Gaia FGK benchmark stars: new candidates at low-metallicities *

K. Hawkins$^1$, P. Jofré$^1$, U. Heiter$^2$, C. Soubiran$^1$, S. Blanco-Cuaresma$^4$, L. Casagrande$^5$, G. Gilmore$^1$, K. Lind$^6$, L. Magrini$^7$, T. Masseron$^1$, E. Pancino$^{7,8,9}$, S. Randich$^1$, and C. C. Worley$^1$

1 Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom
2 Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden
3 Université de Bordeaux - CNRS, LAB - UMR 5804, BP 89, 33270, Floirac, France
4 Observatoire de Genève, Université de Genève, CH-1290 Versoix, Switzerland
5 Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, The Australian National University, ACT 2611, Australia
6 Max Planck Institute for Astronomy Konigstuhl 17, D-69117 Heidelberg, Germany
7 INAF/Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5 50125 Firenze, Italy
8 INAF/Osservatorio Astronomico di Bologna, Via Ranzani 1, 40127 Bologna, Italy
9 ASI Science Data Center, Via del Politecnico snc, 00133, Roma, Italy

ABSTRACT

Context. We have entered an era of large spectroscopic surveys in which we can measure, through automated pipelines, the atmospheric parameters and chemical abundances for large numbers of stars. Calibrating these survey pipelines using a set of “benchmark stars” in order to evaluate the accuracy and precision of the provided parameters and abundances is of utmost importance. The recent proposed set of Gaia FGK benchmark stars of Heiter et al. (2015) has up to five metal-poor stars but no recommended stars within −2.0 < [Fe/H] < −1.0 dex. However, this metallicity regime is critical to calibrate properly.

Aims. In this paper, we aim to add candidate Gaia benchmark stars inside of this metal-poor gap. We began with a sample of 21 metal-poor stars which was reduced to 10 stars by requiring accurate photometry and parallaxes, and high-resolution archival spectra.

Methods. The procedure used to determine the stellar parameters was similar to Heiter et al. (2015) and Jofré et al. (2014) for consistency. The difference was to homogeneously determine the angular diameter and effective temperature ($T_{\text{eff}}$) of all of our stars using the Infrared Flux Method utilizing multi-band photometry. The surface gravity (log $g$) was determined through fitting stellar evolutionary tracks. The [Fe/H] was determined using four different spectroscopic methods fixing the $T_{\text{eff}}$ and log $g$ from the values determined independent of spectroscopy.

Results. We discuss, star-by-star, the quality of each parameter including how it compares to literature, how it compares to a spectroscopic run where all parameters are free, and whether Fe ionisation-excitation balance is achieved.

Conclusions. From the 10 stars, we recommend a sample of five new metal-poor benchmark candidate stars which have consistent $T_{\text{eff}}$, log $g$, and [Fe/H] determined through several means. These stars, which are within −1.3 < [Fe/H] < −1.0, can be used for calibration and validation purpose of stellar parameter and abundance pipelines and should be of highest priority for future interferometric studies.

1. Introduction

Chemodynamical studies of our Galaxy are beginning to use large samples of stars as a result of in multi-object spectroscopic surveys (e.g. Gaia-ESO, APOGEE, GALAH, and others). In particular, the recently launched Gaia satellite will undoubtedly revolutionise our understanding of the Milky Way with accurate parallaxes and proper motions, and accompanying spectral information for more than a billion stars. Combining data from the many multi-object spectroscopic surveys which are already underway, and the rich dataset from Gaia will be the way forward in order to disentangle the full chemo-dynamical history of our Galaxy. One example is the Gaia-ESO Public Spectroscopic Survey (GES, Gilmore et al. 2012; Randich et al. 2013), which aims to provide atmospheric parameters and elemental abundances of more than 10$^5$ stars. Another example is the Australian GALAH survey (De Silva et al. 2015), which will undoubtedly contain large numbers of metal-poor stars, and the Apache Point Galactic Evolution Experiment (APOGEE) survey (Eisenstein et al. 2011), which samples giant stars across a broad range in metallicity. In the future, even larger datasets will be produced, such as the southern 4MOST survey (de Jong et al. 2012) or its complimentary northern survey WEAVE (Dalton et al. 2014).

Our methods to do stellar spectroscopy, in particular, to determine the main atmospheric parameters including effective temperature ($T_{\text{eff}}$), surface gravity (log $g$) and metallicity ([Fe/H]), have necessarily evolved towards a more automatic and efficient way. However, these methods need to be calibrated in order to judge their performance. This calibration can be properly done with a set of well-known stars, or benchmark stars. In addition, the multiple surveys need to be corrected for systematic offsets between them in order to compare results. This work is about the assessment of such stars.

Beside astrometry, Gaia will produce, for most stars, atmospheric parameters of stars through a pipeline named APSIS (Bailey-Jones et al. 2013). For the calibration of APSIS, we have, in previous reports on this subject, defined a set of stars that cover different parts on the Hertzsprung-Russell diagram (HRD) in the FGK spectral range (henceforth Paper I, Heiter et al. 2015). We attempted to cover a wide range in metallicities, such that these stars would represent a large portion of the Gaia observations. We have called this sample the Gaia FGK bench-
mark stars (GBS, Paper I). The $T_{\text{eff}}$ and log $g$ of the current set of GBS have been determined with fundamental relations, independently from spectroscopy, making use of the star’s angular diameter ($\theta_\text{LD}$) and bolometric flux ($F_{\text{bol}}$) combined with its distance (Paper I). The metallicity is then determined by using a homogeneous library of spectra. That library is described in (Blanco-Cuaresma et al. 2014b, henceforth Paper II). This library is analysed to determine the metallicity based on the adopted values for $T_{\text{eff}}$ and log $g$ (Jofré et al. 2014, henceforth Paper III). High spectral resolution analyses not only yield atmospheric parameters but also individual abundances, thus the same library has been used to derive the abundance of 4 alpha elements and 6 iron-peak elements (Jofré et al. 2015, henceforth Paper IV).

These stars have been shown to be an excellent sample to calibrate the stellar parameter determination pipelines of the Gaia-ESO Survey (Smiljanic et al. 2014, Recio-Blanco et al. in prep) or other spectroscopic surveys and studies (e.g. Schönrich & Bergemann 2014; DePascale et al. 2014; Lemasle et al. 2014; De Silva et al. 2015; Boeche & Grebel 2015; Hawkins et al. 2016). However, the calibrations are currently limited by less than a handful of metal-poor main-sequence stars in our initial GBS sample (e.g. see the calibration paper by Smiljanic et al. 2014). The reason is that metal-poor stars are normally further away and thus fainter, making it impossible to measure their $\theta_\text{LD}$ accurately with current interferometric instruments except in very rare cases. The metallicity regime around [Fe/H] $\sim$ −1.0 dex is particularly important because this represents the transition between several Galactic components (e.g. Venn et al. 2004; Nissen & Schuster 2010; Bensby et al. 2014; Hawkins et al. 2015). For example, the halo is thought to have a mean metallicity of −1.5 with a dispersion of 0.50 dex and the thick disk has a mean metallicity of −0.50 dex with a dispersion of 0.25 dex. Thus at [Fe/H] $\sim$ −1.0 dex, the thick disk and halo components are entangled. Therefore, it is critical to calibrate this metallicity regime correctly.

Among the set of current GBS, nearly 20% (6 stars) have radius and bolometric flux estimated indirectly using photometric relations. At least one of the two current (recommended) metal-poor GBS have radius and bolometric flux estimated indirectly using photometric relations. In this paper, we use similar and consistent relations to include more metal-poor stars in a homogeneous way. We do this because for many of these GBS candidates $\theta_\text{LD}$ can not yet be reliably measured with interferometry. In particular, systematic effects might still be the major limitation at the sub-milliarcsec level (e.g. Casagrande et al. 2014).

In the current set of GBS there are a total of five metal-poor stars with [Fe/H] $<-1.0$ dex ($\psi$ Phe, HD122563, HD84937, HD140283, Gmb 1890). However among these five, three have not been recommended for calibration or validation purposes in Paper I. HD140283 was not recommended because of the large uncertainties in the $T_{\text{eff}}$ which is likely a result of a calibrated bolometric flux which had large systematic differences between the photometric and spectroscopic values. Gmb 1890 has a highly uncertain $T_{\text{eff}}$ which could be due to calibration errors in the interferometry (Paper I) and thus it was not recommended. The measured angular diameter of Creevey et al. (2015) yields an effective temperature that is more than 400 K lower than the spectroscopic $T_{\text{eff}}$. Additionally, the cool M giant star $\psi$ Phe was not recommended, in part, because of an uncertain metallicity caused by the inability of the methods employed to properly deal with the molecular features which heavily crowd the spectrum.

This leaves only two metal-poor stars which have metallicities below −2.0 dex and effectively no stars with $-2.0 < [\text{Fe/H}] < -1.0$ dex. We aim to provide a set of new GBS candidates stars inside of the metal-poor gap listed above. These new stars ultimately will allow the astronomical community and spectroscopic surveys to extend their calibrations based on the benchmark stars possibly reaching into the critical regime of $-1.3 < [\text{Fe/H}] < -1.0$ dex. The metallicity distribution of the recommended set for calibration and validation purposes from Paper I (blue histogram) and the additional metal-poor candidate stars (red histogram) are shown in Fig. 1. In the background of that figure is the metallicity distribution of the full recommended sample of stars from the GES iDR4 UVES (black dash-dotted histogram) and GIRAFFE (gray solid histogram) spectra (for more information on UVES see Dekker et al. 2000). A sizable fraction of the stars in the GES iDR4 are in the metal poor regime and thus a proper calibration through metal poor GBS is necessary. The recommended stellar parameters (and metallicity) of the GES iDR4 stars have been determined by spectral analysis of several methods (nodes) whose results have been homogenized and combined (details of this will be published in Hourihane et al. 2016 in prep).

In this fifth work of the series, we define a new set of GBS candidates stars inside of the metal-poor gap. We note that these candidates do not have $\theta_\text{LD}$ measurements and should remain as candidates until an $\theta_\text{LD}$ can be measured directly, at least for a handful, in the near future. In addition, we aim to provide a set of metal-poor stars with predicted $\theta_\text{LD}$ which can be used as the input for future interferometric studies.

As such, this paper is organised in the following way: in Sect. 2 we begin by selecting a sample of relatively bright metal-poor stars that have archival spectra. We then describe the several methods that we have used to determine the $T_{\text{eff}}$ (Sect. 3) and log $g$ (Sect. 4). Fixing these parameters, we determined the metallicity using methods consistent with Paper III, which we describe in Sect. 5. In Sect. 6 we present the results of the
parameter analysis and discuss, star-by-star, the quality of the parameters and recommend a new set of metal-poor benchmark stars. We also compare our results with what is known about these stars in the literature. Finally, in Sect. 7 we summarize our analysis and recommendations.

2. Sample

The initial target list was selected using the PASTEL database requiring the following: (1) $4500 < T_{\text{eff}} < 6500$ K, (2) $-2.0 < [\text{Fe/H}] < -1.0$ dex, and (3) there were at least four $T_{\text{eff}}$ and metallicity estimates in the literature, since 1990, with a standard deviation of less than 100 K and 0.1 dex, respectively. The third criterion was used filter out stars where there are obvious discrepancies in the stellar parameters or the star was ill-behaved in order to maximise the chance that after our analysis, the stars will have metallicities and parameters in the regime of interest. These criteria result a total of 21 stars including Gbm1380 (HD103095). The metallicity distribution of these stars will have metallicities and parameters in the regime that we required that the photometric and parallax information can be found in Table 2. The photometric and parallax information was determined in two ways: (1) using the photometric relationship outlined in Equations 1.2 and 1.3 of Paper I which relates the bolometric flux, $F_{\text{bol}}$, and the $B$-band flux. $F_{\text{bol}}$ is determined by fitting the $B$- and $V$-band photometric calibrations (e.g. van Belle 1999; Kervella et al. 2004; Di Benedetto 2005; Boyajian et al. 2014) with the Stefan-Boltzmann law, and (2) using the IRFM (e.g. Blackwell & Shallis 1977; Blackwell et al. 1979, 1980; Casagrande et al. 2009, 2010, 2012) and spectroscopic (e.g. Gratton et al. 1996; Nissen & Schuster 1997; Gratton et al. 2000; Mishenina et al. 2000; Fulbright 2000; Gratton et al. 2003; Sousa et al. 2011) means. In some cases the spectroscopic $T_{\text{eff}}$ is determined by fitting the wing of the strong Balmer H features, usually Ha or Hg (e.g. Aker et al. 1994; Mashonkina & Gehren 2000; Gehren et al. 2004). Since the distance is known, the log $g$ is largely derived using the parallax (e.g. Gratton et al. 2000; Gehren et al. 2004; Jonsell et al. 2005). However, in some cases the Fe ionisation balance (Axer et al. 1994; Fulbright 2000; Sousa et al. 2011; Ishigaki et al. 2012) or Mg-triplet wing fitting (e.g. Mashonkina & Gehren 2000) has been used. Metallicity is determined from the analysis of iron lines under 1D-LTE approximations in most of the works (e.g. Axer et al. 1994; Fulbright 2000; Jonsell et al. 2005; Valenti & Fischer 2005; Sousa et al. 2011; Ishigaki et al. 2012). Extensive discussions of these works and our results are found in Sect. 6.

3. Determination of effective temperature

The analysis presented here is consistent with the previous papers in the series (Paper I, Paper II, Paper III) allowing the parameters of these metal-poor stars to be added to the GBS sample covering a wide and well sampled parameter space in the HRD.

As in Paper I-IV, we chose stars that have been widely studied in the past. Table 1 indicates there are between 4 – 35 studies for each star. However, as seen below, these studies are very different from each other (using different procedures to determine the stellar parameters) and thus the advantage of this work is to homogenise the stellar parameters with respect to Paper I-IV so that they can be ingested into the current GBS.

The parameters given in Table 1 have been determined through a variety of means. For example, the $T_{\text{eff}}$ has been determined through both photometric (e.g. Alonso et al. 1996a; Nissen et al. 2002; Ramírez & Meléndez 2005; Jonsell et al. 2005; Masana et al. 2006; Reddy et al. 2006; González Hernández & Bonifacio 2009; Casagrande et al. 2010, 2011; Ishigaki et al. 2012) and spectroscopic (e.g. Gratton et al. 1996; Nissen & Schuster 1997; Gratton et al. 2000; Mishenina et al. 2000; Fulbright 2000; Gratton et al. 2003; Sousa et al. 2011) means. In some cases the spectroscopic $T_{\text{eff}}$ is determined by fitting the wing of the strong Balmer H features, usually Ha or Hg (e.g. Aker et al. 1994; Mashonkina & Gehren 2000; Gehren et al. 2004). Since the distance is known, the log $g$ is largely derived using the parallax (e.g. Gratton et al. 2000; Gehren et al. 2004; Jonsell et al. 2005). However, in some cases the Fe ionisation balance (Axer et al. 1994; Fulbright 2000; Sousa et al. 2011; Ishigaki et al. 2012) or Mg-triplet wing fitting (e.g. Mashonkina & Gehren 2000) has been used. Metallicity is determined from the analysis of iron lines under 1D-LTE approximations in most of the works (e.g. Axer et al. 1994; Fulbright 2000; Jonsell et al. 2005; Valenti & Fischer 2005; Sousa et al. 2011; Ishigaki et al. 2012). Extensive discussions of these works and our results are found in Sect. 6.

3.1. Deriving temperature using angular diameter-photometric relationships

To compute the $T_{\text{eff}}$, we used Equation 1 of Paper I which relates the $T_{\text{eff}}$ to the bolometric flux, $F_{\text{bol}}$, and the $\theta_{\text{D}}$. We estimated the $F_{\text{bol}}$ using the photometric relationship outlined in Equations 8 and 9 of Alonso et al. (1995) which rely on the $V$ and $K$ photometry. We note that the photometric relationship to obtain the $F_{\text{bol}}$ required that the $K$ magnitude was in the Johnson rather than 2MASS bandpasses. Thus, we converted the 2MASS photometry (columns 7 and 9 in Table 2) bands into the Johnson system

This star now has spectra available in the ESO archive but was not public when the target selection for this project was completed

This is a spectroscopic binary system in which neither component had an ESO/NARVAL spectrum.
using the following relationship:

\[ K_j = K_{2MASS} - 0.1277(J - K)_{2MASS} + 0.0460, \]  

(1)

where \( K_1, K_{2MASS} \) are the \( K \)-band magnitude in the Johnson and 2MASS systems, respectively. The 2MASS subscript refers to the 2MASS J, K, and (J-K), and the J subscript to Johnson system. This relationship was obtained by combining Equations 6, 7, 13, and 14 from Alonso et al. (1994) and Equations 12 and 14 from Carpenter (2001). The uncertainty in \( K_1 \) was determined by propagating the uncertainty in the 2MASS \( (J - K)_{2MASS} \). We note here that the photometry was corrected for reddening using the values in column 13 of Table 2. These corrections are very small and have the effect of changing the \( \theta_{LD} \) on the order of less than 1% and \( T_{eff} \) by less than 30 K when compared to the raw photometric values.

The \( \theta_{LD} \) was determined directly through photometric relationships. We have made use of four separate \( \theta_{LD} \)-photometric relations in order to test the robustness of this procedure (van Belle 1999; Kervella et al. 2004; Di Benedetto 2005; Boyajian et al. 2014). The first set of calibrations used were taken from the work of van Belle (1999). We determined the angular diameter of all stars by taking the average of the \( \theta_{LD} \) relation (their Equation 2) and \( \theta_{LD} \) relation (their Equation 3). The second set of calibrations, which was used only for the dwarf stars, were taken from the photometric relationships of Kervella et al. (2004). Just as above, we averaged the \( \theta_{LD} \)-\((B-V)\) relation (their Equation 22) and \( \theta_{LD} \)-\((V-K)\) relation (their Equation 23). We note this procedure was used for the \( \theta_{LD} \) for the GBS HD22879 and \( \theta_{LD} \) for Paper I. The third set of calibrations were from Di Benedetto (2005). We computed the \( \theta_{LD} \) of all stars using the \( \theta_{LD} \)-\((B-V)\) relation (their Equations 1 and 2). The final set of calibrations used were from Boyajian et al. (2014). We made use of their \( \theta_{LD} \)-\((B-V)\) relation (their Equation 4) which is only applicable to the dwarf stars.

The results of the \( \theta_{LD} \) and \( T_{eff} \) computed using the various \( \theta_{LD} \)-photometric calibrations above can be found in Fig. 2. In the top panel of Fig. 2, we compare the \( \theta_{LD} \), in milliarcseconds (mas), of each star and relation used. We also plot the \( \theta_{LD} \) computed from the infrared flux method (hereafter IRFM, see Sect. 3.2 and Casagrande et al. 2006, 2010, 2014, for more details). In the middle panel of Fig. 2 we show the relative difference between the four \( \theta_{LD} \)-photometric relations with that computed from the IRFM. In the bottom panel of Fig. 2, we compare the \( T_{eff} \) derived from the different \( \theta_{LD} \)-photometric calibrations and that computed from the IRFM. In most cases the \( \theta_{LD} \) from each of the photometric calibrations are consistent (within 1\( \sigma \)) with each other and the \( \theta_{LD} \) from the IRFM.

As noted by Paper I, we choose to use the \((V-K)\)-\( \theta_{LD} \) relations because they have the smallest dispersion in the fitted relationship (on the order of less than 1%) compared to other photometric colours. These equations are created by relating the \( \theta_{LD} \) of dwarf, subgiant, and giant stars determined via interferometry to their \((V-K)\) colour and \( K_1 \) magnitude (e.g. van Belle 1999; Kervella et al. 2004; Di Benedetto 2005; Boyajian et al. 2014). While it is likely that the brightest star (HD175305) may have direct \( \theta_{LD} \) measurements, most of these stars are dim.

| Star          | RA (J2000) | DEC (J2000) | T_{eff} (K) | \sigma T_{eff} (K) | N | log g (dex) | \sigma log g (dex) | N | [Fe/H] (dex) | \sigma [Fe/H] (dex) | N |
|--------------|------------|-------------|-------------|-------------------|---|-------------|-------------------|---|--------------|-------------------|---|
| BD+264251    | 21:43:57.12 | +27:23:24.00 | 5991        | 97                | 8 | 4.30        | 0.36              | 7 | –1.27        | 0.08              | 7 |
| HD102200     | 11:45:34.24 | –46:03:46.39 | 6119        | 52                | 10| 4.22        | 0.16              | 7 | –1.22        | 0.06              | 7 |
| HD106038     | 12:12:01.37 | +13:50:40.62 | 6012        | 68                | 9 | 4.36        | 0.09              | 4 | –1.31        | 0.04              | 4 |
| HD126681     | 14:27:24.91 | –18:24:40.44 | 5567        | 84                | 21| 4.59        | 0.17              | 12| –1.18        | 0.09              | 12|
| HD175305     | 18:47:06.44 | +74:43:31.45 | 5085        | 58                | 15| 2.49        | 0.25              | 13| –1.43        | 0.07              | 14|
| HD196892     | 20:40:49.38 | –18:37:33.62 | 5954        | 94                | 9 | 4.16        | 0.24              | 8 | –1.03        | 0.08              | 8 |
| HD201891     | 21:11:59.03 | +17:43:39.89 | 5883        | 68                | 35| 4.33        | 0.15              | 28| –1.05        | 0.08              | 28|
| HD218857     | 23:11:24.60 | –16:15:04.02 | 5119        | 40                | 7 | 2.50        | 0.34              | 6 | –0.91        | 0.09              | 7 |
| HD241253     | 05:09:56.96 | +05:33:26.75 | 5879        | 94                | 13| 4.35        | 0.15              | 9 | –1.06        | 0.06              | 9 |
| HD298986     | 10:17:14.88 | –52:29:18.71 | 6177        | 82                | 8 | 4.23        | 0.06              | 5 | –1.33        | 0.04              | 5 |

Table 1. General information on metal-poor benchmark candidates.

Notes. The stellar parameters for each star were compiled using the PASTEL database (Soubiran et al. 2010). The \( T_{eff} \), \( \sigma T_{eff} \), log g, \( \sigma \log g \), [Fe/H], and \( \sigma [\text{Fe/H}] \) represent the mean and dispersion of the stellar parameters from N references in the PASTEL database.
making direct interferometric $\theta_{LD}$ measurements difficult with current instruments. Thus for the moment, we have the only option to rely on the photometric calibrations for $\theta_{LD}$ and $F_{bol}$.

It is important to note that recent studies (e.g. Creevey et al. 2012, 2015) have indicated the $\theta_{LD}$-photometric relationship may underestimate the $\theta_{LD}$ particularly at low metallicities. This is likely because the $\theta_{LD}$-photometric relationship is often only constrained by less than a handful, around 2–3, metal-poor stars (e.g. see Fig. 5 of Kervella et al. 2004). Since the $T_{eff}$ is proportional to $\theta_{LD}^{-0.5}$, underestimating the $\theta_{LD}$ causes the $T_{eff}$ to be overestimated. We also made use of the IRFM because it has the advantage of including not only information from $V$ and $K$ but also a broad range of photometry improving the $T_{eff}$ estimates and predicted $\theta_{LD}$ (see Sect. 3.2).

We are also prompted to use the IRFM because there is a relatively large disagreement (on the order of 10% which causes differences in $T_{eff}$ of more than 300 K) between the $\theta_{LD}$ of the giant stars in our sample using the calibrations of van Belle (1999) and Di Benedetto (2005). The reason for this discrepancy is currently not clear. One explanation is that there are intrinsic errors in the procedures that were used to determine the fitted relations. For example, reddening was not taken into account when relating the photometric colours to $\theta_{LD}$ in the work of van Belle (1999) which in part could cause discrepancies in their fitted relationships. In addition, it is important to note that a weakness of using these relations is that they do not include dependencies on [Fe/H]. As a result, many of these relations perform best around solar metallicity, by construction.

### 3.2. Infrared Flux Method

In addition to the $\theta_{LD}$-photometric relationships used in the previous section, we also made use of the infrared flux method (IRFM). This is one of the least model-dependent techniques to determine effective temperatures in stars, and it was originally devised to obtain stellar angular diameters with an accuracy of a few percent (Blackwell & Shallis 1977; Blackwell et al. 1979, 1980). Our analysis is based on the implementation described in Casagrande et al. (2006, 2010).

The basic idea is to recover for each star its bolometric flux and infrared monochromatic flux, both measured on the Earth. Their ratio is then compared to that obtained from the two same quantities defined on a surface element of the star, i.e., the bolometric flux $\sigma_{bol}$ and the theoretical infrared monochromatic flux. The only unknown parameter in this comparison is $T_{eff}$, which can be obtained (often with an iterative scheme, as described further below). For stars roughly earlier than M-type, the theoretical monochromatic flux is relatively easy to compute because the near infrared region is largely dominated by the continuum, with a nearly linear dependence on $T_{eff}$ (Rayleigh-Jeans regime) and is largely unaffected by other stellar parameters such as metallicity and surface gravity. This minimizes any dependence on model atmospheres, and makes the IRFM complementary to most spectroscopic methods, where instead $T_{eff}$ is often degenerate with gravity and metallicity. Once the bolometric flux and the effective temperature are known, the limb-darkened angular diameter is self-consistently obtained from the IRFM.

Since most of the times fluxes are derived from multi-band photometry, the problem is ultimately reduced to a derivation of...
Fig. 2. Top Panel: The computed $\theta_{LD}$ for each star from the four $\theta_{LD}$-photometric relationships: (K04, Kervella et al. 2004) is represented by red squares, (VB99, van Belle 1999) is represented by black circles, (Be05, Di Benedetto 2005) is represented by green diamonds, (B14, Boyajian et al. 2014) is represented by blue stars. In addition, the infrared flux method (IRFM) is also displayed as (cyan triangles). Middle Panel: $\Delta \theta_{LD}$/$\theta_{LD,adopted}$ for each star. Here $\Delta \theta_{LD} = \theta_{LD} - \theta_{LD, adopted}$. The adopted $\theta_{LD}$ is that computed from the IRFM. Bottom Panel: Comparison of the $T_{\text{eff}}$ for each star, computed from the $\theta_{LD}$-photometric relationships, with the adopted value from the infrared flux method.

Fig. 3. The adopted $T_{\text{eff}}$ (top panel), log $g$ (middle panel), and [Fe/H] (bottom panel) of the metal-poor GBS candidate stars (black closed circles) compared with the values from the literature (open red circles) sourced from the PASTEL catalogue.

4. Determination of surface gravity

The surface gravity was determined using the same procedure as in Paper I. We briefly summarize this method below. The log $g$ was determined using the adopted relationship $g = GM / R^2$ where $G$ is Newton’s gravitational constant, $M$ is the mass of the star and $R$ is its radius. The radius of the star was estimated using the adopted $\theta_{LD}$ which is listed in Col. 4 of Table 3 and the parallax listed in Col. 11 of Table 2. The mass for each star was computed by fitting the stellar parameters to a set of

---

Article number, page 6 of 18
stellar evolutionary tracks. In this case, those of Yonsei-Yale\textsuperscript{4} were used (Yi et al. 2003; Demarque et al. 2004). The fitting procedure is described in Paper I. The luminosity was computed from the bolometric flux and parallax. The $T_{\text{eff}}$ used was the value adopted from the IRFM. Additionally, the input metallicity was initially assumed to be the mean value from the PASTEL database. The log $g$ does not change significantly when using the final metallicity values described in Sect. 6. More details about the specific inputs to the Y$^2$ models, comparison of the masses determined from other stellar evolutionary tracks (e.g. Padova, Bertelli et al. 2008, 2009), and a comparison of the log $g$ determined by this method and others can be found in Sects. 4 and 5 of Paper I. The middle panel of Fig. 3 indicates that the log $g$ values determined in this way are consistent with the literature values from the PASTEL database. The grid of stellar models were interpolated with respect to mass and metallicity. The mass was then determined by minimizing the difference between the interpolated models and the position of the star on the HRD. The main source of uncertainty in the log $g$ determined tends to be from the radius (and thus $\rho_l$) compared to the mass (see Sect. 4.1 and Appendix A of Paper I).

Figure 4 shows the locations of all stars in the HRD, together with Yonsei-Yale evolutionary tracks for different metallicities. Most of the stars cluster around the tracks for $0.8 M_{\odot}$, with two dwarfs and the most metal-poor giant at somewhat lower masses. The mass difference of successive tracks ($0.05 M_{\odot}$) corresponds to the typical uncertainty in mass.

5. Determination of metallicity

To determine the metallicity of the candidates, we analysed their spectra. Because these stars were selected from the PASTEL database, they have been previously studied and thus their spectra can be found in archives (see Table 4). Nine stars have previously been observed in the U580 setup of the UVES instrument and the spectra were downloaded from the ESO archives\textsuperscript{5}.

$$\text{Table 4. Spectra used for this study.}$$

| Star   | I     | Date$_{\text{obs}}$ | SNR   | $R_{\text{in}}$ | Program ID |
|--------|-------|---------------------|-------|----------------|------------|
| BD+264251 U 2003-08-09 286 45254 71.B-0529(A) |
| HD102200 U 2001-03-06 160 51690 67.D-0086(A) |
| HD106038 U 2004-03-28 254 45254 072.B-0585(A) |
| HD126681 U 2000-04-09 240 51690 65.L-0507(A) |
| HD175305 N 2010-03-16 150 80000 ... |
| HD196892 U 2005-10-15 268 45990 076.B-0055(A) |
| HD201891 U 2012-10-18 107 66320 090.B-0605(A) |
| HD218857 U 2001-10-09 102 56990 68.D-0546(A) |
| HD241253 U 2005-10-08 194 56990 076.B-0133(A) |
| HD298986 U 2000-04-09 173 51690 65.L-0507(A) |

Notes. The I, or instrument, is either the U580 setting for the UVES instrument on the Very Large Telescope (denoted by U) or the NARVAL instrument (denoted by N). We note that while the input resolution ($R_{\text{in}} = \lambda/\Delta\lambda$) varies depending on the instrument and setup, we convolved all spectra to a common value of $R = 40000$. In addition, all of the spectra have a spectral coverage of at least 4760 – 6840 Å.

Additionally, one star (HD175305) comes from the archive of the NARVAL spectrograph operated by the Télescope Bernard Lyot\textsuperscript{6}. The spectra were prepared in the same way as Paper II: they were normalised, corrected by radial velocity and convolved to the lowest common resolution ($R = 40000$), in the same fashion as in the rest of the GBS. Note the resolution is lower in this case compared to our previous study because we could not find the whole data set with higher resolution. In all cases, the signal to noise ratio (SNR) of the spectra is better than 100 pixel$^{-1}$. The spectra for these stars have been included in the GBS high-resolution spectral library\textsuperscript{7} and can be publicly accessed (for more details on the library consult Blanco-Cuaresma et al. 2014b).

The analysis was done as in Sects. 4.1, 4.2, and 4.3 of Paper III, namely we used several codes. In addition, we used common input material (spectra, atomic data for the line list, Fe, Fe lines, etc.) and fixed the $T_{\text{eff}}$ and log $g$ to their adopted values determined in Sect. 3 and 4, respectively. We made use of the 1D-LTE MARCS atmosphere models (Gustafsson et al. 2008) and a common set of pre-defined iron lines, which were selected from the “golden lines” for metal-poor stars of Paper III. We considered the lines used for HD140283, HD122563, HD84937, HD22879 and Gmb 1830. Then, by visual inspection, we ensured that these lines were present and unblended, in the the spectra of the new candidate stars, obtaining a final list of 131 FeI and FeII lines (see Table 4 from Paper III for the input atomic data). Individual lines used for each star can be found in Tables A1-A10 of the online material. For clarity and reproducibility, in appendix A, we outline the format of the online material.

In this work, we employed four methods, or nodes, to determine the metallicity. Two methods use the equivalent width (EW) technique which include: (1) Bologna – based on GALA developed by Mucciarelli et al. (2013) and (2) EPINARBO – based on FAMA developed by Magrini et al. (2013). Both of these methods measure the EWs of individual iron features using the DOOp code (Cantat-Gaudin et al. 2014) which is an automated wrapper for the DAOSPEC code (Stetson & Pancino 2006).

\textsuperscript{4} http://www.astro.yale.edu/demarque/yystar.html

\textsuperscript{5} http://archive.eso.org/cms.html

\textsuperscript{6} http://tblegacy.bagn.obs-mip.fr/

\textsuperscript{7} http://www.blanco-cuaresma.com/s/benchmarkstars/
Table 3. Adopted parameters for metal-poor benchmark candidates.

| Star        | $\theta_\text{LD}$ (mas) | $\sigma_{\theta_\text{LD}}$ (mas) | $F_\text{bol}$ ($10^{-11}$ W m$^{-2}$) | $\sigma F_\text{bol}$ ($10^{-11}$ W m$^{-2}$) | $T_\text{eff}$ (K) | $\sigma T_\text{eff}$ (K) | log $g$ (dex) | $\sigma \log g$ (dex) | $v_\text{mic}$ (km s$^{-1}$) | $v_\sin i$ (km s$^{-1}$) | $\sigma v_\sin i$ (km s$^{-1}$) |
|-------------|---------------------------|-----------------------------------|----------------------------------------|------------------------------------------|-----------------|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| *BD+264251  | 0.077                     | 0.001                             | 0.2759                                 | 0.0059                                   | 6129            | 80                | 4.41            | 0.16            | 1.40            | 2.81            | 2.10            |
| HD102200    | 0.138                     | 0.002                             | 0.9062                                 | 0.0041                                   | 6155            | 80                | 4.22            | 0.07            | 1.43            | 1.90            | 4.59            |
| HD106038    | 0.073                     | 0.001                             | 0.2477                                 | 0.0026                                   | 6121            | 80                | 4.55            | 0.14            | 1.39            | 0.00            | 0.00            |
| *HD126681   | 0.128                     | 0.002                             | 0.5543                                 | 0.0050                                   | 5640            | 80                | 4.64            | 0.11            | 1.16            | 0.00            | 0.00            |
| HD175305    | 0.447                     | 0.006                             | 4.3520                                 | 0.6752                                   | 5059            | 80                | 2.53            | 0.14            | 1.54            | 5.01            | 1.92            |
| *HD196892   | 0.178                     | 0.002                             | 1.4130                                 | 0.0233                                   | 6053            | 80                | 4.19            | 0.06            | 1.36            | 0.00            | 0.00            |
| HD201891    | 0.273                     | 0.004                             | 3.1154                                 | 0.0517                                   | 5948            | 80                | 4.30            | 0.04            | 1.29            | 2.93            | 5.31            |
| *HD218857   | 0.194                     | 0.003                             | 0.8841                                 | 0.0161                                   | 5162            | 80                | 2.66            | 0.32            | 1.58            | 3.14            | 6.80            |
| *HD241253   | 0.091                     | 0.001                             | 0.3642                                 | 0.0013                                   | 6023            | 80                | 4.22            | 0.18            | 1.34            | 1.89            | 4.50            |
| HD298986    | 0.073                     | 0.001                             | 0.2691                                 | 0.0067                                   | 6223            | 100               | 4.19            | 0.19            | 1.48            | 4.07            | 5.98            |

Notes. The $\theta_\text{LD}$ were computed as a part of the IRFM. In addition, the $T_\text{eff}$ and $F_\text{bol}$ in this table represents the adopted $T_\text{eff}$ and bolometric flux from the IRFM, respectively. We estimated $v_\text{mic}$ using the GES relationship of Bergemann and Hill. The uncertainty in $v_\text{mic}$ was conservatively assumed to be 0.20 km s$^{-1}$ for all stars. Stars with an asterisk (*) in Col. 1 are currently not recommended (see Sect. 6.1 for a star by star discussion on the recommendations.)

2008). The other two methods use spectral synthesis including: (1) BACCHUS/ULB – developed by T. Masseron (Masseron 2006) which made use of the Turbospectrum synthesis code (Alvarez & Plez 1998; Plez 2012) and (2) iSpec – developed by Blanco-Cuaresma et al. (2014a). For more details on these methods we refer the reader to Sect. 4.3 of Paper III, Sect. 3.3, Table 4 of Paper IV and the development papers cited above. The first three methods were also employed in our previous metallicity determination in Paper III, while all four methods were used to determine abundances of several elements for the GBS sample (Paper IV).

The initial metallicity for the analysis was considered to be [Fe/H] = −1.00 dex for all stars. The macroturbulence parameter, $v_{\text{mic}}$, was determined simultaneously with the iron abundance, in the same way as in Paper III. The macroturbulence parameter, $v_{\text{mic}}$, was set to the value determined by the GES $v_{\text{mic}}$ relationship (e.g. Smiljanic et al. 2014, Paper III, Bergemann et al., in prep).

We conducted a total of eight runs which included: the “main run” fixing the $T_\text{eff}$ and log $g$ and $v_{\text{mic}}$ to their adopted values and six “error” runs where these three fixed values were varied by their $\pm1\sigma$ uncertainties listed in Table 3. This was done to evaluate the impact of the $1\sigma$ uncertainty in the adopted parameters on the [Fe/H]. In addition, each node solved for the stellar parameters independently using its own procedure, in what we define as the “free run” or eighth run. We note that in the free run we do not require the different nodes to use the same procedure (e.g. $\alpha$ clipping outlying Fe lines, tolerances of conversion, line selection etc.). This test was done primarily to see how each node performed when not using fixed $T_\text{eff}$ and log $g$ parameters. We emphasize that some of these nodes, particularly the EW nodes, often require a much larger number of lines for best performance. Thus, we remind the reader that the results of the free analysis simply allow us to quantify, in a different way, the benefit of fixing the $T_\text{eff}$ and log $g$. We refer the reader to Paper III for a extensive discussion on this matter.

A node-to-node comparison of the [Fe/H] can be found in Fig. 5 where we plot the metallicity of each star (including the results for the GBS in Paper III) obtained by each node relative to the mean literature value from PASTEL database. We also sort the stars on the x-axis towards increasing metallicity. The y-axis of the figure is the $\Delta$[Fe/H]$_{\text{lit}}$, which is defined as [Fe/H]−[Fe/H]$_{\text{lit}}$, where [Fe/H] is the metallicity of the star determined by a specific node and [Fe/H]$_{\text{lit}}$ is the mean [Fe/H] from the PASTEL database. The name of the star is indicated on the bottom of the figure. We note that only 3 nodes (ULB, Bologna, and EPINARBO) of Paper III were included in this figure. These nodes are the same as in this work. Fig. 5 indicates that the metallicities from the different nodes for the metal-poor candidates have a standard deviation of 0.028. In addition, the values generally agree well with the literature with a mean offset of +0.04 dex. This is consistent with the offset (+0.04 dex) and standard deviation (0.07 dex) of the FG dwarfs among the GBS (Paper III). The typical node-to-node scatter for the candidate stars are comparable to the GBS in the same $T_\text{eff}$ regime. Again we note that the node abundances for each star were determined by averaging the abundances of each line.

NLTE-corrected metallicities for each star can be found in Col. 2 of Table 5. The uncertainty in [Fe/H] due to the uncertainty in $T_\text{eff}$, log $g$, and $v_{\text{mic}}$ can be found in Cols. 4, 5, and 6, respectively. The difference between the LTE and NLTE-corrected metallicity, $\Delta$(LTE), and the difference between the mean Fe I and Fe II abundance, $\Delta$(ion), is found in Col. 7 and 8 respectively. The line-to-line dispersion of Fe I, Fe II and the number of Fe I and Fe II lines used in the analysis are listed in Cols. 9, 10, 11, and 12, respectively. Table 5 indicates that the difference between Fe I and Fe II can be as high as 0.10 dex in the worst cases. The $\Delta$(ion) values are smaller than for some of the GBS, e.g. HD 122563, where $\Delta$(ion)$_{\text{HD122563}} = −0.19$ dex (Paper III). We note here that HD 122563 is more metal-poor, with [Fe/H] = −2.64, than the stars we consider in this paper. On the other hand, the NLTE corrections, which are on the order of 0.05 dex, are similar to those of the current set of GBS.

We remind the reader that the final metallicity was computed as a mean of NLTE-corrected Fe lines. The NLTE corrections were computed in the same way as Paper III, namely by interpolating over a grid of NLTE corrections outlined in Lind et al. (2012). For this calculation, the adopted parameters were used. When the NLTE correction for a given line is not available the median of the NLTE corrections is assumed. This is both reasonable and reliable because the NLTE corrections per line are very similar for a single star (e.g. Bergemann et al. 2012). The NLTE correction range from +0.020 to +0.064 dex.
Fig. 5. The $\Delta [\text{Fe/H}]_{\text{lit}} = [\text{Fe/H}] - [\text{Fe/H}]_{\text{lit}}$ for each star and node for our GBS candidates and the current GBS stars ordered by metallicity. The node symbols are as follows: (1) iSpec is represented as a blue open hexagon, (2) BACCHUS/ULB is represented as a magenta open diamond, (3) EPINARBO is represented as a green open star, and (4) Bologna is represented as a red open triangle. We do not display the nodes that are used in Paper III and not this work. The typical dispersion between the methods is on the order of $\pm 0.03$ dex while the typical offset between the literature and each method is on the order of $+0.04$ dex.

Table 5. Adopted $[\text{Fe/H}]$ for metal-poor benchmark candidates.

| Star          | $[\text{Fe/H}]$ | $\sigma_{\text{FeI}}$ | $\Delta(T_{\text{eff}})$ | $\Delta(\log g)$ | $\Delta(v_{\text{mic}})$ | $\Delta(\text{LTE})$ | $\Delta(\text{ion})$ | $\sigma_{\text{FeII}}$ | $N_{\text{FeI}}$ | $N_{\text{FeII}}$ |
|--------------|-----------------|----------------------|--------------------------|------------------|--------------------------|-----------------------|-----------------------|----------------------|-----------------|------------------|
| *BD+264251   | -1.23           | 0.07                 | 0.09                     | 0.02             | 0.02                     | -0.03                 | -0.05                 | 0.05                 | 63              | 8                |
| HD102200     | -1.12           | 0.07                 | 0.08                     | 0.01             | 0.01                     | 0.04                  | 0.02                  | 0.07                 | 58              | 8                |
| HD106038     | -1.25           | 0.08                 | 0.08                     | 0.01             | 0.01                     | 0.02                  | -0.03                 | 0.05                 | 66              | 7                |
| *HD126681    | -1.07           | 0.06                 | 0.05                     | 0.01             | 0.01                     | 0.01                  | 0.02                  | 0.05                 | 61              | 7                |
| HD175305     | -1.29           | 0.06                 | 0.06                     | -0.01            | -0.01                    | 0.06                  | 0.08                  | 0.04                 | 56              | 8                |
| *HD196892    | -0.93           | 0.05                 | 0.05                     | 0.01             | 0.01                     | 0.04                  | 0.03                  | 0.05                 | 68              | 8                |
| HD201891     | -0.97           | 0.06                 | 0.06                     | 0.00             | 0.00                     | 0.03                  | 0.07                  | 0.02                 | 68              | 8                |
| *HD218857    | -1.78           | 0.07                 | 0.11                     | 0.04             | 0.04                     | 0.06                  | 0.01                  | 0.05                 | 56              | 8                |
| *HD241253    | -0.99           | 0.06                 | 0.09                     | 0.01             | 0.01                     | 0.03                  | 0.09                  | 0.03                 | 66              | 7                |
| HD298986     | -1.26           | 0.07                 | 0.09                     | 0.01             | 0.02                     | 0.05                  | 0.04                  | 0.05                 | 66              | 6                |

Notes. The $[\text{Fe/H}]$ is the NLTE-corrected and is the recommended value for each star. The $\Delta(T_{\text{eff}})$ is the uncertainty in the $[\text{Fe/H}]$ due to the uncertainty in $T_{\text{eff}}$, $\Delta(\log g)$ is the uncertainty in the $[\text{Fe/H}]$ due to the uncertainty in $\log g$, and $\Delta(v_{\text{mic}})$ is the uncertainty in the $[\text{Fe/H}]$ due to the uncertainty in $v_{\text{mic}}$. $\Delta(\text{LTE})$ is the NLTE-corrected $[\text{Fe/H}]$ minus the LTE $[\text{Fe/H}]$, $\Delta(\text{ion}) = [\text{FeI}/\text{H}] - [\text{FeII}/\text{H}]$. The line-to-line dispersion of Fe I and Fe II are $\sigma_{\text{FeI}}$ and $\sigma_{\text{FeII}}$, respectively. Finally $N_{\text{FeI}}$ and $N_{\text{FeII}}$ are the number of Fe I and Fe II lines used for the analysis, respectively. Stars with an asterisk (*) in Col. 1 are currently not recommended (see Sect. 6.1 for a star-by-star discussion on the recommendations).

For each Fe I and Fe II line, run and star we have four measurements (one for each of the nodes) for the iron abundance, which can be found in the tables online. We note here that the EW measurements for the synthesis methods (ULB/BACCHUS, iSpec) are measured for completeness but are not used to measure the abundances. The Fe abundance for each of the selected “golden” lines, and its computed NLTE correction can also be found as part of the online material. A description of this online material can be found in appendix A.

6. Results and Discussion

In this section, we discuss, on a star-by-star basis the results of the stellar parameter analysis. We discuss the quality of each parameter for each star, separately. In addition, we describe the node-to-node variation in the stellar parameters. Finally we compare the adopted stellar parameters with those determined spectroscopically.

As in Paper III and Paper IV, we selected only the lines that were sufficiently strong to have reliable abundances and sufficiently weak to not saturate, that is, line strength or reduced...
equivalent width (REW) was in the range of $-6.0 \leq \text{REW} \leq -5.0$ where $\text{REW} = \log(\text{EW} / \lambda)$. For this selection the adopted equivalent width (EW) was computed by averaging over the four measurements. Among the selected lines, we computed the mean of the four Fe abundance measurements and calculated its NLTE correction consistent with Paper III and references therein.

To help facilitate the discussion, we plot the final NLTE-corrected abundances for each line and star in Fig. 6 and Fig. 7 using different symbols for neutral and ionised lines. Each star is indicated in a different set of right-left panels. For reference, the star’s name is listed in the right panel and its stellar parameters are indicated in the left panel. The left panels show the abundances as a function of REW while the right panels show the abundances as a function of excitation potential (EP). We performed linear fits to the neutral lines. The slope of the trend and its standard error are indicated at the top of each panel. A slope is considered to be significant if its absolute value is larger than the standard error. We also performed a linear fit to only high EP lines (with EP $\geq 2eV$). We choose this cut because the low-excitation transitions are thought to experience significantly larger departures from 1D, LTE compared to higher excitation transitions (e.g. Bergemann et al. 2012). The red dashed lines correspond to the mean abundances determined from ionised lines.

In Fig. 6 and Fig. 7, we find that three of the ten stars (HD126681, HD 218857, and HD 298986) have significant trends in REW and Fe abundance indicating an potential issue with their $v_{\text{mic}}$. Figs. 6 and 7 also indicates that six of the ten stars have significant trends in the Fe abundance and EP whether using all of the Fe i lines or using just the high-EP lines as suggested by Bergemann et al. (2012).

The criteria for recommending a GBS candidate are as follows: (1) The $T_{\text{eff}}$ derived from IRFM should be consistent with the $\theta_\text{LD}$-photometric calibrations, (2) the $T_{\text{eff}}$ determined via the IRFM and photometric calibrations should be consistent with the spectroscopic $T_{\text{eff}}$ (i.e. the correlation between EP and Fe abundance should be null), (3) the log $g$ determined via isochrone fitting (assuming the $T_{\text{eff}}$ from IRFM) should be consistent with the spectroscopic log $g$ (i.e. the mean abundance of Fe i should equal that of Fe ii). Finally all stars where there is large discrepancies between the recommended parameters and PASTEL (i.e. differences in $T_{\text{eff}}$ more than 500 K, log $g$ larger than 0.5 dex, [Fe/H] larger than 0.5 dex) are flagged as suspicious.

6.1. Star-By-Star Discussion

In this subsection we discuss the results star-by-star. For this discussion, we remind the reader that the adopted $T_{\text{eff}}$, determined via the IRFM, can be found in Col. 6 of Table 3. The adopted log $g$ is determined though relating the $\theta_\text{LD}$ and the mass. The mass is determined through isochrone fitting, using the $Y$ stellar evolutionary tracks, the adopted $T_{\text{eff}}$ and the mean [Fe/H]. The recommended NLTE-corrected [Fe/H], derived using four spectroscopic methods and the adopted $T_{\text{eff}}$ and log $g$, can be found in Col. 2 of Table 5.

We begin the discussion by comparing the adopted $T_{\text{eff}}$ with that of the mean value from the PASTEL database and determined by the four $\theta_\text{LD}$-photometric calibrations (van Belle 1999; Kervella et al. 2004; Di Benedetto 2005; Boyajian et al. 2014).

In addition, we evaluate the spectroscopic validity of the $T_{\text{eff}}$ by ensuring that the trend in the Fe abundance with EP is null. As a diagnostic, we compare the adopted $T_{\text{eff}}$ and $T_{\text{eff}}$ from the free run (described in Sect. 5). We note here that the results of the free run indicate that the EW methods tend to systematically underestimate the $T_{\text{eff}}$ and log $g$. A potential reason for this is that the EW methods are affected by the restriction of lines allowed to be used in this analysis while synthesis methods are less affected by this. In addition, there are stark differences in the EW and synthesis procedures (e.g. sigma-clipping, convergence threshold of the pipeline, etc.) that were not fixed during this test. We stress that this test is not attempting to quantify the performance of EW methods.

We then compare the adopted log $g$ with those determined from various means in the literature and from the free stellar parameter run. We test its validity by confirming that the Fe i and Fe ii abundance agree (ionisation balance). Next we compare the metallicity derived using the adopted $T_{\text{eff}}$ and log $g$ and that from the literature. The [Fe/H] from the literature in most cases assumes LTE while we tabulate the NLTE corrected metallicity. The NLTE correction listed in Table 5 is positive and thus may explain why in Fig. 3 our final NLTE-corrected [Fe/H] (filled black circles) are a bit larger than the literature (open red circles). These NLTE correction are on the order of 0.05 dex. We note here that these corrections are treated as an uncertainty in our results.

We also inspect the trend between REW and Fe abundance as a way to access the quality of the $v_{\text{mic}}$. As a general comment, the $T_{\text{eff}}$ determined using indirect data in all stars is systematically higher than the mean $T_{\text{eff}}$ from the PASTEL database (Fig. 3) and determined spectroscopically (Fig. 8). We compute the combined uncertainty in the [Fe/H] in the same way as Paper I (i.e. by quadratically summing all $\sigma$ and $\Delta$ CoIs. in Table 5).
addition, we remark as to whether the candidate can have direct $\theta_{\text{LD}}$ measurements from current optical or near-infrared interferometers including the the VLT Interferometer or the CHARA array (for a detailed description of such facilities and their $\theta_{\text{LD}}$ limitation see Dravins et al. 2012). Finally, using the above discussion we either recommend or not recommend the star as a new GBS candidate.
BD+264251

The adopted \( T_{\text{eff}} \) of this star is hotter than the mean literature value by 140 K (2\%). It is most discrepant from the \( T_{\text{eff}} \) derived via \((B-V)\) photometry in the work of Mishenina et al. (2000). In addition, the adopted \( T_{\text{eff}} \) for this star is in fair agreement with the temperature derived from the various photometric calibration of angular diameter (van Belle 1999; Kervella et al. 2004; Di Benedetto 2005; Boyajian et al. 2014). The \( T_{\text{eff}} \) from the free run output of this star is between 0.01\% and 6\% smaller than the adopted \( T_{\text{eff}} \) for the iSpec and EPINARBO nodes, respectively. The spectroscopic and adopted \( T_{\text{eff}} \) do not agree which is consistent with the significant trend in the Fe abundance as a function of EP (Fig 6). However, this trend can be resolved by varying the stellar parameters within the uncertainties. In particular, it may be resolved by reducing \( v_{\text{mic}} \) by 0.2 km s\(^{-1}\) (i.e. the assumed uncertainty in the \( v_{\text{mic}} \)).

The adopted \( \log g \) of this star is 0.1 dex (2\%) larger than the mean value from the PASTEL database. The Fe\( \text{i} \) and Fe\( \text{ii} \) lines agree to within 0.05 dex (Table 5). In addition, the most discrepant \( \log g \) from the literature is from Mishenina et al. (2000). In this study the \( \log g \) is derived from the ionisation balance however only making use of 20 Fe\( \text{i} \) and 5 Fe\( \text{ii} \) lines. We not
only make use of a method independent of spectroscopy for our adopted log g, we also find relatively good agreement between 63 Fe I and 8 Fe II lines.

The [Fe/H] derived from the spectrum assuming the adopted $T_{\text{eff}}$ and log g is 0.05 dex (4%) larger than the mean value from the PASTEL database and from 0.10 – 0.40 dex (or 8–32%) larger than the free run output of the ULB/BACCHUS and EPINARBO nodes, respectively. The combined uncertainty in [Fe/H] is on the order of ±0.15 dex. There is no significant correlation between Fe abundance with REW.

We do not recommend this star as a GBS candidate because of the discrepant photometry, ranging a total of 0.15 mag in V, which leads to relatively uncertain $T_\text{eff}$. Additionally, the agreement between the $T_{\text{eff}}$ from the $\theta_{1\text{D}}$-photometric relationships and the IRFM is in worse agreement than all of the other candidates. This uncertainty in $T_{\text{eff}}$ propagates to all other parameters. In addition, the predicted $\theta_{1\text{D}}$ of this star is 0.07 mas and thus will be impossible to measure directly with the current state-of-the-art interferometers (with limits on the order of 0.1 mas with the Cherenkov Telescope Array) and possibly future intensity interferometers (e.g. Fig. 1 of Dravins et al. 2012).

**HD102200**

The adopted $T_{\text{eff}}$ of this star is in excellent agreement (less than 1%) with other spectroscopic and photometric studies (e.g. Mashonkina et al. 2003; Gehren et al. 2004; Jonsell et al. 2005; Sousa et al. 2011). It is also in good agreement with the $T_{\text{eff}}$ derived from the various $\theta_{1\text{D}}$-photometric calibrations. We note that the $T_{\text{eff}}$ from the free run output of this star ranges between less than 0.1% and 6% from the adopted $T_{\text{eff}}$ for the iSpec and EPINARBO nodes, respectively. Additionally, the adopted $T_{\text{eff}}$ is consistent with the spectroscopic $T_{\text{eff}}$. This is indicated by the null trend in the [Fe/H] abundance as a function of EP validating the adopted $T_{\text{eff}}$.

The adopted log g of this star is in excellent agreement (less than 0.5%) with the mean value of studies collated in the PASTEL database. It is also in fair agreement with the free parameter run. The disagreement between the adopted value and the free parameter run ranges between 0.5% and 10% for the EPINARBO and Bologna methods, respectively. There is also very good agreement (within 0.02 dex) between mean Fe I abundance, determined from averaging 58 lines and, the average Fe II abundance, determined by averaging 8 Fe II lines. This indicates that the adopted log g is in good agreement with the spectroscopic log g.

The [Fe/H] derived from the spectrum assuming the adopted $T_{\text{eff}}$ and log g is −0.10 dex (10%) larger than the mean literature value and 0.17 – 0.35 dex (15–30%) larger than the [Fe/H] determined in the free spectroscopic run with the ULB/BACCHUS and EPINARBO methods, respectively (Fig. 8). However, it is important to keep in mind that both the free run and the bulk of the literature assumes LTE. The NLTE correction for this star is on the order of +0.05 dex (see Table 5). The combined uncertainty in [Fe/H] is on the order of ±0.13 dex. There is no significant correlation between Fe abundance with REW.

In light of good agreement between the adopted stellar parameters and the various literature sources, the spectroscopic validation, and the free run output, we recommend this star as a GBS candidate. In addition, its predicted $\theta_{1\text{D}}$ is 0.14 mas (twice as large as BD+264251). However, due to its faintness (V=8.8) it would be very challenging to achieve a direct estimate of the $\theta_{1\text{D}}$ of this star with current interferometers.

**HD106038**

The adopted $T_{\text{eff}}$ of this star agrees well (~ 2%) with the mean value from the literature (Alonso et al. 1996a; Nissen & Schuster 1997; Nissen et al. 2002; Ramírez & Meléndez 2005; Gratton et al. 2003; Casagrande et al. 2011). The most discrepant $T_{\text{eff}}$ from the literature is cooler than the adopted $T_{\text{eff}}$ by -900 K determined via (V−K)−$T_\text{eff}$ relations (Nissen et al. 2002). The adopted $T_{\text{eff}}$ is also in good agreement with the derived from the photometric calibration of $\theta_{1\text{D}}$ (~1%). Additionally, the $T_{\text{eff}}$ from the free run output of this star ranges between less than 0.1% and 6% from adopted $T_{\text{eff}}$ for the iSpec and EPINARBO nodes, respectively. The spectroscopic analysis showed that there is a null trend in the Fe abundance as a function EP. This indicates that the spectroscopic and adopted $T_{\text{eff}}$ are consistent with one another.

The adopted log g of this star is in good agreement (4%) with the mean value from the literature. It is also in good agreement the values determined from the free parameter run (between 0.5 – 18% for the EPINARBO and Bologna methods, respectively). The Fe I and Fe II are consistent with each other within –0.026 dex which indicates that the adopted and spectroscopic log g are in agreement.

The [Fe/H] derived assuming the adopted $T_{\text{eff}}$ and log g is ~ 0.05 dex (4%) larger than the mean from literature and 0.08 – 0.28 dex (15–30%) larger than the [Fe/H] determined from the free run from the ULB and EPINARBO methods, respectively (Fig. 8). In addition, the combined [Fe/H] uncertainty is on the order of 0.13 dex. Finally, There is no significant correlation between Fe abundance with REW.

We have shown that there is good agreement between the adopted stellar parameters and the various literature sources, the spectroscopic validation and the free run output. As a result we recommend this star as a GBS candidate. However, similar to BD+264251, this star has a predicted $\theta_{1\text{D}}$ of 0.07 mas making it impossible to observe with current interferometers.

**HD126681**

The adopted $T_{\text{eff}}$ is in excellent agreement (~ 1.2%) with the typical $T_{\text{eff}}$ found in the literature (e.g. Tomkin et al. 1992; Blackwell & Lynas-Gray 1998; Fullbright 2000; Nissen et al. 2002; Gratton et al. 2003; Reddy et al. 2006; Masana et al. 2006; Sousa et al. 2011). The most discrepant $T_{\text{eff}}$ is from the work of Reddy et al. (2006). The authors determine the $T_{\text{eff}}$ of their sample using Strömgren (b−y) photometry (e.g. Alonso et al. 1996b). However, we note that at the $T_{\text{eff}}$ of this star, the authors show (in their Fig. 6) that the difference in $T_{\text{eff}}$ determined by Strömgren (b−y) photometry and (V−K) photometry has a dispersion of at least 100 K. The adopted $T_{\text{eff}}$ is also consistent with those derived from the photometric calibrations of $\theta_{1\text{D}}$ within 100 K. The $T_{\text{eff}}$ from the free run output of this star ranges between less than 0.1% and 4% from adopted $T_{\text{eff}}$ for the iSpec and EPINARBO nodes, respectively. However, there is a significant trend in the [Fe/H] abundance as a function of EP. This trend cannot be resolved by accounting for the uncertainties in the stellar parameters. This indicates that the adopted $T_{\text{eff}}$ is not in good agreement with the spectroscopic $T_{\text{eff}}$.

The adopted log g of this star is in good agreement with the mean value from the PASTEL database (4%). It is also in good agreement with the free run output (between 0.5 – 18% for the EPINARBO and Bologna methods, respectively). In addition, the mean abundance of Fe I (using 61 neutral lines) and Fe II (using 7 ionised lines) agrees within 0.021 dex.
The [Fe/H] derived from the spectrum assuming the adopted $T_{\text{eff}}$ and log $g$ (described in Sect. 5) is $\sim 0.05$ dex (4%) larger than the mean literature value and 0.08 – 0.28 dex (15–30%) larger than the [Fe/H] determined in the free parameter run by the ULB and EPINARBO methods, respectively (Fig. 8). The NLTE corrections on the Fe abundance are on the order of $+0.02$ dex. The combined uncertainty in the [Fe/H] is on the order of 0.10 dex. We also found a significant correlation between Fe abundance withREW indicating that the $v_{\text{mic}}$ may not be adequate.

We do not recommend this star as a GBS candidate because we cannot validate its $T_{\text{eff}}$ using Fe $i$ ionisation/excitation balance. In addition the $v_{\text{mic}}$ must be changed in order to balance the correlation between Fe abundance andREW. The $\theta_{\text{LD}}$ of this star is on the order of 0.10 mas which would make it out of reach for current interferometers.

---

**HD175305**

The adopted $T_{\text{eff}}$ is in excellent agreement ($\sim 1\%$) with the mean literature value (e.g. Wallerstein et al. 1979; Alonso et al. 1996a; Nissen & Schuster 1997; Fulbright 2000; Burris et al. 2000; Ishigaki et al. 2012). While the most discrepant $T_{\text{eff}}$ in the literature, from Fulbright (2000), is more than 400 K cooler than the adopted $T_{\text{eff}}$, it is an outlier among many other studies. Disregarding this outlier study, the mean difference between the adopted $T_{\text{eff}}$ and the literature is 20 K. The $\theta_{\text{LD}}$ determined from the photometric calibration from van Belle (1999) is larger by nearly a factor of five compared to that of Di Benedetto (2005). This in turn causes the temperature to be discrepant by 250 K ($\sim 4.5\%$) between these calibrations. The adopted $T_{\text{eff}}$ is consistent with Di Benedetto (2005). The discrepancy between these two $\theta_{\text{LD}}$-photometric calibrations is in part what motivated using the IRFM as the adopted procedure. The adopted $T_{\text{eff}}$ and the $T_{\text{eff}}$ derived from the free run output agrees within $2\%$. There is a null trend in the [Fe/H] abundance as a function of EP indicating good agreement between the spectroscopic and adopted $T_{\text{eff}}$.

The adopted log $g$ of this star agrees within $1\%$ of the mean value from the literature and those determined from the free spectroscopic run (less than 15%). While there is an offset of 0.08 dex between the abundance of Fe $i$, determined from 56 neutral Fe lines, and Fe $ii$, determined from 8 ionised Fe lines, it can be resolved by taking into account the uncertainty in $T_{\text{eff}}$ and log $g$.

The [Fe/H] derived from the spectrum assuming the adopted $T_{\text{eff}}$ and log $g$ is $\sim 0.05$ dex (4%) larger than the mean from the PASTEL database and 0.08 – 0.28 dex (15–30%) larger than the [Fe/H] determined in the free spectroscopic run from the ULB and EPINARBO methods, respectively (Fig. 8). The combined uncertainty in the [Fe/H] is on the order of 0.14 dex. There is also a null correlation between the Fe abundance withREW.

Because of the good agreement (less than 2% in $T_{\text{eff}}$, 18% in log $g$ and less than 15% in [Fe/H]) between the various methods (i.e. the adopted, validation through Fe excitation/ionisation balance, free run output, and literature) of determining the stellar parameters, we recommend this star as a GBS candidate. In addition, the relatively large $\theta_{\text{LD}}$ of this star ($0.447 \pm 0.006$ mas), makes it possible to be observed in the near future with current interferometers.

**HD196892**

The adopted $T_{\text{eff}}$ is in good agreement (less than 2%) with the mean literature value (e.g. Axer et al. 1994; Jehin et al. 1999; Thévenin & Idiart 1999; Gratton et al. 2003; Jonsell et al. 2005; Sousa et al. 2011). The most discrepant $T_{\text{eff}}$ is from the work of Axer et al. (1994) where it is derived using Hα, Hβ, Hγ, and Hδ fitting. These authors note that there are likely systematic differences of their $T_{\text{eff}}$ with photometric values from other studies (e.g. Fuhrmann et al. 1994). This may, in part, explain the discrepancy. The $T_{\text{eff}}$ derived from the various photometric calibrations of $\theta_{\text{LD}}$ are consistent within 100 K of the adopted value. We note that the adopted $T_{\text{eff}}$ and that from the free run output agree within 4%. There is a significant trend in the [Fe/H] abundance as a function of EP indicating that the spectroscopic and adopted $T_{\text{eff}}$ disagree. This trend cannot be resolved accounting for the uncertainties in the parameters.

The adopted log $g$ of this star is in excellent agreement (less than 1%) with the mean value from the PASTEL database. In addition, it is consistent with the free run. The mean Fe $i$ abundance, derived using 68 neutral Fe lines is consistent (within 0.03 dex) of the Fe $ii$ abundance, derived from 8 ionised Fe lines. This indicates that the spectroscopic log $g$ is consistent with the adopted value.

The [Fe/H] derived from the spectrum is $\sim 0.1$ dex (10%) larger than the mean literature value and as much as 0.22 dex (25%) larger than the [Fe/H] determined from the free spectroscopic run (Fig. 8). We remind the reader that this is not taking into account the NLTE correction which in this star is on the order of $+0.04$ dex. The combined [Fe/H] uncertainty is on the order of $\pm 0.08$ dex. There is no significant correlation betweenREW and Fe abundance.

We do not recommend this star as a GBS candidate because of the statistically significant trend in Fe $i$ abundance and EP. In particular, this trend cannot be resolved varying the parameters within their uncertainties. In addition, the $\theta_{\text{LD}}$ of this star is on the order of 0.18 $\pm$ 0.002 mas making interferometric $\theta_{\text{LD}}$ measurements very challenging.

**HD201891**

This star has an adopted temperature that is in good agreement ($\sim 1\%$) with the typical value from other studies (e.g. Edvardsson et al. 1993; Fuhrmann et al. 1997; Israelian et al. 1998; Clementini et al. 1999; Thévenin & Idiart 1999; Chen et al. 2000; Zhao & Gehren 2000; Mishenina & Kovtyukh 2001; Qui et al. 2002; Ramírez & Meléndez 2005; Valenti & Fischer 2005; Reddy & Lambert 2008; Casagrande et al. 2011). In fact, of the 35 studies which are listed in the PASTEL database, only 7 have $T_{\text{eff}}$ that differ by more than 100 K from our adopted value. The most discrepant $T_{\text{eff}}$ is 260 K lower (Valenti & Fischer 2005) than the adopted $T_{\text{eff}}$. It is important to note that Valenti & Fischer (2005) determined the $T_{\text{eff}}$ of this star using a spectral fitting procedure. In addition, the $T_{\text{eff}}$ from Valenti & Fischer (2005) are well calibrated around solar $T_{\text{eff}}$ and metallicity, but get increasingly worse at low metallicities and high $T_{\text{eff}}$ (e.g. see Fig 11, top panel of Casagrande et al. 2011). The 1D-LTE assumption under which the $T_{\text{eff}}$ is determined through spectroscopy may also account, in part, for the discrepancy. The adopted $T_{\text{eff}}$ is also in good agreement with the four $\theta_{\text{LD}}$-photometric calibrations. The adopted $T_{\text{eff}}$ and free run $T_{\text{eff}}$ of this star are in fair agreement (within 4%). While HD201891 has a statistically significant correlation between Fe $i$ abundance and EP, this cor-
relation can effectively be resolved by varying the parameters within their uncertainty.

The adopted log g is in excellent agreement with the mean value from the literature. The log g is also consistent (between 0 – 10% level for the ULB and Bologna nodes, respectively) with the free run output. There is a slight discrepancy (at the 0.06 dex level) between the mean abundance neutral Fe (using 68 Fe i lines) and the mean abundance of ionised Fe (using 8 Fe ii lines). This discrepancy can be reduced to ~0.02 dex by varying the parameters within their uncertainties.

The derived [Fe/H] is 0.07 dex (8%) larger than the mean literature value and as much as 0.22 dex (23%) larger than the [Fe/H] from the free run output. The total NLTE correction is on the order of +0.03 dex. The combined uncertainty in [Fe/H] is ±0.10 dex. We also find no significant correlation betweenREW and Fe abundance.

We recommend this star as a GBS candidate. While we noted a statistically significant correlation between the Fe i abundance and EP, this can be resolved by taking into account the uncertainties on the parameters. In addition the discrepancy between the neutral and ionised Fe lines is also reduced to an acceptable level by accounting for the uncertainties in the parameters. Finally HD201891 has a relatively high θ iD, with θ iD = 0.273 ± 0.004, for a dwarf star and thus it may be possible with current interferometers to achieve an θ iD estimate for this star.

**HD218857**

The adopted T eff is in excellent agreement (typically less than 1%) with the literature (Axer et al. 1994; Pilachowski et al. 1996; Burris et al. 2000; Mishenina & Kovtyukh 2001; Ishigaki et al. 2012). The T eff derived from the photometric calibration on angular diameter from Di Benedetto (2005) is in excellent agreement with the adopted T eff. However, the photometric calibration of van Belle (1999), is ~250 K lower than the adopted value. We note that the T eff from the free run output of this star ranges between less than 0.1% and 8% from adopted T eff for the iSpec and EPINARBO nodes, respectively. HD218857 also has a statistically significant correlation between Fe i abundance and EP lines considering both high EP and all EP Fe i lines indicating that the T eff from spectroscopic techniques may be in tension with the values determined in Sect. 3. This trend cannot be resolved by varying the stellar parameters within the uncertainty.

The adopted log g of this star is ~0.1 dex (4%) larger than the typical value from the literature and as much as 1 dex larger (40%) than the value determined from the free run. However, the mean abundance of Fe i, determined using 56 Fe i lines, is within 0.01 dex of the mean abundance of Fe ii determined using 8 Fe ii lines.

The derived [Fe/H] is 0.13 dex (8%) larger than the mean literature value and as much as 0.48 dex (27%) larger than the [Fe/H] determined in the free spectroscopic run. The typical NLTE Fe corrections for this star are on the order of +0.06 dex. The combined uncertainty in [Fe/H] is on the order of 0.16 dex. In addition, we find a significant correlation betweenREW and Fe abundance indicating a potential issue with the \( v_{mic} \).

We do not recommend this star as a GBS candidate because of the significant trend in Fe i abundance and EP as well as the disagreement betweenFe i and Fe ii. This trend cannot be resolved through varying the T eff, log g, [Fe/H], and \( v_{mic} \) within their uncertainties. In addition, the combined uncertainty in [Fe/H] is 0.19 dex. While this uncertainty is on the order of 0.19 dex. This uncertainty is in excellent agreement with the mean value from the literature and as much as 0.22 dex (23%) larger than the [Fe/H] from the free run output. The total NLTE correction is on the order of +0.03 dex. There is no significant correlation betweenREW and Fe abundance.

We do not recommend this star as a GBS candidate because of the significant trend in Fe i abundance and EP as well as the disagreement betweenFe i and Fe ii. This trend cannot be resolved through varying the T eff, log g, [Fe/H], and \( v_{mic} \) within their uncertainties. In addition, the combined uncertainty in [Fe/H] is 0.19 dex. This uncertainty is not consistent with the adopted value. This ionisation imbalance is not resolved taking into account the uncertainties in the parameters. The adopted log g of this star is 0.14 dex (2%) less than the typical value from the literature and as much as 0.45 dex larger (10%) than the value determined from the free run. The mean Fe i abundance, derived using 66 neutral Fe lines does not agree well (at the 0.10 dex level) with the Fe ii abundance, derived from 7 ionised Fe lines. This indicates that the spectroscopic log g is not consistent with the adopted value. This ionisation imbalance is not resolved taking into account the uncertainties in the parameters. The adopted T eff derived is 0.06 dex (6%) larger than the mean literature value and up to 0.24 dex (25%) larger than the [Fe/H] determined in the free spectroscopic run. The NLTE Fe corrections are on the order of +0.03 dex. There is no significant correlation betweenREW and Fe abundance.

**HD298986**

The adopted T eff of this star is in excellent agreement (~ 1.5 – 1.7%) with typical values from other studies (e.g. Axer et al. 1994; Nissen et al. 2002; Mashonkina et al. 2003; Masana et al. 2006; Casagrande et al. 2010, 2011). The adopted T eff also agrees well with those derived from the \( \theta_{iD} \)-photometric calibrations. We note that the T eff from the free run output of this star agrees with the adopted T eff within 5%. Additionally, the adopted T eff is consistent with the spectroscopic T eff, as indicated by a null trend in the [Fe/H] abundance as a function of EP.

The adopted log g of this star is within 0.02 dex (less than 1%) of the typical value from the literature. The uncertainty in the log g is on the order of 0.19 dex. While this uncertainty is on the high end, it is not significantly larger than several current GBS including α Tau, α Cet, and γ Sge. However, these stars are very cool giants. It is also consistent with the spectroscopic
value as indicated by the agreement, on the order of 0.03 dex, of mean abundance of ionised (6 lines) and neutral iron (66 lines).

The derived [Fe/H] agrees within 0.06 dex (5%) of the mean from the PASTEL database and can be as much as 0.29 dex (23%) larger than the [Fe/H] determined from the free run (Fig. 8). The NLTE corrections for Fe are on the order of +0.05 dex. The combined uncertainty in the [Fe/H] is on the order of 0.13 dex. While we do find significant correlation betweenREW and Fe abundance, this is resolved by increasing the $v_{mic}$ within its uncertainty.

Given the good agreement between the adopted values determined semi-independent of spectroscopy and other studies, as well as consistent with Fe i ionisation and excitation balance, we recommend this star as a GBS candidate. The predicted angular diameter of this star is 0.07 mas and is below the detection limit of current interferometers.

A summary of the consistency checks we have outlined above can be found for each star in Table 6.

### 6.2. Recommendations

From the above discussion, we recommend the following metal-poor stars as GBS candidates for calibration and validation purposes: HD102200, HD106038, HD175305, HD201891, and HD298986. A summary of the consistency checks and discussion can be found in Table 6. The other five stars do not pass the primary criteria for good GBS candidates. In most cases, these stars are not recommended due to not being able to validate (through Fe excitation balance) the $T_{eff}$ of the star. The stars BD+264251, HD126681, HD196892, HD218857, and HD241253 are denoted with an asterisk in Table 3 and 5 to indicate that they are not recommended as GBS candidates.

### 7. Summary and Conclusions

In this paper, we make an analysis of a sample of well-studied metal-poor stars in order to evaluate which of them can be included as *Gaia* benchmark stars. The GBS are a necessary set of calibrator stars that have already been invaluable in the era of large spectroscopic surveys. These surveys (e.g. *Gaia*-ESO, GALAH, and others) use them to calibrate their automated stellar parameter pipeline. As the astronomical community continues to lean towards even larger spectroscopic surveys (e.g. 4MOST and WEAVE) the need for improved samples of GBS will increase. Therefore, the aim of this paper was to add stars to the metal-poor gap defined by $-2.0 < [\text{Fe/H}] < -1.0$ dex. We initially began with 21 stars all within the desired metallicity range, however, only 10 stars remained for spectral analysis of which 5 were ultimately recommended for calibration purposes (details on their selection and all quality control cuts can be found in Sect. 2). Six of the ten stars in our sample were initially suggested in Appendix B of Paper I. In this work we, performed an analysis on the stellar parameters that are consistent with the previous set of GBS.

We used up to four $\theta_{LD}$-photometric calibrations to estimate the $\theta_{LD}$ using the broad band photometry available for each star. The bolometric fluxes were computed also using photometric calibrations. This procedure has been also employed for 6 stars (20%) in the current GBS (Paper I). These together were used to determine the $T_{eff}$ of each star using the adopted Stefan-Boltzmann law. The $\theta_{LD}$-photometric calibrations of the two giant stars in our sample produced results that disagreed at the 10% level (leading to a $T_{eff}$ discrepancy of $\sim$300 K). As such, we also employed the IRFM to estimate the $T_{eff}$. We found very good agreement of the $T_{eff}$ between the IRFM and the four $\theta_{LD}$-photometric calibrations. The log $g$ for the stars was computed by fitting a stellar evolutionary track (from the $Y^2$ set).

The ESO and NARVAL archival spectra were then employed to derive the [Fe/H] for the stars. We processed (e.g. continuum normalised, convolved to common resolution of $R = 40000$, etc.) these spectra in the same way as described in Paper II. We used a set of 131 Fe i and Fe ii lines from Paper III and four separate methods (nodes) to compute the [Fe/H]. There were 2 ‘equivalent width’ nodes (EPINARBO and Bologna) and 2 spectral synthesis codes (BACCHUS and iSpec) that were used in Paper III and IV. We employed seven separate runs per node which consisted of: a main run where the $T_{eff}$, log $g$, and $v_{mic}$ were fixed to their adopted value determined from the procedures outlined in Sects. 3 and 4, and six ‘error’ runs which varied each of the three parameters by $\pm1\sigma$ of their uncertainties. The ‘error’ runs were used to evaluate the impact of the uncertainties in the adopted derived stellar parameters on the [Fe/H] analysis.

The final combined metallicity was computed as the average of that from the four nodes. The metallicity-EP and metallicity-REW plots (shown in Fig. 6 and Fig. 7) were used to validate the stellar parameters on the basis of the standard Fe i ionisation/excitation balance method. We also used Fig. 6 and Fig. 7 in our discussion of the results and the star-by-star analysis noting the consistency of the adopted and spectroscopic parameters in Sect. 6.1. We found that five of the ten stars (HD102200, HD106038, HD175305, HD201891, and HD298986) have stellar parameters which are consistent between the photometric methods and the spectroscopic analysis. In Sect. 6, we evaluate the parameters in the context of the literature.

We present, in Table 5, the recommended parameters of the metal-poor GBS candidates and correspond to those which do not have an asterisk. The typical uncertainties in $T_{eff}$, log $g$, and [Fe/H] are $\pm80$ K, $\pm0.14$ dex, and $\pm0.13$ dex, respectively. While these uncertainties are marginally higher compared to the current set of FGK GBS, this is likely a result of not having a direct measurement of the $\theta_{LD}$. We recommend all stars with large angular diameters (particularly HD175305, HD201891, and HD102200) to be included in future interferometric $\theta_{LD}$ studies. In fact, HD175305 and HD201891 can, in principle, be observed with current interferometers (Table 6) and a possible extension of this work is to obtain a direct $\theta_{LD}$ measurement for these two stars. Direct measurement on the $\theta_{LD}$ is what will be
needed to improve their accuracy so that they can take their place among stars with the highest quality parameters to calibrate the next generation of surveys.

The recommended metal-poor candidates in this paper are dominated by stars within the metallicity range of $-1.3 < \frac{\text{[Fe/H]}}{\text{dex}} < -1.0$ dex. This is a critical metallicity regime because it is the interface of several Galactic components, such as the thick disk, the accreted halo, the inner halo and potentially the metal-poor tail of the thin disk. Furthermore, there is a lack of recommended GBS at these metallicities. With this work, we have decreased the $\sim 1$ dex metallicity gap by 30% and provided the astronomical community with these urgently needed calibration stars.

In addition, In Paper IV it was shown that a line-by-line differential approach, whereby the abundance of the star of interest is compared directly with the abundance of a reference star, to derive the metallicity yields more precise results. This could be done with Fe as well to improve the precision of the metallicity values. This was not done in the present work to remain consistent with Paper III which derived the metallicity in an absolute way. Redoing the metallicity analysis of all of the GBS in a differential framework will undoubtedly improve the precision of the derived metallicities and is planned in the near future. Therefore we stress that this work was a first step. We will soon have a new version of the PASTEL catalogue (Soubiran in prep) and more precise parallaxes from Gaia which will certainly significantly increase the number of metal-poor candidate benchmark stars.

Appendix A: Description of online table

For clarity and reproducibility of our analysis we are providing ten online tables. There is one table per star, each of which contains the information, on a line-by-line basis to reproduce this work. These tables have the same format and structure. Table A.1 displays the structure of the online tables which can be found in electronic format the CDS.

| Column | Label | Unit |
|--------|-------|------|
| (1)    | Element |      |
| (2)    | Absorption line wavelength | Å  |
| (3)    | Mean EW | mÅ  |
| (4)    | Mean Abundance (A) | dex |
| (5)    | NLTE correction | dex |
| (6)    | EW (EPI) | mÅ  |
| (7)    | EW (BOL) | mÅ  |
| (8)    | EW (ULB) | mÅ  |
| (9)    | EW (iSpec) | mÅ  |
| (10)   | A(EPI) | dex |
| (11)   | A(BOL) | dex |
| (12)   | A(iSpec) | dex |

Notes. This table is only available in electronic form at CDS. For the EW and abundances, the node is noted in the parentheses. For example EW (EPI) denotes the EW measurement of a specific line from the EPINARBO node while A(BOL) is the log(abundance) of a specific line for the Bologna node. (*) In the online table, the lines with NLTE corrections of $-0.000$ are those that do not have corrections available. This is done for identification purposes. In these cases, the median of the NLTE corrections of the other lines is assumed.

Acknowledgements. We would like to thank the referee for constructive comments that helped improve the presentation of this work. We would like to thank A. Case for discussions that improved this work. K.H. is funded by the British Marshall Scholarship program and the King’s College, Cambridge Studentship. This work was partly supported by the European Union FP7 programme through ERC grant number 320360. U.H. acknowledges support from the Swedish National Space Board (Rymdstyrelsen).

References

Alonso, A., Arribas, S., & Martinez-Roger, C. 1994, A&AS, 107, 365
Alonso, A., Arribas, S., & Martinez-Roger, C. 1995, A&A, 297, 197
Alonso, A., Arribas, S., & Martinez-Roger, C. 1996a, A&AS, 117, 227
Alonso, A., Arribas, S., & Martinez-Roger, C. 1996b, A&A, 313, 873
Alvarez, R. & Plez, B. 1998, A&A, 330, 1109
Axer, M., Fuhrmann, K., & Gehren, T. 1994, A&A, 291, 895
Bailey-Jones, C. A. L., András, R., Arcay, B., et al. 2013, A&A, 559, A74
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Bergemann, M., Lind, K., Collet, R., Magic, Z., & Asplund, M. 2012, MNRAS, 427, 27
Bertelli, G., Girardi, L., Marigo, P., & Nasi, E. 2008, A&A, 484, 815
Bertelli, G., Nasi, E., Girardi, L., & Marigo, P. 2009, A&A, 508, 355
Blackwell, D. E. & Lysen GRay, A. E. 1998, A&AS, 129, 505
Blackwell, D. E., Perford, A. D., & Shalvis, M. J. 1980, A&A, 82, 249
Blackwell, D. E. & Shalvis, M. J. 1977, MNRAS, 180, 177
Blackwell, D. E., Shalvis, M. J., & Selby, M. J. 1979, MNRAS, 188, 847
Blanco-Cuaresma, S., Soubiran, C., Heiter, U., & Jofre, P. 2014a, A&A, 569, A111
Blanco-Cuaresma, S., Soubiran, C., Jofre, P., & Heiter, U. 2014b, A&A, 566, A98, Paper II
Boeche, C. & Grebel, E. K. 2015, ArXiv e-prints
Boyajian, T. S., van Belle, G., & von Braun, K. 2014, AJ, 147, 47
Borris, D. L., Pilachowski, C. A., Armandroff, T. E., et al. 2000, ApJ, 544, 302
Cantat-Gaudin, T., Donati, P., Pancino, E., et al. 2014, A&A, 562, A10
Carpenter, J. M. 2001, AJ, 121, 2851
Casagrande, L., Portinari, L., & Flynn, C. 2006, MNRAS, 373, 13
Casagrande, L., Portinari, L., & Glass, I. S., et al. 2014, MNRAS, 439, 2060
Casagrande, L., Ramirez, I., Meléndez, J., Bessell, M., & Asplund, M. 2010, A&A, 512, A54
Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, A&A, 530, A138
Dekker, H., D’Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, Proc. SPIE, Vol. 4008, Optical and IR Telescope Instrumentation and Detectors, ed. M. Yee & A. F. Moorwood, 534–545
Demarque, P., Woo, J.-H., Kim, Y.-C., & Yi, S. K. 2004, ApJS, 155, 667
Di Benedetto, G. P. 1999, A&A, 339, 858
Di Benedetto, G. P. 2005, MNRAS, 357, 174
Dravins, D., LeBohec, S., Jensen, H., & Nuñez, P. D. 2012, New A Rev., 56, 143
Dravins, D., LeBohec, S., Jensen, H., & Nuñez, P. D. 2012, New A Rev., 56, 143
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Fuhrmann, K., Axer, M., & Gehren, T. 1994, A&A, 285, 585
Fuhrmann, K., Pfeiffer, M., Frank, C., Reetz, J., & Gehren, T. 1997, A&A, 323, 909
Fullbright, J. P. 2000, AJ, 120, 1841
Gehren, T., Liang, Y. C., Shi, J. R., Zhang, H. W., & Zhao, G. 2004, A&A, 413, 1045
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
González Hernández, J. I. & Bonifacio, P. 2009, A&A, 497, 497
Gratton, R. G., Carretta, E., & Castelli, F. 1996, A&A, 314, 191
Gratton, R. G., Carretta, E., Claudi, R., Lucatello, S., & Barbieri, M. 2003, A&A, 404, 187

Article number, page 17 of 18
Gratton, R. G., Sneden, C., Carretta, E., & Bragaglia, A. 2000, A&A, 354, 169
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Hawkins, K., Jofré, P., Masseron, T., & Gilmore, G. 2015, MNRAS, 453, 758
Hawkins, K., Masseron, T., Jofre, P., et al. 2016, ArXiv e-prints:1604.08800
Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, A&A, 582, A91, Paper IV
Ishigaki, M. N., Chiba, M., & Aoki, W. 2012, ApJ, 753, 64
Israelian, G., García López, R. J., & Rebolo, R. 1998, ApJ, 507, 805
Jehin, E., Magain, P., Neuforge, C., et al. 1999, A&A, 341, 241
Jofré, P., Heiter, U., Soubiran, C., et al. 2015, A&A, 582, A91, Paper III
Jonsell, K., Edvardsson, B., Gustafsson, B., et al. 2005, A&A, 440, 321
Kervella, P., Thévenin, F., Di Folco, E., & Ségransan, D. 2004, A&A, 426, 297
Lemasle, B., de Boer, T., Hill, V., et al. 2014, ArXiv e-prints
Lind, K., Bergemann, M., & Asplund, M. 2012, MNRAS, 427, 50
Magrini, L., Randich, S., Friel, E., et al. 2013, A&A, 558, A38
Masana, E., Jordi, C., & Ribas, I. 2006, A&A, 450, 735
Mashonkina, L. & Gehren, T. 2000, A&A, 364, 249
Mashonkina, L., Gehren, T., Travaglio, C., & Borkova, T. 2003, A&A, 397, 275
Masseron, T. 2006, PhD thesis, Observatoire de Paris, France
Meléndez, J., Casagrande, L., Ramírez, I., Asplund, M., & Schuster, W. J. 2010, A&A, 515, L3
Mermilliod, J.-C., Mermilliod, M., & Hauck, B. 1997, A&AS, 124, 349
Mishenina, T. V., Korotin, S. A., Klochkova, V. G., & Panchuk, V. E. 2000, A&A, 353, 978
Mishenina, T. V. & Kovtyukh, V. V. 2001, A&A, 370, 951
Mucciarelli, A., Pancino, E., Lovisi, L., Ferraro, F. R., & Lapenna, E. 2013, ApJ, 766, 78
Nissen, P. E., Primas, F., Asplund, M., & Lambert, D. L. 2002, A&A, 390, 235
Nissen, P. E. & Schuster, W. J. 1997, A&A, 326, 751
Nissen, P. E. & Schuster, W. J. 2010, A&A, 511, L10
Pilachowski, C. A., Sneden, C., & Kraft, R. P. 1996, AJ, 111, 1689
Plez, B. 2012, Turbospectrum: Code for spectral synthesis, astrophysics Source Code Library
Prochaska, J. X., Naumov, S. O., Carney, B. W., McWilliam, A., & Wolfe, A. M. 2000, AJ, 120, 2513
Qin, H.-M., Zhao, G., Takada-Hidai, M., et al. 2002, PASJ, 54, 103
Ramírez, I. & Meléndez, J. 2005, ApJ, 626, 446
Randich, S., Gilmore, G., & Gaia-ESO Consortium. 2013, The Messenger, 154, 47
Reddy, B. E. & Lambert, D. L. 2008, MNRAS, 391, 95
Reddy, B. E., Lambert, D. L., & Allende Prieto, C. 2006, MNRAS, 367, 1329
Schönrich, R. & Bergemann, M. 2014, MNRAS, 443, 698
Smiljanic, R., Korn, A. J., Bergemann, M., et al. 2014, A&A, 570, A122
Soubiran, C., Le Campion, J.-F., Cayrel de Strobel, G., & Caillo, A. 2010, A&A, 515, A111
Sousa, S. G., Santos, N. C., Israeliain, G., et al. 2011, A&A, 526, A99
Stetson, P. B. & Pancino, E. 2008, PASP, 120, 1332
Thévenin, F. & Idart, T. P. 1999, ApJ, 521, 753
Tomkin, J., Lemke, M., Lambert, D. L., & Sneden, C. 1992, AJ, 104, 1568
Valenti, J. A. & Fischer, D. A. 2005, ApJS, 159, 141
van Belle, G. T. 1999, PASP, 111, 1515
van Leeuwen, F. 2007, A&A, 474, 653
Venn, K. A., Irwin, M., Shetrone, M. D., et al. 2004, AJ, 128, 1177
Wallerstein, G., Pilachowski, C., Gerend, D., Baird, S., & Cantera, R. 1979, MNRAS, 186, 691
Yi, S. K., Kim, Y.-C., & Demarque, P. 2003, ApJS, 144, 259
Zhao, G. & Gehren, T. 2000, A&A, 362, 1077