



**HAL**  
open science

## Atomic diffusion and mixing in old stars

Alvin Gavel, Pieter Gruyters, Ulrike Heiter, Andreas Korn, Thomas Nordlander, Kilian Scheutwinkel, Olivier Richard

► **To cite this version:**

Alvin Gavel, Pieter Gruyters, Ulrike Heiter, Andreas Korn, Thomas Nordlander, et al.. Atomic diffusion and mixing in old stars: VII. Abundances of Mg, Ti, and Fe in M 30. *Astronomy and Astrophysics - A&A*, 2021, 652, pp.A75. 10.1051/0004-6361/202140770 . hal-03326489

**HAL Id: hal-03326489**

**<https://hal.science/hal-03326489>**

Submitted on 22 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Atomic diffusion and mixing in old stars

## VII. Abundances of Mg, Ti, and Fe in M 30<sup>★</sup>

Alvin Gavel<sup>1</sup> , Pieter Gruyters<sup>1</sup>, Ulrike Heiter<sup>1</sup>, Andreas J. Korn<sup>1</sup>, Thomas Nordlander<sup>2,3</sup>,  
Kilian H. Scheutwinkel<sup>1</sup>, and Olivier A. Richard<sup>4</sup>

<sup>1</sup> Observational Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden

e-mail: [alvin.gavel@physics.uu.se](mailto:alvin.gavel@physics.uu.se)

<sup>2</sup> Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

<sup>3</sup> ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

<sup>4</sup> Laboratoire Univers et Particules de Montpellier, Université de Montpellier, CNRS, Place Eugène Bataillon, 34095 Montpellier, France

Received 9 March 2021 / Accepted 24 May 2021

### ABSTRACT

**Aims.** We attempt to constrain the efficiency of additional transport or mixing processes that reduce the effect of atomic diffusion in stellar atmospheres.

**Methods.** We apply spectral synthesis methods to spectra observed with the GIRAFFE spectrograph on the VLT to estimate abundances of Mg, Ti, Fe, and Ba in stars in the metal-poor globular cluster M 30. To the abundances we fit trends of abundances predicted by stellar evolution models assuming different efficiencies of additional transport or mixing processes. The fitting procedure attempts to take into account the effects of parameter-dependent systematic errors in the derived abundances.

**Results.** We find that the parameter  $T_0$ , which describes the efficiency of additional transport or mixing processes, can almost certainly be constrained to the narrow range  $\log_{10}(T_0/[K])$  between 6.09 and 6.2. This corresponds to decreased abundances for stars at the main sequence turn-off point compared to the red giant branch by 0.2 dex for Mg, 0.1 dex for Fe, and 0.07 dex for Ti. We also find that while our estimates do have non-negligible systematic errors stemming from the continuum placement and the assumed microturbulence, our method can take them into account.

**Conclusions.** Our results partly amend the results of an earlier paper in this article series, that tentatively used a value of  $\log_{10}(T_0/[K]) = 6.0$  when modelling the Spite plateau of lithium. To more easily distinguish physical effects from systematic errors, we recommend that studies of this kind focus on elements for which the expected surface abundances as functions of effective temperature have a distinct structure and cover a wide range.

**Key words.** globular clusters: individual: M 30 – methods: statistical – stars: abundances – stars: atmospheres – techniques: spectroscopic

## 1. Introduction

To first approximation, one might expect the atmospheres of the stars in a stellar cluster to have identical elemental abundances: They formed from the collapse of a chemically almost homogeneous gas cloud, and the outer layers of a star are mostly unaffected by nucleosynthesis. Closer inspection has shown that this is not quite true: The assumptions that cluster stars form from chemically identical gas and that nucleosynthesis has no effect on observed abundances, are both violated in some circumstances. There are also several processes that cause elements to separate by depth in the stellar atmosphere. This paper focuses on constraining a parameter that describes one of those processes.

The assumption that the stars in a cluster are formed from chemically homogeneous gas would be true if they all formed at exactly the same time. In reality, the formation of a stellar cluster takes time, and by the time the last stars form, several of the first stars have already reached the end of their evolu-

tion and have polluted the remaining gas. This causes the phenomenon of anticorrelations in stellar clusters, where increased abundances of certain elements are statistically correlated with lowered abundances of others, such as Mg with Al and Na with O (Gratton et al. 2001). Such anticorrelations were first described in Cohen (1978), and in Carretta (2006) they were definitively shown to be correlated with the orbital parameters of the stars. This led to the hypothesis that the anticorrelations somehow reflect chemical conditions at the site of formation, which in turn are probably mostly affected by material ejected from the first generation of stars.

The assumption that nucleosynthesis does not affect surface abundances is violated in giant branch stars through the process of dredge-up, in which the convective envelope reaches deep enough into the star to start pulling up materials produced by core hydrogen burning. However, this mainly affects the isotopic ratios of C, N, and O and the abundances of Li and Be (Pagel 2009), none of which we measure in the main part of this study.

Several different processes cause elements to separate with depth in the stellar atmosphere. Some of these processes are well understood and some are still not. The conceptually simplest are

<sup>★</sup> Based on data collected at the ESO telescopes under programme 099.D-0748(A).

the diffusive processes. Over long timescales, there are two dominant processes that tend to cause different species to drift vertically in a macroscopically stationary region of a stellar atmosphere: There is the tendency for gravity to pull heavier species downwards, referred to as ‘gravitational settling’. On the other hand, there is the tendency for radiation to push more opaque species upwards, referred to as ‘radiative acceleration’. In models of old metal-poor stars, gravitational settling tends to be strongest, creating an overall tendency to lower the abundances of metals in the outer layers of the atmosphere, except for some elements that undergo significant levitation around the turn-off point (TOP). The separation of elements is counteracted by macroscopic vertical mass flows – that is, convection (Michaud 1970; Richard et al. 2001). Based on observations, it is known that warm stars have noticeably lower abundances than giant stars. Based on modelling, this is thought to be because warm stars have smaller convective envelopes, while giant stars have large convective envelopes that keep their abundances close to the original values. An overview of this topic is given in the first paper in this article series, Korn et al. (2007).

However, there appears to be at least one more mechanism that affects the separation of elements in the atmosphere. If only the diffusive processes described above, together with convection were important, we would expect to see much stronger trends in abundances as a function of effective temperature than have actually been observed in stellar clusters (Korn et al. 2007). This implies that some other transport process of chemical species is counteracting the effects of diffusion (Michaud et al. 1976; Richard et al. 2001). The most conspicuous case is lithium, for which there is a wide range of  $T_{\text{eff}}$ , called the ‘Spite plateau’, for which abundances are almost constant (Spite & Spite 1982; Richard et al. 2002). In this series of papers, we refer to this mechanism as the ‘additional transport or mixing process’ (AddMix). Most other papers use the term ‘turbulent mixing’ instead. AddMix should not be confused with the term ‘extra mixing’, which is sometimes used for the unrelated process of thermohaline mixing (Ulrich 1972). As the vague name is meant to emphasise, the physical mechanism of AddMix is currently not well understood (Richer et al. 2000). It has been suggested that it might involve mass loss and rotationally induced mixing (Michaud et al. 1983; Proffitt & Michaud 1991; Vick et al. 2013). Essentially, stellar evolution models incorporating AddMix postulate the existence of some sort of turbulent mixing process parametrised in terms of a turbulent diffusion coefficient  $D_{\text{T}}$ , defined as

$$D_{\text{T}} \equiv \omega D_{\text{He}}(T_0) \left( \frac{\rho(T)}{\rho(T_0)} \right)^{-n}, \quad (1)$$

where the proportionality constant  $\omega$  is a real number, the exponent  $n$  is some integer,  $\rho$  is density at a given temperature,  $T_0$  is a reference temperature and  $D_{\text{He}}(T_0)$  is the atomic diffusion coefficient of He at the temperature  $T_0$  (Richer et al. 2000; Richard et al. 2005). We note that while  $D_{\text{He}}$  is included in Eq. (1), based on the physical assumptions made by the modellers, we could be agnostic about what AddMix actually is and treat  $\omega D_{\text{He}}(T_0)$  as simply being a generic coefficient with dimensions of area divided by time.

Since we lack a fundamental physical model to derive them from first principles, the parameters in  $D_{\text{T}}$  have to be determined empirically by fitting model predictions to observations. This has been done by several different studies, including the article series that this paper is part of. Fitting to observed abundances of Li, Be, and B in the Sun has led to the adoption of

$n = 3$ . This simultaneously leads to adopting the proportionality  $\omega = 400$  (Proffitt & Michaud 1991; Richard et al. 2002, 2005). The parameter  $T_0$  still needs to be further constrained. By convention, we label AddMix models in terms of  $\log_{10}(T_0/[\text{K}])$ . This is usually done in a shorthand such that  $T5.9$  should be read as ‘the model for which  $\log_{10}(T_0/[\text{K}]) = 5.9$ ’. The third paper in this article series – Nordlander et al. (2012), henceforth called ‘Paper III’ – arrived at  $T6.0$  for the metal-poor cluster NGC 6397. The sixth paper in this article series – Gruyters et al. (2016), henceforth called ‘Paper VI’ – adopted this value for M 30 (NGC-7099), based on their similar metallicities, but noted that  $T5.8$  provided a better fit to their derived abundances.

The main goal of this article is to provide a more accurate estimate of  $T_0$  by comparing abundance trends predicted for different values of  $T_0$  to derived abundances for M 30. To do this we must also address a problem with interpreting derived abundances: All abundance estimates are affected by systematic errors, which depend to some extent on stellar parameters such as  $T_{\text{eff}}$ . This means that any measured trend will be a difficult-to-disentangle mixture of systematics and the actual physical trend. We have occasionally even encountered scepticism as to what extent measured abundance trends even measure anything other than the systematics of the measurement method. Hence, secondary objectives of the article are to develop a well-defined method for constraining systematic uncertainties, to then use this to definitively show that diffusion processes do have measurable effects, and to clarify how well  $T_0$  can be constrained when systematics are taken into account.

## 2. Observations

We have photometric and spectroscopic observations of 177 stars in the globular cluster M 30. The cluster has been selected on the basis that it is very metal-poor, with  $[\text{M}/\text{H}] = -2.35$ , and relatively little studied. Since it is 6 kpc below the Galactic plane (Goldsbury et al. 2010), it has the fairly low reddening of  $E(B - V) = 0.03 \pm 0.01$  mag (Kains et al. 2013, 2016). Together with the clusters NGC 6397, NGC 6752 and M 4, which have previously been studied in this article series, it forms a sequence of  $[\text{M}/\text{H}]$  from  $-2.35$  to  $-1$ , which allows us to probe AddMix as a function of overall metallicity. M 30 also has the advantage of being far above the galactic plane, minimising the effect of Galactic reddening.

### 2.1. Photometric observations

The stars had previously been observed with *VI* broadband photometry by Peter Stetson (Stetson 2000, 2005). This was used to estimate the stellar parameters needed for abundance determination, as described in Sect. 4.3. A colour–magnitude diagram can be found in Paper VI (Gruyters et al. 2016, Fig. 1).

### 2.2. Spectroscopic observations

The Fibre Large Array Multi Element Spectrograph (FLAMES) is a spectrograph connected to the second Unit Telescope (UT2) at the Very Large Telescope (VLT) facility (Pasquini et al. 2002). This feeds the GIRAFFE<sup>1</sup> spectrograph. The spectra have a resolution  $\lambda/\Delta\lambda = 23\,000$  and were taken with an exposure time around 2775 s. They were observed in the HR6 mode, which covers the wavelength range 4538–4760 Å. They were observed

<sup>1</sup> Despite customarily being spelled in uppercase, the name is not an acronym.

**Table 1.** Lines used to derived stellar abundances.

Species	Line centre [Å]	Line mask [Å]		Segment mask [Å]		log <i>gf</i>
		Start	End	Start	End	
Mg I	4702.9909	4702.00	4704.00	4680	4720	−0.44
Ti II	4563.7574	4562.75	4564.75	4540	4580	−0.69
Ti II	4571.9713	4570.97	4572.97	4550	4590	−0.31
Fe II	4583.8292	4582.70	4584.70	4560	4600	−1.86
Ba II	4554.0290	4553.00	4555.00	4535	4575	0.17

**Notes.** Only Mg4702, Ti4563 and Fe4583 were used for fitting the trend models. All wavelengths are given in Å. The line mask describes the wavelength interval used in the fitting of abundances, and the segment mask describes the entire wavelength region modelled around a line, as described in Sect. 4.4.1.

as part of project 099.D-0748(A) with principal investigator Pieter Gruyters. Several spectra were taken of each star, typically around 30 for faint stars and around 5 for bright stars. We give a full list of the stars in Table B.2.

For each star, the observed spectra were coadded into a single spectrum. As described in Appendix A, this was done before correcting for radial velocity, and also required the removal of sky flux. After this the spectra were visually inspected, and those with issues likely to prevent a meaningful abundance determination were flagged and removed from further analysis. For the abundance derivation we further co-added the spectra for individual stars into group spectra of stars with similar stellar parameters, as described in Appendix B.

### 3. Stellar evolution models

We computed stellar evolution models, including atomic diffusion (with radiative accelerations) and AddMix as described by Eq. (1), using the Montréal/Montpellier stellar evolution code (Turcotte et al. 1998; Richer et al. 2000; Richard et al. 2001). These were used to compute isochrones describing the evolutionary variation in surface chemical composition, giving us expected abundances as functions of effective temperature calculated for six discrete  $T_0$  values:  $T5.8$ ,  $T5.9$ ,  $T5.95$ ,  $T6.0$ ,  $T6.09$ , and  $T6.2$ , assuming a cluster age of 13.5 Gy. With this range, we sample the range of diffusion models used in previous papers of this series, as well as those that are compatible with the empirical constraints arising from a thin and flat Spite plateau – notwithstanding the slight rise in lithium in subgiants prior to the onset of the first dredge-up as observed in NGC 6397, (Korn et al. 2006, 2007; Lind et al. 2009). We show the isochrone in Table 1 together with the estimated stellar parameters of the individual stars, which have been estimated as described in Sect. 4.3. Previous tests have shown that the isochrones are able to reproduce the colour-magnitude diagram of M 92 (NGC 6341), which has similar metallicity to M 30 (VandenBerg et al. 2002).

In these models, the vertical stratification of chemical composition is followed for the elements that have known radiative cross-sections, and evolutionary variations are thus predicted for those elements (Richer et al. 1998). This includes Fe, Ti, and Mg, which our spectra contain lines from. Radiative cross-sections, and hence model predictions, are not available for Ba, which prevented us from fitting to our abundance estimates of that element. We discuss this element in Appendix E.

We show the predicted trends, together with our derived abundances (see Sect. 4), for Mg in Fig. 3, for Ti in Fig. 5, and for Fe in Fig. 6. For Fe and Mg, the expectation is that the abundance

should increase with decreasing effective temperature, while the expectation for Ti is that it should drop and then rise, with a sharper rise the lower  $T_0$ . This would seem to make the Ti trend the easiest to distinguish, but the trend is also much weaker than the trends expected for the other elements.

### 4. Spectroscopic derivation of abundances

To derive abundances for individual spectra, we use a pipeline based on the spectral synthesis code Spectroscopy Made Easy (SME), described in Sect. 4.4.1. The pipeline is similar to the pipeline for determining stellar parameters described in Gavel et al. (2019).

We present abundances using what is informally known as the ‘12-scale’. That is, we show

$$\log \varepsilon(x) \equiv \log_{10}(N_x/N_H) + 12, \quad (2)$$

where  $x$  is some element,  $N_x$  is the number density of the element, and  $N_H$  is the number density of hydrogen.

#### 4.1. Algorithm

The algorithm works by fitting a synthetic spectrum to the observed spectrum. This is done in a wavelength segment around the line in question that has been chosen to be wide enough to allow a reliable continuum determination. The fitting is done over a wavelength segment covering the line.

The fitting involves three steps. The first step is cleaning of contaminants: Before looking at the observed spectrum, we generate two model spectra: One that contains spectral lines for all elements, and one that contains only spectral lines for the element that we intend to fit. Pixels for which the estimated fluxes differ by more than 0.005 are flagged as contaminated, and do not get used in the fitting. The second step is normalisation: The continuum level of the observed spectrum is estimated, as described in Sect. 4.4.1. The final step is fitting: The model spectrum is fitted to the observed spectrum by minimising a  $\chi^2$ -like goodness-of-fit metric, as described in Sect. 4.4.1. Both the abundance of the element in question, and a Gaussian broadening parameter, are free parameters in the fit. The broadening parameter is mostly intended to capture the effects of macroturbulence at the depth of line formation, but can also capture other physical and instrumental effects.

#### 4.2. Line selection

The wavelength region covered by our spectra is relatively line-poor. While this has the benefit of decreasing the number of

contaminants, it gave us relatively few lines to work with. We selected one line of Fe, one line of Mg, two lines of Ti, and one line of Ba, shown in Table 1 together with line masks as described in Sect. 4.4.1 and  $\log gf$  values. The lines are taken from the *Gaia*-ESO line list described in Heiter et al. (2021), which in turn uses values from Ralchenko et al. (2010), Pickering et al. (2001), Wood et al. (2013), Kurucz (2013) and Miles & Wiese (1969). As a shorthand, we will refer to the lines as Mg4702, Ti4563, Ti4571, Fe4583, and Ba4554, respectively. In the main body of this article we only study Mg4702, Ti4563, and Fe4583, but we discuss Ti4571 and Ba4554 in Appendix E.

We note that the Fe II line at 4583.8292 Å overlaps with a weaker Fe I line at 4583.7195 Å. The blend was taken into account in our main analysis, but was not treated completely consistently in the 3D analysis described in Appendix D.

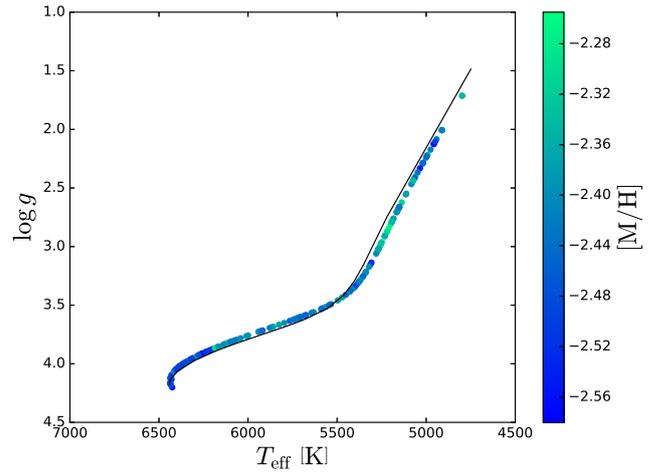
### 4.3. Stellar parameter estimates

We estimated abundances using spectral fitting. This required some estimates of the stellar parameters effective temperature  $T_{\text{eff}}$ , surface gravity  $\log g$ , and metallicity  $[M/H]$  – and in most cases also microturbulence  $v_{\text{mic}}$ . As described in Sect. 4.1, we used a single broadening parameter to capture the combined effects of macroturbulence, stellar rotation, and instrumental broadening – the last of which is expected to dominate.

$T_{\text{eff}}$  and  $\log g$  were determined by projecting the stars onto Victoria isochrones (VandenBerg et al. 2014). In Paper VI, Stetson’s  $V$  magnitude and  $V - I$  colour of the 150 stars studied in that article were used to fit a 13.4 Gyr,  $[M/H] = -2.3$  isochrone for M 30. We show our estimated  $T_{\text{eff}}$ ,  $\log g$ , and  $[M/H]$  (described below) in Fig. 1, together with the stellar isochrone used in model *T6.09*. There is a slight inconsistency between the isochrone used for the parameter estimates and that for the stellar evolution model. To the extent that this matters, it may introduce a slight tendency to underestimate the abundance difference between hot and cold stars. Paper VI also compared Victoria isochrones to BaSTI (Pietrinferni et al. 2013) and Dartmouth isochrones (Dotter et al. 2008) and found that only the Victoria isochrones were capable of reproducing the observed morphology of the cluster (Gruyters et al. 2016, Sect. 3.1). In Appendix C we investigate the effect of a possible offset in  $T_{\text{eff}}$  and the resulting shifts in  $\log g$  and  $v_{\text{mic}}$ .

The overall metallicity  $[M/H]$  for each star was determined using spectral synthesis methods similar to those used in this paper. We note that  $[M/H]$  is not crucial to our analysis: Our spectral synthesis code (described in Sect. 4.4.1) requires some assumed overall metallicity, but the result of fitting an abundance to a specific line should ideally not depend on this assumption. The only way it should be able to affect our abundance estimates is if the lines we used to measure abundances are blended with lines of other elements, which in turn are included in our line list. In that case, the accuracy of the  $[M/H]$  estimate affects how well we can compensate for the blend.

As described in Sect. 4.4.2, the microturbulence parameter  $v_{\text{mic}}$  is necessary when using 1D models of the stellar atmosphere. Normally,  $v_{\text{mic}}$  would be estimated using spectral fitting, but our spectra do not contain enough lines for this. The upcoming Paper VIII in this article series, investigating a sample similar to this, finds that it was not possible to robustly estimate  $v_{\text{mic}}$ . A preliminary discussion of this is found in Nordlander et al. (2017, Sect. 3.4.1). This forced us to use some empirically determined relation for estimating  $v_{\text{mic}}$  from the known stellar parameters. Paper VI in this article series used an empirical rela-



**Fig. 1.** Isochrone used in the *T6.09* model, together with the  $T_{\text{eff}}$ ,  $\log g$ , and  $[M/H]$  of the individual stars estimated as described in Sect. 4.3.

tion that is commonly used within *Gaia*-ESO (Gruyters et al. 2016, Sect. 3.3). The same formula is explicitly described in Gavel et al. (2019, Appendix E). However, it was calibrated on population I stars, and is likely not appropriate for our current sample. Instead, for stars with  $3.5 > \log g$  we used a relation estimated in Sitnova et al. (2015) based on dwarf stars with  $-2.6 \leq [Fe/H] \leq 0.2$ :

$$\frac{v_{\text{mic,dwarf}}}{[\text{km s}^{-1}]} = -0.21 + 0.06 \cdot [Fe/H] + 5.6 \frac{T_{\text{eff}}}{10^4 \text{ K}} - 0.43 \cdot \log g. \quad (3)$$

For stars with  $\log g < 3.5$  we used a relation estimated in Mashonkina et al. (2017) for giant stars with  $-4.0 < [Fe/H] < -1.7$ :

$$\frac{v_{\text{mic,giant}}}{[\text{km s}^{-1}]} = 0.14 - 0.08 \cdot [Fe/H] + 4.9 \frac{T_{\text{eff}}}{10^4 \text{ K}} - 0.47 \cdot \log g. \quad (4)$$

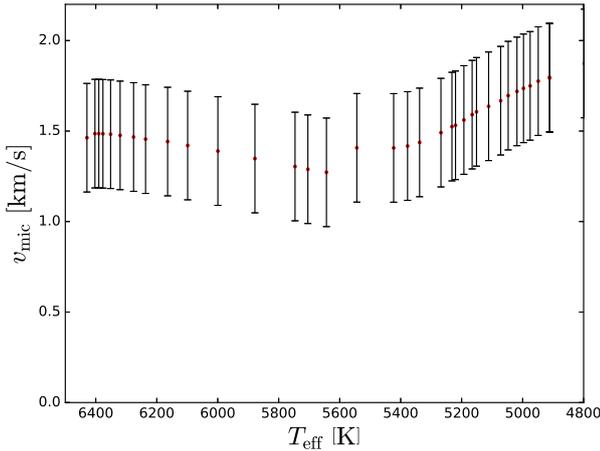
As discussed in Sect. 5.1, this is expected to be one of the two dominant sources of systematic error, which we took into account by varying the  $v_{\text{mic}}$  within  $\pm 0.3 \text{ km s}^{-1}$  around the values predicted by the formula. Figure 2 shows the  $v_{\text{mic}}$  for each spectrum used in the fitting, together with the range of variation. There is a discontinuity at the dwarf-giant boundary, but it is small compared to the variation in  $v_{\text{mic}}$ .

The spectra of individual TOP stars generally have too low S/N to be reliably analysed, especially since we have found S/N-dependent systematic errors, which we discuss in Sect. 5.1. This means that we would prefer to have spectra with approximately equal S/N. As described in Appendix B, we coadd spectra of stars with similar stellar parameters to produce group spectra with  $S/N \approx 100$ .

### 4.4. External software

The analysis pipeline uses several external pieces of software. The pipeline is built around the spectral synthesis code SME. SME in turn requires a line list and a grid of model atmospheres. We describe SME briefly in Sect. 4.4.1.

For the line list, we use the line list that has been compiled for use in the *Gaia*-ESO Survey, based on the best literature values available around 2012–2014, when the survey began (Gilmore et al. 2012; Heiter et al. 2021). We use model atmospheres calculated with the Model Atmospheres with a Radiative



**Fig. 2.** Estimated microturbulence parameter  $v_{\text{mic}}$  for the spectra used in this analysis. The error bars show the range over which each microturbulence was varied in the analysis, as described in Sect. 5.1.

and Convective Scheme (MARCS) code, under spherical symmetry and with  $v_{\text{mic}} = 2 \text{ km s}^{-1}$  for stars with  $\log g \leq 3.5$  dex, and with plane-parallel geometry and  $v_{\text{mic}} = 1 \text{ km s}^{-1}$  for stars with  $\log g > 3.5$  dex (Gustafsson et al. 2008).

In addition, the pipeline makes use of the IDL Astronomy User’s Library (AstroLib) and the Coyote IDL library (Landsman 1993; Fanning 2015). Outside of the pipeline proper, several Python modules were used: Astropy was used to handle spectral files (Astropy Collaboration 2013, 2018). SciPy was used to handle most calculations (Jones et al. 2001). emcee was used to run Goodman and Weare’s Affine Invariant Markov chain Monte Carlo Ensemble Sampler for model fits (Goodman & Weare 2010; Foreman-Mackey et al. 2013). Matplotlib was used to make all plots in this article (Hunter 2007).

#### 4.4.1. SME

SME is a code for synthesising and fitting stellar spectra. The earliest published version is described in Piskunov & Valenti (1996) and the most important changes since then are documented in Piskunov & Valenti (2017). Our pipeline uses SME version 542.

SME calculates synthetic spectra on the fly, based on a line list and a grid of model atmospheres. To determine stellar parameters and/or abundances, it fits synthetic spectra to an observed spectrum (Piskunov & Valenti 2017). The fitting method resembles a  $\chi^2$ -minimisation, but tweaked to be more robust to the line list being incomplete (Piskunov, priv. comm.). When estimating abundances, SME also provides estimates of the uncertainties on the estimates, as described in Sect. 5.3.

We determined line-by-line abundances. This means that for each line, we defined a segment mask and a line mask. The segment mask simply specifies which pixels in the observed spectrum should be synthesised at all. The line mask specifies which pixels should be used in the fitting itself. We show the segment and line masks for each line in Table 1.

Before the abundance determination, the segment is normalised using a linear rescaling that tries to fit the observed spectrum to an initial model spectrum, which is calculated using the first guess of the abundances. Current versions of SME allow doing this by a weighted least-squares minimisation over all pixels in a segment (Piskunov, priv. comm.). Older versions instead

require the pixels used in the fitting to be explicitly assigned to a continuum mask, and only use an unweighted least-squares minimisation (Piskunov & Valenti 1996, Sect. 3.4). We use the newer method in the main body of the article, but since – as discussed in Sect. 5.1 – the continuum normalisation is expected to be one of our two main sources of systematic error, we compare to the results of using the older method in Appendix F.

In SME, the abundances in stars are by default assumed to follow the Solar abundance pattern as estimated by Grevesse et al. (2007), but rescaled by  $[M/H]$ . In addition, in our pipeline, the abundances of the  $\alpha$  elements O, Ne, Mg, Si, S, Ar, Ca, and Ti are increased by 0.4, which is an empirical estimate of  $[\alpha/Fe]$  at this metallicity.

#### 4.4.2. Approximations in radiation transport

By default, SME assumes that the stellar atmosphere is 1D and that radiation transport occurs in local thermodynamic equilibrium (LTE), meaning that atomic level populations can be estimated using Saha–Boltzmann statistics. In some cases we can take into account non-LTE (NLTE) effects, and in some cases we can at least partly take 3D effects into account. As a rule, we cannot do both at once. In our analysis we used 1D LTE for Ti and Fe, and 1D NLTE for Mg and Ba. However, in Appendix D we consider indicative 3D LTE calculations for Ti and Fe.

The 1D approximation is expected to be poor at describing convection near the optical surface since the interaction between convective motion and radiative energy transport creates both temperature inhomogeneities across the surface – which cannot be described in 1D – and a steeper temperature gradient than predicted in 1D. There are also local velocity fields that give rise to a desaturation effect, which we model in 1D as microturbulence with the parameter  $v_{\text{mic}}$ . All these cause the Curves of Growth (COG) for saturated lines – such as the Ti II and Fe II lines in red giants – to be much more accurate in 3D than 1D. On the other hand, the overionisation of Mg I at these metallicities means that the LTE approximation is likely to get worse in 3D since the steeper temperature gradient increases the NLTE effects.

The LTE assumption is expected to give good results for the Ti II and Fe II lines, but not for the Mg I and Ba II lines. The stars in M 30 are very metal-poor, which means that the lack of metal-line opacities causes the local gas and radiation field to decouple. This results in overionisation, meaning that the neutral ground state is depleted relative to the predictions of LTE. The singly ionised majority species – Ti II and Fe II – are relatively immune to this since they have higher ionisation potentials.

For the NLTE computation, we used a grid of departure coefficients that estimate how much the level populations at different depths in the atmosphere are expected to differ between NLTE and LTE. The capacity for using such grids is built into SME (Piskunov & Valenti 2017, Sect. 3). We used two pre-calculated grids, one for Ba and one for Mg. For Mg, we used a grid described in Amarsi et al. (2020), using a model atom described in Osorio et al. (2015).

We performed 3D LTE spectrum synthesis calculations using SCATE (Hayek et al. 2011) on 3D hydrodynamic model atmospheres from the STAGGER grid (Magic et al. 2013). We sampled the temporal evolution by using 20 snapshots, selected at regular intervals, from each hydrodynamic simulation. For comparison, 1D LTE radiative transfer calculations were performed using the same codes on model atmospheres computed with the 1D hydrostatic model atmosphere code ATMO (Magic et al. 2013) which implements the same microphysics as STAGGER. The radiative transfer calculations were performed over a range

of abundances in steps of 0.1 dex. 3D LTE-1D LTE abundance corrections were derived by interpolating the 3D LTE equivalent widths onto the 1D LTE curve of growth.

## 5. Model specification

Our goal is to constrain  $T_0$ , assuming the stellar evolution models described in Sect. 3. We also compared our results to a null model that assumes that abundances do not vary with stellar parameters at all, to test whether we could rule out the possibility that diffusion processes have no appreciable effect on abundances.

It would seem to be simple to compare these models: We could simply fit the derived abundances to the predictions of all models, and then compare the quality of the best fits using the metric of our choice. Unfortunately, things are more complicated than they seem: Both the derived abundances and model predictions have systematic uncertainties of comparable size to the random uncertainties. This means that we cannot simply calculate some goodness-of-fit metric for each value of  $T_0$  and choose the best one. We describe the origins of the systematic uncertainties in Sect. 5.1 and our method for constraining them in Sect. 5.2. In Sect. 5.3 we state our method for estimating statistical uncertainties. In Sect. 5.4 we state the likelihood function that follows from the assumptions made. In Sect. 5.5 we describe the null hypothesis that we used to verify that our results look noticeably different from what would be expected if diffusion processes are inefficient.

### 5.1. Sources of systematic error

In stellar spectroscopy, derived parameters such as elemental abundances often have systematic uncertainties of comparable size to statistical uncertainties. In this specific case, we know that the abundances derived by our pipeline have systematic errors that depend strongly on  $T_{\text{eff}}$  and  $\log g$ . In addition, the derived abundances depend systematically on the S/N, which for observational reasons depend on  $T_{\text{eff}}$  and  $\log g$  as well. Before we could make any well-motivated statement about how well the derived abundances fit any hypothesis, we needed to compensate for or at least constrain these systematic effects. In addition, an even bigger source of systematic uncertainty is that some starting abundance for each element must be assumed in the stellar evolution model, before the evolution over time can be calculated. However, unless the assumed value is very far off from the true value, this simply creates a  $T_{\text{eff}}$ -independent constant offset in the (logarithmic) abundances.

A general overview of sources of systematic uncertainty in stellar spectroscopy can be found in Jofré et al. (2019). In this study we assumed that the dominant sources of systematic uncertainty are the continuum placement and the choice of  $v_{\text{mic}}$ . As described in Sect. 4.1, it is necessary to estimate the continuum level around a spectral line before the line can be used to estimate an elemental abundance. This estimate is never perfect, and relatively small errors can have a large impact on derived abundances. If this simply led to a random scatter in the derived abundances, it would be easy to deal with, but unfortunately it is likely to depend on the stellar parameters. A detailed study supporting this assumption can be found in Jofré et al. (2017). As described in Sect. 4.3 the microturbulence parameter  $v_{\text{mic}}$  must be estimated for those abundance estimations that use a one-dimensional model of the stellar atmosphere, and we have no way of directly measuring it. There are many methods for estimating  $v_{\text{mic}}$  based on  $T_{\text{eff}}$ , and any one of them will systemati-

cally differ from the true  $T_{\text{eff}}$  dependence, imposing a systematic error in the derived abundances.

In principle, there are many other sources of systematic error, due to both observational biases and imperfections in the modelling. However, we to some extent implicitly take them into account, as long as they are much smaller than the two sources of systematic errors that we explicitly take into account. Our method for compensating for the systematic errors only requires some upper bound on their size, while being otherwise agnostic about what the actual source of the errors is. Hence, as long as the additional sources of error are small enough to fit within the bounds set by those two sources of error, it can compensate for them. As an example of an additional source of systematic error (Gruyters et al. 2014, Sect. 3.6 and Fig. 6) – Paper V in this article series – found that this type of study is also affected by systematic errors in abundances derived from weak lines. The underlying reason is that the error estimation described in Sect. 5.3 implicitly assumes that the expected flux depends linearly on the parameter being estimated. However, while the expected flux is approximately linear in the relative number of absorbers  $N_x/N_H$ , we define abundances on a logarithmic scale. For weak lines the pixel noise is proportionally large compared to the change in expected flux given by a change in the estimated abundance, which creates a non-negligible asymmetry in the error on the abundance estimate.

### 5.2. Method for constraining systematic uncertainties

To constrain the systematic uncertainties, we need to make six assumptions. The first assumption is that for a given element, the arbitrariness in the assumed abundances of the primordial gas in the stellar evolution models will affect model abundances in all stars with a scaling factor that is constant over the cluster. The second assumption is that the systematic error not due to the initial abundance choice is due to offsets in the continuum level and arbitrariness in the assumed  $v_{\text{mic}}$ . The third assumption is that by looking at a spectrum, a human spectroscopist can give a reliable estimate of what range the correct continuum level could plausibly lie within. The fourth assumption is that for a given spectrum, the component of the error that is uncorrelated from spectrum to spectrum will be approximately the same no matter how the continuum is placed for that particular spectrum. The fifth assumption is that for a sample of stars such as ours,  $v_{\text{mic}}$  is somewhere within  $\pm 0.3 \text{ km s}^{-1}$  of those predicted by Eqs. (3) and (4), and within this interval the derived abundance depends monotonically on  $v_{\text{mic}}$ . The sixth assumption is that the systematic error does not vary very quickly over the range of stellar parameters present in our sample, so that it can be well approximated with a low-order polynomial.

Assumption one means that we can add an arbitrary constant to the log-abundances for each element. In practise, we shifted the average abundance of the stars below 5500 K to have the same mean abundance as predicted by the model. Below that threshold the models make practically identical predictions. Assumptions two through five allowed us to use the following method for constraining the systematic uncertainty for each abundance estimate: We visually inspected the fitted spectra for each abundance, and noted within which range the continuum could reasonably vary. Doing this, we found that the continuum must be somewhere within 0.5% of SME's best estimate. We re-derived abundances with the continua shifted to the top and bottom of that range. We also re-derived abundances with  $v_{\text{mic}}$  varied by  $0.3 \text{ km s}^{-1}$ , giving us nine derived abundances in total. The exception is in Appendix D, where the 3D-modelling makes

the  $v_{\text{mic}}$  parameter unnecessary. We denote the highest and lowest estimates of the abundance by  $a_i^{\text{max}}$  and  $a_i^{\text{min}}$ , respectively. Assumption six then allows us to model the systematic error  $e_i^{\text{sys}}$  as a polynomial subject to the constraint that for each spectrum  $|e_i^{\text{sys}}| \leq a_i^{\text{max}} - a_i^{\text{min}}$ . Since we do not know how quickly the systematic error may vary, we modelled it with a range of polynomials of increasing order. In Sect. 6.1 we describe how we dealt with the problem that a higher-order polynomial necessarily gives a better fit.

When fitting the stellar evolution models described in Sect. 3, we assumed that the derived stellar abundances follow normal distributions centred around the values predicted by the models, plus the systematic offsets. Section 5.3 describes how the width of the distributions was estimated.

While the dependence of derived abundances on S/N can in principle be handled in the same way as the dependences directly on the stellar parameters, we used a simple method to mostly cancel it out: Instead of deriving abundances for the spectra of individual stars, we coadded spectra with similar stellar parameters together, in such a way that the grouped spectra have approximately the same S/N. Given this, this source of systematic error should to first approximation affect all derived abundances equally, which makes it easier to take into account. We show this coaddition in Appendix B.

### 5.3. Method for estimating statistical uncertainties

We had two sources of discrepancies between our derived and predicted abundances that have to be treated as statistical errors. First of all, pixel noise in our spectra created some amount of scatter in our derived abundances. Secondly, there is some intrinsic random scatter in the true abundances in the stars, which is physically real but uninteresting for our attempts at fitting abundance trends.

Current versions of SME estimate statistical errors using a heuristic algorithm described in Piskunov & Valenti (2017). That algorithm was derived on the assumption that multiple lines are being fitted, and attempts to take into account systematic shifts in the abundance scale due to uncertain line parameters. It is probably not appropriate for estimating the errors in single-line abundances. To get meaningful estimates of the statistical errors, we restored the older error estimation method in our copy of SME. This takes the errors directly from the covariance matrix generated in the fitting, as described in Piskunov & Valenti (1996). This essentially assumed that our statistical errors come entirely from pixel noise. We also imposed a lower bound of 0.05 dex on the errors, since we do not believe any error estimate lower than that is realistic.

We assumed that the intrinsic scatter in abundances is normally distributed with a standard deviation  $\sigma_{X, \text{int}}$  that is specific to each element. We also made the simplifying assumption that within the spectra used to create our group spectra, as described in Appendix B, the variation in abundance is small enough to have an approximately linear effect on the abundances – meaning that in our spectral fitting the group spectra could be expected to behave as though they had abundances with variations  $\sigma_{X, \text{int}} / \sqrt{n_{\text{star}}}$ , where  $n_{\text{star}}$  is the number of stars included in each group spectrum.

Magnesium should have the greatest scatter, since it is affected by anticorrelations. Paper II in this series found intrinsic scatter with a magnitude somewhere below 0.1 dex in the cluster NGC 6397 (Lind et al. 2008, Sect. 4.2, Fig. 7). Paper V got similar results for the cluster NGC 6752 (Gruyters et al.

2014, Fig. 3). Based on this, we adopted 0.1 as a conservative estimate of  $\sigma_{\text{Mg, int}}$ . Fe and Ti are not affected by anticorrelations, but still have some observed scatter (Milone et al. 2017, Sect. 3.3). For Fe, an upper limit of 0.05 dex has been established (Carretta et al. 2009), so we adopted this as a conservative estimate of  $\sigma_{\text{Fe, int}}$ . Since Ti should behave similarly to Fe, we adopted this as our estimate of  $\sigma_{\text{Ti, int}}$  as well.

### 5.4. Likelihood function

We modelled the error by assuming that for each star  $i$  the derived abundance  $a_i$  is given by

$$a_i = a_i^{\text{trend}}(T_0) + a^{\text{offset}} + a_i^{\text{int}} + e_i^{\text{sys}} + e_i, \quad (5)$$

where  $a_i^{\text{trend}}(T_0)$  is the abundance predicted by the model, whether the null model or diffusion,  $a^{\text{offset}}$  is some number that does not vary from star to star,  $a_i^{\text{int}}$  is a normal distribution with mean zero and standard deviation  $\sigma_{X, \text{int}}$ ,  $e_i^{\text{sys}}$  is a polynomial in  $T_{\text{eff}}$  and  $e_i$  is a normal distribution with mean zero and standard deviation  $\sigma_i$  given by SME's uncertainty estimates as described in Sect. 5.3.

The likelihood for a particular  $a_i$  and element  $X$  is then given by

$$L(a_i) = \begin{cases} \frac{\exp\left(-\frac{(a_i^{\text{trend}}(T_0) + e_i^{\text{sys}} - a_i)^2}{2(\sigma_i^2 + \sigma_{X, \text{int}}^2)}\right)}{\sqrt{2\pi(\sigma_i^2 + \sigma_{X, \text{int}}^2)}}, & a_i + e_i^{\text{sys}} \in [a_i^{\text{min}}, a_i^{\text{max}}] \\ 0, & a_i + e_i^{\text{sys}} \notin [a_i^{\text{min}}, a_i^{\text{max}}]. \end{cases} \quad (6)$$

Since the  $a_i$  are uncorrelated (by construction, all correlations are built into trends and systematics) the likelihood for the full sequence of estimates  $\{a_i\}$  is given by

$$L(\{a_i\}) = \prod_{i=0}^n L(a_i). \quad (7)$$

For each line we calculated best fits to several models of the form Eq. (5). We used the  $T5.8$ ,  $T5.9$ ,  $T5.95$ ,  $T6.0$ ,  $T6.09$ ,  $T6.2$ , and null models to describe the true trend  $a_i^{\text{trend}}(T_0)$  and a zeroth to fifth degree polynomial in  $T_{\text{eff}}$  to describe the systematic error  $e_i^{\text{sys}}$ , making  $7 \cdot 6 = 42$  models in total. We did not go beyond that since we expect an error model with six free parameters to be able to fit practically any data.

### 5.5. Definition of null hypothesis

We took as our null hypothesis that the abundances follow a normal distribution independent of the stellar parameters. As the mean of the distribution, we took the solar abundance scaled by the metallicity and expected  $[\alpha/M]$  of the M 30 stars. As the standard deviation of the distribution we used the same  $\sigma_{X, \text{int}}$  as for the stellar evolution models.

## 6. Model comparison

Given the likelihood function Eq. (7), each combination of diffusion trend and degree of the systematic error polynomial has some maximum likelihood  $\hat{L}$ . We want to turn those likelihoods into some quantitative statements about how likely the different values of  $T_0$  are, and also how well our data could be distinguished from what we would see under the null hypothesis that diffusion is inefficient. We compare all models using

the corrected Akaike information criterion ( $AIC_c$ ), described in Sect. 6.1. We also compare the best-fit stellar evolution model to the null hypothesis by using the likelihood ratio, described in Sect. 6.2. We make a caveat about the ‘garden of forking paths’ issue with the statistical hypothesis testing in Sect. 6.3.

### 6.1. Akaike information criterion

Formally, the  $AIC_c$  for a model is defined as

$$AIC_c = 2k + \frac{2k(k+1)}{n-k-1} - 2 \ln \hat{L}, \quad (8)$$

where  $n$  is the number of data points,  $k$  is the number of free parameters in the model and  $\hat{L}$  is the maximum likelihood of the data – that is, the likelihood given the best-fitting model parameters, which in our case are  $T_0$  and the coefficients of the polynomial describing the systematic error. (Some authors choose to drop the factors of 2 from the definition). In a set of models,  $\Delta AIC_c$  for a model is the difference between the  $AIC_c$  for that model and that of the model with the lowest  $AIC_c$ . When presenting results, we will only show  $\Delta AIC_c$ .

A detailed discussion of the definition and interpretation of the  $AIC_c$  can be found in Burnham & Anderson (2003) or Burnham & Anderson (2004). However, we attempt to give an intuitive explanation here: None of the statistical models that we were testing is exactly true, but we wanted to identify which one is closest to the truth. One measure of how ‘close to the truth’ a statistical distribution is, is the Kullback–Leibler divergence (KL divergence) between that distribution and the true distribution. Since we do not know the true distribution, we cannot calculate this directly. However, by looking at the difference  $\Delta AIC_c$  between two models, we can estimate the probability that one model has a lower KL divergence from the truth than the other.

Translating  $\Delta AIC_c$  into these probabilities is not completely straightforward. As a heuristic, if two models differ by  $\Delta AIC_c < 2$ , they have comparable probabilities of being the KL-minimising model, and if they differ by  $\Delta AIC_c > 10$ , the one with lower  $AIC_c$  is virtually certain to be KL-minimising. This can be made more rigorous by introducing Akaike weights, defined as

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta AIC_c^i\right)}{\sum_r \exp\left(-\frac{1}{2}\Delta AIC_c^r\right)}, \quad (9)$$

where the summation is performed over all  $\Delta AIC_c$ . These weights can be seen as Bayesian posteriors for model  $i$  being KL-minimising.

One advantage of this method is that it neither requires nor prevents us from adopting a Bayesian framework. Contrary to common misconception, the AIC and its corrections can be derived on either frequentist or Bayesian grounds. In addition, since it looks at which statistical model is likely to be closest to the true distribution this method does not assume that either model is true. The disadvantage is that we are measuring ‘closeness’ in terms of information loss. While this would be the correct choice if we were only interested in estimating out-of-sample uncertainty – that is, if we only wanted to decide which model to use when predicting abundances in the stars of M 30 that we have not yet looked at – it is not clear that it is the best measure of which of the underlying physical hypotheses is likely to be true.

Using the likelihoods Eq. (7) for each of the 30 model fits we calculated  $AIC_c$  for each model. Since the  $AIC_c$  is additive – as

can be seen from Eq. (8) – we looked at the best fit over all lines, but we also looked at the  $AIC_c$  for the individual lines separately. We could in principle look at combinations of the error models, such as a linear error in Ti but quadratic in Fe, but we do not do so and instead assume that if the error in one line must be described with a polynomial of a certain degree, it must be for all lines.

### 6.2. Likelihood ratio

Aside from the use of the Akaike information criterion to identify the support for different models, we also looked at the likelihood ratios between the diffusion and null models. This to allow us to demonstrate that diffusion processes do have effects that can be distinguished from systematic errors, irrespective of how well  $T_0$  can be constrained. When comparing the null and stellar evolution models, we took the likelihoods for the stellar evolution models identified as the best, in the sense of having the lowest  $AIC_c$ . Since the likelihood is monotonously increasing with the number of parameters describing the systematic error, there is no unique highest-likelihood model.

The mathematical background for this section can be found in Colquhoun (2019). The likelihood ratio is formally defined as

$$L_{10} = \frac{\hat{L}_{\text{AddMix}}}{\hat{L}_{\text{null}}}. \quad (10)$$

Using this, we could calculate the False Positive Rate (FPR). This is a Bayesian quantity, stating the estimated risk of falsely accepting the alternative hypothesis, given some prior probability that it is true. This is a useful metric since it essentially corresponds to what readers often incorrectly take the  $p$ -value to mean, implying that it expresses what most readers looking at  $p$ -values actually want to know. (A detailed discussion of common misconceptions of  $p$ -values can be found in Goodman 2008). Formally, the FPR is defined as

$$FPR \equiv \frac{1}{1 + L_{10} \frac{P(H_{\text{AddMix}})}{1 - P(H_{\text{AddMix}})}}, \quad (11)$$

where  $P(H_{\text{AddMix}})$  is the prior probability that the diffusion hypothesis is true.

If we accept a prior of 0.5 for both the null hypothesis and the diffusion hypothesis, this simplifies to

$$FPR = \frac{1}{1 + L_{10}}. \quad (12)$$

Alternatively, we can accept a  $FPR_{\text{max}} = 0.05$ , and calculate the smallest prior  $P(H_{\text{AddMix}})$  that one could have that would still allow us to accept the hypothesis:

$$P(H_{\text{AddMix}}) = \frac{1 - FPR_{\text{max}}}{1 + (L_{10} - 1) FPR_{\text{max}}}. \quad (13)$$

One advantage of using the likelihood ratio is that it gives us quantitative results that have fairly straightforward interpretations. One disadvantage is that we have to select either  $P(H_{\text{AddMix}})$  or FPR more or less arbitrarily. A second disadvantage is that it implicitly assumes that one or the other of our models is true: In reality, while one or the other is true of the physical hypotheses that diffusion does or does not appreciably affect stellar spectra, the model hypotheses that the stellar abundances are sampled from statistical distributions of the form Eq. (5) are certainly all false. Whatever distribution reality is sampling from, it is guaranteed to be more complex than our models.

### 6.3. A caveat about forking paths

We should point out that neither the values of  $\Delta\text{AIC}_c$  or FPR can be taken entirely at face value. Strictly speaking, statistical measures of this kind can only be interpreted in a straightforward way when the procedure for statistical analysis was decided on before the data were seen. When the analysis was adapted in response to the data, the study has to be seen as exploratory rather than confirmatory, and statistical measures have to be treated as heuristics. This issue is discussed in [Gelman & Loken \(2014\)](#) regarding  $p$ -values specifically, but the argument is general enough to apply to any statistical measure.

In our case, several important discoveries were made partway through the statistical analysis. These include, but were not limited to, finding that there are non-negligible parameter-dependent systematic uncertainties, that spectra of individual stars tend to have too low S/N to be useful, that systematics also depend on S/N and  $v_{\text{mic}}$ , and that there are two very different but arguably reasonable methods for calculating SME uncertainties. Hence, we have to consider this an exploratory study.

## 7. Results and discussion

We estimate abundances for our sample of group spectra of stars from the TOP to the tip of the red giant branch (RGB) of M30, as described in Sect. 4. We use those abundances to fit statistical models that contain both the predictions of stellar evolution models and a model of parameter-dependent systematic errors, as described in Sect. 5. We then compare the models using the Akaike information criterion as a way of fixing the parameter  $T_0$  that describes the efficiency of AddMix processes in the stellar evolution models, as described in Sect. 6.1. We also use  $\text{AIC}_c$  and likelihood ratios to compare the models to the null hypothesis that there are no overall trends – the likelihood ratios being described in Sect. 6.2.

We discuss the results of using all lines together in Sect. 7.1. We discuss the results for Mg, Ti and Fe individually in Sects. 7.2–7.4. We make some concluding remarks in Sect. 8.

### 7.1. All lines

We show all derived abundances in Table B.1, together with other properties of each spectrum. In Table 2 we show the combined  $\Delta\text{AIC}_c$  for all lines, assuming that the true trend is either that predicted by one of our stellar evolution models, or simply flat, and that the systematic error is somewhere between constant and a 5th degree polynomial.

Just by going with the heuristic that models whose  $\Delta\text{AIC}_c$ -values differ by less than 2 are more or less indistinguishable, while models that have  $\Delta\text{AIC}_c$  larger than 10 can probably be ruled out, we can draw the conclusion that the models closest to reality are almost certainly  $T6.09$  and  $T6.2$ , and the null model can be ruled out. Our systematic errors are probably best modelled as either constant or a linear offset in  $T_{\text{eff}}$ , although the second- and third-order polynomials are still plausible. No matter how we model our systematic errors,  $T5.8$ ,  $T5.9$ ,  $T5.95$ , and  $T6.0$  are definitely ruled out.

We show the Akaike weights – as defined in Eq. (9) – for the models in Table 3. Within a Bayesian framework, the weights can be interpreted as posterior probabilities of each model being KL-minimising, but even within a Bayesian framework the caveat in Sect. 6.3 still applies. The weights say essentially the same as the rule-of-thumb interpretations above. The probability is negligible that any model other than  $T6.09$  or  $T6.2$  is KL-minimising.

**Table 2.**  $\Delta\text{AIC}_c$  values for all four lines of Fe, Ti, and Mg together.

$T_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Const.	180	64.8	35.6	16.9	2.59	0.00	58.0
Lin.	134	46.6	25.4	10.7	0.97	2.41	24.4
Quad.	136	52.0	31.7	17.3	6.60	7.42	22.3
Cubic	135	51.1	31.1	17.1	7.62	8.39	25.6
4th	142	59.6	38.8	24.4	14.8	14.5	32.8
5th	149	68.2	47.0	32.0	20.5	19.3	39.8

**Notes.** Each column assumes a particular stellar evolution model, as described in Sect. 3, and each row assumes a particular model of the error, as described in Sect. 5.2. Values above 10 are written in grey, since they represent models that can essentially be ruled out.

**Table 3.** Akaike weight  $w_i$ , as defined by Eq. (9), for all lines combined.

$T_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Const.	0.00	0.00	0.00	0.00	0.12	0.44	0.00
Lin.	0.00	0.00	0.00	0.00	0.27	0.13	0.00
Quad.	0.00	0.00	0.00	0.00	0.02	0.01	0.00
Cubic	0.00	0.00	0.00	0.00	0.01	0.01	0.00
4th	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5th	0.00	0.00	0.00	0.00	0.00	0.00	0.00

**Notes.** Weights below 0.01 are written in grey, since they represent models that, under the Bayesian interpretation of Akaike weights, have a probability below 1% of being KL-minimising.

We show the log-likelihoods of the linear-offset models in Table 4. With these, we can compare the best-fitting null model (with linear systematic error) to the best-fitting stellar evolution model ( $T6.2$  with constant systematic error) outside of the  $\text{AIC}_c$  framework. According to Eq. (12) rejecting the null hypothesis based on this would give us a FPR of  $2.56 \times 10^{-13}$ . Conversely, Eq. (13) states that we should be willing to proceed on the assumption that diffusion processes are real as long as our prior was above  $4.88 \times 10^{-12}$ .

In summary, irrespective of what framework we use, we can conclude that  $\log(T_0/[K])$  is probably somewhere around 6.09–6.2. We also find that in light of the data, we can rule out the null hypothesis that the abundance trends are purely an effect of systematics. The AddMix efficiency  $T6.09$ –6.2 derived here is somewhat higher than that found by Papers I–III in this article series when they investigated the cluster NGC 6397, which is a factor of two more metal-rich than M 30 ([Korn et al. 2007](#); [Lind et al. 2008](#); [Nordlander et al. 2012](#)).

Our findings mostly support the conclusions of Paper VI. That article attempted to investigate the Spite plateau by estimating abundances for lithium and several elements in stars in M 30. They used  $T6.0$  in their analysis, based on Paper III, but tentatively noted that  $T5.8$  actually gave a better fit to their derived calcium abundances. Based purely on our findings,  $T6.0$  was the best choice of the two, but either  $T6.09$  or  $T6.2$  could have been better.

### 7.2. Magnesium

In Table 5 we show  $\Delta\text{AIC}_c$  for the fits to the Mg4702-line. The line clearly favours the null model. However, if it has to choose between one of the other models, it prefers  $T6.2$ , without

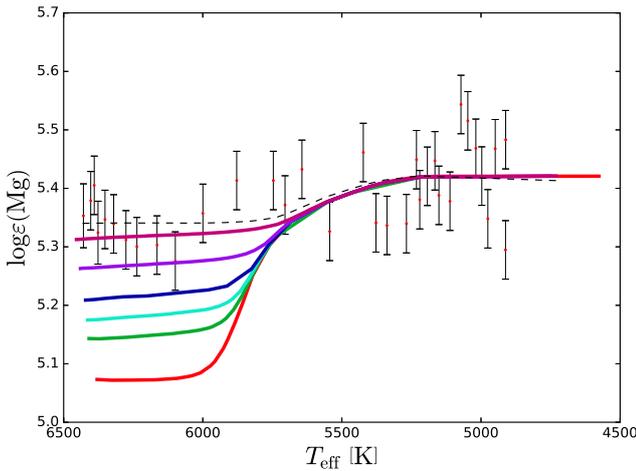
**Table 4.** Maximum (best-fit) log-likelihoods  $\ln \hat{L}$  for all lines combined.

$T_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Const.	44.6	102	117	126	133	134	105
Lin.	70.9	115	125	133	137	137	126
Quad.	73.7	116	126	133	139	138	131
Cubic	78.3	120	130	137	142	142	133
4th	79.4	121	131	138	143	143	134
5th	80.9	121	132	139	145	146	135

**Table 5.**  $\Delta\text{AIC}_c$  values for Mg4702 line, assuming 1D-NLTE.

$T_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Const.	121	56.3	37.2	22.1	6.84	0.00	6.00
Lin.	111	47.5	28.4	13.5	2.19	1.55	2.10
Quad.	110	47.9	29.8	15.8	4.66	3.96	3.60
Cubic	111	48.7	30.4	16.5	6.80	6.71	6.36
4th	113	51.7	33.4	19.3	9.29	8.16	8.97
5th	116	54.7	36.0	21.6	9.94	8.06	10.6

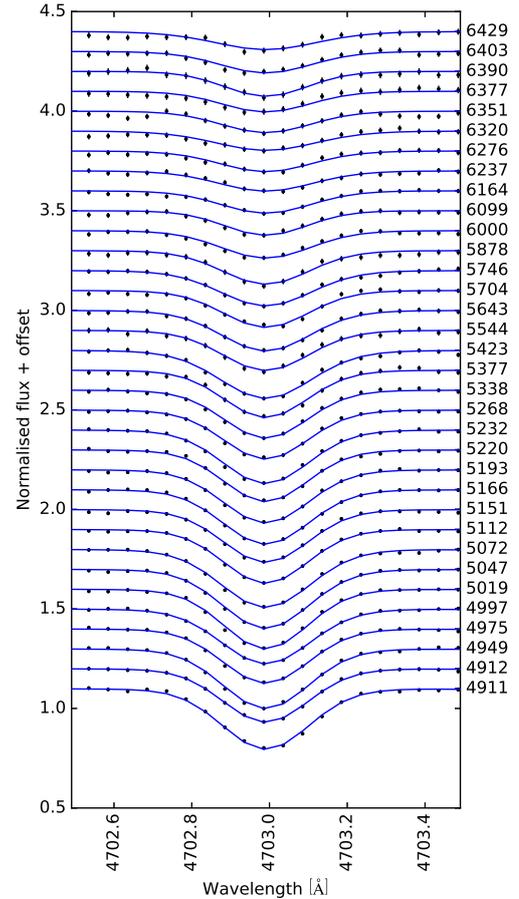
**Notes.** Values above 10 are written in grey, since they represent models that can essentially be ruled out.



**Fig. 3.** Model predictions and estimated abundances for the Mg4702 line. The red curve shows the abundance predictions from  $T_5.8$ , the green those from  $T_5.9$ , the light blue those from  $T_5.95$ , the dark blue those from  $T_6.0$ , the violet those from  $T_6.09$  and the magenta those from  $T_6.2$ . Red dots denote estimated abundances for each group spectrum. The black error bars show the random component of the error. The dashed black curve shows our estimate of the true trend with the best-estimate of the systematic error added.

showing any strong preference for any model of the systematic error. In Fig. 3 we show the abundances as functions of  $T_{\text{eff}}$ , together with the model predictions and the bounds on the systematic errors.

There is an apparent rise in the abundances near the TOP. This coincides with an increase in the width of the systematic uncertainty. We show the observed and modelled flux for the Mg line for each group spectrum in Fig. 4. This shows the likely reason why: The line becomes much shallower near the TOP, which leads to a systematic error separate from the continuum fitting or  $v_{\text{mic}}$ , as described in Sect. 5.1.



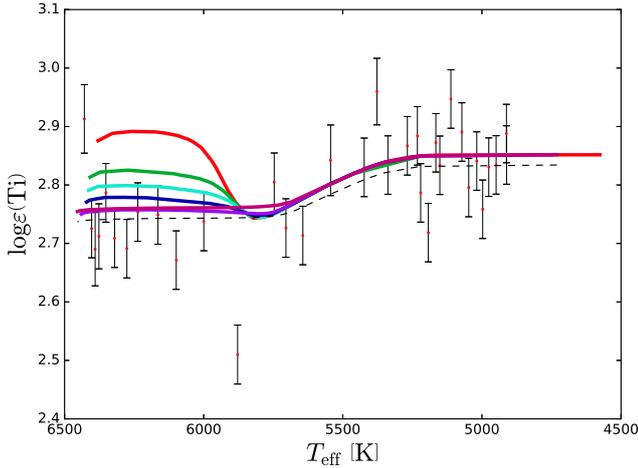
**Fig. 4.** Magnesium line for each group spectrum. Black pluses denote observed fluxes with error bars. Blue curves denote best fits given by SME assuming the  $v_{\text{mic}}$ -values given by Eqs. (3) and (4). The vertical scale is arbitrary. On the right side are shown approximate  $T_{\text{eff}}$ -values of the group spectra.

In summary, it seems that this line on its own cannot rule out the null model, but if some external constraint does, it gives most of the information that constrains  $T_0$ . This is a bit surprising, given that it is the element that is expected to have the largest intrinsic scatter due to anticorrelations. As we will see in Sect. 8, the reason is mostly that it has a more distinct predicted trend than the other lines, and that it is less sensitive to  $v_{\text{mic}}$ .

### 7.3. Titanium

In Fig. 5 we show the abundances as functions of  $T_{\text{eff}}$ , together with the model predictions and the bounds on the systematic errors. The group spectrum with  $T_{\text{eff}} = 5878$  K, which covers the  $\log g$  range 3.666–3.726 dex, looks like an outlier since it has the derived abundance range 2.42–2.62 dex while the abundances for the surrounding spectra lie above 2.7 dex. Based on visual inspection, we believe that this is due to a discontinuity in the observed fluxes at wavelengths close to the Ti4563 line in that spectrum, placing it among those most affected by the systematic error introduced by the continuum fitting (see Sect. 5.1). The same can be said about the hottest spectrum, at  $T_{\text{eff}} = 6429$  K, which covers the  $\log g$  range 4.066–4.21 dex.

In Table 6 we show the  $\Delta\text{AIC}_c$  values for the Ti4563 line. As one might expect from the figure,  $\Delta\text{AIC}_c$  is compatible with any stellar evolution model except  $T_5.8$ . However, it can definitely



**Fig. 5.** Predictions and estimated abundances for the Ti4563 line. Symbols are the same as in Fig. 3.

**Table 6.**  $\Delta\text{AIC}_c$  values for Ti4563 line.

$mT_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Const.	38.2	11.2	4.28	1.51	0.91	2.64	32.5
Lin.	20.6	2.18	0.99	1.51	2.77	4.25	20.8
Quad.	21.1	4.63	3.44	3.61	4.18	5.49	14.8
Cubic	17.9	0.78	0.00	0.42	0.95	1.69	14.5
4th	20.7	3.23	1.77	2.23	3.11	3.89	16.5
5th	20.9	5.68	4.33	4.48	5.06	5.62	19.5

**Notes.** Values above 10 are written in gray, since they represent models that can essentially be ruled out.

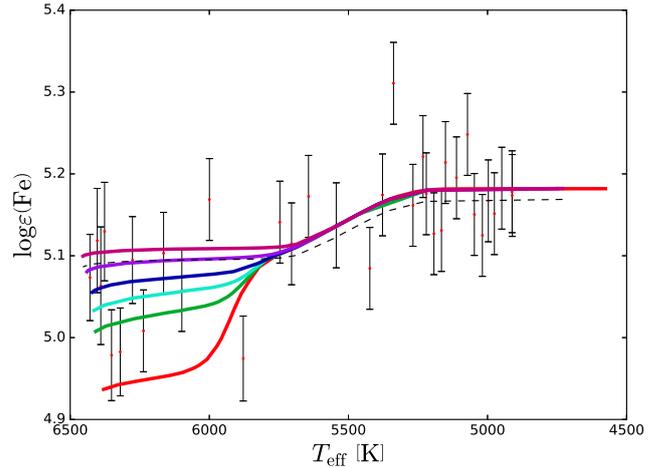
rule out the null model, which the Mg4702 on its own could not do.

#### 7.4. Iron

In Fig. 6 we show the abundances as a function of  $T_{\text{eff}}$ , together with the model predictions and the bounds on the systematic errors. In Table 7 we show  $\Delta\text{AIC}_c$  for the fits. The line turns out to be compatible with every model, including just barely  $T5.8$  and the null model. This seems to be due to a combination of large statistical errors in the derived abundances, and all models except  $T5.8$  and the null model making very similar predictions.

## 8. Summary and conclusion

We have derived stellar abundances of Mg, Ti, and Fe for a sample of stars from the metal-poor cluster M 30 using a spectral fitting method. We have used the derived abundances to fit combinations of stellar evolution models and models of our systematic errors. The stellar evolution models predict various degrees of surface depletion of elements due to atomic diffusion, leading to a temperature dependence of the abundances in the visible stellar atmosphere. The models differ in the assumed efficiency of the AddMix, which is parametrised by a reference temperature  $T_0$ . The models of our systematic errors assume that our derived abundances are offset by a low-order polynomial in the effective temperature, and that this offset comes mostly from errors in the continuum placement and assumed microturbulence. We have then compared the combined fits using the Akaike information criterion and find that the model combina-



**Fig. 6.** Predictions and estimated abundances for the Fe4583 line. Symbols are the same as in Fig. 3.

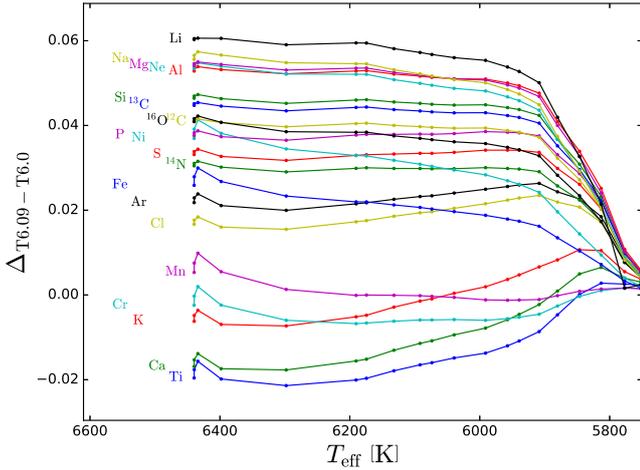
**Table 7.**  $\Delta\text{AIC}_c$  values for Fe4583 line.

$T_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Const.	26.6	3.99	0.77	0.00	1.55	4.07	26.2
Lin.	8.81	3.63	2.68	2.40	2.72	3.32	8.16
Quad.	11.4	6.20	5.20	4.65	4.46	4.68	10.7
Cubic	13.6	8.36	7.42	6.84	6.59	6.70	11.5
4th	14.8	11.3	10.3	9.56	9.13	9.16	14.0
5th	18.0	14.5	13.4	12.7	12.2	12.3	16.5

**Notes.** Values above 10 are written in grey, since they represent models that can essentially be ruled out.

tions labelled as most likely to be closest to reality are all in a small range of  $T_0$ .

The process of developing our statistical method allows us to state some general conclusions about this type of study. There are three properties that a spectral line should have to be useful in constraining  $T_0$  (or some other parameter that stellar evolution models are sensitive to). First of all, the predicted trend should be as sensitive to  $T_0$  as possible – that is,  $\partial a_i^{\text{trend}}(T_0) / \partial T_0$  should be as large as possible. Secondly, this sensitivity should change as quickly with the temperature as possible – that is,  $\partial^2 a_i^{\text{trend}}(T_0) / \partial T_0 \partial T_{\text{eff}}$  should have a large value somewhere, although it does not particularly matter where. This is necessary to allow distinguishing the measured trend from a systematic error: If the stellar evolution models predict that changing  $T_0$  simply raises the trend uniformly, it is very difficult to distinguish that from a systematic error. On the other hand, if the models predict trends that sharply diverge at some specific temperature, and it turns out that the measured abundances do change at that specific temperature by the amount predicted by some value of  $T_0$ , then it is very unlikely that this is because the systematic errors so happen to change sharply at that specific temperature. Thirdly, the measured abundances should have as little sensitivity to  $v_{\text{mic}}$  – and other difficult-to-constrain parameters – as possible. Figure 7 shows the differences  $\Delta_{T6.09-T6.0} \equiv \log \varepsilon(X_{T6.09}) - \log \varepsilon(X_{T6.0})$ , where  $X_{T_x.y}$  is the abundance of element  $X$  assuming AddMix model  $T_x.y$ . Trends are shown for each element that we have models for. In most cases, the elements that have the greatest  $\Delta_{T6.09-T6.0}$  are also those for which the derivative of  $\Delta_{T6.09-T6.0}$  with respect to  $T_{\text{eff}}$  takes the greatest value at some point, such as Li, Na, Mg, Ne, and Al.



**Fig. 7.** Sensitivity of the predicted surface abundances to the choice of  $T_0$ , given here as  $\Delta T_{6.09-T6.0}$  for each element that we have models for.

Based on this, we recommend that studies of this sort focus on light elements. In particular, the Na 8183–94 Å doublet and the Al 13123–13150 Å lines are measurable in first-generation TOP stars at this metallicity, while still not becoming strongly saturated on the RGB. These elements have the additional advantage that they have well-known atom data, and their electron structure is simple enough to make 3D NLTE calculations feasible.

We find that we can rule out the null hypothesis of no abundance variation, and that our stellar evolution model with atomic diffusion and additional mixing with an intermediate strength of  $\log(T_0/[K]) = 6.09-6.2$  is most consistent with the derived abundances. This allows us to slightly amend one important result of Paper VI: The authors found an average lithium abundance on the Spite-plateau of  $2.21 \pm 0.12$  dex, corresponding to an initial abundance of  $A(\text{Li})_{\text{init}} = 2.48 \pm 0.10$  dex, which is considerably below the value of 2.72 dex predicted by current models of Big Bang nucleosynthesis (Pitrou et al. 2018). However, they did so provisionally, based on a  $T_{6.0}$  model. For the  $T_{6.09-6.2}$  models, it would correspond to an initial abundance of 2.42–2.46 dex. Surprisingly, the gap to the BBN lithium abundance is larger than the ones found in other globular clusters studied in this series of papers.

Even though a stellar solution to the cosmological lithium problem remains possible and even likely (see Meléndez et al. 2010 and Gao et al. 2020 for independent pieces of evidence), the physics of stellar internal mixing is too poorly understood to achieve a fully satisfactory agreement with precision-cosmology BBN predictions at present. The physical effects that give rise to additional mixing below the outer convection zone in Population II stars need to be modelled in the framework of hydrodynamics.

**Acknowledgements.** AG, UH and AJK acknowledge support from the Swedish National Space Agency and the Knut and Alice Wallenberg Foundation. Parts of this research were conducted by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. PG thanks the European Science Foundation (ESF) for support in the framework of EuroGENESIS. OR acknowledges the “Programme National de Physique Stellaire” (PNPS) of the CNRS/INSU co-funded by the CEA and the CNES, France, for financial support. This work is in part based upon work from the “ChETEC” COST Action (CA16117), supported by COST (European Cooperation in Science and Technology). Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 099.D-0748(A).

## References

- Amarsi, A. M., Lind, K., Osorio, Y., et al. 2020, *A&A*, 642, A62
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123
- Burnham, K. P., & Anderson, D. R. 2003, *Model Selection and Multimodel Inference*, 2nd edn. (New York: Springer-Verlag)
- Burnham, K. P., & Anderson, D. R. 2004, *Soc. Methods Res.*, 33, 261
- Carretta, E. 2006, *AJ*, 131, 1766
- Carretta, E., Bragaglia, A., Gratton, R., D’Orazi, V., & Lucatello, S. 2009, *A&A*, 508, 695
- Cohen, J. G. 1978, *ApJ*, 223, 487
- Colquhoun, D. 2019, *Am. Statist.*, 73, 192
- Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, *ApJS*, 178, 89
- ESO 2020, ESO Archive Science Portal <http://archive.eso.org/scienceportal/home>
- Fanning, D. W. 2015, *Coyote’s Guide to IDL Programming*, <http://www.idlcoyote.com/>
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Gao, X., Lind, K., Amarsi, A. M., et al. 2020, *MNRAS*, 497, L30
- Gavel, A., Gruyters, P., Heiter, U., et al. 2019, *A&A*, 629, A74
- Gelman, A., & Loken, E. 2014, *Am. Sci.*, 102, 460
- Gilmore, G., Randich, S., Asplund, M., et al. 2012, *The Messenger*, 147, 25
- Goldsbury, R., Richer, H. B., Anderson, J., et al. 2010, *AJ*, 140, 1830
- Goodman, S. 2008, *Seminars Hematol.*, 45, 135
- Goodman, J., & Weare, J. 2010, *Comm. Appl. Math. Comput. Sci.*, 5, 65
- Gratton, R. G., Bonifacio, P., Bragaglia, A., et al. 2001, *A&A*, 369, 87
- Grevesse, N., Asplund, M., & Sauval, A. J. 2007, *Space Sci. Rev.*, 130, 105
- Gruyters, P., Nordlander, T., & Korn, A. J. 2014, *A&A*, 567, A72
- Gruyters, P., Lind, K., Richard, O., et al. 2016, *A&A*, 589, A61
- Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, *A&A*, 486, 951
- Häussler, B. 2020, *Very Large Telescope Paranal Science Operations FLAMES User Manual*, 106th edn.
- Hayek, W., Asplund, M., Collet, R., & Nordlund, Å. 2011, *A&A*, 529, A158
- Heiter, U., Lind, K., Bergemann, M., et al. 2021, *A&A*, 645, A106
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Jofré, P., Heiter, U., Worley, C. C., et al. 2017, *A&A*, 601, A38
- Jofré, P., Heiter, U., & Soubiran, C. 2019, *ARA&A*, 57, 571
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open Source Scientific tools for Python*
- Kains, N., Bramich, D. M., Arellano Ferro, A., et al. 2013, *A&A*, 555, A36
- Kains, N., Bramich, D. M., Arellano Ferro, A., et al. 2016, *A&A*, 588, C2
- Korn, A. J., Grundahl, F., Richard, O., et al. 2006, *Nature*, 442, 657
- Korn, A. J., Grundahl, F., Richard, O., et al. 2007, *ApJ*, 671, 402
- Kurucz, R. L. 2013, *Robert L. Kurucz Online Database of Observed and Predicted Atomic Transitions*
- Landsman, W. B. 1993, in *Astronomical Data Analysis Software and Systems II*, eds. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, *ASP Conf. Ser.*, 52, 246
- Lind, K., Korn, A. J., Barklem, P. S., & Grundahl, F. 2008, *A&A*, 490, 777
- Lind, K., Primas, F., Charbonnel, C., Grundahl, F., & Asplund, M. 2009, *A&A*, 503, 545
- Magic, Z., Collet, R., Asplund, M., et al. 2013, *A&A*, 557, A26
- Mashonkina, L., Jablonka, P., Pakhomov, Y., Sitnova, T., & North, P. 2017, *A&A*, 604, A129
- Meléndez, J., Casagrande, L., Ramírez, I., Asplund, M., & Schuster, W. J. 2010, *A&A*, 515, L3
- Michaud, G. 1970, *ApJ*, 160, 641
- Michaud, G., Charland, Y., Vauclair, S., & Vauclair, G. 1976, *ApJ*, 210, 447
- Michaud, G., Tarasick, D., Charland, Y., & Pelletier, C. 1983, *ApJ*, 269, 239
- Miles, B. M., & Wiese, W. L. 1969, *Atomic Data*, 1, 1
- Milone, A. P., Piotto, G., Renzini, A., et al. 2017, *MNRAS*, 464, 3636
- Nordlander, T., Korn, A. J., Richard, O., & Lind, K. 2012, *ApJ*, 753, 48
- Nordlander, T., Gruyters, P., Richard, O., & Korn, A. J. 2017, *Galactic Archeology in Metal-poor Stars*
- Osorio, Y., Barklem, P. S., Lind, K., et al. 2015, *A&A*, 579, A53
- Pagal, B. E. J. 2009, *Nucleosynthesis and Chemical Evolution of Galaxies* (Cambridge: Cambridge University Press)
- Pasquini, L., Avila, G., Blecha, A., et al. 2002, *The Messenger*, 110, 1
- Pickering, J. C., Thorne, A. P., & Perez, R. 2001, *ApJS*, 132, 403
- Pietrinferni, A., Cassisi, S., Salaris, M., & Hidalgo, S. 2013, *A&A*, 558, A46
- Piskunov, N. E., & Valenti, J. A. 1996, *A&AS*, 118, 595
- Piskunov, N. E., & Valenti, J. A. 2017, *A&A*, 597, A16
- Pitrou, C., Coc, A., Uzan, J.-P., & Vangioni, E. 2018, *Phys. Rep.*, 754, 1
- Proffitt, C. R., & Michaud, G. 1991, *ApJ*, 371, 584
- Ralchenko, Y., Kramida, A., Reader, J., & NIST ASD Team 2010, *NIST Atomic Spectra Database (ver. 4.0.0)* [Online]
- Richard, O., Michaud, G., & Richer, J. 2001, *ApJ*, 558, 377

- Richard, O., Michaud, G., Richer, J., et al. 2002, *ApJ*, 568, 979
- Richard, O., Michaud, G., & Richer, J. 2005, *ApJ*, 619, 538
- Richer, J., Michaud, G., Rogers, F., et al. 1998, *ApJ*, 492, 833
- Richer, J., Michaud, G., & Turcotte, S. 2000, *ApJ*, 529, 338
- Sitnova, T., Zhao, G., Mashonkina, L., et al. 2015, *ApJ*, 808, 148
- Spite, F., & Spite, M. 1982, *A&A*, 115, 357
- Stetson, P. B. 2000, *PASP*, 112, 925
- Stetson, P. B. 2005, *PASP*, 117, 563
- Turcotte, S., Richer, J., Michaud, G., Iglesias, C. A., & Rogers, F. J. 1998, *ApJ*, 504, 539
- Ulrich, R. K. 1972, *ApJ*, 172, 165
- VandenBerg, D. A., Richard, O., Michaud, G., & Richer, J. 2002, *ApJ*, 571, 487
- VandenBerg, D. A., Bergbusch, P. A., Ferguson, J. W., & Edvardsson, B. 2014, *ApJ*, 794, 72
- Vick, M., Michaud, G., Richer, J., & Richard, O. 2013, *A&A*, 552, A131
- Wehrhahn, A. 2020, *Continuum and Radial Velocity Determination*, <https://github.com/AWehrhahn/SME/tree/master/src/pysme>
- Wood, M. P., Lawler, J. E., Sneden, C., & Cowan, J. J. 2013, *ApJS*, 208, 27

## Appendix A: Spectra of individual stars

All spectra were downloaded from the ESO archive (ESO 2020). At this stage, the stellar spectra had been cleaned of cosmics and heliocentric correction had been applied, but the sky spectrum had not been removed (Haeussler 2020). For each observing night we also had 19–20 spectra observed with sky fibres. We averaged these together and subtracted this estimate of the sky flux from the stellar spectra observed on that night. After that, we coadded the individual exposures for each star. This was implemented in a Python module that itself makes use of the modules SciPy and Astropy (Jones et al. 2001;

Astropy Collaboration 2013, 2018). Once the coaddition is finished, the radial velocity was estimated by comparison to the star HD 140283, using an algorithm by Ansgar Wehrhahn: First a cross-correlation is used to derive a first guess. Then a  $\chi^2$ -minimisation is performed, on the assumption that the spectra are now close enough for this not to get stuck in a local minimum (Wehrhahn 2020).

A full list of the observed stars is shown in Table B.2, together with the number of spectra observed for each star, the estimated S/N for the coadded spectrum, and the stellar parameters estimated as described in Sect. 4.3.

## Appendix B: Group spectra

For reasons explained in Sect. 5, we coadded spectra of stars with similar stellar parameters to produce group spectra with roughly equal S/N. For each one of these, we took the best stellar

parameter choice to be the mean of the stellar parameters of the constituent spectra. We show the properties of the group spectra in Table B.1, together with the derived abundances. The S/N of those single-star spectra sets the minimum S/N that all other group spectra are intended to match.

**Table B.1.** Stellar parameters and line abundances for each group spectrum.

log $g$ range	$T_{\text{eff}}$ [K]	[M/H]	$n$	$S/N$	Ba4554	Fe4583	Mg4702	Ti4563	Ti4571
2.005–2.007	4911	–2.36	1	120	0.01–0.55	5.03–5.33	5.37–5.60	2.60–3.13	2.83–3.40
2.007–2.083	4912	–2.42	1	104	0.04–0.57	5.04–5.33	5.20–5.41	2.63–3.17	2.96–3.53
2.083–2.126	4949	–2.49	2	148	0.02–0.56	5.04–5.34	5.36–5.58	2.59–3.11	2.96–3.52
2.126–2.22	4975	–2.41	2	154	0.00–0.55	5.01–5.31	5.25–5.45	2.59–3.10	2.87–3.44
2.22–2.283	4997	–2.45	2	138	0.00–0.55	5.03–5.32	5.32–5.53	2.52–3.04	2.83–3.39
2.283–2.331	5019	–2.44	2	134	0.04–0.58	4.99–5.28	5.37–5.58	2.60–3.11	2.79–3.35
2.331–2.414	5047	–2.44	2	119	–0.00 to 0.55	5.02–5.30	5.42–5.62	2.56–3.07	2.93–3.50
2.414–2.467	5072	–2.34	3	129	0.20–0.68	5.11–5.41	5.44–5.65	2.65–3.16	2.94–3.50
2.467–2.661	5112	–2.37	3	124	0.18–0.67	5.06–5.35	5.29–5.47	2.71–3.22	2.91–3.45
2.661–2.676	5151	–2.43	3	106	0.10–0.63	5.08–5.37	5.31–5.48	2.61–3.09	2.96–3.50
2.676–2.768	5166	–2.40	4	117	0.09–0.61	5.01–5.28	5.36–5.54	2.64–3.14	2.94–3.49
2.768–2.842	5193	–2.35	5	110	–0.07 to 0.50	5.01–5.27	5.34–5.51	2.52–2.97	2.75–3.25
2.842–2.909	5220	–2.31	3	117	0.09–0.61	5.06–5.32	5.31–5.46	2.59–3.03	2.89–3.40
2.909–2.96	5232	–2.36	2	114	0.11–0.62	5.10–5.37	5.37–5.53	2.67–3.14	2.83–3.34
2.96–3.153	5268	–2.37	8	107	0.10–0.61	5.05–5.30	5.27–5.41	2.67–3.11	2.88–3.38
3.153–3.275	5338	–2.41	6	94	0.06–0.58	5.19–5.46	5.27–5.40	2.65–3.06	2.90–3.38
3.275–3.341	5377	–2.42	5	106	–0.09 to 0.48	5.07–5.31	5.28–5.41	2.76–3.20	2.88–3.35
3.341–3.459	5423	–2.44	9	94	–0.23 to 0.36	4.99–5.21	5.40–5.53	2.65–3.05	2.72–3.14
3.459–3.532	5544	–2.44	4	83	–0.41 to 0.17	5.03–5.27	5.27–5.39	2.68–3.06	2.74–3.17
3.532–3.589	5643	–2.42	4	83	–0.33 to 0.27	5.06–5.31	5.38–5.49	2.58–2.89	2.72–3.12
3.589–3.61	5704	–2.45	4	87	–0.32 to 0.32	5.01–5.24	5.32–5.43	2.60–2.90	2.65–3.01
3.61–3.666	5746	–2.43	6	88	–0.56 to –0.04	5.03–5.28	5.36–5.48	2.67–2.99	2.64–2.96
3.666–3.726	5878	–2.42	5	86	–0.50 to –0.01	4.89–5.08	5.35–5.48	2.42–2.62	2.72–3.07
3.726–3.802	6000	–2.40	5	86	–0.68 to –0.37	5.07–5.29	5.30–5.42	2.63–2.88	2.68–2.97
3.802–3.832	6099	–2.42	4	78	–0.59 to –0.23	4.97–5.17	5.22–5.34	2.58–2.79	2.72–3.00
3.832–3.883	6164	–2.39	8	91	–0.68 to –0.30	5.01–5.22	5.24–5.38	2.65–2.88	2.65–2.89
3.883–3.917	6237	–2.51	8	85	–0.69 to –0.39	4.93–5.11	5.23–5.38	2.67–2.86	2.73–2.99
3.917–3.95	6276	–2.48	8	84	–0.78 to –0.37	5.00–5.21	5.24–5.39	2.61–2.79	2.71–2.98
3.95–3.972	6320	–2.50	6	77	–0.82 to –0.24	4.90–5.08	5.27–5.42	2.63–2.80	2.69–2.93
3.972–4.003	6351	–2.48	8	82	–0.83 to –0.60	4.90–5.07	5.28–5.45	2.70–2.91	2.77–3.02
4.003–4.025	6377	–2.50	7	79	–0.64 to –0.21	5.04–5.24	5.26–5.40	2.63–2.81	2.72–2.94
4.025–4.038	6390	–2.49	6	72	–0.57 to –0.33	4.97–5.18	5.35–5.47	2.61–2.79	2.82–3.06
4.038–4.066	6403	–2.50	7	71	–0.70 to 0.41	5.01–5.24	5.32–5.45	2.65–2.82	2.73–2.96
4.066–4.21	6429	–2.50	8	72	–0.94 to –0.70	4.99–5.16	5.26–5.48	2.81–3.04	2.93–3.20

**Notes.** For each line the smallest and greatest abundance found when varying the continuum level and assumed  $v_{\text{mic}}$  is shown. Statistical errors are generally in the range 0.03–0.05, much smaller than the systematic errors. This includes abundances that are never actually used in the model fitting, such as for the Ba4554 line.

**Table B.2.** Individual stars, the number  $n$  of spectra coadded for each star, and the best estimates of S/N and stellar parameters.

Star no.	$n$	$S/N$	$T_{\text{eff}}$ [K]	$\log g$	[M/H]	Comment
18	33	47	5372	3.29	-2.445	
21	28	44	5562	3.51	-2.4	
36	33	36	6380	4.02	-2.435	
44	28	37	5940	3.73	-2.4	
101	33	37	6390	4.03	-2.44	
108	33	39	6263	3.92	-2.505	
182	28	48	5741	3.62	-2.48	
226	33	57	5472	3.43	-2.36	
227	5	45	5194	2.79	-2.315	
234	5	43	5192	2.79	-2.355	
248	28	43	6240	3.90	-2.525	
288	28	37	6307	3.95	-2.435	Removed from analysis
305	28	37	6375	4.01	-2.5	Removed from analysis
384	28	57	5404	3.35	-2.47	
385	33	45	6367	4.00	-2.49	
392	5	28	5357	3.26	-2.28	
429	33	52	5859	3.68	-2.365	
445	28	49	5826	3.67	-2.38	
450	28	40	6348	3.99	-2.5	Removed from analysis
455	28	41	6221	3.89	-2.515	
518	5	16	6408	4.05	-2.51	Removed from analysis
669	5	29	5353	3.25	-2.45	
679	28	39	6344	3.98	-2.505	
682	28	42	6333	3.97	-2.43	
751	5	67	5065	2.41	-2.445	
848	33	50	6139	3.84	-2.4	
864	15	35	5926	3.72	-2.44	
889	28	41	6286	3.93	-2.4	
898	28	55	5497	3.46	-2.38	
902	5	18	6165	3.85	-2.37	Removed from analysis
956	33	110	5232	2.91	-2.315	
1048	5	62	5114	2.55	-2.41	
1062	5	68	5079	2.46	-2.295	
1089	28	38	6397	4.04	-2.54	
1118	28	35	6436	4.17	-2.52	
1289	28	42	6190	3.87	-2.3	
1297	5	18	6232	3.90	-2.49	
1307	28	39	6416	4.07	-2.55	
1363	28	45	6065	3.80	-2.425	
1481	33	47	6242	3.90	-2.485	
1507	28	47	5678	3.59	-2.405	
1552	33	57	5716	3.61	-2.52	
1766	28	42	6257	3.91	-2.515	
1875	5	97	4958	2.13	-2.485	
1878	5	60	5072	2.44	-2.275	
1910	5	73	5083	2.47	-2.4	
1945	5	72	4956	2.12	-2.575	
2002	28	40	6397	4.04	-2.51	
2016	5	37	5268	3.02	-2.45	
2047	28	43	6364	4.00	-2.49	
2050	5	39	5260	3.00	-2.31	
2064	28	53	5766	3.63	-2.46	
2069	5	26	5405	3.35	-2.48	
2164	28	36	6387	4.03	-2.525	
2247	5	33	5307	3.14	-2.515	
2282	3	28	5273	3.04	-2.335	
2423	28	50	5941	3.73	-2.38	
2504	5	24	5452	3.41	-2.45	
2530	5	88	4993	2.22	-2.49	

Table B.2. continued.

Star no.	$n$	$S/N$	$T_{\text{eff}}$ [K]	$\log g$	[M/H]	Comment
2534	5	54	5156	2.68	-2.395	
2663	28	43	6131	3.83	-2.405	
2877	5	50	5158	2.69	-2.425	
2965	28	39	6351	3.99	-2.575	
3040	5	72	5060	2.40	-2.33	
3099	5	103	4911	2.01	-2.36	
3196	5	38	5248	2.96	-2.255	
3242	28	50	5708	3.60	-2.42	
3307	5	46	5192	2.79	-2.4	
3309	28	39	6380	4.02	-2.54	Removed from analysis
3310	28	41	6350	3.99	-2.39	
3485	28	53	5684	3.59	-2.435	
3528	5	136	4798	1.71	-2.345	Removed from analysis
3594	28	40	6401	4.04	-2.53	
3634	28	40	6395	4.04	-2.52	Removed from analysis
3711	5	54	5151	2.66	-2.425	
3776	28	41	6131	3.83	-2.395	
3872	28	34	6437	4.12	-2.465	
4007	28	49	6042	3.78	-2.395	
4021	28	36	6406	4.05	-2.44	
4338	28	36	6430	4.10	-2.5	
4383	5	38	5252	2.97	-2.375	
4385	5	80	5020	2.29	-2.425	
4509	28	45	5674	3.58	-2.45	
4577	28	63	5383	3.31	-2.395	
4622	28	44	6332	3.97	-2.47	
4630	5	21	6270	3.92	-2.485	Removed from analysis
4653	5	46	5210	2.84	-2.265	
4656	28	49	5919	3.72	-2.48	Removed from analysis
4689	5	27	5401	3.34	-2.425	
4779	5	25	5431	3.39	-2.5	
4801	5	42	5232	2.91	-2.405	
4901	5	31	5339	3.22	-2.48	
4903	5	52	5151	2.66	-2.525	
4907	28	48	6010	3.76	-2.38	
4985	5	80	5001	2.24	-2.405	
5098	28	97	5230	2.91	-2.36	
5171	5	104	4941	2.08	-2.41	
5191	28	51	6095	3.81	-2.48	
5205	5	91	4912	2.01	-2.415	
5231	28	43	6394	4.03	-2.4	
5331	28	37	6326	3.97	-2.525	
5350	28	39	6416	4.07	-2.41	
5363	28	48	6079	3.80	-2.42	
5364	28	32	6426	4.20	-2.55	
5379	5	43	5186	2.77	-2.375	
5390	28	44	6201	3.88	-2.52	
5410	5	83	5017	2.28	-2.46	
5457	28	42	6276	3.93	-2.475	
5458	28	44	6316	3.96	-2.55	
5473	5	17	6361	4.00	-2.475	Removed from analysis
5523	5	89	4993	2.22	-2.34	
5712	5	54	5165	2.71	-2.395	
5719	28	39	6366	4.00	-2.56	
5732	28	49	6115	3.82	-2.35	
5733	28	51	5632	3.56	-2.385	
5792	5	21	5797	3.65	-2.39	Removed from analysis
5804	5	51	5183	2.76	-2.375	
5885	5	57	5152	2.67	-2.33	

Table B.2. continued.

Star no.	$n$	$S/N$	$T_{\text{eff}}$ [K]	$\log g$	[M/H]	Comment
5950	28	47	6234	3.90	-2.475	
5976	5	21	6256	3.91	-2.565	
5980	5	28	5375	3.30	-2.4	
5987	28	41	6281	3.93	-2.45	
6006	15	26	6438	4.16	-2.5	Removed from analysis
6044	5	73	5035	2.33	-2.545	
6099	5	41	5255	2.98	-2.325	
6105	28	35	6433	4.18	-2.48	Removed from analysis
6142	5	43	5220	2.87	-2.295	
6218	28	52	5728	3.61	-2.375	
6231	5	30	5344	3.23	-2.35	
6239	28	44	6268	3.92	-2.58	
6258	5	20	6280	3.93	-2.505	
6292	5	29	5365	3.28	-2.375	
6309	28	45	5709	3.60	-2.425	
6313	28	76	5321	3.17	-2.39	
6328	28	39	6388	4.03	-2.53	
6419	5	18	6360	4.00	-2.495	Removed from analysis
6441	28	41	6382	4.02	-2.495	
6460	5	30	5281	3.06	-2.4	
6519	5	18	6171	3.86	-2.4	Removed from analysis
6614	2	17	5411	3.36	-2.39	Removed from analysis
6621	28	45	6186	3.87	-2.3	
6654	28	52	5590	3.53	-2.44	
6678	5	20	5721	3.61	-2.43	Removed from analysis
6714	5	44	5199	2.81	-2.285	
6730	5	24	5428	3.38	-2.365	
67415	28	74	5393	3.33	-2.49	
200065	33	33	6386	4.03	-2.535	Removed from analysis
200096	28	36	6329	3.97	-2.45	Removed from analysis
200249	5	17	6213	3.88	-2.485	Removed from analysis
200275	5	23	5404	3.35	-2.495	
200512	28	40	6312	3.95	-2.545	
200791	28	44	5725	3.61	-2.425	
200809	28	43	5861	3.68	-2.43	
200824	5	53	5137	2.62	-2.31	
200868	5	22	5313	3.15	-2.495	
200907	28	43	5582	3.53	-2.515	
201211	28	40	6106	3.82	-2.425	
201548	30	51	5534	3.49	-2.46	Removed from analysis
201705	28	35	6402	4.05	-2.485	
201887	28	35	6408	4.05	-2.465	
201979	28	34	6386	4.02	-2.455	
201997	33	40	6284	3.93	-2.44	Removed from analysis

## Appendix C: Impact of assumed $T_{\text{eff}}$

**Table C.1.** Change in stellar parameters and line abundances given by shifting the temperature upwards by 100 K, for three representative group spectra.

Group spec.	RGB	SGB	TOP
$\Delta T_{\text{eff}}$ [K]	+100	+100	+100
$\Delta \log g$ [dex]	+0.28	+0.06	+0.17
$\Delta v_{\text{mic}}$ [km s <sup>-1</sup> ]	-0.09	+0.03	-0.02
$\Delta \log \varepsilon(\text{Fe}4583)$ [dex]	+0.12	+0.02	+0.08
$\Delta \log \varepsilon(\text{Mg}4702)$ [dex]	+0.05	+0.07	+0.04
$\Delta \log \varepsilon(\text{Ti}4563)$ [dex]	+0.19	+0.04	+0.09

**Table C.2.** Derived abundance for the TOP spectrum covering the  $\log g$  range 4.038–4.066 dex, and the resulting  $T_{\text{eff}}$  in the RGB spectrum covering the  $\log g$  range 2.007–2.083 dex when requiring that line to have the same abundance.

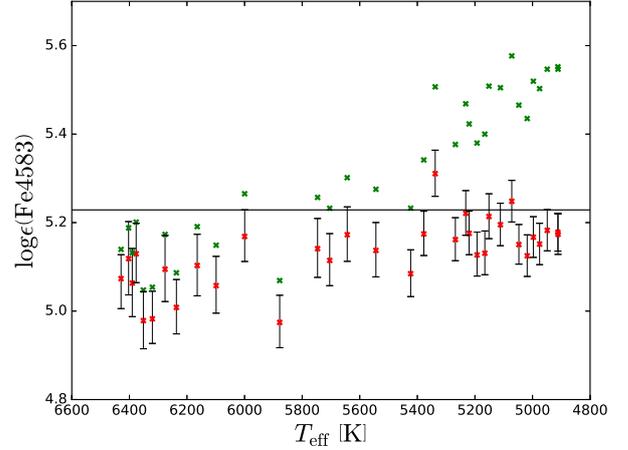
Line	Fe4583	Mg4702	Ti4563
TOP $\log \varepsilon(x)$ [dex]	5.13	5.31	2.86
Corresponding RGB $T_{\text{eff}}$ [K]	4643	5668	4447

In the main body of this article, we assumed that the continuum normalisation and assumed  $v_{\text{mic}}$  dominate over all other sources of systematic error. As a test of this assumption, we examined the impact of the assumed  $T_{\text{eff}}$  on the derived abundances. For three representative group spectra – at the bottom, middle, and top of the temperature range – we observed the effect of increasing the temperature by 100 K. We dub the group spectra ‘TOP’, ‘SGB’, and ‘RGB’, respectively. As the TOP spectrum, we selected the spectrum covering the  $\log g$  range 3.95–3.972, which at 6320 K is still cold enough that it can be shifted upwards by a 100 K without falling off the isochrone entirely.

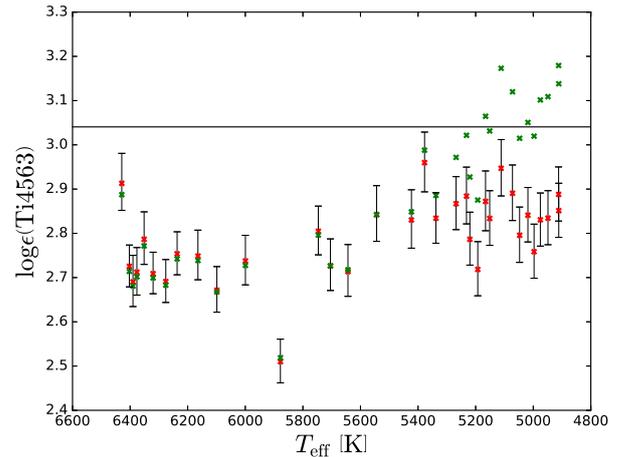
In Table C.1 we show the shift in stellar parameters and the resulting abundances for the three spectra when applying a shift of 100 K to the effective temperature. Compared to the full range covered by shifting the continuum level and assumed  $v_{\text{mic}}$ , as shown in Table B.1, we see that this source of error is indeed small by comparison.

As an additional check, we tested if there was any plausible  $T_{\text{eff}}$  that would cause the RGB spectrum to assume the same abundance as the TOP spectrum. We used the same RGB spectrum as before, but for the TOP spectrum we took that covering the  $\log g$  range 4.038–4.066 dex. We derived the line abundances for the TOP spectrum and then fitted  $T_{\text{eff}}$  for the RGB spectrum while keeping the abundance at this value. Throughout we kept the broadening parameter fixed at a value corresponding to  $v_{\text{mic}} = 8$  km s<sup>-1</sup> and resolution  $\lambda/\Delta\lambda = 23\,000$ . We found that keeping the broadening parameter free caused it to be driven up to implausibly high values – corresponding to a  $v_{\text{mic}}$  in excess of 30 km s<sup>-1</sup> – when doing the  $T_{\text{eff}}$  fitting. We show the TOP abundance and resulting RGB  $T_{\text{eff}}$  in Table C.2. Since the necessary shifts are both very large and inconsistent, we conclude that the difference in abundance cannot be purely an artefact of an offset in the  $T_{\text{eff}}$  scale.

## Appendix D: Comparison to 3D LTE



**Fig. D.1.** Derived abundances from the Fe4583 line. Red crosses show 1D LTE results and green crosses show 3D LTE results. Error bars denote the possible systematic error due to uncertainty in the continuum fitting. The horizontal line denotes the initial abundance assumed in the stellar evolution model.



**Fig. D.2.** Derived abundances from the Ti4563 line. Red crosses show 1D LTE results and green crosses show 3D LTE results. Error bars denote the possible systematic error due to uncertainty in the continuum fitting. The horizontal line denotes the initial abundance assumed in the stellar evolution model.

To test the robustness of our results, we repeated our analysis using the grids of 3D corrections for Fe and Ti described in Sect. 4.4.2. We find that the grids predict slightly lowered abundances at the warm end and strongly raised at the cold end. We show the change for the Fe4583 line in Fig. D.1 and for the Ti4563 line in Fig. D.2.

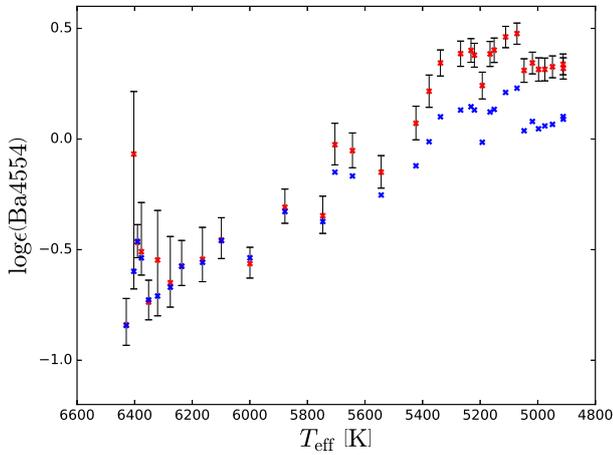
As can be seen in the plots, the abundance trends do not appear to level out at low temperatures, which all stellar evolution models predict that they should. Since the stellar evolution models make identical predictions at the cold end, these clearly unphysical abundances do not actually affect the results of the analysis, but they make us reluctant to trust the 3D-corrected abundances for the warmer stars. We show the resulting  $\Delta\text{AIC}_c$  values in Table D.1. The results are mostly consistent with the physical part of Table 3, in that it clearly prefers T6.09 over all other models. However, it needs more parameters to model the systematic error accurately.

**Table D.1.** Akaike weight  $w_i$ , as defined by Eq. (9), for all lines combined, after applying 3D corrections where applicable.

$T_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Constant	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Linear	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Quadratic	0.00	0.00	0.00	0.00	0.02	0.00	0.00
Cubic	0.00	0.00	0.00	0.04	0.83	0.00	0.00
4th	0.00	0.00	0.00	0.00	0.10	0.00	0.00
5th	0.00	0.00	0.00	0.00	0.00	0.00	0.00

**Notes.** Weights below 0.01 are written in grey, since they represent models that, under the Bayesian interpretation of Akaike weights, have a probability below 1% of being KL-minimising.

## Appendix E: Lines left out of the analysis

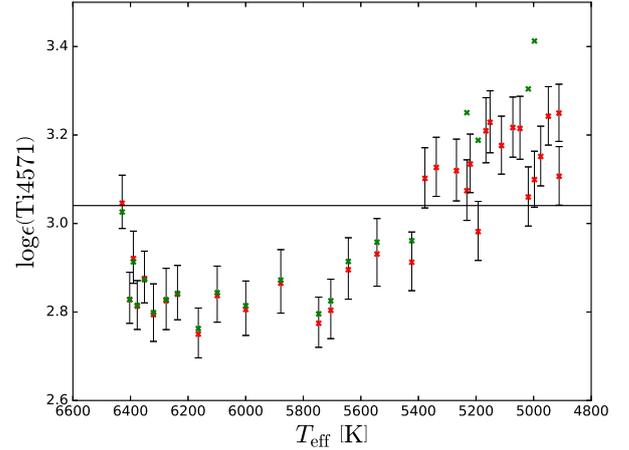


**Fig. E.1.** Derived abundances for barium, based on the Ba4554 line. Red crosses show 1D LTE and blue crosses show 1D NLTE. Error bars denote the possible systematic error due to uncertainty in the continuum fitting.

Aside from the three lines discussed in the main body of the article, we fitted the lines Ba4554 and Ti4571, described in Table 1. The Ba4554 line was left out from the start for the simple reason that we do not have a stellar evolution model for this element, so it cannot be used to estimate  $T_0$ . We only analysed it for completeness, since it was a clearly resolved line in the wavelength range we observed. We originally intended to include the Ti4571 line, but discovered that for the giant stars, it becomes heavily saturated and loses almost all its sensitivity to the abundance.

In Fig. E.1, we show the derived abundances for Ba4554 assuming 1D LTE and 1D NLTE, with systematic error bars given by the highest and lowest possible choice of continuum. Based on this, we can see that there is a clear trend and that it levels out at low temperatures, as one would expect. The overall magnitude of the observed trend is smaller in NLTE, however.

In Fig. E.2, we show the derived abundances for Ti4571 assuming 1D LTE and 3D LTE, with systematic error bars given by the highest and lowest possible choice of continuum. At low temperatures the trend does not flatten out, but rises sharply in



**Fig. E.2.** Derived abundances for titanium, based on the Ti4571 line. Red crosses denote 1D LTE and green crosses denote 3D LTE. Green crosses are missing for those stars for which the parameters and derived 1D LTE abundance fell outside of the convex hull of the grid. Error bars denote the possible systematic error due to uncertainty in the continuum fitting. The horizontal line denotes the assumed initial abundance in the stellar evolution model.

an unphysical way. In some cases the stars fall outside the convex hull of the grid of 3D corrections, and have to be left out completely.

## Appendix F: Impact of continuum fitting algorithm

**Table F.1.** Akaike weights  $w_i$  for all lines combined, using the continuum fitting algorithm described in Gavel et al. (2019).

$T_0$ :	5.8	5.9	5.95	6.0	6.09	6.2	Null
Const.	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lin.	0.00	0.00	0.00	0.01	0.54	0.19	0.00
Quad.	0.00	0.00	0.00	0.00	0.13	0.06	0.00
Cubic	0.00	0.00	0.00	0.00	0.05	0.02	0.00
4th	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5th	0.00	0.00	0.00	0.00	0.00	0.00	0.00

**Notes.** Weights below 0.01 are written in grey, since they represent models that, under the Bayesian interpretation of Akaike weights, have a probability below 1% of being KL-minimising.

Current versions of SME offer two ways of estimating the continuum level: One that minimises the  $\chi^2$ -distance between a preliminary model spectrum and pixels covered by a continuum mask and one that compares relative line depths. The latter algorithm is a recent addition to SME, and is that used in the main body of the article. To get an estimate of the sensitivity of our analysis to the choice of algorithm, we compared our results to results using the old algorithm. We defined the continuum mask using the algorithm described in Gavel et al. (2019). In Table F.1 we show the Akaike weights using this algorithm. Comparing to Table 3, the preference is still for  $\log_{10}(T_0/[\text{K}])$  in the range 6.09–6.2. The main difference is that  $T_{6.0}$  is now ranked as just barely possible, and the systematic error can no longer be modelled as a robust offset. Based on this, we conclude that our analysis is robust to changes in the continuum fitting algorithm.