. Our method will be applied to Gaia DR3 data and the results will be published in a forthcoming paper.

Acknowledgements. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 138713388 – SFB 881 (“The Milky Way System”, subproject A04). KL acknowledges funds from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant agreement no. 852977) and funds from the Knut and Alice Wallenberg foundation. This research was supported by the Australian Research Council Centre of Excellence for All Sky Astrometry in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work was supported by computational resources provided by the Australian Government through the National Computational Infrastructure (NCI) under the National Computational Merit Allocation Scheme and the ANU Merit Allocation Scheme (project y89).

References

Aguado, D. S., Youakim, K., González Hernández, J. I., et al. 2019, MNRAS, 490, 2241
Alvarez, R., & Plez, B. 1998, A&A, 330, 1109
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8
Andrae, R., Fouesneau, M., Sordo, R., et al. 2022, A&A, in press, https://doi.org/10.1051/0004-6361/202243462
Arentsen, A., Starkenburg, E., Aguado, D. S., et al. 2021, MNRAS, 505, 1239
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Asstralad, T. L. 2015, Technical Report GAIA-CS-TN-MPIA-TLA-001, Max Planck Institute for Astronomy, Heidelberg

Beers, T. C., & Christlieb, N. 2005, ARA&A, 43, 531
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86
Christlieb, N., Schönrock, T., Frebel, A., et al. 2008, A&A, 484, 721
Creevey, O. L., Sordo, R., Pailler, F., et al. 2022, A&A, in press, https://doi.org/10.1051/0004-6361/202243688
Da Costa, G. S., Bessell, M. S., Mackey, A. D., et al. 2019, MNRAS, 489, 5900
Fitzpatrick, E. L. 1999, PASP, 111, 63
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2022, A&A, accepted, [arXiv:2206.05992]
Frebel, A. 2010, Astron. Nachr., 331, 474
Frebel, A., & Norris, J. E. 2013, in Planets, Stars and Stellar Systems (Springer), 55
Frey, B., & Norris, J. E. 2015, ARA&A, 53, 631
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Liu, C., Baier-Jones, C. A. L., Sordo, R., et al. 2012, MNRAS, 426, 2463
Masseron, T., Plez, B., Van Eck, S., et al. 2014, A&A, 571, A47
Nordlander, T., Bessell, M. S., Da Costa, G. S., et al. 2019, MNRAS, 488, L109
Placco, V. M., Frebel, A., Beers, T. C., & Stancliffe, R. J. 2014, ApJ, 797, 21
Plez, B. 2012, Turbospectrum: Code for spectral synthesis
Rainbokiva, T., Prskov, N., Kurucz, R. L., et al. 2015, Phys. Scr, 90, 054005
Snedden, C., Cowan, J. J., & Gallino, R. 2008, A&A, 46, 241
Starkenburg, E., Martin, N., Youakim, K., et al. 2017, MNRAS, 471, 2587
Youk, J., Beers, T. C., Placco, V. M., et al. 2016, ApJ, 833, 20
Youakim, K., Starkenburg, E., Martin, N. F., et al. 2020, MNRAS, 492, 4986

Fig. 11. Calculation of the success rate, the completeness (left plot), and the contamination (right plot) for threshold values [Fe/H]_{inf} = −3.5, −3.0, −2.5, −2.0. The contamination was calculated with reference to the threshold synthetic spectra, whereas the completeness and the success rate were calculated with reference to model spectra with [Fe/H] ≤ −3.
Appendix A: Additional figures and data.

The figures in the appendix are described in Sections 2.3 and 3.2. Figures A.1 and A.2 describe the way we developed our method, whereas Figures A.3 and A.4 detail the dependence of our method on effective temperature and relative carbon abundance. In Table A.1 we provide the entire dataset that we used to produce Figure 4, so that the interested readers can apply our method themselves.

Fig. A.1. Contour plots where the third dimension are iso-flux lines, in particular for the \( f_{\text{CaHK/H\beta}} \) flux ratio. Those iso-\( f_{\text{CaHK/H\beta}} \) lines are temperature-sensitive up to about \([\text{Fe/H}] = -3.5\) dex, and then they become metallicity-sensitive, hence we can use them to find a relation from which we can infer \([\text{Fe/H}]\). Because the metallicity sensitivity starts at \([\text{Fe/H}] = -3.5\), we cannot distinguish iron abundances below that threshold, but rather identify whether they are below or above it.
Table A.1. Noise-free $f_{\text{CaHK/H}\beta}$ and $f_{\text{G/CaNIR}}$ flux ratios and their respective astrophysical parameters.

| a/a | $f_{\text{G/CaNIR}}$ | $f_{\text{CaHK/H}\beta}$ | $T_{\text{eff}}$ (K) | $\log g$ (dex) | $[\text{Fe/H}]$ (dex) | $[\text{C/Fe}]$ (dex) |
|-----|-----------------|-----------------|----------------|--------------|----------------|----------------|
| 0   | 0.353852        | 0.056188        | 4800.0         | 5.0          | -0.5          | -0.5          |
| 1   | 0.361829        | 0.056728        | 4800.0         | 5.0          | -0.5          | -1.0          |
| 2   | 0.365205        | 0.056951        | 4800.0         | 5.0          | -0.5          | -1.5          |
| 3   | 0.337023        | 0.054872        | 4800.0         | 5.0          | -0.5          | 0.0           |
| 4   | 0.288912        | 0.050126        | 4800.0         | 5.0          | -0.5          | 0.5           |
| 5   | 0.187117        | 0.038496        | 4800.0         | 5.0          | -0.5          | 1.0           |
| 6   | 0.122764        | 0.032674        | 4800.0         | 5.0          | -0.5          | 1.5           |
| 7   | 0.086716        | 0.033368        | 4800.0         | 5.0          | -0.5          | 2.0           |
| 8   | 0.324806        | 0.050360        | 4800.0         | 5.0          | 0.0           | -0.5          |
| 9   | 0.333351        | 0.050936        | 4800.0         | 5.0          | 0.0           | -1.0          |
| 10  | 0.336977        | 0.051179        | 4800.0         | 5.0          | 0.0           | -1.5          |
| 11  | 0.305943        | 0.048967        | 4800.0         | 5.0          | 0.0           | 0.0           |
| 12  | 0.227209        | 0.042632        | 4800.0         | 5.0          | 0.0           | 0.5           |
| 13  | 0.144015        | 0.034670        | 4800.0         | 5.0          | 0.0           | 1.0           |
| 14  | 0.097575        | 0.032297        | 4800.0         | 5.0          | 0.0           | 1.5           |
| 15  | 0.073156        | 0.034265        | 4800.0         | 5.0          | 0.0           | 2.0           |

8963 0.471143   0.097089  5000.0  1.5  -7.0  0.0
8964 0.471170   0.097088  5000.0  1.5  -7.0  0.5
8965 0.471178   0.097102  5000.0  1.5  -7.0  1.0
8966 0.471137   0.097087  5000.0  1.5  -7.0  1.5
8967 0.471192   0.097109  5000.0  1.5  -7.0  2.0
8968 0.471005   0.097071  5000.0  1.5  -7.0  2.5
8969 0.470596   0.097043  5000.0  1.5  -7.0  3.0
8970 0.469710   0.096964  5000.0  1.5  -7.0  3.5
8971 0.466643   0.096713  5000.0  1.5  -7.0  4.0
8972 0.458394   0.095966  5000.0  1.5  -7.0  4.5
8973 0.439977   0.094247  5000.0  1.5  -7.0  5.0
8974 0.409192   0.090923  5000.0  1.5  -7.0  5.5
8975 0.368257   0.084704  5000.0  1.5  -7.0  6.0

Notes. Figure 4, as well as our model, was produced from this dataset.

Fig. A.2. We fit Equation (2) in the metallicity-$f_{\text{CaHK/H}\beta}$ space for each temperature-surface gravity pair from our parameter space (see Figure 2). Some of those fits are shown in the above panels.
Fig. A.3. Temperature dependence of the performance of our method is twofold. When \([\text{C/Fe}] \leq 0\), the uncertainty is slightly better for lower temperatures up to \([\text{Fe/H}] \sim -3.5\), and above that, it becomes almost independent of \(T_{\text{eff}}\). When \([\text{C/Fe}] > 0\), the pattern is the same as above for \([\text{Fe/H}] \) up to \(\sim -3.5\); whereas, above \(-3\) dex, lower temperatures render higher uncertainties.

Fig. A.4. Computation of \(\sigma_{\text{[Fe/H]}}\) for all temperatures corresponding to \(\log g = 3.5\) dex, and all bins of \([\text{C/Fe}]\). The manner in which the performance of our method depends on the relative carbon abundances relates to whether or not \([\text{C/Fe}] \leq 0\) dex: at and below a Solar value, \(\sigma_{\text{[Fe/H]}}\) is practically independent of carbon, but above it the uncertainty rises as carbon-enhancement increases. The last effect lessens as temperature rises. The same exercise was performed for all surface gravities of our parameter space, and the results were similar.