

University of Groningen

## The Pristine Inner Galaxy Survey (PIGS) II

Arentsen, Anke; Starkenburg, Else; Martin, Nicolas F.; Aguado, David S.; Zucker, Daniel B.; Allende Prieto, Carlos; Hill, Vanessa; Venn, Kim A.; Carlberg, Raymond G.; González Hernández, Jonay I.

*Published in:*  
Monthly Notices of the Royal Astronomical Society

*DOI:*  
[10.1093/mnras/staa1661](https://doi.org/10.1093/mnras/staa1661)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Arentsen, A., Starkenburg, E., Martin, N. F., Aguado, D. S., Zucker, D. B., Allende Prieto, C., Hill, V., Venn, K. A., Carlberg, R. G., González Hernández, J. I., Mashonkina, L. I., Navarro, J. F., Sánchez-Janssen, R., Schultheis, M., Thomas, G. F., Youakim, K., Lewis, G. F., Simpson, J. D., Wan, Z., ... O'Connell, J. E. (2020). The Pristine Inner Galaxy Survey (PIGS) II: Uncovering the most metal-poor populations in the inner Milky Way. *Monthly Notices of the Royal Astronomical Society*, 496(4), 4964-4978. <https://doi.org/10.1093/mnras/staa1661>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).








The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# The Pristine Inner Galaxy Survey (PIGS) II: Uncovering the most metal-poor populations in the inner Milky Way

Anke Arentsen <sup>1</sup>★, Else Starkenburg,<sup>1</sup> Nicolas F. Martin,<sup>2,3</sup> David S. Aguado,<sup>4</sup> Daniel B. Zucker,<sup>5</sup> Carlos Allende Prieto,<sup>6,7</sup> Vanessa Hill,<sup>8</sup> Kim. A. Venn <sup>9</sup>, Raymond G. Carlberg,<sup>10</sup> Jonay I. González Hernández,<sup>6,7</sup> Lyudmila I. Mashonkina,<sup>11</sup> Julio F. Navarro,<sup>5</sup> Rubén Sánchez-Janssen <sup>12</sup>, Mathias Schultheis,<sup>10</sup> Guillaume F. Thomas,<sup>13</sup> Kris Youakim <sup>1</sup>, Geraint F. Lewis <sup>14</sup>, Jeffrey D. Simpson <sup>15</sup>, Zhen Wan <sup>14</sup>, Roger E. Cohen,<sup>16</sup> Doug Geisler<sup>17,18,19</sup> and Julia E. O’Connell<sup>17</sup>

*Affiliations are listed at the end of the paper*

Accepted 2020 June 6. in original form 2020 May 7

## ABSTRACT

Metal-poor stars are important tools for tracing the early history of the Milky Way, and for learning about the first generations of stars. Simulations suggest that the oldest metal-poor stars are to be found in the inner Galaxy. Typical bulge surveys, however, lack low metallicity ( $[\text{Fe}/\text{H}] < -1.0$ ) stars because the inner Galaxy is predominantly metal-rich. The aim of the Pristine Inner Galaxy Survey (PIGS) is to study the metal-poor and very metal-poor (VMP,  $[\text{Fe}/\text{H}] < -2.0$ ) stars in this region. In PIGS, metal-poor targets for spectroscopic follow-up are selected from metallicity-sensitive *CaHK* photometry from the CFHT. This work presents the  $\sim 250 \text{ deg}^2$  photometric survey as well as intermediate-resolution spectroscopic follow-up observations for  $\sim 8000$  stars using AAOmega on the AAT. The spectra are analysed using two independent tools: ULYSS with an empirical spectral library, and FERRE with a library of synthetic spectra. The comparison between the two methods enables a robust determination of the stellar parameters and their uncertainties. We present a sample of 1300 VMP stars – the largest sample of VMP stars in the inner Galaxy to date. Additionally, our spectroscopic data set includes  $\sim 1700$  horizontal branch stars, which are useful metal-poor standard candles. We furthermore show that PIGS photometry selects VMP stars with unprecedented efficiency: 86 per cent/80 per cent (lower/higher extinction) of the best candidates satisfy  $[\text{Fe}/\text{H}] < -2.0$ , as do 80 per cent/63 per cent of a larger, less strictly selected sample. We discuss future applications of this unique data set that will further our understanding of the chemical and dynamical evolution of the innermost regions of our Galaxy.

**Key words:** techniques: photometric – techniques: spectroscopic – stars: fundamental parameters – stars: Population II – Galaxy: bulge – Galaxy: stellar content.

## 1 INTRODUCTION

The oldest, metal-poor stars that still exist today formed in relatively pristine environments and are unique probes of the conditions in the early Universe. These stars carry the imprint of the first supernovae and are local windows into the first episodes of star formation. Studying their chemistry in detail advances our understanding of the first generations of stars and their supernovae, and, combined with kinematic information, metal-poor stars provide clues to the early formation history of the Milky Way and its different components (see reviews by Freeman & Bland-Hawthorn 2002; Tolstoy, Hill & Tosi 2009; Frebel & Norris 2015).

The  $\Lambda$  cold dark matter simulations of Diemand, Madau & Moore (2005) predict that, if any of the very first stars have survived to the present day, 30–60 per cent of them would be within 3 kpc of the Galactic centre. Similarly, the simulations of Tumlinson (2010)

suggest that metal-poor stars below  $[\text{Fe}/\text{H}]^1 = -3.5$  (all formed at redshifts  $> 15$ ) would be most frequently found within the bulge, because dark matter haloes grow inside-out. These stars are not necessarily born in the bulge, but they have ended up in the central regions because of their dynamical history. Using hydrodynamical simulations, Starkenburg et al. (2017a) also find that the fraction of stars that are both metal-poor ( $[\text{Fe}/\text{H}] < -2.5$ ) and old (formed within the first 0.8 Gyr) is highest towards the centres of galaxies, even though the relative fractions of just metal-poor stars or just old stars increase with galactic radius. Hence, a clear picture emerges that the central part of our Galaxy is a privileged location to look for the oldest and most metal-poor stars.

A complication in studying old, metal-poor stars in the Galactic bulge is that the overwhelming majority of stars are metal-rich ( $[\text{Fe}/\text{H}] > -0.5$ ). This metal-rich component has been studied in

<sup>1</sup> $[\text{X}/\text{Y}] = \log(N_{\text{X}}/N_{\text{Y}})_{*} - \log(N_{\text{X}}/N_{\text{Y}})_{\odot}$ , where the asterisk subscript refers to the considered star, and N is the number density.

\* E-mail: aarentsen@aip.de

detail in previous surveys such as the GIRAFFE Inner Bulge Survey (GIBS; Zoccali et al. 2014), the Bulge Radial Velocity Assay (BRAVA; Kunder et al. 2012), the Abundances and Radial velocity Galactic Origins Survey (ARGOS; Ness et al. 2012; Freeman et al. 2013), and the Apache Point Observatory Galactic Evolution Experiment (APOGEE; e.g. Schultheis et al. 2017; García Pérez et al. 2018; Zasowski et al. 2019). The general conclusion from these previous studies is that the Galactic bulge is a rotating, boxy/peanut-shaped pseudo-bulge, with little room for a pressure-supported, classical bulge (e.g. Howard et al. 2009; Shen et al. 2010; Hill et al. 2011; Ness et al. 2013b). If there is a classical component, it can only contain a fraction of the mass of the entire bulge and it is likely more dominant at lower metallicities.

The kinematics of the metal-poor tail of the bulge are still largely unexplored, with the exception of some RR Lyrae studies (e.g. Dékány et al. 2013; Kunder et al. 2016). The ARGOS survey of bulge red clump stars, with a survey strategy aimed to be metallicity-unbiased, found only 16 out of 14 150 stars with  $-2.6 < [\text{Fe}/\text{H}] < -2.0$ , and only 84 stars with  $[\text{Fe}/\text{H}] < -1.5$  (Ness et al. 2013a). This emphasizes that targeted surveys are necessary for the discovery of large samples of metal-poor stars in the inner Galaxy.

There have been several efforts to find very metal-poor ( $[\text{Fe}/\text{H}] < -2.0$ ) stars in the inner Galaxy using photometry. Casey & Schlafman (2015) employed a combination of infrared 2MASS and WISE photometry. These authors found and studied three  $[\text{Fe}/\text{H}] < -2.7$  stars with high-resolution spectroscopy, but it was limited to bright stars only ( $V < 14$ ). Another study by Koch et al. (2016), used a filter focused on the Ca II K line to select candidate metal-poor stars. They found four stars with  $-2.7 < [\text{Fe}/\text{H}] < -2.3$ , among them the first carbon-enhanced metal-poor star in the inner Galaxy.

So far, the largest effort to search for metal-poor stars in the inner Galaxy has been the Extremely Metal-poor Bulge stars with AAOmega (EMBLA) survey (Howes et al. 2014, 2015, 2016), which used SkyMapper photometry. The SkyMapper filter set includes a narrow-band  $v$  filter containing the Ca H&K lines. These lines are good metallicity indicators, and in combination with broad-band photometry, metal-poor stars can be selected efficiently using this  $v$  filter (Wolf et al. 2018). The EMBLA team has performed low-resolution follow-up spectroscopy of photometrically selected candidates and identified  $\sim 150$  stars with  $[\text{Fe}/\text{H}] < -2.5$  (L. Howes, private communication). From this sample, 37 stars were selected for high-resolution spectroscopic follow-up, and of those, 30 have  $[\text{Fe}/\text{H}] \leq -2.5$ , and 9 have  $[\text{Fe}/\text{H}] \leq -3.0$ . These stars are the first known extremely metal-poor stars in the inner Galaxy. The recent Chemical Origins of Metal-poor Bulge Stars (COMBS) survey also analysed stars selected using SkyMapper photometry. Their high-resolution spectroscopic sample contains five stars with  $[\text{Fe}/\text{H}] < -2.0$  (Lucey et al. 2019).

A striking difference between the metal-poor EMBLA stars in the inner Galaxy and typical Galactic halo stars is that only a small fraction of EMBLA stars appear to be enhanced in carbon (Howes et al. 2015, 2016). In the Galactic halo, the number of stars with enhanced carbon increases with decreasing metallicity. These carbon-enhanced metal-poor (CEMP) stars comprise 15–20 per cent of stars with  $[\text{Fe}/\text{H}] < -2$  and up to 70 per cent of stars with  $[\text{Fe}/\text{H}] < -4$  (Yong et al. 2013; Placco et al. 2014). One of the reasons for the lack of CEMP stars in EMBLA could be that the survey is biased against carbon-rich stars. The SkyMapper  $v$  filter includes a region of the spectrum where a CN-band can be present in carbon-enhanced stars, which makes such stars appear more metal-rich and fail the selection for follow-up (Da Costa et al. 2019).

The low number of known metal-poor stars in the bulge, the apparent lack of CEMP stars in the inner Galaxy, the possibility

to uncover the oldest, metal-poor stars in the Milky Way, and the opportunity to study the metal-poor tail of the Galactic bulge are all good reasons to expand the search for metal-poor stars in the inner Galaxy. The Pristine Inner Galaxy Survey (PIGS) aims to build an unprecedentedly large sample of old, metal-poor stars for detailed chemodynamical studies. The Pristine survey (Starkenburg et al. 2017b, hereafter S17b) is very efficient at uncovering metal-poor stars photometrically. It has reported success rates of 56 per cent and 23 per cent from spectroscopic follow-up studies for finding stars in the Galactic halo with  $[\text{Fe}/\text{H}] < -2.5$  and  $[\text{Fe}/\text{H}] < -3.0$ , respectively (Youakim et al. 2017; Aguado et al. 2019; Venn et al. 2020). In the Pristine survey, a narrow-band Ca H&K filter on MegaCam at the Canada–France–Hawaii Telescope (CFHT) is used to provide photometric metallicities down into the extremely metal-poor regime. The  $CaHK$  photometry is combined with available broad-band photometry for colour/temperature information to compute these photometric metallicities. The Pristine filter is narrower than the SkyMapper filter, and should be less biased against CEMP stars (see fig. 2 in S17b).

Compared to the halo, the inner Galaxy has a much higher density of (mainly metal-rich) stars. It is therefore significantly more crowded, there is a larger foreground population, and there is much more (inhomogeneous) reddening due to the presence of dust. These challenges, together with the unique science case, have led to a change in strategy between the main Pristine survey and PIGS, which we will describe in this paper. The first scientific results from PIGS were published in a Letter in this journal (Arentsen et al. 2020), and it focused solely on the kinematics of the metal-poor ( $[\text{Fe}/\text{H}] < -1.0$ ) tail of the inner Galaxy using a sample of several thousand low- and intermediate resolution spectroscopic observations.

In this work, we present the PIGS data collection and analysis in detail and focus on our results and success rates to find and analyse the most metal-poor stars in the inner Galaxy. The photometric  $CaHK$  observations and selection of metal-poor candidates are described in Section 2. The analysis of low-/medium resolution spectroscopic follow-up observations are described in Section 3, the results for the stellar parameters are discussed in Section 4, and the survey performance is determined in Section 5. A summary of our PIGS survey and future plans are provided in Section 6.

## 2 PHOTOMETRIC OBSERVATIONS

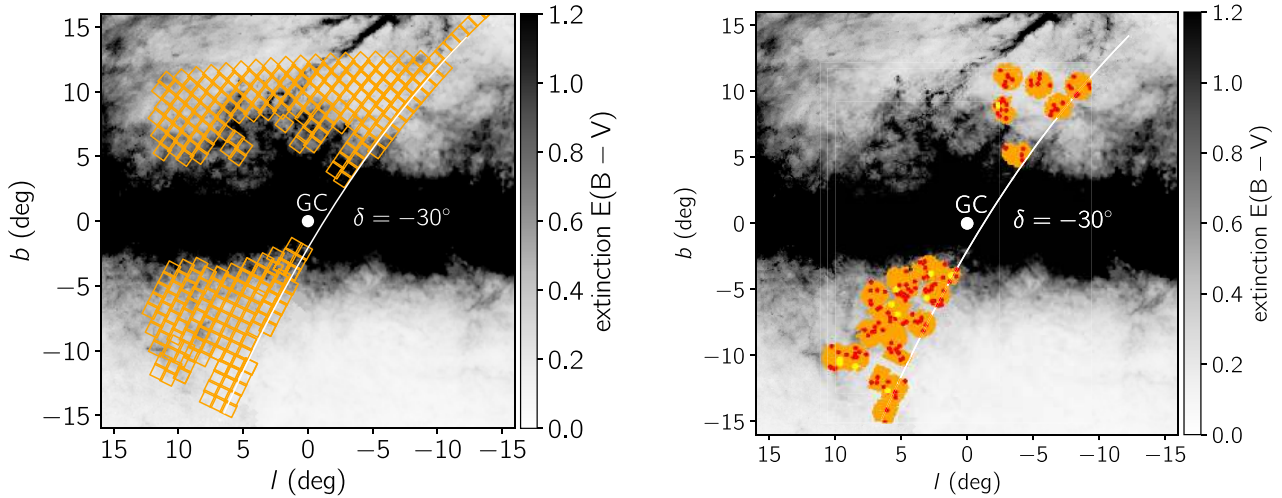
In this section, the photometric PIGS data are presented. We discuss the data reduction and calibration, and the target selections for the spectroscopic follow-up.

With MegaCam on CFHT, we have obtained  $\sim 250$  square degrees of metallicity-sensitive  $CaHK$  photometry in the Galactic bulge region. The footprint covers a region of  $|l|$  and  $|b| \lesssim 12^\circ$  above a declination of  $-30^\circ$  (because we observe from the Northern hemisphere), and we limited it to areas with  $E(B - V) \lesssim 0.7$  to avoid regions with too much extinction. A pilot program was initiated in 2017A, and extended to a larger program in 2018A and 2019A. The final photometric PIGS footprint is shown in the left-hand panel of Fig. 1. A small overlap between neighbouring MegaCam fields is included for calibration purposes.

### 2.1 Data reduction and calibration

#### 2.1.1 Reduction

The PIGS MegaCam images are processed similarly to the Pristine images, as presented in S17b. All downloaded images are pre-processed by CFHT with the ELIXIR pipeline, whose role is to



**Figure 1.** PIGS footprints, on top of the extinction map from Green et al. (2018) (where available and reliable, which is above a declination of  $-30^\circ$  and at  $|b| > 2^\circ$ ) and Schlegel, Finkbeiner & Davis (1998) (elsewhere). A truncation of the colour bar at  $E(B - V) = 1.2$  is set for clarity. Left-hand panel: photometric footprint, where each square is one MegaCam field. Right-hand panel: spectroscopic footprint, located within the photometric footprint from 2017 and 2018. Stars with  $[\text{Fe}/\text{H}] < -2.5$  are marked in red, and those with  $[\text{Fe}/\text{H}] < -3.0$  in yellow.

remove instrumental signatures from the data (Magnier & Cuillandre 2004). The images were processed with a version of the Cambridge Astronomical Survey Unit pipeline (CASU; Irwin & Lewis 2001) tailored to MegaCam images. The astrometry is performed on a CCD to CCD basis and using 2MASS (Skrutskie et al. 2006) for reference. While we could also use the *Gaia* DR2 (Gaia Collaboration 2016, 2018) astrometry, the shallower 2MASS data works well for the densely populated regions of the survey and has the benefit of converging more quickly on a good astrometric solution, accurate at the 0.1 arcsec level. This accuracy is more than enough to cross-match the sources to *Gaia* DR2 positions, and we use the *Gaia* positions for the stars throughout this paper. With the CASU pipeline, aperture photometry is carried out on the individual CCDs to build a catalogue of sources that are at least  $5\sigma$  above the background threshold.

The following quality cuts are made on the *Gaia* cross-matched catalogue to create our final non-calibrated catalogues:

- (i) *casu\_flag* = -1 or *casu\_flag* = -2
- (ii) *Gaia phot\_variable\_flag*  $\neq$  VARIABLE
- (iii) *Gaia phot\_bp\_mean\_mag* and *phot\_rp\_mean\_mag*  $> 0$
- (iv) *Gaia duplicated\_source* = false

The first cut only keeps point-sources from the CASU *CaHK* photometry data reduction. The other three are quality cuts on the *Gaia* data, removing variable stars, stars that do not have *BP* and/or *RP* photometry, and stars that are flagged to be duplicated sources. The last flag proved to be unnecessary in hindsight. The full catalogue consists of more than 34 million objects with a *Gaia* DR2 counterpart (there are fewer unique stars; some stars have multiple entries due to the overlap between fields).

### 2.1.2 Calibration

The (non-calibrated) catalogue produced after the data reduction steps described above needs to be calibrated with several additional steps. First, because the MegaCam images show subtle variations in magnitude that depend on the location in the field of view, a flat-field correction is applied as determined in section 2.4.2 of

S17b. Secondly, all MegaCam fields are put on the same scale. S17b used the determination of the stellar locus in the SDSS dereddened  $((g - i)_0, (\text{CaHK} - g)_0)$  colour-colour space to shift all fields to the scale of one reference field. The stellar locus consists mainly of metal-rich foreground stars. Unfortunately this method does not work for our footprint towards the inner Galaxy. The stellar locus is less well-defined since we look through a larger part of the disc, and, additionally, correcting for 3D extinction is a challenge. Thus, our MegaCam pointings are placed to create an overlap between neighbouring *CaHK* fields of  $\sim 0.1^\circ$ , and these overlapping regions are used for a relative photometric calibration. All *Gaia* DR2 source.id's that have multiple *CaHK* matches within 1 arcsec are identified, and we assume that these sources are the same stars observed in two (or more) different MegaCam fields. These are then used to determine the relative offsets between fields.

The 2017 and 2018 data sets and calibration are primarily used to select stars for our spectroscopic follow-up program (described in Section 3). As an initial (crude) calibration, we set one central field as the reference and started by determining offsets for the directly neighbouring fields with respect to the reference. For each pair of fields, all stars in common were used to compute the minimum  $\chi^2$  and the best offset. This process was repeated for each of those fields and their neighbours that had not yet been calibrated, until all fields are on the same scale. In this calibration, each field is calibrated with respect to only one other field.

In 2019, the calibration was changed to a more homogeneous approach, minimizing the  $\chi^2$  for the overlapping regions of all fields simultaneously (using only the stars with *CaHK* uncertainties less than 0.01), using the iterative method of Magnier et al. (1992) to find the best solution. We have since re-calibrated all our photometry using this approach, including the fields in the 2017 and 2018 footprint.

It is important to note that since we perform a relative calibration, the *CaHK* photometry is not on the same scale as the *CaHK* photometry in S17b. It is only homogeneous for continuous regions; thus, fields with  $b > 0^\circ$  and  $b < 0^\circ$  are not necessarily on the same scale as the two regions do not overlap.

## 2.2 Target selection for spectroscopic follow-up

### 2.2.1 Choice of broad-band photometry

The calibrated *CaHK* photometry was combined with broad-band photometry to determine the best metal-poor candidates. The *CaHK* lines are not only sensitive to metallicity, but they also depend strongly on the temperature of the star, which can be approximated using broad-band colours. In S17b, SDSS photometry was used, but there is no SDSS photometry towards the Galactic bulge (York et al. 2000). Instead, *Gaia* DR2 (Gaia Collaboration 2018) and Pan-STARRS1 (PS1; Chambers et al. 2016) photometry were used to select targets. To select giant stars that are roughly at the distance of the bulge (for the typical extinction in our fields), we limited our magnitude range to  $13.5 < G < 16.5$  when using *Gaia*, and  $14.0 < g < 17.0$  when using PS1 (see Section 3.1). An additional cut was imposed on the *Gaia* parallaxes ( $\varpi$ ) to remove foreground stars. This cut has slightly changed during the progress of the survey, from no cut in our spectroscopic follow-up pilot, in 2017 August, to  $\varpi > 0.25$  mas in 2018A,  $\varpi - \Delta\varpi > 0.30$  mas in 2018B and  $\varpi - \Delta\varpi > 0.25$  mas in 2019A.

The use of broad-band colour systems has evolved as well as we gained further insight by collecting larger spectroscopic samples. In 2018, after the pilot program, we enthusiastically switched from PS1 photometry to *Gaia* DR2 photometry when it became available, because the latter has the advantage of being more precise and having all-sky coverage. The bulk of our spectroscopic follow-up to date has thus been selected with the help of *Gaia* photometry. Further analysis of the results of these *Gaia*-selected runs, however, showed that the PS1 photometry performs better for our purposes, especially in regions with high extinction. The reason for this is that even though the *Gaia* photometry is much more precise than PS1, it is much harder to correct for extinction. We have used *Gaia* extinction coefficients determined by F. Anders<sup>2</sup> for stars of  $T_{\text{eff}} = 5250$  K and an extinction of  $A_V = 1.5$  (which corresponds to an  $E(B - V) \approx 0.5$ ), together with the Green et al. (2018) reddening map. Because the *Gaia* filters are so broad, their extinction coefficients depend non-negligibly on the amount of extinction itself and on the temperature of the star. The coefficients are therefore not constant throughout our footprint, which covers a range of extinction values.

For these reasons, we switched back to using PS1 for our follow-up in 2019 and indeed noticed that it reduced the contamination from more metal-rich stars. The PS1 filters are narrower and therefore less dependent on the temperature of the star and the extinction. Additionally, the Green et al. (2018) reddening map has been made using PS1 photometry and they provide up-to-date PS1 extinction coefficients to go with this map, making the broad-band de-reddened photometry more self-consistent. For the *CaHK* photometry, the extinction coefficient 3.924 from S17b was used, which is on the Schlafly & Finkbeiner (2011, hereafter SF11) scale.

Green et al. (2018) provide a 3D map that can be queried at different distances. For our purposes, we assume that all dust is located in the foreground disc (which should be valid in the PIGS latitude range), and query the map at a distance of 8 kpc. Another assumption is that the extinction law is constant (and correct). However, Schlafly et al. (2016) have shown that there appear to be spatial variations, which have not yet been characterized in

detail. Finally, the extinction correction is limited by the pixel scale of the map used, while the extinction may vary on even smaller scales.

In this paper, whenever photometry is shown (in Figs 2 and 9), the updated PS1 reddening map by Green et al. (2019, hereafter G19) was used. The difference between the extinction coefficients for the PS1 *g* and *r* filters provided by the authors and those from SF11 is significant (0.35 for both), due to differences in the adopted reddening vector and scaling of their map. The extinction coefficients for *CaHK* and the *Gaia* photometry are, however, on the SF11 scale. To convert the *Gaia* photometry to the G19 reddening scale, 0.35 was added to the  $A_V$  as used in Footnote 2 (assuming that the difference is the same as for the PS1 *g* and *r* filters, which overlap in wavelength range with *V*). The *CaHK* wavelength range lies outside the PS1 coverage and is bluer (more affected by extinction), therefore extrapolation is likely invalid for this filter. To estimate the G19 *CaHK* coefficient, we instead tested a range of coefficients and determined for which value the dispersion in our photometric metallicity bins was the smallest (see Section 5.2 for a description of the photometric metallicities). A minimum was found for  $A_{\text{CaHK}} = 4.5$ , which was therefore adopted as the coefficient to use with the G19 map.

### 2.2.2 Metallicity-sensitive photometric selection

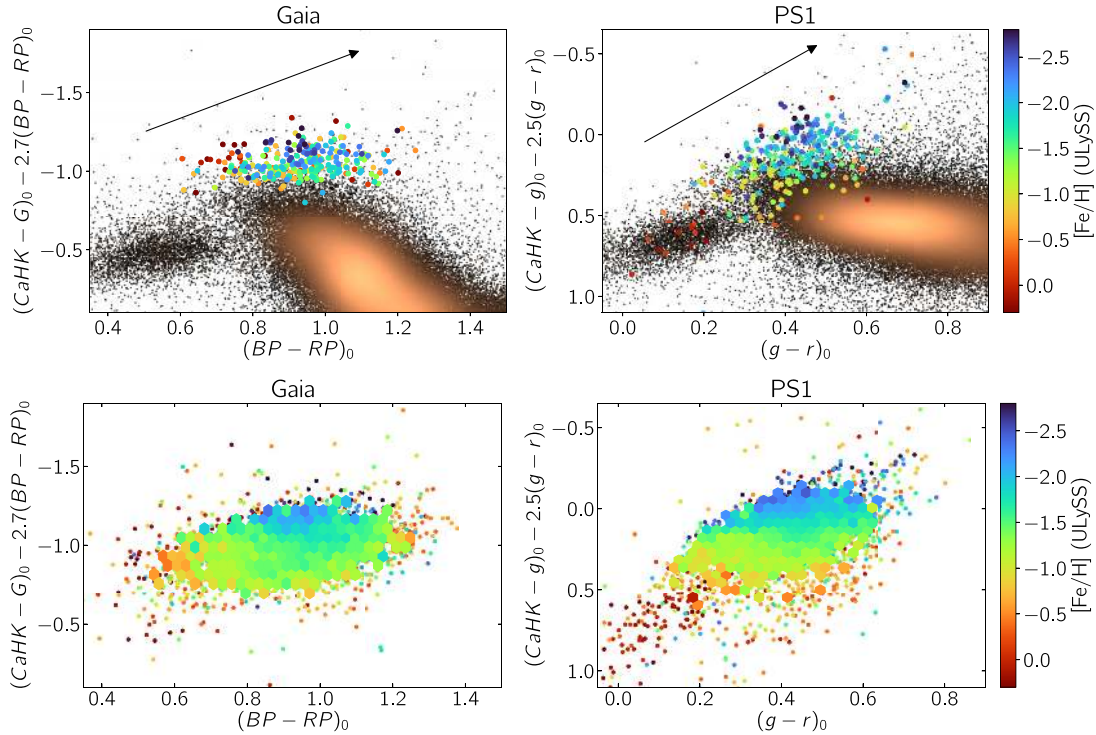
In S17b, photometric metallicities for individual stars were determined from the *CaHK* and broad-band photometry. This was not done for PIGS for multiple reasons: our *CaHK* photometry is not all on the same absolute scale, the method had not been calibrated for *Gaia* broad-band photometry, and the photometric metallicities would in any case be much less precise than in the halo because of uncertainties in the reddening correction.

Instead, a relative selection was applied in each follow-up field. An example selection for follow-up with *Gaia* is shown in the top left-hand panel of Fig. 2, for a two-degree field of view. In this Pristine colour–colour diagram, the metal-poor stars are roughly expected to lie towards the top. Targets were therefore simply selected by their vertical location in this diagram, within a certain  $(BP - RP)_0$  range corresponding to the expected location of metal-poor giant stars. The coloured points are stars followed up spectroscopically, which have been colour-coded according to their spectroscopic metallicity (for details on the spectroscopy, see the next section). The top-right-hand panel of the figure shows the same stars, but now with PS1 broad-band photometry instead of *Gaia*. The contamination of metal-rich stars would have been lower, if this field had been selected using PS1. In the bottom row of the figure we present the same diagrams, now for all stars in our spectroscopic follow-up program with  $b < 0^\circ$  (a homogeneously calibrated region of the photometry). There is a much clearer gradient with  $[\text{Fe}/\text{H}]$  in the case of PS1 photometry compared to the *Gaia* photometry. The main culprit of this difference is the extinction, which – as explained above – is difficult to correct for in the broad-band *Gaia* filters.

## 3 SPECTROSCOPIC FOLLOW-UP OBSERVATIONS

We performed low-/intermediate-resolution spectroscopic follow-up of our metal-poor candidates with the 3.9 m Anglo-Australian Telescope (AAT). The follow-up was done with the dual-beam AAOmega spectrograph (Saunders et al. 2004) with the Two Degree Field (2dF) multi-object instrument, which has roughly 360 fibres for science targets (plus  $\sim 40$  sky and guide fibres) over a 2-degree

<sup>2</sup>These coefficients are  $A_G/A_V = 0.718$ ,  $A_{BP}/A_V = 0.9785$  and  $A_{RP}/A_V = 0.576$ . They were derived by interpolating between coefficients for stars of 4000 and 6500 K. The Schlafly & Finkbeiner (2011) value for  $A_V = 2.742$  ( $R_V = 3.1$ ) was used.



**Figure 2.** Pristine colour–colour diagrams combining our *CaHK* photometry with broad-band photometry from *Gaia* (left-hand panel) and PS1 (right-hand panel). Top row: stars within our selected magnitude range and passing the parallax cut for a two-degree field of view centred on  $(l, b) = (2.4^\circ, -5.2^\circ)$ . We show our spectroscopic follow-up for this field, where the selection was originally made using *Gaia*. These stars are colour-coded by their ULYSS [Fe/H] (for details on the spectroscopic analysis, see Section 3.4). The extinction vector for  $E(B - V) = 0.45$  (the average in this field) is indicated by the arrow in the top left. Bottom row: 4300 stars in the Southern spectroscopic PIGS sample ( $b < 0^\circ$ ) that pass the ULYSS quality criteria. Each hexbin contains at least five stars and is coloured by its average ULYSS [Fe/H]. Single stars are plotted as additional points.

diameter field of view (Lewis et al. 2002; Sharp et al. 2006). To date, 23 fields have been observed, yielding spectra for 8350 stars. The right-hand panel of Fig. 1 shows the current spectroscopic PIGS footprint. Three fields were observed in service mode in 2017 August and 2018 June, the rest was observed during three observing runs in 2018 June, August, and 2019 April. The observations of two fields in 2017 August were our spectroscopic pilot program.

In this section, our observing strategy, the data reduction, the determination of radial velocities, and two independent methods of analysing the spectra are described. We first describe the derivation of  $T_{\text{eff}}$ ,  $\log g$  and [Fe/H] with ULYSS (Koleva et al. 2009), which is a data-driven method using empirical reference model spectra. In our second analysis, the employed method is FERRE (Allende Prieto et al. 2006) with a grid of synthetic model spectra to determine  $T_{\text{eff}}$ ,  $\log g$ , [Fe/H], and [C/Fe]. The parameters from ULYSS for the sample in this work were used in Arentsen et al. (2020), although we note that if the FERRE parameters had been used instead this would not have altered the results of that paper.

### 3.1 Observing strategy

For each field, the top 200 candidates from the photometry were provided and they were given the highest priority in the configuration of 2dF. Generally,  $\sim 95$  per cent of those candidates were allocated a fibre. The next best 200 candidates were provided with intermediate priority and the next best 200 candidates with low priority to fill the remaining  $\sim 160$  science fibres. Stars already in the EMBLA survey were removed from our selection in most fields. 25 fibres

were allocated to sky positions, the input catalogue for these was created by hand using the Aladin tool (Bonnarel et al. 2000) and PS1 images. Finally there are 8 fibres for guide stars, which were selected from *Gaia* DR2.

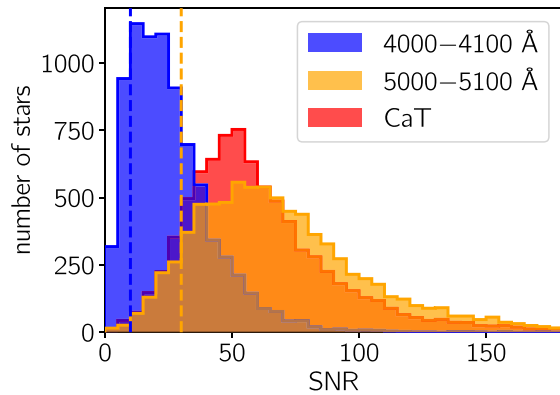
The stars were observed with the 580V and 1700D gratings, resulting in blue spectra at a resolving power  $R \sim 1300$  covering  $\sim 3700\text{--}5500 \text{ \AA}$ , and near-infrared calcium triplet (CaT) spectra at  $R \sim 11\,000$  covering  $\sim 8400\text{--}8800 \text{ \AA}$ . The CaT spectra were used to determine radial velocities, and both spectra together to determine stellar parameters and metallicities.

The recommended magnitude range for stars in one 2dF field is  $\sim 3$  magnitudes. When using *Gaia* photometry, stars between  $G = 13.5$  and 16.5 were selected, and when using PS1, the selected stars were between  $g = 14.0$  and 17.0. We aimed for a minimum signal-to-noise ratio (SNR) of 20 in the blue spectra, which meant integrating for 2 h on each field. Because there were some fibre throughput issues (see below), this was not always successful for the faintest stars. The SNR in different regions of the spectra is summarized in Fig. 3.

### 3.2 Data reduction

The spectra were reduced using the AAT 2DFDR<sup>3</sup> (v. 6.46) package. The standard settings for the 580V and 1700D spectra were used respectively, with the exception of two parameters that determine how spectra from multiple observations are combined.

<sup>3</sup><https://aat.anu.edu.au/science/software/2dfdr>



**Figure 3.** Distribution of the SNR in different regions of the AAT spectra. The CaT range is from  $\sim 8400\text{--}8800\text{ \AA}$ . The  $\text{SNR}_{4000\text{--}4100} = 10$  and  $\text{SNR}_{5000\text{--}5100} = 30$  quality cuts (see Section 3.4) are indicated by the vertical dashed lines.

The first setting which was changed is the `FLUX_WEIGHT` parameter, which determines how much weight is given to each individual exposure when combining to a single spectrum. The standard setting determines a frame zero-point by taking the median of the zero-points for each fibre. These fibre zero-points are derived using the difference between an empirical brightness from a smoothed version of the spectra and the expected brightness from provided external photometry. The frame zero-points are used to determine the relative weights of multiple exposures. This, however, is inappropriate when there is variation in throughput across the field. From the 2DFDR quality plots, we concluded such a variation is indeed present in most of our fields. There are often regions on the plate that have a reduced throughput. Although these regions are not always in the same location, they are usually in the outer regions of the plate. The effect is described in some detail in the appendix of Li et al. (2019). The root cause of the problem is the accuracy of the fibre placement, but exactly why the fibre placement is bad is unknown. There have been efforts at the AAT to resolve these problems, but so far no solution has been found (T. S. Li and C. Lidman, private communication). Because of the observed variation in throughput, we therefore decided to use the flux weighting by object instead of by frame, where each object is assigned its own weight in the combination of multiple frames.

Secondly, the `ADJUST_CONTINUUM` option was turned off. It was originally introduced into the software to fix a problem with the continuum of early 2dF spectra, which was solved in 1999. The 2DFDR documentation suggests it is worth using for data taken after the problem was fixed (it is the default), but it can also be turned off. We found that the default setting sometimes produced spurious, unphysical calcium triplet line shapes, therefore this option was disabled.

### 3.3 Radial velocities

Radial velocities were determined from the CaT spectra using the `FXCOR` package in IRAF.<sup>4</sup> The CaT spectra were cross-correlated with synthetic spectra, which were created using the MARCS (Model

Atmospheres in Radiative and Convective Scheme) stellar atmospheres and the Turbospectrum spectral synthesis code (Alvarez & Plez 1998; Gustafsson et al. 2008; Plez 2008). A preliminary velocity was derived using a template with  $T_{\text{eff}} = 5000\text{ K}$ ,  $\log g = 2.5$ , and  $[\text{Fe}/\text{H}] = -2.0$ . After zero-shifting the spectra, we derive stellar parameters with ULYSS (see Section 3.4). To improve the radial velocities, they were re-derived using the template that is the closest match to the ULYSS parameters, from the following grid:  $T_{\text{eff}} = [4500, 5000, 5500]\text{ K}$ ,  $\log g = 2.5$ , and  $[\text{Fe}/\text{H}] = [0.0, -1.0, -2.0, -3.0]$ . From the `FXCOR` formal uncertainties and a test of repeated observations, the radial velocity uncertainties are estimated to be of the order of  $2\text{ km s}^{-1}$ .

## 3.4 Spectral analysis with ULYSS

### 3.4.1 Method

For the first determination of stellar atmospheric parameters ( $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ ) the University of Lyon Spectroscopic analysis Software (ULYSS,<sup>5</sup> Koleva et al. 2009, version 1.3.1) was used. ULYSS is a full-spectrum fitting package that determines the best parameters by minimizing the  $\chi^2$  of the residuals of the fit between an observed spectrum and model spectra. The model library used is the empirical MILES library (Sánchez-Blázquez et al. 2006; Falcón-Barroso et al. 2011) with a resolving power of  $\sim 2200$ , for which Prugniel, Vauglin & Koleva (2011) have computed a ULYSS interpolator. We use the version of the interpolator as updated by Sharma, Prugniel & Singh (2016), who improved it for cool stars. Recently, Arentsen et al. (2019) have used ULYSS with this interpolator to determine stellar atmospheric parameters for DR2 of the X-shooter Spectral Library (Gonneau et al. 2020). The reliability of the method has been tested extensively in the aforementioned works, and it has been found to be robust for FGK stars with  $-2.5 < [\text{Fe}/\text{H}] < 0.5$  and  $0 < \log g < 5$ . Details on how the method works can be found in Prugniel et al. (2011), it is briefly summarized here.

The model spectrum  $S(\lambda)$  in ULYSS is described by:

$$S(\lambda) = P_n(\lambda) \times G(v_r, \sigma) \otimes \text{TGM}(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], \lambda), \quad (1)$$

where  $P_n(\lambda)$  is a series of Legendre polynomials up to degree  $n$ , and  $G(v_r, \sigma)$  is a Gaussian broadening function described by the radial velocity  $v_r$  and the broadening width  $\sigma$ . The  $\sigma$  is mainly related to the resolution of the spectra and to the rotational velocity of a star, sometimes it is also increased to smooth the spectra if no good fit can be found otherwise. The polynomial  $P_n(\lambda)$  normalizes the spectra, correcting for, e.g. the effects of the detector response and dust extinction. The TGM function represents the interpolator, which is a polynomial function that creates spectra from a reference library for a given combination of  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . In the fit, the free parameters are  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ,  $v_r$ ,  $\sigma$  and the coefficients of  $P_n(\lambda)$ .

The spectra reduced to the rest-frame with our preliminary CaT radial velocities were used, therefore  $v_r$  only absorbs uncertainties in the velocity. The degree of  $P_n(\lambda)$  was set to 10, which is enough to correct for the response function of the detector and the extinction, but not so high that it starts affecting the parameters. The spectra were fit between 3800 and 5400  $\text{\AA}$ . The region of the CH *G*-band was masked between 4260 and 4325  $\text{\AA}$ , since carbon can be anomalous in metal-poor stars and it should be avoided that this drives the fit.

<sup>4</sup>IRAF (Image Reduction and Analysis Facility) is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under contract with the National Science Foundation.

<sup>5</sup>ULYSS is available at <http://ulyss.univ-lyon1.fr>

The MILES library does not contain the region of the CaT, so the blue arm was fit alone.

We run ULYSS with a grid of initial guesses to avoid getting trapped in local minima. The starting grid is  $T_{\text{eff}} = [4000, 5000, 6000]$  K,  $\log g = [2.0, 4.0]$ , and  $[\text{Fe}/\text{H}] = [-2.4, -1.7, -0.3]$ , and the solution with the smallest  $\chi^2$  is finally adopted as the best solution. Furthermore, the /CLEAN option was turned on, which rejects regions where the fit is bad due, for example, to bad pixels or features which are not in the model spectra such as strong carbon bands.

Some spectra do not pass through this first run or have extremely high  $\sigma$  values. They were rerun with a multiplicative polynomial of degree  $n = 30$ , which gives solutions for 89 additional stars. The main reason that these spectra fail, is that they either have a strong continuum shape that cannot be corrected with  $n = 10$  or they are very carbon rich, in which case the higher polynomial degree tries to correct for the carbon features.

Example fits for seven stars are presented in Fig. 4, in blue.

### 3.4.2 Selection of reliable results

The atmospheric parameters were determined for 8089 PIGS stars. To select reliable solutions, the following criteria were used:

- (i)  $\text{SNR}_{4000-4100} > 10$  or  $\text{SNR}_{5000-5100} > 30$
- (ii) signal-to-residual ratio (SRR)  $> 15$
- (iii)  $\sigma < 150 \text{ km s}^{-1}$
- (iv) fraction of rejected pixels in the fit between 4600 and 4800 Å  $< 10$  per cent and between 5000 and 5200 Å also  $< 10$  per cent
- (v) not double-lined

which leaves 7126 stars. The SNR cut was determined using repeated observations: above these SNR values, the dispersion in the results for individual observations is almost independent of SNR. This cut removes 677 stars, a relatively large fraction because of the flux-loss issues described in Section 3.2. The second and third cuts were set after inspecting the fits of stars that passed the SNR criterion, but had low SRR or very high  $\sigma$  which typically indicates a bad fit. Then the fraction of rejected pixels in the fit in the regions of two strong carbon bands were used to weed out a number of carbon-enhanced stars that had not been removed yet by the other cuts. The ULYSS results are not reliable for such stars. These three cuts remove an additional 136 stars. The final cut removes stars with double-lined CaT spectra, based on a cross-correlation function (CCF) analysis. We cross-correlate the observed spectrum and the radial velocity synthetic template, and flag stars that have two or more clear peaks in the CCF (if the secondary is at least 50 per cent as strong as the primary). This removes a final 150 stars. Given the crowding towards the bulge, the double-lined stars are probably more likely to be chance-alignments than spectroscopic binaries.

### 3.4.3 Stellar parameter uncertainties

The formal uncertainties provided by ULYSS might underestimate the statistical uncertainties if the uncertainty spectra are not perfect. To investigate this, a repeat-observation test as in Arentsen et al. (2019) was used. For each independent pair of observations of the same star, we computed the following weighted difference in the stellar parameters:

$$\Delta P_{w,i} = \frac{P_{1,i} - P_{2,i}}{\sqrt{\epsilon_{1,i}^2 + \epsilon_{2,i}^2}}, \quad (2)$$

where P represents  $T_{\text{eff}}$ ,  $\log g$  or  $[\text{Fe}/\text{H}]$ ,  $\epsilon$  represents the formal uncertainty given by ULYSS, and 1 and 2 indicate two observations of the same star. The distribution of  $\Delta P$  should be close to Gaussian, with a standard deviation  $\sigma_{\Delta P} = 1$  if the uncertainties are properly scaled. A smaller/larger  $\sigma_{\Delta P}$  indicates over-/underestimated uncertainties.

We select two fields, one with four sub-exposures (F\_273.7-27.1) and one with six sub-exposures (F\_284.0\_30.0). The stars in these fields are a representative sub-sample of the total. After applying the quality cuts for reliable solutions from the previous section, 2850 independent pairs of observations remain, which were used for the test. For all three parameters the distribution of  $\Delta P$  is found to be close to Gaussian, with a standard deviation of 1.2. This indicates that the uncertainties are somewhat underestimated, therefore we multiply the formal ULYSS uncertainties by this factor to derive proper internal uncertainties. The median of the internal uncertainties for the reliable set of ULYSS parameters are 32 K, 0.08 dex and 0.04 dex for  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$ , respectively.

## 3.5 Spectral analysis with FERRE

### 3.5.1 Method

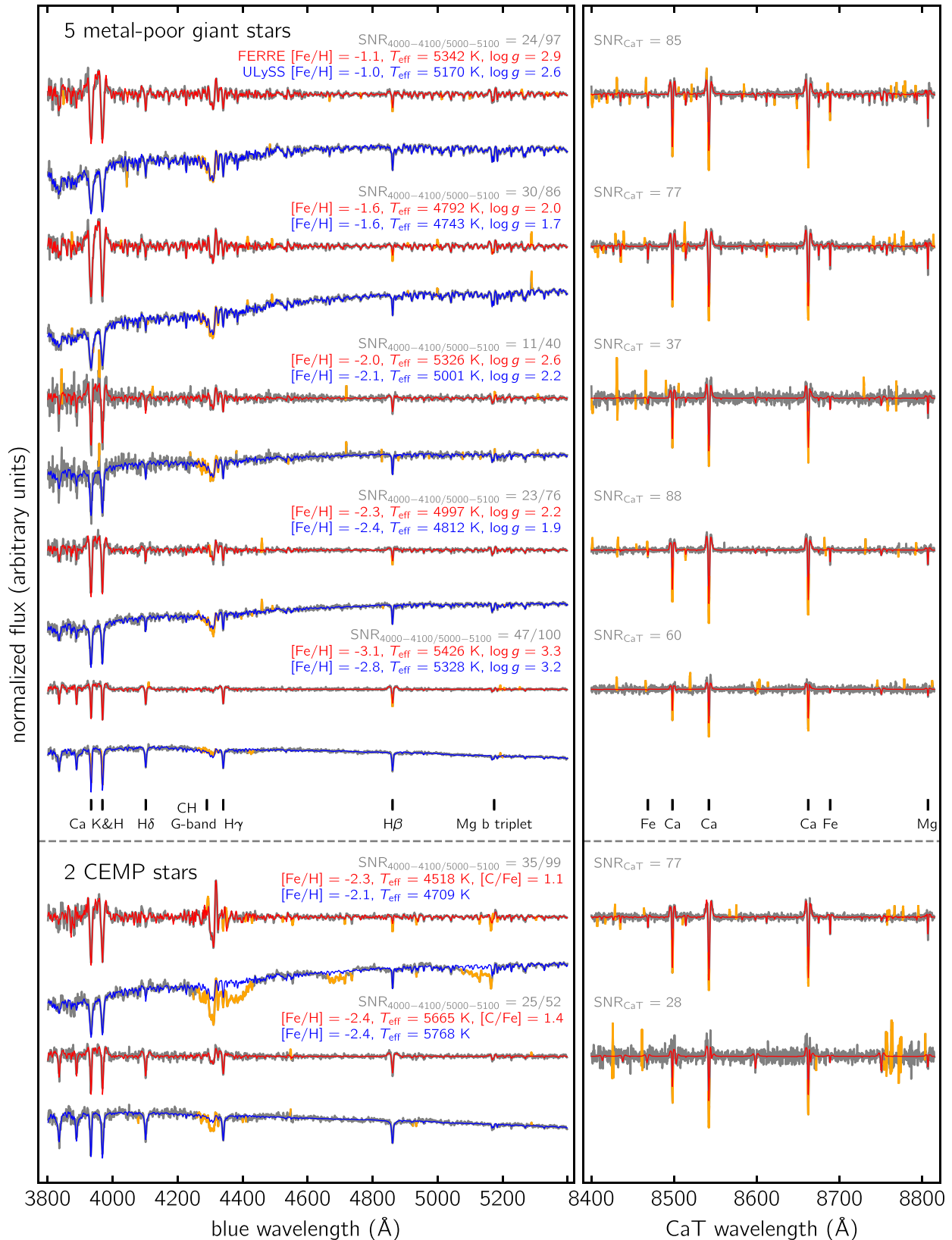
The spectra were additionally analysed with the FERRE<sup>6</sup> code (Allende Prieto et al. 2006). FERRE performs full-spectrum fitting against a synthetic stellar model library, interpolating between the different nodes (the cubic algorithm was selected for the interpolation). In addition, the Boender-Timmer-Rinnoy-Kan (BTRK) algorithm was used as the global optimisation technique already included in the FERRE architecture. For each fit four different runs have been performed to avoid getting trapped in a local minimum. The  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$  and  $[\text{C}/\text{Fe}]$  were derived simultaneously. In this work, we mainly focus on  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . The results with regards to  $[\text{C}/\text{Fe}]$  merit a separate discussion that is the subject of a forthcoming paper (Arentsen et al., in preparation). The synthetic spectra in the reference grid were computed with the ASSET code (Koesterke, Allende Prieto & Lambert 2008), with the model atmospheres computed using the Kurucz code, described in Mészáros et al. (2012), assuming a microturbulence of  $2 \text{ km s}^{-1}$ . The grid is similar to the carbon-rich ultra metal-poor (CRUMP) grid presented in Aguado et al. (2017), which has also been used in previous *Pristine* analyses (Youakim et al. 2017; Aguado et al. 2019), with two important differences: our current grid starts at  $T_{\text{eff}} = 4500$  K instead of 4750 K, and it extends up to  $[\text{Fe}/\text{H}] = +0.5$  compared to  $-2.0$  previously. The grid ranges and grid steps are the following:

- (i)  $4500 \text{ K} < T_{\text{eff}} < 7000 \text{ K}$ ,  $\Delta T_{\text{eff}} = 250 \text{ K}$
- (ii)  $1.0 < \log g < 5.0$ ,  $\Delta \log g = 0.5$
- (iii)  $-4.0 < [\text{Fe}/\text{H}] < +0.5$ ,  $\Delta [\text{Fe}/\text{H}] = 0.5$
- (iv)  $-1.0 < [\text{C}/\text{Fe}] < +3.0$ ,  $\Delta [\text{C}/\text{Fe}] = 1.0$

where for all nodes  $[\alpha/\text{Fe}] = +0.4$ . This value of  $[\alpha/\text{Fe}]$  is reasonable for bulge stars with  $[\text{Fe}/\text{H}] < -1.0$  or even  $-0.5$  (Barbuy, Chiappini & Gerhard 2018), but may result in biased results for more metal-rich stars. However, as the focus of this work is on metal-poor stars, this grid is the best choice for our science case. One of the advantages of the FERRE code is that it is able to fit different regions of the spectrum simultaneously even if they have different resolutions. That fact allows us to use all the information contained in both arms, fitting the blue and the CaT spectra together. We found,

<sup>6</sup>FERRE is available from <http://github.com/callendeprieto/ferre>





**Figure 4.** Example fits for ULySS (blue) and FERRE (red) for five metal-poor giants and two CEMP stars. For both analyses, the observed (normalized) spectrum is shown in grey. Regions in the observed spectra which are masked in the fit are indicated in orange, these contain both a priori masks (the CH G-band in ULySS and the calcium triplet line cores in FERRE) and pixels that are masked iteratively. Several of the strongest spectral features have been indicated. For each star, the stellar parameters from both analyses are shown.

for instance, that including the CaT spectra reduces the scatter in the determined  $\log g$  values considerably, likely because the wings of the calcium lines are sensitive to gravity. The blue and CaT spectra were corrected to rest-frame velocities using the CaT radial velocities. Several artefacts were removed from the spectra which are the result of dead pixels, and usually occur in the same place in the spectrum. The model spectra were smoothed to  $R = 1300$  and  $R = 11\,000$  for the blue and CaT spectra, respectively. The observed and model spectra were normalized using a running mean filter of 30 pixels.

We ran FERRE twice. First FERRE is run to get an estimate of the stellar parameters and a corresponding model spectrum for each observed spectrum. The cores of the calcium triplet lines were masked because they are strongly affected by non-local thermodynamical equilibrium (NLTE) processes.<sup>7</sup> The residuals between the observed and model spectrum were used to flag bad pixels. In the blue spectra, the standard deviation  $\sigma_f$  between model and observed spectrum was determined in chunks of 200 Å, because the SNR varies strongly throughout. Pixels were flagged as bad if they are discrepant by more than  $3\sigma_f$ , and their neighbouring pixels were added if they are discrepant by more than  $2\sigma_f$ . Flagged regions are typically dead pixels that had not yet been removed, data reduction artefacts or other outliers. For the near-infrared spectra a single  $\sigma_f$  was computed since the SNR does not vary, and the pixels were flagged in the same way. The threshold was increased to  $5\sigma_f$  around the three calcium triplet lines and the magnesium line at 8806 Å, to avoid these lines being removed entirely. The SNR is generally high enough that uncertainties in the spectral modelling start to affect the fit of the lines. Typically more pixels are flagged in the CaT spectra compared to the blue spectra because of faulty sky subtraction, which affects the infrared more. When we run FERRE for the second time, the uncertainties on the pixels flagged as bad are increased to  $10^9$ , in order to set their weight in the fit to practically zero. This second run resulted in significantly better fits for many stars. Example fits for seven stars are presented in Fig. 4, in red.

### 3.5.2 Selection of reliable results

Reliable stellar parameters were selected using the following criteria:

- (i)  $\text{SNR}_{4000-4100} > 10$  or  $\text{SNR}_{5000-5100} > 30$
- (ii)  $\log \chi^2 < -2.4$
- (iii) not double-lined

which leaves 7186 stars. The SNR cut is the same as for the ULYSS analysis, as is the cut on double-lined CaT spectra. The cut on the  $\log \chi^2$  of the fit has been set to remove stars that are outliers by  $> +3\sigma$  from the median. It removed 60 stars on top of the two other cuts.

<sup>7</sup>The line wings beyond 1.5 Å from the line centre can be safely used in LTE analyses, therefore the central 1.5 Å of each line were masked. However, because of the running mean normalization, some of the NLTE information from the core may be spread to the neighbouring pixels. To test whether this significantly affects the analysis, we compare the results from FERRE runs with the blue arm only and with both arms together. There are no significant systematic changes in the metallicities between the two runs, although there are some systematic differences in  $\log g$ . It is unclear whether this is because the CaT spectra add necessary information to get better  $\log g$  values, or whether this is a NLTE effect. Since there are no significant changes in the metallicities, we assume that it is sufficient to mask the cores of the lines.

### 3.5.3 Stellar parameter uncertainties

The same method was used as with the ULYSS analysis to investigate the magnitude of the formal FERRE uncertainties, using 2785 independent pairs of observations. We find close-to-Gaussian distributions of  $\Delta P$  (see equation 2) with a standard deviation of 0.1 for each of the three main parameters ( $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$ ) and 0.07 for  $[\text{C}/\text{Fe}]$ . This means that the formal FERRE uncertainties are severely overestimated, by a factor of  $\sim 10$ . No satisfactory explanation for this has been found. After re-scaling the formal uncertainties into internal uncertainties using the factors derived from the repeated-observations test, the values appear reasonable. The median internal FERRE uncertainties are 105 K, 0.34 dex, 0.09 dex, and 0.12 dex for  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and  $[\text{C}/\text{Fe}]$ , respectively. This is larger than for the ULYSS analysis.

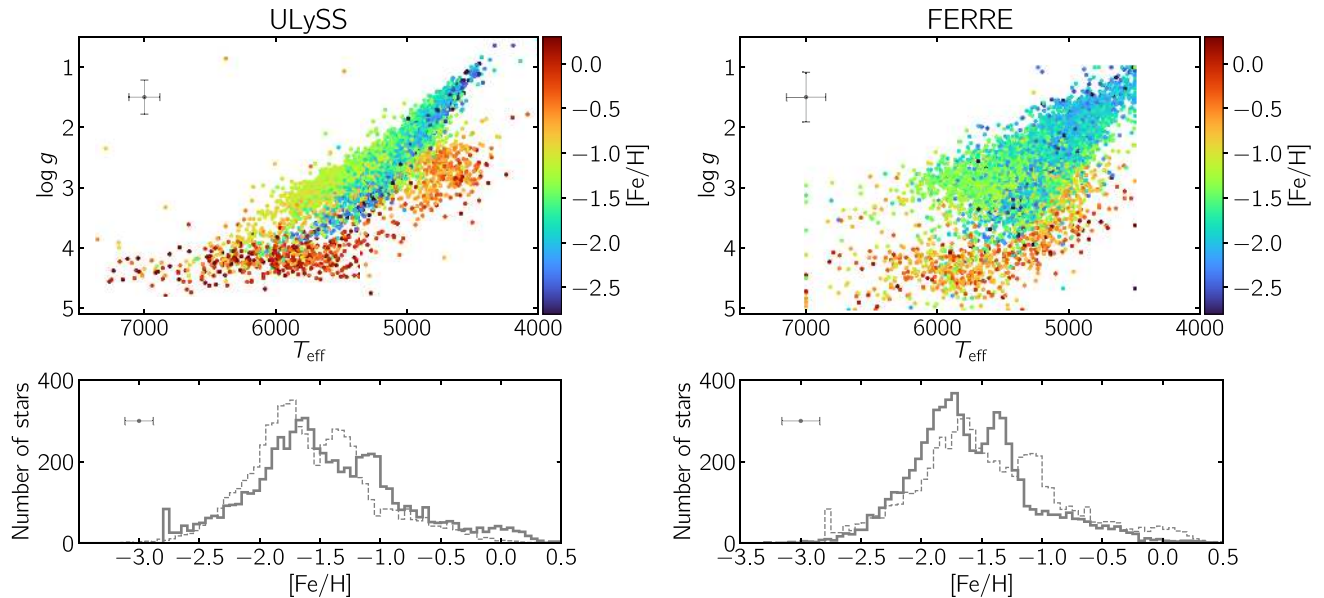
## 4 STELLAR PARAMETER RESULTS

The results of the ULYSS and FERRE analyses of the PIGS low-/intermediate-resolution spectra are presented here. They are compared to each other to identify regions where the methods have issues and to derive the total uncertainties. The results are also compared to those in the SDSS APOGEE data base for a subsample of stars.

### 4.1 Metallicity distribution and Kiel diagram

There are 7073 PIGS stars that passed the quality criteria for both the ULYSS and FERRE analysis methods. The Kiel diagrams from ULYSS and FERRE for these stars are shown in the top row of Fig. 5, the metallicity distributions in the bottom row. There are several noteworthy features in the Kiel diagrams and metallicity distributions, seen in both analyses. There are a number of metal-rich dwarf stars ( $[\text{Fe}/\text{H}] > -0.5$ ,  $\log g > 3.5$ ) in our sample, which are in the sample mainly due to the *Gaia* selection (as can be seen in Fig 2; and more of these entered our selection box in higher extinction regions), or due to the selection testing in our two pilot fields. However, most of the stars in the sample are giants, consisting of three types. First, there are the metal-poor red giant branch (RGB) stars which were the target of our survey ( $[\text{Fe}/\text{H}] < -1.5$ ). These are the most numerous, and correspond to the largest peak in the metallicity distributions in Fig. 5. Secondly, there are the hotter horizontal branch (HB) stars, with slightly higher metallicities ( $-1.5 < [\text{Fe}/\text{H}] < -1.0$ ). These are unintentionally part of our selection because they are relatively blue in colour and therefore also enter our selection box. They are less numerous than the RGB stars, but constitute a significant fraction of the sample and make up the second peak of the metallicity distribution. These stars are extremely useful, since they are metal-poor standard candles. Finally, there is a small number of metal-rich giants ( $[\text{Fe}/\text{H}] > -1.0$ ), the result of ‘normal’ contamination in our photometric selection. There is such a large number of metal-rich stars in the bulge that it is to be expected that some of them scatter into our selection box by accident, especially in higher extinction regions.

There are clear differences between the ULYSS and FERRE analyses, which are the result of different strengths and limitations of the two methods. The features in the ULYSS Kiel diagram are narrower, which we attribute to the data-driven nature of this method. Data-driven methods have two properties that contribute to creating such narrow features: the reference library contains only ‘physical’ stars that do indeed exist, and the differences in stellar parameters between each of the library spectra is irregular, i.e. not constant. The latter aids in the quality of the results by learning



**Figure 5.** Top row: Kiel diagram, colour-coded by  $[\text{Fe}/\text{H}]$  from ULYSS (left-hand panel) and FERRE (right-hand panel). Each panel shows the stars that pass the quality criteria for both methods. Bottom row: metallicity distributions, where dashed histogram indicates the distribution with the other method. Median final uncertainties for metal-poor ( $[\text{Fe}/\text{H}] < -1.0$ ) giants are shown in the top left corner of each panel. There are three grid effects visible in this figure: the  $[\text{Fe}/\text{H}]$  limit in the ULYSS analysis is  $-2.8$ , and the FERRE grid limits are  $4500 < T_{\text{eff}} < 7000$  K and  $1 < \log g < 5$ .

to interpolate between smaller offsets, unlike the use of synthetic grids, which have regular grid step spacing, thereby missing non-linear interpolation opportunities. Also, synthetic grids cover all possible combinations of stellar parameters (even those that should not exist), providing a larger parameter space for exploration, even non-physical regimes. And finally, most synthetic spectral grids are missing relevant physics, such as non-LTE corrections for some elements which can affect the stellar parameters (e.g. Kovalev et al. 2019).

Significant efforts are underway to create and/or extend existing empirical stellar spectral libraries to include more of the rarer stars (e.g. XSL, Gonneau et al. 2020), although they remain limited. In the PIGS sample, this limitation is particularly relevant for the HB stars and for the most metal-poor stars. In the MILES library and its interpolator (which is used in the ULYSS analysis), only a limited number of HB stars are present, all with similar metallicities ( $[\text{Fe}/\text{H}] \approx -1.1$ ; Prugniel et al. 2011). It is therefore suspicious that in the ULYSS analysis, we find a peak of HB stars at exactly the same metallicity. The FERRE analysis should be unbiased in this regime, and finds the HB metallicities to be  $\sim 0.3$  dex lower. There are also only a handful of very metal-poor stars ( $[\text{Fe}/\text{H}] < -2.5$ ) in MILES. The limit of the interpolator is at  $[\text{Fe}/\text{H}] = -2.8$ , hence the spike in the ULYSS metallicity distribution there. In contrast, the FERRE synthetic grid goes to  $[\text{Fe}/\text{H}] = -4.0$ , therefore the results for the metal-poor tail are expected to be more trustworthy.

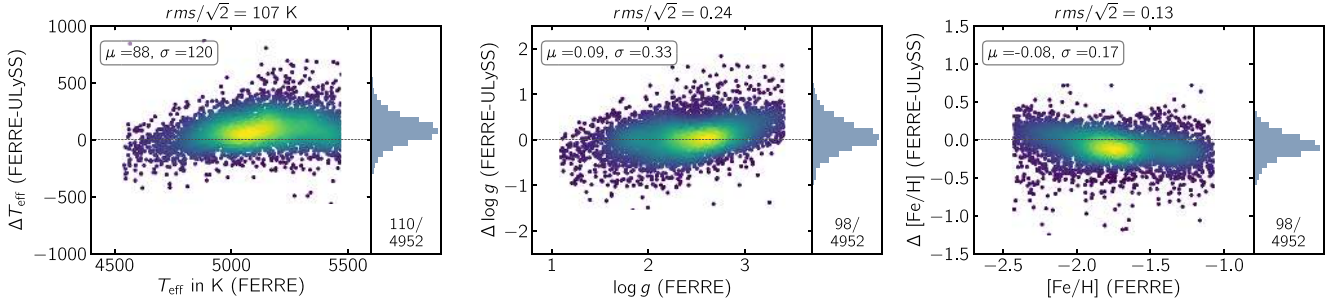
We highlight three final differences between the ULYSS and FERRE analyses. First, the metal-rich dwarfs are more metal-rich in ULYSS than in FERRE. This is due to the limitations of the FERRE grid chosen for this analysis, which only contains  $\alpha$ -enhanced spectra. In our selected magnitude range, dwarf stars lie between us and the bulge and are likely part of the disc, and metal-rich disc stars are typically not enhanced in  $\alpha$ -elements. Therefore our FERRE analysis does not derive the proper metallicity for these stars. Secondly, there are a number of relatively metal-rich stars that ULYSS classifies as giants with  $\log g \sim 3.0$ , and FERRE as dwarfs/sub-giants

( $\log g > 3.5$ ). Inspection of the spectra of these stars shows that they have Ca H&K in emission: these are active stars. This also explains why they ended up as contamination in our photometric selection: the emission filled in the Ca II H&K absorption lines, making them look metal-poor in the *CaHK* photometry. Retrospectively, we found that many of these stars can be removed using a *Gaia* DR2 selection cut based on a combination of the mean flux over error for the *G*-band and the number of repeated observations. Finally, there are two sequences for the giants, which have different metallicities between ULYSS and FERRE. In the ULYSS analysis, there are a number of stars with slightly higher  $[\text{Fe}/\text{H}]$  and lower  $\log g$  on the hot side of the RGB. In the FERRE analysis, these stars also have lower  $\log g$  but similar  $[\text{Fe}/\text{H}]$  to the main RGB stars. These stars might be early asymptotic giant branch (AGB) stars. This hypothesis is strengthened by slightly lower carbon abundances for these stars (early AGB stars have lowered carbon abundances due to previous internal mixing processes on the RGB). This observation will be discussed in more detail in future work (Arentsen et al., in preparation).

#### 4.2 Total uncertainties

In Sections 3.4.3 and 3.5.3 we provided internal uncertainties for the ULYSS and FERRE analyses. These indicate the precision of the derived stellar parameters in both methods. We can use the comparison between ULYSS and FERRE to determine the *external* precision, which can be combined with the internal uncertainties to derive total uncertainties.

The difference in stellar parameters between the two methods is presented in Fig. 6, in a region of the parameter space for which both methods are expected to behave well (based on the observations in the previous section). There are some systematic differences between the two analyses. The FERRE temperatures are on average 88 K higher than those from ULYSS, where the discrepancy for the higher temperature stars is slightly larger. The mean difference for  $\log g$  is 0.09 dex higher for FERRE, and it is somewhat larger for the



**Figure 6.** Comparison of the parameters between ULYSS and FERRE colour-coded by density, for giant stars ( $\log g < 3.5$ ) cooler than 5500 K and with  $-2.5 < [\text{Fe}/\text{H}] < -1.0$ . The zero-line is indicated. The mean difference and standard deviation are indicated in the top left-hand corner of each main panel, after discarding  $3\sigma$  outliers. The number of outliers versus the total number of stars are indicated on the bottom of the histogram panels. Finally, the title of each panel includes the corresponding  $rms/\sqrt{2}$ , which is quadratically added to the internal uncertainties from Section 3.4.3 and 3.5.3 to derive total uncertainties for both analyses.

highest  $\log g$  stars. Finally, the mean difference in  $[\text{Fe}/\text{H}]$  is 0.08 dex (where FERRE is more metal-poor), varying slightly with  $[\text{Fe}/\text{H}]$ . The standard deviations around the mean are 120 K, 0.33 dex, and 0.17 dex for  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$ , respectively.

A better measure of the external precision than the standard deviation around the mean can be made using the root-mean-square ( $rms$ ), which includes the biases in the deviations from the mean. The external uncertainty can be estimated as  $rms/\sqrt{2}$ , which assumes that both methods contribute equally to the dispersion. These values have been indicated in the title of each panel in Fig. 6, and are 107 K, 0.24 dex, and 0.13 dex for  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ , respectively (after discarding  $3\sigma$  outliers). Total uncertainties can now be derived by quadratically adding the internal ULYSS or FERRE uncertainties and the  $rms/\sqrt{2}$ . This results in the following median total uncertainties for  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  in the ULYSS analysis: 111 K, 0.25 dex, and 0.14 dex, respectively, and in the FERRE analysis: 149 K, 0.41 dex, and 0.16 dex, respectively.

The external uncertainties on  $[\text{C}/\text{Fe}]$  cannot be derived in the same way, since it is not a parameter in the ULYSS analysis. Instead, they are estimated by scaling the  $[\text{Fe}/\text{H}]$  external uncertainties. In the FERRE analysis, the internal uncertainties on  $[\text{C}/\text{Fe}]$  are roughly 1.5 times larger than the  $[\text{Fe}/\text{H}]$  internal uncertainties. Assuming that the external uncertainties follow the same trend, an uncertainty floor of  $1.5 \times 0.13$  dex = 0.20 dex is added to derive total uncertainties on  $[\text{C}/\text{Fe}]$ . The median of the total uncertainties for  $[\text{C}/\text{Fe}]$  is 0.23 dex.

### 4.3 Comparison with APOGEE

One additional comparison was made with fully external data, by comparing our derived stellar parameters with APOGEE for stars in common between the surveys. Because most bulge surveys contain mainly metal-rich stars, there generally is a lack of stars in common with this work. For instance, in APOGEE DR16 (SDSS-IV Collaboration 2020), there are only a handful of stars in common with PIGS. This is because APOGEE selects stars ‘randomly’ with respect to metallicity, which in the bulge results in a predominantly metal-rich sample. However, we are collaborating with the bulge Cluster APOGEE Survey (CAPOS; Geisler et al., in preparation) to obtain APOGEE follow-up of more PIGS stars. To date, one CAPOS field which includes PIGS targets has been fully observed (two additional fields are scheduled to be observed/finished) resulting in a total of 69 stars in common between PIGS and APOGEE DR16 + CAPOS with good stellar parameters from both surveys (62 of these come from

CAPOS, 7 from DR16). These stars have SNR in APOGEE between 30 and 150, and median uncertainties in  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{M}/\text{H}]$  are 134 K, 0.09 dex, and 0.03 dex, respectively.

The AAT observations of the CAPOS field are unfortunately among the worst in the PIGS sample due to relatively bad weather conditions and the sky subtraction is suboptimal. This affects our FERRE analysis more, because that includes the CaT spectra where the sky subtraction is more important. Additionally, the APOGEE pipeline struggles to provide good stellar parameters for very metal-poor stars (see e.g. fig. 9 of Leung & Bovy 2019, which shows that the median absolute error in  $[\text{Fe}/\text{H}]$  rises quickly for  $[\text{Fe}/\text{H}] < -1.7$ , even in very high signal-to-noise spectra, until it reaches about 0.4 dex at  $[\text{Fe}/\text{H}] = -2.3$ ). The APOGEE-PIGS comparison should therefore not be overinterpreted, but it is still a useful additional indication of our external precision.

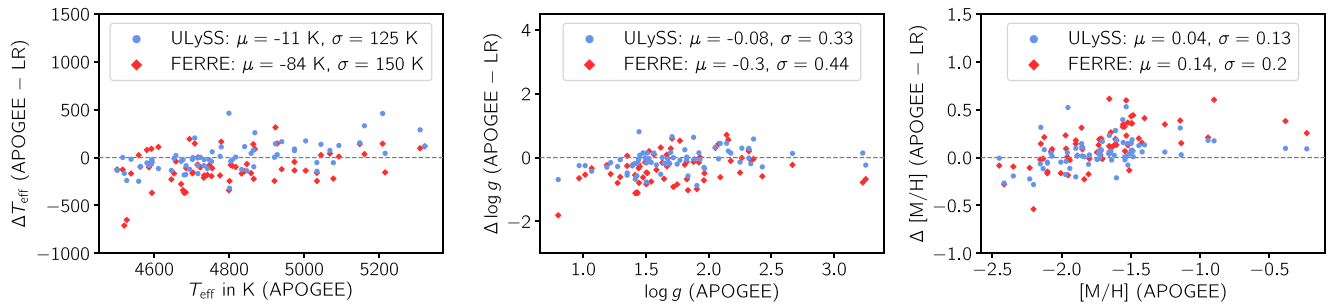
The comparison between the PIGS and APOGEE parameters is presented in Fig. 7. The APOGEE  $[\text{M}/\text{H}]$  was used as metallicity, since for many metal-poor stars no  $[\text{Fe}/\text{H}]$  is available or uncertain. The dispersion in the parameters between APOGEE and the ULYSS analysis is comparable to that of the internal comparison (Fig. 6), whereas the dispersions for our FERRE analysis are somewhat larger, as expected due to the lower quality CaT spectra. The biases in the ULYSS-APOGEE comparison are negligible, but in the FERRE analysis they are more significant, and in the same direction as in the internal comparison (our FERRE analysis having larger  $T_{\text{eff}}$ , larger  $\log g$  and lower  $[\text{Fe}/\text{H}]$ ). This could suggest that the biases in the internal comparison are mainly due to FERRE.

It is important to realize that this comparison with APOGEE is only in a limited range of the parameter space. However, it does show that the typical total uncertainties we have derived from our analysis appear to be realistic.

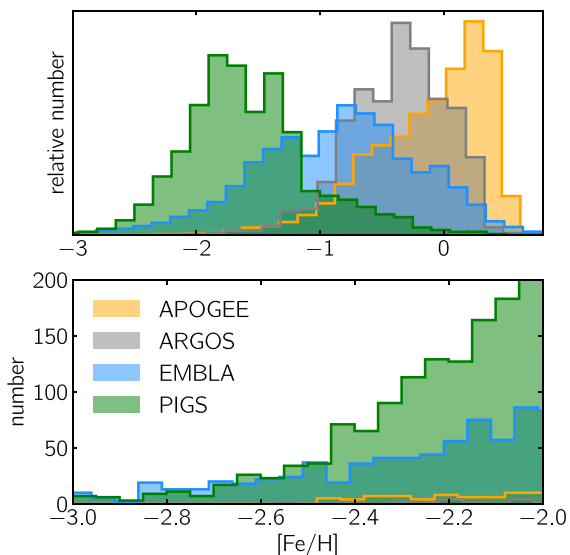
## 5 DISCUSSION OF PIGS SURVEY PERFORMANCE

### 5.1 Metallicity distribution

Due to the nature of the PIGS selection, the sample is significantly more metal-poor than typical inner Galaxy surveys like for example ARGOS (Ness et al. 2013a) and APOGEE (SDSS-IV Collaboration 2020). The PIGS metallicity distribution relative to these two inner Galaxy surveys is presented in the top panel of Fig. 8. The metallicity distribution from the low-resolution EMBLA survey sample (Howes



**Figure 7.** Comparison of the parameters between APOGEE and ULYSS and FERRE, for cool giants ( $T_{\text{eff}} < 5500$ ,  $\log g < 3.5$ ). The mean difference and standard deviation, after discarding  $3\sigma$  outliers, are indicated in the legend of each panel.



**Figure 8.** Top panel: Relative metallicity distributions of the PIGS, EMBLA, ARGOS, and APOGEE inner Galaxy surveys. For PIGS, the FERRE  $[\text{Fe}/\text{H}]$  for the AAT sample is shown, with the footprint as in Fig. 1. The EMBLA metallicities are those for their AAT sample (provided by L. Howes, private communication), with a footprint  $|l| \lesssim 12^\circ$  and  $3^\circ \lesssim |b| \lesssim 12^\circ$ . The ARGOS sample consists of all fields with field centres  $|l|, |b| \leq 10^\circ$  (metallicities provided by M. Ness, private communication). The APOGEE sample shows the  $[\text{M}/\text{H}]$  of all DR16 stars in the region  $|l|, |b| < 12^\circ$  (SDSS-IV Collaboration 2020). Bottom panel: absolute number of VMP stars in the same samples.

et al. 2016) is also included. It is clear that the PIGS sample is unique compared to existing bulge surveys in its large number of low-metallicity stars.

The PIGS sample contains  $\sim 1300/160/10$  stars with  $[\text{Fe}/\text{H}] < -2.0/-2.5/-3.0$ , respectively (from the FERRE analysis). The ten most metal-poor stars are spread between  $-3.5 < [\text{Fe}/\text{H}] < -3.0$ . This is the largest sample of confirmed very metal-poor (VMP,  $[\text{Fe}/\text{H}] < -2.0$ ) stars towards the inner Galaxy, as shown in the bottom panel of Fig. 8. It is also larger than the sample of VMP stars published by Aguado et al. (2019) for the main *Pristine* survey, although their sample contains many more stars with  $[\text{Fe}/\text{H}] < -2.5$  compared to PIGS. In EMBLA there is a similar number of stars with  $[\text{Fe}/\text{H}] < -2.5$  as in PIGS, but the total size of their sample is twice as large as PIGS. There is only one star in common between PIGS and EMBLA with  $[\text{Fe}/\text{H}] < -2.5$  and twenty with  $[\text{Fe}/\text{H}] < -2.0$ , because in most fields we removed the EMBLA stars from our selection. The PIGS stars with  $[\text{Fe}/\text{H}] < -2.5$  are

distributed relatively homogeneously over the different fields, see the right-hand panel of Fig. 1.

In addition to the VMP stars, the PIGS sample contains  $\sim 4900$  stars with  $-2.0 < [\text{Fe}/\text{H}] < -1.0$ , an unprecedentedly large sample of metal-poor inner Galaxy stars to be used to study the metal-poor tail of the Galactic bulge. A first investigation of this population was presented in Arentsen et al. (2020), using the metallicities and radial velocities to investigate the rotational signature in the inner Galaxy as a function of metallicity. That work shows that the rotational signal around the Galactic centre decreases with decreasing  $[\text{Fe}/\text{H}]$ , while the velocity dispersion increases. Rotation then becomes negligible for the most metal-poor stars, providing a strong constraint for any models attempting to explain the formation and evolution of the inner Galaxy.

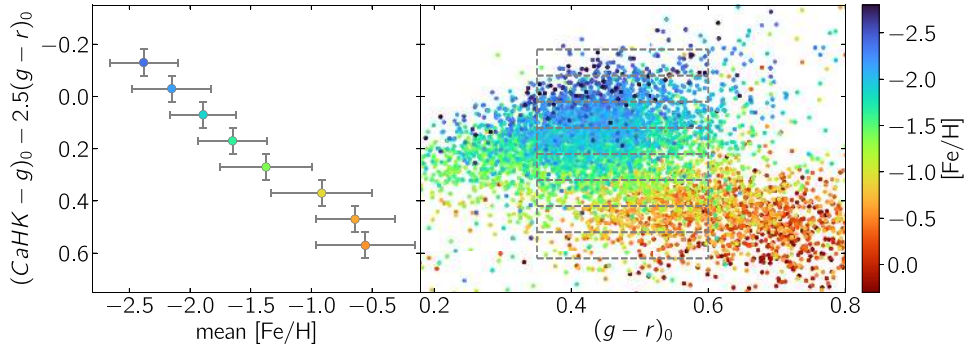
More detailed investigations of this metal-poor sample, including analysis of the orbits of these stars using (future) *Gaia* data, are planned for future work. Stars of special interest for dynamical studies are the  $\sim 1700$  HB stars, for which, being standard candles, accurate distances may be inferred.

## 5.2 Survey success rates

Using our spectroscopic follow-up sample and spectroscopic observations from the literature, we investigate how well we can produce photometric metallicities in PIGS. For individual stars, photometric metallicities are likely not good, for the reasons described in Section 2. Instead, the average metallicity and the metallicity dispersion were determined in sub-regions of the *Pristine* colour–colour diagram and we investigate how clean a selection can be made for intermediate metal-poor (IMP,  $[\text{Fe}/\text{H}] < -1.5$ ) and VMP ( $[\text{Fe}/\text{H}] < -2.0$ ) stars.

In this test, the FERRE spectroscopic metallicities were used instead of those from ULYSS, because they are typically better for the VMP stars. We have seen that FERRE is biased for  $[\text{Fe}/\text{H}] > -0.5/-1.0$  because the synthetic grid used only includes alpha-enhanced spectra (see Section 3.5). However, for the purpose of the tests in this section we included data from three additional spectroscopic surveys that have many more stars with higher metallicities, making the small contribution of metal-rich PIGS stars irrelevant. The three surveys which were added are EMBLA (low-resolution sample, L. Howes, private communication), ARGOS (M. Ness, private communication) and APOGEE DR16 (SDSS-IV Collaboration 2020).

In the right-hand panel of Fig. 9 the *Pristine* colour–colour diagram is shown including all spectroscopic data samples described above. For stars with  $0.35 < (g - r)_0 < 0.6$ , where the discrimination between different metallicities appears to work best and where most of the stars of interest are located, it is sliced along the y-axis in



**Figure 9.** Mean metallicity in bins of the y-axis of the Pristine colour–colour diagram (left-hand panel) and the corresponding diagram with the indicated bins (right-hand panel). Error bars in the left-hand panel are the  $1\sigma$  dispersion in each bin for  $[\text{Fe}/\text{H}]$ , and for the y-axis they indicate the size of the bin. Included are stars from PIGS (using their FERRE  $[\text{Fe}/\text{H}]$ ), and stars from EMBLA, APOGEE DR16 (using  $[\text{M}/\text{H}]$ ) and ARGOS. Only stars with  $E(B - V) < 0.6$  are included.

bins of 0.1 mag. The left-hand panel shows the mean  $[\text{Fe}/\text{H}]$  and the dispersion in  $[\text{Fe}/\text{H}]$  in each bin. Only stars with  $E(B - V) < 0.6$  are used (corresponding to roughly Galactic  $b < -4.5^\circ$ ) to avoid strong outliers in very high extinction regions. Additionally, a selection cut using the number of repeated observations and mean flux over error for the  $G$ -band photometry in *Gaia* DR2 was included, to remove stars with variable photometry from the PIGS sample.

Fig. 9 shows that there is an almost linear relation between the y-axis colour and the metallicity in this regime, which confirms that our box selection for follow-up is a good way of selecting the most metal-poor stars. The  $[\text{Fe}/\text{H}]$  dispersion in the bins ranges from 0.30 to 0.45 dex, and the typical  $[\text{Fe}/\text{H}]$  difference between bins is  $\sim 0.2$  dex or more. The largest dispersion is in the intermediate metallicity range, possibly because it contains the largest mixture of stars of different surveys (and each has been analysed differently), or because the metallicity-sensitivity in this regime of the diagram is more non-linear.

We use the relative number of IMP and VMP stars in our best selection bins to estimate the PIGS selection efficiency for metal-poor stars in the inner Galaxy. In the most metal-poor bin, 86 per cent and 80 per cent of the stars have  $[\text{Fe}/\text{H}] < -2.0$ , for regions with  $E(B - V) < 0.4$  and  $> 0.4$ , respectively. If, instead, the two most metal-poor bins are combined the number of stars is increased five-fold (to 500). In these bins, 80 per cent of the stars have  $[\text{Fe}/\text{H}] < -2.0$  and 98 per cent have  $[\text{Fe}/\text{H}] < -1.5$ , for regions with  $E(B - V) < 0.4$ . This goes down to 63 per cent and 93 per cent, respectively, in  $E(B - V) > 0.4$  regions. These efficiencies do not change significantly when using  $(g - i)$  as the effective temperature indicator instead of  $(g - r)$ . One might expect the former to work better because the  $g$  and  $i$  filters cover a longer wavelength baseline, and should thus yield better temperature estimates. However, because they are further apart, the extinction correction of the difference between the two filters is likely less good. These two effects appear to roughly cancel each other in PIGS.

We compare our efficiency to the results from Casey & Schlafman (2015), using 2MASS and WISE photometry to identify VMP stars in the bulge region, who report a success rate of 20 per cent of their candidates having  $-3.0 < [\text{Fe}/\text{H}] < -2.0$ . The EMBLA survey has not published success rates, but the metallicity distribution of their low-resolution follow-up sample peaks around  $[\text{Fe}/\text{H}] = -1.0$  (Howes et al. 2016, see also Fig. 8), whereas PIGS peaks at lower metallicities. Additionally, as discussed in Section 5.1, their sample has a similar number of stars with  $[\text{Fe}/\text{H}] < -2.5$ , but twice as many stars in total. We therefore conclude that our selection is more

efficient. Taken together, this shows that the efficiency of PIGS for selecting metal-poor stars in the bulge region is unprecedented.

## 6 SUMMARY AND OUTLOOK

In this paper, we have presented the Pristine Inner Galaxy Survey. The  $\sim 250 \text{ deg}^2$  of photometric  $\text{CaHK}$  observations in the bulge region, the data reduction and calibration, and the low-/intermediate resolution spectroscopic follow-up of  $\sim 8000$  metal-poor candidate stars were described in detail. Robust stellar parameters and metallicities were determined for the spectroscopic data using two completely independent analysis methods: (1) ULYSS, a full-spectrum fitting package using an empirical library, and (2) FERRE, a full-spectrum fitting method employing a synthetic stellar model library. Our analysis with ULYSS is restricted to the blue arm only, whereas FERRE can fit both blue and red arms of our observations simultaneously. The application of two methods allowed us to make an in-depth comparison between the merits and weaknesses of both methods in the context of analysing metal-poor stars, some of which are inherent to the empirical/synthetic nature of their model spectra. The use of these two complementary methods further enables the estimation of an external uncertainty on top of the internal uncertainties of each of the methods separately. An additional, completely independent, comparison has been carried out with overlapping stars observed with APOGEE.

As a result of its effective selection of metal-poor targets, PIGS data at the moment includes 1300 stars with spectroscopic  $[\text{Fe}/\text{H}] < -2.0$ , the largest sample of confirmed very metal-poor stars in the inner Galaxy to date. Combining spectroscopy and photometry, we tested the performance of our photometric selection of metal-poor stars. An efficient selection of  $[\text{Fe}/\text{H}] < -2.0$  stars can be made; PIGS has a 86 per cent/80 per cent (intermediate/high extinction) success rate using the most conservative selection (producing only a small sample), and a success rate of 80 per cent/63 per cent for a sample five times larger. PIGS is therefore the perfect source of metal-poor candidate targets for future spectroscopic surveys in the inner Galaxy, such as for instance the multi-object spectroscopic 4MOST Milky Way Disc and Bulge Low-Resolution Survey (4MIDABLE-LR; Chiappini et al. 2019).

Additionally, our sample includes a large sample of metal-poor horizontal branch stars. These stars are interesting in their own right, especially because of their standard candle nature.

Some follow-up work using the sample and analysis presented here is already underway. The third PIGS paper will present carbon

abundance results from the FERRE analysis, including an investigation of the sensitivity of the Pristine filter to the carbon abundance, results on the carbon-enhanced metal-poor stars in the inner Galaxy, and an investigation of the evolution of the carbon abundance in metal-poor giant stars. A further paper in the PIGS series will focus in more detail on the dynamical properties of the metal-poor stars in the inner Galaxy using (future) *Gaia* data. We plan to derive distances and orbital properties for the spectroscopic metal-poor sample ( $[\text{Fe}/\text{H}] < -1.0$ ). Additionally, as proper motions in the inner Galaxy become more reliable with future *Gaia* data releases, it may be possible to include the much larger photometric samples of metal-poor stars in kinematic analyses. Finally, we note that there is an ongoing high-resolution spectroscopic follow-up of PIGS stars that will be described elsewhere.

In the coming years, large samples of metal-poor inner Galaxy stars, such as provided through PIGS and other (future) surveys (for example *Gaia* DR3+ and 4MOST; de Jong et al. 2019) will provide a unique view of the central regions of the Milky Way. The chemistry and kinematics of these stars will lead to insights into the early history of the Milky Way, and large samples of them will allow for the discovery of rare stars whose abundance patterns teach us about previous, early generations of stars, all the way back to the First Stars.

## ACKNOWLEDGEMENTS

We thank Friedrich Anders for discussions on the extinction in *Gaia*, and for sharing his *Gaia* extinction coefficients. We thank Ting S. Li for valuable discussions and efforts regarding the fibre throughput issues of 2dF. We thank Louise Howes and Melissa Ness for sharing their spectroscopic catalogues with us, this was very helpful in the early development of PIGS and for this paper. We thank Gal Matijević for performing the cross-match between the *Pristine CaHK* photometry and *Gaia* DR2.

We thank the Australian Astronomical Observatory, which have made these observations possible. We acknowledge the traditional owners of the land on which the AAT stands, the Gamilaraay people, and pay our respects to elders past and present. Based on data obtained at Siding Spring Observatory (via programs S/2017B/01, A/2018A/01, OPTICON 2018B/029 and OPTICON 2019A/045, PI: A. Arentsen).

Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii.

Based on observations obtained through the Chilean National Telescope Allocation Committee through programs CN2017B-37, CN2018A-20, CN2018B-46, CN2019A-98, and CN2019B-31.

AA, ES, and KY gratefully acknowledge funding by the Emmy Noether program from the Deutsche Forschungsgemeinschaft (DFG). NFM gratefully acknowledges support from the French National Research Agency (ANR) funded project ‘Pristine’ (ANR-18-CE31-0017) along with funding from INSU/CNRS through the Programme National Galaxies et Cosmologie and through the CNRS grant PICS07708. DA thanks the Leverhulme Trust for financial support. KAV thank the Natural Sciences and Engineering Research Council for funding through the Discovery Grants program and the CREATE training program on New Technologies for Canadian Observatories. DBZ acknowledges the support of the Australian Research Council through Discovery Project grant DP180101791. CAP and JIGH acknowledge financial support from the Spanish Min-

istry of Science and Innovation (MICIIN) under the project MICIIN AYA2017-86389-P. JIGH also acknowledges the Spanish MICIIN under 2013 Ramón y Cajal program MICIIN RYC-2013-14875. DG gratefully acknowledges support from the Chilean Centro de Excelencia en Astrofísica y Tecnologías Afines (CATA) BASAL grant AFB-170002. DG also acknowledges financial support from the Dirección de Investigación y Desarrollo de la Universidad de La Serena through the Programa de Incentivo a la Investigación de Académicos (PIA-DIDULS). Horizon 2020: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 730890. This material reflects only the authors views and the Commission is not liable for any use that may be made of the information contained therein.

The authors thank the International Space Science Institute, Bern, Switzerland for providing financial support and meeting facilities to the international team ‘Pristine’.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen’s University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This research has made use of ‘Aladin sky atlas’ developed at CDS, Strasbourg Observatory, France (Bonnarel et al. 2000). It also made extensive use of the MATPLOTLIB (Hunter 2007), PANDAS (McKinney 2010), ASTROPY (Astropy Collaboration 2013; Price-Whelan et al. 2018) and DUSTMAPS (Green 2018) Python packages, and of TOPCAT (Taylor 2005).

This paper is based on observations made with the Canada-France-Hawaii Telescope (CFHT) and the Anglo-Australian Telescope (AAT).

## REFERENCES

- Aguado D. S., González Hernández J. I., Allende Prieto C., Rebolo R., 2017, *A&A*, 605, A40  
 Aguado D. S. et al., 2019, *MNRAS*, 490, 2241  
 Allende Prieto C., Beers T. C., Wilhelm R., Newberg H. J., Rockosi C. M., Yanny B., Lee Y. S., 2006, *ApJ*, 636, 804

- Alvarez R., Plez B., 1998, *A&A*, 330, 1109
- Arentsen A. et al., 2019, *A&A*, 627, A138
- Arentsen A. et al., 2020, *MNRAS*, 491, L11
- Astropy Collaboration, 2013, *A&A*, 558, A33
- Barbuy B., Chiappini C., Gerhard O., 2018, *ARA&A*, 56, 223
- Bonnarel F. et al., 2000, *A&AS*, 143, 33
- Casey A. R., Schlaufman K. C., 2015, *ApJ*, 809, 110
- Chambers K. C. et al., 2016, preprint (arXiv:1612.05560)
- Chiappini C. et al., 2019, *Messenger*, 175, 30
- Da Costa G. S. et al., 2019, *MNRAS*, 489, 5900
- de Jong R. S. et al., 2019, *Messenger*, 175, 3
- Dékány I., Minniti D., Catelan M., Zoccali M., Saito R. K., Hempel M., Gonzalez O. A., 2013, *ApJ*, 776, L19
- Diemand J., Madau P., Moore B., 2005, *MNRAS*, 364, 367
- Falcón-Barroso J., Sánchez-Blázquez P., Vazdekis A., Ricciardelli E., Cardiel N., Cenarro A. J., Gorgas J., Peletier R. F., 2011, *A&A*, 532, A95
- Frebel A., Norris J. E., 2015, *ARA&A*, 53, 631
- Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
- Freeman K. et al., 2013, *MNRAS*, 428, 3660
- Gaia Collaboration, 2016, *A&A*, 595, A1
- Gaia Collaboration, 2018, *A&A*, 616, A1
- García Pérez A. E. et al., 2018, *ApJ*, 852, 91
- Gonneau A. et al., 2020, *A&A*, 634, A133
- Green G., 2018, *J Open Source Softw.*, 3, 695
- Green G. M. et al., 2018, *MNRAS*, 478, 651
- Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, *ApJ*, 887, 93 (G19)
- Gustafsson B., Edvardsson B., Eriksson K., Jørgensen U. G., Nordlund Å., Plez B., 2008, *A&A*, 486, 951
- Hill V. et al., 2011, *A&A*, 534, A80
- Howard C. D. et al., 2009, *ApJ*, 702, L153
- Howes L. M. et al., 2014, *MNRAS*, 445, 4241
- Howes L. M. et al., 2015, *Nature*, 527, 484
- Howes L. M. et al., 2016, *MNRAS*, 460, 884
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Irwin M., Lewis J., 2001, *New Astron. Rev.*, 45, 105
- Koch A., McWilliam A., Preston G. W., Thompson I. B., 2016, *A&A*, 587, A124
- Koesterke L., Allende Prieto C., Lambert D. L., 2008, *ApJ*, 680, 764
- Koleva M., Prugniel P., Bouchard A., Wu Y., 2009, *A&A*, 501, 1269
- Kovalev M., Bergemann M., Ting Y.-S., Rix H.-W., 2019, *A&A*, 628, A54
- Kunder A. et al., 2012, *AJ*, 143, 57
- Kunder A. et al., 2016, *ApJ*, 821, L25
- Leung H. W., Bovy J., 2019, *MNRAS*, 483, 3255
- Lewis I. J. et al., 2002, *MNRAS*, 333, 279
- Li T. S. et al., 2019, *MNRAS*, 490, 3508
- Lucey M. et al., 2019, *MNRAS*, 488, 2283
- Magnier E. A., Cuillandre J. C., 2004, *PASP*, 116, 449
- Magnier E. A., Lewin W. H. G., van Paradijs J., Hasinger G., Jain A., Pietsch W., Truemper J., 1992, *A&AS*, 96, 379
- McKinney W., 2010, in van der Walt S., Millman J., eds, *Proceedings of the 9th Python in Science Conference*. p. 56
- Mészáros S. et al., 2012, *AJ*, 144, 120
- Ness M. et al., 2012, *ApJ*, 756, 22
- Ness M. et al., 2013a, *MNRAS*, 432, 2092
- Ness M. et al., 2013b, *MNRAS*, 430, 836
- Placco V. M., Frebel A., Beers T. C., Stancliffe R. J., 2014, *ApJ*, 797, 21
- Plez B., 2008, *Phys. Scripta Vol. T*, 133, 014003
- Price-Whelan A. M. et al., 2018, *AJ*, 156, 123
- Prugniel P., Vauglin I., Koleva M., 2011, *A&A*, 531, A165
- Sánchez-Blázquez P. et al., 2006, *MNRAS*, 371, 703
- Saunders W. et al., 2004, in Moorwood A. F. M., Iye M., eds, *Proc. SPIE Conf. Ser. Vol. 5492, Ground-based Instrumentation for Astronomy*. SPIE, Bellingham, p. 389
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103 (SF11)
- Schlafly E. F. et al., 2016, *ApJ*, 821, 78
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Schultheis M. et al., 2017, *A&A*, 600, A14
- SDSS-IV Collaboration, 2020, *ApJS*, 249, 3
- Sharma K., Prugniel P., Singh H. P., 2016, *A&A*, 585, A64
- Sharp R. et al., 2006, in McLean I. S., Iye M., eds, *Proc. SPIE Conf. Ser. Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy*. SPIE, Bellingham, p. 62690G
- Shen J., Rich R. M., Kormendy J., Howard C. D., De Propris R., Kunder A., 2010, *ApJ*, 720, L72
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Starkenburger E. et al., 2017a, *MNRAS*, 471, 2587
- Starkenburger E., Oman K. A., Navarro J. F., Crain R. A., Fattahi A., Frenk C. S., Sawala T., Schaye J., 2017b, *MNRAS*, 465, 2212 (S17b)
- Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, *ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV*. Astron. Soc. Pac., San Francisco, p. 29
- Tolstoy E., Hill V., Tosi M., 2009, *ARA&A*, 47, 371
- Tumlinson J., 2010, *ApJ*, 708, 1398
- Venn K. A. et al., 2020, *MNRAS*, 492, 3241
- Wolf C. et al., 2018, *PASA*, 35, e010
- Yong D. et al., 2013, *ApJ*, 762, 27
- York D. G. et al., 2000, *AJ*, 120, 1579
- Youakim K. et al., 2017, *MNRAS*, 472, 2963
- Zasowski G. et al., 2019, *ApJ*, 870, 138
- Zoccali M. et al., 2014, *A&A*, 562, A66
- <sup>1</sup>Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany
- <sup>2</sup>Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France
- <sup>3</sup>Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
- <sup>4</sup>Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- <sup>5</sup>Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia
- <sup>6</sup>Instituto de Astrofísica de Canarias, Vía Láctea, E-38205 La Laguna, Tenerife, Spain
- <sup>7</sup>Universidad de La Laguna, Departamento de Astrofísica, E-38206 La Laguna, Tenerife, Spain
- <sup>8</sup>Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Blvd de l'Observatoire, F-06304 Nice, France
- <sup>9</sup>Department of Physics and Astronomy, University of Victoria, Victoria, BC V8W 3P2, Canada
- <sup>10</sup>Department of Astronomy and Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada
- <sup>11</sup>Institute of Astronomy of the Russian Academy of Sciences, Pyatnitskaya st. 48, 119017 Moscow, Russia
- <sup>12</sup>UK Astronomy Technology Centre, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK
- <sup>13</sup>NRC Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada
- <sup>14</sup>Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, NSW 2006, Australia
- <sup>15</sup>School of Physics, UNSW, Sydney, NSW 2052, Australia
- <sup>16</sup>Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
- <sup>17</sup>Departamento de Astronomía, Universidad de Concepción, Casilla 160-C, Concepción, Chile
- <sup>18</sup>Instituto de Investigación Multidisciplinario en Ciencia y Tecnología, Universidad de La Serena, Avenida Raúl Bitrán S/N, La Serena, Chile
- <sup>19</sup>Departamento de Astronomía, Facultad de Ciencias, Universidad de La Serena, Av. Juan Cisternas 1200, La Serena, Chile

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.