Unveiling systematic biases in the 1D LTE excitation–ionization balance of Fe for FGK stars: a novel approach to determination of stellar parameters

Gregory R. Ruchti,[1][*][†] Maria Bergemann,[1] Aldo Serenelli,[2] Luca Casagrande[3] and Karin Lind[1]

[1]*Max Planck Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany*
[2]*Instituto de Ciencias del Espacio (CSIC-IEEC), Facultad de Ciencias, Campus UAB, E-08193 Bellaterra, Spain*
[3]*Research School of Astronomy & Astrophysics, Mount Stromlo Observatory, The Australian National University, ACT 2611, Australia*

ABSTRACT

We present a comprehensive analysis of different techniques available for the spectroscopic analysis of FGK stars and provide a recommended methodology which efficiently estimates accurate stellar atmospheric parameters for large samples of stars. Our analysis includes a simultaneous equivalent width analysis of Fe I and Fe II spectral lines, and for the first time, utilizes on-the-fly non-local thermodynamic equilibrium (NLTE) corrections of individual Fe I lines. We further investigate several temperature scales, finding that estimates from Balmer line measurements provide the most accurate effective temperatures at all metallicities. We apply our analysis to a large sample of both dwarf and giant stars selected from the Radial Velocity Experiment (RAVE) survey. We then show that the difference between parameters determined by our method and that by the standard 1D LTE excitation–ionization balance of Fe reveals substantial systematic biases: up to 400 K in effective temperature, 1.0 dex in surface gravity and 0.4 dex in metallicity for stars with [Fe/H] ~ -2.5. This has large implications for the study of the stellar populations in the Milky Way.

Key words: stars: abundances – stars: late-type – stars: Population II – radiative transfer – methods: data analysis – techniques: spectroscopic.

1 INTRODUCTION

The fundamental atmospheric (effective temperature, surface gravity and metallicity) and physical (mass and age) parameters of stars provide the major observational foundation for chemo-dynamical studies of the Milky Way and other galaxies in the Local Group. With the dawn of large spectroscopic surveys to study individual stars, such as SEGUE (Yanny et al. 2009), RAVE (Steinmetz et al. 2006), Gaia-ESO (Gilmore et al. 2012) and HERMES (Barden et al. 2008), these parameters are used to infer the characteristics of different populations of stars that comprise the Milky Way.

Stellar parameters determined by spectroscopic methods are of key importance. The only way to accurately measure metallicity is through spectroscopy, which thus underlies photometric calibrations (e.g. Holmberg, Nordström & Andersen 2007; An et al. 2009; Árnadóttir, Feltzing & Lundström 2010; Casagrande et al. 2011), while high-resolution spectroscopy is also used to correct the low-resolution results (e.g. Carollo et al. 2010). The atmospheric

parameters can all be estimated from a spectrum in a consistent and efficient way. This also avoids the problem of reddening inherent in photometry since spectroscopic parameters are not sensitive to reddening. The spectroscopic parameters can then be used alone or in combination with photometric information to fit individual stars to theoretical isochrones or evolutionary tracks to determine the stellar mass, age and distance of a star.

A common method for deriving the spectroscopic atmospheric parameters is to use the information from Fe I and Fe II absorption lines under the assumption of hydrostatic equilibrium (HE) and local thermodynamic equilibrium (LTE). Many previous studies have used some variation of this technique (e.g. ionization or excitation equilibrium) to determine the stellar atmospheric parameters and abundances, and henceforth distances and kinematics, of FGK stars in the Milky Way. For example, some have used this procedure to estimate the effective temperature, surface gravity and metallicity of a star (e.g. Fulbright 2000; Prochaska et al. 2000; Johnson 2002), while others use photometric estimates of effective temperature in combination with the ionization equilibrium of the abundance of iron in LTE to estimate surface gravity and metallicity (e.g. McWilliam et al. 1995; François et al. 2003; Bai et al. 2004; Allende Prieto et al. 2006; Lai et al. 2008).

---

[*] Current address: Lund Observatory, Department of Astronomy and Theoretical Physics, Box 43, SE-22100, Lund, Sweden.
[†] E-mail: gruchti@MPA-Garching.MPG.DE

However, both observational (e.g. Fuhrmann 1998; Ivans et al. 2001; Ruchti et al. 2011; Bruntt et al. 2012) and theoretical evidence (e.g. Thévenin & Idiart 1999; Asplund 2005; Mashonkina et al. 2011) suggest that systematic biases are present within such analyses due to the breakdown of the assumption of LTE. More recently, Bergemann et al. (2012b) and Lind, Bergemann & Asplund (2012) quantified the effects of non-local thermodynamic equilibrium (NLTE) on the determination of surface gravity and metallicity, revealing very substantial systematic biases in the estimates at low metallicity and/or surface gravity. It is therefore extremely important to develop sophisticated methods, which reconcile these effects in order to derive accurate spectroscopic parameters.

This is the first in a series of papers, in which we develop new, robust methods to determine the fundamental parameters of FGK stars and then apply these techniques to large stellar samples to study the chemical and dynamical properties of the different stellar populations of the Milky Way. In this work, we utilize the sample of stars selected from the RAVE survey originally published in Ruchti et al. (2011, hereafter R11) to formulate the methodology to derive very accurate atmospheric parameters. We consider several temperature scales and show that the Balmer line method is the most reliable among the different methods presently available. Further, we have developed the necessary tools to apply on-the-fly NLTE corrections[1] to Fe I lines, utilizing the grid described in Lind et al. (2012). We verify our method using a sample of standard stars with interferometric estimates of effective temperature and/or *Hipparcos* parallaxes. We then perform a comprehensive comparison to standard 1D, LTE techniques for the spectral analysis of stars, finding significant systematic biases.

2 SAMPLE SELECTION AND OBSERVATIONS

NLTE effects in iron are most prominent in low-metallicity stars (Lind et al. 2012; Bergemann et al. 2012b). We therefore chose the metal-poor sample from R11 for our study. These stars were originally selected for high-resolution observations based on data obtained by the RAVE survey in order to study the metal-poor thick disc of the Milky Way. Spectral data for these stars were obtained using high-resolution echelle spectrographs at several facilities around the world.

Full details of the observations and data reduction of the spectra can be found in R11. Briefly, all spectrographs delivered a resolving power greater than 30 000 and covered the full optical wavelength range. Further, nearly all spectra had signal-to-noise ratios greater than 100:1 per pixel. The equivalent widths (EWs) of both Fe I and Fe II lines, taken from the line lists of Fulbright (2000) and Johnson (2002), were measured using the ARES code (Sousa et al. 2007). However, during measurement quality checks, we found that the continuum was poorly estimated for some lines. We therefore determined EWs for these affected lines using hand measurements.

3 STELLAR PARAMETER ANALYSES

We computed the stellar parameters for each star using two different methods.

In the first method, which is commonly used in the literature, we derived an effective temperature, $T_{\text{LTE-Fe}}$, surface gravity, log $g_{\text{LTE-Fe}}$,

metallicity, $[\text{Fe/H}]_{\text{LTE-Fe}}$, and microturbulence, $v_{\text{LTE-Fe}}$, from the ionization and excitation equilibrium of Fe in LTE. This is hereafter denoted as the LTE-Fe method. We used an iterative procedure that utilized the MOOG analysis program (Sneden 1973) and 1D, plane-parallel ATLAS-ODF model atmospheres from Kurucz[2] computed under the assumption of LTE and HE (Castelli & Kurucz 2003, and references therein). In our procedure, the stellar effective temperature was set by minimizing the magnitude of the slope of the relationship between the abundance of iron from Fe I lines and the excitation potential of each line. Similarly, the microturbulent velocity was found by minimizing the slope between the abundance of iron from Fe I lines and the reduced EW of each line. The surface gravity was then estimated by minimizing the difference between the abundance of iron measured from Fe I and Fe II lines. Iterations continued until all of the criteria above were satisfied. Finally, $[\text{Fe/H}]_{\text{LTE-Fe}}$ was chosen to equal the abundance of iron from the analysis. Our results for this method are described in Section 4.

The second method, denoted as the NLTE-Opt method, consists of two parts. First, we determined the optimal effective temperature estimate, T_{Opt}, for each star (see Section 5 for more details). Then, we utilized MOOG to compute a new surface gravity, log $g_{\text{NLTE-Opt}}$, metallicity, $[\text{Fe/H}]_{\text{NLTE-Opt}}$, and microturbulence, $v_{\text{NLTE-Opt}}$. This was done using the same iterative techniques as the LTE-Fe method, that is the ionization balance of the abundance of iron from Fe I and Fe II lines.

There are, however, three important differences. First, the stellar effective temperature was held fixed to the optimal value, T_{Opt}. Secondly, we restricted the analysis to Fe lines with excitation potentials above 2 eV, since these lines are less sensitive to 3D effects as compared to the low-excitation lines (see the discussion in Bergemann et al. 2012b). Thirdly, the abundance of iron from each Fe I line was adjusted according to the NLTE correction for that line at the stellar parameters of the current iteration in the procedure. The NLTE corrections were determined using the NLTE grid computed in Lind et al. (2012) and applied on-the-fly via a wrapper program to MOOG. Note that the NLTE calculations presented in Lind et al. (2012) were analogously calibrated using the ionization equilibria of a handful of well-known stars. Our extended sample, including more stars with direct measurements of surface gravity and effective temperature, provides support for the realism of this calibration. The grid extends down to log $g = 1$. We imposed a routine which linearly extrapolated the NLTE corrections to below this value. The results of extrapolations were checked against NLTE grids presented in Bergemann et al. (2012a) and no significant differences were found. Further, Lind et al. (2012) found very small NLTE corrections for Fe II lines. We therefore do not apply any correction to the Fe II lines.

Iterations continued until the difference between the average abundance of iron from the Fe II lines and the NLTE-adjusted Fe I lines was in agreement (within ±0.05 dex) and the slope of the relationship between the reduced EW of the Fe I lines and their NLTE-adjusted iron abundance was minimized. Sections 5 and 6 describe our final stellar parameter estimates for this method.

4 INITIAL LTE-FE PARAMETERS

The initial LTE-Fe stellar parameters for our sample stars are listed in Table 1. Residuals in the minimizations of this technique gave typical internal errors of 0.1 dex in both log $g_{\text{LTE-Fe}}$ and $[\text{Fe/H}]_{\text{LTE-Fe}}$ and ~55 K in $T_{\text{LTE-Fe}}$. As we show in the following sections, these

---

[1] 'NLTE correction' refers to the difference between the abundance of iron computed in LTE and NLTE obtained from a line with a given equivalent width.

[2] See http://kurucz.harvard.edu/.

Table 1. Atmospheric stellar parameter data.

| Star | LTE-Fe | | | | | $T_{\rm IRFM}$ (K) | $\sigma_{T_{\rm IRFM}}$ (K) | $E(B-V)$ | $T_{\rm R11}$ (±140 K) | $T_{\rm Bal}$ (K) | $\sigma_{T_{\rm Bal}}$ (K) | $T_{\rm eff}$ (K) | $\sigma_{T_{\rm eff}}$ (K) | NLTE-Opt | | |
	$T_{\rm eff}$ (K)	$\sigma_{T_{\rm eff}}$ (K)	$\log g$ (±0.1)	[Fe/H] (±0.1)	v									$\log g$ (±0.1)	[Fe/H] (±0.1)	v
C0023306−163143	5128	58	2.40	−2.63	1.3	–	–	–	5528	5400	100	5443	101	3.20	−2.29	0.9
C0315358−094743	4628	40	1.51	−1.40	1.5	–	–	–	4722	4800	100	4774	89	2.06	−1.31	1.6
C0408404−462531	4466	40	0.25	−2.25	2.2	–	–	–	4600	–	–	4600	90	1.03	−2.10	2.1
C0549576−334007	5151	50	2.53	−1.94	1.3	–	–	–	5379	5400	100	5393	82	3.16	−1.70	1.1
C1141088−453528	4439	40	0.39	−2.42	2.1	–	–	–	4592	4500	200	4562	123	1.10	−2.28	1.9

This table is published in its entirety in the electronic edition of the MNRAS. A portion is shown here for guidance regarding its form and content.
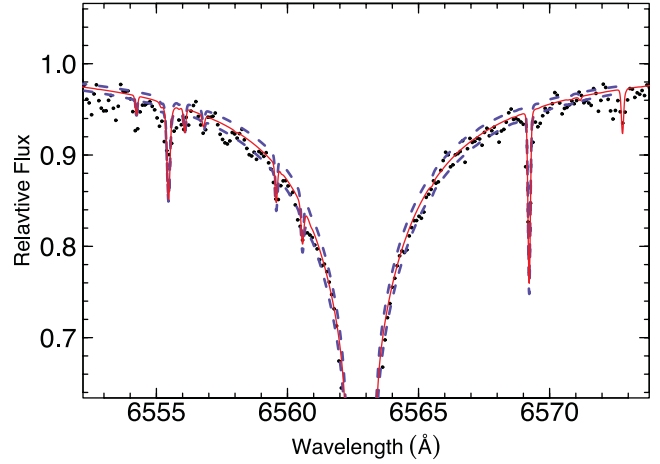

Figure 1. Example fit to Hα in the spectrum of the metal-poor ([Fe/H] $\sim$ −1.0) dwarf, J142911.4−053131. The solid, red curve shows the best fit to the data (at 5700 K), while the blue, dashed curves represent ±100 K around the fit.

small internal errors can be quite misleading as they are not representative of the actual accuracy of stellar parameter estimates. Often, especially in metal-poor stars, estimates of $T_{\rm eff}$, $\log g$ and [Fe/H] that result from this method are far too low when compared to other more accurate data (cf. R11).

5 EFFECTIVE TEMPERATURE OPTIMIZATION

It was found in Bergemann et al. (2012b) and Lind et al. (2012) that taking into account NLTE in the solution of excitation equilibrium does not lead to a significant improvement of the stellar effective temperature. This was also supported by our test calculations for a sub-sample of stars. Fe I lines formed in LTE or NLTE are still affected by convective surface inhomogeneities and overall different mean temperature/density stratifications, which are most prominent in strong low-excitation Fe I lines (Shchukina, Trujillo Bueno & Asplund 2005; Bergemann et al. 2012b). Using 1D hydrostatic models with either LTE or NLTE radiative transfer thus leads to effective temperature estimates that are too low when the excitation balance of Fe I lines is used (see below). It is therefore important that the stellar effective temperature be estimated by other means.

5.1 Three effective temperature scales

We used three different methods to compute the effective temperature.

The first estimate, $T_{\rm Bal}$, was derived from the wings of the Balmer lines, which is among the most reliable methods available for the effective temperature determination of FGK stars (e.g. Fuhrmann, Axer & Gehren 1993; Fuhrmann 1998; Barklem et al. 2002; Cowley & Castelli 2002; Gehren et al. 2006; Mashonkina et al. 2008). The only restriction of this method is that for stars cooler than 4500 K, the wings of H I lines become too weak to allow reliable determination of $T_{\rm eff}$. Profile fits of Hα and Hβ lines were performed by careful visual inspection of different portions of the observed spectrum in the near and far wings of the Balmer lines which were free of contaminant stellar lines. Figs 1 and 2 show two example fits to Hα. Note that the Balmer lines were self-contained within a single order in each spectrum. Therefore, we did not use neighbouring orders for the continuum normalization.
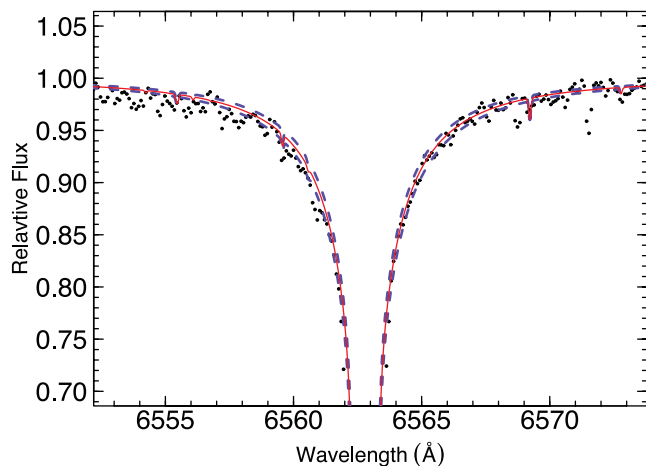
Figure 2. Example fit to Hα in the spectrum of the metal-poor ([Fe/H] ∼ −2.3) giant, J230222.8−683323. The solid, red curve shows the best fit to the data (at 5200 K), while the blue, dashed curves represent ±100 K around the fit.

Table 2. Effective temperatures determined from direct interferometric measurements of angular diameters.

HD	log g	[Fe/H]	T_{Int} (K)	$\sigma_{T_{\text{Int}}}$ (K)	T_{Bal} (K)	$\sigma_{T_{\text{Bal}}}$ (K)	Ref.
6582	4.50	−0.70	5343	18	5295	100	a
107 00	4.50	−0.50	5376	22	5320	100	a
220 49	4.50	0.00	5107	21	5050	100	a
228 79	4.23	−0.86	5786	16	5800	100	b*
276 97	2.70	0.00	4897	65	4900	100	c
283 05	2.00	0.00	4843	62	4800	100	c
291 39	1.22	−0.22	3871	48	4000	200	c
499 33	4.21	−0.42	6635	18	6530	100	b
614 21	3.90	−0.10	6555	17	6500	100	a
625 09	2.88	0.12	4858	60	4870	100	c
849 37	4.00	−2.00	6275	17	6315	100	b*
855 03	2.50	0.30	4433	51	4450	100	b
100 407	2.87	−0.04	5044	33	5025	100	b
102 870	4.00	0.20	6062	20	6075	100	a
121 370	4.00	0.20	5964	18	5975	100	a
122 563	1.65	−2.50	4598	42	4650	100	d
124 897	1.60	−0.54	4226	53	4240	200	c
140 283	3.70	−2.50	5720	29	5775	100	b*
140 573	2.00	0.00	4558	56	4610	100	c
150 680	4.00	0.00	5728	24	5795	100	a
161 797	4.00	0.20	5540	27	5550	100	a
215 665	2.25	0.12	4699	71	4800	100	c

References for T_{Int}: a – Cayrel et al. (2011); b – *Gaia* calibration stars from Heiter (private communication) (those marked with a [*] have angular diameters determined using the surface–brightness relations from Kervella et al. 2004); c – Mozurkewich et al. (2003); d – Creevey et al. (2012).

Theoretical profiles were computed using the SIU code with `MAFAGS-ODF` model atmospheres (Fuhrmann 1998; Grupp 2004a). Same as `ATLAS-ODF` (Section 3), the MAFAGS models were computed with Kurucz opacity distribution functions; thus, the differences between the model atmosphere stratifications are expected to be minimal in our range of stellar parameters. For self-broadening of H lines, we used the Ali & Griem (1965) theory. As shown by Grupp (2004a) this method successfully reproduces the Balmer line spectrum of the Sun within 20 K and provides accurate stellar parameters that agree very well with *Hipparcos* astrometry (Grupp 2004b). The errors are obtained directly from profile fitting, and they are largely internal, ±50 to 100 K.

A key advantage of the Balmer lines is that they are insensitive to interstellar reddening, which affects photometric techniques (see below). However, the Balmer line effective temperature scale could be affected by systematic biases, caused by the physical limitation of the models. The influence of deviations from LTE in the H I line formation in application to cool metal-poor stars was studied by Mashonkina et al. (2008). Comparing our results to the NLTE estimates by Mashonkina et al. (2008) for the stars in common, we obtain ΔT_{eff}(our − M08) = −70 K (HD 122 563 metal-poor giant), ΔT_{eff}(our − M08) = 50 K (HD 849 37, metal-poor turn-off). The difference is clearly within the T_{eff} uncertainties. On the other hand, it should be kept in mind that the atomic data for NLTE calculations for hydrogen are of insufficient quality and, at present, do not allow accurate quantitative assessment of NLTE effects in H, as elaborately discussed by Barklem (2007). Likewise, the influence of granulation is difficult to assess. Ludwig et al. (2009) presented 3D effective temperature corrections for Balmer lines for a few points on the HRD, for which 3D radiative-hydrodynamics simulations of stellar convection are available. For the Sun,[3] they find ΔT_{eff}(3D − 1D) ≈ 35 K, and for a typical metal-poor subdwarf with [Fe/H] =−2, ΔT_{eff}(3D − 1D) of the order of 50 to 80 K (average over Hα, Hβ and Hγ). However, in the absence of consistent 3D NLTE calculations, it is not possible to tell whether 3D and NLTE effects will amplify or cancel for FGK stars. Thus, we do not apply any theoretical corrections to our Balmer effective temperatures.

Currently, the only way to understand whether our Balmer T_{eff} scale is affected by systematics is by comparing with independent methods, in particular interferometry. We, therefore, computed the Balmer T_{eff} for several nearby stars with direct and indirect interferometric angular diameter measurements. The results are listed in Table 2, while we plot the difference between our Balmer estimate and that from interferometry in Fig. 3. Both T_{eff} scales show an agreement of 3 ± 60 K for stars with [Fe/H] > −1, while the Balmer estimate is ∼50 K warmer than T_{Int} at the lowest metallicities. These differences are well within the combined errors in the interferometric and Balmer measurements. This suggests that deviations from 1D HE *and* LTE are either minimal or affect both interferometric and Balmer T_{eff} in exactly the same way. Also note that, for the stars in common with Cayrel et al. (2011), our estimates are fully consistent.

For the second method, we utilized the Tycho-2 and 2MASS photometry of each star to compute effective temperature estimates, T_{IRFM}, using the infrared flux method (IRFM), as presented in Casagrande et al. (2010, hereafter C10). Note that in C10 the IRFM calculations were applied only to dwarfs and subgiants and validated by comparison with a large body of interferometric angular diameters. However, the same code can be safely applied to lower surface gravities, as shown by comparison with newly determined angular diameters for giants (Huber et al. 2012). The advantage of this method is that it is much less sensitive to model assumptions that are required for spectroscopic analyses. However, the quality of the photometric data used to compute IRFM effective temperatures, as well as interstellar reddening, can still largely affect the result. In our case reddening has been estimated using the Drimmel, Cabrera-Lavers & López-Corredoira (2003) map, with distances
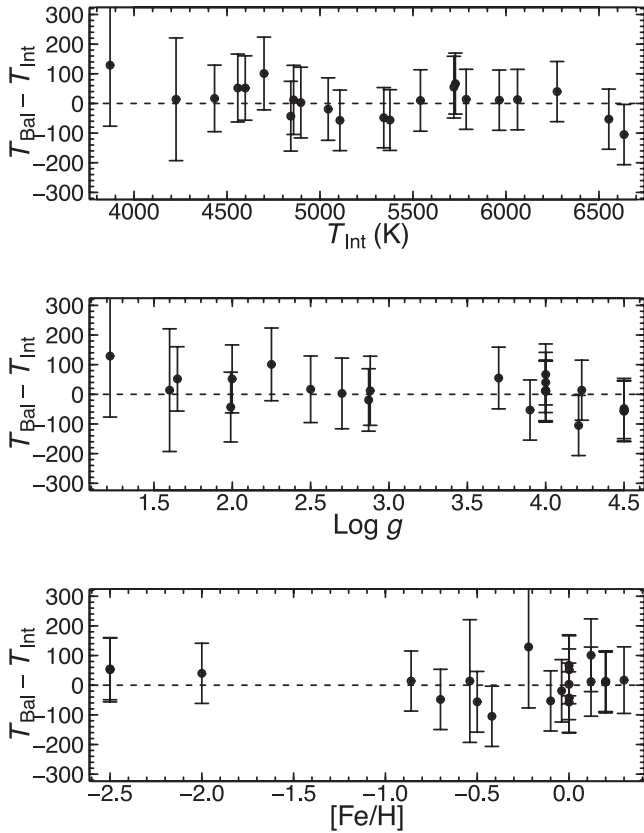
---

[3] Here, we use their results obtained with $\alpha_{\text{MLT}} = 0.5$ consistent with the `MAFAGS-ODF` model atmospheres adopted here.

Figure 3. Comparison of the Balmer effective temperature estimate to that from interferometry, as listed in Table 2. The Balmer effective temperature estimates agree with the interferometric estimate to within ∼100 K.

derived from our spectroscopic log *g*. Typical values of $E(B - V)$ are around 0.05 mag, although for some of the brightest giants the value can be considerably larger (see Table 1).

Finally, we chose the effective temperature estimates from R11 (denoted as T_{R11}) as our third effective temperature scale. These estimates were based on the R11 calibration, which was derived from the trend between $[Fe/H]_{LTE-Fe}$ and the difference between T_{LTE-Fe} and the 2MASS $J - K_S$ photometric effective temperature for several globular cluster stars, *Hipparcos* stars and low-reddened stars in the R11 sample. In principle, this method should yield similar effective temperature estimates to that of the IRFM, since it utilizes the colour-temperature transformations presented in González Hernández & Bonifacio (2009), which were based upon their IRFM calculations. The advantage is that the R11 calibration relies on the $J - K_S$ colour, which is less sensitive to reddening ($E(J - K_S) \sim 0.5E(B - V)$). However, note that the $J - K_S$ colour correlates only mildly with T_{eff}, and thus calibrations involving that index exhibit a rather larger internal dispersion (in our case 139 K for dwarfs and 94 K for giants) when compared to IRFM effective temperature estimates for standard stars. Further, effective temperatures computed using the calibration in C10 are typically ∼40 K hotter than those computed in González Hernández & Bonifacio (2009). A detailed explanation for this discrepancy is given in C10.

5.2 Comparisons

We next applied each of the above methods to our sample stars, the values of which can be found in Table 1. Note that for several stars, the Balmer lines fell in the middle of the order of the spectrum.
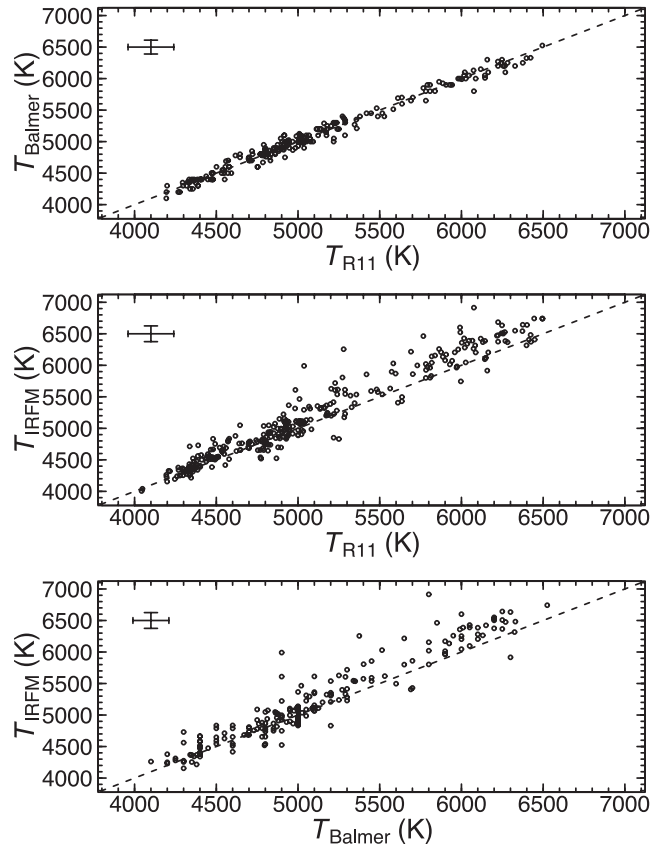

Figure 4. Comparison of the three different T_{eff} scales. Both the estimate from R11 (T_{R11}) and that from the Balmer lines (T_{Bal}) agree within ∼80 K, while that from the IRFM (T_{IRFM}) is systematically higher than both and shows an increasing dispersion with increasing effective temperature. Error bars, which show the mean error in each T_{eff} scale, are displayed in the upper-left corner.

The continuum cannot be determined with sufficient accuracy in such regions. We therefore did not measure the Balmer lines for those stars. Further, not all stars in our sample have Tycho-2 photometry estimates. We were unable to compute an IRFM effective temperature estimate for these stars.

Fig. 4 shows the comparisons between the three effective temperature estimates when applied to our sample stars. The estimates from T_{Bal} and T_{R11} show remarkable agreement, with a difference of only 1 ± 79 K. The IRFM effective temperatures, however, are systematically higher than both T_{Bal} and T_{R11} (by 119 ± 215 and 113 ± 186 K, respectively), with an increasing dispersion towards hotter effective temperatures.

It is possible that inherent NLTE and 3D effects could be influencing the Balmer effective temperature scale; however, we see excellent agreement with T_{R11} and interferometric measurements. As stated previously, the effective temperature estimates computed in C10 are about 40 K warmer than those computed in González Hernández & Bonifacio (2009). Further, the $J - K_S$ calibration has a large internal dispersion. However, the difference between T_{IRFM} and T_{R11} extends well beyond these limits.

It is possible that the uncertainty in the interstellar reddening may be systematically affecting the IRFM estimates. In order to test the accuracy of the estimates of reddening from the Drimmel et al. (2003) map, we also tried to measure $E(B - V)$ using the interstellar Na D lines. However, the majority of the stars in our sample had multi-component interstellar Na D features, or the feature was not
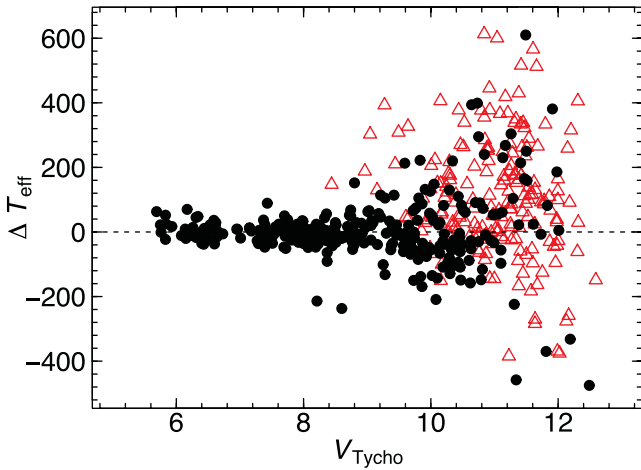
Figure 5. An illustration of the difference between different effective temperature estimates as a function of V_{Tycho}. The black, filled circles represent the difference between $T(V_{\text{Tycho}} - K_{\text{s}})$ and $T(V_{\text{JC}} - K_{\text{s}})$ for the sample of stars in C10. The red, open triangles show the difference between T_{IRFM} and T_{Bal} for the stars in our sample. For $V_{\text{Tycho}} \geq 9$, both samples display an increased scatter in ΔT_{eff}. This clearly indicates that the uncertainty in V_{Tycho} is poorly constrained for those stars.

discernible from the stellar Na lines. Note that, for several of the stars in which we could measure single-component interstellar Na D lines, the Drimmel et al. estimates and the Na D estimates were on average different by <0.02 mag, which translates to a difference of <100 K in effective temperature. Given these differences, reddening alone cannot account for the very large differences (≥ 300 K) for many stars.

Instead, the large scatter mostly likely arises from the poor quality of the Tycho-2 magnitudes for stars fainter than $V_{\text{Tycho}} \sim 9$. For the kind of stars analysed in this work, B and V magnitudes are the dominant contributors to the bolometric flux, as compared to the infrared 2MASS magnitudes. Should a star be matched with a brighter (dimmer) source in B and V, then the bolometric flux will be over-estimated (underestimated) by a very large amount, and the IRFM will return a systematically higher (lower) T_{IRFM} estimate.

Using the sample in C10, it is possible to compute T_{eff} using both the Tycho-2 $V_{\text{Tycho}} - K_{\text{s}}$ and Johnson–Cousins $V_{\text{JC}} - K_{\text{s}}$ calibrations. We plot the difference between $T(V_{\text{Tycho}} - K_{\text{s}})$ and $T(V_{\text{JC}} - K_{\text{s}})$ as a function of the Tycho-2 V magnitude for the C10 sample (black points) in Fig. 5. In addition, we have over-plotted the difference between T_{IRFM} and T_{Bal} for the stars in our present sample (red triangles). As shown in the figure, both samples exhibit a large scatter in the difference in effective temperature estimates at $V_{\text{Tycho}} \geq 9$. Further, the hotter stars are on average among the faintest stars in our sample and thus have larger errors in V_{Tycho}. This is a clear indicator that the poor quality of the Tycho-2 photometric measurements for our stars is responsible for the discrepancy between T_{IRFM} and T_{Bal}.

5.3 Final effective temperature estimates

From the comparisons above, Balmer line measurements provide the most reliable effective temperature estimates for all stars in our sample. In addition, T_{R11} exhibits small differences with respect to T_{Bal} across all effective temperatures, as illustrated in Fig. 4. In contrast, IRFM effective temperatures appear to show a large dispersion, which we attribute to large errors in Tycho-2 photometry,

as well as uncertainty in the reddening. We therefore adopted the mean of only T_{Bal} and T_{R11}, weighted according to the internal errors from each method, as our final T_{Opt} estimate. This also serves to reduce the internal error on the final optimal effective temperature estimate, which was typically $\lesssim 100$ K.

As noted previously, we could not measure the Balmer lines in several of the stars in our sample. For those stars, we adopted the T_{R11} estimate. The authors of R11 adopted an error of 140 K in their T_{R11} estimate, which was derived from the residuals in their calibration. However, the comparison between T_{Bal} and T_{R11} in Fig. 4 suggests that this value was overestimated. The mean error in T_{Opt}, for those stars with both a T_{Bal} and T_{R11} estimate, was 90 K. We therefore adopted this value for stars with only a single T_{R11} estimate. Our final T_{Opt} values and corresponding errors can be found in Table 1.

6 SURFACE GRAVITY AND [Fe/H] IN NLTE

Using the final values of T_{Opt} described in the previous section, we derived the remaining stellar parameters using the NLTE-Opt method described in Section 3. We first validated this methodology by applying our analysis to a sample of 18 'standard' stars, which have *Hipparcos* parallaxes. The spectra for these stars were obtained for the analysis in Fulbright (2000), and are of similar quality to our sample. Both the LTE-Fe and NLTE-Opt atmospheric parameters for each standard star are given in Table 3. Using the *Hipparcos* parallax and an estimate of the bolometric correction derived from the bolometric flux relations presented in González Hernández & Bonifacio (2009), we also computed an 'astrometric surface gravity' ($g_\pi = 4\pi G M \sigma T^4 / L$) for each star, which is listed as $\log g_\pi$ in Table 3. Note that we computed an astrometric surface gravity using other various flux relations (Alonso, Arribas & Martinez-Roger 1995; C10; Torres 2010), finding results within $\lesssim 0.1$ dex of that computed using the relation in González Hernández & Bonifacio (2009). Further, we assumed a mass of $0.8 \, \text{M}_\odot$ for [Fe/H] < -1 and $0.9 \, \text{M}_\odot$ for [Fe/H] ≥ -1. However, a difference of $0.1 \, \text{M}_\odot$ will only change the astrometric surface gravity by ~ 0.05 dex. The NLTE-Opt surface gravity estimates show a remarkable 0.02 ± 0.11 dex agreement with the astrometric surface gravity, while the LTE-Fe estimates are too low by -0.32 ± 0.39 dex.

Given the agreement, we applied the above analysis to our sample stars. The final NLTE-Opt estimates for surface gravity and metallicity, as well as for the microturbulence, can be found in Table 1. We adopted 0.1 dex error in both the surface gravity and metallicity, based on our comparisons with the standard stars above.
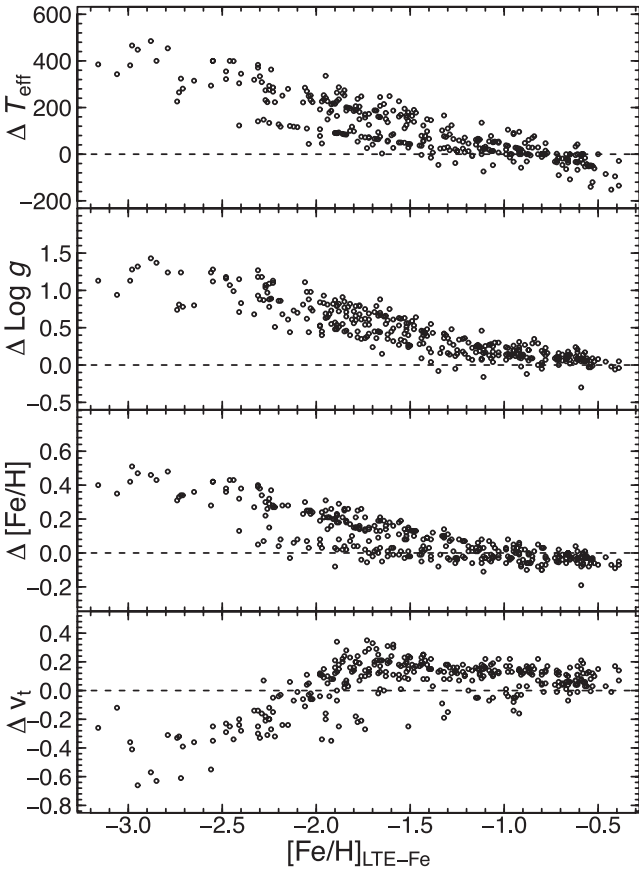
7 NLTE-Opt VERSUS LTE-Fe

In Fig. 6, we compare our final NLTE-Opt stellar parameters to those derived using the LTE-Fe method. The differences in the estimates of effective temperature, surface gravity, metallicity and microturbulence all display clear trends with decreasing metallicity. The microturbulent velocity is underestimated by ~ 0.1–0.2 km s^{-1} until [Fe/H]$_{\text{LTE-Fe}} \sim -1.8$, where $v_{\text{LTE-Fe}}$ becomes larger than $v_{\text{NLTE-Opt}}$. The differences between T_{eff} range from 200 to 400 K for metal-poor giants, and -50 to -100 K for dwarfs. The differences for $\log g$ and [Fe/H] reach a factor of 30 in surface gravity ($\Delta \log g = 1.5$ dex) and a factor of 3 in metallicity (Δ [Fe/H] $= 0.5$ dex) at [Fe/H] ~ -2.5.

Fig. 7 illustrates how the different LTE-Fe and NLTE-Opt results can change the position of each star in the $\log g$ versus $\log (T_{\text{eff}})$ plane. In addition, we have included several evolutionary tracks,

Table 3. *Hipparcos* star surface gravity comparisons.

HD err	$T_{\text{LTE-Fe}}$ ($\sim\pm60$ K)	$\log g_{\text{LTE-Fe}}$ (±0.1)	$[\text{Fe/H}]_{\text{LTE-Fe}}$ (±0.1)	T_{Opt} ($<\pm100$ K)	$\log g_{\text{NLTE-Opt}}$ (±0.1)	$[\text{Fe/H}]_{\text{NLTE-Opt}}$ (±0.1)	$\log g_{\pi}$ (±0.1)
228 79	5726	4.04	−0.92	5817	4.27	−0.89	4.33
246 16	5084	3.34	−0.62	5071	3.40	−0.69	3.29
593 74	5741	4.04	−0.96	5877	4.33	−0.88	4.49
849 37	6137	3.58	−2.34	6374	4.18	−2.11	4.15
108 317	4922	1.89	−2.58	5367	3.04	−2.14	3.14
111 721	4956	2.52	−1.37	5091	2.93	−1.29	2.70
122 956	4569	1.15	−1.75	4750	1.94	−1.61	2.03
134 169	5868	4.03	−0.77	5924	4.20	−0.74	4.03
140 283	5413	2.81	−2.79	5834	3.71	−2.41	3.73
157 466	6070	4.41	−0.34	6002	4.37	−0.41	4.35
160 693	5808	4.29	−0.47	5749	4.24	−0.55	4.31
184 499	5740	4.11	−0.58	5766	4.23	−0.57	4.08
193 901	5555	3.94	−1.18	5775	4.39	−1.01	4.57
194 598	5814	4.02	−1.23	5991	4.39	−1.10	4.27
201 891	5676	3.89	−1.21	5871	4.30	−1.06	4.30
204 155	5696	3.94	−0.71	5733	4.08	−0.69	4.03
207 978	6343	3.93	−0.62	6294	4.02	−0.62	3.96
222 794	5588	3.99	−0.66	5604	4.08	−0.66	3.91

Figure 6. Comparison of stellar parameters derived using the LTE-Fe method and the NLTE-Opt stellar parameters versus $[\text{Fe/H}]_{\text{LTE-Fe}}$. The difference in effective temperature, surface gravity and metallicity shows a large systematic increase with decreasing metallicity. The dual trends seen in ΔT_{eff}, $\Delta\log g$ and $\Delta[\text{Fe/H}]$ are a result of the R11 effective temperature calibration, in which the authors found that stars with effective temperatures less than 4500 K only required a small correction to $T_{\text{LTE-Fe}}$. Therefore, these stars stand out in the plots.
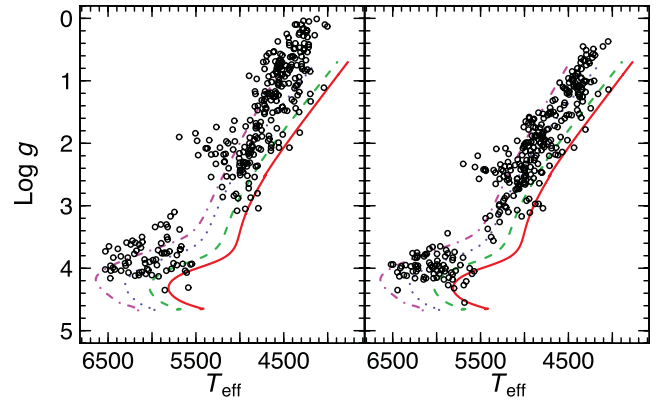
Figure 7. Stellar positions in the $\log g$ versus T_{eff} plane in LTE-Fe (left-hand panel) and NLTE-Opt (right-hand panel). The curves shown are evolutionary tracks computed using the GARSTEC code (Weiss & Schlattl 2008). Each track was computed assuming a mass of $0.8\,\text{M}_{\odot}$ and an [Fe/H] of −0.5 (solid, red), −1.0 (long-dashed, green), −1.5 (short-dashed, blue) and −2.5 (dot–dashed, magenta). Note that those stars around $\log g = 2.0$ that lie away from the tracks are most likely horizontal branch stars.

computed using the GARSTEC code (Weiss & Schlattl 2008), for comparison. Generally, the NLTE-Opt estimates of surface gravity and effective temperature trace the morphology of the theoretical tracks much more accurately. Several features are most notable. The NLTE-Opt parameters lead to far less stars that lie on or above the tip of the red giant branch (RGB), and more stars occupy the middle or lower portion of the RGB. Also, stars at the turn-off and subgiant branch are now more consistent with stellar evolution calculations.

Figs 6 and 7 further prompted us to determine the relative importance of the effective temperature scale versus the NLTE corrections for gravities and metallicities in the NLTE-Opt method. We singled out the effect of the NLTE corrections by deriving additional, LTE-Opt surface gravity and metallicity estimates using LTE iron abundances combined with our T_{Opt} estimate. Note that, as with the NLTE-Opt method, Fe lines which have an excitation potential below 2 eV were excluded. The comparison between these LTE-Opt estimates and the final NLTE-Opt estimates is shown in Fig. 8.
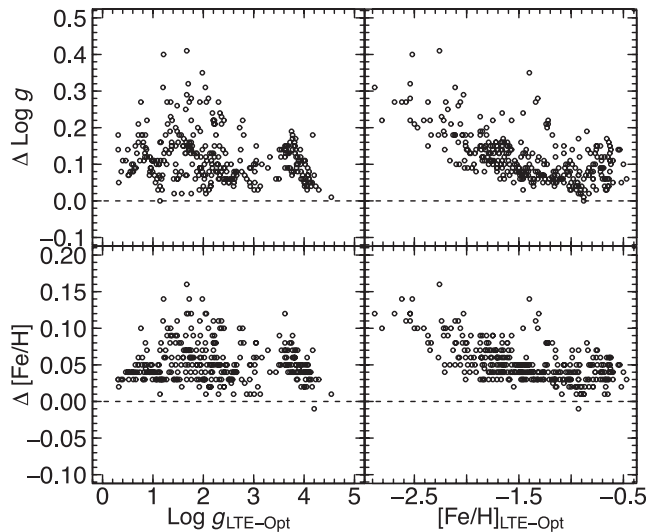
Figure 8. Comparison of LTE-Opt parameters to NLTE-Opt parameters. This shows the underlying effects of the NLTE corrections. Typically, the LTE-Opt surface gravity and metallicity estimates are systematic underestimated by about 0.1 and 0.05 dex, respectively. However, they become increasingly more underestimated to about 0.15 and 0.3 dex, respectively, at the lowest metallicities.

As evident from this figure, solving for ionization equilibrium in NLTE also leads to *systematic* changes in the $\log g$ and [Fe/H], such that LTE gravities are underestimated by 0.1 to 0.3 dex, whereas the error in metallicity is about 0.05 to 0.15 dex. These effects are consistent with that seen in Lind et al. (2012).

We thus conclude that reliable effective temperatures are necessary to avoid substantial biases in a spectroscopic determination of $\log g$ and [Fe/H], such as displayed in Fig. 6. We have shown here that, at present, the excitation balance of Fe I lines with 1D hydrostatic model atmospheres in LTE does not provide the correct effective temperature scale, supporting the results by Bergemann et al. (2012b). In contrast, Balmer lines provide such a scale. Furthermore, NLTE effects on ionization balance are necessary to eliminate the discrepancy between Fe I and Fe II lines, an effect that is present, regardless of the adopted T_{eff}. Only in this way is it possible to determine accurate surface gravity and metallicity from Fe lines.

8 CONCLUSION

In this work, we explore several available methods to determine effective temperature, surface gravity and metallicity for late-type stars. The methods include the excitation and ionization balance of Fe lines in LTE and NLTE, semi-empirically calibrated photometry (R11) and the IRFM. Applying these methods to the large set of high-resolution spectra of metal-poor FGK stars selected from the RAVE survey, we then devise a new efficient strategy which provides robust estimates of their atmospheric parameters. The principal components of our method are (i) Balmer lines to determine effective temperatures, (ii) NLTE ionization balance of Fe to determine $\log g$ and [Fe/H] and (iii) restriction of the Fe I lines to that with the lower level excitation potential greater than 2 eV to minimize the influence of 3D effects (Bergemann et al. 2012b).

A comparison of the new NLTE-Opt stellar parameters to that obtained from the widely used method of 1D LTE excitation–ionization of Fe, LTE-Fe, reveals significant *systematic biases* in the latter. The difference between the NLTE-Opt and LTE-Fe pa-

rameters systematically increases with decreasing metallicity, and can be quite large for the metal-poor stars: from 200 to 400 K in T_{eff}, 0.5 to 1.5 dex in $\log g$ and 0.1 to 0.5 dex in [Fe/H]. These systematic trends are largely influenced by the difference in the estimate of the stellar effective temperature, and thus, a reliable effective temperature scale, such as the Balmer scale, is of critical importance in any spectral parameter analysis. However, a disparity between the abundance of iron from Fe I and Fe II lines still remains. It is therefore necessary to include the NLTE effects in Fe I lines to eliminate this discrepancy.

The implications of the very large differences between the NLTE-Opt and LTE-Fe estimates of atmospheric parameters extend beyond that of just the characterization of stars by their surface parameters and abundance analyses. Spectroscopically derived parameters are often used to derive other fundamental stellar parameters such as mass, age and distance through comparison to stellar evolution models. The placement of a star along a given model will be largely influenced by the method used to determine the stellar parameters. For example, distance scales will change, which could affect the abundance gradients measured in the Milky Way (e.g. R11), as well as the controversial identification of different components in the MW halo (Schönrich, Asplund & Casagrande 2011; Beers et al. 2012). We explore this in greater detail in the next paper of this series (Serenelli et al. 2012).

REFERENCES

Ali A. W., Griem H. R., 1965, Phys. Rev., 140, 1044
Allende Prieto C., Beers T. C., Wilhelm R., Newberg H. J., Rockosi C. M., Yanny B., Lee Y. S., 2006, ApJ, 636, 804
Alonso A., Arribas S., Martinez-Roger C., 1995, A&A, 297, 197
An D. et al., 2009, ApJ, 707, L64
Árnadóttir A. S., Feltzing S., Lundström I., 2010, A&A, 521, A40
Asplund M., 2005, ARA&A, 43, 481
Bai G. S., Zhao G., Chen Y. Q., Shi J. R., Klochkova V. G., Panchuk V. E., Qiu H. M., Zhang H. W., 2004, A&A, 425, 671
Barden S. C. et al., 2008, SPIE, 7014, 149
Barklem P. S., 2007, A&A, 466, 327
Barklem P. S., Stempels H. C., Allende Prieto C., Kochukhov O. P., Piskunov N., O'Mara B. J., 2002, A&A, 385, 951
Beers T. C. et al., 2012, ApJ, 746, 34
Bergemann M., Kudritzki R.-P., Plez B., Davies B., Lind K., Gazak Z., 2012a, ApJ, 751, 156
Bergemann M., Lind K., Collet R., Magic Z., Asplund M., 2012b, MNRAS, 427, 27
Bruntt H. et al., 2012, MNRAS, 423, 122
Carollo D. et al., 2010, ApJ, 712, 692

Casagrande L., Ramírez I., Meléndez J., Bessell M., Asplund M., 2010, A&A, 512, A54 (C10)

Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, A&A, 530, A138

Castelli F., Kurucz R. L., 2003, IAUS, 210, 20P

Cayrel R., van't Veer-Menneret C., Allard N. F., Stehlé C., 2011, A&A, 531, A83

Cowley C. R., Castelli F., 2002, A&A, 387, 595

Creevey O. L. et al., 2012, A&A, 545, A17

Drimmel R., Cabrera-Lavers A., López-Corredoira M., 2003, A&A, 409, 205

François P. et al., 2003, A&A, 403, 1105

Fuhrmann K., 1998, A&A, 338, 161

Fuhrmann K., Axer M., Gehren T., 1993, A&A, 271, 451

Fulbright J. P., 2000, AJ, 120, 1841

Gehren T., Shi J. R., Zhang H. W., Zhao G., Korn A. J., 2006, A&A, 451, 1065

Gilmore G. et al., 2012, Msngr, 147, 25

González Hernández J. I., Bonifacio P., 2009, A&A, 497, 497

Grupp F., 2004a, A&A, 420, 289

Grupp F., 2004b, A&A, 426, 309

Holmberg J., Nordström B., Andersen J., 2007, A&A, 475, 519

Huber D. et al., 2012, ApJ, 760, 32

Ivans I. I., Kraft R. P., Sneden C., Smith G. H., Rich R. M., Shetrone M., 2001, AJ, 122, 1438

Johnson J. A., 2002, ApJS, 139, 219

Kervella P., Thévenin F., Di Folco E., Ségransan D., 2004, A&A, 426, 297

Lai D. K., Bolte M., Johnson J. A., Lucatello S., Heger A., Woosley S. E., 2008, ApJ, 681, 1524

Lind K., Bergemann M., Asplund M., 2012, MNRAS, 427, 50

Ludwig H.-G., Behara N. T., Steffen M., Bonifacio P., 2009, A&A, 502, L1

Mashonkina L. et al., 2008, A&A, 478, 529

Mashonkina L., Gehren T., Shi J.-R., Korn A. J., Grupp F., 2011, A&A, 528, A87

McWilliam A., Preston G. W., Sneden C., Searle L., 1995, AJ, 109, 275

Mozurkewich D. et al., 2003, AJ, 126, 2502

Prochaska J. X., Naumov S. O., Carney B. W., McWilliam A., Wolfe A. M., 2000, AJ, 120, 2513

Ruchti G. R. et al., 2011, ApJ, 737, 9 (R11)

Schönrich R., Asplund M., Casagrande L., 2011, MNRAS, 415, 3807

Serenelli A., Bergemann M., Ruchti G. R., 2012, MNRAS, submitted

Shchukina N. G., Trujillo Bueno J., Asplund M., 2005, ApJ, 618, 939

Sneden C., 1973, ApJ, 184, 839

Sousa S. G., Santos N. C., Israelian G., Mayor M., Monteiro M. J. P. F. G., 2007, A&A, 469, 783

Steinmetz M. et al., 2006, AJ, 132, 1645

Thévenin F., Idiart T. P., 1999, ApJ, 521, 753

Torres G., 2010, AJ, 140, 1158

Weiss A., Schlattl H., 2008, Ap&SS, 316, 99

Yanny B. et al., 2009, AJ, 137, 4377

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table 1. Atmospheric stellar parameter data (http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/sts319/-/DC1).

Please note: Oxford University Press are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TeX/LaTeX file prepared by the author.