Gaia FGK benchmark stars: opening the black box of stellar element abundance determination

P. Jofré[1,2], U. Heiter[3], C. C. Worley[1], S. Blanco-Cuaresma[4], C. Soubiran[5], T. Masseron[1], K. Hawkins[1,6],
V. Adibekyan[7], S. Buder[8], L. Casamiquela[9], G. Gilmore[1], A. Hourihane[1], and H. Tabernero[10]

1. Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
   e-mail: pjofre@ast.cam.ac.uk
2. Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile
3. Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden
4. Observatoire de Genève, Université de Genève, 1290 Versoix, Switzerland
5. Laboratoire d'Astrophysique de Bordeaux, Univ. Bordeaux, CNRS, UMR 5804, 33615 Pessac, France
6. Department of Astronomy, Columbia University, 550 W 120th St, New York, NY 10027, USA
7. Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
8. Max-Planck Institute for Astronomy, 69117 Heidelberg, Germany
9. Departament d'Astronomia i Meteorologia, Universitat de Barcelona, ICC/IEEC, 08007 Barcelona, Spain
10. Dpto. Astrofísica, Facultad de CC. Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

ABSTRACT

Gaia and its complementary spectroscopic surveys combined will yield the most comprehensive database of kinematic and chemical information of stars in the Milky Way. The *Gaia* FGK benchmark stars play a central role in this matter as they are calibration pillars for the atmospheric parameters and chemical abundances for various surveys. The spectroscopic analyses of the benchmark stars are done by combining different methods, and the results will be affected by the systematic uncertainties inherent in each method. In this paper, we explore some of these systematic uncertainties. We determined line abundances of Ca, Cr, Mn and Co for four benchmark stars using six different methods. We changed the default input parameters of the different codes in a systematic way and found, in some cases, significant differences between the results. Since there is no consensus on the correct values for many of these default parameters, we urge the community to raise discussions towards standard input parameters that could alleviate the difference in abundances obtained by different methods. In this work, we provide quantitative estimates of uncertainties in elemental abundances due to the effect of differing technical assumptions in spectrum modelling.

Key words. stars: general – line: profiles – methods: data analysis – Galaxy: abundances

1. Introduction

In the times of the first data release of *Gaia*, Galactic astronomy is experiencing a revolution due to the millions of spectra of individual Milky Way stars becoming available to the community thanks to the several on-going and future high-resolution spectroscopic surveys. Some prominent ones are APOGEE (Ahn et al. 2014), *Gaia*-ESO (Gilmore et al. 2012, hereafter GES), GALAH (De Silva et al. 2015), 4MOST (de Jong et al. 2012) and WEAVE (Dalton et al. 2012). Because these datasets have different selection functions attempting to survey stars of differing nature, the spectra are different from each other and therefore different abundance analysis pipelines are developed independently, each affected by their own systematic uncertainties. In order to extract the best information out of these rich datasets, it became important to combine the complementary data from these surveys and therefore the attempts of cross-calibrating surveys became a principal activity in the field (Feltzing 2015). Cross-calibration can best be done when considering stars observed by multiple surveys (Bland-Hawthorn & Gerhard 2016; Ho et al. 2017).

The *Gaia* FGK benchmark stars (GBS; Heiter et al. 2015a, hereafter Paper I) have become a crucial resource in the cross-calibration of stellar parameterisation methods and stellar population surveys. Provided as a spectral library

(Blanco-Cuaresma et al. 2014b, hereafter Paper II) with which different instruments and resolutions can be simulated, key studies have extended the GBS known characteristics from stellar parameters ($T_{\rm eff}$ and $\log g$, Paper I) to metallicity (Jofré et al. 2014, hereafter Paper III), α and Fe–peak element abundances (Jofré et al. 2015b, hereafter Paper IV) and extensions of the sample towards low metallicities (Hawkins et al. 2016a, Paper V). See also Jofré (2016) for a summary.

The GBS spectra are of exquisite quality, with excellent signal-to-noise ratio (S/N) and very high resolution. Thus, their analysis has provided the community with key reference parameters and element abundances. The GBS are being considered as calibrators by GES (Randich et al. 2013), GALAH (De Silva et al. 2015) and RAVE (Kunder et al. 2017). APOGEE has also observed some of them in order to evaluate results. In addition, some independent works on spectral analyses are using the GBS as a tool to assess the quality of results. Examples are the AMBRE project (e.g. Worley et al. 2016) and the introduction of new methods (Schönrich & Bergemann 2014; Boeche & Grebel 2016; Hawkins et al. 2016b; Teixeira et al. 2016).

In order to avoid systematic uncertainties due to a given method, the results of the spectral properties of the GBS come from combining several different methods. This follows the spirit

Table 1. Atmospheric parameters of the four GBS from Papers I and III.

Star	T_{eff} K	$\log g$ dex	[Fe/H] dex	v_{mic} km s^{-1}	$v \sin i$ km s^{-1}	v_{mac} km s^{-1}
Sun	5771 ± 01	4.44 ± 0.00	0.03 ± 0.05	1.06 ± 0.18	1.6	4.2
Arcturus	4286 ± 35	1.64 ± 0.09	–0.52 ± 0.08	1.58 ± 0.12	3.8	5.0
61 Cyg A	4374 ± 22	4.63 ± 0.04	–0.33 ± 0.38	1.07 ± 0.04	0.0	4.2
HD 22879	5868 ± 89	4.27 ± 0.03	–0.86 ± 0.05	1.05 ± 0.19	4.4	5.4

Notes. v_{mac} was used only as an input parameter by `SME` and `iSpec`.

of the GES parametrisation strategy (Smiljanic et al. 2014). Although the expectation is that this strategy yields more precise results, the combination of different methods is not trivial as discrepancies between results can be large and usually very difficult to improve. While discussions of this problem are found thanks to GES efforts in Smiljanic et al. (2014) and in Papers III and IV, as well as in Hinkel et al. (2014) and Hawkins et al. (2016b), other dedicated works in this direction have also been reported. The latter come from massive efforts concluding from workshops, where experts met and analysed common stars with different methods (Lebzelter et al. 2012; Hinkel et al. 2016).

In particular, in Paper IV we found that differences in spectral analyses methods can lead to differences in abundances of more than 0.3 dex, even when stellar parameters and atomic data were kept fixed. To overcome this problem, in Paper IV we combined the abundances in a differential approach. Differential analyses with respect to the Sun have shown to significantly improve the accuracy of results (e.g. Bensby et al. 2014; Nissen 2015; Jofré et al. 2015c; Tucci Maia et al. 2016). Since the GBS are very different from one another, in Paper IV they were separated into groups and the abundances were derived differentially, taking one reference star within each group. The reference stars were then carefully analysed differentially with respect to the Sun.

The differential approach, however, does not solve the issue of having occasional disagreements between methods that cannot be justified given the same material and input parameters used for the analysis, and does not provide clues towards the cause of these discrepancies. Thus, with the team determining abundances of GBS, we proposed to investigate what might cause these differences. The aim is essentially to "open the black box" of each pipeline and investigate which of the many hidden assumptions in spectral analysis are responsible for occasional large abundance differences. Since for many of the technical assumptions in such analyses there is no consensus on the correct approach, this study also aims to provide guidelines for a realistic error estimation.

In February, 2016, we organised a 4-day workshop at the Institute of Astronomy in Cambridge, UK, which was attended by scientists representing six out of the eight methods that have been used throughout the GBS work series. Among these methods, three were also part of the workshop summarised in Hinkel et al. (2016). Here, we report our findings. We describe first the data we used in Sect. 2, in Sect. 3 we describe our methods and in Sect. 4 we discuss each test and present the results. We finalise with a summary and concluding remarks in Sect. 5.

2. Input data

2.1. Stellar sample

Four GBS were chosen for this investigation. They are representative of the groups *FG*-dwarfs, *K*-dwarfs, *FGK*-giants and

metal-poor: Sun, 61 Cyg A, Arcturus and HD 22879 and are used as reference stars in Paper IV for the of the abundance determination. Table 1 summarises their fundamental T_{eff} and $\log g$ from Paper I and recommended (NLTE) [Fe/H] determined in Paper III as well as the adopted v_{mic} and $v \sin i$. Their high S/N high-resolution spectra were taken from the GBS library described in Paper II. These spectra were previously used for the study of GBS abundances as determined by several methods with common inputs (Papers III and IV). In particular, the spectra used here were selected from Paper IV and come from NARVAL for 61 Cyg A, the spectrum of Hinkle & Wallace (2005) for the Sun and Arcturus, and HARPS for HD 22879. All were normalised, convolved to a common resolution of 70 000 and corrected by radial velocity (RV) as described in Paper II (see also Paper IV for details).

2.2. Lines

Four lines, Ca I, Cr I, Co I and Mn I, were chosen to make the tests. Out of the many lines used in Paper IV for determining the abundances of these elements, we searched for those that were flagged with "Y" in Table C36 in the *FG*-dwarfs, *K*-dwarfs, *FGK*-giants and metal-poor groups. The "Y" flag signifies that at least 50% of the stars in a given group used that line for the final abundance of that element. The atomic data were taken from the fifth version of the line list created for the *Gaia*-ESO survey (Heiter et al. 2015b). All have $\log gf$ values that have been evaluated to be of good quality (for details see Heiter et al. 2015b). Their atomic data are listed in Table 2. The two elements, Co I and Mn I, both have hyperfine structure (HFS). Those components are also listed.

Their profiles are plotted for each star in Fig. 1. One can see that the lines are mostly clean and of weak to moderate strength. Note that the Cr line in the metal-poor star HD 22789 is very weak and could not be measured by all methods. Unfortunately, due to the intrinsic difference between the four stars, this was the only Cr line available that could be analysed by at least one method in all four stars. The abundances derived in Paper IV for each element, line and star studied in the present work are indicated in Table 3. In addition, we consider a reference value for comparison, indicated in Table 4. This value corresponds to the average of the abundances obtained by all methods and is taken from the third column of Tables C2, C8, C20 and C33 of Paper IV for 61 Cyg A, Arcturus, HD 22879 and the Sun, respectively. It is important to state here that during this work this value is used solely as a zero point. It is beyond the scope of this work to reproduce that result or provide another updated reference value.

2.3. Atmospheric models

The atmospheric models are those employed by the analysis of the spectra in GES, as well as in Paper III and Paper IV. These are
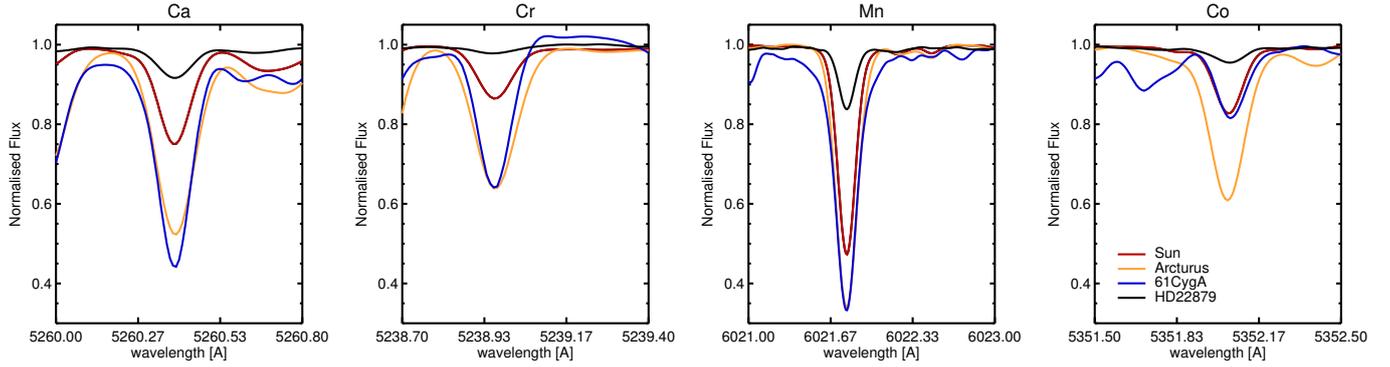
Fig. 1. Observed profiles of the four different lines that were selected and analysed, for each of the four GBS.

Table 2. Atomic line data of the four chosen lines and the additional Ca line used for the blending test (Sect. 4.7) published in Paper IV as well as the HFS for Co I and Mn I.

Element	Ion	λ	Ex.Pot	$\log gf$
Ca	1	5260.3870	2.521	−1.719
Ca	1	6455.598	2.523	−1.290
Cr	1	5238.9610	2.709	−1.270
Mn	1	6021.7933	3.075	−0.054
Co	1	5352.0397	3.576	0.060
HFS				
Mn	1	6021.7210	3.075	−2.756
Mn	1	6021.7470	3.075	−1.539
Mn	1	6021.7490	3.075	−2.404
Mn	1	6021.7690	3.075	−1.363
Mn	1	6021.7720	3.075	−2.279
Mn	1	6021.7780	3.075	−0.621
Mn	1	6021.7880	3.075	−1.337
Mn	1	6021.7910	3.075	−2.358
Mn	1	6021.7950	3.075	−0.761
Mn	1	6021.8020	3.075	−1.404
Mn	1	6021.8080	3.075	−0.919
Mn	1	6021.8110	3.075	−1.580
Mn	1	6021.8170	3.075	−1.103
Mn	1	6021.8210	3.075	−1.325
Mn	1	6021.8210	3.075	−1.610
Co	1	5351.8931	3.576	−3.244
Co	1	5351.9244	3.576	−2.855
Co	1	5351.9526	3.576	−1.736
Co	1	5351.9770	3.576	−1.532
Co	1	5351.9985	3.576	−1.459
Co	1	5352.0191	3.576	−0.581
Co	1	5352.0321	3.576	−0.662
Co	1	5352.0509	3.576	−0.503
Co	1	5352.0696	3.576	−0.569

Notes. For Mn, we used the modified line list (see Sect. 4.1). Ca: all data from Smith & Raggett (1981); Cr: wavelength and energies from Wallace & Hinkle (2009), $\log gf$ from Sobeck et al. (2007); Mn: wavelength, energies, $\log gf$ from Den Hartog et al. (2011), HFS constants for lower level from Handrich et al. (1969), for upper level from Brodzinski et al. (1987); Co: wavelength and energies from Kurucz (2008), loggf from Cardon et al. (1982), HFS constants from Pickering (1996).

the MARCS models (Gustafsson et al. 2008), which are computed under the one-dimensional (1D), local thermodynamical equilibrium (LTE) assumption and assume the standard composition for α-enhancement with respect to iron abundance.

3. Methods

We considered six methods in total, all previously used in Papers III and IV. Three methods are based on synthesis and three on EWs. These are summarised in Table 5 and briefly described below. For more details, we refer the reader to Papers III and IV, as well as Smiljanic et al. (2014) for the GES analysis of the UVES spectra.

3.1. Synthesis methods

3.1.1. Université Libre Bruxelles (ULB)

The results of this method have been provided by T. Masseron and are based on BACCHUS (Masseron et al. 2016), which was developed in Brussels. In this work, the method was based on finding the synthetic spectrum from Turbospectrum (Plez 2012) that best reproduces the observed spectrum with χ^2 minimisation between synthesis and observation. Each line is treated independently, and therefore a pseudo-continuum is used to locally normalise the spectral region of 20 Å around that given line using a linear function. The method is also able to automatically determine the mask (spectral regions) that will be used for calculating the χ^2.

3.1.2. Spectroscopy Made Easy (SME)

SME was developed in Uppsala and in the USA (Valenti & Piskunov 1996; Piskunov & Valenti 2016) and has been used by S. Buder for this work. SME is written in IDL and performs a χ^2 minimisation using synthetic spectra computed on-the-fly. The radiative transfer code is an integral part of the SME package. The fitting is done in pre-defined regions, which can be a spectral segment, an absorption line or simply a region of the line. The synthesis is done locally around that line, in a segment that is pre-defined by the user. The observed spectrum is normalised locally for that given segment, taking pre-defined continuum regions and fitting a straight line to the selected high-flux points in these regions.

3.1.3. Interactive spectroscopic framework (iSpec)

iSpec was developed by Blanco-Cuaresma et al. (2014a) and was used by K. Hawkins for this work. iSpec follows the procedure of SME but is written in python and uses the radiative transfer code SPECTRUM (Gray & Corbally 1994) in this case. The continuum is also determined locally, but fits a function to the pre-defined continuum regions that can be either a spline or a

Table 3. Absolute abundances, A(X) and equivalent widths (EW) of the four chosen lines in each GBS from Paper IV, for the methods considered in this work.

Star	Element		Line	EW [mÅ]				A(X)				
	Symbol	Z	[Å]	PORTO	BOL	UCM	ULB	iSpec	PORTO	BOL	UCM	ULB
Sun	Ca	20	5260.38	34.30	30.30	–	34.31	6.33	6.25	6.23	–	6.31
	Cr	24	5238.96	15.30	14.00	15.50	16.84	5.60	–	5.43	5.47	5.52
	Mn	25	6021.81	90.60	–	–	59.38	5.74	5.49	–	–	5.38
	Co	27	5352.04	24.90	21.40	24.20	–	4.85	4.76	4.78	4.81	4.82
Arcturus	Ca	20	5260.38	74.20	70.40	–	80.27	5.93	5.72	5.70	–	5.87
	Cr	24	5238.96	49.90	–	50.60	51.79	5.01	4.83	–	4.84	4.87
	Mn	25	6021.81	145.00	135.70	–	93.98	4.89	–	4.92	–	4.61
	Co	27	5352.04	70.20	61.50	–	–	4.57	4.47	4.45	–	4.36
61 Cyg A	Ca	20	5260.38	83.50	69.00	–	88.05	6.05	5.73	5.75	–	5.99
	Cr	24	5238.96	44.30	32.20	–	32.00	5.16	–	4.89	–	4.88
	Mn	25	6021.81	148.10	132.00	–	–	5.33	–	5.00	–	–
	Co	27	5352.04	25.50	15.60	–	–	4.61	–	4.31	–	3.03
HD 22879	Ca	20	5260.38	12.50	9.90	–	12.53	5.85	–	5.65	–	5.76
	Cr	24	5238.96	–	–	–	2.72	4.96	–	–	–	4.68
	Mn	25	6021.81	24.80	24.40	–	–	4.40	–	4.29	–	–
	Co	27	5352.04	5.40	–	–	–	4.23	–	–	–	4.09

Table 4. Reference absolute abundances of the four chosen lines for each GBS, taken from Paper IV.

	A(Ca)	A(Cr)	A(Mn)	A(Co)
Sun	6.28	5.51	5.34	4.83
Arcturus	5.82	4.97	4.41	4.45
61 Cyg A	5.87	5.11	4.83	4.10
HD 22879	5.75	4.77	4.08	4.11

Table 5. Summary of spectroscopic methods employed in this work.

Name	Approach	Radiative transfer code	Wrapper
iSpec	synth	SPECTRUM	iSpec
ULB	synth	Turbospectrum	BACCHUS
SME	synth	sme_synth	SME
Porto	EW	MOOG	
BOL	EW	SYNTHE	GALA
UCM	EW	MOOG	StePar

Notes. The name of the method, the approach (EW: equivalent width, synth: synthesis), the radiative transfer code employed and the wrapper code that uses the radiative transfer code (if applicable) are indicated.

polynome of an order specified by the user. Here, a second order polynomial was used for the local normalisation.

3.2. EWs

3.2.1. Bologna (BOL)

This method has been developed at the University of Bologna and the results obtained with it in this work were provided by L. Casamiquela. In this method, two steps are performed with two different codes. First, the EWs themselves are measured using DOOp (Cantat-Gaudin et al. 2014), which is an automatic wrapper for DAOSPEC (Stetson & Pancino 2008). The second step is to input the EWs into GALA (Mucciarelli et al. 2013),

which uses the radiative transfer code SYNTHE (Kurucz 2005) and optimises the physical parameters using the classical spectroscopic methods based on iron lines. Afterwards it measures the line abundances. Continuum subtraction in this case is performed for the determination of EWs globally, as described in Stetson & Pancino (2008).

3.2.2. Universidad Computense de Madrid (UCM)

This method has been used by H. Tabernero for this work. As for BOL, two different codes are considered: EWs are determined with TAME (Kang & Lee 2012) and the abundances with MOOG (Sneden 1973) version 2014. The continuum is also normalised globally for the EW determination. TAME does not allow determination of EWs without normalising the spectra, that is, TAME always performs a re-normalisation of spectra for the EW determination. The code StePar (Tabernero et al. 2012) is employed as a wrapper code to call MOOG and derive abundances in an automatic way.

3.2.3. University of Porto (Porto)

This method has been used by V. Adibekyan for this work and is based on ARES v2 code (Sousa et al. 2015) for the EW determination and on the 2014 version of the MOOG radiative transfer code for the abundances. Measurement of the EW is done locally around each spectral line by performing a local normalisation of spectra. As for TAME (above), ARES does not allow EW determination without re-normalisation of the spectra. For details on derivation of stellar parameters and chemical abundances with ARES+MOOG we refer the reader to Sousa (2014), Adibekyan et al. (2012b).

4. Results

As in Papers III and IV, our results were obtained by fixing the stellar parameters (T_{eff}, $\log g$, [Fe/H], v_{mic}, $v\sin i$) to the

recommended values (see Table 1). We performed eight different tests, listed below.

- *Linelist*: to understand the effect on final abundances that different methods have in dealing with a shifted wavelength of an element in the linelist.
- *Continuum normalisation*: to study how the different methods normalise their data.
- *Hyperfine structure*: to quantify this, often neglected, quantum effect on final abundances.
- *Microturbulence*: to study the dependence of the final abundances on this parameter.
- *α-element abundances and continuum opacities*: to study how the different methods treat the α-enhancement in the spectral modelling.
- *Atmospheric model interpolation*: to test the impact of using different interpolated atmosphere models on abundances.
- *Blends*: to study the impact of subtle blends on the final abundances by different methods.
- *Radiative transfer codes*: to compare the radiative transfer codes using otherwise identical data.

We aim at making a systematic study of these effects, which are commonly not discussed in the literature because they are believed to have a small impact on the final abundances. To do so, we determined abundances for each of these tests, leaving the remaining input parameters fixed. We highlight here that this is by no means a complete list of possible tests to investigate effects of abundance determination using different methods. There are many other effects that are not explored in this work but should be explored in the future in a similar way as those here. One example is the effect of using self-consistent model atmospheres when abundances of some elements are more extreme, such as in the case of second generation globular cluster stars. Other examples are the use of different source data for the line modelling, such as oscillator strengths, line opacities in the source function, the partition functions and the damping approximations. For the latter, in our case we used the damping constants provided in the GES linelist, that is, van der Waals constants of Barklem et al. (2000) for Ca, Cr and Co, and the semi-empirical calculations of Kurucz (2007) for Mn. In this work we have kept this information fixed but further effects of abundances in this direction are interesting to explore.

The results from the eight tests performed here are discussed separately in the following sections.

4.1. Linelist

We found that the wavelength of the selected Mn feature was incorrect in the GES linelist version 5.0. This was noticed because no extra local RV correction was applied to the synthesis methods and therefore the core of the Mn line in the synthesis of each star was slightly shifted (by the same amount) with respect to the observation. An example is shown for Arcturus in the upper panels of Fig. 2, in which we show the syntheses of iSpec and SME with and without hyperfine structure splitting (HFS) along with the observed profiles of Arcturus. To build the new hyperfine structure list, we considered the weighted mean position of the line (hence the weighted median upper and lower energy levels of the transition; the HFS constants A and B of each energy level; the angular momentum quantum number of the levels (J) and the nuclear spin of the atom (I) and the total $\log gf$). Applying the Casimir formula (Casimir 1936; Schwartz 1955) and the Russell-Saunders formulae (Condon & Shortley 1935) for relative line strength allows us to derive each component and its
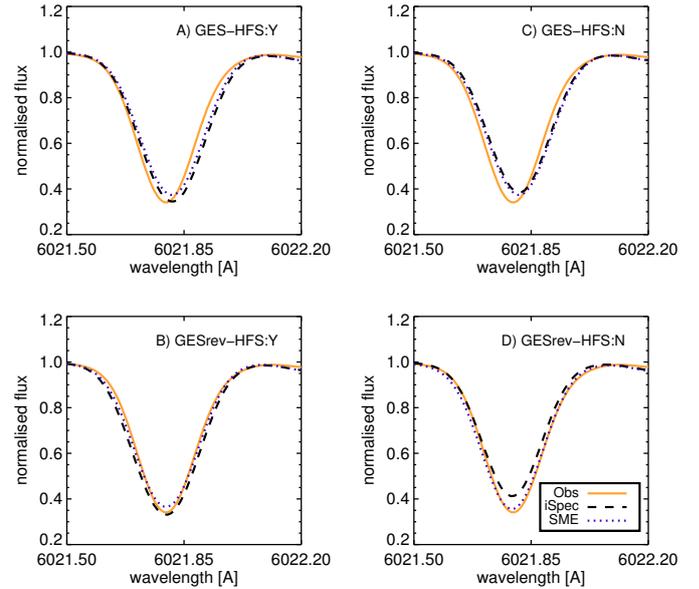

Fig. 2. Profiles of Arcturus of Mn line. Each panel represents the syntheses of SME and iSpec corresponding to the final abundance in each of the runs.

respective $\log gf$. Although the HFS constants and $\log gf$ values are those from the GES linelist, the HFS components in the GES linelist around the Mn line are binned to 10 mÅ, while here no binning was applied. This allows us to test not only the effect of the overall line shift, but also the apparent number of components of the HFS. The lower panels of Fig. 2 show the syntheses of SME and iSpec of Arcturus with the revised line data along with the observations. One can see that the core of the line agrees in the new case.

In automatic determination of elemental abundances of large high-resolution spectroscopic surveys, it is highly probable that some atomic transitions will have laboratory wavelengths that do not perfectly agree with the observations. Furthermore, it is known that atmospheric motions can cause slight shifts in the position of atomic lines, also producing a slight disagreement between the laboratory wavelength and the observation (Molaro & Monai 2012). Finally, sometimes wavelength calibration problems exist, shifting some lines from the laboratory (see also extensive discussion on the impact of determining stellar parameters and abundances when issues of wavelength calibration arise in Hinkel et al. 2016). Since many methods perform only a global RV correction, abundances determined from shifted lines can be more uncertain. In this section we quantify this effect.

Furthermore, since Mn is an element that is affected by HFS, the splitting in the line list can be binned in different ways. In addition to the different wavelength position of the core of the line, the GES linelist included six components, while the corrected line included 15 components. This allows us to study how the number of components considered for the HFS splitting affect the final abundance.

The abundances of Mn were derived four times for each star by each method as follows:

1. no HFS, linelist version 5.0 (incorrect wavelength);
2. no HFS, GES linelist version 6 (correct wavelength);
3. with HFS, GES linelist version 5.0 (incorrect wavelength);
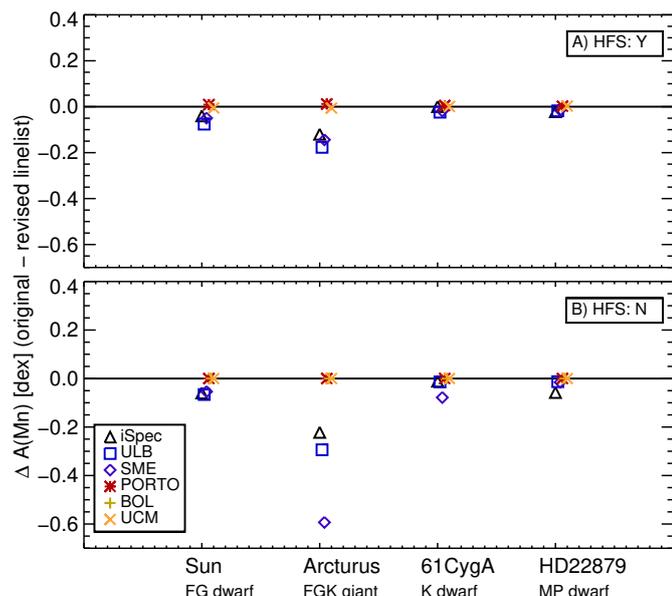4. with HFS, GES linelist version 6 (correct wavelength).

Fig. 3. Difference in Mn abundances per star per method between the original Mn wavelength and the correct Mn wavelength for **A)** considering HFS of Mn; and **B)** not considering HFS.

The results of these four runs are shown in Fig. 3, in which the abundances determined from the old and the new linelist were compared for each star and method. In the upper panel, we compare the results when HFS is taken into account (hereafter HFS:Y), while in the lower panel we compare the results when HFS is not taken into account (hereafter HFS:N).

In agreement with the results obtained in Hinkel et al. (2016), we found that when considering HFS:N, we can see in the bottom panel of Fig. 3 that the EW methods (PORTO, UCM and BOL) do not show a difference in the derived abundance when the linelist presents a small shift in wavelength. This is not surprising, as abundances determined from EW consider the curve of growth, which relates the strength of a line with the abundance. Thus, a small shift in the line list does not affect the measurement of the line strength, that is, the EW itself. For the spectrum synthesis methods (SME, iSpec and ULB), on the other hand, a difference in the derived abundance is seen. This is also expected since synthesis methods perform a pixel-by-pixel fitting between model and observation and therefore it is important that both profiles are properly aligned. The magnitude of the difference varies between the stars and methods, where weak lines, such as the metal-poor star, show small differences of the order of 0.06 dex by iSpec and 0.01 dex by SME and ULB, and the stronger lines, such as the one of the giant, can produce significant differences of up to 0.6 dex by SME. Note that even in the case of the Sun, a difference of 0.06 dex by ULB in retrieved abundances can arise for a bad position in a given line. In the right hand panels of Fig. 2, one can see how the fit improves from Panel C to D, in particular for SME, when the line is properly aligned for synthesis and observation.

When considering HFS:Y, we can see in Panel A of Fig. 3 that even the EW methods are slightly affected. Although the same EW is used in all cases, the different components in the HFS produce an abundance that differs by up to 0.01 dex in the case of Arcturus when using the PORTO method. We note here that BOL results are not shown, because this method cannot take into account HFS. For the synthesis methods, the differences are less significant than when we consider HFS:N, which

is due to a smaller overall shift of model and observation. Interestingly, the extreme difference of 0.6 dex shown by SME in the HFS:N case drops to less than 0.1 dex for the HFS:Y case. This can be seen in the left hand panels of Fig. 2, in which the left wing of the line is less affected by this shift when the synthesis takes HFS into account.

It is instructive to comment that this issue is well-known and therefore most synthesis methods have a way to deal with it. For example, SME has a functionality implemented, which locally corrects the radial velocity of the star, in order to align the synthetic with the observed profile. Other methods lacking this implementation, such as ULB and iSpec in this work, but also several other ones reported in the literature, especially those based on a grid of synthetic spectra (e.g. Recio-Blanco et al. 2006; García Pérez et al. 2016), require a more detailed preparation of the line list.

A common approach is to generate "astrophysical calibrations" to correct oscillator strengths, damping coefficients, and, in several cases, the central wavelength of the line. Such calibrations are performed by fitting synthetic spectra to spectra of well-known stars (the Sun and Arcturus for instance) while determining the aforementioned line list parameters (see e.g. Shetrone et al. 2015; Boeche & Grebel 2016). This ensures that the line will provide accurate synthetic spectra for the analysis. Another option is to reject bad lines based on the confidence of the atomic data from the laboratory measurements, as well as the good reproducibility of the line profile of the synthesis of these well-known stars (Heiter et al. 2015b). This is the procedure implemented, for instance, by GES. In both approaches, a fraction of lines with inaccurate parameters may remain after processing, due to the large number of lines that are observed by current high-resolution spectrographs. We noticed here that even in our highly detailed analysis of GBS based on the careful preparation of the GES linelist, the Mn line had a core wavelength that was shifted. Such issues are best assessed with an extra correction, such as that made by SME. This is, however, difficult for grid-based methods, such as MATISSE (Recio-Blanco et al. 2006) or the ASPCAP pipeline of APOGEE (García Pérez et al. 2016).

Conclusion for line list test. EW methods are robust for shifts in wavelength in the linelist, but synthesis methods can be significantly affected and easily produce differences of 0.2 dex. To overcome this issue, synthesis methods should always perform an extra local wavelength correction to ensure proper alignment of model and observation, as well use a well-tested line list. In addition, binning of HFS components by wavelength may have an effect on abundances, although it seems to be small in the case of EW methods. Further studies of a wider variety of elements and transitions are needed to assess this issue. Henceforth, tests will be made with the revised Mn linelist.

4.2. Continuum normalisation

The vast majority of the spectral analysis methods for determining atmospheric parameters and chemical abundances need to perform a normalisation of the continuum flux. This fundamental step is done in several different ways, such as by fitting global polynomials to the pseudo-continuum (Allende Prieto et al. 2006; Stetson & Pancino 2008; Jofré et al. 2010; García Pérez et al. 2016) or more local functions using synthetic spectra or linear fits, such as the synthesis methods employed in this work. Other methods do not normalise the continuum and perform the parameter determination, adding an extra
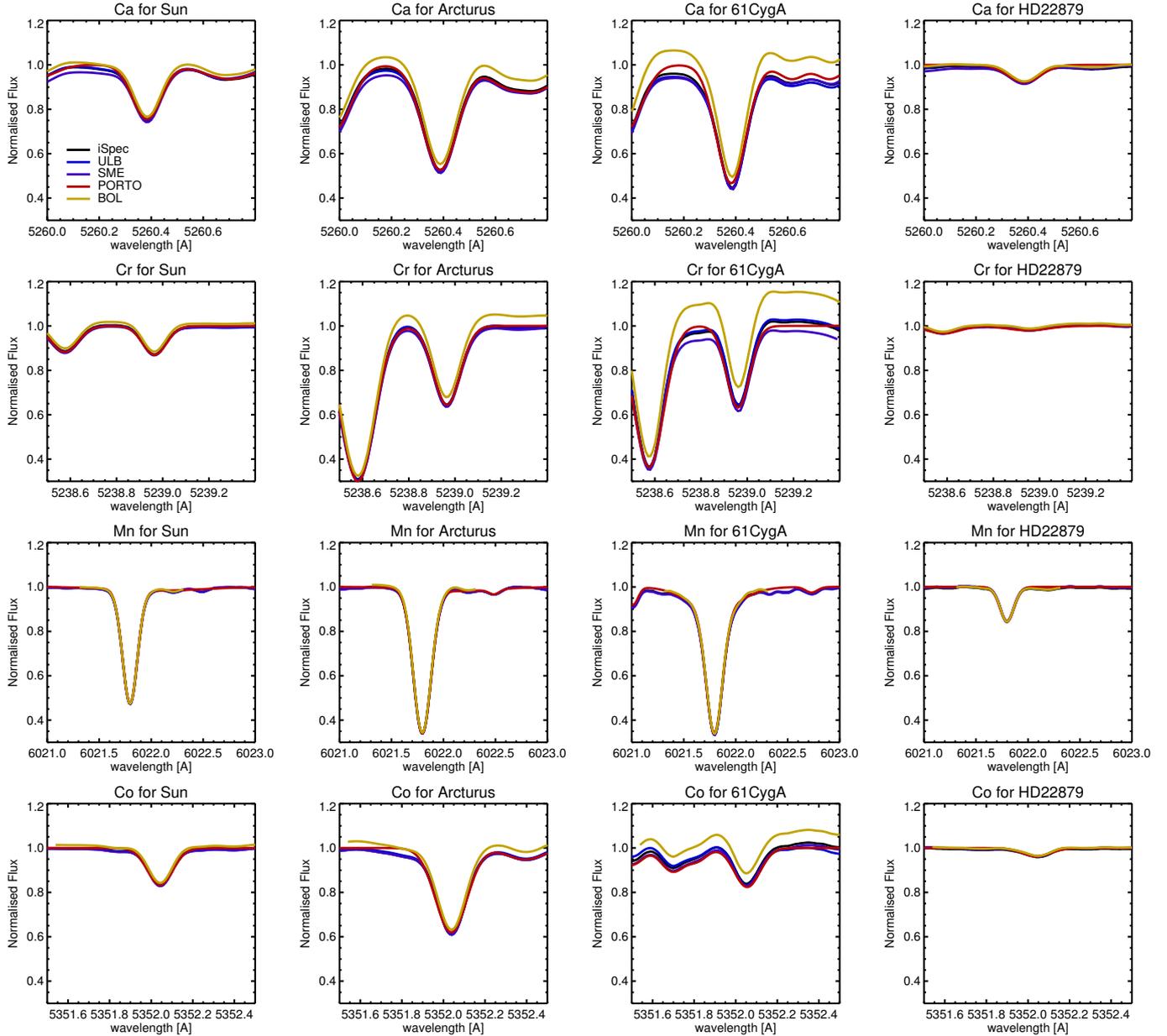
Fig. 4. Spectra of all GBS and lines used in this work as provided after the global normalisation performed with `iSpec`. Re-normalised observed spectra are shown with different colours for the different methods, as indicated in the legend.

parameter to the fit for this function at each pixel (Katz et al. 1998; Koleva et al. 2009). Since there is still no consensus on the best way to treat the continuum flux of the data, it is well known that the measured abundances carry an error due to normalisation. However, since this process is done in such a large variety of ways, it is not always possible to quantify this source of error in determination of elemental abundances.

In this test, we attempt with our methods to estimate typical uncertainties due to continuum normalisation. To do so, the abundances of each element were determined four times for two different normalisation procedures (S and B, see below), as follows:

1. each method performed its own normalisation to the provided spectrum S ($free\_S$);
2. each method analysed a fixed spectrum S and performed no further normalisation (fix_S);

3. each method performed its own normalisation to the provided spectrum B ($free\_B$);
4. each method analysed a fixed spectrum B and performed no further normalisation (fix_B).

Here the abundances of Mn and Co were determined without considering HFS. This allowed a comparison of all methods that only concentrated on the continuum normalisation issue. We further highlight here that the spectra used in this work have already been pre-normalised as described in Paper II, that is, using the synthetic spectra with stellar parameters of the star to find the continuum points and then fitting cubic splines to the local pseudo-continuum. Thus, any further normalisation test essentially meant performing a re-normalisation of the provided spectrum. In this sense, when we refer to the S spectrum, we consider the original spectrum as normalised by `iSpec`, while the B spectrum refers to the spectrum of `iSpec` re-normalised by `BOL`. We mention that `BOL` re-normalised the entire continuum, including
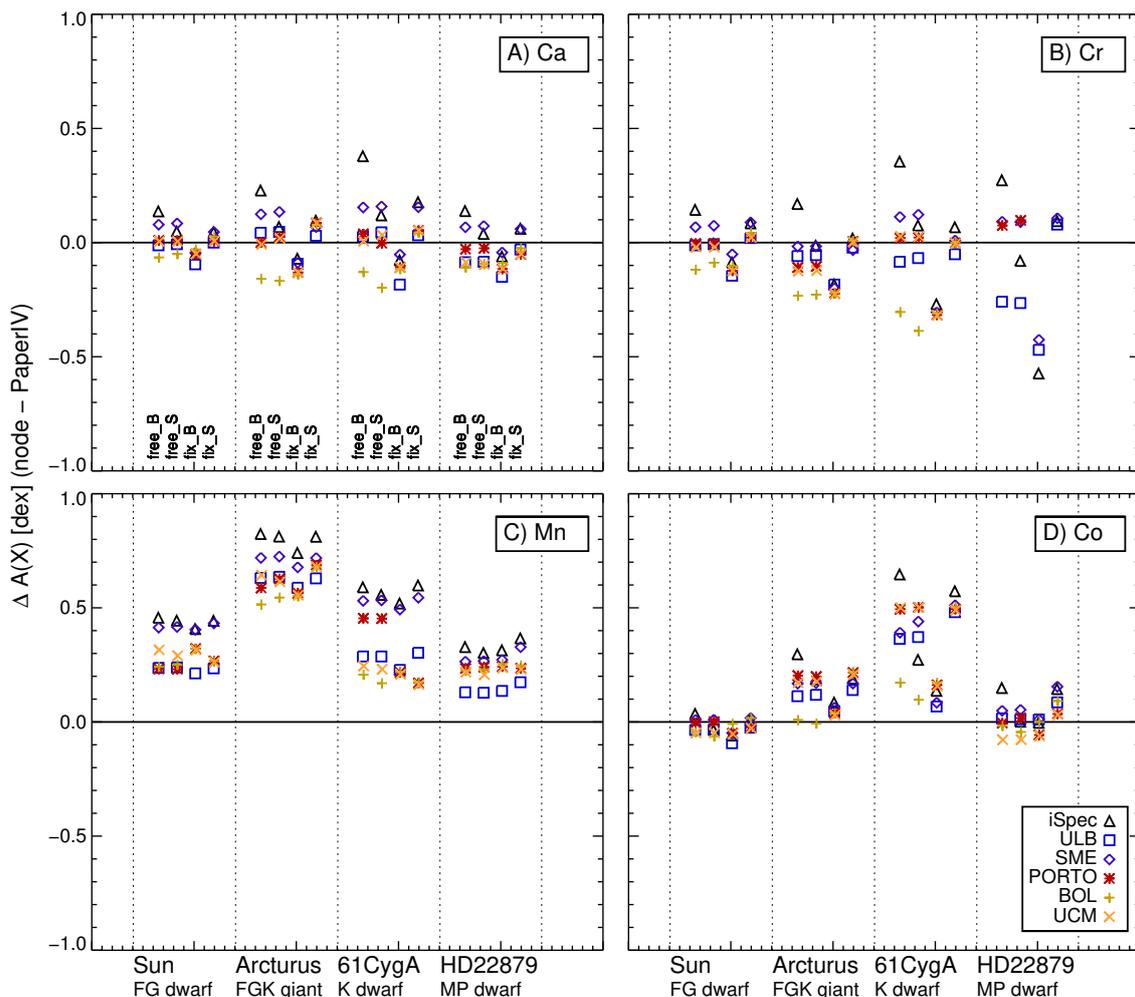
Fig. 5. Difference in element abundance derived by each method for each normalisation test compared to the abundances of Table 4 for each element in a different panel.

the Cr line in HD 22879, but did not measure abundances because the line is too weak. We also mention that in the *free* test, the methods `iSpec` and `BOL` performed a re-normalisation of their already normalised spectrum. In the case of the *S* spectrum, the continuum was normalised globally using a synthetic spectrum with the stellar parameters of the star (see Paper II), which could be further normalised locally around the line under study. In the case of `BOL`, an iterative process in measuring EWs and fitting the pseudo-continuum is performed (Stetson & Pancino 2008), which allows a re-normalisation of the spectrum *B* by `BOL` in this particular test.

We also note that re-normalising the pseudo-continuum is not unusual in the field. Several GES nodes perform a re-normalisation onto the data products (see Smiljanic et al. 2014, for node description). Indeed, some methods, such as ARES (`PORTO`) or TAME (`UCM`), are not able to determine EWs without normalising and therefore, whenever they analyse a normalised spectrum (such as in this case) they re-normalise. To overcome this issue in this and the following tests using fixed continuum, `PORTO` and `UCM` determined the abundances considering the EWs measured by `BOL`, which is able to determine EWs from fixed normalised spectra.

Figure 4 compares the freely re-normalised observed spectral profiles produced by each method. The 16 panels show each element per star. For the Sun and the metal-poor star HD 22879, the

normalised spectrum produced by each method is almost identical to the input spectrum. For the Ca feature for both stars, there is some slight variation in the placement of the continuum to the left of the spectral line.

For Arcturus and the K dwarf 61 Cyg A there are more gross differences in the normalisation. `BOL`, in particular, consistently places the continuum lower than the other methods for these stars for the Ca, Cr and Co spectral features. This is due to the effective continuum placement technique (see Stetson & Pancino 2008, Sect. 3.2) that places a depressed continuum that tries to take into account, in a statistical way, the unrecognised flux deficits and excesses due to contamination effects. For Mn, `BOL` provides a continuum placement matching those of the other methods and a good fit to the input observed spectrum. The normalisation of the spectral features in 61 Cyg A shows the most difference between the methods. `PORTO` underestimates the continuum placement for the Ca feature, while `SME` overestimates the continuum placement for the Cr feature.

One can see from all panels how the global fit performed by `BOL` is very different from the rest, with a systematic underestimation of the continuum with respect to the rest of the methods. This is the reason why we decided to consider this normalised spectrum for comparing *B* and *S*.

Figure 5 shows the abundances for each star derived in this test compared to the reference element abundances listed in

Table 4. Each panel shows one element and each star is separated by the dashed vertical lines. The four tests are shown next to one another, following the same order indicated in Panel A.

When fixing the continuum (two right-hand tests for each star/element) we can see that the node-to-node scatter in the abundances decreases. Furthermore, this scatter does not seem to depend heavily on the input normalised spectrum, which remains very similar for the fix_B and fix_S cases. The zero point, however, significantly changes, with the results obtained from the fix_B spectrum being systematically lower than those obtained from the fix_S spectrum. This can be explained by looking at the profiles of both spectra, where the continuum of the B spectrum is systematically higher than the S spectrum (see Fig. 4). In the case of each method performing its own normalisation to input spectra, we see in the left hand tests for each star/element in Fig. 5 that the node-to-node scatter is higher when the B spectrum is used with respect to the S spectrum. This suggests that using the spectrum that requires larger continuum corrections by each method (the B spectrum) results in a larger range of abundances.

It is interesting to note that when comparing the $free$ and fix tests using B spectrum, the node-to-node scatter decreases significantly, but also the mean absolute value changes. The latter is because when the nodes apply their normalisation for fix_B, the fit to the observations improves and the abundances are closer to that of the method that provided the observations (BOL). While the node-to-node scatter of the S spectrum from the $free$ to the fix test improves, the zero point does not change significantly. Because the S spectrum is already normalised well to the continua provided in Paper IV, adding a correction to the normalisation in this case introduces only noise in the measurement. The mean fix_S is close to iSpec, analogous to the finding that the mean fix_B is close to BOL.

Another interesting feature to note from Fig. 5 is that for Arcturus and 61 Cyg A, a large node-to-node scatter for all determination of abundances is obtained in the $free$ test, reflecting the well-known challenge of normalising cool stars, which are very crowded by absorption lines (e.g. Lebzelter et al. 2012, as well as Papers III and IV for extensive discussions). On the contrary, HD 22879 has a smaller node-to-node scatter, since it presents few and weak lines, facilitating continuum normalisation. Still, freeing the continuum can produce larger node-to-node scatter for very weak lines such as Cr (see Fig. 4), than when the continuum is fixed.

Finally, we also highlight here the significant differences in the abundances determined in this work with respect to those of Table 4, such as Mn for all stars and Co for 61 Cyg A. One difference is that here, HFS was neglected, while the reference value is only from methods with HFS. See Fig. 11 in Paper IV showing the size of the effect for the Sun, which is similar to that seen here. See also Sect. 4.3 and Fig. 6. Another difference is that here we employ the revised line list while in Paper IV the old line list was employed, and the abundance of Mn was provided by ULB among other methods. As discussed in Sect. 4.1, ULB does not perform a further wavelength correction to the line. We recall that the final abundances for Mn of GBS are based on several Mn lines but a revision of this line should be performed in future releases of abundances of GBS. In this context, we also remind the reader that we do not intend to correct nor reproduce the reference abundances of Paper IV in this work, here we only use them as a comparative value for assessing the node-to-node scatter. The final Mn abundances for the GBS in Paper IV follow a different strategy of performing differential abundance with respect to the Sun, aiming at reducing the systematic uncertainties

such as continuum placement (see Paper IV for extensive discussion). The values used here are the average of the absolute abundances obtained by each method that considered HFS, and do not reflect the final differential abundances suitable for calibration purposes obtained in Paper IV.

Conclusion for continuum normalisation test. The agreement between methods is improved if the continuum of a spectrum is fixed, although the final absolute values highly depend on which continuum is applied to the data. Therefore, a careful continuum normalisation should be performed, which then should be kept fixed for abundance determination. Hereafter our tests are be made with the fix_S method, which means that abundances determined by UCM and PORTO consider the EWs determined by BOL.

4.3. Hyperfine structure splitting (HFS)

Current high-resolution spectrographs such as UVES, APOGEE and GALAH can fully resolve many features in absorption lines. On top of the basic atomic data of wavelength and oscillator strength, information on line broadening due to quantum effects such as HFS can be disentangled. Since HFS causes the line profile to increase in width and the peak intensity of the line to decrease, the line appears asymmetric and can no longer be described by a simple Voigt profile (Blackwell-Whitehead et al. 2005). Neglecting this results in the incorrect measurement of the wavelength and a miscalculation of the equivalent width or synthetic spectrum, leading to an incorrect calculation of the abundance of this given element.

Manganese is known to be strongly affected by HFS producing differences in retrieved abundances of up to 0.6 dex (e.g. del Peloso et al. 2005, North et al. 2012, and Paper IV). Studies of HFS effects on other odd-Z elements such as Co also show that differences in derived abundances of Co can be up to 0.1 dex (del Peloso et al. 2005; Boyarchuk et al. 2008, see also Paper IV). Although the majority of modern studies of elemental abundances take this effect into account, there are still some cases in which HFS is not implemented (e.g. Neves et al. 2009; Jofré et al. 2015a; Adibekyan et al. 2015, to name a few). One example in our work is the BOL method, which derives abundances from EWs with the radiative transfer code SYNTHE. This code does not support calculations of HFS. For this reason, the BOL method is not considered in the discussion of this section.

Here we attempt with our methods to estimate typical uncertainties due to HFS. To do so, the abundances of Mn and Co were determined twice, as follows:

1. each method determined abundances considering hyperfine structure splitting (HFS:Y);
2. each method determined abundances while neglecting hyperfine structure splitting (HFS:N).

In Fig. 6, we compare the abundances obtained by each method for each star for Mn and Co, where we plot the differences of the abundances obtained from both runs for Mn in the upper panel and Co in the lower panel. Each method is represented by a different symbol and colour as indicated in the legend, using the same symbols and colours as in previous figures. We can see that for both cases, HFS:N systematically overestimates the abundance by a value that depends on the spectral type and element. Mn shows large differences of up to 0.4 dex in the case of Arcturus. The differences in abundances between HFS:N and
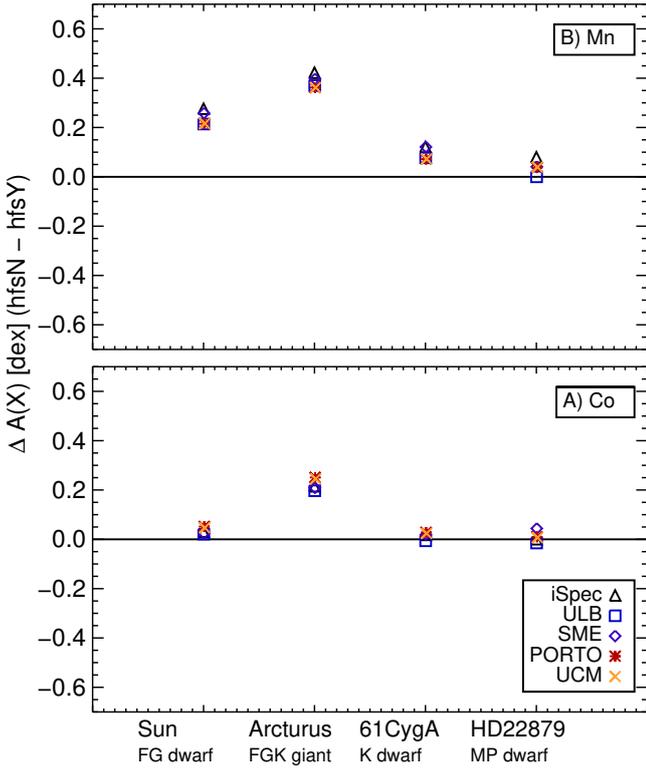
Fig. 6. Difference in Mn and Co abundance derived by each method with and without HFS.

Fig. 7. Difference in Mn with the reference value for each method with and without HFS.

HFS:Y are similar for all methods, with the metal-poor star presenting no differences at all for ULB. The fact that the differences are significantly larger for Arcturus than for HD 22879 demonstrates how the HFS effect depends on line shape: stronger, more asymmetric lines with broader wings result in larger differences in abundances when the modelled line is split into its HFS components. We remark here that, even for the Sun, a difference of 0.2 dex in Mn abundance can be obtained due to HFS.

The HFS of Co is more subtle than for Mn, and therefore for Co, the differences are in general smaller than for Mn, below 0.05 dex in most cases. Arcturus, which is the star with the strongest lines overall (see Fig. 1) is an exception, presenting a difference of more than 0.2 dex.

Figure 7 shows another striking result: while the abundances obtained under HFS:Y are systematically lower than the results obtained under HFS:N, the node-to-node scatter does not seem to significantly decrease. This also shows that even if HFS is taken into account, the abundances might be uncertain due to the HFS treatment, binning of the different components in the line list (see Sect. 4.1) and the proper modelling of lines that strongly deviate from having Gaussian or Voigt profiles. The line under study, even in the case of the Sun, can produce 0.2 dex differences in abundances obtained by different methods.

Conclusion for HFS. Overabundances of the order of 0.4 dex can appear in strong Mn lines in giants if HFS is neglected, but even for weak lines, overabundances of 0.15 dex can be obtained. The HFS effect of Co is less significant but even for giants can be of 0.2 dex. These differences are much larger than accuracies of 0.05 dex in iron-peak elements needed for disentangling different stellar populations, star formation histories and constraining chemical evolution models as well as nucleosynthesis yields (e.g. Feltzing et al. 2007; Nomoto et al. 2013). We would
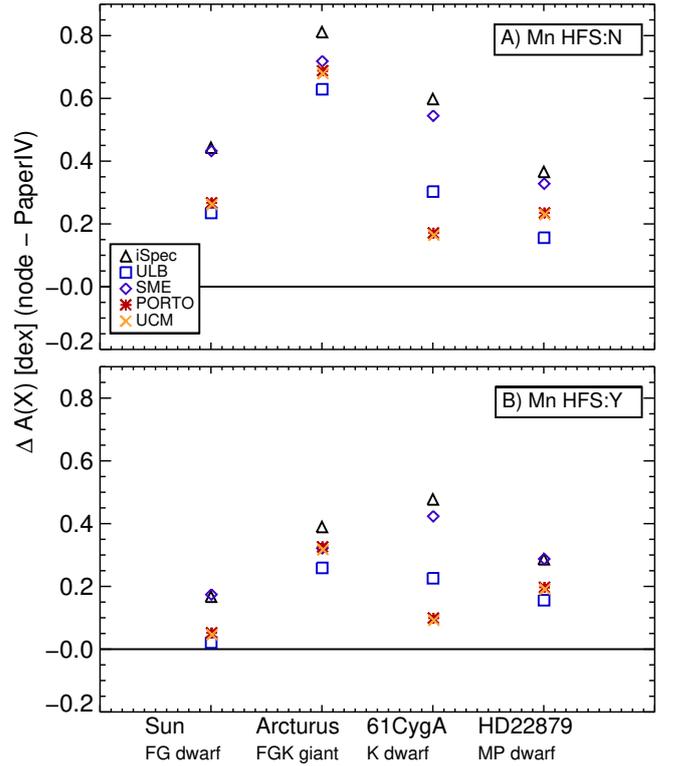
like to point out that in the fourth release of GES (Hourihanne et al., in prep.), HFS is not taken into account by all the nodes, adding, in some cases, uncertainties of the order discussed here. The results of this work have motivated the GES node UCM (see Smiljanic et al. 2014) to include HFS for future GES data releases, which should improve the accuracy of the derived abundances of this survey.

4.4. Microturbulence

When deriving atmospheric parameters and chemical abundances using 1D model atmospheres, the microturbulence (v_{mic}) parameter also needs to be determined. In 1D models, this parameter is meant to account for some of the turbulent motions in the atmosphere that cause the spectral lines to be broadened. A realistic description of this physical mechanism leading to an accurate modelling of the line profiles can only be done in 3D. This is computationally expensive and currently only a small number of these accurate models are available to the community (Magic et al. 2013). The stronger the line, the more it is affected by this broadening and thus by v_{mic}. Because in spectral synthesis calculations under 1D v_{mic} is not physical, each code can employ slightly different values for v_{mic} given an otherwise defined set of stellar parameters (see Paper III). Moreover, it has been found that the microturbulence derived by minimising the trend of Fe line abundances against reduced equivalent width decreases with increasing metallicity (e.g. by Hawkins et al. 2016b, using APOGEE spectra of giant stars).

In this section we attempt, using our methods, to estimate typical uncertainties due to different values derived or assumed for v_{mic}. To do so, the abundances of all elements were determined six times fixing values of v_{mic} = [0.8, 1.0, 1.2, 1.4, 1.6, 1.8] km s^{-1}. The results for all stars,
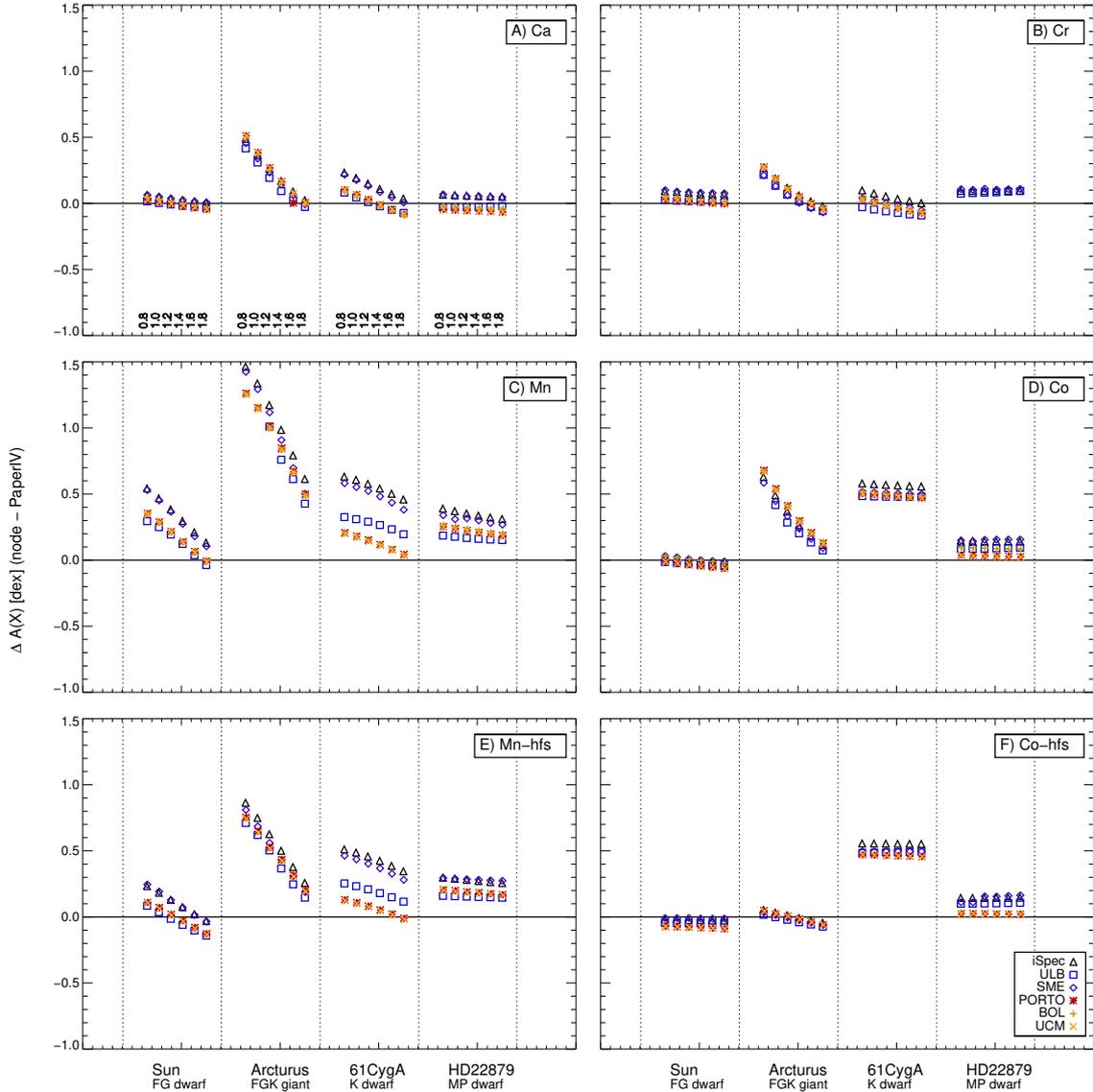
Fig. 8. Difference in element abundance derived by each method for each value of microturbulence compared to the reference abundances for each element.

elements (with and without HFS for Mn and Co) and methods are shown in Fig. 8. Panel A shows the abundances of Ca for all methods and stars and, as in Fig. 5, the results of all runs are plotted side-by-side. The values of v_{mic} are indicated at the bottom of the panel and they follow the same order in each star and the rest of the panels. The methods are indicated by different colours and symbols, which are explained in the legend and are the same as in previous figures. Panel B shows the results for Cr, Panels C and D show the results for Mn and Co neglecting HFS, and Panels E and F show the results for Mn and Co when HFS is considered.

The strong dependency of derived abundance on v_{mic} can be seen for almost every star, element and method. Exceptions are Cr and Ca for HD 22879 and Co for all stars except Arcturus. When there is a dependency, it behaves in every case in the same way; the larger the v_{mic} value, the smaller the obtained abundance. This is because a large v_{mic} value will imply a stronger line and therefore less abundance is needed to model a line of a given strength. Note that for this test we varied the microturbulence by a large amount (1 km s^{-1}). For example, differences of 0.2 km s^{-1}, as found in Paper III, correspond to differences of 0.2 km s^{-1}, as found in Paper III, correspond to differences

0.1 dex in abundance for strong lines such as those of Arcturus but of 0.05 dex or less for weaker lines.

All methods have a very similar dependency with respect to v_{mic}. In each panel, we see that the strongest dependency of v_{mic} and abundances is found for Arcturus, because it presents the strongest lines of the stars in our sample. Likewise, the weakest dependency of abundances on v_{mic} is seen for HD 22879. For the weakest lines of HD 22879, such as the Cr and Co line, the dependency of the derived abundance is almost null. Interestingly, a weaker dependency of v_{mic} with abundances is seen for Mn and Co when HFS is taken into account. This is due to the splitting into several components; v_{mic} plays a less important role as each component acts as a weak line.

Conclusion for microturbulence test. There is a dependency of the abundances on the adopted value of v_{mic}, where larger abundances are derived for smaller values of v_{mic}. This relation is well known and several empirical v_{mic} relations exist in the literature (e.g. Valenti & Fischer 2005; Adibekyan et al. 2012a; Ramírez et al. 2013), which can be used when the spectra

do not allow for determination of this parameter simultaneously with the other stellar parameters (see discussions in e.g. Smiljanic et al. 2014; Hawkins et al. 2016b). As in the case of the continuum normalisation, the absolute value of the abundance depends on the adopted value of v_{mic}, but since this parameter is not a physical parameter per se, it is difficult to find an absolute value for v_{mic} that would account for the correct line broadening in every method. We also found that if HFS is taken into account, strong lines become less dependent on v_{mic}.

4.5. Enhancement of α-elements and continuum opacities

When solving the radiative transfer equations to derive chemical abundances of a star, there are three main inputs. The model atmosphere, providing the pressure and temperature of the gas at different depths in the atmosphere; the line list, providing the atomic (and molecular) data regarding the interactions between radiation and matter in the atmosphere; and the chemical composition, providing the distribution of the different elements in the gas and hence the ionisation equilibrium of atoms and the dissociation equilibrium of molecules. Moreover, chemical composition is key for calculating continuous absorption, scattering coefficients and partition functions. Usually, the chemical composition of the Sun is adopted; in our work this being that of Grevesse et al. (2007). This composition can then be scaled appropriately to different stellar metallicities, which are usually driven by the model atmosphere taken into account and should be consistent with the composition of the model atmosphere.

However, the abundances of the so-called α-elements (referring to O, Ne, Mg, Si, S, Ar, Ca and Ti) are different for different Galactic stars, due to the different supernova Type II rates at different moments and parts of the Milky Way (see Nomoto et al. 2013, for a review). The general trend is that at solar metallicities, α-elements are solar-scaled ($[\alpha/\mathrm{Fe}] = 0$) and the α-element abundance linearly increases towards lower metallicities, reaching a plateau of $[\alpha/\mathrm{Fe}] \sim 0.4$ at $[\mathrm{Fe}/\mathrm{H}] \sim -1$ (see e.g. Edvardsson et al. 1993).

Non-solar abundance patterns can change the overall opacities and therefore the radiative transfer equations can lead to different solutions for the continuum flux as well as for abundances. Carbon-enhanced stars are a good example of this effect, as this element is very abundant in stars, its interactions with photons occur over a large frequency range, and it contributes to continuous absorption in several ways. Its impact in the model atmosphere has been discussed in Gustafsson et al. (2008), in which continuous opacities of C I, C II, C $^-$, as well as CH and CO have been taken into account, and appear to be significantly different for carbon-enhanced metal-poor stars. The enhancement of α-elements is more subtle than for carbon. Although the effects in the temperature-pressure structures in model atmospheres are minor, for cooler stars these effects can become significant, in particular due to TiO absorption. At the same time, electron contributions from Mg and Ca increase, raising H $^-$ opacities. The higher electron pressure causes a decrease in densities and thereby convective flux, which results in a stronger temperature gradient in the deep atmosphere. Therefore, special atmosphere models of α-enhanced stars are nowadays available to the community. This α-enhancement in the MARCS models concerns continuous opacities of O I, O II, O $^-$, Mg I, Mg II, Si I, Si II, Ca I, Ca II as well as TiO. In particular, the model grid employed in this work, as well as in the GES and the ASPCAP pipeline, follows the general trend of the Milky Way, that is, it does not take into account outlier stars such as some accreted halo stars that have α-abundances lower than

the plateau value of 0.4 mentioned above (Nissen & Schuster 2010; Hawkins et al. 2014), or higher α-abundances at solar metallicities (Adibekyan et al. 2012b; Nissen 2015).

In the abundance analysis of α-enhanced stars, α-enhanced model atmospheres should be employed together with consistent input abundances for the radiative transfer calculation. Unfortunately, in the literature, the consistency in composition of the model atmosphere and the radiative transfer calculation are not always clearly documented. Furthermore, it is not clear how the different radiative transfer codes compare when this scaling in the abundances of α-elements is taken into account or when the abundances are assumed to be solar-scaled.

In principle, this applies to every element, but beyond C, N, α- and iron-peak elements, the number of other elements in a stellar atmosphere is too small to significantly affect the overall opacities and radiative transfer equation solutions. In Fig. 9, we show the syntheses performed with SME and ULB considering α-enhanced and solar-scaled abundances for the three stars that are more metal-poor than the Sun, for Cr, which is the bluest line analysed here, and for Mn, which is the reddest line. The syntheses were done without the inclusion of line absorption, and are intended to show the impact on the continuum flux and the continuum opacities due to α-enhancement.

In each panel of Fig. 9, we plot the absolute flux of the syntheses around a spectral window. The units of the flux are indicated in the y-axis. The mean difference between α-enhanced and solar-scaled (in percentage) of the flux in the spectral window for each method is indicated in the legend. We can see that although the same models are used for the synthesis computation, the treatment of opacities due to α-elements is different for both radiative transfer codes, in which the continuum flux of SME is systematically lower than the continuum flux of ULB. Furthermore, the differences between α-enhanced and solar-scaled syntheses are systematically larger in the case of SME with respect to ULB, up to an order of magnitude for Arcturus.

As a next step, we used our methods to estimate typical uncertainties in derived abundances due to this effect. Thus, we performed a test in which the abundances were derived twice, as follows:

1. each method determined abundances scaling the abundances according to the standard α-enhancement of the Milky Way (α-enhanced);
2. each method determined abundances considering solar-scaled abundances (solar).

In both cases, the same model-atmosphere was used, that is, following the standard α-enhancement as a function of metallicity of the Milky Way. In Fig. 10, we show the results obtained in this test. Each panel corresponds to an element, in which Co and Mn were determined with and without HFS. Each method is represented by a different symbol, consistent throughout all figures. The differences between the results obtained considering solar abundances and α-enhanced are shown for each star. For the Sun, the abundance difference is zero, since $[\alpha/\mathrm{Fe}] = 0$ in both cases.

For more metal-poor stars, differences in abundances of up to 0.1 dex were found when comparing the two cases of chemical composition except for the methods using MOOG (PORTO and UCM). MOOG itself does not seem to calculate the electron number density, it rather uses the values read from the model atmosphere throughout its calculations. Since the model atmospheres were not changed in the test, the electron number density did not change either, and thus the continuous H $^-$ opacity is the same.
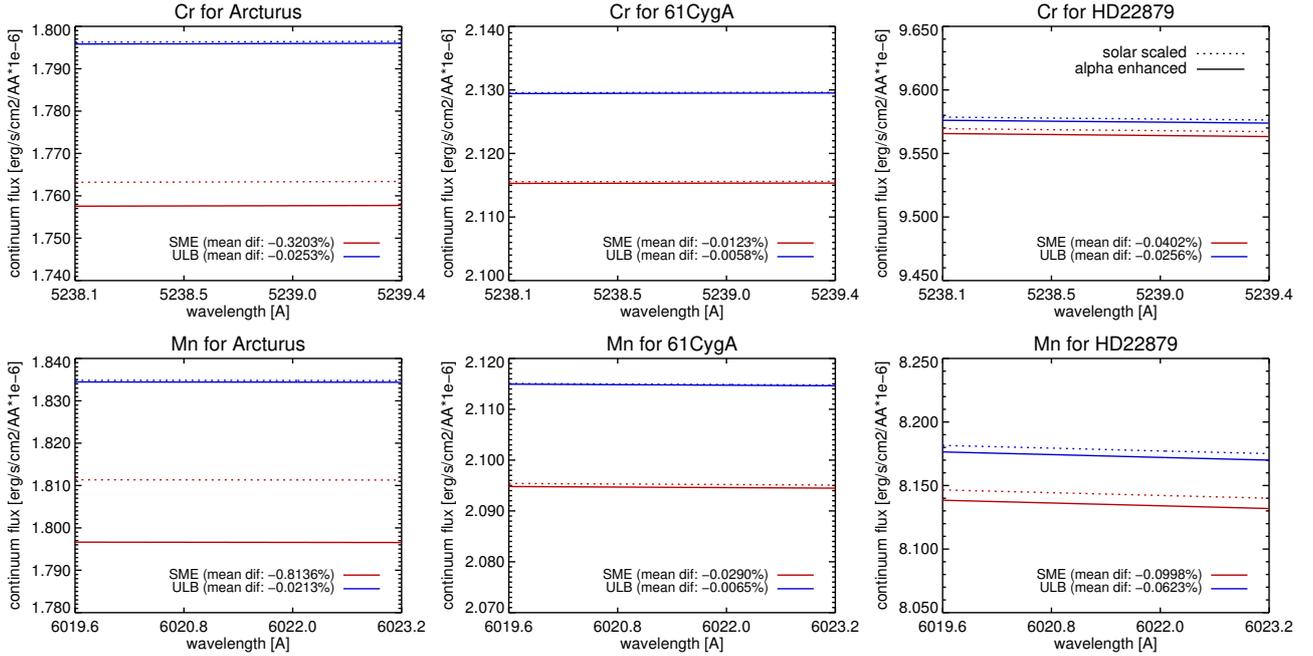
Fig. 9. Syntheses performed with SME and ULB considering α-enhanced and solar-scaled composition. The syntheses were performed without line absorption, and show how the continuum fluxes change when α-enhancement is taken into account or not. The Sun has been excluded from this figure as, by definition, it has zero α-enhancement and the syntheses are identical. The bluest and the reddest line regions analysed in this work were chosen for this illustration. The difference of the fluxes (solar-scaled – α-enhanced) in percent are shown at the bottom of each panel. The syntheses with α-enhancement are plotted with continuous lines while the syntheses with solar-scaled abundances with dotted lines. The syntheses computed with SME are plotted with red colour, while the syntheses computed with ULB (Turbospectrum) are plotted with blue colour.

We verified that all other radiative transfer codes used in this work (sme_synth, Turbospectrum, SPECTRUM, WIDTH9) include their own computation of the electron number density (or electron pressure) while solving the equilibrium equations, taking into account the electron contributions from the ionisation of different species[1]. This is not necessarily a problem for the abundance calculations with MOOG, since the model atmospheres include α-enhancement and the corresponding electron pressure. It means that MOOG does not need to adjust the abundances, but all others should do so, to be consistent with the model atmospheres.

Interestingly, the most metal-poor star HD 22879, which should be the most α-enhanced star of the sample, does not show the largest difference in results. The cool stars Arcturus and 61 Cyg A show a systematic difference in abundance, being higher for α-enhanced composition with respect to solar. This is probably due to the lower continuum flux obtained for the alpha-enhanced composition (Fig. 9). However, the differences of the metal-poor warm stars have a different sign than the cool stars. This suggests that the changes in pressure structure affect the abundances in the opposite direction than the changes in continuum opacities.

Conclusion for α-enhancement and continuum opacities. When α-abundances of stars are different from solar, it is important that the abundances are consistent with the model-atmosphere. While MOOG does this automatically, this is not the case in most of the classical methods for abundance determination. If the abundances are not accordingly scaled, differences of up to 0.1 dex can arise in retrieved abundances of cool stars.
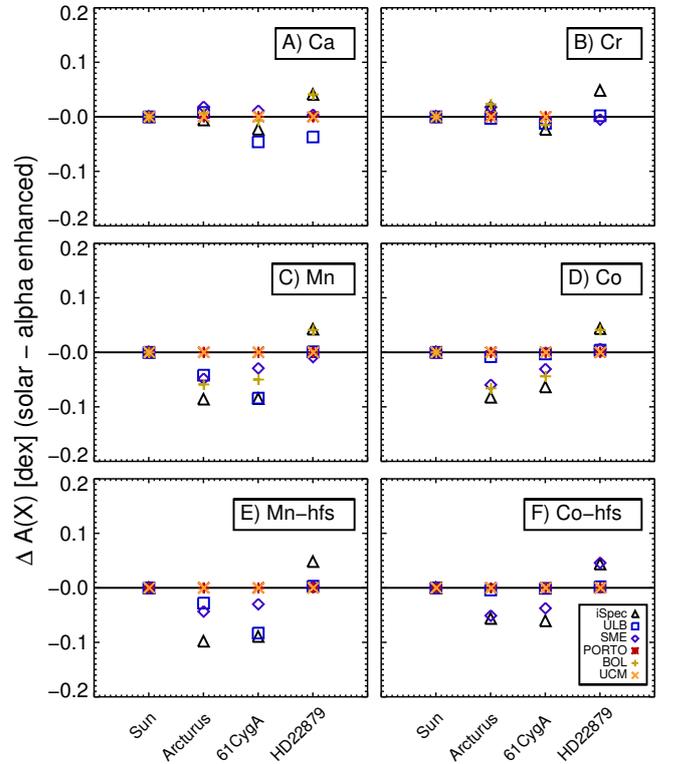

Fig. 10. Difference in element abundance derived by each method for different α-element treatments.

This uncertainty is larger than aimed for current pipelines to determine abundances in large spectroscopic surveys. We recommend to always specify the chemical composition adopted in the
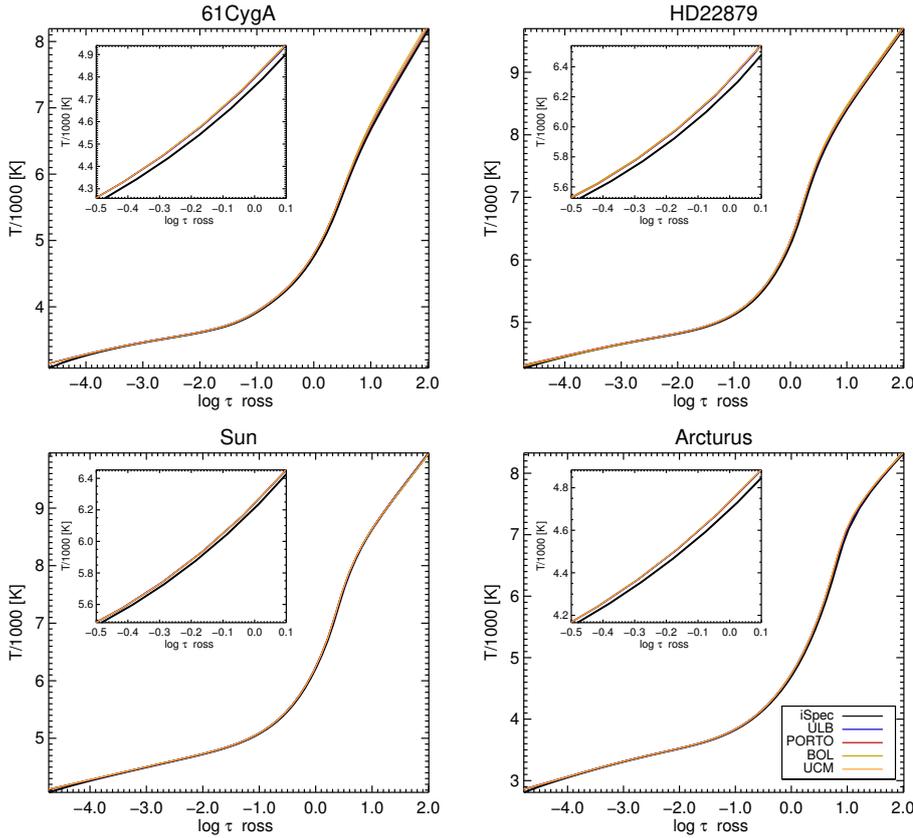
---

[1] See e.g. SPECTRUM manual p. 33 (and file density9.c); WIDTH9/ ATLAS9 subroutine NELECT; Turbospectrum file tryck.f.

Fig. 11. Comparison of interpolated model temperatures as a function of Rosseland optical depth. The profiles for all methods except for `iSpec` lie on top of each other.

radiative transfer calculations when publishing abundances. The size and sign of the abundance difference seems to vary stellar parameters in a complex way, depending on the relative importance of opacity and pressure effects. Further studies for a wider variety of stars and transitions are needed to assess this issue in more detail.

4.6. Atmospheric model interpolation

When deriving stellar parameters and abundances, methods require an appropriate model atmosphere, which is usually taken from a grid of models, such as MARCS, as we have used in this case. Other widely-used 1D-LTE model atmosphere grids are created with ATLAS (Kurucz 1979) and PHOENIX (Husser et al. 2013). More sophisticated model grids considering 3D have also become available recently, such as the Stagger grid (Magic et al. 2013). These grids cover the parameter space at fixed steps in T_{eff}, $\log g$ and [Fe/H], with the 3D grids having larger steps than the 1D grids as the models are computationally more expensive to generate. In either case, it is very common that when deriving abundances, the model atmosphere computed for the exact parameters of the star is not available, so the radiative transfer code operates considering an interpolated model atmosphere of the exact stellar parameters. How this interpolation is done is rarely documented in spectroscopic methods. Here, ULB and BOL follow the interpolation procedure of Masseron (2006), which is the same as the interpolation procedure adopted by the MARCS grid described in Gustafsson et al. (2008). PORTO and UCM employ interpolation code available on the MARCS models web page, and therefore these models should be essentially the same as those of BOL and ULB. SME is interpolating the model atmospheres in a very similar way to Masseron (2006) with a trilinear interpolation between eight corner models surrounding

the target parameters. The procedure has three steps, first obtaining four models at the requested [Fe/H], followed by two models with the desired $\log g$, from which a single output model is produced by interpolation in T_{eff}. On the other hand, `iSpec` first does a bilinear interpolation in T_{eff} and $\log g$ between neighbouring models, followed by a linear interpolation in [Fe/H].

In order to compare interpolated models, we first transformed the model data produced by each method onto a common depth scale, where possible. ULB provided both the optical depth at 500 nm (τ_{500}) and the Rosseland optical depth (τ_{ross}), while SME included only τ_{500}. The remaining methods provided the Rosseland opacity κ_{ross} and the mass variable ρdx, which were converted to τ_{ross} by integration. Further conversions involved those between number densities and pressures, and between quantities given in linear or logarithmic form.

In Fig. 11 we show the interpolated model profiles of all methods with different colours. For this illustration we show the temperature (T) as a function of τ_{ross}. Each panel indicates a star, and the sub-panel shows a zoom of the model around $\log \tau_{ross} \sim 0$. The models produced by SME are not included in this figure, but were compared with those of ULB on the τ_{500} scale and found to be identical. Here we define two models as equal when they differ by less than 0.1%. Therefore, we assume that the results from the comparison shown in Fig. 11 can also be applied for the models interpolated by SME.

From Fig. 11, we can see that the models of ULB, BOL, UCM and PORTO are equal while at $\tau_{ross} \sim 1$, `iSpec` is slightly different with respect to the rest of the models. For the Sun, we find differences of -30 K for $T \sim 6230$ K, corresponding to 0.5%, for Arcturus we find a difference of -40K for $T \sim 4730$K, corresponding to 0.8%, for HD 22879 we see a difference of -60 K for $T \sim 6310$ K, corresponding to 1% and for 61 Cyg A the difference is of -40 K for $T \sim 4800$ K, corresponding to 0.8%.
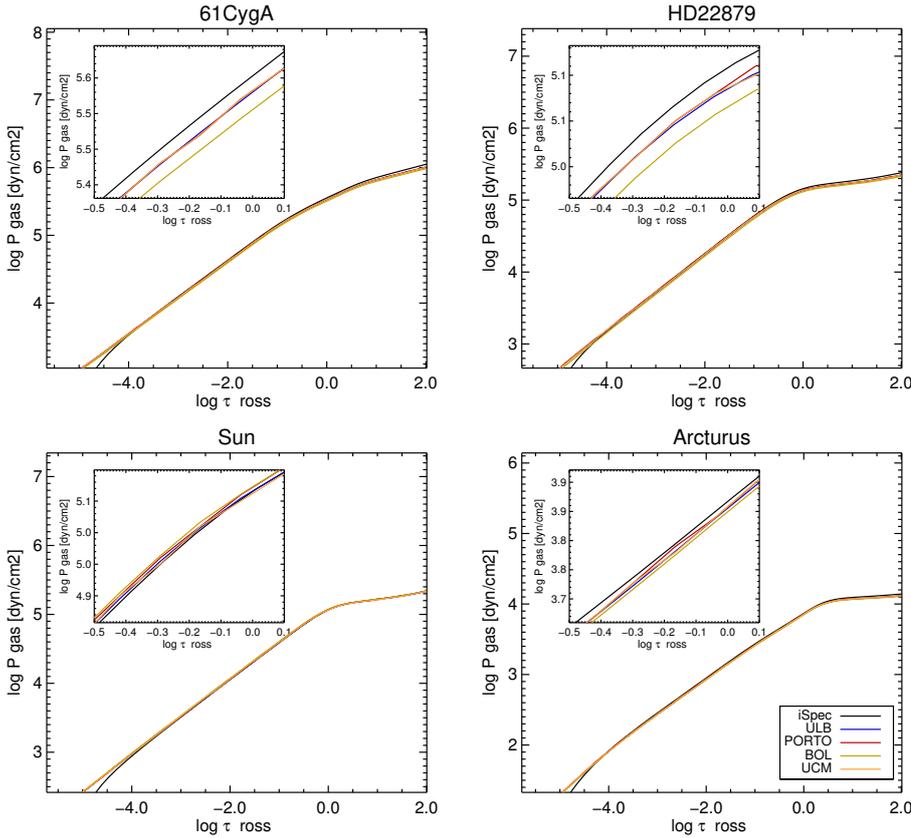
Fig. 12. Comparison of interpolated model gas pressures as a function of Rosseland optical depth.

The gas pressure is plotted as a function of $\log \tau_{ross}$ for all models except SME in Fig. 12. As in previous comparisons, we found that ULB and SME are the same and therefore we can assume that the comparison of SME with respect to the rest of the methods is the same as the comparison of ULB with respect to the rest of the methods. In contrast to the temperature profile, we see that there are more significant differences in the line-forming region between the different interpolated models. For the Sun, we see a difference of 2% to 5%, for Arcturus this is 5% and for HD 22879 and 61 Cyg A this difference is 10%.

Here we estimate typical uncertainties in abundances due to different interpolations of atmosphere models. To do so, the abundances were determined twice, as follows:

1. the methods determined abundances performing their own interpolation of model atmospheres, that is, using the model profiles shown in Figs. 11 and 12 (*free_model*);
2. the methods determined abundances using a fixed model (*fixed_model*).

The *fixed_model* for each star was created using iSpec. The reason is that the latest version of iSpec (Blanco-Cuaresma et al., in prep.; see also Blanco-Cuaresma et al. 2016, for further details) has now included all 1D LTE radiative transfer codes used in the present work, namely sme_synth, MOOG, Turbospectrum, SPECTRUM and SYNTHE9. Therefore, with iSpec it is possible to store the same interpolated model into different formats that can be then applied by the different methods considered in this work. Furthermore, the interpolation procedure of iSpec differs more from the others, making the comparison of retrieved abundances more significant in this case.

In Fig. 13, we illustrate the results of this test. In each panel we show the difference obtained using the pre-defined atmosphere model of iSpec compared to that obtained when an

individual interpolation is performed. Because of the construction of this test, we do not plot the iSpec results as they are the same in both cases. For the other methods, we see that typical differences are of less than 0.04 dex for all stars and lines. Although the interpolated temperature structures were found to be very similar for all methods, the abundance differences vary considerably. For example, while ULB is only very slightly affected by the input model in most cases, SME and BOL seem to be most affected, in particular for the metal-poor and the cool stars.

The methods however respond to the same direction in almost all cases, showing that the dependency of the model when solving the radiative transfer equations is consistent among methods. The magnitude of the dependency is also consistent with the general difference obtained between methods, that is, the largest difference in the gas pressure versus τ_{ross} was obtained for 61 Cyg A of 10%, which displays the largest difference in abundances obtained by all methods. We comment here that this test was done for stars with parameters that are well inside the grid of atmospheres. The impact on abundances obtained for stars with parameters on the edge of the grid was not explored since the stars used in this work have parameters that lie well within the grid of models. The extrapolation of atmosphere models might cause larger differences in abundances than found here.

A potential test, that was not performed here but is of similar nature would be to study the difference of abundances derived when standard model atmospheres or models with self-consistent chemical composition are used. For example, the influence of Mg, Al and Na as electron donors or C in molecular equilibrium can be significant on the structure of the atmosphere of a red giant. This can affect the abundances derived in, for example, second generation globular cluster stars where the α-elements O, Mg and C are depleted whereas Al and Na
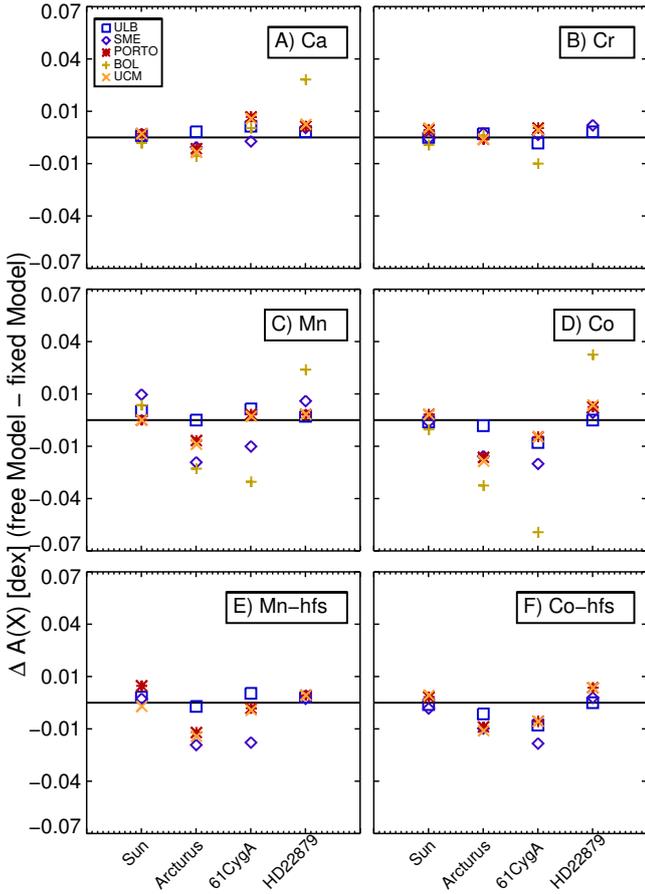
Fig. 13. Difference in element abundances derived by each method by performing its own interpolation of model atmospheres or considering a pre-defined model.
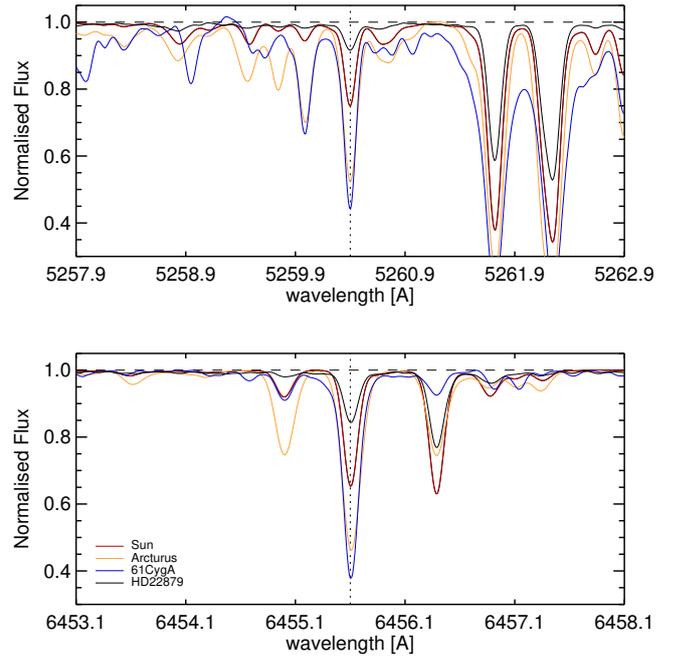

Fig. 14. Observed profiles of two Ca lines in the four GBS. The *upper panel* shows the profiles for the Ca line analysed throughout this work while the *lower panel* shows the profiles for a cleaner Ca line.

are enhanced. To quantify this effect would require model atmospheres calculated for the exact compositions of second generation globular cluster stars.

Conclusion on model interpolation. Even if models are interpolated following very similar approaches, subtle differences of approximately 1% in the temperature and of 5% in gas profiles are found. This translates to differences of the order of 0.02 dex in the derived abundances, although this number depends on the method considered. Such differences are of the order of magnitude of uncertainties accepted by big surveys, but are larger than those aimed for very high precision abundance studies such as those of Nissen (2015) or Spina et al. (2016) of 0.01 dex.

4.7. Blends

Although abundance determination from spectral lines aims at analysing lines that are free of blends and isolated enough to identify robust continuum points, it is difficult to find perfect clean lines for all types of stars (see also extensive discussion in Paper IV). Indeed, even in our careful choice of optical lines for this work, it was not possible to find perfectly clean and unblended lines for all four different GBS. The example of the Ca line is seen in the upper panel of Fig. 14, which shows the region around the Ca I $\lambda5260$ Å line. One can see that the right wing of the line is blended, even in the metal-poor star HD 22879. This blend corresponds to a blend of Si I and Mn I.

The lower panel of Fig. 14 shows another Ca I line ($\lambda6455$ Å), which is placed at a less crowded region and does not show strong blends on its wings, except for the left wing of 61 Cyg A.

Although synthesis methods can cope better with blends by disregarding the blended regions and synthesising only part of the lines, this is not always done as these blends need to be previously identified, with the amount of blend varying from star to star. The EW methods suffer more evidently from this effect as the total area covered by the line is larger than the actual area filled by the corresponding element, yielding larger abundances (Stetson & Pancino 2008).

In this test, we attempt, with our methods, to estimate typical uncertainties in abundance of a given element due to different blend levels in different lines. To do so, the abundances of Ca were determined twice as follows:

1. each method analysed the Ca I line $\lambda5260$ Å;
2. each method analysed the Ca I line $\lambda6455$ Å.

The atomic data for the Ca I line $\lambda6455$ Å are from the same source as for the Ca I $\lambda5260$ Å line (Smith & Raggett 1981).

In Fig. 15, we illustrate the results of this test, where we plot the difference between the results of each method and the value obtained for the line in Paper IV as reference. Panel A shows the results obtained for the $\lambda5260$ Å line and Panel B the results obtained for the $\lambda6455$ Å line. In Panel C we show the difference in the abundances obtained from each line.

We do not find a significant difference in the node-to-node scatter for the more blended $\lambda5260$ Å line with respect to the less blended $\lambda6455$ Å line. The larger scatter among methods is obtained for cooler stars for both lines. The differences obtained between both lines are in general very small for all methods and stars except for the three EW methods for 61 Cyg A with a difference of the order of 0.15 dex. This suggests that abundances derived from EWs are only affected in the most extreme case of blending (see Fig. 14), while the synthesis method is robust with
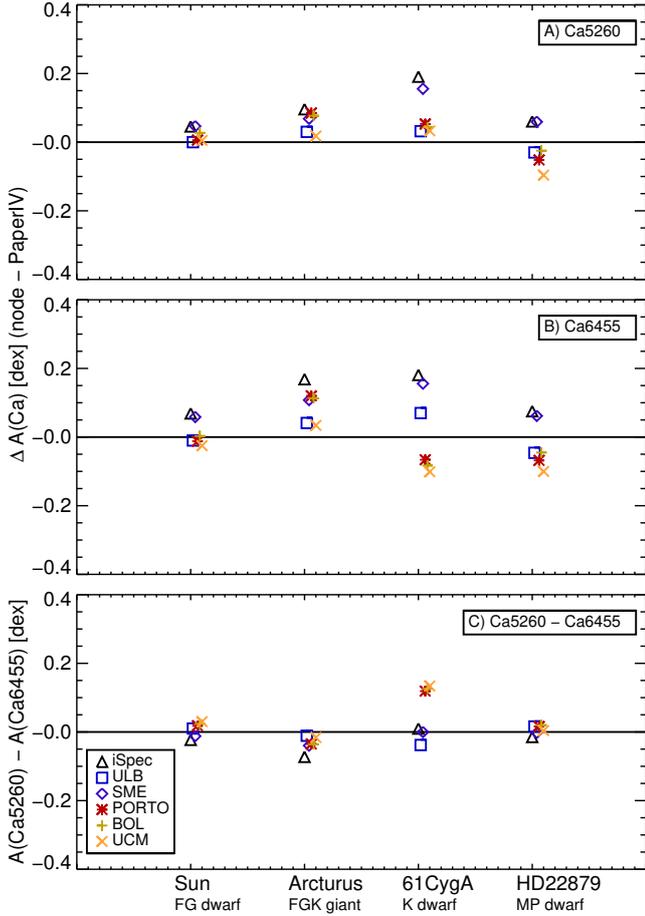
Fig. 15. Panels **A)** and **B)**: differences in Ca abundances with respect to a reference value (Paper IV) obtained from two lines with different blend levels. Panel **A)** shows the results of the more blended line. Panel **C)** shows the differences between the abundances from both lines for each node.
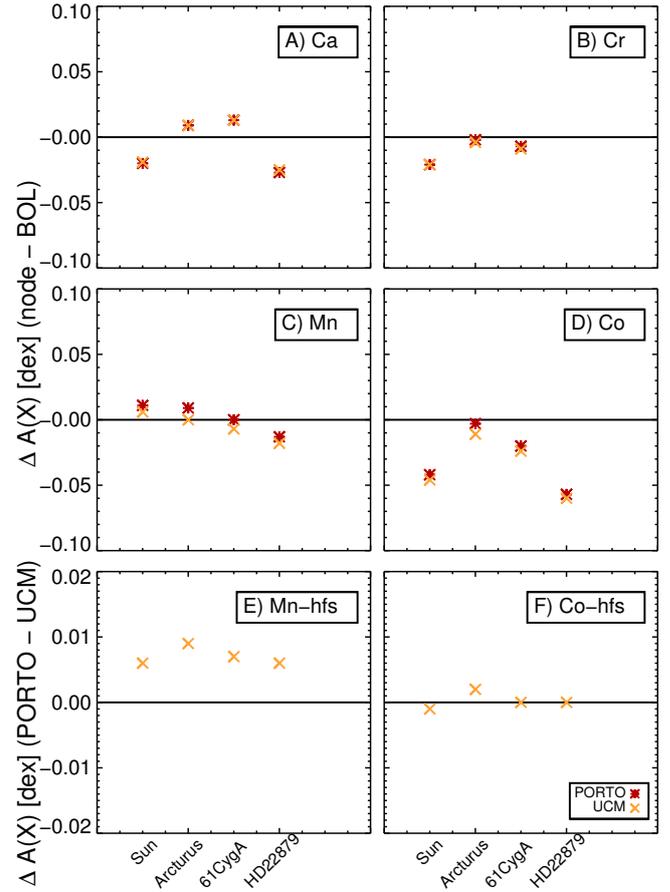

Fig. 16. Abundances obtained for same EW and different radiative transfer codes. Only methods using EWs (BOL, UCM and PORTO) are taken into account. These methods use the radiative transfer codes SYN-THE (BOL) and MOOG (UCM and PORTO).

respect to blends in all cases. On the other hand, the method-to-method scatter in absolute abundances for 61 Cyg A is actually smaller for the more blended line (Fig. 15, Panels A and B), indicating that the effect of the blends is working in the direction opposite to other differences between methods.

Conclusion for blends. From this test, we conclude that synthesis and EW methods result in similar abundances for lines with moderate blend levels. The Ca I λ5260 Å line may serve as an example for the limiting case. The differences obtained from method to method are therefore dominated by other uncertainties, such as those discussed throughout this paper. It is worth noting that this conclusion is obtained from this test only, applied to these two lines in the four stars mentioned. See also Adibekyan et al. (2015) for further discussions on this matter.

4.8. Radiative transfer codes

In this work, we employ six different methods that can be divided into two different fundamental methods: equivalent widths and syntheses. The differences among EW methods could be due to a difference in the measured EW, either by continuum placement, or by the line fitting procedure, among others. The differences among syntheses could also be due to different values adopted for the instrumental resolution, the macroturbulence parameter $v_{\rm mac}$ parameter or the specific spectral region to synthesise. If

the aforementioned quantities are fixed then one can explore the differences obtained in abundances due to the radiative transfer codes, which is what we explore in this section. For that, we separate the analysis into these two fundamental types, namely EWs and syntheses.

4.8.1. EWs

In this section, we explore the effects of using different radiative transfer codes when the same EW is assumed. More specifically, we compare the results obtained by BOL, UCM and PORTO. The radiative transfer codes to compare are SYNTHE as used by BOL, which calls this code from GALA (Mucciarelli et al. 2013); and MOOG as used by UCM with the code StePar (Tabernero et al. 2012) or by PORTO, as described in Paper III and Paper IV.

For this comparison, we fixed the EWs of BOL, which have been used for all the tests of this paper except those presented in Sect. 4.2 for the continuum normalisation test. In Fig. 16, we show the differences obtained in the abundances of all elements between BOL and the other two EW methods. In other words, the figure shows the difference between SYNTHE and two versions using MOOG. For the bottom Panels E and F, we show the differences between PORTO and UCM, to see the difference under HFS.

In general, we can see that the abundances derived by UCM and PORTO are more similar to each other than when compared to the abundances with BOL. This is not surprising as UCM and
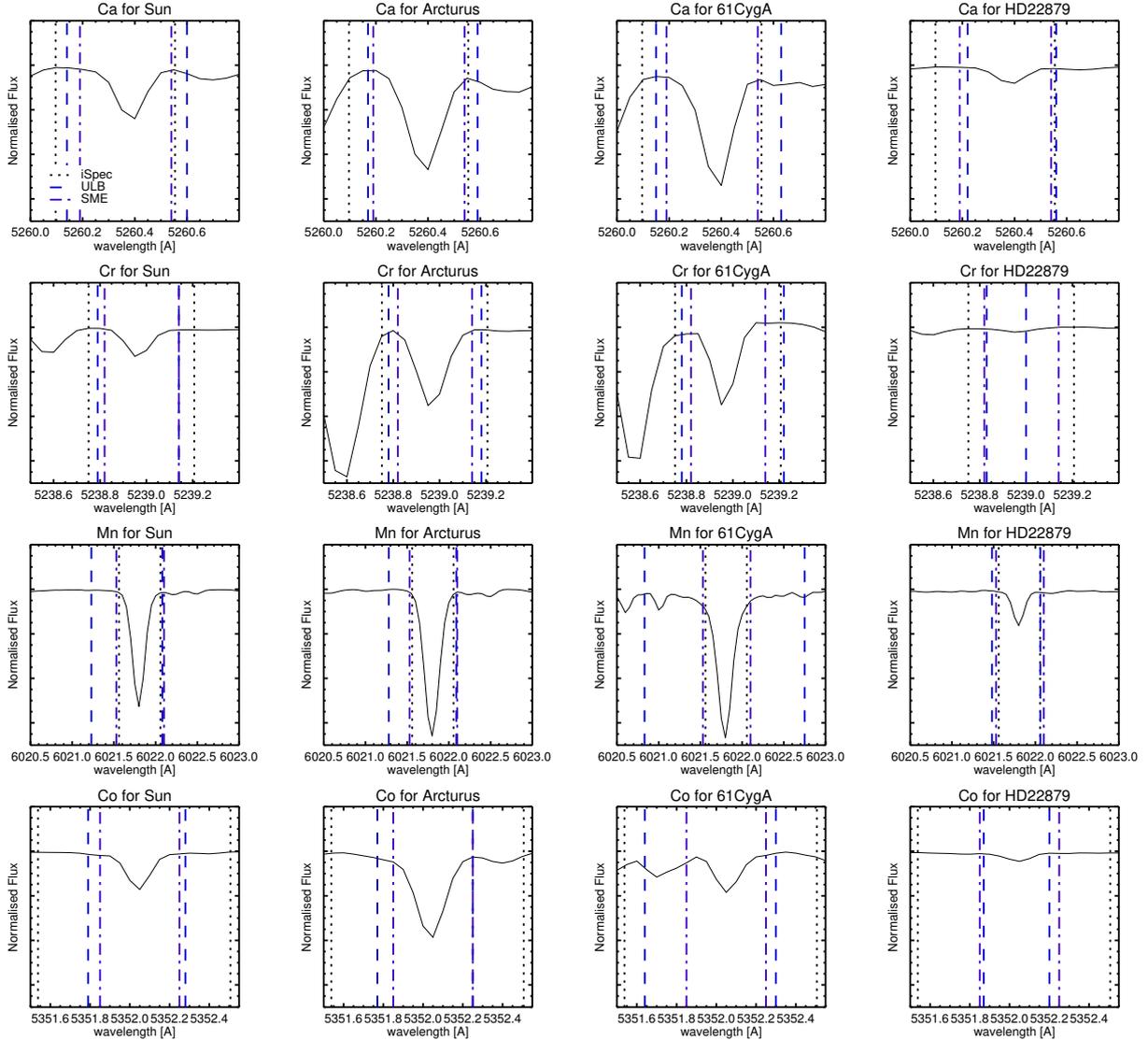
Fig. 17. Line regions employed by the three synthesis methods considered in this work are enclosed by the vertical lines of the corresponding colour. The line profile of each star and element is also shown for reference.

PORTO use the same radiative transfer code, MOOG. For Ca and Cr (Panels A and B of Fig. 16), the differences between UCM and PORTO are negligible for most of the stars, while the differences of them with respect to BOL are less than 0.05 dex. For Mn and Co without HFS (Panels C and D of Fig. 16) the differences between BOL and UCM/PORTO can reach up to 0.06 dex, as in the case of Co for the metal-poor star. Even when the same radiative transfer code (MOOG) is used, relative differences of 0.01 dex appear in the case of Arcturus for Mn and Co with or without HFS (Panels C to F). This hints towards other differences in how the codes are performing the fitting by the different methods rather than the radiative transfer code itself.

We finally note that the weakest lines (those of the Sun and HD 22879) have the largest difference between SYNTHE and MOOG results. This can be explained by the high sensitivity of the line strength of weak lines to abundance, that is, the steep slope of the curve of growth at the weak-line end.

4.8.2. Syntheses

In this test, we attempted, with our methods, to estimate typical uncertainties caused by using different radiative transfer codes

when synthesis methods are considered. We compare the results obtained with the methods ULB, SME and iSpec, which use the radiative transfer codes Turbospectrum (Plez 2012), sme_synth (Valenti & Piskunov 1996) and SPECTRUM (Gray & Corbally 1994), respectively. An equivalent to fix the EW of the previous section in this case would mean to fix the spectral region (mask) where the fitting is performed for the abundance determination. In addition, each method performs a convolution to model different broadening effects in a line profile. In this section, we explore these two parameters.

We first compare the masks (regions) considered by the three methods in each case. This is shown in Fig. 17, where in each panel, we plot the line profile of a different star and element. The masks are represented as vertical lines with different colours and lines for each method.

In general, the regions agree very well within the three methods, with a few exceptions such as the masks for Co of iSpec or the Mn mask for 61 Cyg A of ULB. While iSpec employs the same mask for a given element, ULB and SME adjust the mask according to the properties of the line profile of each star individually. This procedure is also possible for iSpec in other applications (Blanco-Cuaresma et al. 2015), but was not implemented
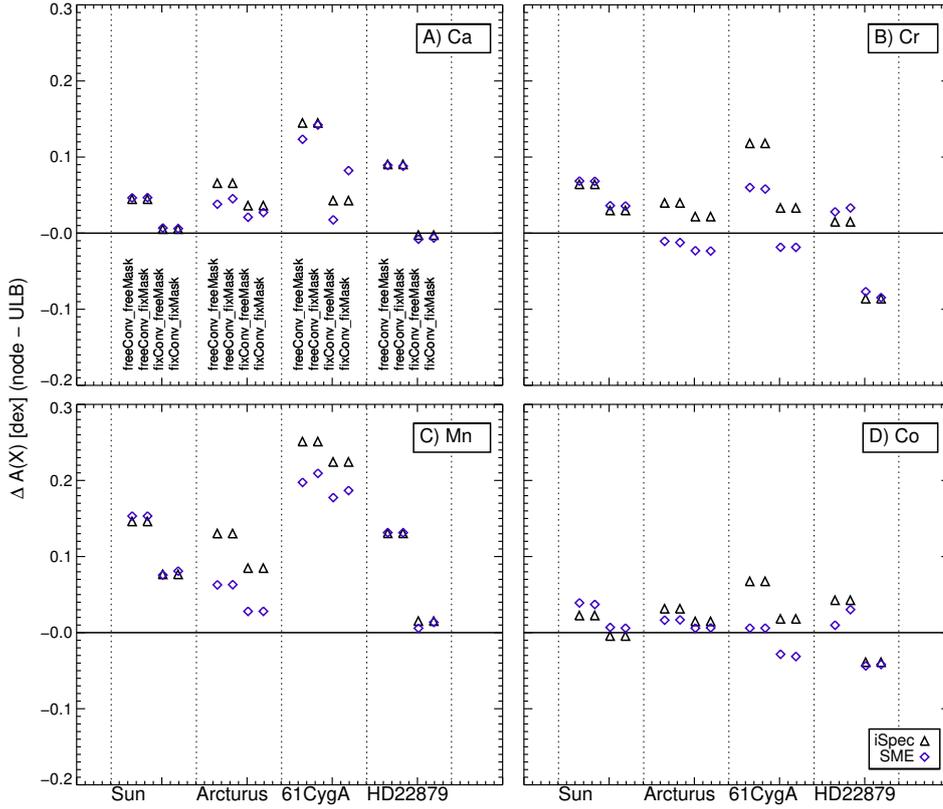
Fig. 18. Abundances obtained for identical masks/convolutions but different synthesis codes. Only methods using synthesis (ULB, SME and iSpec) are taken into account. These methods use the radiative transfer codes Turbospectrum (ULB), sme_synth (SME) and SPECTRUM (iSpec).

in this work. The criteria for adjusting masks is clearly different between SME and ULB, sometimes obtaining significantly different regions for performing the fitting, such as Mn of 61 Cyg A or the weak Cr line of HD 22879. We note that the methods have independently developed a procedure to choose masks, but the final outcome is in most cases very similar, that is, less than 1 Å difference.

Regarding broadening parameters, these account for a better reproduction of the instrumental profile, which is usually treated as a Gaussian smoothing of the synthetic spectrum to match the resolution of the observed spectrum. Line broadening can also be caused by stellar rotation, macroturbulence and other effects. Unlike microturbulence, these broadening mechanisms do not change the strength of a spectral line. In iSpec and SME, rotational broadening and macroturbulent broadening are applied after the synthetic spectrum has been convolved with the appropriate line-spread functions given by the instrumental resolution.

While SME and iSpec need the three parameters (v_{mac}, projected equatorial velocity $v\sin i$ and the resolution R) as independent input for the computation of synthetic spectra, ULB employs only one convolution parameter. While SME and iSpec considered $R = 70\,000$, and $v\sin i$ and v_{mac} as listed in Table 1, ULB considered an overall convolution parameter of 5.8, 8.0, 4.6 and 5.9 km s^{-1} for the Sun, Arcturus, 61 Cyg A and HD 22879, respectively.

To study the effect of both broadening and mask choice, in the final results we determined the abundances four times as follows:

1. each method determined abundances with the same convolution and fixed mask ($fixConv\_fixMask$);
2. each method determined abundances with the same convolution and free mask ($fixConv\_freeMask$);
3. each method determined abundances with a different convolution and fixed mask ($freeConv\_fixMask$);

4. each method determined abundances with different convolution and free mask ($freeConv\_freeMask$).

Here, HFS was considered for the analysis of the Mn and Co lines. The masks and the convolution parameters were always fixed by ULB, and therefore our analysis consisted of comparing the results of ULB with SME and iSpec. Thus, for the fixed tests, v_{mac} and $v\sin i$ were considered to be zero by iSpec and SME, and the resolution was considered to correspond to the convolution parameter of ULB, which corresponds to $R = c/v_{ULB} = 51\,688, 37\,474, 65\,172$ and $50\,812$ for the Sun, Arcturus, 61 Cyg A and HD 22879, respectively. When convolution or the mask were free, it meant that SME and iSpec used their own parameters.

In Fig. 18, we show the results of this test. In each panel we show one element, and each star is separated by vertical dashed lines. The four tests are indicated for each star in each panel, following the same order as in Panel A. We compared the results of SME and iSpec with respect to ULB in each case.

When comparing the $freeMask$ with the $fixMask$ cases, we see that iSpec obtains essentially the same results. The same happens for most of the masks adopted by SME, with notable exceptions being the Ca and Mn masks for 61 Cyg A, which differ more significantly (see Fig. 17). This test is telling us that the methods are very robust in the abundance measurement when different masks are employed. A notable case is Co for iSpec, presenting negligible differences when two very different masks are considered. This might be due to the fact that results will not be affected if regions not sensitive to the abundance are considered in the fit.

A more notable difference is obtained when the broadening parameters (convolution) differ. Both SME and iSpec show very similar differences to ULB, but we remind the reader that both methods use the same parameters. There is a general tendency for decreasing the final abundance when the ULB convolution

Table 6. Summary of tests and uncertainties.

Test	Expected σ dex	Max σ dex	Min σ dex	Comment
1. Line list	0.05	0.6	0.0	EWs are not affected by position of core of line but number of HFS components might be important
2. Continuum	0.3	0.6	0.05	Absolute abundance is very dependent but methods should not re-normalise
3. HFS	0.08	0.4	0.0	Differences in abundances between HFS:N and HFS:Y are similar for all methods
4. v_{mic}	0.2	1.2	0.01	Maximum difference corresponds to 1 km s^{-1} range in v_{mic}, dependency decreases when HFS is considered.
5. α-enhancement	0.02	0.1	0.001	Cool stars are more affected than metal-poor warm stars.
6. Atmosphere model interpolation	0.01	0.04	0.002	Obtained from models with 1% difference in temperature and 5–10% difference in gas pressure.
7. Blends	0.02	0.1	0.0	Maximum difference is found for EW methods for cool dwarf star.
8.1. Same EWs	0.02	0.06	0.001	Weak lines are more affected.
8.2. Same syntheses	0.07	0.12	0.001	Convolution is more important than the choice of mask.

Notes. For each test, we list the typical (median) uncertainty obtained from all results in a given test (expected σ), as well as the maximum difference (Max σ) and the minimum difference (Min σ).

parameter is taken into account. This suggests that the overall convolution of ULB has a larger value than the total convolution of the other two methods. A broader line will have a weaker core, imitating a lower abundance. This is because the core of a weak line (lines with EWs of approx. 120 mÅ (that is, a reduced EW of approx. $\log(\mathrm{EW}/\lambda) = -4.8$ in the optical, see Paper III) or less) is usually the region with the largest sensitivity to the abundance and where most of the data points to perform the fitting lie. We see that in both cases, the increase in broadening translates to a decrease in abundance. This input difference drives differences of the order of 0.05 dex in abundances for most of the cases, but could reach more than 0.1 dex in some cases, such as Mn for HD 22879.

Finally, we see that even adopting the same masks and convolution parameters (rightmost points for each star in Fig. 18), the differences between ULB and SME or iSpec/ are not fully alleviated. While in some cases this indeed happened, as in the abundance of Ca for the Sun and HD 22879, this did not happen for the same stars for Cr, for which the differences increased. We might attribute this uncertainty to an intrinsic difference between Turbospectrum, SPECTRUM and sme_synth, or to some other effect such as the fitting procedure. For SME and iSpec, syntheses of a given line are computed on-the-fly for different abundance values and a χ^2 is calculated between the synthetic and observed profile. If the χ^2 is higher than the desired convergency parameter, new syntheses are computed considering abundance values towards the direction in which χ^2 was better. This process is repeated until convergency is achieved. The fitting procedure of ULB is slightly different. Only five syntheses are computed with different abundance values and a χ^2 is calculated for each of them. A likelihood function is then obtained and the solution is found via interpolation of that function.

To check if this different approach produces discrepancies of up to 0.1 dex, one would have to modify the fitting procedures of the codes, which goes beyond the scope of this work.

Conclusion for radiative transfer codes.   When the same EWs are used, differences of up to 0.05 dex can be obtained when different radiative transfer codes are used. For lines affected by HFS, differences are more notable, even if the same radiative transfer code is employed. We conclude that the treatment of HFS is an important driver of differences between same radiative transfer codes, which can reach 0.18 dex for Mn.

The choice of mask had, in most of the cases, very little impact on the determination of abundances, even if the masks differed significantly. On the other hand, the convolution parameters can produce differences of up to 0.1 dex.

5. Summary and concluding remarks

The goal of this work was to explore the impact on the derived stellar abundances produced by some of the technical assumptions of different methods commonly used in the field. For this, participants with expertise in the methods employed in our previous works on spectral analysis of GBS (Papers III and IV) were considered. In this work we discussed the differences in abundances obtained by six methods, three methods based on equivalent widths (BOL, UCM and PORTO) and three methods based on syntheses (ULB, SME and iSpec).

We analysed four GBS, each of them being a reference star of a group of GBS as defined in Jofré et al. (2015b): the Sun representing *FG*-dwarfs, Arcturus representing *FGK*-giants, 61 Cyg A representing one of the *K*-dwarfs, and HD 22879 representing metal-poor stars. For each star, we determined the abundance of four elements, using one line of Ca, Cr, Mn and Co for each. These lines were selected from Jofré et al. (2015b) because they were analysed by several methods for these stars.

By fixing the stellar parameters and using the same observed spectrum, atmosphere model and line list, we performed eight different tests with the aim of exploring potential sources of systematic uncertainties in the derived abundances. These are non-exhaustive and further tests should be explored in future works on this matter. A summary of these tests with the main conclusions is given below and in Table 6. In the table, we list each test and give a quote of the typical uncertainty considering all methods and stars, as well as the minimum and maximum uncertainty.

These uncertainties are the median, minimum and maximum difference of all results analysed in the given test.

- *Linelist*: we investigated the impact on deriving abundances of Mn using two line lists, one of them having the core of the line slightly shifted with respect to the observations. Both line lists also had different binning for HFS. We found that when HFS was not taken into account, the EW methods showed no difference, while the synthesis method showed a difference of up to 0.6 dex. When HFS was taken into account, typical differences of 0.05 dex were found in the case of syntheses, while a difference of less than 0.01 dex was found for the case of EW methods.
- *Continuum normalisation*: we investigated the impact of considering different approaches for normalising the spectra of all elements on deriving abundances. We took two spectra normalised by two methods and determined abundances by fixing the continuum or allowing the methods to re-normalise. We found that the method-to-method scatter decreased by up to 0.5 dex when the continuum was kept fixed. We also found offsets of 0.5 dex (Co for 61 Cyg A) in the final absolute abundances. We conclude that the continuum placement is key for an absolute result of the retrieved abundances, and that when combining different methods, the agreement improves if the continuum is kept fixed.
- *Hyperfine structure*: we compared the results obtained for the abundances of Mn and Co considering and neglecting HFS. We found that differences of up to 0.4 dex can be obtained if HFS of Mn is not considered while a difference of 0.1 dex is obtained if the HFS of Co is not considered. While such effects have been reported in the past, HFS is still often neglected in modern spectroscopic studies.
- *Microturbulence*: we derived abundances changing the broadening parameter v_{mic}. We found that all methods have the same dependency of the derived abundances on v_{mic}, in which the strongest lines have the strongest dependencies. In the most extreme case, the final abundances can differ by 1.2 dex when v_{mic} is different by a value of 1 km s^{-1}. Typical differences of 0.2 km s^{-1} correspond to differences of 0.1 dex in abundance for strong lines ($\log \mathrm{EW}/\lambda \sim -4.8$) and of 0.05 dex for weaker lines.
- *α-element abundances and continuum opacities*: differences of less than 0.1 dex are obtained whether the abundances with the corresponding α-enhancement of the atmosphere model are scaled or not. Cool stars are more affected than metal-poor stars, reflecting how α-abundances affect the overall continuum opacities.
- *Atmospheric model interpolation*: we compared the interpolated atmosphere models by the different methods and found differences of up to 10% in the gas pressure at $\tau_{\mathrm{ross}} = 1$ for 61 Cyg A. We derived abundances of all elements taking two different interpolated models, and found that these differences translate to differences in the abundances of less than 0.02 dex in most cases. These differences depend on the method employed.
- *Blends*: we investigated the impact of abundances derived from two Ca lines that have different levels of small blends. We found no significant differences in the final abundances.
- *Radiative transfer codes*: we investigated the differences of radiative transfer codes by (1) fixing the EWs and (2) fixing the masks and convolution parameters when synthesis methods are used to derive abundances. We found maximal differences of 0.06 dex for weak lines in the first case and

maximal differences of 0.12 dex for cool stars in the second case.

Based on our analysis, we argue that if independent pipelines are combined to provide a unique set of final results of abundances, then a fixed value for the continuum flux, as well as for the v_{mic} parameter, should be adopted. The derivation of these input data must be properly documented and provided to the community for better reproducibility of results. Furthermore, since it is common that observed and theoretical lines can have slight differences in the central wavelength, synthesis methods should take this into account by shifting the observed lines to a local rest frame.

Thanks to the tests performed for this workshop we were able to improve two technical aspects in the spectral analysis of the GBS and the GES: we improved the line Mn I λ6021 Å in the GES line list, which will be adopted in its future version. In addition, the GES/GBS methods `UCM` and `PORTO` improved with the new functionality of treating HFS, which will be adopted in future GES data releases. In this work, we provide quantitative estimates of uncertainties in elemental abundances due to typical differences in technical assumptions in spectrum modelling. These may serve as a guideline when estimating uncertainties for pipelines based on a single method.

The *Gaia*-ESO survey homogenises the results of work by multiple analysis teams, in which these types of issues have been startlingly revealed. Surveys currently under development, taking into account the lessons learnt from *Gaia*-ESO, are inclined to preferentially use the results from only one analysis pipeline (see overview of e.g. Allende Prieto 2016), thereby sidestepping the issues explored here. However, it is clear that combining surveys with the much anticipated *Gaia* standard will yield the most comprehensive database of kinematic and chemical measurements of the Milky Way stellar populations to-date (Feltzing 2015; Bland-Hawthorn & Gerhard 2016). It is at this inter-survey level that the effects discribed here will return in full force, and therefore it is essential that we work to understand and resolve such issues now, or risk rendering these impressive datasets eternally and unsatisfactorily disconnected.

References

Adibekyan, V., Figueira, P., Santos, N. C., et al. 2015, A&A, 583, A94
Adibekyan, V. Z., Delgado Mena, E., Sousa, S. G., et al. 2012a, A&A, 547, A36
Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012b, A&A, 545, A32

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17
Allende Prieto, C. 2016, Astron. Nachr., 337, 837
Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, ApJ, 636, 804
Barklem, P. S., Piskunov, N., & O'Mara, B. J. 2000, A&AS, 142, 467 (BPM)
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Blackwell-Whitehead, R. J., Pickering, J. C., Pearse, O., & Nave, G. 2005, ApJS, 157, 402
Blanco-Cuaresma, S., Soubiran, C., Heiter, U., & Jofré, P. 2014a, A&A, 569, A111
Blanco-Cuaresma, S., Soubiran, C., Jofré, P., & Heiter, U. 2014b, A&A, 566, A98
Blanco-Cuaresma, S., Soubiran, C., Heiter, U., et al. 2015, A&A, 577, A47
Blanco-Cuaresma, S., Nordlander, T., Heiter, U., et al. 2016, The 19th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, Uppsala, Sweden, 22
Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529
Boeche, C., & Grebel, E. K. 2016, A&A, 587, A2
Boyarchuk, A. A., Antipova, L. I., & Pakhomov, Y. V. 2008, Astron. Rep., 52, 630
Brodzinski, T., Kronfeldt, H.-D., Kropp, J.-R., & Winkler, R. 1987, Zeitschrift fur Physik D Atoms Molecules Clusters, 7, 161
Cantat-Gaudin, T., Donati, P., Pancino, E., et al. 2014, A&A, 562, A10
Cardon, B. L., Smith, P. L., Scalo, J. M., Testerman, L., & Whaling, W. 1982, ApJ, 260, 395
Casimir, H. B. G. 1936, in On the Interaction between Atomic Nuclei and Electrons (Springer), 201
Condon, E., & Shortley, G. 1935, The Theory of Atomic Spectra (CUP)
Dalton, G., Trager, S. C., Abrams, D. C., et al. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, Proc. SPIE, 8446, 84460
de Jong, R. S., Bellido-Tirado, O., Chiappini, C., et al. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, Proc. SPIE, 8446, 84460
De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, MNRAS, 449, 2604
del Peloso, E. F., Cunha, K., da Silva, L., & Porto de Mello, G. F. 2005, A&A, 441, 1149
Den Hartog, E. A., Lawler, J. E., Sobeck, J. S., Sneden, C., & Cowan, J. J. 2011, ApJS, 194, 35 (DLSSC)
Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101
Feltzing, S. 2015, ArXiv e-prints [arXiv:1506.08642]
Feltzing, S., Fohlman, M., & Bensby, T. 2007, A&A, 467, 665
García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, AJ, 151, 144
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
Gray, R. O., & Corbally, C. J. 1994, AJ, 107, 742
Grevesse, N., Asplund, M., & Sauval, A. J. 2007, Space Sci. Rev., 130, 105
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Handrich, E., Steudel, A., & Walther, H. 1969, Phys. Lett. A, 29, 486
Hawkins, K., Jofré, P., Gilmore, G., & Masseron, T. 2014, MNRAS, 445, 2575
Hawkins, K., Jofré, P., Heiter, U., et al. 2016a, A&A, 592, A70
Hawkins, K., Masseron, T., Jofré, P., et al. 2016b, A&A, 594, A43
Heiter, U., Jofré, P., Gustafsson, B., et al. 2015a, A&A, 582, A49
Heiter, U., Lind, K., Asplund, M., et al. 2015b, Phys. Scr, 90, 054010
Hinkel, N. R., Timmes, F. X., Young, P. A., Pagano, M. D., & Turnbull, M. C. 2014, AJ, 148, 54
Hinkel, N. R., Young, P. A., Pagano, M. D., et al. 2016, ApJS, 226, 4
Hinkel, K., & Wallace, L. 2005, in Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis, eds. T. G. Barnes, III, & F. N. Bash, ASP Conf. Ser., 336, 321
Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2017, ApJ, 836, 5
Husser, T.-O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6
Jofré, P. 2016, Astron. Nachr., 337, 859

Jofré, P., Panter, B., Hansen, C. J., & Weiss, A. 2010, A&A, 517, A57
Jofré, P., Heiter, U., Soubiran, C., et al. 2014, A&A, 564, A133
Jofré, E., Petrucci, R., Saffe, C., et al. 2015a, A&A, 574, A50
Jofré, P., Heiter, U., Soubiran, C., et al. 2015b, A&A, 582, A81
Jofré, P., Mädler, T., Gilmore, G., et al. 2015c, MNRAS, 453, 1428
Kang, W., & Lee, S.-G. 2012, MNRAS, 425, 3162
Katz, D., Soubiran, C., Cayrel, R., Adda, M., & Cautain, R. 1998, A&A, 338, 151
Koleva, M., Prugniel, P., Bouchard, A., & Wu, Y. 2009, A&A, 501, 1269
Kunder, A., Kordopatis, G., Steinmetz, M., et al. 2017, AJ, 153, 75
Kurucz, R. L. 1979, ApJS, 40, 1
Kurucz, R. L. 2005, Mem. Soc. Astron. It. Suppl., 8, 76
Kurucz, R. L. 2007, Robert L. Kurucz on-line database of observed and predicted atomic transitions, http://kurucz.harvard.edu/atoms
Kurucz, R. L. 2008, Robert L. Kurucz on-line database of observed and predicted atomic transitions, http://kurucz.harvard.edu/atoms
Lebzelter, T., Heiter, U., Abia, C., et al. 2012, A&A, 547, A108
Magic, Z., Collet, R., Asplund, M., et al. 2013, A&A, 557, A26
Masseron, T. 2006, Ph.D. Thesis, Observatoire de Paris, France
Masseron, T., Merle, T., & Hawkins, K. 2016, Astrophysics Source Code Library [record ascl:1605.004]
Molaro, P., & Monai, S. 2012, A&A, 544, A125
Mucciarelli, A., Pancino, E., Lovisi, L., Ferraro, F. R., & Lapenna, E. 2013, ApJ, 766, 78
Neves, V., Santos, N. C., Sousa, S. G., Correia, A. C. M., & Israelian, G. 2009, A&A, 497, 563
Nissen, P. E. 2015, A&A, 579, A52
Nissen, P. E., & Schuster, W. J. 2010, A&A, 511, L10
Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, ARA&A, 51, 457
North, P., Cescutti, G., Jablonka, P., et al. 2012, A&A, 541, A45
Pickering, J. C. 1996, ApJS, 107, 811
Piskunov, N., & Valenti, J. A. 2016, A&A, 597, A16
Plez, B. 2012, Astrophysics Source Code Library [record ascl:1205.004]
Ramírez, I., Allende Prieto, C., & Lambert, D. L. 2013, ApJ, 764, 78
Randich, S., Gilmore, G., & Gaia-ESO Consortium. 2013, The Messenger, 154, 47
Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, MNRAS, 370, 141
Schönrich, R., & Bergemann, M. 2014, MNRAS, 443, 698
Schwartz, C. 1955, Phys. Rev., 97, 380
Shetrone, M., Bizyaev, D., Lawler, J. E., et al. 2015, ApJS, 221, 24
Smiljanic, R., Korn, A. J., Bergemann, M., et al. 2014, A&A, 570, A122
Smith, G., & Raggett, D. S. J. 1981, J. Phys. B At. Mol. Phys., 14, 4015
Sneden, C. A. 1973, Ph.D. Thesis, The University of Texas at Austin, USA
Sobeck, J. S., Lawler, J. E., & Sneden, C. 2007, ApJ, 667, 1267 (SLS)
Sousa, S. G. 2014, in Determination of Atmospheric Parameters of B-, A-, F- and G-type Stars, eds. E. Niemczura, B. Smalley, & W. Pych (Springer Int. Publ.), 297
Sousa, S. G., Santos, N. C., Adibekyan, V., Delgado-Mena, E., & Israelian, G. 2015, A&A, 577, A67
Spina, L., Meléndez, J., Karakas, A. I., et al. 2016, A&A, 593, A125
Stetson, P. B., & Pancino, E. 2008, PASP, 120, 1332
Tabernero, H. M., Montes, D., & González Hernández, J. I. 2012, A&A, 547, A13
Teixeira, G. D. C., Sousa, S. G., Tsantaki, M., et al. 2016, A&A, 595, A15
Tucci Maia, M., Ramírez, I., Meléndez, J., et al. 2016, A&A, 590, A32
Valenti, J. A., & Fischer, D. A. 2005, ApJS, 159, 141
Valenti, J. A., & Piskunov, N. 1996, A&AS, 118, 595
Wallace, L., & Hinkle, K. 2009, ApJ, 700, 720
Worley, C. C., de Laverny, P., Recio-Blanco, A., Hill, V., & Bijaoui, A. 2016, A&A, 591, A81