

## *Gaia* FGK benchmark stars: new candidates at low metallicities<sup>★,★★</sup>

K. Hawkins<sup>1</sup>, P. Jofré<sup>1</sup>, U. Heiter<sup>2</sup>, C. Soubiran<sup>3</sup>, S. Blanco-Cuaresma<sup>4</sup>, L. Casagrande<sup>5</sup>, G. Gilmore<sup>1</sup>, K. Lind<sup>6</sup>,  
L. Magrini<sup>7</sup>, T. Masseron<sup>1</sup>, E. Pancino<sup>7,8,9</sup>, S. Randich<sup>7</sup>, and C. C. Worley<sup>1</sup>

<sup>1</sup> Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK  
e-mail: khawkins@ast.cam.ac.uk

<sup>2</sup> Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden

<sup>3</sup> Université de Bordeaux – CNRS, LAB – UMR 5804, BP 89, 33270 Floirac, France

<sup>4</sup> Observatoire de Genève, Université de Genève, 1290 Versoix, Switzerland

<sup>5</sup> Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, The Australian National University, ACT 2611, Australia

<sup>6</sup> Max-Planck Institute for Astronomy Konigstuhl 17, 69117 Heidelberg, Germany

<sup>7</sup> INAF/Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Firenze, Italy

<sup>8</sup> INAF/Osservatorio Astronomico di Bologna, via Ranzani 1, 40127 Bologna, Italy

<sup>9</sup> ASI Science Data Center, via del Politecnico snc, 00133 Roma, Italy

Received 7 February 2016 / Accepted 26 May 2016

### 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 has up to five metal-poor stars but no recommended stars within  $-2.0 < [\text{Fe}/\text{H}] < -1.0$  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 the previous works in this series 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  $[\text{Fe}/\text{H}]$  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 I 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  $[\text{Fe}/\text{H}]$  determined through several means. These stars, which are within  $-1.3 < [\text{Fe}/\text{H}] < -1.0$ , can be used for calibration and validation purpose of stellar parameter and abundance pipelines and should be of highest priority for future interferometric studies.

**Key words.** stars: fundamental parameters – techniques: spectroscopic – standards

## 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 ( $[\text{Fe}/\text{H}]$ ), 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

\* Based on data obtained from the ESO Science Archive Facility under request number pdjofre132105.

\*\* Full Appendix tables are only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](http://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/592/A70>

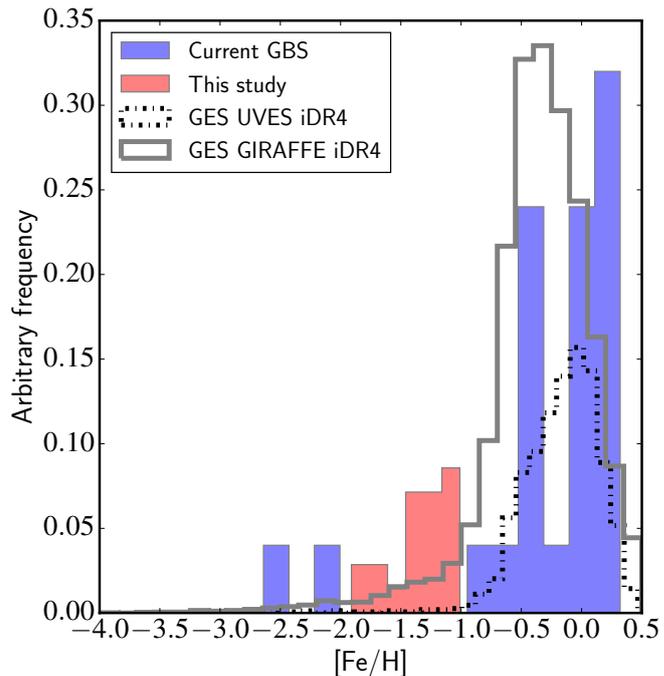
offsets between them in order to compare results. This work is about the assessment of such stars.

Beside astrometry, *Gaia* will produce, for most stars, atmospheric parameters of stars through a pipeline named APSIS (Bailer-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 (Heiter et al. 2015, henceforth Paper I). 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 benchmark 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-Cuadros 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; De Pascale et al. 2014; Lemasle et al. 2014; De Silva et al. 2015; Boeche & Grebel 2016; 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  $[\text{Fe}/\text{H}] \sim -1.0$  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  $[\text{Fe}/\text{H}] \sim -1.0$  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  $[\text{Fe}/\text{H}] < -1.0$  dex ( $\psi$  Phe, HD 122563, HD 84937, HD 140283, Gmb 1890). However among these five, three have not been recommended for calibration or validation purposes in Paper I. HD 140283 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



**Fig. 1.** The  $[\text{Fe}/\text{H}]$  distribution of the current GBS sample from Paper III (blue filled histogram) and the selected sample of metal-poor GBS candidates (red filled histogram). The GES iDR4 metallicity distribution from the UVES sample and GIRAFFE sample are shown as a black dash-dotted histogram and gray solid histogram, respectively.

the photometric and spectroscopic values. Gmb 1830 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  $-2.0 < [\text{Fe}/\text{H}] < -1.0$  dex. We aim to provide a set of new GBS candidate 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  $-1.3 < [\text{Fe}/\text{H}] < -1.0$  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 Hourihan et al., in prep.).

In this fifth work of the series, we define a new set of GBS candidate 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  $\theta_{LD}$  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)  $-2.0 < [\text{Fe}/\text{H}] < -1.0$  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 (HD 103095). The metallicity distribution of these 21 stars can be found as the red histogram in Fig. 1. We further required there to be known *BVJHK* photometry with defined uncertainties less than 0.15 mag in order to compute accurate photometric  $T_{\text{eff}}$ . This criterion reduced the sample to 17 stars, removing BD+053640, HD 199289, HD 134440. We also required there to be a known, and non-negative, parallax with a relative uncertainty better than 50%. This criterion removed four stars (HD 206739, HD 204543, HD 063791, HD 083212). Finally, we required spectra in the ESO and NARVAL archives. This last criterion removed four stars (HD 21581<sup>1</sup>, HD 023439A/B<sup>2</sup>). The stars BD+053640, HD 206739, HD 063791, and HD 083212 also do not have high-resolution spectra in the ESO/NARVAL archives.

This reduced the sample to the final version of 10 selected stars. We note that after the above cuts, we have mostly selected stars with  $-1.3 < [\text{Fe}/\text{H}] < -1.0$  dex, which is highly appropriate given that the interface of the thick disk, accreted halo, and possibly even the thin disk is within the regime (e.g. Nissen & Schuster 2010; Bensby et al. 2014; Hawkins et al. 2015) and there is a lack of such stars in the current GBS. Throughout the rest of this paper, we will consider and discuss only these 10 stars.

In Table 1, we present the basic collected information for the new candidates including the sky position (J2000 right ascension, RA, and declination, Dec) and the mean and dispersion of the stellar parameters taken from PASTEL. Additionally, the photometric and parallax information can be found in Table 2. Their *B*- and *V*-band photometry were taken from the General Catalogue of Photometric Data (henceforth GCPD, Mermilliod et al. 1997). Where the *B*- and *V*-band photometry was not

defined in the GCPD catalogue the Simbad database was used. The  $J_{2\text{MASS}}$  and  $K_{2\text{MASS}}$  magnitudes are sourced from the 2MASS catalogue (Cutri et al. 2003). The adopted reddening values,  $E(B - V)$  were taken from Meléndez et al. (2010), Casagrande et al. (2010, 2011). The parallax for each star was adopted from the updated analysis of the HIPPARCOS catalogue (van Leeuwen 2007). The tables have been separated by those stars which have been selected for further analysis and those which have not for clarity.

The final sample that we focus on in this paper contains those 10 metal-poor stars, all covering the metallicity regime that we are most interested in, namely metal-poor ( $[\text{Fe}/\text{H}] \sim -1.0$  dex) stars with an emphasis on dwarf stars. More than half of these stars were suggested in Appendix B of Paper I. The analysis presented here is consistent with the previous papers in the series (Papers I–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, 2000, 2003; Nissen & Schuster 1997; Mishenina et al. 2000; Fulbright 2000; 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  $H\alpha$  or  $H\beta$  (e.g. Axer 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

$T_{\text{eff}}$  was determined in two ways: (1) using  $\theta_{LD}$ -photometric calibrations (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. 2006, 2010). In Sect. 3.1 we describe the first procedure and in Sect. 3.2 we discuss the second procedure.

### 3.1. Deriving temperature using angular diameter-photometric relationships

To compute the  $T_{\text{eff}}$ , we used Eq. (1) of Paper I which relates the  $T_{\text{eff}}$  to the bolometric flux,  $F_{\text{bol}}$ , and the  $\theta_{LD}$ . We estimated

<sup>1</sup> This star now has spectra available in the ESO archive but was not public when the target selection for this project was completed.

<sup>2</sup> This is a spectroscopic binary system in which neither component had an ESO/NARVAL spectrum.

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

Star	RA (J2000)	Dec (J2000)	$T_{\text{eff}}$ (K)	$\sigma T_{\text{eff}}$ (K)	$N$	$\log g$ (dex)	$\sigma \log g$ (dex)	$N$	[Fe/H] (dex)	$\sigma[\text{Fe}/\text{H}]$ (dex)	$N$
Selected											
BD+264251	21:43:57.12	+27:23:24.00	5991	97	8	4.30	0.36	7	-1.27	0.08	7
HD 102200	11:45:34.24	-46:03:46.39	6119	52	10	4.22	0.16	7	-1.22	0.06	7
HD 106038	12:12:01.37	+13:15:40.62	6012	68	9	4.36	0.09	4	-1.31	0.04	4
HD 126681	14:27:24.91	-18:24:40.44	5567	84	21	4.59	0.17	12	-1.18	0.09	12
HD 175305	18:47:06.44	+74:43:31.45	5085	58	15	2.49	0.25	13	-1.43	0.07	14
HD 196892	20:40:49.38	-18:47:33.28	5954	94	9	4.16	0.24	8	-1.03	0.08	9
HD 201891	21:11:59.03	+17:43:39.89	5883	68	35	4.33	0.15	28	-1.05	0.08	28
HD 218857	23:11:24.60	-16:15:04.02	5119	40	7	2.50	0.34	6	-1.91	0.09	7
HD 241253	05:09:56.96	+05:33:26.75	5879	94	13	4.35	0.15	9	-1.06	0.06	9
HD 298986	10:17:14.88	-52:29:18.71	6177	82	8	4.23	0.06	5	-1.33	0.04	5
Not selected											
BD+053640	18:12:21.88	+05:24:04.41	5051	83	9	4.59	0.15	5	-1.20	0.10	5
HD 021581	03:28:54.48	-00:25:03.11	4889	61	11	2.15	0.20	6	-1.67	0.08	7
HD 023439A	03:47:02.12	+41:25:38.12	5059	73	13	4.51	0.11	13	-1.06	0.08	13
HD 023439B	03:47:02.63	+41:25:42.56	4808	70	6	4.55	0.09	6	-1.04	0.09	6
HD 063791	07:54:28.72	+62:08:10.76	4715	73	9	1.75	0.07	9	-1.68	0.08	9
HD 083212	09:36:19.95	-20:53:14.75	4512	55	15	1.37	0.32	13	-1.46	0.06	13
HD 103095	11:52:58.76	+37:43:07.23	5071	76	44	4.65	0.17	36	-1.34	0.10	38
HD 134440	15:10:12.96	-16:27:46.51	4817	82	16	4.61	0.11	11	-1.44	0.08	11
HD 199289	20:58:08.52	-48:12:13.45	5895	60	12	4.36	0.23	9	-1.01	0.06	10
HD 204543	21:29:28.21	-03:30:55.37	4667	66	15	1.30	0.22	11	-1.80	0.10	12
HD 206739	21:44:23.94	-11:46:22.84	4662	33	8	1.70	0.00	5	-1.58	0.02	5

**Notes.** The stellar parameters for each star were compiled using the PASTEL database (Soubiran et al. 2010). The  $T_{\text{eff}}$ ,  $\sigma T_{\text{eff}}$ ,  $\log g$ ,  $\sigma \log g$ , [Fe/H], and  $\sigma[\text{Fe}/\text{H}]$  represent the mean and dispersion of the stellar parameters from  $N$  references in the PASTEL database.

**Table 2.** Photometry and parallax of metal-poor benchmark candidates.

Star	$B$ (mag)	$\sigma B$ (mag)	$V$ (mag)	$\sigma V$ (mag)	$N$	$J_{2\text{MASS}}$ (mag)	$\sigma J$ (mag)	$K_{2\text{MASS}}$ (mag)	$\sigma K$ (mag)	$\pi$ (mas)	$\sigma \pi$ (mas)	$E(B - V)$ (mag)
Selected												
BD+264251	10.52	0.03	10.05	0.03	2	8.98	0.02	8.64	0.02	9.03	1.68	0.007
HD 102200	9.21	0.01	8.76	0.00	2	7.69	0.02	7.38	0.02	13.00	0.98	0.005
HD 106038	10.63	0.01	10.16	0.01	9	9.11	0.03	8.76	0.02	9.98	1.57	0.003
HD 126681	9.90	0.01	9.31	0.01	4	8.04	0.02	7.63	0.02	21.04	1.12	0.000
HD 175305	7.93	0.02	7.17	0.01	6	5.61	0.02	5.06	0.02	6.39	0.36	0.000
HD 196892	8.73	0.02	8.23	0.02	5	7.18	0.03	6.82	0.02	16.15	0.93	0.000
HD 201891	7.89	0.01	7.38	0.02	10	6.25	0.02	5.93	0.02	29.10	0.64	0.000
HD 218857	9.60	0.11	8.88	0.11	3	7.40	0.02	6.87	0.02	3.21	1.09	0.019
HD 241253	10.24	0.00	9.72	0.00	2	8.64	0.03	8.29	0.02	8.66	1.77	0.001
HD 298986	10.46	0.03	10.03	0.03	4	9.04	0.02	8.74	0.02	6.61	1.41	0.004
Not selected												
BD+053640	11.16	...	10.43	...	0	8.85	0.04	8.34	0.04	15.58	1.82	...
HD 021581	9.54	0.01	8.71	0.00	2	6.98	0.02	6.41	0.02	4.03	1.00	...
HD 023439A	8.93	0.01	8.18	0.02	4	6.62	0.02	6.12	0.02	46.65	2.63	...
HD 023439B	9.26	0.03	8.77	0.01	4	6.95	0.02	6.35	0.02	50.72	1.02	...
HD 063791	8.81	0.02	7.90	0.01	0	6.05	0.04	5.43	0.02	1.07	0.73	...
HD 083212	9.39	0.04	8.32	0.03	3	6.31	0.02	5.61	0.02	0.96	0.77	...
HD 103095	7.19	0.02	6.44	0.02	2	4.94	0.20	4.37	0.03	109.99	0.41	...
HD 134440	10.22	0.02	9.43	0.01	1	...	...	...	...	35.14	1.48	...
HD 199289	8.82	...	8.30	...	1	7.18	0.02	6.84	0.02	18.95	0.76	...
HD 204543	9.17	0.02	8.30	0.01	0	6.46	0.02	5.78	0.02	-0.13	1.08	...
HD 206739	9.45	0.03	8.60	0.02	0	6.70	0.02	6.03	0.02	1.93	1.17	...

**Notes.** The  $B$  and  $V$  magnitudes were sourced from the GCPD catalogue (Mermilliod et al. 1997) with  $N$  number of references. In cases where the  $B$  and  $V$  were not found in the GCPD catalogue, they were taken from the Simbad database. The  $J_{2\text{MASS}}$  and  $K_{2\text{MASS}}$  magnitudes are sourced from 2MASS (Cutri et al. 2003). All parallaxes were taken from a reanalysis of the HIPPARCOS catalogue (van Leeuwen 2007). The adopted reddening values were taken from Meléndez et al. (2010), Casagrande et al. (2010, 2011).

the  $F_{\text{bol}}$  using the photometric relationship outlined in Eqs. (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 (Cols. 7 and 9 in Table 2) bands into the Johnson system using the following relationship:

$$K_J = K_{2\text{MASS}} - 0.1277(J - K)_{2\text{MASS}} + 0.0460, \quad (1)$$

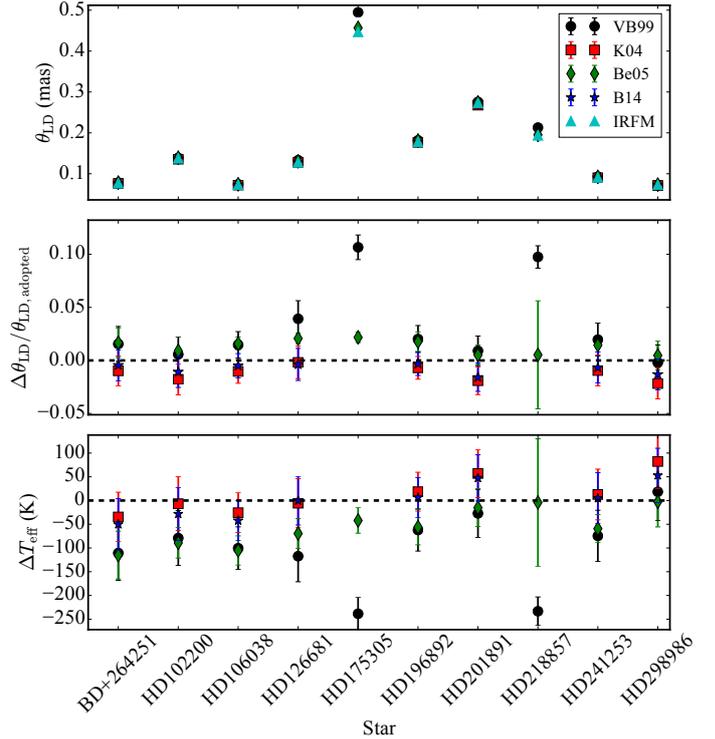
where  $K_J$ ,  $K_{2\text{MASS}}$  are the  $K$ -band magnitude in the Johnson and 2MASS systems, respectively. The 2MASS subscript refers to the 2MASS  $J$ ,  $K$ , and  $(J - K)$ , and the  $J$  subscript to Johnson system. This relationship was obtained by combining Eqs. (6), (7), (13), and (14) from [Alonso et al. \(1994\)](#) and Eqs. (12) and (14) from [Carpenter \(2001\)](#)<sup>3</sup>. The uncertainty in  $K_J$  was determined by propagating the uncertainty in the  $K_{2\text{MASS}}$  and  $(J - K)_{2\text{MASS}}$ . We note here that the photometry was corrected for reddening using the values in Col. 13 of Table 2. These corrections are very small and have the effect of changing the  $\theta_{\text{LD}}$  on the order of less than 1% and  $T_{\text{eff}}$  by less than 30 K when compared to the raw photometric values.

The  $\theta_{\text{LD}}$  was determined indirectly through photometric relationships. We have made use of four separate  $\theta_{\text{LD}}$ -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  $\theta_{\text{LD}}(B - V)$  relation (their Eq. (2)) and  $\theta_{\text{LD}}(V - K)$  relation (their Eq. (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  $\theta_{\text{LD}}(B - V)$  relation (their Eq. (22)) and  $\theta_{\text{LD}}(V - K)$  relation (their Eq. (23)). We note this procedure was used for the  $\theta_{\text{LD}}$  for the GBS HD 22879 and  $\epsilon$  For in Paper I. The third set of calibrations were from [Di Benedetto \(2005\)](#). We computed the  $\theta_{\text{LD}}$  of all stars using the  $\theta_{\text{LD}}(B - V)$  relation (their Eqs. (1) and (2)). The final set of calibrations used were from [Boyajian et al. \(2014\)](#). We made use of their  $\theta_{\text{LD}}(B - V)$  relation (their Eq. (4)) which is only applicable to the dwarf stars.

The results of the  $\theta_{\text{LD}}$  and  $T_{\text{eff}}$  computed using the various  $\theta_{\text{LD}}$ -photometric calibrations above can be found in Fig. 2. In top panel of Fig. 2, we compare the  $\theta_{\text{LD}}$ , in milliarcseconds (mas), of each star and relation used. We also plot the  $\theta_{\text{LD}}$  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  $\theta_{\text{LD}}$ -photometric relations with that computed from the IRFM. In the bottom panel of Fig. 2, we compare the  $T_{\text{eff}}$  derived from the different  $\theta_{\text{LD}}$ -photometric calibrations and that computed from the IRFM. In most cases the  $\theta_{\text{LD}}$  from each of the photometric calibrations are consistent (within  $1\sigma$ ) with each other and the  $\theta_{\text{LD}}$  from the IRFM.

As noted by Paper I, we choose to use the  $(V - K)$ - $\theta_{\text{LD}}$  relationships 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

<sup>3</sup> We note that Eqs. (12) and (14) from [Carpenter \(2001\)](#) have been updated in 2003. These updates can be found at <http://www.astro.caltech.edu/~jmc/2mass/v3/transformations/>. The difference between the 2001 and 2003 values is negligible. For example, the mean difference in  $K_J$  is 0.005 mag leading to a change in  $T_{\text{eff}}$  on the order of less than 8 K.



**Fig. 2.** *Top panel:* the computed  $\theta_{\text{LD}}$  for each star from the four  $\theta_{\text{LD}}$ -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:*  $\Delta\theta/\theta_{\text{adopted}}$  for each star. Here  $\Delta\theta_{\text{LD}} = \theta_{\text{LD}} - \theta_{\text{LD,adopted}}$ . The adopted  $\theta_{\text{LD}}$  is that computed from the IRFM. *Bottom panel:* comparison of the  $T_{\text{eff}}$  for each star, computed from the  $\theta_{\text{LD}}$ -photometric relationships, with the adopted value from the infrared flux method.

$\theta_{\text{LD}}$  of dwarf, subgiant, and giant stars determined via interferometry to their  $(V - K)_J$  colour and  $K_J$  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 (HD 175305) may soon have direct  $\theta_{\text{LD}}$  measurements, most of these stars are dim, making direct interferometric  $\theta_{\text{LD}}$  measurements difficult with current instruments. Thus for the moment, we have the only option to rely on the photometric calibrations for  $\theta_{\text{LD}}$  and  $F_{\text{bol}}$ .

It is important to note that recent studies (e.g. [Creevey et al. 2012, 2015](#)) have indicated the  $\theta_{\text{LD}}$ -photometric relationship may underestimate the  $\theta_{\text{LD}}$  particularly at low metallicities. This is likely because the  $\theta_{\text{LD}}$ -photometric relationship are 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  $T_{\text{eff}}$  is proportional to  $\theta_{\text{LD}}^{-0.5}$ , underestimating the  $\theta_{\text{LD}}$  causes the  $T_{\text{eff}}$  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  $T_{\text{eff}}$  estimate and predicted  $\theta_{\text{LD}}$  (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  $T_{\text{eff}}$  of more than 300 K) between the  $\theta_{\text{LD}}$  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  $[\text{Fe}/\text{H}]$ . As a result, many of these relations perform best around solar metallicity, by construction.

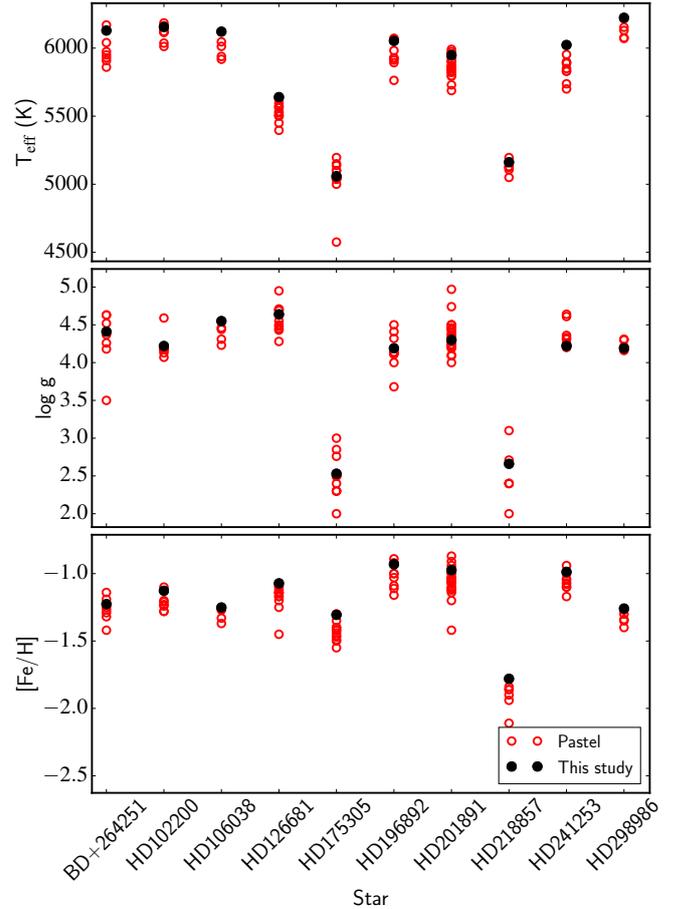
### 3.2. Infrared flux method

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  $\sigma T_{\text{eff}}^4$  and the theoretical infrared monochromatic flux. The only unknown parameter in this comparison is  $T_{\text{eff}}$ , 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  $T_{\text{eff}}$  (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  $T_{\text{eff}}$  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 fluxes in physical units, i.e. it depends on the photometric absolute calibration. Without exaggeration, this is the most critical point when implementing the IRFM, since it sets the zero-point of the  $T_{\text{eff}}$  scale. In our case, the absolute calibration has been anchored using solar twins, and the zero-point of the resulting effective temperature scale thoroughly tested (Casagrande et al. 2010, 2014; Datson et al. 2012, 2014).

For the sake of this work, the bolometric flux was recovered using multi-band photometry (Johnson-Cousins  $BV(RI)_C$  and 2MASS  $J_{2\text{MASS}}, H_{2\text{MASS}}, K_{2\text{MASS}}$ ) and the flux outside of these bands estimated using theoretical model fluxes from Castelli & Kurucz (2004). For each star  $[\text{Fe}/\text{H}]$  and  $\log g$  were fixed to the GBS recommended values. Whereas an iterative procedure was adopted in  $T_{\text{eff}}$ , starting with an initial guess, and iterating the IRFM until convergence within 1 K was reached. Despite all candidate GBS being relatively nearby, some of them might be slightly affected by extinction. When available, we adopted the reddening values derived from interstellar Na I D lines (Meléndez et al. 2010) or from Casagrande et al. (2010) or Casagrande et al. (2011) for the remaining cases.

Ultimately, we adopted the  $T_{\text{eff}}$  computed from the IRFM, as opposed to the  $\theta_{LD}$ -photometric calibrations, in large part because it provides a robust estimate of the  $\theta_{LD}$  for the two problematic giant stars making use of the available full broad band photometry rather than the  $(V - K)$  colour. In addition, in this way we have  $\theta_{LD}$  and  $T_{\text{eff}}$  from both giant and dwarf stars that are computed using a homogenous framework. We note that

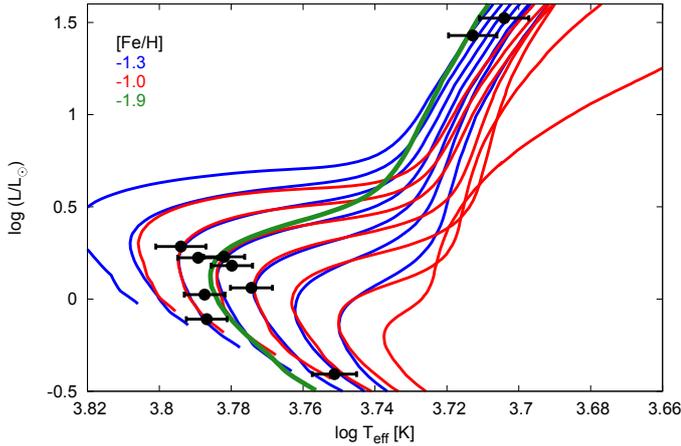


**Fig. 3.** The adopted  $T_{\text{eff}}$  (top panel),  $\log g$  (middle panel), and  $[\text{Fe}/\text{H}]$  (bottom panel) of the metal-poor GBS candidate stars (black closed circles) compared with the values from the literature (open red circles) sourced from the PASTEL catalogue.

for all dwarf stars, except for BD+264251, the IRFM temperature agrees very well with all four  $\theta_{LD}$ -photometric relationships described in Sect. 3.1 within the  $1\sigma$  uncertainty. This is also the case for the giant stars when considering the calibration of Di Benedetto (2005) but not that of van Belle (1999). The top panel of Fig. 3 shows the adopted  $T_{\text{eff}}$  from the IRFM with respect to literature values obtained from the PASTEL database. It indicates that the  $T_{\text{eff}}$  determined from the IRFM are systematically larger, by  $\sim 60$  K, compared to the mean  $T_{\text{eff}}$  from the PASTEL database. However, the IRFM is favored in this study over the PASTEL database because it is internally consistent. The reason for this minor discrepancy is unclear. The adopted  $F_{\text{bol}}$ ,  $\theta_{LD}$ , and  $T_{\text{eff}}$  and their uncertainties determined via the IRFM can be found in Table 3.

## 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. 1 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 stellar



**Fig. 4.** HRD showing the stellar parameters of the metal-poor benchmark candidates (black circles with error bars) superimposed on *Yonsei-Yale* evolutionary tracks for  $[\alpha/\text{Fe}] = +0.3$ , and  $[\text{Fe}/\text{H}] = -1.3$  (blue lines) and  $[\text{Fe}/\text{H}] = -1.0$  (red lines). Tracks are shown for masses increasing from  $0.60 M_{\odot}$  to the right to  $0.90 M_{\odot}$  to the left, in steps of  $0.05 M_{\odot}$ . The green line shows a track for  $[\text{Fe}/\text{H}] = -1.9$  and a mass of  $0.70 M_{\odot}$ , corresponding to the properties of the giant star HD 218857.

evolutionary tracks. In this case, those of *Yonsei-Yale*<sup>4</sup> were used (Y<sup>2</sup>, 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<sup>2</sup> 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  $\theta_{\text{LD}}$ ) 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<sup>5</sup>. Additionally, one star (HD 175305) comes from the archive of

the NARVAL spectrograph operated by the T el escope Bernard Lyot<sup>6</sup>. 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 ( $R = 40\,000$ ), 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  $100 \text{ pixel}^{-1}$ . The spectra for these stars have been included in the GBS high-resolution spectral library<sup>7</sup> 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 I, Fe II 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 HD 140283, HD 122563, HD 84937, HD 22879 and Gmb 1830. Then, by visual inspection, we ensured that these lines were present and unblended, in the spectra of the new candidate stars, obtaining a final list of 131 Fe I and Fe II lines (see Table 4 from Paper III for the input atomic data). Individual lines used for each star can be found at the CDS. 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 2008). The other two methods use spectral synthesis including: (1) BACCHUS/ULB – developed by 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  $[\text{Fe}/\text{H}] = -1.00$  dex for all stars. The macroturbulence parameter,  $v_{\text{mac}}$ , was determined simultaneously with the iron abundance, in the same way as in Paper III. The microturbulence 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  $\pm 1\sigma$  uncertainties listed in Table 3. This was done to evaluate the impact of the  $1\sigma$  uncertainty in the adopted parameters on the  $[\text{Fe}/\text{H}]$ . In addition, each node solved for the stellar parameters independently using its own procedure, in what we

<sup>4</sup> <http://www.astro.yale.edu/demarque/yystar.html>

<sup>5</sup> <http://archive.eso.org/cms.html>

<sup>6</sup> <http://tblegacy.bagn.obs-mip.fr/>

<sup>7</sup> <http://www.blancocuaresma.com/s/benchmarkstars/>

**Table 3.** Adopted parameters for metal-poor benchmark candidates.

Star	$\theta_{LD}$ (mas)	$\sigma\theta_{LD}$ (mas)	$F_{bol}$ ( $10^{-11} \text{ W m}^{-2}$ )	$\sigma F_{bol}$ ( $10^{-11} \text{ W m}^{-2}$ )	$T_{eff}$ (K)	$\sigma T_{eff}$ (K)	$\log g$ (dex)	$\sigma \log g$ (dex)	$v_{mic}$ ( $\text{km s}^{-1}$ )	$v \sin i$ ( $\text{km s}^{-1}$ )	$\sigma v \sin i$ ( $\text{km s}^{-1}$ )
*BD+264251	0.077	0.001	0.2759	0.0059	6129	80	4.41	0.16	1.40	2.81	2.10
HD 102200	0.138	0.002	0.9062	0.0041	6155	80	4.22	0.07	1.43	1.90	4.59
HD 106038	0.073	0.001	0.2477	0.0026	6121	80	4.55	0.14	1.39	0.00	0.00
*HD 126681	0.128	0.002	0.5543	0.0050	5640	80	4.64	0.07	1.16	0.00	0.00
HD 175305	0.447	0.006	4.3520	0.6752	5059	80	2.53	0.14	1.54	5.01	1.92
*HD 196892	0.178	0.002	1.4130	0.0233	6053	80	4.19	0.06	1.36	0.00	0.00
HD 201891	0.273	0.004	3.1154	0.0517	5948	80	4.30	0.04	1.29	2.93	5.31
*HD 218857	0.194	0.003	0.8841	0.1625	5162	80	2.66	0.32	1.58	3.14	6.80
*HD 241253	0.091	0.001	0.3642	0.0013	6023	80	4.22	0.18	1.34	1.89	4.50
HD 298986	0.073	0.001	0.2691	0.0067	6223	100	4.19	0.19	1.48	4.07	5.98

**Notes.** The  $\theta_{LD}$  were computed as a part of the IRFM. In addition, the  $T_{eff}$  and  $F_{bol}$  in this table represents the adopted  $T_{eff}$  and bolometric flux from the IRFM, respectively. We estimated  $v_{mic}$  using the GES relationship of Bergemann and Hill. The uncertainty in  $v_{mic}$  was conservatively assumed to be  $0.20 \text{ km s}^{-1}$  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).

**Table 4.** Spectra used for this study.

Star	I	Date <sub>obs</sub>	SNR ( $\text{pixel}^{-1}$ )	$R_{in}$	Program ID
BD+264251	U	2003-08-09	286	45 254	71.B-0529(A)
HD 102200	U	2001-03-06	160	51 690	67.D-0086(A)
HD 106038	U	2004-03-28	254	45 254	072.B-0585(A)
HD 126681	U	2000-04-09	240	51 690	65.L-0507(A)
HD 175305	N	2010-03-16	150	80 000	...
HD 196892	U	2005-10-15	268	45 990	076.B-0055(A)
HD 201891	U	2012-10-18	107	66 320	090.B-0605(A)
HD 218857	U	2001-10-09	102	56 990	68.D-0546(A)
HD 241253	U	2005-10-08	194	56 990	076.B-0133(A)
HD 298986	U	2000-04-09	173	51 690	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 ( $R_{in} = \lambda/\Delta\lambda$ ) varies depending on the instrument and setup, we convolved all spectra to a common value of  $R = 40\,000$ . In addition, all of the spectra have a spectral coverage of at least 4760–6840 Å.

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.  $\sigma$ -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  $T_{eff}$  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  $T_{eff}$  and  $\log g$ . We refer the reader to Paper III for an extensive discussion on this matter.

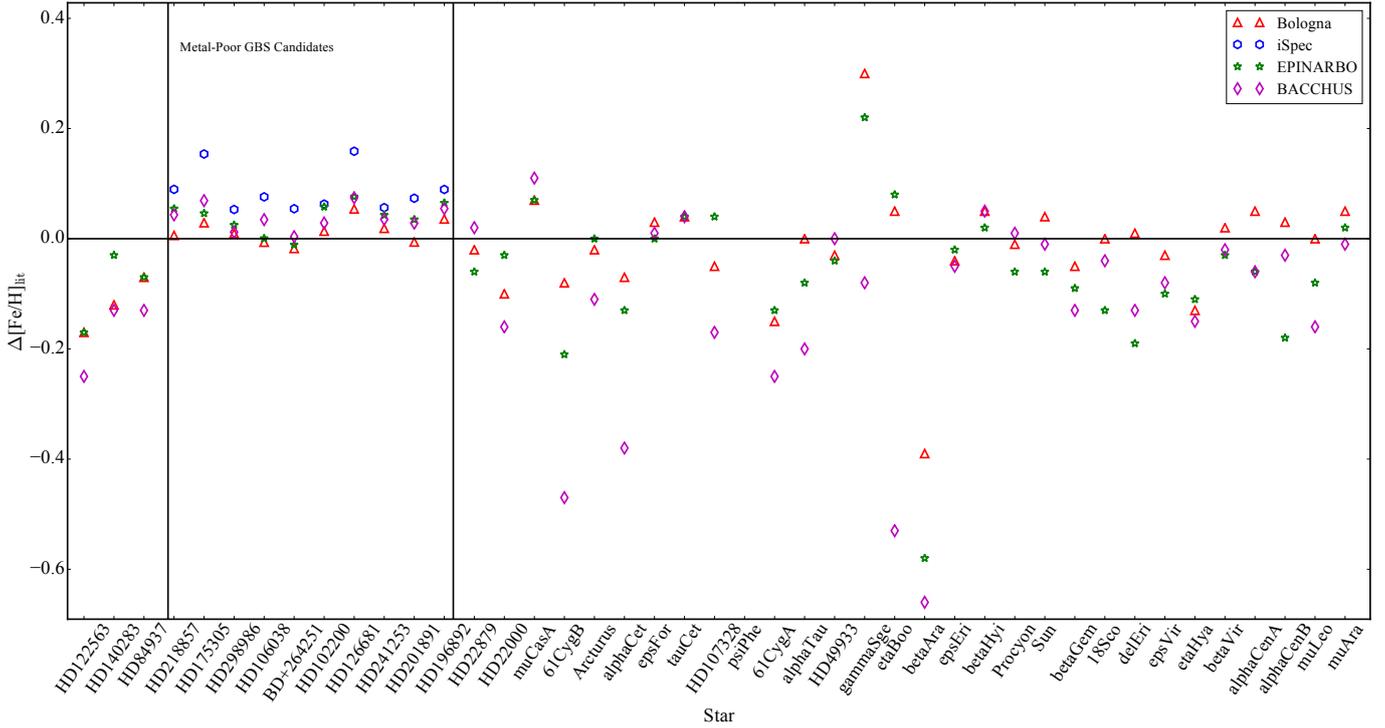
A node-to-node comparison of the [Fe/H] can be found in Fig. 5 where we plot the metallicity of each star (including the results for the GBS in Paper III) obtained by each node relative to the mean literature value from PASTEL database. We also sort the stars on the  $x$ -axis towards increasing metallicity. The  $y$ -axis of the figure is the  $\Delta[\text{Fe}/\text{H}]_{lit}$ , which is defined as  $[\text{Fe}/\text{H}] - [\text{Fe}/\text{H}]_{lit}$ , where [Fe/H] is the metallicity of the star determined by a specific node and  $[\text{Fe}/\text{H}]_{lit}$  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. Figure 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  $T_{eff}$  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  $T_{eff}$ ,  $\log g$ , and  $v_{mic}$  can be found in Cols. 4–6, respectively. The difference between the LTE and NLTE-corrected metallicity,  $\Delta(\text{LTE})$ , and the difference between the mean Fe I and Fe II abundance,  $\Delta(\text{ion})$ , is found in Cols. 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. 3, 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  $\Delta(\text{ion})$  values are smaller than for some of the GBS, e.g. HD 122563, where  $\Delta(\text{ion})_{HD122563} = -0.19$  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.

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 at the CDS. 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”



**Fig. 5.** The  $\Delta[\text{Fe}/\text{H}]_{\text{lit}} = [\text{Fe}/\text{H}] - [\text{Fe}/\text{H}]_{\text{lit}}$  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 the each method is on the order of  $+0.04$  dex.

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

Star	$[\text{Fe}/\text{H}]$	$\sigma_{\text{FeI}}$	$\Delta(T_{\text{eff}})$	$\Delta(\log g)$	$\Delta(v_{\text{mic}})$	$\Delta(\text{LTE})$	$\Delta(\text{ion})$	$\sigma_{\text{FeII}}$	$N_{\text{FeI}}$	$N_{\text{FeII}}$
*BD+264251	-1.23	0.07	0.09	0.02	0.02	0.03	-0.05	0.05	63	8
HD 102200	-1.12	0.07	0.08	0.01	0.01	0.04	0.02	0.07	58	8
HD 106038	-1.25	0.08	0.08	0.01	0.01	0.02	-0.03	0.05	66	7
*HD 126681	-1.07	0.06	0.05	0.01	0.01	0.01	0.02	0.05	61	7
HD 175305	-1.29	0.06	0.06	-0.01	-0.01	0.06	0.08	0.04	56	8
*HD 196892	-0.93	0.05	0.05	0.01	0.01	0.04	0.03	0.05	68	8
HD 201891	-0.97	0.06	0.06	0.00	0.00	0.03	0.07	0.02	68	8
*HD 218857	-1.78	0.07	0.11	0.04	0.04	0.06	0.01	0.05	56	8
*HD 241253	-0.99	0.06	0.09	0.01	0.01	0.03	0.09	0.03	66	7
HD 298986	-1.26	0.07	0.09	0.01	0.02	0.05	0.04	0.05	66	6

**Notes.** The  $[\text{Fe}/\text{H}]$  is the NLTE-corrected and is the recommended value for each star. The  $\Delta(T_{\text{eff}})$  is the uncertainty in the  $[\text{Fe}/\text{H}]$  due to the uncertainty in  $T_{\text{eff}}$ ,  $\Delta(\log g)$  is the uncertainty in the  $[\text{Fe}/\text{H}]$  due to the uncertainty in  $\log g$ , and  $\Delta(v_{\text{mic}})$  is the uncertainty in the  $[\text{Fe}/\text{H}]$  due to the uncertainty in  $v_{\text{mic}}$ .  $\Delta(\text{LTE})$  is the NLTE-corrected  $[\text{Fe}/\text{H}]$  minus the LTE  $[\text{Fe}/\text{H}]$ .  $\Delta(\text{ion}) = [\text{Fe I}/\text{H}] - [\text{Fe II}/\text{H}]$ . The line-to-line dispersion of Fe I and Fe II are  $\sigma_{\text{Fe I}}$  and  $\sigma_{\text{Fe II}}$ , respectively. Finally  $N_{\text{Fe I}}$  and  $N_{\text{Fe II}}$  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).

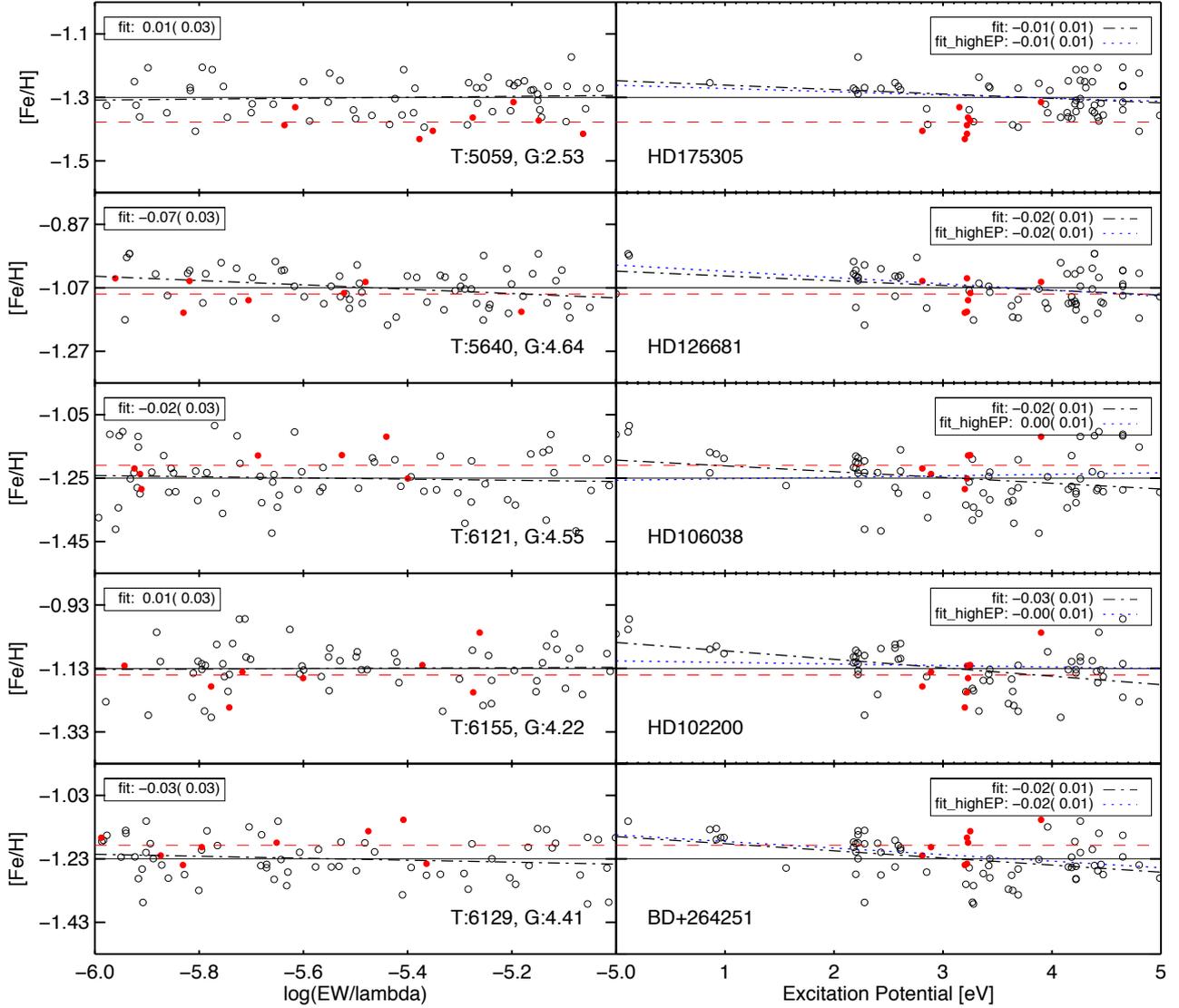
lines, and its computed NLTE correction can also be found at the CDS. A description of this online material can be found in Appendix A.

## 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 Papers III and 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  $-6.0 \leq REW \leq -5.0$  where  $REW = \log(EW/\lambda)$ . 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 Figs. 6 and 7 using



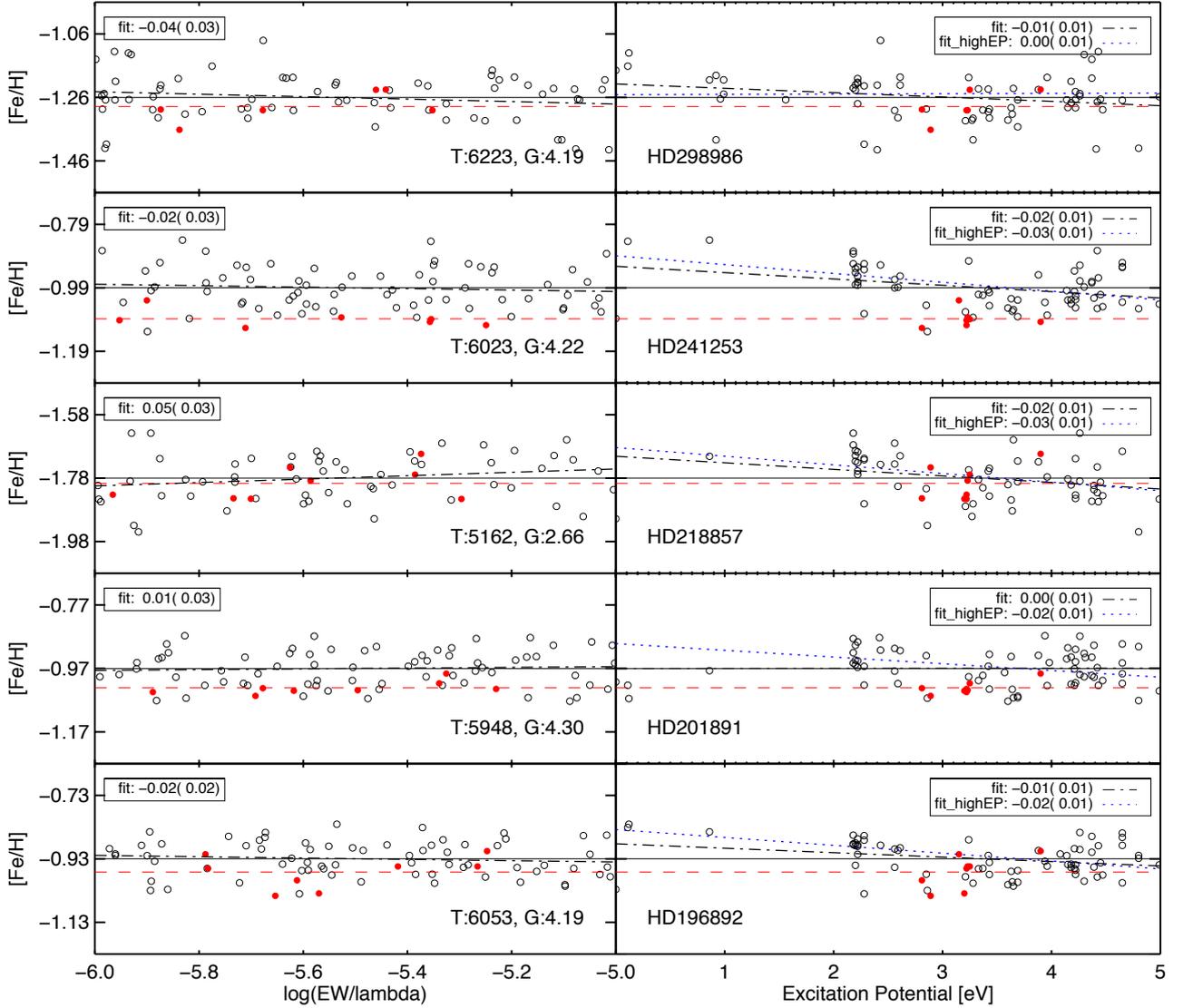
**Fig. 6.** Final iron abundances as a function of REW (*left panels*) and EP (*right panels*) for five stars (BD+264251, HD 102200, HD 106038, HD 126681, HD 175305) analysed in this work. Open circles indicate neutral lines while filled red circles indicated ionised lines. A linear regression fit to the neutral lines is performed for all lines (indicated with a black dash-dotted line) and for high-EP lines ( $EP \geq 2$  eV, indicated by a blue dotted line). 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. The effective temperature and surface gravity for each star is also indicated at the bottom of the left panels for reference. Dashed red line indicates the mean of the ionised lines.

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  $EP \geq 2$  eV). 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 Figs. 6 and 7, we find that three of the ten stars (HD 126681, HD 218857, and HD 298986) have significant

trends in REW and Fe abundance indicating an potential issue with their  $v_{\text{mic}}$ . Figures 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,  $[\text{Fe}/\text{H}]$  larger than 0.5 dex) are flagged as suspicious.



**Fig. 7.** Final iron abundances as a function of REW (*left panels*) and EP (*right panels*) for the remaining five stars (HD 196892, HD 201891, HD 218857, HD 241253, HD 298986) analysed in this work. Symbols used are the same as in Fig. 6.

### 6.1. Star-by-star discussion

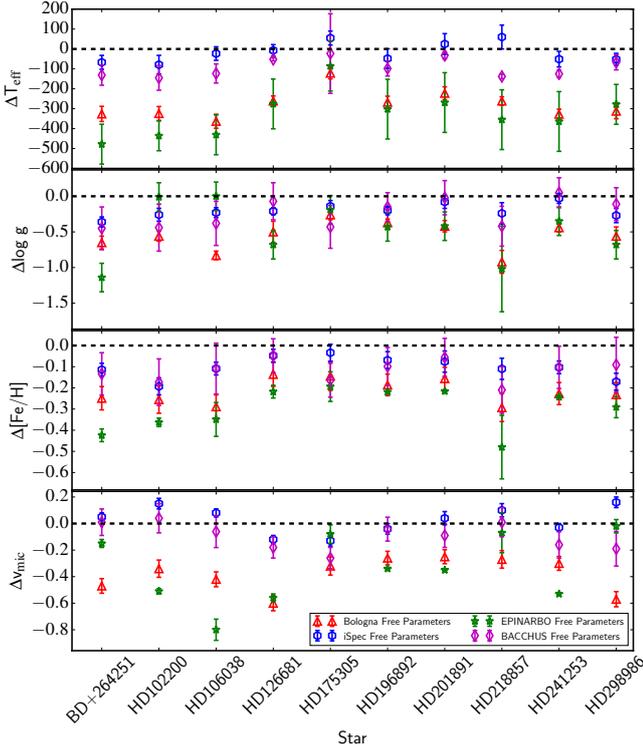
In this subsection we discuss the results star-by-star. For this discussion, we remind the reader that the adopted  $T_{\text{eff}}$ , determined via the IRFM, can be found in Col. 6 of Table 3. The adopted  $\log g$  is determined through relating the  $\theta_{\text{LD}}$  and the mass. The mass is determined through isochrone fitting, using the  $Y^2$  stellar evolutionary tracks, the adopted  $T_{\text{eff}}$  and the mean  $[\text{Fe}/\text{H}]$ . The recommended NLTE-corrected  $[\text{Fe}/\text{H}]$ , 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  $[\text{Fe}/\text{H}]$  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  $[\text{Fe}/\text{H}]$  (filled black circles) are a bit larger than the literature (open 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



**Fig. 8.** The  $\Delta T_{\text{eff}}$ ,  $\Delta \log g$ ,  $\Delta[\text{Fe}/\text{H}]$ , and  $\Delta v_{\text{mic}}$  computed from the free run (described in Sect. 5) for each star from top to bottom, respectively. We note here that the  $\Delta$  represents the difference of the node and adopted values for each parameter. For example,  $\Delta \log g = \log g_{\text{node}} - \log g_{\text{adopted}}$ . The BACCHUS (ULB) method is denoted as the open magenta diamond, the iSpec method is denoted as the open blue hexagon, the EPINARBO method is denoted as an open green star and the Bologna method is denoted as an open red triangle.

$T_{\text{eff}}$  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  $[\text{Fe}/\text{H}]$  in the same way as Paper I (i.e. by quadratically summing all  $\sigma$  and  $\Delta$  Cols. in Table 5). In addition, we remark as to whether the candidate can have direct  $\theta_{\text{LD}}$  measurements from current optical or near-infrared interferometers including 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  $0.2 \text{ km s}^{-1}$  (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 I and Fe 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 I and 5 Fe 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  $[\text{Fe}/\text{H}]$  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  $[\text{Fe}/\text{H}]$  is on the order of  $\pm 0.15$  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  $\theta_{\text{LD}}$ -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  $\theta_{\text{LD}}$  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).

#### HD 102200

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  $\theta_{\text{LD}}$ -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  $[\text{Fe}/\text{H}]$  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  $[\text{Fe}/\text{H}]$  derived from the spectrum assuming the adopted  $T_{\text{eff}}$  and  $\log g$  is  $\sim 0.10$  dex (10%) larger than the mean literature value and 0.17–0.35 dex (15–30%) larger than the  $[\text{Fe}/\text{H}]$  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  $[\text{Fe}/\text{H}]$  is on the order of  $\pm 0.13$  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  $\theta_{LD}$  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  $\theta_{LD}$  of this star with current interferometers.

#### HD 106038

The adopted  $T_{\text{eff}}$  of this star agrees well ( $\sim 2\%$ ) 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  $\sim 200$  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  $\theta_{LD}$  ( $\sim 1\%$ ). 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  $\sim 0.05$  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  $\theta_{LD}$  of 0.07 mas making it impossible to observe with current interferometers.

#### HD 126681

The adopted  $T_{\text{eff}}$  is in excellent agreement ( $\sim 1.2\%$ ) with the typical  $T_{\text{eff}}$  found in the literature (e.g. Tomkin et al. 1992; Blackwell & Lynas-Gray 1998; Fulbright 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 Strömgen ( $b-y$ ) 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 Strömgen ( $b-y$ ) 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  $\theta_{LD}$  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 with REW 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 and REW. The  $\theta_{LD}$  of this star is on the order of 0.10 mas which would make it out of reach for current interferometers.

#### HD 175305

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 outlying study, the mean difference between the adopted  $T_{\text{eff}}$  and the literature is 20 K. The  $\theta_{LD}$  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  $\theta_{LD}$ -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 with REW.

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  $\theta_{LD}$  of this star ( $0.447 \pm 0.006$  mas), makes it possible to be observed in the near future with current interferometers.

### HD 196892

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 H $\alpha$ , H $\beta$ , H $\gamma$ , and H $\delta$  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  $\theta_{LD}$  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 between REW 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  $\theta_{LD}$  of this star is on the order of  $0.18 \pm 0.002$  mas making interferometric  $\theta_{LD}$  measurements very challenging.

### HD 201891

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  $\theta_{LD}$ -photometric calibrations. The adopted  $T_{\text{eff}}$  and free run  $T_{\text{eff}}$  of this star are in fair agreement (within 4%). While HD 201891 has a statistically significant correlation between Fe I abundance and EP, this correlation 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  $\sim 0.02$  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  $\pm 0.10$  dex. We also find no significant correlation between REW 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 HD 201891 has a relatively high  $\theta_{LD}$ , with  $\theta_{LD} = 0.273 \pm 0.004$ , for a dwarf star and thus it may be possible with current interferometers to achieve an  $\theta_{LD}$  estimate for this star.

### HD 218857

The adopted  $T_{\text{eff}}$  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  $T_{\text{eff}}$  derived from the photometric calibration on angular diameter from Di Benedetto (2005) is in excellent agreement with the adopted  $T_{\text{eff}}$ . However, the photometric calibration of van Belle (1999), is  $\sim 250$  K lower than the adopted value. We note that the  $T_{\text{eff}}$  from the free run output of this star ranges between less than 0.1% and 8% from adopted  $T_{\text{eff}}$  for the iSpec and EPINARBO nodes, respectively. HD 218857 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  $T_{\text{eff}}$  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  $\sim 0.1$  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 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 between REW and Fe abundance indicating a potential issue with the  $v_{\text{mic}}$ .

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 uncertain  $\log g$ . This trend cannot be resolved through varying

the  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and  $v_{\text{mic}}$  within their uncertainties). In addition, the typical uncertainties in the parameters (particularly the uncertainty in  $\log g$ ) of this star are quite large compared to the other stars. However, the star is rather faint ( $V = 8.9$ ) making interferometric  $\theta_{\text{LD}}$  measurements very challenging if not impossible.

#### HD 241253

The adopted  $T_{\text{eff}}$  of this star is 150 K (3%) larger than the typical literature value (e.g. Axer et al. 1994; di Benedetto 1998; Prochaska et al. 2000; Nissen et al. 2002; Gehren et al. 2004; Mashonkina et al. 2003; Masana et al. 2006; Reddy et al. 2006; Reddy & Lambert 2008). The  $T_{\text{eff}}$  is most discrepant with the literature at the 350 K level (Reddy et al. 2006). As we noted above these authors determine the  $T_{\text{eff}}$  of their sample using Strömgen ( $b - y$ ) photometry. Interestingly, these authors revise the  $T_{\text{eff}}$  of the star two years later (Reddy & Lambert 2008) which makes it consistent with our adopted  $T_{\text{eff}}$ . The  $T_{\text{eff}}$  determined using the  $\theta_{\text{LD}}$ -photometric calibration are in good agreement with the adopted value (less than 2%). The  $T_{\text{eff}}$  from the free run output of this star is in moderate agreement with the adopted  $T_{\text{eff}}$  (within 7%). However, this star has a statistically significant correlation between Fe I abundance and EP lines considering both high EP and all EP Fe I lines. This correlation cannot be resolved by varying the stellar parameters within the uncertainty.

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  $[\text{Fe}/\text{H}]$  derived is 0.06 dex (6%) larger than the mean literature value and up to 0.24 dex (25%) larger than the  $[\text{Fe}/\text{H}]$  determined in the free spectroscopic run. The NLTE Fe corrections are on the order of +0.03 dex. There is no significant correlation between REW 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 between Fe I and Fe II. This trend cannot be resolved through varying the  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and  $v_{\text{mic}}$  within their uncertainties. In addition, we cannot achieve ionisation balance accounting for the uncertainties in the parameters. This star has a predicted  $\theta_{\text{LD}}$  that is 0.09 mas and thus is impossible to achieve with current interferometers.

#### HD 298986

The adopted  $T_{\text{eff}}$  of this star is in excellent agreement ( $\sim 1.5$ – $1.7\%$ ) 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  $T_{\text{eff}}$  also agrees well with those derived from the  $\theta_{\text{LD}}$ -photometric calibrations. We note that the  $T_{\text{eff}}$  from the free run output of this star agrees with the adopted  $T_{\text{eff}}$  within 5%. Additionally, the adopted  $T_{\text{eff}}$  is consistent with the spectroscopic  $T_{\text{eff}}$ , as indicated by a null trend in the  $[\text{Fe}/\text{H}]$  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

**Table 6.** Summary of star-by-star consistency check.

Star	$T_{\text{eff}}$	$\log g$	$[\text{Fe}/\text{H}]$	$v_{\text{mic}}$	$\theta_{\text{LD}}$
*BD+264251	×	✓	✓	✓	I
HD 102200	✓	✓	✓	✓	I
HD 106038	✓	✓	✓	✓	I
*HD 126681	×	✓	✓	×	I
HD 175305	✓	✓	✓	✓	P
*HD 196892	×	✓	✓	✓	I
HD 201891	✓	✓	✓	✓	P
*HD 218857	×	✓	✓	×	I
*HD 241253	×	×	✓	✓	I
HD 298986	✓	✓	✓	✓	I

**Notes.** In this table the ✓ represents a star that has “passed” (or × for “failed”) a consistency check for the  $T_{\text{eff}}$  (Col. 2),  $\log g$  (Col. 3),  $[\text{Fe}/\text{H}]$  (Col. 4), and  $v_{\text{mic}}$  (Col. 5) parameters. In addition, we remark whether the  $\theta_{\text{LD}}$  of the star is possible (P) or impossible (I) to directly measure with current (or near future) interferometers.

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  $\alpha$  Tau,  $\alpha$  Cet, and  $\gamma$  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  $[\text{Fe}/\text{H}]$  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  $[\text{Fe}/\text{H}]$  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  $[\text{Fe}/\text{H}]$  is on the order of 0.13 dex. While we do find significant correlation between REW and Fe abundance, this is resolved by increasing the  $v_{\text{mic}}$  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: HD 102200, HD 106038, HD 175305, HD 201891, and HD 298986. 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  $T_{\text{eff}}$  of the star. The stars BD+264251, HD 126681, HD 196892, HD 218857, HD 241253 are denoted with an astrix 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  $-2.0 < [\text{Fe}/\text{H}] < -1.0$  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  $\theta_{\text{LD}}$ -photometric calibrations to estimate the  $\theta_{\text{LD}}$  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  $T_{\text{eff}}$  of each star using the adopted Stefan-Boltzmann law. The  $\theta_{\text{LD}}$ -photometric calibrations of the two giant stars in our sample produced results that disagreed at the 10% level (leading to a  $T_{\text{eff}}$  discrepancy of  $\sim 300$  K). As such, we also employed the IRFM to estimate the  $T_{\text{eff}}$ . We found very good agreement of the  $T_{\text{eff}}$  between the IRFM and the four  $\theta_{\text{LD}}$ -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  $[\text{Fe}/\text{H}]$  for the stars. We processed (e.g. continuum normalised, convolved to common resolution of  $R = 40\,000$ , 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  $[\text{Fe}/\text{H}]$ . There were 2 “equivalent width” nodes (EPINARBO and Bologna) and 2 spectral synthesis codes (BACCCHUS and iSpec) that were used in Papers III and IV. We employed seven separate runs per node which consisted of: a main run where the  $T_{\text{eff}}$ ,  $\log g$ , and  $v_{\text{mic}}$  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  $\pm 1\sigma$  of their uncertainties. The “error” runs were used to evaluate the impact of the uncertainties in the adopted derived stellar parameters on the  $[\text{Fe}/\text{H}]$  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 Figs. 6 and 7) were used to validate the stellar parameters on the basis of the standard Fe I ionisation/excitation balance method. We also used Figs. 6 and 7 in our discussion of the results and the star-by-star analysis noting the consistence of the adopted and spectroscopic parameters in Sect. 6.1. We found that five of the ten stars (HD 102200, HD 106038, HD 175305, HD 201891, and HD 298986) 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  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  are  $\pm 80$  K,  $\pm 0.14$  dex, and  $\pm 0.13$  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  $\theta_{\text{LD}}$ . We recommend all stars with large angular diameters (particularly HD 175305, HD 201891, and

HD 102200) be included in future interferometric  $\theta_{\text{LD}}$  studies. In fact, HD 175305 and HD 201891 can, in principle, be observed with current interferometers (Table 6) and a possible extension of this work is to obtain a direct  $\theta_{\text{LD}}$  measurement for these two stars. Direct measurement on the  $\theta_{\text{LD}}$  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  $-1.3 < [\text{Fe}/\text{H}] < -1.0$  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.

*Acknowledgements.* We would like to thank the referee for constructive comments that helped improved the presentation of this work. We would like to thank A. Casey 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., & Martínez-Roger, C. 1994, *A&AS*, 107, 365  
 Alonso, A., Arribas, S., & Martínez-Roger, C. 1995, *A&A*, 297, 197  
 Alonso, A., Arribas, S., & Martínez-Roger, C. 1996a, *A&AS*, 117, 227  
 Alonso, A., Arribas, S., & Martínez-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  
 Bailer-Jones, C. A. L., Andrae, 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., & Lynas-Gray, A. E. 1998, *A&AS*, 129, 505  
 Blackwell, D. E., & Shallis, M. J. 1977, *MNRAS*, 180, 177  
 Blackwell, D. E., Shallis, M. J., & Selby, M. J. 1979, *MNRAS*, 188, 847  
 Blackwell, D. E., Petford, A. D., & Shallis, M. J. 1980, *A&A*, 82, 249  
 Blanco-Cuaresma, S., Soubiran, C., Heiter, U., & Jofré, P. 2014a, *A&A*, 569, A111  
 Blanco-Cuaresma, S., Soubiran, C., Jofré, P., & Heiter, U. 2014b, *A&A*, 566, A98 (Paper II)  
 Boeche, C., & Grebel, E. K. 2016, *A&A*, 587, A2  
 Boyajian, T. S., van Belle, G., & von Braun, K. 2014, *AJ*, 147, 47  
 Burris, 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., Ramírez, 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
- Casagrande, L., Portinari, L., Glass, I. S., et al. 2014, *MNRAS*, 439, 2060
- Castelli, F., & Kurucz, R. L. 2004, ArXiv e-prints [arXiv:astro-ph/0405087]
- Chen, Y. Q., Nissen, P. E., Zhao, G., Zhang, H. W., & Benoni, T. 2000, *A&AS*, 141, 491
- Clementini, G., Gratton, R. G., Carretta, E., & Sneden, C. 1999, *MNRAS*, 302, 22
- Creevey, O. L., Thévenin, F., Boyajian, T. S., et al. 2012, *A&A*, 545, A17
- Creevey, O. L., Thévenin, F., Berio, P., et al. 2015, *A&A*, 575, A26
- Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, *VizieR Online Data Catalog*, II/246
- Dalton, G., Trager, S., Abrams, D. C., et al. 2014, in Ground-based and Airborne Instrumentation for Astronomy V, *Proc. SPIE*, 9147, 91470
- Datson, J., Flynn, C., & Portinari, L. 2012, *MNRAS*, 426, 484
- Datson, J., Flynn, C., & Portinari, L. 2014, *MNRAS*, 439, 1028
- de Jong, R. S., Bellido-Tirado, O., Chiappini, C., et al. 2012, in SPIE Conf. Ser., 8446
- De Pascale, M., Worley, C. C., de Laverny, P., et al. 2014, *A&A*, 570, A68
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, 449, 2604
- Dekker, H., D’Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, in Optical and IR Telescope Instrumentation and Detectors, eds. M. Iye, & A. F. Moorwood, *Proc. SPIE*, 4008, 534
- Demarque, P., Woo, J.-H., Kim, Y.-C., & Yi, S. K. 2004, *ApJS*, 155, 667
- di Benedetto, G. P. 1998, *A&A*, 339, 858
- Di Benedetto, G. P. 2005, *MNRAS*, 357, 174
- Dravins, D., LeBohec, S., Jensen, H., & Nuñez, P. D. 2012, *New Astron. Rev.*, 56, 143
- Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, *A&A*, 275, 101
- 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
- Fulbright, 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., Sneden, C., Carretta, E., & Bragaglia, A. 2000, *A&A*, 354, 169
- Gratton, R. G., Carretta, E., Claudi, R., Lucatello, S., & Barbieri, M. 2003, *A&A*, 404, 187
- 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, in press, DOI: 10.1051/0004-6361/201628812
- Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, *A&A*, 582, A49 (Paper I)
- 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. 2014, *A&A*, 564, A133 (Paper III)
- Jofré, P., Heiter, U., Soubiran, C., et al. 2015, *A&A*, 582, A81 (Paper IV)
- 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, *A&A*, 572, A88
- 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, Ph.D. 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., & Kovtyukh, V. V. 2001, *A&A*, 370, 951
- Mishenina, T. V., Korotin, S. A., Klochkova, V. G., & Panchuk, V. E. 2000, *A&A*, 353, 978
- Mucciarelli, A., Pancino, E., Lovisi, L., Ferraro, F. R., & Lapenna, E. 2013, *ApJ*, 766, 78
- Nissen, P. E., & Schuster, W. J. 1997, *A&A*, 326, 751
- Nissen, P. E., & Schuster, W. J. 2010, *A&A*, 511, L10
- Nissen, P. E., Primas, F., Asplund, M., & Lambert, D. L. 2002, *A&A*, 390, 235
- 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
- Qui, 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., Israelian, G., et al. 2011, *A&A*, 526, A99
- Stetson, P. B., & Pancino, E. 2008, *PASP*, 120, 1332
- Thévenin, F., & Idiart, 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., & Canterna, 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

## Appendix A: Description of online tables

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 at the CDS.

**Table A.1.** Online table format.

Column	Label	Unit
(1)	Element	
(2)	Absorption line wavelength	Å
(3)	Mean EW	mÅ
(4)	Mean abundance (A)	dex
(5)	NLTE correction <sup>a</sup>	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(ULB)	dex
(13)	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. <sup>(a)</sup> 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.