A&A 570, A122 (2014)
DOI: 10.1051/0004-6361/201423937
© ESO 2014

Astronomy & Astrophysics

The *Gaia*-ESO Survey: The analysis of high-resolution UVES spectra of FGK-type stars⋆,⋆⋆

R. Smiljanic[1,2], A. J. Korn[3], M. Bergemann[4,5], A. Frasca[6], L. Magrini[7], T. Masseron[8], E. Pancino[9,10], G. Ruchti[11], I. San Roman[12], L. Sbordone[13,14,15], S. G. Sousa[16,17], H. Tabernero[18], G. Tautvaišienė[19], M. Valentini[20], M. Weber[20], C. C. Worley[21,5], V. Zh. Adibekyan[16], C. Allende Prieto[22,23], G. Barisevičius[19], K. Biazzo[6], S. Blanco-Cuaresma[24], P. Bonifacio[25], A. Bragaglia[9], E. Caffau[13,25], T. Cantat-Gaudin[26,27], Y. Chorniy[19], P. de Laverny[21], E. Delgado-Mena[16], P. Donati[9,28], S. Duffau[13,14,15], E. Franciosini[7], E. Friel[29], D. Geisler[12], J. I. González Hernández[18,22,23], P. Gruyters[3], G. Guiglion[21], C. J. Hansen[13], U. Heiter[3], V. Hill[21], H. R. Jacobson[30], P. Jofre[24,5], H. Jönsson[11], A. C. Lanzafame[6,31], C. Lardo[9], H.-G. Ludwig[13], E. Maiorca[7], Š. Mikolaitis[19,21], D. Montes[18], T. Morel[32], A. Mucciarelli[28], C. Muñoz[12], T. Nordlander[3], L. Pasquini[1], E. Puzeras[19], A. Recio-Blanco[21], N. Ryde[11], G. Sacco[7], N. C. Santos[16,17], A. M. Serenelli[33], R. Sordo[26], C. Soubiran[24], L. Spina[7,34], M. Steffen[20], A. Vallenari[26], S. Van Eck[8], S. Villanova[12], G. Gilmore[5], S. Randich[7], M. Asplund[35], J. Binney[36], J. Drew[37], S. Feltzing[11], A. Ferguson[38], R. Jeffries[39], G. Micela[40], I. Negueruela[41], T. Prusti[42], H-W. Rix[43], E. Alfaro[44], C. Babusiaux[25], T. Bensby[11], R. Blomme[45], E. Flaccomio[40], P. François[25], M. Irwin[5], S. Koposov[5], N. Walton[5], A. Bayo[43,46], G. Carraro[47], M. T. Costado[44], F. Damiani[30], B. Edvardsson[3], A. Hourihane[5], R. Jackson[39], J. Lewis[5], K. Lind[5], G. Marconi[47], C. Martayan[47], L. Monaco[47], L. Morbidelli[7], L. Prisinzano[40], and S. Zaggia[26]

(Affiliations can be found after the references)

Received 3 April 2014/ Accepted 1 September 2014

ABSTRACT

Context. The ongoing *Gaia*-ESO Public Spectroscopic Survey is using FLAMES at the VLT to obtain high-quality medium-resolution Giraffe spectra for about 10^5 stars and high-resolution UVES spectra for about 5000 stars. With UVES, the Survey has already observed 1447 FGK-type stars.

Aims. These UVES spectra are analyzed in parallel by several state-of-the-art methodologies. Our aim is to present how these analyses were implemented, to discuss their results, and to describe how a final recommended parameter scale is defined. We also discuss the precision (method-to-method dispersion) and accuracy (biases with respect to the reference values) of the final parameters. These results are part of the *Gaia*-ESO second internal release and will be part of its first public release of advanced data products.

Methods. The final parameter scale is tied to the scale defined by the *Gaia* benchmark stars, a set of stars with fundamental atmospheric parameters. In addition, a set of open and globular clusters is used to evaluate the physical soundness of the results. Each of the implemented methodologies is judged against the benchmark stars to define weights in three different regions of the parameter space. The final recommended results are the weighted medians of those from the individual methods.

Results. The recommended results successfully reproduce the atmospheric parameters of the benchmark stars and the expected $T_{\rm eff}$-log g relation of the calibrating clusters. Atmospheric parameters and abundances have been determined for 1301 FGK-type stars observed with UVES. The median of the method-to-method dispersion of the atmospheric parameters is 55 K for $T_{\rm eff}$, 0.13 dex for log g and 0.07 dex for [Fe/H]. Systematic biases are estimated to be between 50−100 K for $T_{\rm eff}$, 0.10−0.25 dex for log g and 0.05−0.10 dex for [Fe/H]. Abundances for 24 elements were derived: C, N, O, Na, Mg, Al, Si, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Mo, Ba, Nd, and Eu. The typical method-to-method dispersion of the abundances varies between 0.10 and 0.20 dex.

Conclusions. The *Gaia*-ESO sample of high-resolution spectra of FGK-type stars will be among the largest of its kind analyzed in a homogeneous way. The extensive list of elemental abundances derived in these stars will enable significant advances in the areas of stellar evolution and Milky Way formation and evolution.

Key words. methods: data analysis – surveys – stars: abundances – stars: fundamental parameters – stars: late-type

1. Introduction

Following the seminal paper of Burbidge et al. (1957), it is now well established that the vast majority of chemical elements are produced inside stars. These elements and their isotopes are synthesized by various processes in stars of different masses and of different generations (Wallerstein et al. 1997, for a review).

Modern astrophysics strives to trace the processes of synthesis and dispersion of chemical elements, and uses them to decode the history of formation and evolution of planets, of stars, and of the Galaxy. Multi element abundance information is a key requirement in this context, as the abundances of distinct elements are shaped by different physical processes. The investigation of

⋆ Based on observations made with the ESO/VLT, at Paranal Observatory, under program 188.B-3002 (The *Gaia*-ESO Public Spectroscopic Survey, PIs Gilmore and Randich).

⋆⋆ Appendices are available in electronic form at http://www.aanda.org

large samples of long-lived stars, formed in different places and times in the Galaxy, is needed to put together a complete picture of Galactic and stellar evolution.

Obtaining the spectroscopic data to achieve this goal is demanding. To determine accurate, detailed elemental abundances, we need high-resolution, high signal-to-noise (S/N) spectra with broad wavelength coverage. For robust statistics and to cover all Galactic populations, the observation of large stellar samples, including faint stars beyond the solar neighborhood, is needed. To achieve this, a number of spectroscopic surveys are now being conducted and/or planned such as, for example, the Apache Point Observatory Galactic Evolution Experiment (APOGEE, Ahn et al. 2014), the GALactic Archaeology with HERMES, the High Efficiency and Resolution Multi-Element Spectrograph (GALAH, Zucker et al. 2012), the Large sky Area Multi-Object fiber Spectroscopic Telescope (LAMOST) Experiment for Galactic Understanding and Exploration (LEGUE, Deng et al. 2012), the RAdial Velocity Experiment (RAVE, Steinmetz et al. 2006; Kordopatis et al. 2013), the Sloan Extension for Galactic Understanding and Exploration (SEGUE, Yanny et al. 2009), and the *Gaia*-ESO Survey (Gilmore et al. 2012; Randich & Gilmore 2013).

The *Gaia*-ESO Survey[1] is an ambitious public spectroscopic survey that is obtaining medium- and high-resolution spectra of more than 10^5 stars. The observations started on December 31, 2011 and are carried out at the Very Large Telescope (VLT), at the Paranal Observatory, Chile. All the data collected by the Survey are homogeneously reduced and analyzed by the *Gaia*-ESO consortium. Public catalogs with astrophysical parameters will be made available to the community.

The Survey targets represent all major Galactic components (halo, bulge, thin and thick disks) and include a large number of open clusters, selected to cover the parameter space of age, total stellar mass, distance, and metallicity. The targets include early- and late-type stars (from O- to M-type), giants, dwarfs, and pre-main-sequence stars.

Observations are conducted with the FLAMES (Fiber Large Array Multi-Element Spectrograph) multi fiber facility (Pasquini et al. 2002). Medium-resolution spectra ($R \sim 20\,000$) of about $\sim 10^5$ stars are being obtained with Giraffe and high-resolution spectra ($R \sim 47\,000$) of about ~5000 stars are being obtained with UVES (Ultraviolet and Visual Echelle Spectrograph, Dekker et al. 2000).

1.1. The Gaia-ESO release papers

This paper is part of a series that presents a complete description of the *Gaia*-ESO Survey, in preparation for its first public release of advanced data products. The Survey is organized in different working groups (WGs) that deal with all the relevant tasks, from target selection and observations, to data analysis and data archiving. While it is beyond the goals of this paper to describe the Survey's internal organization, we provide an overview of the release papers for clarity and completeness.

Two papers will provide the Survey overview describing science goals, observation plan, team organization, target selection strategy, and data release schedules. The Milky Way part of the Survey will be presented in Gilmore et al. (in prep.), and the open clusters part of the Survey will be presented in Randich et al. (in prep.). The data and procedures used to select probable member stars to be observed in each selected open cluster will be presented in Bragaglia et al. (in prep.). Description of the data reduction aspects will be presented in Lewis et al. (in prep.) for the Giraffe spectra, and are described in Sacco et al. (2014) for the UVES spectra.

The analysis of different types of stars is performed by different WGs. The analysis of the Giraffe spectra of FGK-type stars will be described in Recio-Blanco et al. (in prep.). The analysis of pre-main-sequence stars will be described in Lanzafame et al. (2014). The analysis of OBA-type stars, which are all observed in young open clusters, will be described in Blomme et al. (in prep.). The analysis of non standard objects and outliers will be part of Gilmore et al. (in prep.). The analysis of the UVES spectra of FGK-type stars is the topic of the present paper.

A considerable effort is dedicated to the observation of a comprehensive set of targets for internal and external calibration of the Survey parameter scale. Calibration targets include open and globular cluster stars, *Gaia*[2] benchmark stars, and stars from CoRoT fields (Convection Rotation and planetary Transits, Baglin et al. 2006). Selection and observation of these targets will be described in Pancino et al. (in prep.). Because the Survey includes the analysis of different types of stars, additional steps are needed to homogenize the final results, correcting systematic effects where needed. This additional step is taken to ensure that the results for early- and late-type stars, for dwarfs, giants, and pre-main-sequence stars are all on a single consistent scale. This Survey wide homogenization process will be discussed in François et al. (in prep.)[3].

1.2. The UVES analysis

This paper describes the analysis of the UVES spectra of FGK-type stars in the *Gaia*-ESO Survey conducted within Working Group 11 (WG11) and as implemented for the first release of advanced data products. The products resulting from this analysis include: equivalent widths (EWs) of spectral lines, stellar atmospheric parameters, and elemental abundances.

The analysis process in the Survey is performed in cycles, following the data reduction of newly observed spectra. Each new analysis cycle improves upon the last, as some of the input data is updated (e.g., atomic and molecular data), as teams improve their analysis methods, and as the method used to define the final recommended set of atmospheric parameters and abundances evolves.

We have now completed the analysis of two internal data releases (hereafter iDR). An iDR consists of reduced data that are ready to be analyzed and is initially available only within the *Gaia*-ESO consortium. New iDRs happen roughly every six months, after which a new analysis cycle is started.

The second internal release (iDR2) included a revision in data formats of all the observations done by the Survey, superseding iDR1. There were also significant differences between the analysis strategy applied to the iDR1 and iDR2 data sets. The discussion presented here will concentrate on the analysis of iDR2. These are the results that will be part of the first *Gaia*-ESO public release, together with the results of iDR3.

The analysis of the iDR3 data set, which is an incremental release, is currently ongoing. The iDR3 is incremental because it only includes new observations, completed after iDR2 was made

[1] http://www.gaia-eso.eu

[2] See http://sci.esa.int/gaia/ for more details on the European Space Agency (ESA) *Gaia* space mission.

[3] This final Survey wide homogenization uses as an anchoring point the results of the analysis of UVES spectra of FGK-type stars that is discussed in this paper. Therefore, our results are currently not changed by the final homogenization.

available. Exactly the same analysis strategy that we applied to iDR2 is being applied to iDR3.

For completeness, we also present the analysis of iDR1 in an Appendix. There, we discuss the main differences between the analysis implementation for these two iDRs. The first *Gaia*-ESO science verification papers were based on iDR1, and it is therefore important to document how this analysis was conducted.

We stress again that the *Gaia*-ESO spectrum analysis is under continous development. With improvements in the analysis, the complete Survey data set will be reanalyzed. Therefore, future releases of data products will supersede previous releases. Below is a description of iDR2:

- *Internal Data Release 2 (iDR2)*: this data release consisted of all spectra obtained from the beginning of the Survey up to the end of June 2013, and additional archival data that were included for calibration purposes. For the WG11 analysis, it included a total of 1708 spectra of 1447 FGK-type stars (multiple exposures of benchmark stars were analyzed separately, see Sect. 7.1). From these stars, 1412 were observed by *Gaia*-ESO, 35 of them were obtained from data archives, and 22 of them had both *Gaia*-ESO and archival spectra. The astrophysical results obtained from the analysis of the iDR2 data set will be part of the first *Gaia*-ESO public release of advanced data products. The public release will be available through a dedicated *Gaia*-ESO Survey science archive[4] hosted by the Wide Field Astronomy Unit (WFAU) of the Institute for Astronomy, Royal Observatory, Edinburgh, UK. The results of iDR2 supersede the science verification results of iDR1 presented in the Appendix.

For the analysis of the UVES spectra of FGK-type stars, we implemented multiple parallel methodologies as opposed to adopting one single analysis pipeline. The main advantage of a multiple analysis strategy in a broad survey like ours is that we can identify the different pipelines that perform better in different regions of the parameter space. We are therefore not constrained by the limitations of a single pipeline, which would introduce different systematics in different regions of the parameter space. In addition, with multiple analyses we can quantify the precision of the spectroscopic analyses, by reviewing how well the multiple pipelines agree in each star of the sample. Nevertheless, this strategy also adds a level of complexity to the understanding of the results. A single pipeline would be internally more homogeneous and provide results that are easier to reproduce and correct when (and if) needed.

In this paper, we present a comparison of these multiple pipelines applied to iDR2. Our final parameter scale is built by implementing a homogenization process that ties it to the fundamental scale defined by the *Gaia* benchmark stars. Different pipelines provide better results in different regions of the parameter space. Homogeneity is ensured by guaranteeing that the final results reproduce the "real" parameters of the reference stars well in each of the parameter space regions. We discuss how we use the multiple analyses to define the precision of our results, how the benchmarks are used to define the accuracy of the results, and present the limitations of the final catalog. This is a technical paper describing the spectrum analysis and its results. The scientific implications of the results will be discussed elsewhere.

The paper is structured as follows. Section 2 presents a summary of the sample of FGK-type stars that is analyzed within WG11. In Sect. 3, we present the general characteristics of the spectroscopic data used in the analysis. Section 4 describes the general properties of our multiple-analyses strategy and our homogenization procedure. It is followed by Sect. 5, where we describe the common tools that have been defined for the analysis. The subsequent sections present each of the data products determined in our analysis, discussing the method-to-method comparisons and, when applicable, comparing the final recommended results to the reference parameters of the calibrators. Equivalent widths are discussed in Sect. 6, stellar atmospheric parameters in Sect. 7, and elemental abundances in Sect. 8. Section 9 summarizes the analysis and highlights the scientific value of the data produced here. Two appendices complete the paper. Appendix A contains the details of the individual methodologies employed by each of the Nodes[5] involved in the data analysis. Appendix B presents the science verification analysis of iDR1 and discusses the differences between that and the one implemented for iDR2.

2. The FGK-type stars observed with UVES

The main late-type targets observed with UVES in the *Gaia*-ESO Survey are FG-type dwarfs in the solar neighborhood and clump giants in old (age >1 Gyr) and intermediate-age (0.1 Gyr < age < 1 Gyr) open clusters. In addition, the following targets are also present: 1) candidate clump giants in the inner disk and bulge; 2) K-type giants in the outer regions of the disk; 3) main-sequence and PMS stars in young clusters and in close by intermediate-age clusters; 4) field stars in the line of sight of open clusters; 5) giants in a few globular clusters observed for calibration purposes; 6) giants and dwarfs in fields observed by the CoRoT satellite, used here for calibration purposes. Figure 1 shows how the stars that are part of iDR2 are distributed in the $T_{\rm eff} - \log g$ plane (computed as described in Sect. 7).

The observations of the Milky Way solar neighborhood targets with UVES are made in parallel with Giraffe observations. This means that the exposure times are planned according to the observations being executed with Giraffe fibers (targets down to $V = 19$ mag). These UVES targets are chosen according to their near-infrared colors to be FG-dwarfs/turnoff stars with magnitudes down to $J = 14$ mag. The goal is to observe a sample of ~5000 FG-type stars within 2 kpc of the Sun to derive the detailed kinematic-multielement distribution function of the solar neighborhood. This sample includes mainly thin and thick disk stars, of all ages and metallicities, but also a small fraction of local halo stars.

The target selection is based on 2MASS (Skrutskie et al. 2006) photometry (point sources with quality flags "AAA"). A box is defined in a color-magnitude diagram with limits $12 < J < 14$ and $0.23 < (J - K) < 0.45 + 0.5 \times E(B - V)$. The Schlegel et al. (1998) maps are used to determine the extinction $E(B - V)$. The targets selected before April 2012 had a brightest cut on $J = 11$ instead of 12. When there were not enough targets, the red edge was extended. When there were too many potential targets an algorithm selected roughly the same number of stars per magnitude bin with the rest being marked as lower priority. A complete discussion of field target selection will be given in Gilmore et al. (in prep.).

In open clusters, while we use Giraffe to target complete samples of members down to $V = 19$ mag, with the UVES fibers we observe key brighter objects (down to $V = 16.5$ mag). The spectra are used for accurate atmospheric parameters and

[4] http://ges.roe.ac.uk/index.html

[5] Following the adopted *Gaia*-ESO Survey terminology, each of the independent analysis groups is referred to as a different analysis "Node".

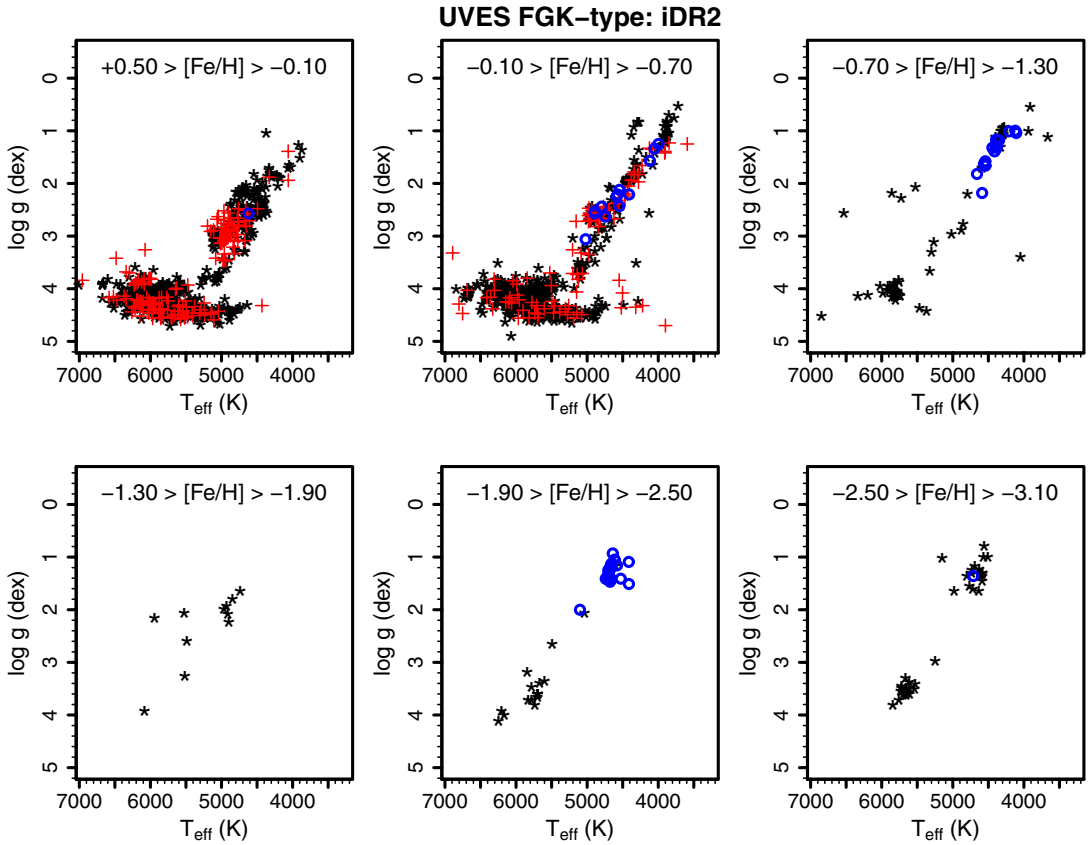

Fig. 1. Distribution of FGK-type stars from iDR2 in the $T_{\rm eff} - \log g$ plane. The panels are divided according to metallicity. Black stars are field stars, red crosses are stars observed in open cluster fields, and blue circles are stars observed in globular cluster fields.

abundances determination. For old and intermediate-age clusters, the UVES fibers are allocated mostly to red-clump giants. Main-sequence stars are also observed in closeby intermediate-age clusters. In young clusters, we also use the UVES fibers to observe selected main-sequence and PMS stars. These objects are first analyzed by the PMS analysis WG (see Lanzafame et al. 2014). Those stars that are considered to be normal FGK-type stars are later added to our analysis sample (i.e., PMS stars without veiling, non-cluster members, and main-sequence stars). In the clusters, exposure times are planned for observations being executed with the UVES fibers. Close to ~1000 FGK-type stars should be observed with UVES in clusters by the end of the Survey. The information obtained with the UVES spectra will enable the robust chemical characterization of the clusters, the study of possible star-to-star chemical variations, and will be critical inputs for studies of stellar evolution.

3. The data

Late-type stars are observed with UVES in the setup centered at 580 nm. The spectrum is exposed onto two CCDs, resulting in a wavelength coverage of 470−684 nm with a gap of ~5 nm in the center. The FLAMES-UVES fibers have an aperture on the sky of 1″, resulting in a resolving power of $R = 47\,000$.

The UVES data are reduced with the ESO UVES pipeline and dedicated scripts described in Sacco et al. (2014). Some data products are already constrained at this stage: the radial velocity ($v_{\rm rad}$) and its potential variation, and a first guess of the projected rotational velocity ($v \sin i$). The spectra are delivered to the analysis groups in a multiple-extensions FITS format. The data are made available through an operational database hosted by the Cambridge Astronomical Survey Unit (CASU) of the Institute of Astronomy at the University of Cambridge, UK.

Different versions of the spectra available include: 1) wavelength calibrated, sky subtracted, and heliocentric corrected merged spectra; 2) continuum normalized version of the previous spectra; and 3) individual single orders, wavelength calibrated, sky subtracted, and heliocentric corrected. The inverse variance of the spectra listed before are also available. The auxiliary data collected during the sample selection phase (such as photometry and proper motions) or derived during the data reduction phase (such as $v_{\rm rad}$ and $v \sin i$) are also provided. A correction of telluric features is not implemented yet.

The distribution of the S/N per pixel of the iDR2 data is shown in Fig. 2. The stars from the solar neighborhood sample, from open cluster fields, and the calibration targets are shown separately. The use of the calibration targets in the analysis is discussed in Sect. 7.

In addition to the *Gaia*-ESO sample, iDR2 also includes the library of high-resolution, high S/N observed spectra compiled by Blanco-Cuaresma et al. (2014). We analyzed spectra of 30 *Gaia* benchmark stars taken from this library. The *Gaia* benchmark stars are defined as well-known bright stars for which well-determined $T_{\rm eff}$ and $\log g$ values are available from direct methods, independent from spectroscopy (Heiter et al., in prep.). Their metallicities are well constrained from a careful spectroscopic study (Jofré et al. 2014), applying some of the same analysis methods used in the *Gaia*-ESO Survey.

As described in Sect. 7, the analysis of these benchmark stars is used to test the internal accuracy of the *Gaia*-ESO analysis and as an anchor for the scale of the *Gaia*-ESO parameters. In addition, these stars will be used as a first-level calibration

Table 1. Number of FGK-type stars observed with UVES and part of the iDR2 data set.

Gaia-ESO Type	Stars	Comments
Total	1447	*Gaia*-ESO and archival data.
Gaia-ESO	1412	*Gaia*-ESO only, no archival data.
GES_MW	941	Stars from Milky Way fields.
GES_CL	314	Stars from open cluster fields.
GES_SD	157	Calibration targets.
AR_SD	55	Calibrators from archival data (M 67 and the Blanco-Cuaresma et al. library).
GES_SD_BM	20	Benchmark stars with *Gaia*-ESO spectra.
GES_SD_PC	2	Peculiar stars templates.
GES_SD_GC	51	Stars from calibration globular clusters.
GES_SD_OC	23	Stars from calibration open clusters.
GES_SD_CR	55	Stars from the CoRoT fields.

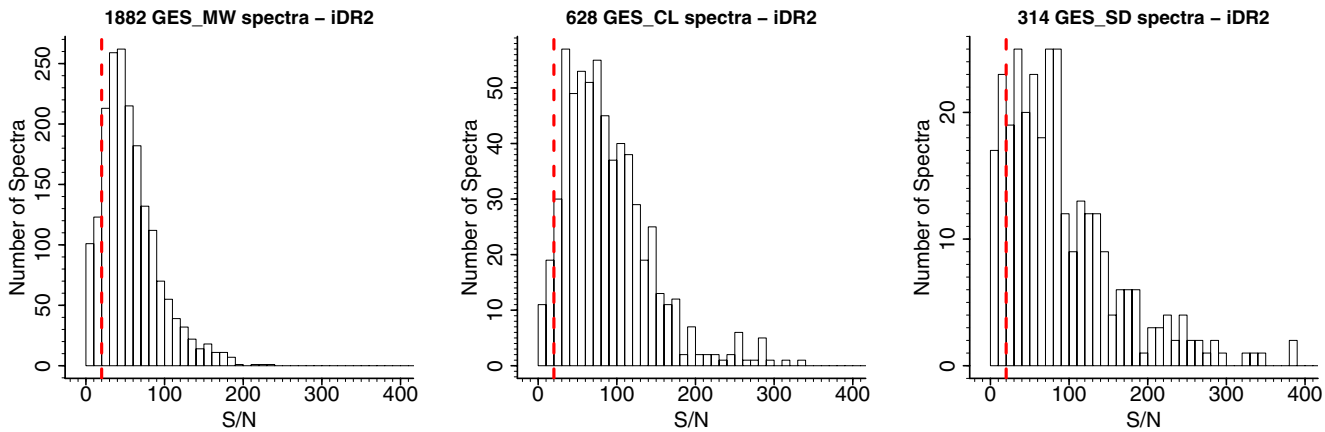

Fig. 2. Distribution of the median S/N per pixel of the spectra observed with UVES that are part of iDR2 (1412 FGK-type stars). Each of the two UVES spectrum parts (from each CCD) is counted separately (thus, two spectra per star). The red dashed line indicates $S/N = 20$. The samples of the solar neighborhood (GES_MW), open clusters (GES_CL), and calibration targets (GES_SD) are shown separately. Only final stacked spectra were included in this plot; the single exposure spectra, even when analyzed, are not counted.

for the *Gaia* results (Bailer-Jones et al. 2013). Their inclusion in our sample is a step toward guaranteeing a high degree of homogeneity between the results of *Gaia*-ESO and *Gaia*. Similarly, other large spectroscopic surveys can use these stars (and other stars of the *Gaia*-ESO calibration sample) to compare their astrophysical parameters scale with ours (see Pancino 2012; and Pancino et al., in prep.). This effort can eventually lead to a global scale of astrophysical parameters across different large spectroscopic surveys. Table 1 summarizes the number of stars included in the iDR2 data set.

4. The analysis strategy

Because of their high-resolution and large wavelength coverage, the UVES spectra allow for the determination of a large number of quantities. The list includes the stellar atmospheric parameters: effective temperature ($T_{\rm eff}$), surface gravity ($\log g$), microturbulence (ξ); the stellar metallicity [Fe/H][6]; elemental abundances for as many elements as the S/N and astrophysical parameters permit; and chromospheric activity indicators[7], where relevant.

In this section, we summarize the general strategy of our spectroscopic analysis, including quality control and homogenization steps. We only present the analysis strategy adopted during the iDR2 analysis used to compute the quantities that will be included in the first public release. The strategy applied during the iDR1 analysis for science verification differed from the current strategy in several respects. These differences are discussed in Appendix B.

4.1. Multiple pipelines strategy

The *Gaia*-ESO Survey consortium includes specialists in many major state-of-the-art standard and special-purpose spectrum analysis methodologies currently employed in the literature. This provides the unique opportunity of applying multiple parallel pipelines to the same large data set. This approach has two main advantages:

1. No single pipeline is optimal to analyze all stellar types that are included in our sample (e.g., giants vs. dwarfs; metal-poor vs. metal-rich stars). With multiple pipelines we can identify and use those that give best results in different regions of the parameter space. All types of objects can thus be properly analyzed, even if they require special treatment.
2. We can investigate and quantify different sources of errors, including method-dependent effects. This gives a robust

[6] The metallicity as an atmospheric parameter refers to the global content of metals in the stellar photosphere. Usually the Fe abundance is used as a proxy of the metallicity. That is true for some of the analysis methodologies employed here, but for others a global metallicity value is determined (see each method in Appendix A).

[7] Chromospheric activity indicators have not yet been derived from the spectra discussed here. The calculation of these quantities is planned and the methods used will be discussed in papers describing future data releases.

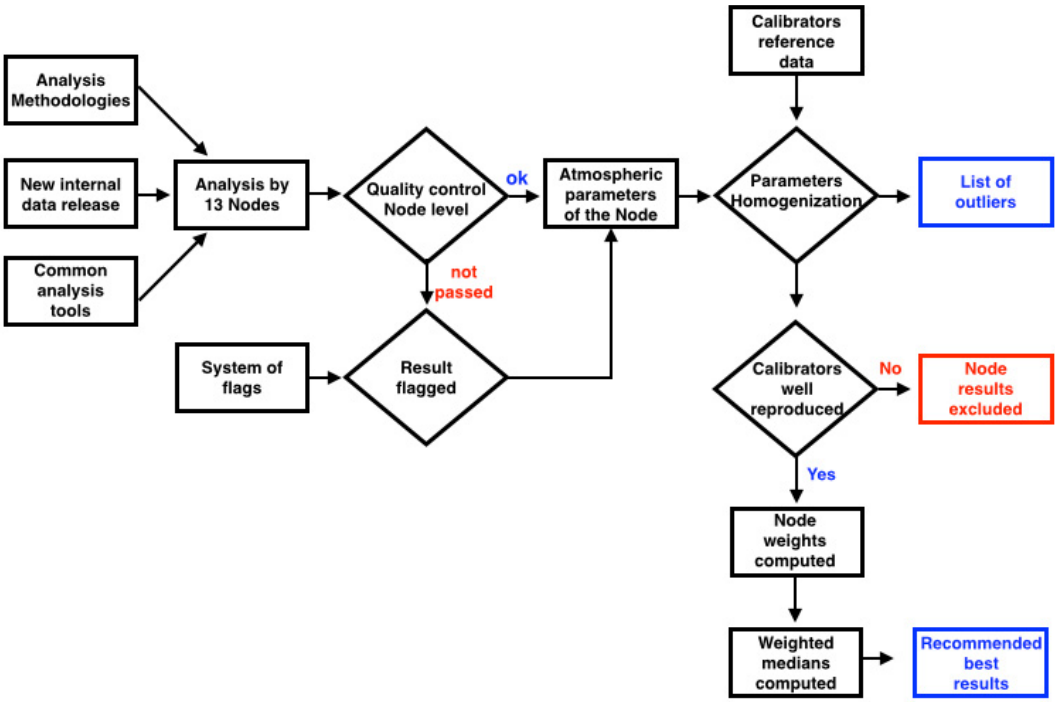

Fig. 3. Example of the analysis flow for the atmospheric parameters with the main tasks and decisions indicated.

measurement of what is the precision of spectroscopic analyses. This is an opportunity to find, on a star-by-star basis, the degree to which their absolute parameters can be trusted.

Both are invaluable advantages in a survey with targets spanning a broad range in atmospheric parameters. We mention in passing that a system of multiple analyses has been implemented to deal with the *Gaia* data (Bailer-Jones et al. 2013). Obviously, the multiple pipelines strategy also adds some complexity to the analysis process and the understanding of the results. In addition, the comparison between the analyses cannot capture all the systematic uncertainties in spectroscopic analyses such as those caused by the limitations of the model atmospheres or by, e.g., ignoring non-LTE effects in the atmospheric parameters (as is the case of most of the methods implemented here apart from the LUMBA Node – Sect. A.8).

To guarantee that we can deliver the best possible results, with well quantified uncertainties, we have established a series of critical tests to evaluate the results, and to bring them to a single parameter scale. Figure 3 depicts the flow of our analysis strategy, as applied to the atmospheric parameters. We discuss now the general properties of these steps, and specific results for each data product are presented in Sect. 6 for EWs, in Sect. 7 for atmospheric parameters, and in Sect. 8 for the abundances.

4.2. Node analyses

The spectroscopic analysis is performed by 13 different analysis Nodes. The methodologies and codes used by each Node are described in detail in Appendix A and summarized in Table 2. The implementation and limitations of each code are discussed in detail elsewhere (references are given in the Appendix).

We stress that all Nodes analyze the same data, as data reduction is a step completed independently from the spectroscopic analysis. In addition, a number of "common tools" have been defined to guarantee some degree of homogeneity in the end results. These tools include: the use of a common line list (of atomic and molecular lines), the use of one single set of model atmospheres, and the analysis of common calibration targets. These constraints are shared with other *Gaia*-ESO WGs, particularly with the WG responsible for the analysis of Giraffe spectra of FGK-type stars. These steps are also taken to facilitate the job of putting the full *Gaia*-ESO results into a single homogeneous scale (see François et al., in prep.). In addition, for use when needed by some analysis methodologies, a microturbulence calibration is recommended and a synthetic spectrum library computed with the same line list is available. More details on these tools are given in Sect. 5, or will be discussed in forthcoming publications.

Each Node performs a first quality control of their own results. They identify objects where the analysis has failed, and investigate the limits to which their results can be trusted. A dictionary of flags is used at this stage. The flags include the possibility of identifying: i) phenomenological peculiarities (e.g., emission lines, multiplicity, or fast rotation); ii) stellar classification remarks, indicating for example a particular evolutionary stage (e.g. white dwarfs, post-AGBs) or properties like strong lines caused by carbon enhancement; and iii) technical issues, such as problems with data reduction, signal-to-noise, or analysis convergence issues. The flags will be part of the released products and the complete dictionary will be described elsewhere (Gilmore et al., in prep. and in the release documentation).

4.3. Parameters homogenization

By parameter homogenization, we mean the procedure of checking the performance of the Node analyses and establishing the final recommended values. At this homogenization step, stars for which only a few Nodes (three or less) have provided parameters have the spectra individually checked. For the vast majority of these cases, the reasons for the analysis failure is easily detected (e.g., fast rotation, emission lines, data reduction issues). A list of outliers is produced, including the appropriate flags.

Table 2. Summary of the analysis methodologies used by each Node involved in WG11.

Node	Contact	Codes	Type of method
Bologna	E. Pancino	DAOSPEC and GALA	Equivalent widths
Catania (OACT)	A. Frasca	ROTFIT	Library of observed spectra
CAUP	S. Sousa	ARES and MOOG	Equivalent widths
Concepcion	S. Villanova	DAOSPEC and GALA	Equivalent widths
EPInArBo	L. Magrini	DAOSPEC and FAMA	Equivalent widths
IAC-AIP	C. Allende Prieto	FERRE	Library of synthetic spectra
Liège	T. Morel	GAUFRE	Equivalent widths
LUMBA	S. Feltzing	SGU based on SME	Synthetic spectra computed on the fly
Nice	V. Hill	MATISSE	Library of synthetic spectra
Paris-Heidelberg	L. Sbordone	MyGIsFOS	Library of synthetic spectra
UCM	D. Montes	ARES and StePar	Equivalent widths
ULB	S. Van Eck	BACCHUS	Synthetic spectra computed on the fly
Vilnius	G. Tautvaišienė	DAOSPEC and MOOG	Equivalent widths

Atmospheric parameters for these stars are not provided, and the list is forwarded to the WG responsible for outlier objects for further investigation.

To critically evaluate the performance of the Nodes, we use a series of calibrators. The *Gaia* benchmark stars, a set of ~30 stars with well-defined fundamental parameters (Heiter et al., in prep.), are the first level of calibration. They are also used as an anchor to define the final scale of the *Gaia*-ESO parameters. For the second level of calibration we use a series of open and globular clusters, where the consistency of the $T_{\rm eff}$ vs. log g values can be checked. Another level of calibration will be possible with the stars observed by the CoRoT satellite, for which asteroseismic log g values are being computed. This third check will be implemented in future releases.

The performance with respect to the benchmarks is judged separately in three regions of the parameter space: 1) *metal-rich dwarfs*: stars with [Fe/H] > -1.00 and log $g > 3.5$; 2) *metal-rich giants*: stars with [Fe/H] > -1.00 and log $g \leq 3.5$; and 3) *metal-poor stars*: stars with [Fe/H] ≤ -1.00. Node results that fail the tests with the calibrators are excluded. For the remaining results, we define weights according to how well they can reproduce the reference parameters of these stars.

These weights are used to compute a weighted-median value for each atmospheric parameter. The weighted medians are adopted as the recommended best value of the atmospheric parameters. Medians are used as they are robust against outliers, minimizing the influence of less consistent results. The weights help to select the best methods in each region of the parameter space, and to force the scale to reproduce the real parameters of the benchmark stars. This is a significant advantage of our approach compared to using the Sun as sole reference star.

A weighted-median approach is also used for the abundances. The difference is that, apart from the Sun, there are no fundamental references of stellar abundances. We thus combined the individual Node values using the same weights defined for the atmospheric parameters. Weighted medians were computed on a line-by-line basis. The final abundance of an element is the median of the line values. In Sects. 6–8, we discuss in detail the approach used to define the final recommended values of EWs, atmospheric parameters, and abundances, respectively.

When using the *Gaia*-ESO results, the final recommended values with their uncertainties in terms of accuracy and precision are preferable. These are the values that have been critically evaluated and calibrated to the system defined by the *Gaia* benchmarks.

5. Common analysis tools

5.1. Line list

The *Gaia*-ESO Survey line list is a compilation of experimental and theoretical atomic and molecular data. As with the analysis strategy, the line list will keep evolving, being updated and improved before new analysis cycles. The details of this compilation, and the full line list will be provided in a separated publication (Heiter et al., in prep.).

Version 4.0 of the line list was used to analyze the iDR2 data. The list of molecules includes: C_2 ($^{12,13}C^{12,13}C$), CaH, ^{12,13}CH, ^{12,13}CN, FeH, MgH, NH, OH, SiH, ^{46,47,48,49,50}TiO, VO, and ^{90,91,92,94,96}ZrO. Atomic transitions needed for both spectrum synthesis and equivalent width analysis are included. Where needed, isotopic shifts and hyperfine structure (HFS) were included (for Sc I, V I, Mn I, Co I, Cu I, Ba II, Eu II, La II, Pr II, Nd II, Sm II). Some atomic oscillator strengths have been newly calculated for the survey (Ruffoni et al. 2014). Collisional broadening by hydrogen is considered following the theory developed by Anstee & O'Mara (1991) and Barklem & O'Mara (1998), where available, including some new broadening computations still unpublished that will be discussed in Heiter et al. (in prep.).

The line-list group critically reviewed the lines used for the EWs analyses and designed a system of flags that is made available together with the line list (also to be published). The flags indicate the quality of the transition probability and the blending properties of the line, as evaluated in the spectra of the Sun and of Arcturus.

It is perhaps necessary to stress here that while all Nodes have access to the same *Gaia*-ESO "master" line list, this does not mean that the methods make use of the same selected subsample of spectral lines. The choice of lines used to constrain the parameters and abundances is made by each Node according to the details of their methodology. As is common, some groups using EWs prefer to select a restricted set of the best lines, while others prefer to rely on the statistical properties of many lines. Other groups prefer to use strong lines such as Hα to assist in constraining the parameters. In addition to that, there are the methods that rely on fitting large portions of the observed spectra in comparison with synthetic spectra. These methods need more extensive line lists, not only those useful for an EW analysis. This is to emphasize that, even though a common line list is adopted, there is still considerable freedom as to how this line list is finally employed by each Node.

5.2. Model atmosphere

For model atmospheres we adopted the MARCS grid of Gustafsson et al. (2008). The grid consists of spherically-symmetric models complemented by plane-parallel models for stars of high surface gravity (between $\log g = 3.0$ and 5.0, or 5.5 for the cooler models). It assumes hydrostatic equilibrium, local thermodynamic equilibrium (LTE), and uses the mixing-length theory of convection. The MARCS models assume solar abundances of Grevesse et al. (2007) and are α-enhanced at low metallicities.

We remark here that the coverage of the MARCS grid in the metal-poor regime is sometimes incomplete. Some of the analysis methods need to be able to interpolate among a grid of models on the fly. For metal-poor stars, it often happens that some of the models needed for this interpolation are not available. These methods will then fail when the border of the grid is reached. This aspect introduces one additional complication to the analysis of metal-poor stars.

Within *Gaia*-ESO, we decided to list the abundances in the "$\log \epsilon$" format[8], without assuming a solar composition. Nevertheless, when metallicities as an atmospheric parameter in the format [Fe/H][9] are quoted in this work, we adopt the solar Fe abundance of Grevesse et al. (2007), $\log \epsilon(\mathrm{Fe})_\odot = 7.45$, unless otherwise noted.

5.3. Spectrum library

For analysis methodologies that make use of pre-computed synthetic spectra, a *Gaia*-ESO library of synthetic spectra is provided. Here we only provide a short description of the library, a complete discussion is given in Recio-Blanco et al. (in prep.).

The synthetic spectra were calculated using the same software used to compute the AMBRE grid of synthetic spectra (de Laverny et al. 2012). The spectra have $R \sim 300\,000$ and cover the whole wavelength region of the UVES setup with a sampling of 0.004 Å. Spectra with different degrees of alpha-enhancement were computed, to account for the feedback of important α-element electron donors on the atmospheric structure. This grid was computed using the complete *Gaia*-ESO line list (atoms + molecules). With each update of the line list, a new grid is computed.

5.4. Microturbulence calibration

A *Gaia*-ESO microturbulence calibration is provided and recommended for those methods that do not derive this parameter from the spectrum analysis. It is used by a few Nodes in the analysis of the UVES spectra, but is more extensively used in the analysis of Giraffe spectra (Recio-Blanco et al., in prep.) because of the reduced number of clean Fe I lines available for constraining this parameter.

These relations are based on the UVES science verification results (obtained as described in Appendix B), on the parameters of the *Gaia* benchmark stars described in Jofré et al. (2014), and on globular cluster data from literature sources. Three relations were derived, for different types of stars, and are valid for $4000 < T_{\mathrm{eff}}$ (K) < 7000, $0.0 < \log g$ (dex) < 5.0, and $-4.5 <$ [Fe/H] (dex) $< +1.0$. A full discussion of these relations will be presented in Bergemann et al. (in prep.).

[8] $\log \epsilon(X) = \log [N(X)/N(\mathrm{H})] + 12$, i.e., a logarithmic abundance by number on a scale where the number of hydrogen atoms is 10^{12}.

[9] $[\mathrm{A/B}] = \log [N(\mathrm{A})/N(\mathrm{B})]_\star - \log [N(\mathrm{A})/N(\mathrm{B})]_\odot$

6. Equivalent widths

Some of the analysis methodologies described in Appendix A rely on the measurement of EWs to determine both the stellar atmospheric parameters and elemental abundances of the stars. The measurements of these EWs are going to be released as part of the *Gaia*-ESO data products. Equivalent widths will be given only for lines effectively used by at least one Node in their analysis. The tables that will be released to the community will include for each line: the average EW, the multiple-measurement dispersion, the number of measurements, and the flags (where applicable).

In this section, we discuss a comparison of multiple measurements of the EWs in the spectra that are part of the iDR2 data set. The EWs are available for 1265 out of the 1268 stars observed by the *Gaia*-ESO Survey for which atmospheric parameters were determined.

The Nodes making use of the traditional EW-based analysis method are: Bologna, CAUP, Concepcion, EPInArBo, Liège, UCM, and Vilnius. The Liège Node measures EWs with the GAUFRE code, however, their measurements were lost because of a computer problem. Therefore, the discussion in this section concentrates on results from only two codes that measure EWs automatically: ARES (Sousa et al. 2007) and DAOSPEC (Stetson & Pancino 2008, 2010). Currently, only DAOSPEC returns a value for the EW measurement error.

Figure 4 shows the comparison between the EWs of lines of different elements measured by different groups in two stars. They represent an easy and a hard case for measuring EWs. One star is a metal-poor dwarf observed with high S/N per pixel (~260); the other star is a metal-rich giant observed with relatively low S/N per pixel (~50). In this and other plots in this section, we compare the multiple measurement scatter with the statistical uncertainty in the EW measurement given by the Cayrel (1988) "formula" (Eq. (7) of that article). This formula gives the EW uncertainty due to random noise when fitting the line profile with a Gaussian. This value is used here as a reference for the "expected uncertainty" but does not take into account all possible sources of error, such as continuum placement.

In Fig. 4, the EWs measured with the same code by different Nodes (left plots using ARES and center-left plots using DAOSPEC) tend to agree to within 2 or 3σ, although systematic differences might be present in some cases. When comparing the EWs measured with ARES (CAUP and UCM Nodes) with those measured by DAOSPEC (Bologna and Vilnius Nodes), center-right and right plots of Fig. 4, it is noticeable that the scatter increases. There seems to be no trend between the ΔEWs and the EWs themselves. Such trends could produce biases in the determination of the microturbulence.

Figure 5 depicts the behavior of $\overline{\sigma_{\mathrm{EW}}}$. For each spectral line used for abundance determination in a given star, the average value of the multiple determinations of its EW is computed together with its standard deviation. For each star, we define $\overline{\sigma_{\mathrm{EW}}}$ as the mean of all the standard deviations of the lines measured in that star. Figure 5 shows that for the majority of the stars, the measurements tend to agree to a level that is better than the expected statistical uncertainty given by the S/N of the spectra. About 70% of the stars have the blue part of the spectrum with median S/N per pixel below 70. For this S/N, the expected 2σ uncertainty of the EWs is of the order of 3 mÅ. As shown in the left panel of Fig. 5, about ~13.7% of the stars have $\overline{\sigma_{\mathrm{EW}}}$ above that. In a few cases, however, it can reach up to ~15 mÅ. A more detailed comparison with the S/N expectation, right panel

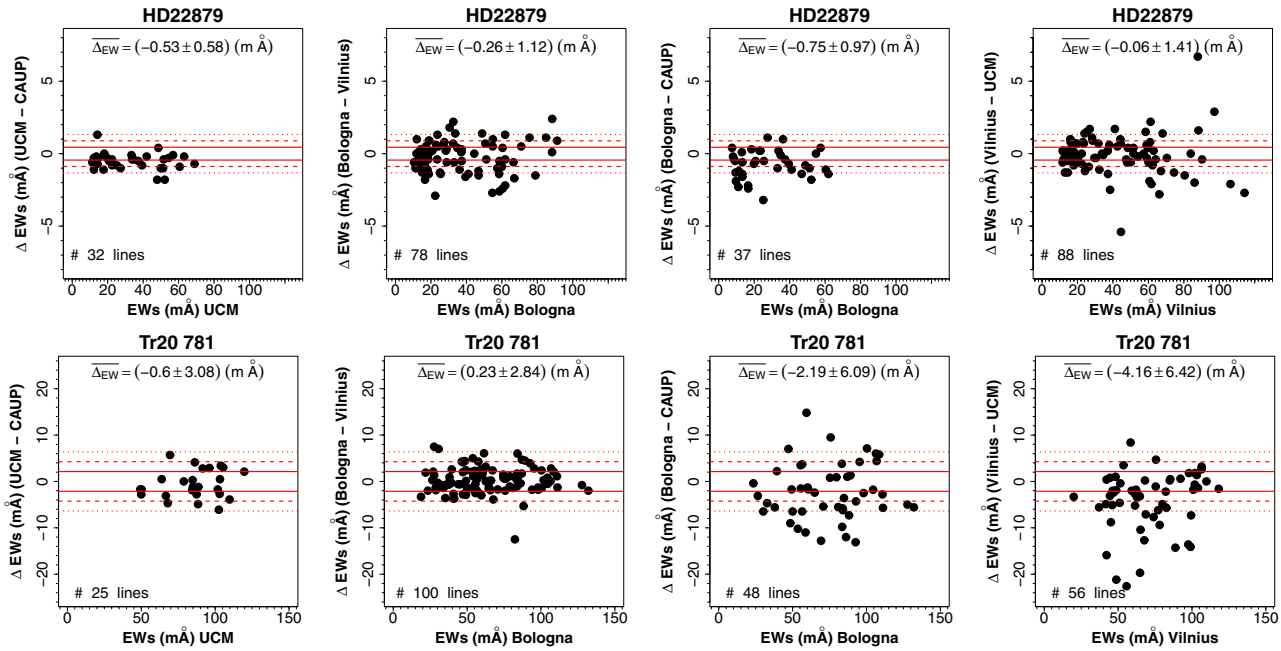

Fig. 4. Comparison between equivalent widths measured by different Nodes for two stars. *Top row*: star HD 22879, a benchmark star used for calibration with $T_{\rm eff}$ = 5786 K, $\log g$ = 4.23, and [Fe/H] = −0.90. Median values of the S/N per pixel are 239 and 283 for the blue and red part of the spectra, respectively. Red lines indicate the typical 1σ (solid line), 2σ (dashed line), and 3σ (dotted line) uncertainty of the EW computed with the Cayrel (1988) formula, adopting $FWHM$ = 0.190 Å, pixel size =0.0232 Å, and S/N = 260. *Bottom row*: a clump giant in the open cluster Trumpler 20 (Trumpler 20 MG 781 in the numbering system of McSwain & Gies 2005), with $T_{\rm eff}$ = 4850 K, $\log g$ = 2.75, and [Fe/H] = +0.15. Median values of the S/N per pixel are 36 and 68 for the blue and red part of the spectra, respectively. Red lines indicate the typical 1σ (solid line), 2σ (dashed line), and 3σ (dotted line) uncertainty of the EW computed with the Cayrel (1988) formula, adopting $FWHM$ = 0.190 Å, pixel size =0.0232 Å, and S/N = 50. In each panel, the average difference of the EWs and its dispersion are also given.

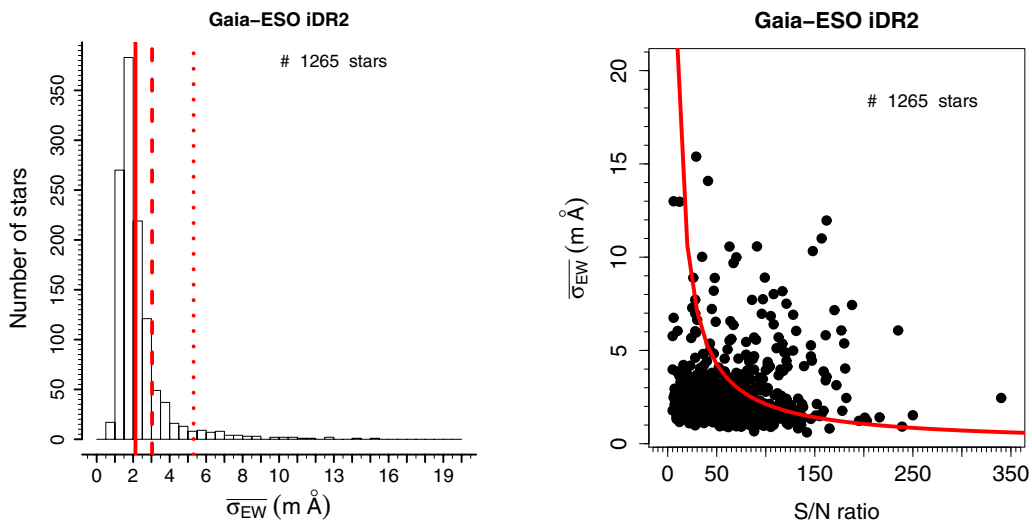

Fig. 5. *Left panel*: histogram of $\overline{\sigma_{\rm EW}}$ per star, taking into account the measurements of all Nodes. Also shown are lines indicating the 2σ uncertainty calculated with Cayrel (1988) formula for S/N = 40 (dotted line at 5.31 mÅ), S/N = 70 (dashed line at 3.04 mÅ), and S/N = 100 (solid line at 2.12 mÅ). *Right panel*: dependence of $\overline{\sigma_{\rm EW}}$ with respect to the median of the S/N per pixel. Also shown is the expected 2σ value given by the Cayrel (1988) formula (as a red line).

of the same figure, shows that the quantity $\overline{\sigma_{\rm EW}}$ is above the 2σ expectation for about 11.7% of the stars.

The cases with higher dispersion might be related to different issues that make the measurement of EWs difficult (e.g., low temperature, high-metallicity, and/or broad lines). Other problems contributing to increase the scatter in the measurements include: the different ways that the continuum is defined in each code (global vs. local continuum for DAOSPEC and ARES respectively); the presence of reduction artefacts; unrecognized binarity in the spectra; the residual wavelike pattern in the continuum caused by problems with the blaze-function correction, as sometimes seen in high S/N echelle spectra[10]; and the free

[10] We note in particular that HD 22879, which is used as an example in Fig. 4 suffers from this issue. This will perhaps more seriously affect DAOSPEC than ARES, as DAOSPEC performs a global fit of the continuum for the whole spectrum. Therefore, the expected uncertainty computed with the Cayrel (1988) formula for the EW measurements should be taken as a lower limit.

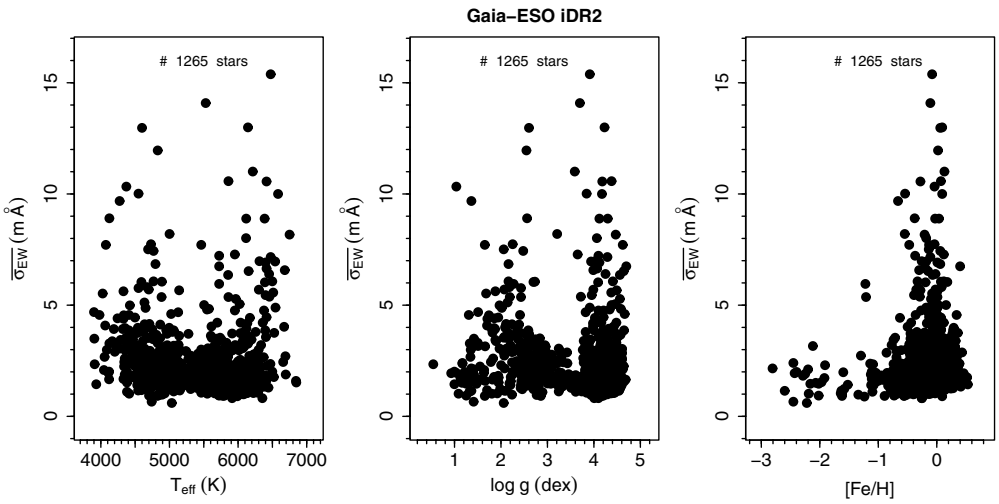

Fig. 6. Mean of all the standard deviations of the EW measurements in a star, $\overline{\sigma_{\rm EW}}$, as a function of the atmospheric parameters.

parameters in each code that need to be adjusted for the measurements. Therefore, the scatter in the measurement of EWs is not just statistical in nature.

In Fig. 6, $\overline{\sigma_{\rm EW}}$ is plotted against the atmospheric parameters of the stars. In Fig. 7, we plot $\overline{\sigma_{\rm EW}}$ against the rotational velocity ($v \sin i$) of the stars. Not all stars have an estimate of $v \sin i$, because this measurement fails in some cases (see Sacco et al. 2014). The figures show that most of the stars where $\overline{\sigma_{\rm EW}} > 5$ mÅ tend to be metal-rich objects. In addition, some of these stars have cool temperatures and many display high rotation. All these factors increase the uncertainty with which EWs can be measured with automatic methods.

It is not the scope of this section to delve into the details of why a perfect agreement between multiple measurements of the same line is not obtained. ARES and DAOSPEC are fully described in dedicated publications where both codes are compared with each other and with other codes. We thus refer the reader to Sousa et al. (2007), Stetson & Pancino (2008), and Cantat-Gaudin et al. (2014a) for these detailed discussions.

This section documents how we did the measurements and discusses the quality of the results and their limitations. For the majority of the stars, the scatter in the multiple measurements compares well with the statistical uncertainty estimated with the Cayrel (1988) formula. Thus, the EW measurements for these stars do not seem to be affected by additional sources of error. For the remaining stars, multiple factors play a role, some of which were identified above. Abnormalities in the spectra will be flagged in the final catalog.

We have, however, identified the general regions of the parameter space where problems are likely to occur. We are working to improve the analysis of these stars, and expect to provide improved results for future releases. A satisfactory agreement between the multiple measurements of EWs is obtained for about 88% of the stars discussed here. For those stars with higher scatter in the EWs, we also expect a large scatter in the comparison of the atmospheric parameters and abundances. However, not all analysis methodologies make use of EWs. As we discuss in the following sections, all values of atmospheric parameters and abundances are given with an estimate of the method-to-method dispersion. This is a measurement of the precision of these values. Values with increased dispersion are more uncertain and should not be given the same weight as more precise results.

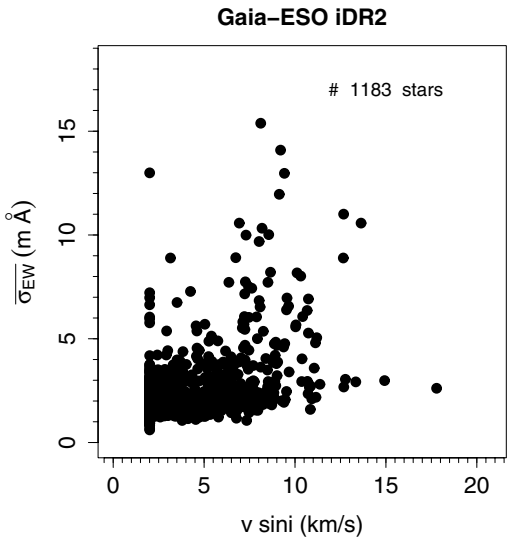

Fig. 7. Mean of all the standard deviations of the EW measurements in a star, $\overline{\sigma_{\rm EW}}$, as a function of the rotational velocity of the star.

7. Atmospheric parameters

As presented in Appendix A, the methods used to derive atmospheric parameters differ from Node to Node. They range from the standard use of EWs of Fe lines to different algorithms that use libraries of observed and/or synthetic spectra.

Once the different Nodes have finalized the first step of the spectroscopic analysis, we face the challenge of putting all the results together, understanding the differences and systematics, and producing a single list with the best, recommended values of the four atmospheric parameters ($T_{\rm eff}$, $\log g$, ξ, and [Fe/H]).

In the analysis of *Gaia*-ESO data, we aim to understand both the precision and accuracy with which the atmospheric parameters can be determined. The dispersion among the results from different methodologies is a good indication of the precision of the values. The accuracy is judged using the comprehensive set of calibrators observed by the Survey: the *Gaia* benchmark stars and the calibration clusters. In addition to those, for subsequent releases we expect to use giants that have asteroseismic-estimated gravities, determined using CoRoT light curves, to help in the calibration effort (see, e.g., Morel & Miglio 2012).

Table 3. Reference parameters of the benchmark stars.

Star	$T_{\rm eff}$ (K)	$\sigma_{T_{\rm eff}}$ (K)	$\log g$ (dex)	$\sigma_{\log g}$ (dex)	[Fe/H] (dex)	$\sigma_{\rm [Fe/H]}$ (dex)	Parameter Space	Remark
Sun	5777	1	4.44	0.01	0.00	0.01	MRD	Used in iDR2; only archival data
Arcturus	4247	28	1.59	0.04	−0.53	0.01	MRG	Only archival data
Procyon	6545	82	3.99	0.02	−0.04	0.01	MRD	Problems with order merging
18 Sco	5747	29	4.43	0.01	0.01	0.01	MRD	Used in iDR2
61 Cyg A	4339	22	4.43	0.16	−0.33	0.03	MRD	Only archival data
61 Cyg B	4045	20	4.53	0.04	−0.38	0.02	MRD	Only archival data
Alf Cen A	5847	68	4.31	0.02	0.24	0.01	MRD	Used in iDR2
Alf Cet	3796	64	0.91	0.08	−0.45	0.05	MRG	Cool star; only archival data
Alf Tau	3927	39	1.22	0.10	−0.37	0.02	MRG	Cool star; only archival data
Bet Ara	4172	48	1.01	0.13	−0.05	0.04	MRG	Used in iDR2
Bet Gem	4858	55	2.88	0.05	0.12	0.01	MRG	Only archival data
Bet Hyi	5873	38	3.98	0.02	−0.07	0.01	MRD	Used in iDR2
Bet Vir	6083	17	4.08	0.01	0.21	0.01	MRD	Used in iDR2; problems with order merging
Del Eri	5045	59	3.77	0.02	0.06	0.01	MRD	Used in iDR2
Eps Eri	5050	25	4.60	0.03	−0.10	0.01	MRD	Only archival data
Eps For	5069	59	3.45	0.05	−0.62	0.01	MRD	Used in iDR2
Eps Vir	4983	56	2.77	0.01	0.13	0.01	MRG	Only archival data
Eta Boo	6105	19	3.80	0.02	0.30	0.01	MRD	Used in iDR2; $v \sin i \simeq 12.7$ km s^{-1}
Gam Sge	3807	48	1.05	0.10	−0.16	0.04	MRG	Used in iDR2; cool star
Ksi Hya	5044	33	2.87	0.01	0.14	0.01	MRG	Used in iDR2
Mu Ara	5845	29	4.27	0.02	0.33	0.01	MRD	Used in iDR2
Mu Leo	4474	52	2.50	0.07	0.26	0.02	MRG	Used in iDR2
Tau Cet	5331	15	4.44	0.02	−0.50	0.01	MRD	Used in iDR2
HD 22879	5786	16	4.23	0.02	−0.88	0.01	MRD	Used in iDR2
HD 49933	6635	18	4.21	0.03	−0.46	0.01	MRD	Used in iDR2; $v \sin i \simeq 10.0$ km s^{-1}
HD 84937	6275	17	4.11	0.06	−2.09	0.02	MPS	Used in iDR2; metal-poor star; only archival data
HD 107328	4496	53	2.11	0.07	−0.34	0.01	MRG	Used in iDR2
HD 122563	4587	54	1.61	0.07	−2.74	0.01	MPS	Used in iDR2; metal-poor star
HD 140283	5720	29	3.67	0.04	−2.43	0.02	MPS	Used in iDR2; metal-poor star
HD 220009	4275	50	1.43	0.10	−0.75	0.01	MRG	Used in iDR2

Notes. $T_{\rm eff}$ and $\log g$ are direct determinations (see Heiter et al., in prep.). Metallicities were derived by Jofré et al. (2014). The metallicity uncertainty listed here only reflects the standard deviation of the mean abundance of the Fe I lines. Also given is the parameter space group to which the star belongs (MRD, MRG, or MPS – see text.)

In the subsections that follow below, we describe how the recommended atmospheric parameters for the iDR2 data set were determined. These results will be part of the first *Gaia*-ESO public release. We determined the results used in the first few *Gaia*-ESO science verification papers in a slightly different way (Appendix B). We start the discussion presenting the use of the main calibrators employed in the *Gaia*-ESO analysis.

7.1. The Gaia benchmark stars

The parameters ($T_{\rm eff}$ and $\log g$) of these well-known bright stars are available from direct methods or from calibrations that are independent of spectroscopy (see Heiter et al., in prep.). Jofré et al. (2014) determined the metallicities adopted here as reference using these same parameters. Table 3 compiles the reference parameters of 30 benchmark stars. The spectra analyzed include both new *Gaia*-ESO observations and the spectrum library of Blanco-Cuaresma et al. (2014).

The atmospheric parameter scale of the *Gaia*-ESO results is tied to the system defined by these benchmark stars. This is a considerable improvement with respect to the standard approach of using the Sun as the only reference. The *Gaia* benchmark stars are distributed across the parameter space, meaning that we can choose better references for stars that are not solar like.

7.1.1. The accuracy of the Node results

We divided the benchmark stars into three groups to judge separately the accuracy of the results in different corners of the parameter space. The groups were: 1) *metal-rich dwarfs* (MRD): stars with [Fe/H] > −1.00 and $\log g > 3.5$ (contains 11 benchmark stars); 2) *metal-rich giants* (MRG): stars with [Fe/H] > −1.00 and $\log g \leq 3.5$ (contains 7 benchmark stars); and 3) *metal-poor stars* (MPS): stars with [Fe/H] ≤ −1.00 (contains three benchmark stars). Only one group of metal-poor stars was defined because only three benchmark stars with [Fe/H] ≤ −1.00 are available.

Some Nodes had difficulty analyzing the archival data. Because the spectra were obtained with different spectrographs, they were made available in a different format than the standard *Gaia*-ESO format. The analysis problem was a shortcoming caused by the use of automatic pipelines designed to deal with a large amount of data in the same format. Thus, to judge the Node results accuracy for iDR2, we decided to use: 1) the results of 19 benchmark stars observed by *Gaia*-ESO; 2) the analysis of a FLAMES spectrum of the Sun[11]; and 3) the analysis of the archival spectrum of HD 84937 (one of the few metal-poor stars in this list).

[11] Obtained on the evening twilight sky and available here http://www.eso.org/observing/dfo/quality/GIRAFFE/pipeline/solar.html

Table 4. Average difference between the Node result for the *Gaia* benchmark stars and the reference values in each region of the parameter space.

Node	MRD $\Delta(T_{\rm eff})$ (K)	MRD $\Delta(\log\ g)$ (dex)	MRD Num. of stars	MRG $\Delta(T_{\rm eff})$ (K)	MRG $\Delta(\log\ g)$ (dex)	MRG Num. of stars	MPS $\Delta(T_{\rm eff})$ (K)	MPS $\Delta(\log\ g)$ (dex)	MPS Num. of stars
Bologna	46	0.13	11	163	0.40	7	–	–	0
CAUP	93	0.21	8	193	0.42	4	–	–	0
Concepcion	150	0.28	8	162	0.48	5	87	1.11	1
EPINARBO	57	0.14	10	74	0.31	7	167	0.35	1
IACAIP	131	0.16	9	114	0.22	7	82	0.23	1
Liege	186	0.22	8	208	0.62	7	–	–	0
LUMBA	81	0.14	11	139	0.39	5	165	0.07	3
Nice	78	0.26	11	82	0.30	5	59	0.20	3
OACT	169	0.19	10	159	0.37	7	–	–	0
ParisHeidelberg	71	0.12	10	91	0.34	5	87	0.43	1
UCM	123	0.11	11	465	0.94	6	–	–	0
ULB	–	–	–	–	–	–	–	–	–
Vilnius	59	0.09	11	184	0.51	6	2	1.10	1

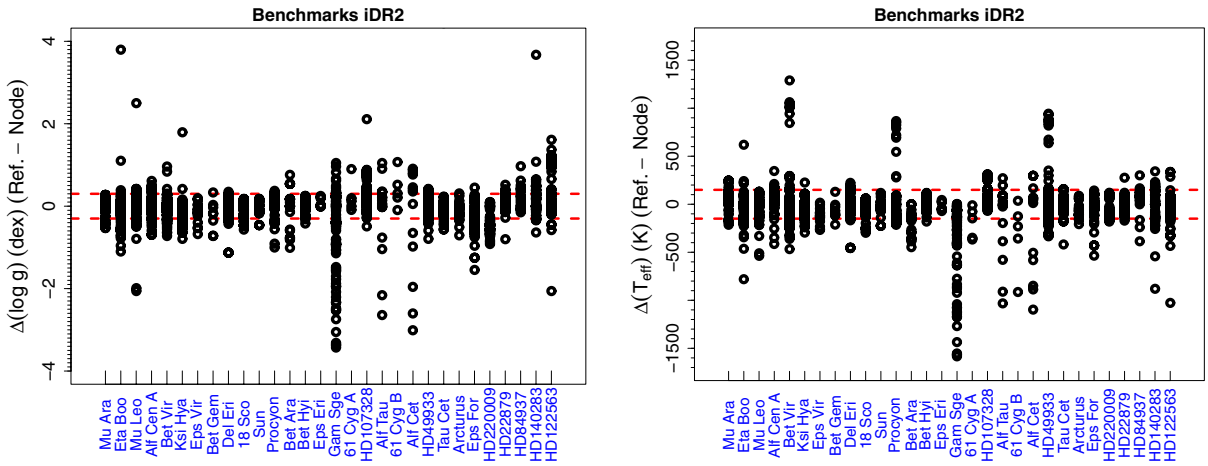

Fig. 8. All Node results for the 30 benchmark stars included in iDR2. The stars are ordered by decreasing metallicity from left to right. *Left panel*: difference between reference and Node value of log g. Red dashed lines indicate an interval of ±0.30 dex. *Right panel*: difference between reference and Node value of $T_{\rm eff}$. Red dashed lines indicate an interval of ±150 K.

For each Node, in each of the three areas of the parameter space, we calculate the average quadratic difference between the reference and the derived atmospheric parameters (only $T_{\rm eff}$ and log g) of the stars. If this average quadratic difference is within ±100 K and ±0.20 dex of the reference values, the Node results are considered to be very accurate (in that region of the parameter space). These average differences per Node are listed in Table 4. This table shows that different Nodes succeeded in analyzing a different number of stars in each region of the parameter space.

The ULB Node uncovered problems with their analysis during the process of homogenization. As there was no time to recompute the atmospheric parameters so close to the end of the analysis cycle, the Node decided to withdraw its results. The results from the Liège Node were found to be uncertain for the "metal-poor stars" group. The results from this Node for this region of the parameter space were not used and the values are not included in Table 4. The OACT Node did not analyze the metal-poor benchmarks. Their method needs observed spectra of metal-poor stars which are currently lacking in the library used as reference. The Nodes Bologna, CAUP, and UCM encountered other problems when analyzing these benchmark stars. As weights for the MPS region of the parameter space are not available for these Nodes, their results for metal-poor stars were not used.

Systematic biases are one component that can affect the accuracy of the results, making the results seem less accurate. They can in principle be corrected for, so that the unbiased results would agree better with the reference atmospheric parameters. For iDR2, however, bias correction was not implemented. This improvement will be implemented for future releases.

Figure 8 shows a comparison between all the Node results for the benchmark stars with respect to their reference $T_{\rm eff}$ and log g values. All results are shown, which includes the analysis of single exposure spectra of the stars, many of which have low S/N per pixel (<20). So the full range in the values displayed does not translate directly to the real uncertainty of the analysis. The final accuracy was judged only on the results for the final coadded spectra. Most of the results tend to be in reasonable agreement with the reference values, but outliers are present. Clear problems appear in some special cases: *i)* Gam Sge, Alf Cet, and Alf Tau are cool stars, with $T_{\rm eff} \lesssim 4000$ K, which almost all Nodes have difficulties in analyzing; *ii)* Procyon and Bet Vir have spectra with reduction problems; *iii)* Eta Boo and HD 49933 are relatively fast rotators ($v \sin i \geqslant 10\,{\rm km\,s^{-1}}$); and *iv)* very metal-poor stars HD 84937, HD 122563, and HD 140283.

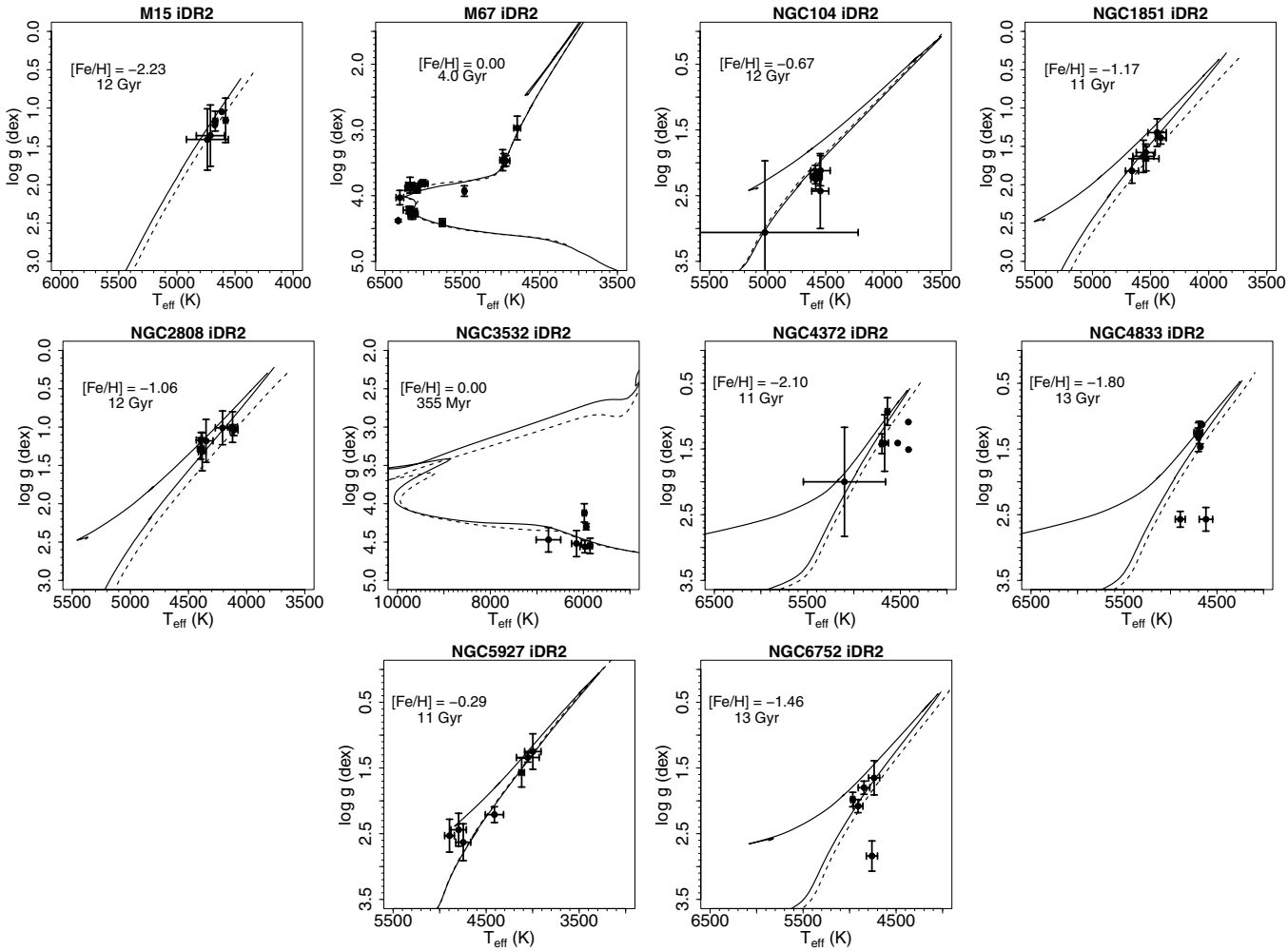

Fig. 9. iDR2 recommended parameters of the stars of the calibration clusters in the $T_{\rm eff}-\log g$ plane. No attempt was made to identify non member stars, i.e., the plots include all stars observed in the field of the clusters. Ages and metallicities were taken from the catalog of Harris (1996, and online updates) for the globular clusters and from the WEBDA database for the open clusters. The isochrones were computed with the web-tool of the PARSEC group (Bressan et al. 2012, all with solar-scaled composition), solid lines, and with BeSPP (Bellaterra Stellar Parameter Pipeline, Serenelli et al. 2013, α-enhanced below [Fe/H] = −0.80) which uses the GARSTEC stellar evolution code (Weiss & Schlattl 2008), dashed lines. Error bars represent the method-to-method dispersion of each atmospheric parameter (see Sect. 7.3).

This comparison already indicates the regions of the parameter space where the results derived here have increased uncertainty, i.e., very cool stars ($T_{\rm eff}$ < 4000 K), metal-poor stars ([Fe/H] ⩽ −2.0), and fast rotators.

7.2. Calibration clusters

A list of open and globular clusters are being observed by *Gaia*-ESO for calibration purposes (see Pancino et al., in prep.). Among other uses, they can serve as a second level of calibration to assess the physical consistency of the results. The calibration clusters used for iDR2 included the globular clusters M 15, NGC 104, NGC 1851, NGC 2808, NGC 4372, NGC 4833, NGC 5927, and NGC 6752 and the open clusters M 67 (with archival data), NGC 3532, and NGC 6705 (both with *Gaia*-ESO data). More of these calibration clusters have been and will be observed as the Survey progresses. They will be added to the calibration effort for future releases.

The observed stars were red giants in the globular clusters, cool main-sequence stars in NGC 3532, and AB-type stars in the open cluster NGC 6705[12]. The AB-type stars were selected to be used as a control sample for comparison between the analysis of FGK-type stars and the analysis of OBA-type stars (see details in Blomme et al., in prep.). Unfortunately, most of these stars turned out to be fast rotators and results for them were deemed uncertain and were excluded during quality control.

The physical consistency of the atmospheric parameters of cluster stars can be judged by comparing the derived values with those expected for an isochrone calculated with the chemical composition and age of that cluster. If the results follow an non physical relation in the diagram, they will be excluded and the Node results disregarded in that part of the parameter space. Although a few stars were identified for which the parameters do not exactly follow the isochrones, in most cases the agreement was deemed acceptable within the uncertainties. To illustrate that no grossly wrong parameters were found, Fig. 9 compares the final recommended atmospheric parameters of the stars observed in the calibration clusters with isochrones computed with literature values for age and metallicity. In these plots, we did not remove stars that might be nonmembers of the clusters.

[12] NGC 6705 was also observed for science goals, in this case the targets were FGK-type stars.

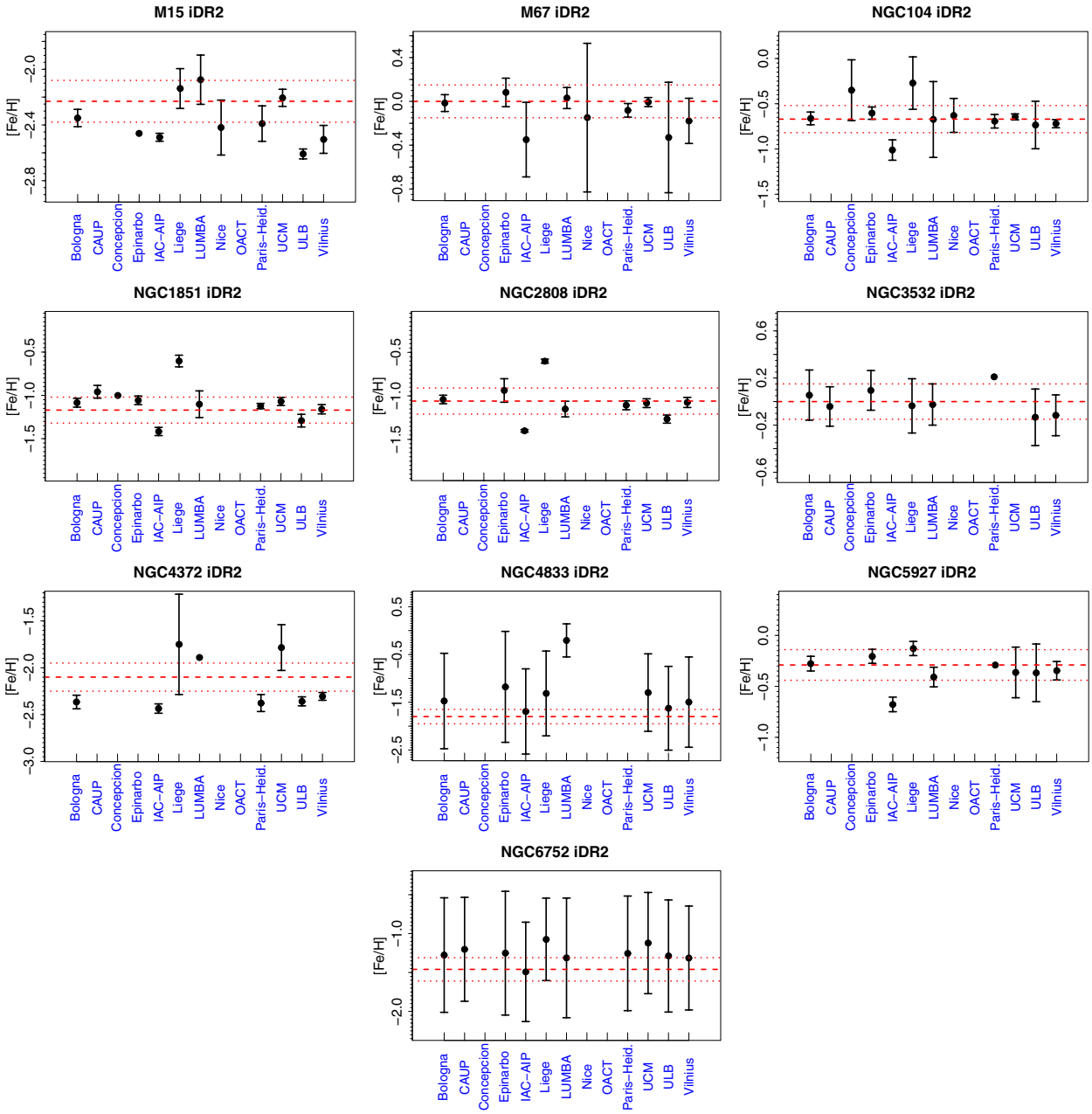

Fig. 10. Average metallicity obtained by the Nodes in comparison with a literature estimate of each calibrating cluster. Error bars indicate the standard error of the mean metallicity value of all stars in the cluster field analyzed by that given Node. Large error bars are thus caused by the presence of nonmember stars with very different metallicities. Different Nodes might have analyzed different number of stars in each cluster. Red dashed line is the literature metallicity of the cluster, taken from the catalog of Harris (1996, and online updates) for the globular clusters and from the WEBDA database for the open clusters. Dotted lines indicate a variation of ±0.15 dex in the metallicity.

That is part of scientific analyses that will be presented elsewhere. The results reproduce well the predicted slope of the red giant branches of the clusters. The M 67 open cluster is particularly interesting, as data of main sequence, turnoff, and giant stars was available. All these evolutionary regions are very well reproduced by the results. The differences between the two isochrone sets on the red giant branch are explained by the fact that PARSEC isochrones have solar-scaled composition, while GARSTEC isochrones are α-enhanced below [Fe/H] = −0.80, to be consistent with the α-enhancement observed in the metal-poor clusters.

Another consistency test is the metallicity determination for the stars in a given cluster. Assuming that all observed stars are cluster members and that there is no metallicity dispersion, it is expected that a given Node should recover very similar metallicities for all stars. Of course, these conditions are sometimes not fulfilled[13]. In Fig. 10, the mean metallicity obtained by each Node for each of the calibrating clusters is shown in comparison with the literature value. The error bars in the plot are the standard deviation of the mean. Nonmembers were not removed, and different Nodes were able to analyze a different number of stars in each cluster. Therefore, the understanding of each individual Node result in this plot is complex, but the general behavior is very informative. In most cases, the dispersion in the

[13] For example, NGC 1851 seems to present a small dispersion in metallicity and star-to-star variations of s-process elements (Yong & Grundahl 2008; Carretta et al. 2011).

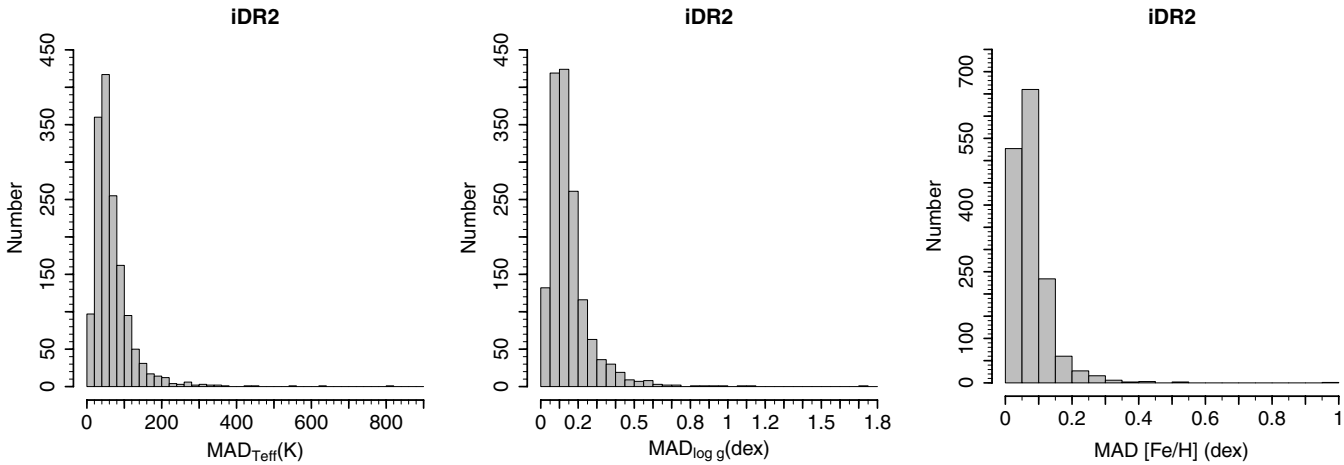

Fig. 11. Histograms showing the distribution of the method-to-method dispersion of the atmospheric parameters of the 1517 results obtained in iDR2 (some stars have multiple results, as single exposure spectra were analyzed sometimes). The dispersion is only computed if at least three Nodes provided results for that given star. *Left*: the dispersion of $T_{\rm eff}$. *Center*: the dispersion of log g. *Right*: the dispersion of [Fe/H].

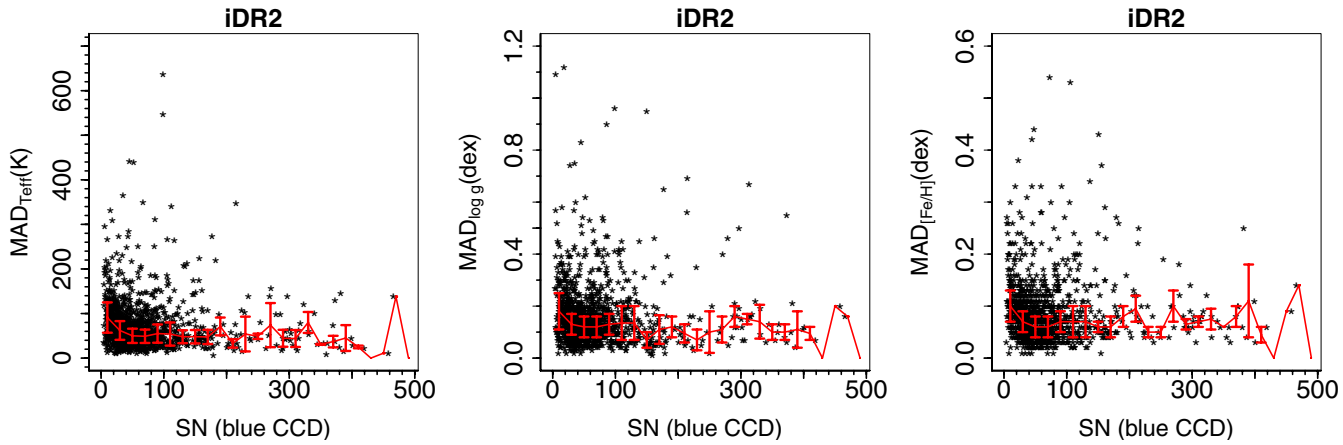

Fig. 12. Dependency of the method-to-method dispersion on the median S/N per pixel of the spectra. The S/N of the bluer part of the spectrum is used as reference. The red line connects the median value in each bin of S/N (in steps of 20). The error bars in the line represent the median absolute deviation.

metallicity values of a given Node is small and the average of the multiple Nodes agree within the dispersion bars. Cases like NGC 4833 and NGC 6752, where the dispersion within a given Node is large, are probably caused by nonmembers with very different metallicities from that of the cluster.

7.3. Method-to-method dispersion

To compare the results of different Nodes and quantify the method-to-method dispersion of each parameter we decided to use the median and the associated MAD (median absolute deviation). The MAD is defined as the median of the absolute deviations from the median of the data and is given by:

$$\mathrm{MAD} = \mathrm{median}_i(|X_i - \mathrm{median}_j(X_j)|). \quad (1)$$

For the iDR2 results, the histograms of the method-to-method dispersions are shown in Fig. 11. The median values of the method-to-method dispersion are 55 K, 0.13 dex, and 0.07 dex for $T_{\rm eff}$, log g, and [Fe/H], respectively. The third quartile of the distribution has values of 82 K, 0.19 dex, and 0.10 dex for $T_{\rm eff}$, log g, and [Fe/H], respectively. These values indicate an overall excellent agreement between the multiple methods for 75% of the results available in iDR2.

This agreement is obtained on absolute values of the parameters rather than on relative values, as we do not implement differential analyses. It should not be surprising that all the different methods do not yield exactly the same results given all the different factors involved in the analysis. Examples are the different ways to constrain the atmospheric parameters and the physics included in each different analysis code.

We recall here that the method-to-method dispersion is a measure of the precision of the results, i.e., the degree to which multiple methodologies can agree on the atmospheric parameters of a star. They are not the physical uncertainty of the values.

Figure 12 shows how the method-to-method dispersion of each atmospheric parameter ($T_{\rm eff}$, log g, and [Fe/H]) depends on the S/N of the spectrum. The plots show that there is a general trend toward larger disagreements being found for smaller values of S/N. Although outliers are found at any S/N value, the dispersion tends to increase only for the lowest values of SN (<40). Otherwise, for $S/N > 40$ it tends to stabilize around a constant value (~50 K, 0.13 dex, and 0.07 dex for $T_{\rm eff}$, log g, and [Fe/H], respectively). Perhaps more surprisingly, the plots also suggest that good agreement between different methods can be

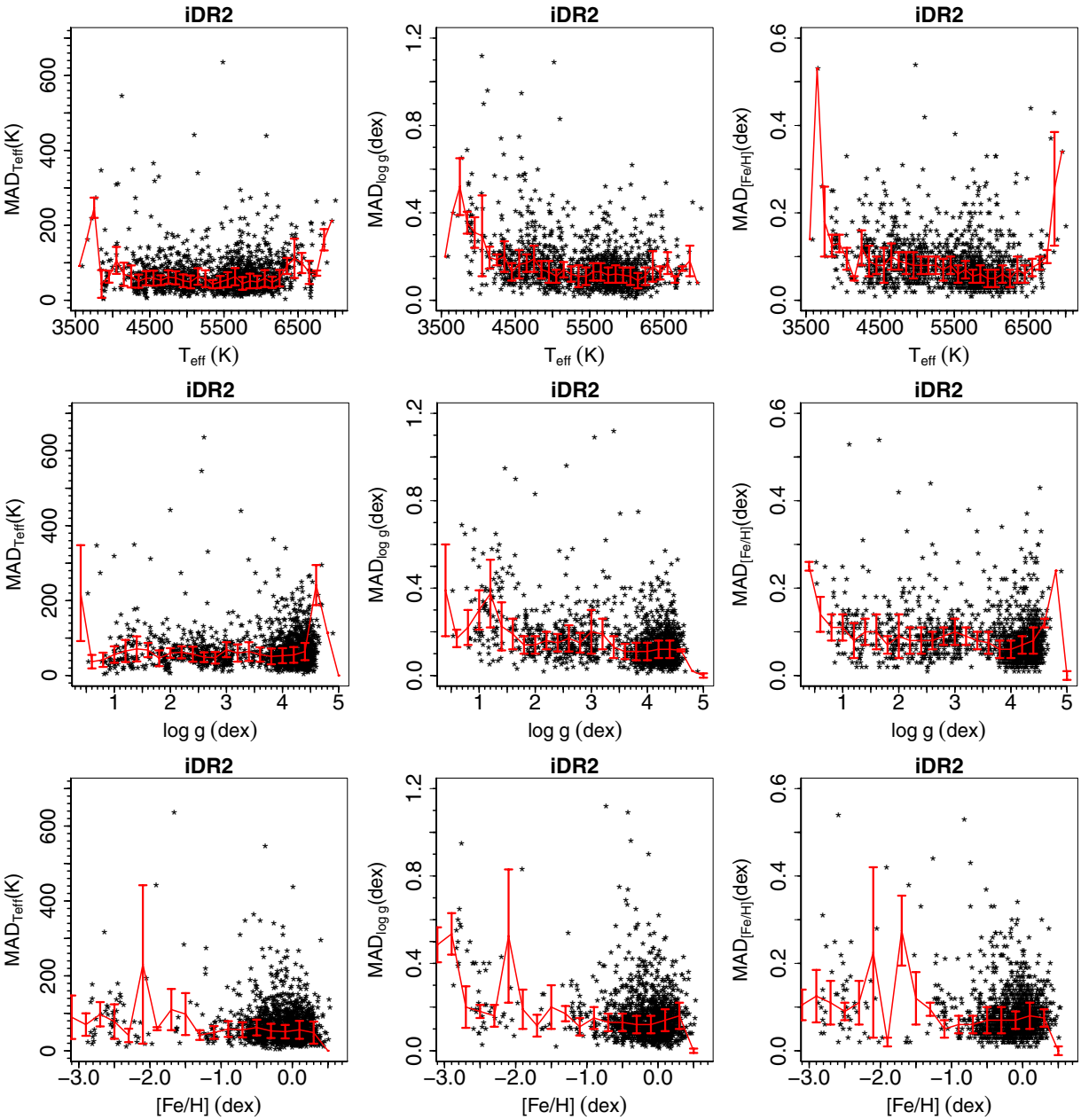

Fig. 13. Dependency of the method-to-method dispersion with respect to the atmospheric parameters of the stars. The red line connects the median value in each bin of 100 K (*top row*), 0.20 dex (*middle row*), and 0.20 dex (*bottom row*). The error bars in the line represent the median absolute deviation.

found even if the S/N is low, as the corners of low S/N and small dispersion in the panels are well populated.

Figure 13 shows how the method-to-method dispersion of each atmospheric parameter depends on the atmospheric parameters themselves. Most of the panels do not show any significant trend. There seems to be an increase in the dispersion of the log g values for cool stars (<4000 K) and for metal-poor stars ([Fe/H] < −2.00 or −2.50), although part of it might be caused by low-number statistics. This suggests that precise results are found across almost the full parameter range of the stars analyzed here. It also indicates that to select good results, cuts in the atmospheric parameters themselves are not needed, and that cuts in the dispersion values are sufficient. Overall, these comparisons show that the bulk of the results are of very good quality.

7.4. The recommended values

In this section, we describe the procedure used to define the recommended values of the atmospheric parameters of each star. The first step was a zeroth-order quality control of the results of each Node. Results that were excluded are those i) with very large error bars (above 900 K for $T_{\rm eff}$ and/or 1.50 dex for log g); ii) with microturbulence value equal to or below 0.00 km s^{-1}; iii) with surface gravity value above 5.00 dex; iv) where the final Node result was the same as the input values of the method, indicating that the automatic analysis failed to converge; and v) flagged as having other convergence problems.

Next, we used the results of the benchmark stars to weight the performance of each Node in the three different regions of the parameter space defined before: 1) metal-rich dwarfs; 2) metal-rich giants; and 3) metal-poor stars. For the benchmark stars

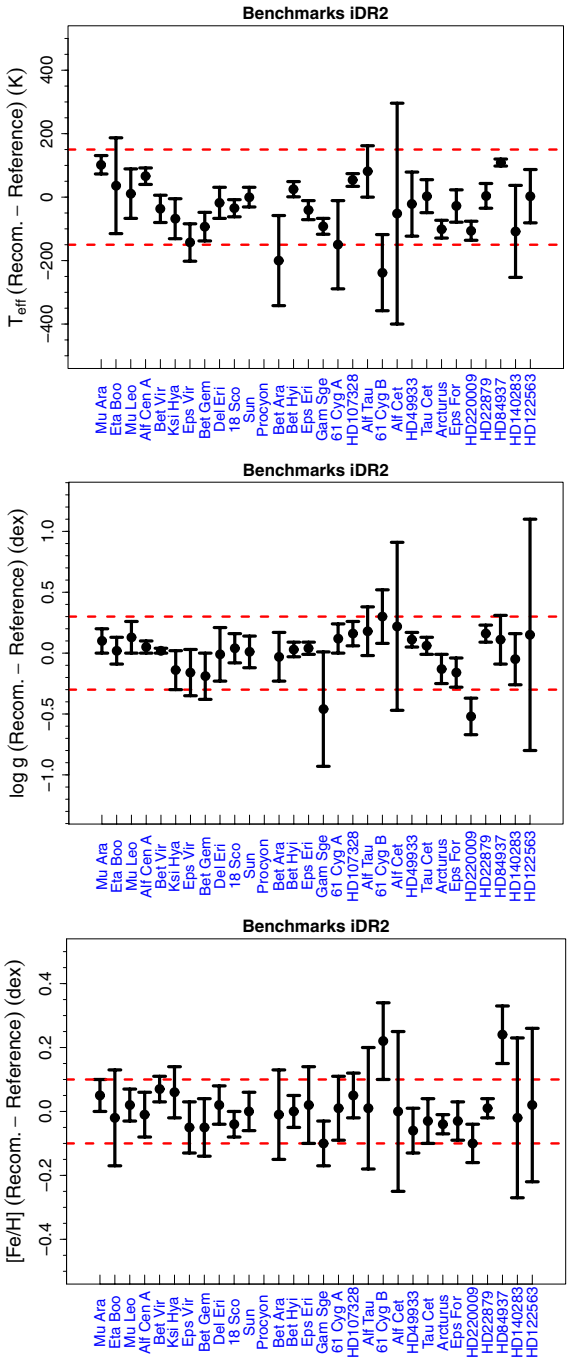

Fig. 14. Difference between the recommended values of $T_{\rm eff}$, log g, and [Fe/H] for the benchmark stars of iDR2 and the reference values. The error bars are the method-to-method dispersions. The stars are sorted in order of decreasing [Fe/H] (*left to right*). The dashed red lines indicate limits of ±150 K for $T_{\rm eff}$, of ±0.30 dex for log g, and of ±0.10 dex for [Fe/H].

in each one of these regions, we computed for each Node the average difference between the parameters it derived ($T_{\rm eff}$ and log g) and the reference parameters (Table 4).

These numbers are a measurement of the accuracy with which each Node can reproduce the reference atmospheric parameters in each region of the parameter space. They were then used to assign weights to the Node results. If the average difference of the Node results was within 100 K for $T_{\rm eff}$ and within 0.20 dex for log g, the Node was assigned a weight of *1.00*. Thus, we are assuming that all Nodes that reproduce the values

Table 5. Node weights per region of the parameter space.

Node	MRD	MRG	MPS
Bologna	1.000	0.546	–
CAUP	0.971	0.495	–
Concepcion	0.694	0.495	0.306
EPINARBO	1.000	0.781	0.585
IACAIP	0.862	0.901	0.935
Liege	0.676	0.386	–
LUMBA	1.000	0.602	0.758
Nice	0.870	0.794	1.000
OACT	0.741	0.585	–
ParisHeidelberg	1.000	0.746	0.637
UCM	0.893	0.214	–
ULB	–	–	–
Vilnius	1.000	0.457	0.308

within these margins are equally accurate and their results should be fully taken into account. Nodes that are less accurate than that are assigned worse weights, in a linear scale, by dividing the average difference of its parameters by 100 K or 0.20 dex, for $T_{\rm eff}$ and log g, respectively, and then averaging these values.

The weights are computed per Node and per region of the parameter space (Table 5). The results of each star are then combined in a weighted median, taking into account the Node weight of the parameter-space region to which they belong. For that, the multiple estimates of the parameter are ranked and interpolation used to find the 50% weighted percentile. The weighted median is:

$$\text{wei_median} = \text{Param}_k + \frac{50 - P_k}{P_{k+1} - P_k}(\text{Param}_{k+1} - \text{Param}_k), \quad (2)$$

where P_k is the percentile rank of parameter k:

$$P_k = \frac{100}{\text{Sum}_n}\left(\text{Sum}_k - \frac{\text{normal}_k}{2}\right), \quad (3)$$

where the weights are normalized on a star by star basis:

$$\text{Normal}_i = \frac{\text{weight}_i}{\sum_{i=1}^{n} \text{weight}_i}, \quad (4)$$

the total sum of weights is then:

$$\text{Sum}_n = \sum_{k=1}^{n} \text{weight}_k = 1.0, \quad (5)$$

and the partial sum of the weights is:

$$\text{Sum}_i = \sum_{k=1}^{i} \text{weight}_k. \quad (6)$$

Thus, for iDR2 the steps to obtain the recommended parameters can be summarized as:

1. A zeroth order quality control is performed, removing very uncertain results.
2. The accuracy of the Node results is judged using the available benchmark stars as reference. Weights are assigned, according to how well the Nodes can reproduce the reference values in a given region of the parameter space.
3. Further consistency tests are conducted using the calibration clusters.

Table 6. Outcome of the analysis of the iDR2 data.

Gaia-ESO type	Number of stars	Comment
Analyzed stars	1447	*Gaia*-ESO and archival data
Stars with results	1301	*Gaia*-ESO and archival data
Stars with results	1268	Only *Gaia*-ESO data
GES_MW	906	Milky Way fields
GES_CL	233	Open clusters fields
GES_SD	129	Calibration targets
AR	33	Archival data

Notes. Number of FGK-type stars observed with UVES with atmospheric parameters determined.

Table 7. Systematic errors of the atmospheric parameters for the iDR2 data set.

Type of star	$\sigma_{T_{\rm eff}}$	$\sigma_{\log\ g}$	$\sigma_{\rm [Fe/H]}$
Metal-rich dwarfs	50 K	0.10 dex	0.05 dex
Metal-rich giants	100 K	0.25 dex	0.05 dex
Metal-poor stars	50 K	0.15 dex	0.10 dex

4. The weighted-median value of the validated results is adopted as the recommended value of that parameter.
5. The MAD is adopted as an indicator of the method-to-method dispersion (analysis precision).
6. The number of results on which the recommended value is based is also reported.

Table 6 summarizes the number of stars for which atmospheric parameters were determined during iDR2. The analysis of about 10% of the stars was not completed for different reasons (e.g., high-rotation, double-lined signatures, too low S/N, emission lines).

A comparison of the recommended values of the atmospheric parameters of the benchmark stars (computed as described above) with the reference values is shown in Fig. 14. The recommended atmospheric parameters of the benchmark stars agree well with the reference values for the majority of the stars, i.e., within ±150 K for $T_{\rm eff}$, ±0.30 dex for log g, and ±0.10 dex for [Fe/H]. The results become more uncertain than that for cooler stars ($T_{\rm eff} \lesssim 4200$ K), as seen for HD 220009, Bet Ara, 61 Cyg B, Alf Cet, and Gam Sge.

The comparison with the benchmark stars together with the results for the clusters (see Fig. 9) illustrates the general good quality of the *Gaia*-ESO recommended results. These final recommended results are those whose use we advise for scientific publications. In Fig. 1, we show the final $T_{\rm eff}-\log g$ plane of the stars included in the iDR2 results.

7.5. Systematic errors

We estimate the systematic errors of the atmospheric parameters in iDR2 using the *Gaia* benchmark stars. These errors are a measurement of the systematic difference between reference and recommended values of the atmospheric parameters. In other words, they are the biases and measure the average accuracy of the *Gaia*-ESO atmospheric parameters. These values are provided in addition to the method-to-method dispersion, as they quantify a different kind of uncertainty of the results.

The systematic errors were computed in the three regions of the parameter space defined before (i.e., metal-rich dwarfs, metal-rich giants, and metal-poor stars). They are the average of the absolute value of the difference between the reference and recommended parameters for the benchmark stars in each of these regions. To avoid reporting unrealistically small values, we adopt as lower limit values of 50 K, 0.10 dex, and 0.05 dex for $T_{\rm eff}$, log g, and [Fe/H], respectively. We do that because: 1) we are reporting average values; 2) the benchmark stars were observed with much higher S/N than the typical *Gaia*-ESO target, and these values could be S/N dependent; and 3) the reference parameters themselves have errors, which were not taken into account in this calculation. The final values are listed in Table 7.

7.6. The effect of the number of Nodes

The number of Nodes contributing to the final recommended parameters varies from star to star. There are different reasons for that, including difficulties for a given method to deal with a certain kind of star. That raises the question of how homogeneous the results are as a whole.

To answer that question, we ran the following test for the stars that have results from ten or more Nodes. First, we randomly select a number of results for that star. Second, we compute what would be the final recommended parameters based on those selected results only, using the same weighted-median approach. We repeat the random selection 1000 times, to build a distribution of the final results (and to understand which results are more likely and what is the full range of possible values). The exercise was repeated, varying the number of Nodes contributing to the final results from three to eleven (12 is the maximum number of Nodes).

The results are plotted in Fig. 15 for two stars, HD 22879 and Tr20 781 (also discussed in Sect. 6). For each case of different number of Nodes, a boxplot with the distribution of the final weighted medians is shown. The red lines indicate the recommended parameter (using all available results) and its method-to-method dispersion.

The comparison shows that:

1. Irrespective of the number of Nodes used, most of the time, the weighted median of the random selection will agree with the final recommended value within the uncertainties.
2. Nevertheless, the fewer the number of Nodes used, the broader the distributions get, meaning that the chance of a spurious final recommended parameter increases.
3. When the number of Nodes increases, the distribution tends to get narrower.

These comparisons indicate that, if outliers are not present, the majority of the recommended results based on few Nodes will agree well with those based on many Nodes. Fluctuations on the final value are mostly within the uncertainties. This is a very important result lending confidence to our final recommended values. It stems from the effort to tie the final parameter scale to the

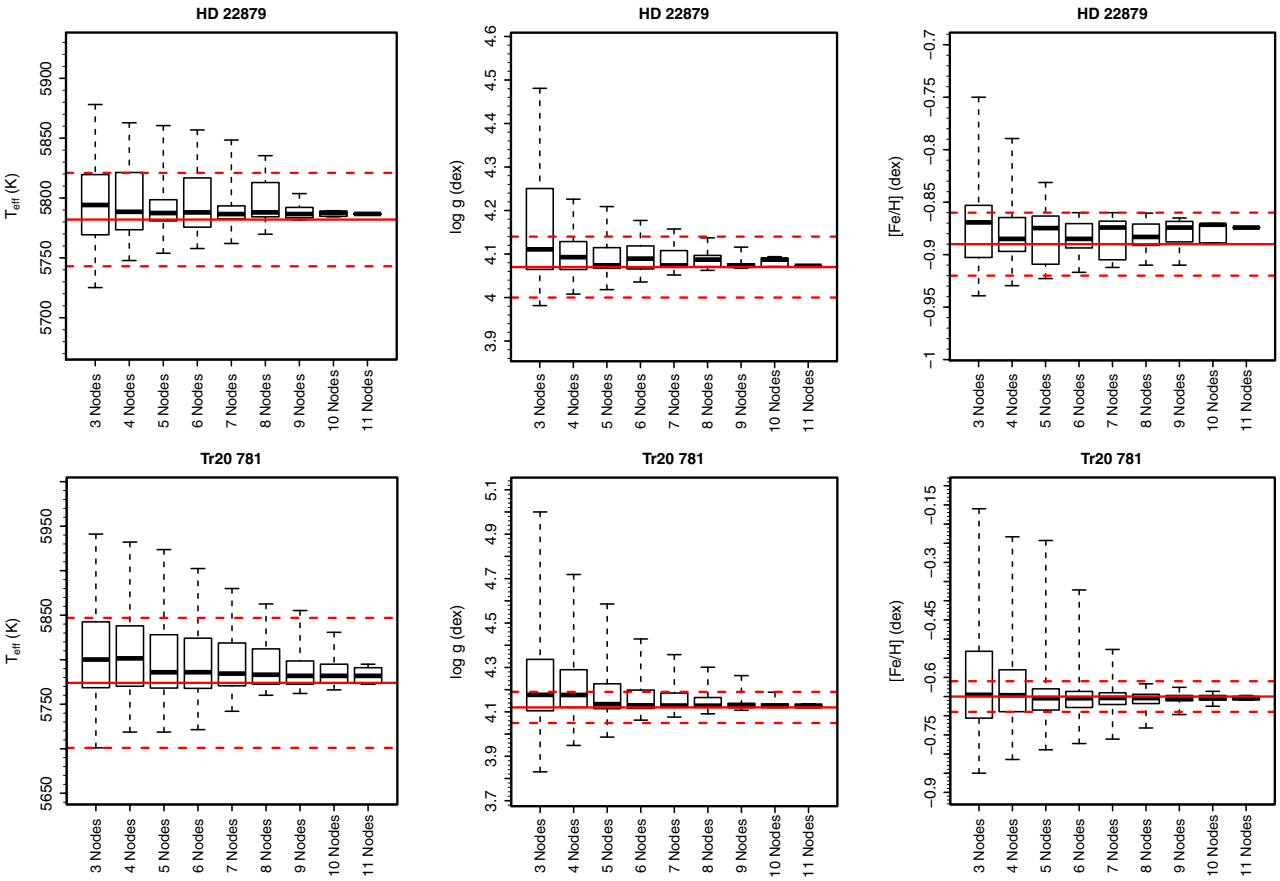

Fig. 15. Histograms of the median values of randomly selected Node results. These histograms display the effect on the recommended parameters caused by changing the number of Nodes that contribute to the final value (see text for full explanation). Red solid lines indicate the final recommended parameter, red dashed lines indicate the final 1σ method-to-method dispersion. The top panels show the case of HD 22879, the bottom panels the case of Tr20 781. In each boxplot, the thicker solid line indicates the median of the distribution, the box extends from the first to the third quartile, and the dashed lines extend to the extreme values.

Gaia benchmark stars. The results as a whole are homogeneous, within their quoted precision.

Nevertheless, some outlier Node results might be present. In the presence of outliers, the chance that the recommended value loses accuracy increases as fewer Nodes are used. For a large number of results, the median is a very robust measurement not affected by the presence of outliers.

The strength of using multiple analyses is highlighted here, as they help to uncover the outliers and minimize their effect on the final recommended value. A higher number of Nodes is also needed to better constrain the confidence on the precision of the final result.

To select the best quality results, cuts on the precision should be enough for most applications. When further accuracy is needed, we recommend a cut based on the number of Nodes providing results. This cut will remove most of the results that have a higher chance of being far from the correct parameter value. The plots and tests seem to indicate that by using five Nodes we can ensure that the majority of the results (>50%) will be close to the real value.

8. Abundances

For iDR2, eight different Nodes[14] computed abundances based solely on their own atmospheric parameters. This was the case because, as shown during the analysis of iDR1, there is no significant difference between these and abundances computed based on the recommended atmospheric parameters (as discussed in Appendix B).

As for the atmospheric parameters, the final recommended abundances are weighted medians from the values obtained by the Nodes. We combined the abundances on a line-by-line basis, adopting the same Node weights defined for the atmospheric parameters. We applied the following conditions to select the best results before combining the abundances, to guarantee that information was available to robustly estimate the precision of the results:

1. Only elemental species analyzed by at least three Nodes were considered.
2. The Node abundances of a given species, at a given star, were combined only if that star was analyzed by at least three Nodes.
3. Each spectral line was only considered if at least three Nodes provided abundances based on that line.
4. When information of the EWs was available, only lines with $5 \leq EW$ (mÅ) ≤ 120 were used. Exceptions were sodium ($5 \leq EW$ (mÅ) ≤ 140) and barium ($5 \leq EW$ (mÅ) ≤ 250)[15].

[14] The Nodes are: Bologna, CAUP, Concepcion, EPInArBo, LUMBA, Paris-Heidelberg, UCM, and Vilnius.

[15] As pointed out by the referee, selecting the lines for deriving abundances and for atmospheric parameters based on the measurements themselves may bias the results. At the lower EW limit, lines for which

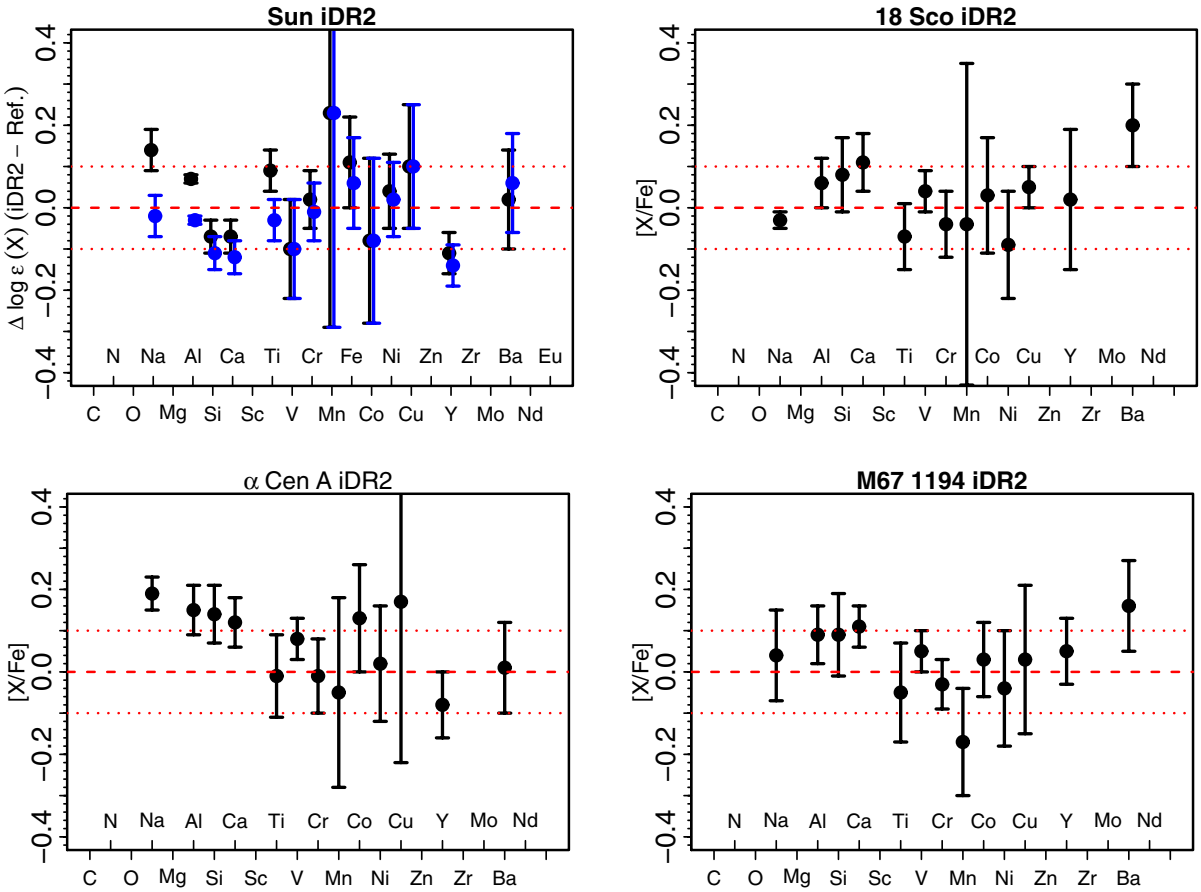

Fig. 16. Abundance pattern of the Sun and selected solar analogs. *Top left*: the results for the FLAMES solar spectrum for which $T_{\rm eff}$ = 5826 K, $\log g$ = 4.50, [Fe/H] = −0.03, and ξ = 1.05 km s^{-1} were derived. In black, the iDR2 results are compared to the solar abundances of Grevesse et al. (2007) and, in blue, to Grevesse & Sauval (1998). *Top right*: abundance pattern of the solar twin 18 Sco as derived here in comparison to the reference solar abundances computed in this work. For this star, the following atmospheric parameters were derived: $T_{\rm eff}$ = 5782 K, $\log g$ = 4.39, [Fe/H] = 0.05, and ξ = 1.04 km s^{-1}. *Bottom left*: abundance pattern of the solar analog α Cen A. The reference solar abundances are those derived here. For this star, the following atmospheric parameters were derived: $T_{\rm eff}$ = 5781 K, $\log g$ = 4.26, [Fe/H] = 0.25, and ξ = 1.21 km s^{-1}. *Bottom right*: abundance pattern of the solar twin in the M 67 open cluster, star YBP 1194, as derived here. The reference solar abundances are those derived here. For this star, the following atmospheric parameters were derived: $T_{\rm eff}$ = 5759 K, $\log g$ = 4.41, [Fe/H] = −0.01, and ξ = 1.06 km s^{-1}. In all plots, error bars are the method-to-method dispersion. Dashed lines represent $\Delta\log \epsilon(X)$ = 0.00 dex or [X/Fe] = 0.00 dex, dotted lines represent $\Delta\log \epsilon(X)$ = ±0.10 dex or [X/Fe] = ±0.10 dex. Abundances from this work are from the neutral species, except for Sc, Y, Zr, Ba, Nd, and Eu, for which the abundances are from the ionized species, and from C and N for which the abundances are from molecules.

5. If, for a given species at a given star, abundances from 20 or more different spectral lines were available, we removed those that are flagged as blended in the *Gaia*-ESO line list.
6. If, before applying the weighted median, the total number of spectral lines with abundances (for a given species at a given star) is more than 20, a 2σ clipping from the mean value was applied. (The total number of lines is counted across all Nodes, therefore if eight Nodes provide abundances for five lines each, it counts as 40 lines for the clipping.)
7. The weighted median abundance of each spectral line is computed.
8. The median value of multiple lines is adopted as the recommended abundance.

The exceptions are C, N, and O. One single Node (Vilnius) computed these abundances using the recommended atmospheric parameters. This choice was made to properly take into account the chemical equilibrium of the molecules. Carbon abundances were computed from C_2 molecules, nitrogen from CN molecules, and oxygen from the forbidden line at 6300 Å (see Sect. 5.1 for the line list description).

The iDR2 results include abundances computed in this way for the following 24 elements: C, N, O, Na, Mg, Al, Si, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Mo, Ba, Nd, and Eu, for a least a few stars[16]. Of the 1268 stars with atmospheric parameters observed by the Survey, we derived abundances of at least 15 different elements for 1079 and for at least ten elements for 1203 stars. This sample of FGK-type stars is already one of the largest of its kind in which abundances for so many elements have been determined. We stress that the list of abundances includes elements formed in different nucleosynthetic channels, i.e., s-process, r-process, Fe-peak, light, and α-elements, providing an unprecedented data set of great scientific value.

Abundances of a number of additional elements were provided by some Nodes, but were finally excluded when the

the inference overestimates the EW have a higher chance of being included than those with underestimated EW (and the opposite for the upper edge). This choice will lead to insignificant biases in high S/N data, but might become important for low S/N data and/or when the spectral lines are very weak.

[16] Abundances of Mo are available for 66 stars, of Nd for 111 stars, of Zr for 159 stars, and of Eu for 228 stars. All other abundances are available for more than 920 stars.

Table 8. Solar abundances derived in iDR2 in comparison to the solar abundances of Grevesse et al. (2007).

Element	Abundance This work	Abundance Grevesse et al.	Element	Abundance This work	Abundance Grevesse et al.
C	–	8.39	Mn	5.62 ± 0.52	5.39
N	–	7.78	Fe	7.56 ± 0.11	7.45
O	–	8.66	Co	4.84 ± 0.20	4.92
Na	6.31 ± 0.05	6.17	Ni	6.27 ± 0.09	6.23
Mg	–	7.53	Cu	4.31 ± 0.15	4.21
Al	6.44 ± 0.01	6.37	Zn	–	4.60
Si	7.44 ± 0.04	7.51	Y	2.10 ± 0.05	2.21
Ca	6.24 ± 0.04	6.31	Zr	–	2.58
Sc	3.29 ± 0.11	3.17	Mo	–	1.92
Ti	4.99 ± 0.05	4.90	Ba	2.19 ± 0.12	2.17
V	3.90 ± 0.12	4.00	Nd	–	1.45
Cr	5.66 ± 0.07	5.64	Eu	–	0.52

Notes. Abundances from this work are from the neutral species, except for Sc, Y, and Ba, for which they are from the ionized species.

conditions listed above were applied. These abundances are not part of the iDR2 recommended results because without multiple determinations it is not possible to estimate their precision. The list includes Li, S, Sr, La, Ce, Pr, Sm, Gd, and Dy. These abundances might still be used by the *Gaia*-ESO consortium for scientific applications, but they are not on the final *Gaia*-ESO iDR2 abundance scale. Instead, the abundances are on the scale defined by the individual Node results that they are part of. Whenever such abundances are used in a publication, this difference will be stressed.

8.1. The Sun and solar analogs

In Fig. 16, we show the abundance pattern of the Sun, and the solar twins/analogs 18 Sco (Porto de Mello & da Silva 1997), α Cen A, and M 67 1194[17] (Önehag et al. 2011), as computed here. The solar abundance pattern is compared to both the solar abundances of Grevesse & Sauval (1998) and Grevesse et al. (2007). The other stars are compared to the solar abundances derived in this work (the solar abundances are given in Table 8). With a few exceptions, the solar values derived here agree with reference solar abundances to within ±0.10 dex. The abundances of the three other stars mostly agree with the solar abundances to within ±0.10 dex. For some elements solar abundances are not listed, either because of weak lines (e.g., CNO) or because they were computed by a reduced number of Nodes, and therefore did not fulfil the criteria discussed above for combining the abundances. Where solar abundances were not derived, we recommend the use of those from Grevesse et al. (2007) for compatibility with the adopted model atmospheres.

8.2. Trends with metallicity

In Fig. 17, we show the trend with metallicity for the [X/Fe] ratio of a few selected elements. All elements display a behavior with metallicity in agreement with what has been established by earlier works (see, e.g., Edvardsson et al. 1993; Venn et al. 2004; Soubiran & Girard 2005; Reddy et al. 2006; Adibekyan et al. 2012; Bensby et al. 2014, and references therein). In these plots, we selected only the best quality results, excluding abundances where the method-to-method dispersion is above 0.20 dex. The inclusion of these extra abundances tends to increase the scatter in each plot. The figures are only included to illustrate which abundances have been derived, and that the general behavior seems correct. Proper scientific discussion requires a full investigation, which is not the goal of this release paper.

8.3. Iron abundances and metallicities

The data products determined here include both an [Fe/H] value determined during the derivation of the atmospheric parameters and abundances derived from Fe I and Fe II lines. The [Fe/H] value is a combination of the values used by the Nodes to constrain the metallicity of the model atmosphere adopted for a given star. For some methodologies, this metallicity is the Fe abundance, while for others it is a global value of the metal content, referred to as [M/H]. For deriving the recommended value of the metallicity as an atmospheric parameter, no distinction was made between [Fe/H] and [M/H]. For about 75% of the stars, results of eight or more Nodes were used to compute [Fe/H].

The abundances derived from Fe I and Fe II lines are calculated using the line-by-line abundances of the Nodes. Only five of the Nodes provided abundances of the iron lines. Since the final values of the Fe abundances and of [Fe/H] are computed in different ways, it is important to check if they are consistent. Such a comparison is shown in Fig. 18. We only show stars where the method-to-method dispersion of the Fe I abundances is below 0.20 dex. Dwarfs and giants are displayed separately to check for possible systematic effects in stars of different evolutionary stages.

If the [Fe/H] values are used to compute Fe abundances, adopting $\log \epsilon(\mathrm{Fe})_\odot = 7.45$ from Grevesse et al. (2007), an average offset between 0.05 and 0.08 dex is found with respect to the listed Fe I abundance, with a scatter of the order of 0.10 dex. This offset is of similar magnitude to the difference between the Solar Fe I abundance derived here ($\log \epsilon(\mathrm{Fe})_\odot = 7.56$, Table 8) and the value from Grevesse et al. (2007). In other words, if the stellar Fe I abundance is used with our Solar Fe abundance, the [Fe/H] would agree with the [Fe/H] value of the atmospheric parameters. The offset between our Fe abundances and the Fe abundance of Grevesse et al. seems to be constant throughout the whole sample, except for metal-poor stars. Therefore, we again advise that care is needed when using the results for the metal-poor stars. Some of the most metal-rich stars (dwarfs and giants) also show a poor agreement. These are difficult cases to analyze

[17] Star NGC 2682 YBP 1194 with identification number from Yadav et al. (2008).

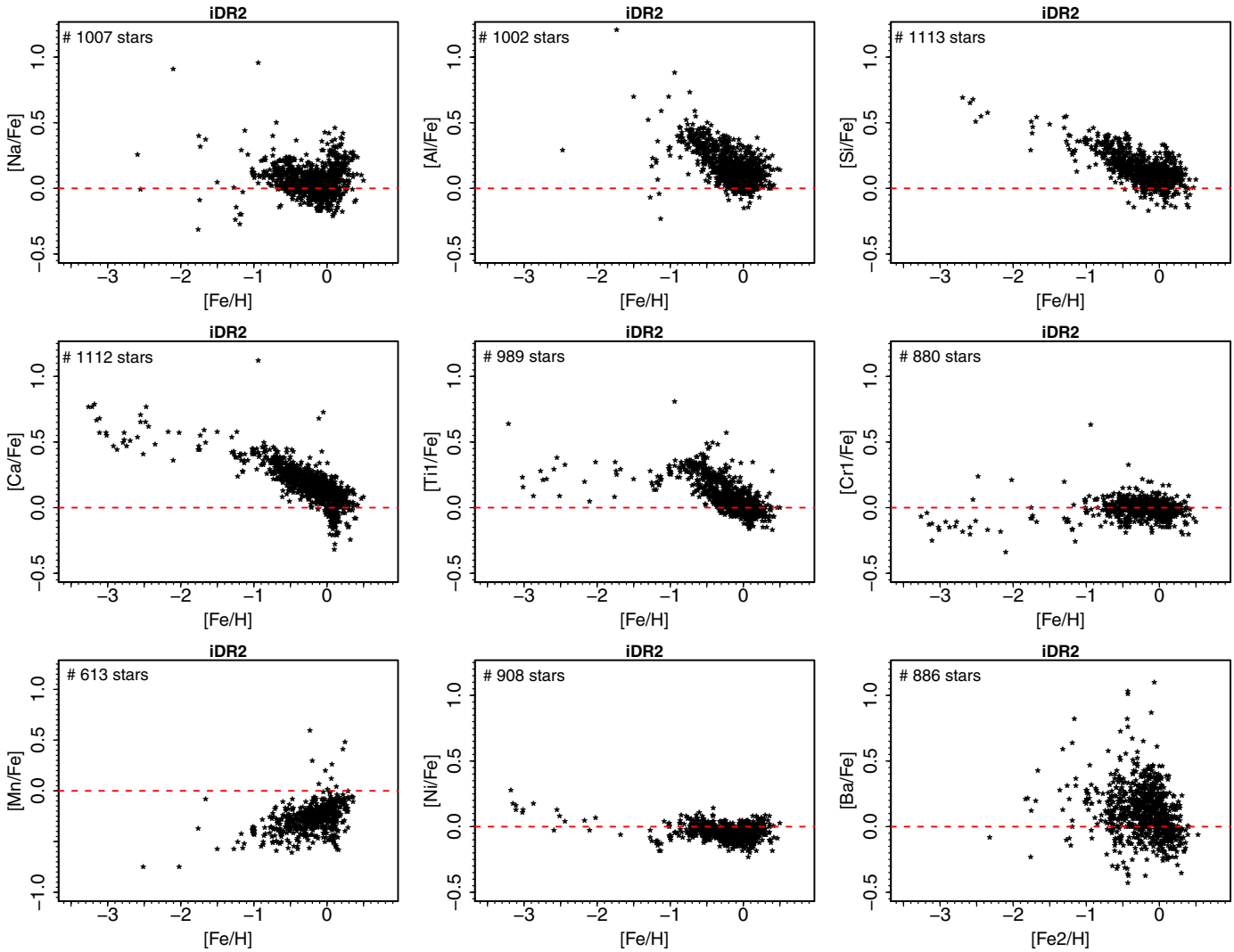

Fig. 17. Abundances trends with metallicity for a few selected elements. Only results where the method-to-method dispersion is below 0.20 dex are plotted. Note that the Mn I plot has a different scale. All abundances shown are from the neutral species, except for the Ba plot, where Ba II and Fe II are used.

because of the increased importance of line blends and should also be treated with care. We are working on improving the analysis for future releases and expect improvements for these stars.

Figure 19 compares the average Fe abundances obtained from Fe I and Fe II lines in stars where the method-to-method dispersion of both Fe I and Fe II is below 0.20 dex. Dwarfs and giants are displayed separately, but the general behavior is similar. A good agreement exists between Fe I and Fe II values for almost all stars. Average offsets are small (~0.02−0.03 dex) and the scatter also seems to be within the uncertainties except, once again, for the most metal-poor stars of the sample. Ionization equilibrium is, however, an invalid assumption for metal-poor giants because of non-LTE effects and possible departures of real atmospheres from model atmospheres. As was seen previously in Table 4, most of the EW methods, which enforce ionization equilibrium, failed in the analysis of the metal-poor benchmark stars. Therefore, their results in this region of the parameter space were not used. The EW methods that did manage to perform the analysis show a huge difference between the log g derived enforcing ionization equilibrium and the fundamental log g of the benchmark stars. The methods that do not enforce ionization equilibrium (the ones that look for best fitting synthetic spectra) better reproduce the real log g of the benchmark stars. Therefore, the lack of agreement between Fe I and Fe II is likely to be the correct behavior and not a problem.

8.4. Method-to-method dispersion

The method-to-method dispersion of the abundances can be used as an indicator of the precision with which the results were derived. In Fig. 20, we show the histogram of the MADs of a few selected elements. In most cases, the majority of the results show very good agreement among the multiple determinations. Usually, the majority of the results have MAD below 0.20 dex. The agreement is worse for ionized species (like Ti II, Cr II, Ba II, and Eu II) and/or those that have important hyperfine structure (like Mn and Cu).

In Fig. 21, we show the behavior of the method-to-method dispersion of Ti II as a function of the atmospheric parameters. No correlation is apparent in these plots and the behavior is the same for the other ionized species. The surface gravity is perhaps the parameter that is harder to constrain (see Sect. 7.3). Deviations from the ionization equilibrium, problems with the lines of FeII (which are usually weaker and/or blended), and issues with atomic data are probably behind the increased

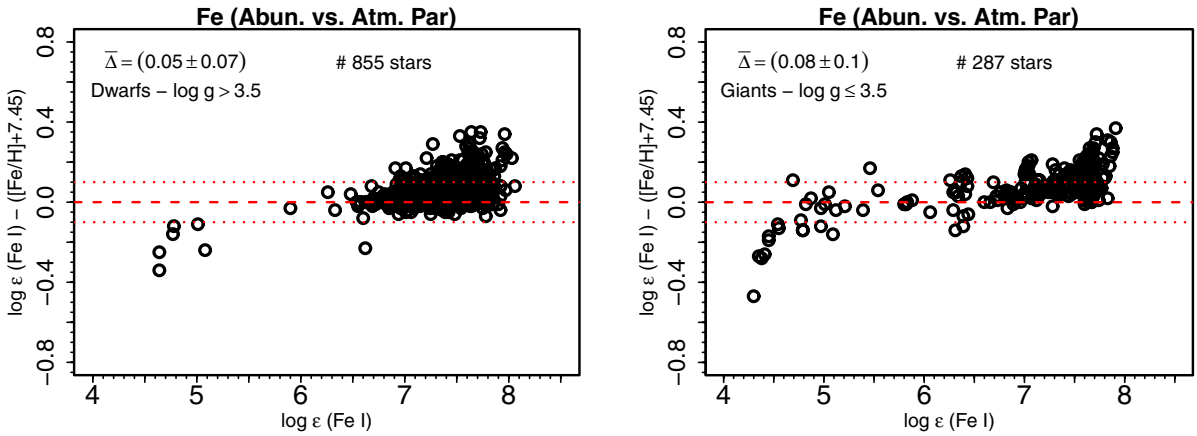

Fig. 18. Comparison between the abundances of Fe I and the metallicity as an atmospheric parameter [Fe/H], in dwarfs (*left panel*) and giants (*right panel*). The metallicity [Fe/H] is put in the log ϵ scale by adding the solar Fe abundance from Grevesse et al. (2007), log $\epsilon(\mathrm{Fe})_{\odot} = 7.45$. Only results where the method-to-method dispersion of Fe I is 0.20 dex or less are shown. Dotted lines indicate a difference of ±0.10 dex, dashed lines represents Fe I equal to [Fe/H].

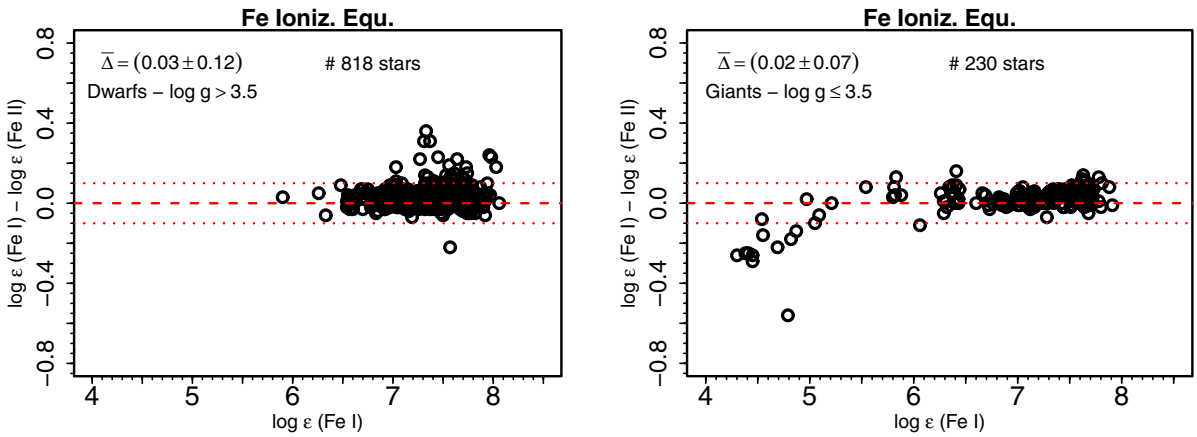

Fig. 19. Comparison between the abundances of Fe I and Fe II, to check for the ionization equilibrium, in dwarfs (*left panel*) and giants (*right panel*). Only results where the method-to-method dispersion of both Fe I and Fe II is 0.20 dex or less are shown. Dotted lines indicate a difference of ±0.10 dex, dashed lines represents Fe I equal to Fe II.

method-to-method dispersion of these elements. We are working on improving the log g determination for future releases. In particular, the Survey observed stars in the CoRoT fields, for which asteroseismic log g values are being determined, and those will be used as reference for the next releases. With better surface gravity values, we expect more precise abundances of the ionized species.

For the iDR2 abundances, we adopt the MADs as the typical uncertainties. This is akin to using the standard deviation of multiple lines of the same element, as is commonly done in the literature. The third quartile of the method-to-method dispersion distribution is below 0.15 dex for Na I, Al I, Si I, Ca I, Sc II, and Ba II. These are the elements for which the quality of the abundances is higher. For the other elements, the third quartile of the method-to-method dispersion distribution is between 0.15 and 0.20 dex for Mg I, Ti I, V I, Zr I, Mo I; between 0.20 and 0.25 dex for Sc I, Ti II, Cr I, Cr II, Ni I, Cu I, Y II, Zr II. For the remaining species, Mn I, Co I, Zn I, Nd II, and Eu II, it is between 0.25 and 0.35 dex. These last elements have more uncertain abundances and should be used with care.

The most robust abundances are those where the method-to-method dispersion is smaller, as that means that the abundances computed by different groups agreed well. Therefore, we recommend that abundances for scientific purposes are chosen carefully, taking the method-to-method dispersion into account. An upper value of 0.20 dex seems to be a reasonable compromise between number of stars and precision. More stringent cuts should be considered if needed.

9. Summary

This paper describes the analysis of high-resolution UVES spectra of FGK-type stars obtained by the *Gaia*-ESO Survey. The analyses of other types of stars and/or spectra are described elsewhere (Blomme et al., in prep.; Lanzafame et al. 2014; Recio-Blanco et al., in prep.). These data are used to derive precise and accurate values of atmospheric parameters and detailed elemental abundances.

Multiple methods are used to determine these quantities. A single pipeline would be more homogeneous internally but, in a broad survey like *Gaia*-ESO, might introduce different systematics in different regions of the parameter space. The parameter scale is tied to the scale defined by the *Gaia* benchmark stars, a set of well-studied stars with fundamental atmospheric parameters determined independently from spectroscopy. In addition, a set of open and globular clusters is used to evaluate the physical soundness of the results. Each of the implemented methodologies is judged against the benchmarks to define weights in three

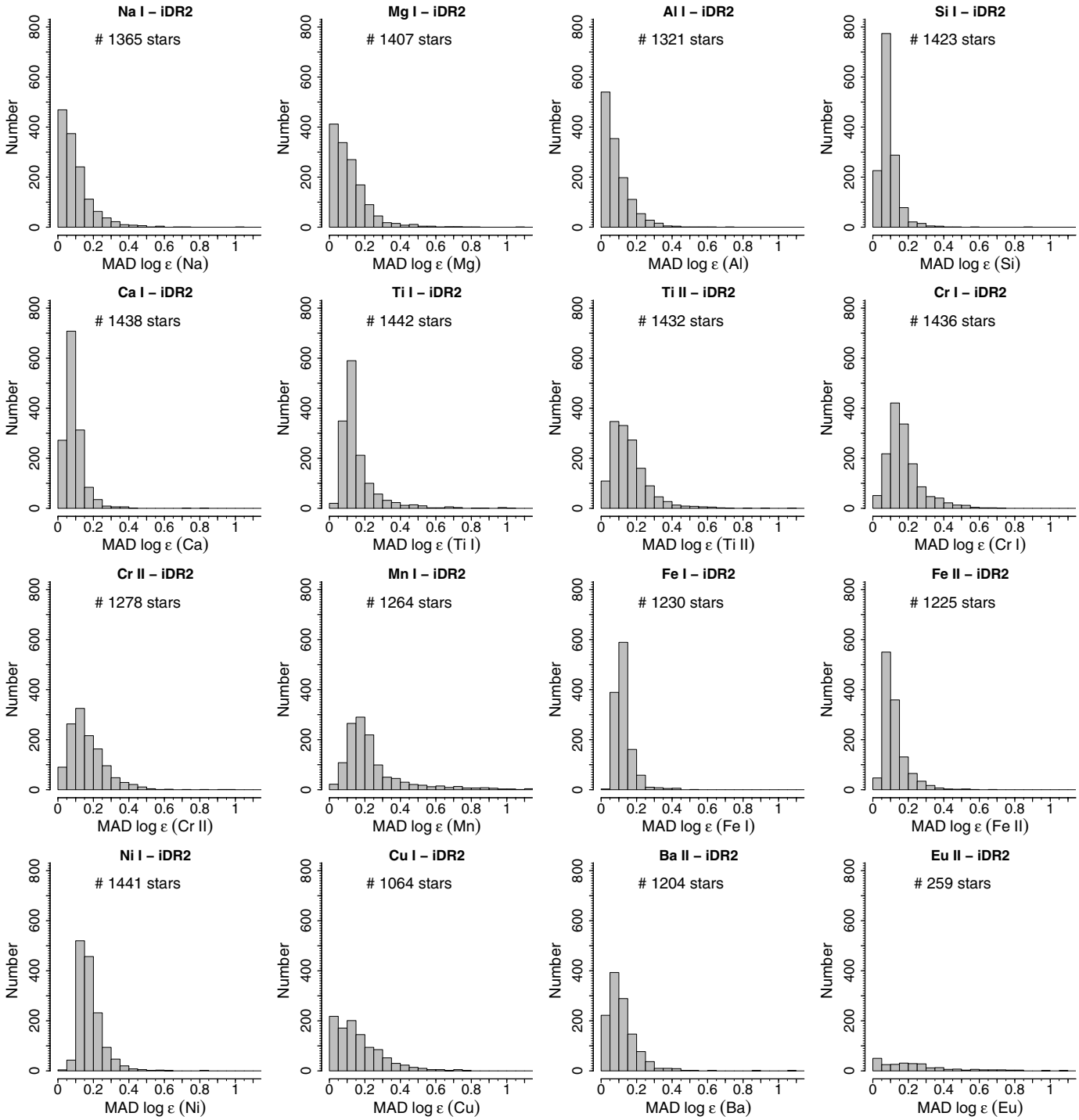

Fig. 20. Histograms with the method-to-method dispersion of selected species included in the iDR2 results.

different regions of the parameter space: i) metal-rich dwarfs; ii) metal-rich giants; and iii) metal-poor stars. The final recommended results are the weighted medians of those from the individual methods. We quantify the precision of the results by means of the method-to-method dispersion, a unique *Gaia*-ESO product. These results are only possible thanks to the massive combined efforts of all the scientists involved in the spectrum analysis and would be hard to quantify outside such a large collaboration.

The work described here is part of the analysis effort conducted to prepare the upcoming public catalog of *Gaia*-ESO advanced data products. The analysis of two iDRs has been completed. These internal data releases happen roughly every six months when a new analysis cycle is launched. The full analysis cycle takes between three and four months to be completed. The data products from the iDR3 analysis (at the time of writing still ongoing) will be included in the public release alongside the results of the iDR2 analysis.

Only the best recommended parameters and abundances, processed as described here, and later subjected to the final Survey wide homogenization (see François et al., in prep.), will be present in the public catalog. The identification, presentation, and discussion of individual scientific topics based on these results is left to the many scientific teams involved in the

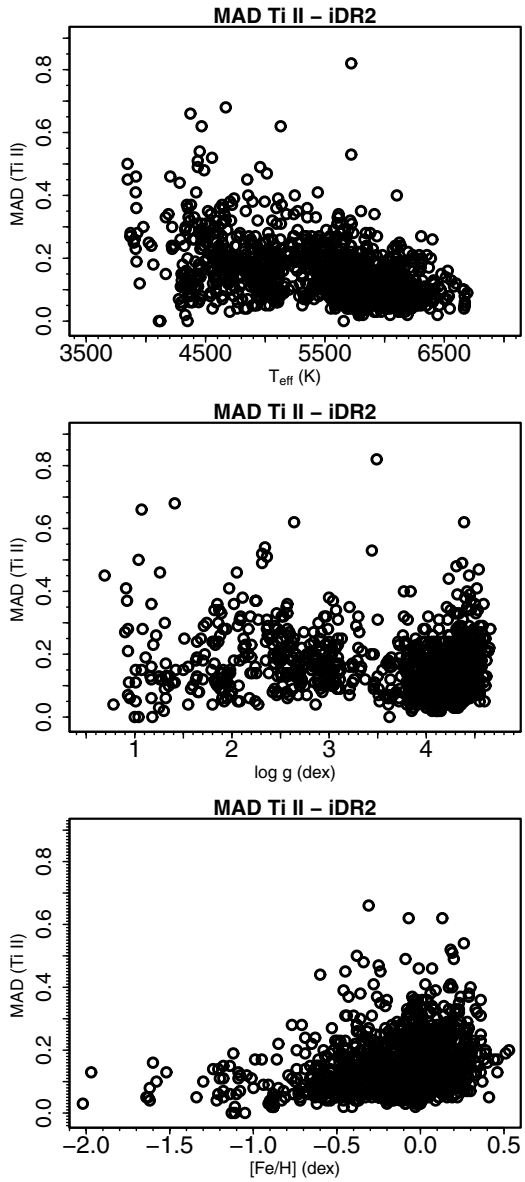

Fig. 21. Method-to-method dispersion of the Ti II abundances as a function of the atmospheric parameters ($T_{\rm eff}$, log g, and [Fe/H]). The dispersion of other ionized species behave in a similar way.

Gaia-ESO Survey. Because of that, we refrained from presenting indepth scientific discussions using the results described here. The tables with public release results will be available through the ESO data archive[18], as is already the case for the first batch of reduced *Gaia*-ESO spectra, as well as through a dedicated *Gaia*-ESO Survey science archive[19] hosted by the Wide Field Astronomy Unit (WFAU) of the Institute for Astronomy, Royal Observatory, Edinburgh, UK. This science archive is designed to provide functionalities beyond those available at the ESO archive.

In the latest internal release (iDR2), we derived atmospheric parameters of 1301 FGK-type stars. For 75% of these stars, the multiple determinations of $T_{\rm eff}$, log g, and [Fe/H] for the same star agree to better than 82 K, 0.19 dex, and 0.10 dex, respectively. The tests and comparisons presented here indicate that care is needed with the results of both cool ($T_{\rm eff}$ < 4000 K) or metal-poor stars ([Fe/H] < −2.00). Abundances for 24 elements were derived in at least a few stars: C, N, O, Na, Mg, Al, Si, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Mo, Ba, Nd, and Eu. We derived abundances of at least 15 different elements for 1079 stars and for at least ten elements for 1203 stars. For the abundances, the majority of the multiple determinations agree to better than 0.20 dex.

The list includes abundances of elements formed in all nucleosynthetic channels. This highlights the unique value of the sample being analyzed here. This is only possible thanks to the high-quality, high-resolution, and large wavelength coverage of the UVES data. The exciting potential of these results is exemplified by the variety of early science papers being produced by the *Gaia*-ESO collaboration (e.g., Bergemann et al. 2014; Cantat-Gaudin et al. 2014b; Donati et al. 2014; Friel et al. 2014; Magrini et al. 2014; Spina et al. 2014).

The value of the *Gaia*-ESO science products will be further enhanced when the results of the *Gaia* mission (Perryman et al. 2001) become available. *Gaia* will provide parallaxes, proper motions, and spectrophotometric metallicities for ~10^9 stars and radial velocities for ~150 million stars. The chemical information coming from the *Gaia* spectra are, however, limited: metallicities ([Fe/H]) and abundances for a few elements, mostly α-elements (Ca, Si, Ti), will be obtained for roughly two million stars brighter than $V \leqslant 12-13$ mag (Wilkinson et al. 2005). Abundances of elements formed by other nucleosynthetic channels (s-process, r-process, Fe peak elements, light elements) in fainter stars, covering a larger volume in the Galaxy, require additional observations from ground-based observatories, such as those being carried out within the ongoing *Gaia*-ESO Survey.

The sample of high-resolution spectra of FGK-type stars discussed here is already among the largest of its kind analyzed in a homogeneous way. The results will enable significant advances in the areas of stellar evolution and Milky Way formation and evolution.

Acknowledgements. R.Sm., through an ESO fellowship, has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 229517. A.J.K. and U.H. acknowledges support by the Swedish National Space Board (SNSB) through several grants. D.G. gratefully acknowledges support from the Chilean BASAL Centro de Excelencia en Astrofisica y Tecnologias Afines (CATA) grant PFB-06/2007. E.C., H.G.L., L.Sb. and S.D. acknowledge financial support by the Sonderforschungsbereich SFB 881 "The Milky Way System" (subprojects A2, A4, A5) of the German Research Foundation (DFG). I.S.R. gratefully acknowledges the support provided by the Gemini-CONICYT project 32110029. L.Sb. and S.D. acknowledge the support of Project IC120009 "Millennium Institute of Astrophysics (MAS)" of Iniciativa Científica Milenio del Ministerio de Economía, Fomento y Turismo de Chile. M.V. acknowledges financial support from Belspo for contract PRODEX COROT. P.B. acknowledges support from the PNCG of INSU CNRS. Part of the computations have been performed on the "Mesocentre SIGAMM" machine, hosted by Observatoire de la Cote d'Azur. P.d.L., V.H. and A.R. acknowledge the support of the French Agence Nationale de la Recherche under contract ANR-2010-BLAN- 0508-01OTP and from the "Programme National de Cosmologie et Galaxies" (PNCG) of CNRS/INSU, France. S.G.S, E.D.M., and V.Zh.A. acknowledge support from the Fundação para a Ciência e Tecnologia (Portugal) in the form of grants SFRH/BPD/47611/2008, SFRH/BPD/76606/2011, SFRH/BPD/70574/2010, respectively. S.Vi. gratefully acknowledges the support provided by FONDECYT reg. n. 1130721. T.B. was funded by grant No. 621-2009-3911 from The Swedish Research Council. T.Mo. acknowledges financial support from Belspo for contract PRODEX GAIA-DPAC. We acknowledge the support from INAF and Ministero dell' Istruzione, dell' Università' e della Ricerca (MIUR) in the form of the grant "Premiale VLT 2012". This work was partly supported by the European Union FP7 program through ERC grant number 320360 and by the Leverhulme Trust through grant RPG-2012-541. The results presented here benefit from discussions held during the *Gaia*-ESO workshops and conferences supported by the ESF (European Science Foundation) through the GREAT Research Network Program. This research has made use of the SIMBAD database, operated at CDS, Strasbourg,

[18] http://archive.eso.org/wdb/wdb/adp/phase3_spectral/form?phase3_collection=GaiaESO

[19] http://ges.roe.ac.uk/index.html

France, of NASA's Astrophysics Data System, of the compilation of atomic lines from the Vienna Atomic Line Database (VALD), and of the WEBDA database, operated at the Department of Theoretical Physics and Astrophysics of the Masaryk University.

References

Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012, A&A, 545, A32
Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17
Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, ApJ, 636, 804
Alonso, A., Arribas, S., & Martínez-Roger, C. 1999, A&AS, 140, 261
Alvarez, R., & Plez, B. 1998, A&A, 330, 1109
Anstee, S. D., & O'Mara, B. J. 1991, MNRAS, 253, 549
Baglin, A., Auvergne, M., Boisnard, L., et al. 2006, in COSPAR Meeting, 36th COSPAR Scientific Assembly, 36, 3749
Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, A&A, 559, A74
Barklem, P. S., & O'Mara, B. J. 1998, MNRAS, 300, 863
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Bergemann, M., Lind, K., Collet, R., Magic, Z., & Asplund, M. 2012, MNRAS, 427, 27
Bergemann, M., Ruchti, G. R., Serenelli, A., et al. 2014, A&A, 565, A89
Blanco-Cuaresma, S., Soubiran, C., Jofré, P., & Heiter, U. 2014, A&A, 566, A98
Boyajian, T. S., von Braun, K., van Belle, G., et al. 2012, ApJ, 757, 112
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Burbidge, E. M., Burbidge, G. R., Fowler, W. A., & Hoyle, F. 1957, Rev. Mod. Phys., 29, 547
Cantat-Gaudin, T., Donati, P., Pancino, E., et al. 2014a, A&A, 562, A10
Cantat-Gaudin, T., Vallenari, A., Zaggia, S., et al. 2014b, A&A, 569, A17
Carretta, E., Lucatello, S., Gratton, R. G., Bragaglia, A., & D'Orazi, V. 2011, A&A, 533, A69
Casagrande, L., Ramírez, I., Meléndez, J., Bessell, M., & Asplund, M. 2010, A&A, 512, A54
Cayrel, R. 1988, in The Impact of Very High S/N Spectroscopy on Stellar Physics, eds. G. Cayrel de Strobel, & M. Spite, IAU Symp., 132, 345
Cayrel, R., Depagne, E., Spite, M., et al. 2004, A&A, 416, 1117
de Laverny, P., Recio-Blanco, A., Worley, C. C., & Plez, B. 2012, A&A, 544, A126
Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, in Optical and IR Telescope Instrumentation and Detectors, eds. M. Iye, & A. F. Moorwood, SPIE Conf. Ser., 4008, 534
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, Res. Astron. Astrophys., 12, 735
Donati, P., Cantat Gaudin, T., Bragaglia, A., et al. 2014, A&A, 561, A94
Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101
Frasca, A., Alcalá, J. M., Covino, E., et al. 2003, A&A, 405, 149
Frasca, A., Guillout, P., Marilli, E., et al. 2006, A&A, 454, 301
Friel, E. D., Donati, P., Bragaglia, A., et al. 2014, A&A, 563, A117
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
González Hernández, J. I., & Bonifacio, P. 2009, A&A, 497, 497
Grevesse, N., & Sauval, A. J. 1998, Space Sci. Rev., 85, 161
Grevesse, N., Asplund, M., & Sauval, A. J. 2007, Space Sci. Rev., 130, 105
Guillout, P., Klutsch, A., Frasca, A., et al. 2009, A&A, 504, 829
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Harris, W. E. 1996, AJ, 112, 1487
Hinkle, K., Wallace, L., Valenti, J., & Harmer, D. 2000, Visible and Near Infrared Atlas of the Arcturus Spectrum 3727-9300 A
Jofré, P., Heiter, U., Soubiran, C., et al. 2014, A&A, 564, A133
Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, AJ, 146, 134
Kupka, F., Dubernet, M.-L., & VAMDC Collaboration. 2011, Baltic Astron., 20, 503
Kurucz, R. 1993, ATLAS9 Stellar Atmosphere Programs and 2 km s^{-1} grid. Kurucz CD-ROM No. 13. Cambridge, Mass.: Smithsonian Astrophysical Observatory, 13
Kurucz, R. L. 2005, Mem. Soc. Astron. It. Supp., 8, 14
Lanzafame, A. C., Frasca, A., Damiani, F., et al. 2014, A&A, submitted
Magrini, L., Randich, S., Friel, E., et al. 2013, A&A, 558, A38
Magrini, L., Randich, S., Romano, D., et al. 2014, A&A, 563, A44
Masseron, T. 2006, Ph.D. Thesis, Observatoire de Paris, France
McSwain, M. V., & Gies, D. R. 2005, ApJS, 161, 118
Mészáros, S., & Allende Prieto, C. 2013, MNRAS, 430, 3285
Morel, T., & Miglio, A. 2012, MNRAS, 419, L34
Mucciarelli, A., Pancino, E., Lovisi, L., Ferraro, F. R., & Lapenna, E. 2013, ApJ, 766, 78
Nelder, J. A., & Mead, R. 1965, Computer Journal, 7, 308
Neves, V., Santos, N. C., Sousa, S. G., Correia, A. C. M., & Israelian, G. 2009, A&A, 497, 563
Önehag, A., Korn, A., Gustafsson, B., Stempels, E., & Vandenberg, D. A. 2011, A&A, 528, A85
Pancino, E. 2012, Proc. of Calibration and Standardization of large surveys and missions in Astronomy and Astrophysics, ASP Conf. Ser., submitted [arXiv:1206.6291]
Pasquini, L., Avila, G., Blecha, A., et al. 2002, The Messenger, 110, 1
Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
Plez, B. 2012, Turbospectrum: Code for spectral synthesis, astrophysics Source Code Library
Porto de Mello, G. F., & da Silva, L. 1997, ApJ, 482, L89
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in C. The art of scientific computing (Cambridge University Press)
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2002, Numerical recipes in C++ : the art of scientific computing (Cambridge University Press)
Prugniel, P., & Soubiran, C. 2001, A&A, 369, 1048
Ramírez, I., & Meléndez, J. 2005, ApJ, 626, 465
Randich, S., & Gilmore, G. 2013, The Messenger, 154, 47
Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, MNRAS, 370, 141
Reddy, B. E., Lambert, D. L., & Allende Prieto, C. 2006, MNRAS, 367, 1329
Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2012, ApJ, 748, 93
Ruffoni, M. P., Den Hartog, E. A., Lawler, J. E., et al. 2014, MNRAS, 441, 3127
Sacco, G. G., Morbidelli, L., Franciosini, E., et al. 2014, A&A, 565, A113
Sbordone, L., Bonifacio, P., Castelli, F., & Kurucz, R. L. 2004, Mem. Soc. Astron. It. Suppl., 5, 93
Sbordone, L., Caffau, E., Bonifacio, P., & Duffau, S. 2014, A&A, 564, A109
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Serenelli, A. M., Bergemann, M., Ruchti, G., & Casagrande, L. 2013, MNRAS, 429, 3645
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Sneden, C. A. 1973, Ph.D. Thesis, The University OF Texas at Austin
Soubiran, C., & Girard, P. 2005, A&A, 438, 139
Soubiran, C., Le Campion, J.-F., Cayrel de Strobel, G., & Caillo, A. 2010, A&A, 515, A111
Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., & Monteiro, M. J. P. F. G. 2007, A&A, 469, 783
Sousa, S. G., Santos, N. C., Mayor, M., et al. 2008, A&A, 487, 373
Sousa, S. G., Santos, N. C., Israelian, G., et al. 2011, A&A, 526, A99
Sousa, S. G., Santos, N. C., Adibekyan, V., et al. 2014, A&A, 561, A21
Spina, L., Randich, S., Palla, F., et al. 2014, A&A, 567, A55
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Stetson, P. B., & Pancino, E. 2008, PASP, 120, 1332
Stetson, P. B., & Pancino, E. 2010, DAOSPEC: An Automatic Code for Measuring Equivalent Widths in High-resolution Stellar Spectra, astrophysics Source Code Library
Tabernero, H. M., Montes, D., & González Hernández, J. I. 2012, A&A, 547, A13
Valenti, J. A., & Piskunov, N. 1996, A&AS, 118, 595
Valenti, J. A., Piskunov, N., & Johns-Krull, C. M. 1998, ApJ, 498, 851
Valentini, M., Morel, T., Miglio, A., Fossati, L., & Munari, U. 2013, in EPJ Web of Conf. 43, 3006
Venn, K. A., Irwin, M., Shetrone, M. D., et al. 2004, AJ, 128, 1177
Wallace, L., Hinkle, K., & Livingston, W. 1998, An atlas of the spectrum of the solar photosphere from 13 500 to 28 000 cm^{-1} (3570 to 7405 A)
Wallerstein, G., Iben, Jr., I., Parker, P., et al. 1997, Rev. Mod. Phys., 69, 995
Weiss, A., & Schlattl, H. 2008, Ap&SS, 316, 99
Wilkinson, M. I., Vallenari, A., Turon, C., et al. 2005, MNRAS, 359, 1306
Worley, C. C., de Laverny, P., Recio-Blanco, A., et al. 2012, A&A, 542, A48
Yadav, R. K. S., Bedin, L. R., Piotto, G., et al. 2008, A&A, 484, 609
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
Yong, D., & Grundahl, F. 2008, ApJ, 672, L29
Zucker, D. B., de Silva, G., Freeman, K., Bland-Hawthorn, J., & Hermes Team. 2012, in Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, eds. W. Aoki, M. Ishigaki, T. Suda, T. Tsujimoto, & N. Arimoto, ASP Conf. Ser., 458, 421

[1] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
e-mail: rsmiljanic@ncac.torun.pl
[2] Department for Astrophysics, Nicolaus Copernicus Astronomical Center, ul. Rabiańska 8, 87-100 Toruń, Poland
[3] Department of Physics and Astronomy, Division of Astronomy and Space Physics, Uppsala University, Box 516, 75120 Uppsala, Sweden
e-mail: andreas.korn@physics.uu.se

[4] Max Planck Institute for Astrophysics, Karl-Schwarzschild Str. 1, 85741 Garching, Germany
[5] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
[6] INAF – Osservatorio Astrofisico di Catania, via S. Sofia 78, 95123 Catania, Italy
[7] INAF – Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Florence, Italy
[8] Université Libre de Bruxelles, Campus Plaine, CP 226, Boulevard du Triomphe, 1050 Bruxelles, Belgium
[9] INAF – Osservatorio Astronomico di Bologna, via Ranzani 1, 40127 Bologna, Italy
[10] ASI Science Data Center, via del Politecnico SNC, 00133, Roma, Italy
[11] Lund Observatory, Department of Astronomy and Theoretical Physics, Box 43, 221 00 Lund, Sweden
[12] Departamento de Astronomía, Universidad de Concepción, Casilla 160 Concepción, Chile
[13] Zentrum für Astronomie der Universität Heidelberg, Landessternwarte, Königstuhl 12, 69117 Heidelberg, Germany
[14] Millennium Institute of Astrophysics, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile
[15] Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile
[16] Centro de Astrofísica, Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal
[17] Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal
[18] Dept. Astrofísica, Facultad de CC. Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain
[19] Institute of Theoretical Physics and Astronomy, Vilnius University, Goštauto 12, 01108 Vilnius, Lithuania
[20] Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany
[21] Laboratoire Lagrange (UMR 7293), Université de Nice-Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France
[22] Instituto de Astrofísica de Canarias, C via Lactea s/n, 38200 La Laguna, Tenerife, Spain
[23] Dept. Astrofísica, Universidad de La Laguna (ULL), 38206 La Laguna, Tenerife, Spain
[24] LAB UMR 5804, Univ. Bordeaux et CNRS, 33270 Floirac, France
[25] GEPI, Observatoire de Paris, CNRS, Univ. Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France
[26] INAF – Osservatorio Astronomico di Padova, Vicolo Osservatorio 2 35122 Padova, Italy
[27] Dipartimento di Fisica e Astronomia "G. Galilei" Università degli Studi di Padova. via Marzolo 8, 35131 Padova, Italy
[28] Dipartimento di Fisica & Astronomia, Universita' di Bologna, Viale Berti/Pichat 6/2, 40127 Bologna, Italy
[29] Department of Astronomy, Indiana University, Bloomington, IN 47405, USA
[30] Kavli Institute for Astrophysics & Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[31] Astrophysics Section, Department of Physics and Astronomy, University of Catania, via S. Sofia 78, 95123 Catania, Italy
[32] Institut d'Astrophysique et de Géophysique, Université de Liège, Allée du 6 Août, Bât. B5c, 4000 Liège, Belgium
[33] Institute of Space Sciences (IEEC-CSIC), Campus UAB, Fac. Ciències, Torre C5 parell 2, 08193 Bellaterra, Spain
[34] Dipartimento di Fisica e Astronomia, Università di Firenze, via Sansone, 1, 50019 Sesto Fiorentino (FI), Italy
[35] Research School of Astronomy and Astrophysics, Australian National University, Cotter Road, Weston Creek, ACT 2611, Australia
[36] Rudolf Peierls Centre for Theoretical Physics, Keble Road, Oxford OX1 3NP, UK
[37] Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, Hatfield, AL10 9AB, UK
[38] Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK
[39] Astrophysics Group, Research Institute for the Environment, Physical Sciences and Applied Mathematics, Keele University, Keele, Staffordshire ST5 5BG, UK
[40] INAF – Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, 90134 Palermo, Italy
[41] Departamento de Física, Ingeniería de Sistemas y Teoría de la Señal, Universidad de Alicante, Apdo. 99, 03080 Alicante, Spain
[42] ESA, ESTEC, Keplerlaan 1, PO Box 299, 2200 AG Noordwijk, The Netherlands
[43] Max Planck Institute für Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[44] Instituto de Astrofísica de Andalucía-CSIC, Apdo. 3004, 18080 Granada, Spain
[45] Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
[46] Instituto de Física y Astronomía, Universidad de Valparaíso, 2340000 Valparaíso, Chile
[47] European Southern Observatory, Alonso de Cordova, 3107 Vitacura, Santiago de Chile, Chile

Pages 28 to 38 are available in the electronic edition of the journal at http://www.aanda.org

Appendix A: Nodes and methods

The UVES data of late-type stars are analyzed in parallel by 13 different Nodes. The details of each analysis methodology and the codes employed are described in the subsections below. Table 2 summarizes some characteristics of the methodology employed by each Node.

A.1. Bologna

The Bologna Node employs the classical EW method for determining atmospheric parameters and abundances. The atmospheric parameters are determined by erasing any trend of the abundances of the iron lines with excitation potential and with EW, and by minimizing the difference between the abundances given by ionized and neutral iron lines. A final health check of the method is provided by verifying that no significant trend of iron abundances with wavelength is present. Abundances are derived for each absorption line of the species of interest.

To measure EWs, the automated FORTRAN code DAOSPEC (Stetson & Pancino 2008, 2010) is used. DAOSPEC is designed to measure EWs in high-resolution ($R \geq 15\,000$) high-S/N stellar spectra (≥30). Upon request, the code normalizes the spectrum by adjusting, iteratively, polynomials to the residuals spectrum (i.e., a spectrum obtained by removing all measured lines from the original spectrum). DAOSPEC provides a global uncertainty of the fit in the form of an average root mean square (rms) of the residuals spectrum, a radial velocity measurement (with its 1σ spread and the number of lines on which it is based), and the EWs with their uncertainty and quality parameters.

DAOSPEC can be somewhat difficult to configure, especially when many spectra with different properties, i.e., S/N, line crowding, full width half maximum (FWHM) and exact spectral coverage, need to be measured in a short time, as is the case for *Gaia*-ESO. Therefore, the code is executed through a wrapper that automatically configures many of its parameters, providing all the statistics and graphical tools to explore the results and correct the deviant cases. This wrapper program is called DOOp (DAOSPEC Option Optimizer Pipeline, Cantat-Gaudin et al. 2014a).

Finally, to derive the atmospheric parameters and elemental abundances automatically, the code GALA[20] is used (Mucciarelli et al. 2013). GALA is based on the Kurucz suite of abundance calculation codes (Kurucz 2005; Sbordone et al. 2004). GALA can run starting from a random first guess of the atmospheric parameters and converges rapidly to meaningful solutions for spectra with the resolution, S/N, and wavelength coverage of the UVES spectra analyzed here. GALA performs a rejection of too weak or too strong absorption lines (the limits are set around the $\log(EW/\lambda) \simeq -4.7$ and -5.9, depending on the star), selects only lines having a certain measurement error (cutting above 5−20%, depending on the spectrum), and performs a sigma-clipping rejection in abundance (set to 2.5σ). GALA provides uncertainties on the atmospheric parameters and on the derived abundances, both in the form of a 1σ spread of the abundances of each line (together with the number of lines used for each species) and in the form of errors on the abundances induced by the uncertainties on the atmospheric parameters (using the prescription of Cayrel et al. 2004, in the case of the present analysis).

[20] GALA is freely distributed at the Cosmic-Lab project Web site, http://www.cosmic-lab.eu/Cosmic-Lab/Products.html

A.2. Catania (OACT)

The Catania Node uses the code ROTFIT, developed by Frasca et al. (2003, 2006) in IDL[21] software environment. The code originally performed only an automatic MK spectral classification and $v \sin i$ measurement minimizing the χ^2 of the residual ($observed - reference$) spectra. The reference spectra come from an adopted spectrum library and are artificially broadened by convolution with rotational profiles of increasing $v \sin i$. The code was later updated for evaluating the atmospheric parameters $T_{\rm eff}$, $\log g$, and [Fe/H] with the adoption of a list of reference stars with well-known parameters (e.g., Guillout et al. 2009).

Unlike codes based on the measurements of EWs and curves of growth, ROTFIT can be applied to the spectra of FGK-type stars with relatively high rotational velocity ($v \sin i \geq 20$ km s^{-1}), where the severe blending of individual lines either hampers or absolutely prevents the use of the above methods. Nevertheless, the analysis was limited to stars with $v \sin i \leq 300$ km s^{-1}.

A reference library composed of 270 high-resolution ($R = 42\,000$) spectra of slowly-rotating FGKM-type stars available in the ELODIE archive (Prugniel & Soubiran 2001) was used. For most of these reference stars, basically those with spectral type in the range from mid-F to late-K, the atmospheric parameters have been redetermined by L. Spina using the EPInArBo methodology (see Sect. A.5). For the remaining few stars, either the recent values tabulated in the PASTEL catalog (Soubiran et al. 2010) or derived in the works of Rojas-Ayala et al. (2012) and Boyajian et al. (2012) for M-type dwarfs, were used. Although the parameter space is not regularly sampled, the reference stars cover all regions relevant for analysis of FGK-type stars with [Fe/H] ≥ −2.0.

Segments of the spectra with 100 Å each are analyzed independently. Spectral regions heavily affected by telluric lines and the cores of Balmer lines, that can be contaminated by chromospheric emission, are excluded. The final stellar parameters $T_{\rm eff}$, $\log g$, [Fe/H], and $v \sin i$, are the averages of the results of each i-th spectral segment weighted according to the χ^2_i of the fit and to the amount of information contained in the segment, which is expressed by the total line absorption $f_i = \int (F_\lambda/F_{\rm C} - 1) {\rm d}\lambda$. The uncertainties of $T_{\rm eff}$, $\log g$, [Fe/H], and $v \sin i$ are the standard errors of the weighted means added in quadrature to the average uncertainties of the stellar parameters of the reference stars evaluated as ±50 K, ±0.1 dex, ±0.1 dex, and ±0.5 km s^{-1} for $T_{\rm eff}$, $\log g$, [Fe/H], and $v \sin i$, respectively. Moreover, the MK spectral type and luminosity class of the star is also provided. They are defined as those of the reference star that more frequently matches the target spectrum in different spectral regions.

A.3. CAUP

The CAUP Node determines the stellar atmospheric parameters ($T_{\rm eff}$, $\log g$, ξ) and the metallicity automatically, with a method used in previous works now adapted to the *Gaia*-ESO Survey (e.g., Sousa et al. 2008, 2011). The method is based on the excitation and ionization balance of iron lines using [Fe/H] as a proxy for the metallicity. The list for the iron lines used to constrain the parameters was selected from the *Gaia*-ESO line list using a new procedure described in detail in Sousa et al. (2014).

[21] IDL (Interactive Data Language) is a registered trademark of ITT Visual Information Solutions.

The EW of the lines are automatically determined using the ARES[22] code (Sousa et al. 2007), following the approach of Sousa et al. (2008, 2011) to adjust ARES according to the S/N of each spectrum.

The stellar parameters are computed assuming LTE using the 2002 version of MOOG (Sneden 1973) and the MARCS grid of models. For that purpose, the interpolation code provided with the MARCS grid was modified to produce an output model readable by MOOG. Moreover, a wrapper program was implemented on the interpolation code to automatize the method.

As damping prescription, the Unsöld approximation multiplied by a factor recommended by the Blackwell group (option 2 within MOOG) was used. The atmospheric parameters are inferred from the previously selected Fe I-Fe II line list. A minimization algorithm, the Downhill Simplex Method (Press et al. 1992), is used to find the best parameters. In order to identify outliers caused by incorrect EW values, a 3σ clipping of the Fe I and Fe II lines is performed after a first determination of the stellar parameters. After this clipping, the procedure is repeated without the rejected lines. The uncertainties in the stellar parameters are determined as in previous works (Sousa et al. 2008, 2011).

Individual abundances are derived using the same tools and methodology as described above, but using the 2010 version of MOOG (see Neves et al. 2009; Adibekyan et al. 2012, for details). The line list for elements other than Fe was selected through the cross-matching between the line list used by Adibekyan et al. (2012) and the line list provided by *Gaia*-ESO. The atomic data from the *Gaia*-ESO Survey was adopted. The errors for the abundances represent the line-to-line scatter.

A.4. Concepcion

The Concepcion Node uses the abundances from Fe I and Fe II lines to obtain atmospheric parameters using the classical EW method. The atmospheric parameters are determined by satisfying the excitation and ionization equilibrium, and by minimizing trends of abundance with EW. The spectroscopic optimization of all the atmospheric parameters is achieved simultaneously.

The EWs are determined with the automatic code DAOSPEC (see description in Sect. A.1). The code adopts a saturated Gaussian function to fit the line profile and a unique value for the FWHM for all the lines. The input values of FWHM are derived manually using the IRAF[23] task splot, leaving DAOSPEC free to readjust the values according to the global residual of the fitting procedure. The measurement of EWs is repeated by using the optimized FWHM value as a new input value until convergence is reached at a level of 5%. The EWs are measured after a renormalization of the continuum, to remove any residual trends introduced during the data reduction.

GALA is used to determine the atmospheric parameters and elemental abundances (see description of GALA in Sect. A.1). Starting from an initial guess of atmospheric parameters, GALA converges rapidly to a meaningful solution. Finally, it computes accurate internal errors for each atmospheric parameter and abundance. When the initial set of parameters are poorly known or in cases of large uncertainties, the guess working-block of GALA is used. This working-block verifies the initial parameters quickly by exploring the parameters space in a coarse grid, saving a large amount of time. In addition, the errors in the EW measurement obtained from DAOSPEC are provided as an input, so the best model atmosphere is computed taking the abundance uncertainties of the individual lines into account.

A.5. EPInArBo

The EPInArBo (ESO-Padova-Indiana-Arcetri-Bologna) Node performs the spectral analysis with the codes DOOp and FAMA (Fast Automatic Moog Analysis, Magrini et al. 2013)[24]. The former (described in Sect. A.1) makes it more convenient to measure EWs in hundreds of spectra in a single batch. The latter is an automation of the 1D-LTE code MOOG and allows the determination of stellar parameters and individual element abundances.

The EWs are measured after a renormalization of the continuum. Each line is measured using a Gaussian fit. Equivalent widths between 20−120 mÅ were used for Fe I and Fe II lines and between 5−120 mÅ for lines of other elements.

FAMA uses the EWs of Fe I and Fe II to derive stellar parameters ($T_{\rm eff}$, $\log g$, [Fe/H], and ξ). A set of first-guess parameters are first produced using the available photometric data and information from the target selection, using the following steps:

1. A first guess estimation of $T_{\rm eff}$ is given by the Alonso et al. (1999) and Casagrande et al. (2010) relations for both cluster and field stars.
2. The cluster parameters, such as distance, age, and reddening, available in the reports prepared by *Gaia*-ESO WG 4 (cluster stars target selection, see Bragaglia et al., in prep.) are used to fix the surface gravity.
3. For the field stars, the information available from target selection is used (i.e., whether the star was a turn-off dwarf or bulge/inner-disk giant) to set a first guess gravity.

The stellar parameters are obtained by searching iteratively for the three equilibria, excitation, ionization, and the trend between $\log n$(Fe I) and $\log (EW/\lambda)$, i.e., with a series of recursive steps starting from a set of initial atmospheric parameters and arriving at a final set of atmospheric parameters that fulfills the three equilibrium conditions.

The convergence criterion is set using information on the quality of the EW measurements, i.e., the minimum reachable slopes are linked to the quality of the spectra, as expressed by the dispersion $\sigma_{\rm FeI}$ around the average value $\langle \log n({\rm FeI}) \rangle$. This is correct in the approximation that the main contribution to the dispersion is due to the error in the EW measurement rather than to inaccuracy in atomic parameters, e.g., the oscillator strengths ($\log gf$).

A.6. IAC-AIP

The IAC-AIP Node employs the optimization code FERRE to identify the combination of atmospheric parameters of a synthetic model that best matches each observed spectrum. FERRE searches for the best solution in a χ^2 sense using the Nelder-Mead algorithm (Nelder & Mead 1965), and the model evaluation is sped up by holding a precomputed grid in memory and

[22] ARES can be downloaded at http://www.astro.up.pt/~sousasag/ares/

[23] IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under cooperative agreement with the National Science Foundation.

[24] FAMA is available from http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/558/A38

interpolating within it. The algorithm is the same as described by Allende Prieto et al. (2006) for the analysis of SDSS/SEGUE data and by Ahn et al. (2014) for the analysis of SDSS/APOGEE spectra. Model interpolations are carried out with cubic Bezier splines, whose accuracy has been studied in detail by Mészáros & Allende Prieto (2013). For each spectrum, five searches initialized at randomly chosen points on the parameter space are performed and the best solution is retained.

The adopted grid of model spectra was not the one described in Sect. 5.3. It was calculated using the code Turbospectrum (Alvarez & Plez 1998; Plez 2012) based on MARCS model atmospheres with the VALD3 line list (Kupka et al. 2011), with updates on $\log gf$ values according to the *Gaia*-ESO line list version 3.0. The parameter range covered by the grid is: $T_{\rm eff}$ = 3000−7000 K, $\log g$ = 0.0−5.0, [Fe/H] = −2.5−+1.0, $v \sin i$ = 1−128 km s^{-1}, ξ = 0.5−4 km s^{-1}, and [α/Fe] = −0.4−+0.4. The model spectra were smoothed by Gaussian convolution to the resolving power of the observations (R = 47 000). To speed up the analysis, the [α/Fe] is tied to the overall metallicity of each star, i.e., with enhanced [α/Fe] ratios at low metallicity, while ξ is tied to both $T_{\rm eff}$ and $\log g$ according to the *Gaia*-ESO microturbulence relation for the iDR1 analysis.

We use all of the individual UVES orders before they are merged, excluding only regions with many telluric lines and the core of the Hα line. The continuum for both the observations and the models is set by cutting the spectra into 2 Å wide chunks, dividing each chunk by its mean value, and all spectra are weighted according to their variance. All observations are shifted to rest wavelength. When only one value of $v_{\rm rad}$ is available in the header, this value was used. If two values were present (one for each CCD), the average value was used. In case no velocity was available, a cross correlation using a hot template star ($T_{\rm eff}$ = 7000 K, $\log g$ = 2) spun up to 50 km s^{-1} was used to derive the radial velocity. If this failed, a value of 0.0 km s^{-1} was used.

A.7. Liège

The Liège Node performs the analysis using the GAUFRE tool (Valentini et al. 2013). GAUFRE is a C++ code that performs the determination of atmospheric parameters and abundances in an automatic way. The tool is made up of several subprograms with specific tasks (see Valentini et al. 2013, for details). For the *Gaia*-ESO Survey UVES spectra, GAUFRE-EW is used. This subprogram determines $T_{\rm eff}$, $\log g$, [M/H], and ξ, in an iterative way using the EWs of Fe lines.

The starting point is the normalization of the spectrum and the measurement of the EWs of every line present in the input line list (when detectable). The program selects a spectral range of 3−4 Å around the line center and the spectrum is then fitted with a polynomial function in order to determine the continuum and the line position. At this stage several parameters, such as the degree of the function and the amplitude of the spectral range to fit, can be defined by the user.

The program then feeds MOOG with the measured EWs and an appropriate MARCS model atmosphere. Within the errors, the MOOG results must satisfy four conditions: fulfill the Fe ionization and excitation equilibria, show no dependence between the Fe I abundances and log (EW/λ), and, finally, yield a mean metallicity identical to that of the adopted model atmosphere. The appropriate MARCS model atmosphere is derived by interpolating within the MARCS grid.

The program iterates until the four conditions are fulfilled. The Downhill Simplex Method (Nelder & Mead 1965; Press et al. 2002) is adopted for estimating the new set of atmospheric parameters at each iteration. The starting point of the process can be determined by the user. Photometric temperatures using Ramírez & Meléndez (2005) and, when available, $\log g$ from asteroseismology were adopted. When no information from photometry or asteroseismology is available, the starting point is set to $T_{\rm eff}$ = 5000 K, $\log g$ = 4.0 dex, [M/H] = 0.0 dex, and ξ = 1.0 km s^{-1}.

The uncertainty in $T_{\rm eff}$ is derived from the standard deviation of the least-squares fit of the Fe I abundance vs. excitation potential. The uncertainty in $\log g$ is determined by propagating the uncertainty in $T_{\rm eff}$. The uncertainty in ξ is calculated based on the standard deviation of the least-squares fit of the Fe I abundance vs. log (EW/λ). The uncertainty in [Fe/H] takes into account the uncertainties in $T_{\rm eff}$, $\log g$, ξ, and the line-to-line scatter of the Fe I abundances.

A.8. LUMBA

The LUMBA (Lund-Uppsala-MPA-Bordeaux-ANU) Node uses a stellar parameter and abundance pipeline (hereafter referred to as SGU) that is based upon the SME (Spectroscopy Made Easy) spectrum synthesis program (Valenti & Piskunov 1996)[25]. SME is a suite of IDL and C++ routines developed to compute theoretical spectra and perform a χ^2 fit to observed spectra. The code assumes LTE and plane-parallel geometry. Chemical equilibrium for molecules is determined as described in Valenti et al. (1998).

A detailed description of the SGU pipeline will be published elsewhere (Bergemann et al. 2014c, in prep.). Briefly, in the SGU pipeline, synthetic spectra are computed in predefined wavelength segments, which are 5 to 20 Å wide. The selected line list is a reduced version of the *Gaia*-ESO version 3.0 line list and includes the atomic and molecular blends relevant for the analysis of FGKM-type stars. Basic stellar parameters are determined iteratively, exploring the full parameter space in $T_{\rm eff}$, $\log g$, [Fe/H], micro- and macro-turbulence. The number of iterations varies, depending on the stellar parameters, value of the goodness-of-fit test (χ^2), and convergence. The main purpose of SGU is to control the sequence of steps that defines the parameters to solve for in the current iteration, and specify the wavelength regions to include in the test statistics. Usually, three to four steps for dwarfs and subgiants, and two steps for giants are used. The wavelength regions (referred to as "masks") to be included in the χ^2 fit also vary, depending on the step. The masks cover the lines of H I (H$_\beta$ and H$_\alpha$), Mg I triplet at 5170 Å, and a carefully selected set of Fe lines. In total, about 60 diagnostic Fe I and Fe II transitions are used. The merged not normalized *Gaia*-ESO spectra are used with a runtime continuum normalization. For the abundance analysis, special masks were developed, which cover the lines of selected elements. For iDR1, atmospheric parameters were computed assuming LTE. For iDR2, the pipeline was modified to include NLTE corrections in Fe (Bergemann et al., in prep.). That resulted in improved stellar parameters (especially log g) for low-metallicity stars. Further, the effects were quite small for more metal-rich stars. Abundances are determined in the last step using stellar parameters from the previous runs.

Errors in the other stellar parameters are estimated from internal SME errors based on S/N and Fe line-to-line scatter (but in

[25] http://www.stsci.edu/~valenti/sme.html

many cases, lines of different elements were used to derive stellar parameters, including H and Mg), combined with the spread in differences between our results for the benchmark stars library and those values that have been deemed acceptable.

A.9. Nice

The Nice Node analysis is based on the automated stellar parametrization pipeline developed for the AMBRE Project (Worley et al. 2012). At the core of the pipeline is the stellar parametrization algorithm MATISSE (MATrix Inversion for Spectrum SynthEsis), developed at the Observatoire de la Côte d'Azur primarily for use in the *Gaia* RVS (Radial Velocity Spectrometer) stellar parametrization pipeline (Recio-Blanco et al. 2006), and the *Gaia*-ESO synthetic spectrum grid (see Sect. 5.3).

MATISSE is a local multilinear regression method that simultaneously determines the stellar parameters (θ) of an observed spectrum $O(\lambda)$ by the projection of the spectrum onto vector functions $B_\theta(\lambda)$. A $B_\theta(\lambda)$ function is constructed as an optimal linear combination of the local synthetic spectra $S(\lambda)$. The stellar parameters determined by the Nice Node are $T_{\rm eff}$, $\log g$, a global metallicity [M/H], and a global α-element abundance over iron ([α/Fe]: α = O, Ne, Mg, Si, S, Ar, Ca, and Ti).

To minimize the impact of mismatches between the observed and synthetic spectra, a solar flux spectrum (Wallace et al. 1998) and an Arcturus spectrum (Hinkle et al. 2000) are compared with corresponding *Gaia*-ESO synthetic spectra in the UVES spectral range. About 24% of the UVES domain is discarded because of telluric/instrumental contamination. A further 4% is discarded for differences between the observed and synthetic normalized fluxes greater than 10% for the Sun or 20% for Arcturus. These limits grossly reject discrepant spectral features (errant lines or blatant mismatched regions) between the observed and synthetic spectra. The resulting comparison prior to any normalization optimisation shows for the remainder that 95% (resp. 80%) of the pixels have less than 5% difference between the Sun (resp. Arcturus) and the corresponding synthetic spectrum, while 94% of the pixels have flux differences less than 10% in the case of Arcturus. As MATISSE uses all the available pixels for the parameter determination, any few discrepant pixels remaining after the full iterative normalisation have little effect on the result.

The final wavelength domain totals 1447 Å between 4790 Å and 6790 Å with 18 080 pixels at a sampling of 0.08 Å/px. The synthetic spectra are convolved with a Gaussian kernel ($FWHM$ = 0.2254 Å) for a resolution range from $R \sim 21\,000$ (4790 Å) to $R \sim 30\,000$ (6790 Å). The observed spectra are convolved to the same resolution using a transformation FWHM based on the measured spectral FWHM and grid FWHM.

The Nice pipeline consists of spectral processing ($v_{\rm rad}$ correction; cleaning/slicing/convolution; normalization to synthetic spectra), and stellar parameter determination by MATISSE (SPC stage in Fig. 4 of Worley et al. 2012). At each iteration of these last two stages, improved estimates of the stellar parameters provide new synthetic spectra for use in the normalization until there is convergence on the final stellar parameters.

Calibration and validation of the pipeline was undertaken using three key samples: the *Gaia*-ESO Benchmarks (see Sect. 7.1); the spectral library of Jofré et al. (2014); and the AMBRE:UVES#580 PASTEL data set (Worley et al., in prep.), a sample of 2273 slit spectra that have high quality spectroscopic stellar parameters cited in the PASTEL catalog. These three samples were used to calibrate the convolution and normalization in the spectral processing by comparison of processed spectra with synthetic spectra and by comparing the MATISSE parameters with the accepted parameters for each sample.

A.10. Paris-Heidelberg

The Paris-Heidelberg Node uses the automatic parameter determination and abundance analysis code MyGIsFOS (Sbordone et al. 2014). MyGIsFOS strictly replicates a "traditional", or "manual", parameter determination and abundance analysis method in a fully automated fashion. To do so, MyGIsFOS determines EWs and abundances for a number of Fe I and Fe II features, and looks for the atmospheric parameters ($T_{\rm eff}$, $\log g$, ξ) that satisfy the excitation and ionization equilibrium, and that minimize trends of abundance with EW. MyGIsFOS uses a precomputed grid of synthetic spectra instead of relying on on-the-fly synthesis or on a priori EW measurements. By fitting against synthetic spectra, MyGIsFOS can use moderately blended features in abundance measurements, or treat directly HFS-affected lines. A summary of how MyGIsFOS works follows:

1. A grid of synthetic spectra varying (in the most general case) in $T_{\rm eff}$, $\log g$, ξ, [Fe/H], and [α/Fe] is provided to the code together with the input spectra (for which initial guess parameters have to be provided), and a set of spectral "regions" to be used either as pseudocontinuum ranges (for normalization) or as spectral features of various kinds (e.g., Fe I lines).
2. The observed spectrum, and each spectrum in the synthetic grid, are pseudonormalized using the continuum intervals, then the synthetic grid is collapsed (by interpolation) at the initial guess values for $T_{\rm eff}$, $\log g$, ξ and [α/Fe], leaving a grid whose sole dimension is [Fe/H].
3. The provided Fe I and Fe II lines are fitted (by χ^2-minimization) against the collapsed grid, deriving best-fit Fe I abundances for each line. EWs are also measured in the process. In a series of nested loops, the aforementioned diagnostics (excitation and ionization equilibrium, etc.) are evaluated, and if needed, the stellar parameters are altered, and the whole process repeated, until convergence is achieved.
4. To measure abundances of other elements, the respective features are fitted against the same grid, collapsed at the final values of $T_{\rm eff}$, $\log g$, ξ, and thus varying in [Fe/H]. The best fitting metallicity value is used as the element [X/H] (this is in principle inconsistent but leads to generally accurate values, see Sbordone et al. 2014). A special case is α-elements, which are measured first after $T_{\rm eff}$, $\log g$, and ξ have been set, and used to estimate the last grid parameter, [α/Fe]. The derived value of [α/Fe], if different enough from the estimated, triggers a new estimation of the other parameters. Finally, all the other elements are measured.

After processing, the output is examined for signs of problems: nonconverging objects are checked individually and eventually rerun. MyGIsFOS does not estimate or vary the spectrum broadening: the grid is provided broadened at the nominal resolution of R = 47 000. Stars showing extra-broadening (essentially moderately rotating objects) are detected by inspecting the quality of line fits, and reprocessed with appropriate broadening.

For GESviDR1Final, $T_{\rm eff}$ was *not* iterated within MyGIsFOS, since this was not yet implemented. Instead, $T_{\rm eff}$ was determined from the available photometry by applying the González Hernández & Bonifacio (2009) relations. Full $T_{\rm eff}$ iteration is now in place and was used in the analysis of iDR2. In addition, MyGIsFOS is using the *Gaia*-ESO grid

of synthetic spectra that does not include a ξ dimension, but relies on a single, precalibrated ξ value depending on $T_{\rm eff}$, $\log g$, and [Fe/H]. Thus, MyGIsFOS is not determining ξ for the *Gaia*-ESO data. In the future, when a new grid of synthetic spectra with the ξ dimension is available, this quantity will also be determined.

A.11. UCM

The UCM Node employs the automatic code StePar (Tabernero et al. 2012) to determine the stellar atmospheric parameters ($T_{\rm eff}$, $\log g$, ξ) and metallicity. StePar computes the stellar atmospheric parameters using MOOG (v.2002). Although designed to make use of a grid of Kurucz ATLAS9 plane-parallel model atmospheres (Kurucz 1993), StePar has been now modified to operate with the spherical and nonspherical MARCS models.

The atmospheric parameters are inferred from a previously selected Fe I-Fe II line list. The code iterates until it reaches the excitation and ionization equilibrium and minimizes trends of abundance with $\log(\mathrm{EW}/\lambda)$. StePar employs a Downhill Simplex Method (Press et al. 1992). The function to minimize is a quadratic form composed of the excitation and ionization equilibrium conditions. The code performs a new simplex optimization until the metallicity of the model and the iron abundance are the same.

Uncertainties for the stellar parameters are derived as described in Tabernero et al. (2012). In addition, a 3σ rejection of the Fe I and Fe II lines is performed after a first determination of the stellar parameters. StePar is then rerun without the rejected lines.

The EW determination of all the lines was carried out with the ARES code. The approach of Sousa et al. (2008) to adjust the parameters of ARES according to the S/N of each spectrum was followed. Regarding the individual abundances, two line lists were prepared: one for dwarfs ($\log g \geq 4.0$) and one for giants ($\log g \leq 4.0$). To get the individual abundances, the EWs are fed to MOOG and then a 3σ-clipping for each chemical element is performed.

A.12. ULB

The ULB Node uses the BACCHUS (Brussels Automatic Code for Characterizing High accUracy Spectra) code, which consists of three different modules designed to derive EWs, stellar parameters, and abundances, respectively. The current version relies on (i) a grid of MARCS model atmospheres; (ii) a specific procedure for interpolating among the model atmosphere thermodynamic structure within the grid (Masseron 2006); and (iii) the radiative transfer code Turbospectrum.

The stellar parameters determination relies on a list of selected Fe lines. The first step consists in determining accurate abundances for the selected lines using the abundance module for a given set of $T_{\rm eff}$ and $\log g$ values. The abundance determination module proceeds in the following way: (i) a spectrum synthesis, using the full set of (atomic and molecular) lines, is used for local continuum level finding (correcting for a possible spectrum slope); (ii) cosmic and telluric rejections are performed; (iii) local S/N is estimated; and (iv) a series of flux points contributing to a given absorption line is selected. Abundances are then derived by comparing the observed spectrum with a set of convolved synthetic spectra characterized by different abundances. Four different diagnostics are used: χ^2 fitting, core line intensity comparison, global goodness-of-fit estimate, and EW comparison. A decision tree then rejects the line, or accepts the line and keeps the best matching abundance.

The second step consists in deducing the EWs of Fe lines using the second module. One asset of the code is precisely this computation of EWs from best-matching synthetic spectra, because the EW of only the considered line is taken into account (excluding the contribution from nearby, blending lines). Indeed, EWs are computed not directly on the observed spectrum, but internally from the synthetic spectrum with the best-matching abundance. This way, the information about the contribution of blending lines is known, allowing a clean computation of the EW of the line of interest.

The last step of the procedure consists in injecting the derived EWs in Turbospectrum to compute abundances for a grid of 27 neighbor model atmospheres (including three values of effective temperature, three of gravity, and three of microturbulence velocity), covering the parameter space of interest. For each model, the code computes the slopes of abundances against excitation potential and against EWs, as well as Fe I and Fe II lines abundances.

The final parameters are determined by requesting that the ionization equilibrium is fulfilled, and that simultaneously null slopes for abundances against excitation potential and against EWs are obtained. The whole procedure is iterated once per star, after a first guess of stellar parameters has been refined and a new seed model computed.

A.13. Vilnius

The Vilnius Node uses a traditional EW based method for the stellar parameters determination. Effective temperature is derived by minimizing the slope of abundances obtained from Fe I lines with respect to the excitation potential. Surface gravity is determined by forcing the measured Fe I and Fe II lines to yield the same [Fe/H] value. Microturbulence is determined by forcing Fe I abundances to be independent of the EWs of the lines. A custom wrapper software was developed to measure EWs, and compute the main atmospheric parameters and abundances automatically.

Equivalent widths were measured using the DAOSPEC software. The atomic and molecular data provided by the *Gaia*-ESO line list group were used. Only lines corresponding to the best quality criteria (flags provided together with the line list) were used. Different subsamples of lines were used for giant stars and for metal-poor stars.

The stellar atmospheric parameters were computed using MOOG (v.2010) and the MARCS atmospheric models. The interpolation code provided with the MARCS grid was modified to make possible an automatic selection of the required sets of models and the extraction of the final interpolated model in the WEBMARCS format for MOOG.

The wrapper code performs an iterative sequence of abundance calculations using a simultaneous quadratic minimization of: (i) abundance dependency on the line excitation potential; (ii) difference between neutral and ionized iron abundances; and (iii) scatter of neutral iron abundances. Iterations were performed on each step until a stable solution was reached. The minimization procedure was based on the Nelder-Mead method (Nelder & Mead 1965). During this iterative procedure, the code searches for possible outliers in abundances determined using different lines. Every resulting abundance for every single line that departed from the mean by more than 2σ was flagged as outlier and was omitted from further calculations.

Table B.1. Number of FGK-type stars observed with UVES and part of the iDR1 data set.

Gaia-ESO Type	Stars	Comments
Total	508	*Gaia*-ESO only, no archival data.
GES_MW	305	Stars from Milky Way fields.
GES_CL	133	Stars from open cluster fields.
GES_SD	70	Calibration targets.

A starting point was selected randomly in a vicinity of $T_{\rm eff}$ = 5500 K, $\log g$ = 4.0, [Fe/H] = −0.5 and ξ = 1.5 km s^{-1}. The final values of atmospheric parameters for a specific star do not depend on the starting point of the calculations. The final abundances of all other elements were derived omitting possible outliers using a 2σ criteria.

The uncertainties of the stellar parameter were determined using error estimations of the line profile fitting and the standard deviations of the abundances. The uncertainty for the effective temperatures was estimated by obtaining the boundary temperature values of the possible satisfactory parameter space, using the error of the linear regression fit. The uncertainty of the gravity was obtained using the possible boundary values of $\log g$, using the standard deviations of the abundances from Fe I and Fe II lines. The uncertainty of the microturbulence velocity is obtained by employing the error of the standard deviation of the neutral iron abundances. The [Fe/H] standard deviation is adopted as the metallicity uncertainty.

Appendix B: The science verification analysis

The science verification analysis was the first full analysis cycle of the Survey. The first few papers with *Gaia*-ESO data are based on results of this first analysis (e.g., Bergemann et al. 2014; Cantat-Gaudin et al. 2014b; Donati et al. 2014; Friel et al. 2014; Magrini et al. 2014; Spina et al. 2014). We therefore believe it is important to document the details, achievements, and shortcomings of this analysis. We document in particular the differences between this analysis and the analysis of iDR2, described in the main text. The data analyzed was part of the first internal data release (iDR1), described below:

- *Internal Data Release 1 (iDR1)*: this data release consists of spectra obtained up to the end of June 2012 and includes spectra of 576 FGK-type stars observed with UVES. Of these stars, 68 are part of young open clusters (age <100 Myr). They were not analyzed by WG11 but by the working group responsible for pre-main-sequence stars (Lanzafame et al. 2014). For the moment, the results have been released only internally to the *Gaia*-ESO collaboration and are referred to as *GESviDR1Final* (*Gaia*-ESO Survey verification internal data release one). We point out that the reduced spectra for part of the stars observed in the first six months are already available through the ESO data archive[26].

The S/N distribution of the iDR1 data are shown in Fig. B.1. Table B.1 summarizes the number of stars part of iDR1. Figure B.2 shows how the stars targeted in the first six months of *Gaia*-ESO observations are distributed in the $T_{\rm eff}-\log g$ plane. Atmospheric parameters were determined for 421 stars out of the 508 in the sample. For the remaining stars, the analysis failed for different reasons (low S/N, fast rotation, reduction artefacts, etc.). Flags will be provided indicating the reason for the failure.

In the sections that follow, we discuss separately the data products determined in the analysis of the iDR1 data, i.e., EWs (Appendix B.1), stellar atmospheric parameters (Appendix B.2), and elemental abundances (Appendix B.3). The differences between this analysis and that of iDR2 are highlighted.

B.1. EWs in iDR1

The Nodes that determined EWs for the iDR1 data set were: Bologna, CAUP, Concepcion, EPInArBo, UCM, ULB, and Vilnius. Three codes were used to measure EWs automatically: ARES, BACCHUS (T. Masseron, unpublished, see Sect. A.12), and DAOSPEC. Of these codes, BACCHUS was not included in the iDR2 discussion.

We include here figures similar to those discussed in the main text about iDR2. A comparison between these plots can show the evolution of the measurements between one iDR and the next.

Figure B.3 shows the comparison between the EWs of Fe I lines measured by different groups in the two stars discussed in Sect. 6 (the metal-poor dwarf HD 22879 with $S/N \sim 260$ and the metal-rich giant Trumpler 20 MG 781 with $S/N \sim 50$).

The EWs measured with the same code by different Nodes (left and center-left plots in Fig. B.3) tend to agree to within 2σ, although a systematic difference is present in some case. When comparing the EWs measured with ARES and DAOSPEC (center-right plots in Fig. B.3), it is noticeable that the scatter increases. As discussed before, this is probably related to the different ways that the continuum is defined in each code (global vs. local continuum for DAOSPEC and ARES, respectively).

The comparison between BACCHUS and the other two codes (right plots in Fig. B.3) show systematic differences that are under investigation. BACCHUS measures the EWs not from the observed spectrum, but from a best fitting synthetic spectrum once the abundance and the parameters are fixed. It removes from the line the contribution of any known blending feature that is included in the line list. The synthetic line is computed in 1D LTE, using all the line information possible: line broadening, HFS, and blends. In this sense, the BACCHUS EWs should be the more robust measurements (assuming that the atmospheric parameters are perfectly known and that the blends are perfectly synthesized). The continuum placement might be another source of error. BACCHUS fits the continuum relying on the synthetic spectrum, adapting it from star to star, and from wavelength region to wavelength region. However, if the continuum match is poor around the measured line, the continuum may be wrong, and so will the final abundance and EW. The issue is complex and we are investigating the causes of the discrepancies and improving the measurements for future releases.

Figure B.4 depicts the behavior of $\overline{\sigma_{\rm EW}}$ measured in GESviDR1Final. For each Fe I line of a star, the average value of the EW is computed, together with its standard deviation. For each star, we define $\overline{\sigma_{\rm EW}}$ as the mean of all the standard deviations of the Fe I lines in that star. For most stars, the standard deviations are small (<3 mÅ), with a few cases reaching up to ~20 mÅ. Figure B.4 shows that for the majority of the stars, the multiple measurements of EWs tend to agree within the expected statistical uncertainty given by the S/N of the spectra.

In Fig. B.5, $\overline{\sigma_{\rm EW}}$ is plotted against the atmospheric parameters, $T_{\rm eff}$, $\log g$, and [Fe/H]. Most of the stars where

[26] http://archive.eso.org/wdb/wdb/adp/phase3_spectral/form?phase3_collection=GaiaESO

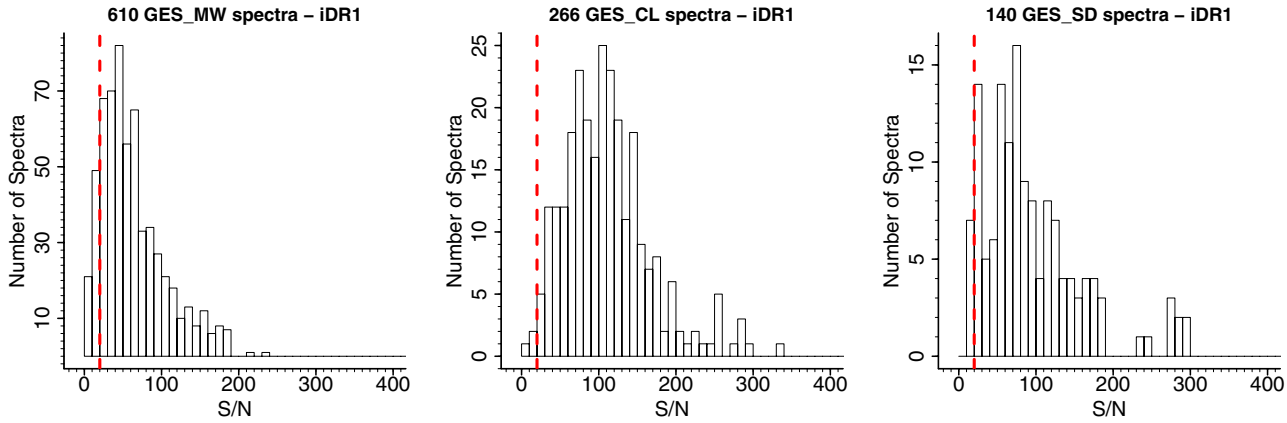

Fig. B.1. Distribution of the median S/N of the spectra in iDR1 (508 FGK-type stars) observed with UVES. Each of the two UVES spectrum parts (from each CCD) is counted separately (thus, two spectra per star). Red dashed lines indicate $S/N = 20$. Samples of the solar neighborhood (GES_MW), open clusters (GES_CL), and calibration targets (GES_SD) are shown separately.

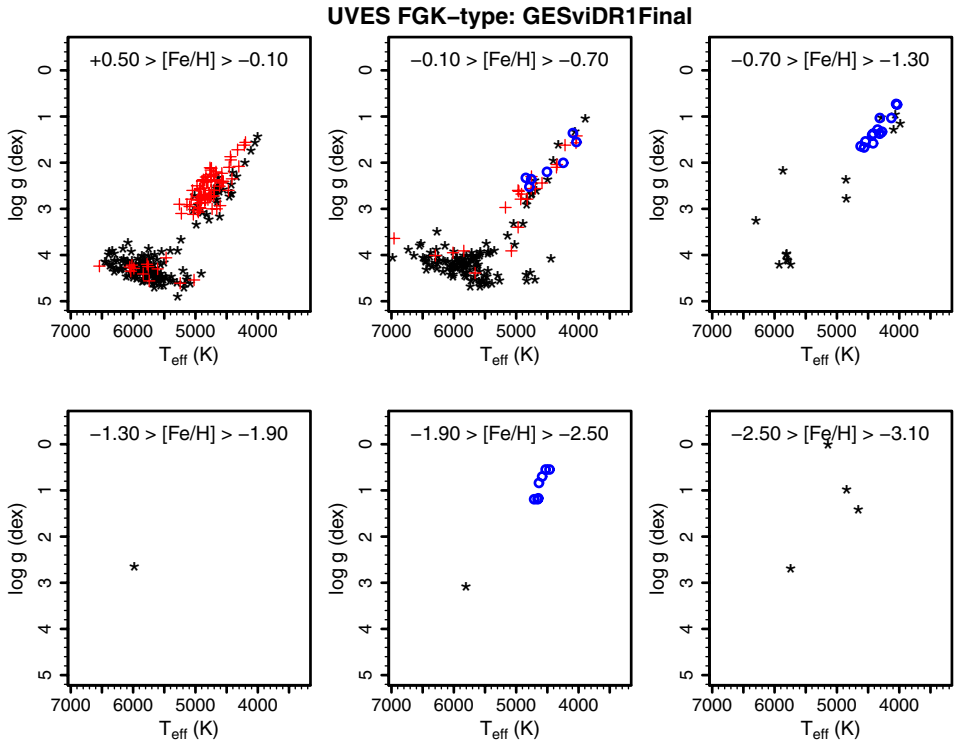

Fig. B.2. Distribution of 421 FGK-type stars from GESviDR1Final in the $T_{\rm eff} - \log g$ plane. The stars were observed with UVES during the first six months of the Survey and had atmospheric parameters determined as described in this paper. The panels are divided according to metallicity. Black stars represent field stars, red crosses stars observed in open-cluster fields, and blue circles stars observed in globular-cluster fields.

$\overline{\sigma_{\rm EW}} > 10$ mÅ tend to be warm, metal-rich subgiants or dwarfs. Many of these stars display significant rotation (Fig. B.6).

The ULB results for EWs (using the BACCHUS code), and finally for atmospheric parameters and abundances for iDR2, were not used to compute the final recommended values that will be released, as the Node withdrew its results.

B.2. Atmospheric parameters in iDR1

B.2.1. Benchmark stars

For the iDR1 analysis, only eight benchmark stars were available[27] and they did not cover the parameter space as well as the 21 stars used in iDR2. The accuracy of the Node results was judged by evaluating if the Node could reproduce $T_{\rm eff}$ and $\log g$ of most benchmark stars to within ±150 K and ±0.30 dex, respectively. If yes, the Node results were considered to be accurate. If not, the Node results were disregarded. In practice, only the results of one Node were discarded.

For iDR1, weights were not computed and the parameter space was not divided in three regions. The individual results were then combined using a simple median. The comparison with the fundamental parameters of the benchmark stars ensures that the final parameters are also in the scale defined by them, to within the accuracy level adopted above (±150 K for $T_{\rm eff}$ and ±0.30 dex for $\log g$).

B.2.2. Calibration clusters

The number of calibration clusters available during the iDR1 analysis was also smaller. Four calibration globular clusters

[27] The stars were: Bet Vir, Eta Boo, Gam Sge, Ksi Hya, HD 22879, HD 107328, HD 122563, and HD 140283.

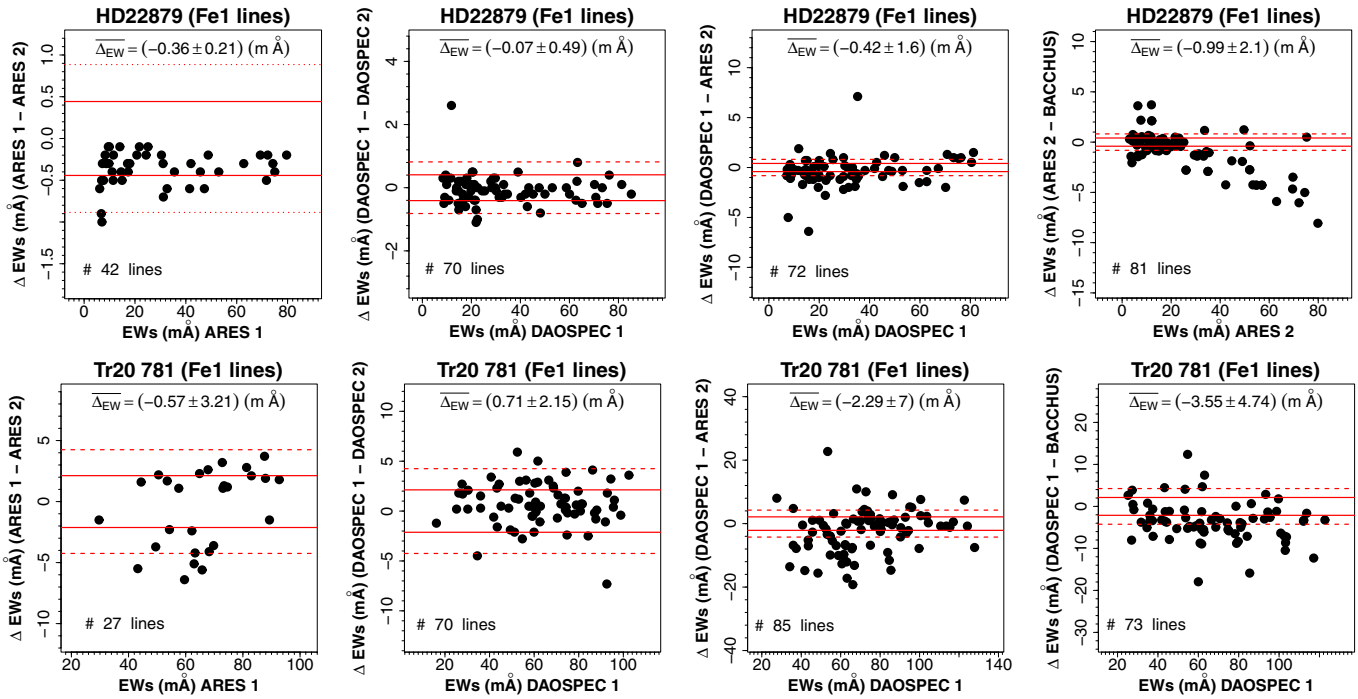

Fig. B.3. Comparison between Fe I equivalent widths measured by different Nodes for two stars. *Top row*: star HD 22879, a benchmark star used for calibration with $T_{\rm eff}$ = 5786 K, log g = 4.23, and [Fe/H] = −0.90. The median S/N of its spectra are 239 and 283 for the blue and red part of the spectra, respectively. Red lines indicate the typical 1σ (solid line) and 2σ (dashed line) uncertainty of the EW computed with the Cayrel (1988) formula, adopting $FWHM$ = 0.190 Å, pixel size = 0.0232 Å, and S/N = 260. *Bottom row*: a clump giant in the open cluster Trumpler 20 (Trumpler 20 MG 781 in the numbering system of McSwain & Gies 2005), with $T_{\rm eff}$ = 4850 K, log g = 2.75, and [Fe/H] = +0.15. The median S/N of its spectra are 36 and 68 for the blue and red part of the spectra, respectively. Red lines indicate the typical 1σ (solid line) and 2σ (dashed line) uncertainty of the EW computed with the Cayrel (1988) formula, adopting $FWHM$ = 0.190 Å, pixel size = 0.0232 Å, and S/N = 50. In each panel, the average difference of the EWs and its dispersion are also given.

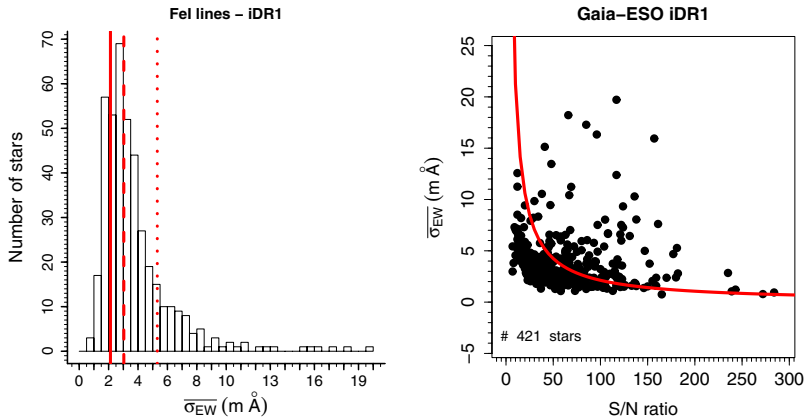

Fig. B.4. *Left panel*: histogram of $\overline{\sigma_{\rm EW}}$ per star, taking into account the measurements of all Nodes. Also shown are lines indicating the 2σ uncertainty calculated with Cayrel (1988) formula for S/N = 40 (dotted line at 5.31 mÅ), S/N = 70 (dashed line at 3.04 mÅ), and S/N = 100 (solid line at 2.12 mÅ). *Right panel*: the dependence of $\overline{\sigma_{\rm EW}}$ with respect to S/N. Also shown is the expected 2σ value given by the Cayrel (1988) formula (as a red line).

(NGC 1851, NGC 2808, NGC 4372, and NGC 5927) and one calibration open cluster (NGC 6705) were analyzed. The NGC 6705 AB-type stars were mostly found to be fast rotators. The results for them were deemed uncertain and were excluded during quality control. In Fig. B.7, we show the final recommended parameters of the stars observed in the cluster fields in comparison with isochrones. The agreement is very good, lending confidence in the final recommended parameters of iDR1.

B.2.3. Method-to-method dispersion

As for iDR2, we compare the results of different Nodes and quantify the method-to-method dispersion of each parameter using the associated median absolute deviation (MAD). The MAD is defined as the median of the absolute deviations from the median of the data.

For the GESviDR1Final results, the median of the method-to-method dispersion is 78 K, 0.17 dex, and 0.07 dex for $T_{\rm eff}$, log g, and [Fe/H], respectively. These values are slightly larger than for iDR2. The third quartile of the distribution has values of 108 K, 0.23 dex, and 0.10 dex for $T_{\rm eff}$, log g, and [Fe/H], respectively. Histograms of these dispersions are shown in Fig. B.8. For $T_{\rm eff}$, the dispersion is within reasonable expectations. For the surface gravity, the dispersion is perhaps too high. However, the surface gravity is a quantity notoriously difficult to derive for

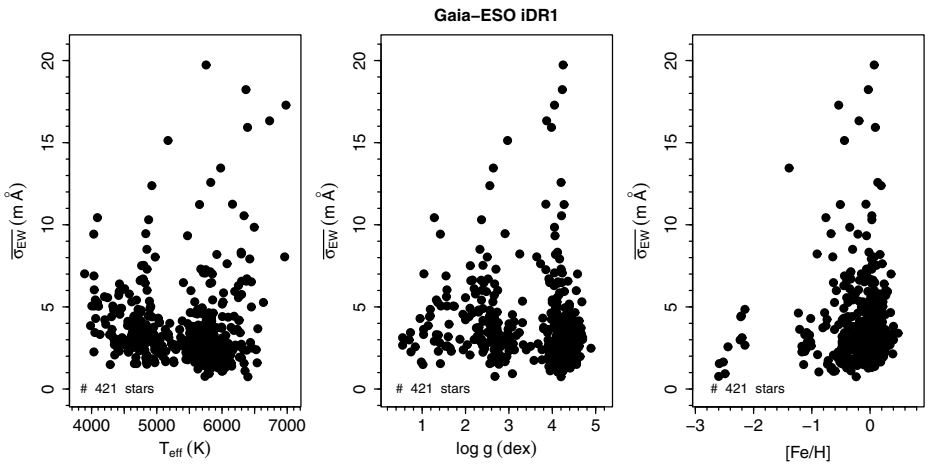

Fig. B.5. Mean of all the standard deviations of the Fe I lines in a star, $\overline{\sigma_{\rm EW}}$, as a function of the atmospheric parameters.

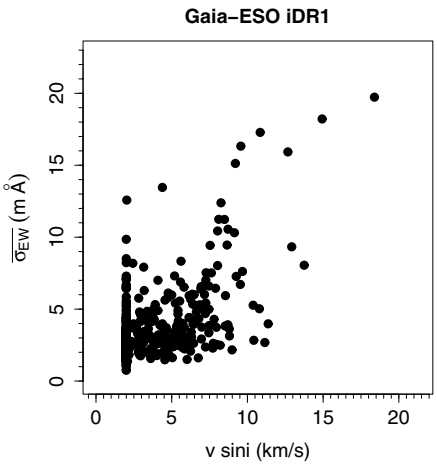

Fig. B.6. Mean of all the standard deviations of the Fe I lines in a star, $\overline{\sigma_{\rm EW}}$, as a function of the rotational velocity of the star.

field stars with uncertain distances. For the metallicity, there is a very good agreement among the Nodes.

B.2.4. Recommended atmospheric parameters

Based on comparisons of individual Node results with the calibrators, as shown above, the following scheme has been adopted to calculate the recommended values of atmospheric parameters of the FGK-type stars with UVES spectra for iDR1:

1. The accuracy of the Node results is judged using the eight available benchmark stars as reference, with a tolerance of ±150 K and ±0.30 dex, for $T_{\rm eff}$ and $\log g$, respectively.
2. Further consistency tests of the Node results are conducted using the calibration clusters.
3. Nodes which fail to reproduce the reference atmospheric parameters of most of the benchmark stars, or that produce unreliable results for the calibration clusters, are disregarded.
4. The median value of the validated results is adopted as the recommended value of that parameter. The median should minimize the effect of eventual outlier results.
5. The MAD is computed to quantify the method-to-method dispersion (analysis precision) and is adopted as an indicator of the uncertainties.
6. The number of results on which the recommended value is based is also reported.

Table B.2. Outcome of the analysis of the iDR1 data.

Gaia-ESO type	Number of stars	Comment
Analyzed stars	508	
Stars with results	421	
GES_MW	271	Milky Way fields.
GES_CL	98	Open clusters fields.
GES_SD	52	Calibration targets.

Notes. Number of FGK-type stars observed with UVES and with atmospheric parameters in the GESviDR1Final internal release.

Table B.2 summarizes the number of stars for which atmospheric parameters were determined during the science verification analysis and are part of the GESviDR1Final internal release. The analysis of a fraction of the stars (~17%) was not completed for different reasons (e.g., high-rotation, double-lined signatures, too low S/N).

A comparison of the recommended values of the atmospheric parameters of the benchmark stars (computed as described above) with the reference values is shown in Fig. B.9. All recommended values of $T_{\rm eff}$ are within ±150 K of the reference values. Good agreement is also present for $\log g$ (within ±0.30 dex), except for HD 140283, a metal-poor subgiant (two spectra of this star were analyzed separately and thus it appears twice in the plot). Gravity values for metal-poor stars are known to be affected by NLTE effects (see, e.g., Bergemann et al. 2012), therefore, it is no surprise that the results of LTE-based analyses shown here are discrepant when compared to the reference values, since the latter are independent from spectroscopy. The results included in GESviDR1Final for metal-poor stars should be used with care. The recommended [Fe/H] values agree with the reference values to within ±0.15 dex.

B.3. Abundances

As for the atmospheric parameters, the elemental abundances were computed in different ways for the iDR1 and the iDR2 datasets.

Multiple determinations of the abundances were conducted. All Nodes that have tools for abundance determinations performed the analysis in parallel for all the stars. For iDR1, the Nodes were asked to compute abundances using two sets of atmospheric parameters for each star, i.e., i) the atmospheric parameters derived by the Node itself; and ii) the set of recommended atmospheric parameters, computed as described above.

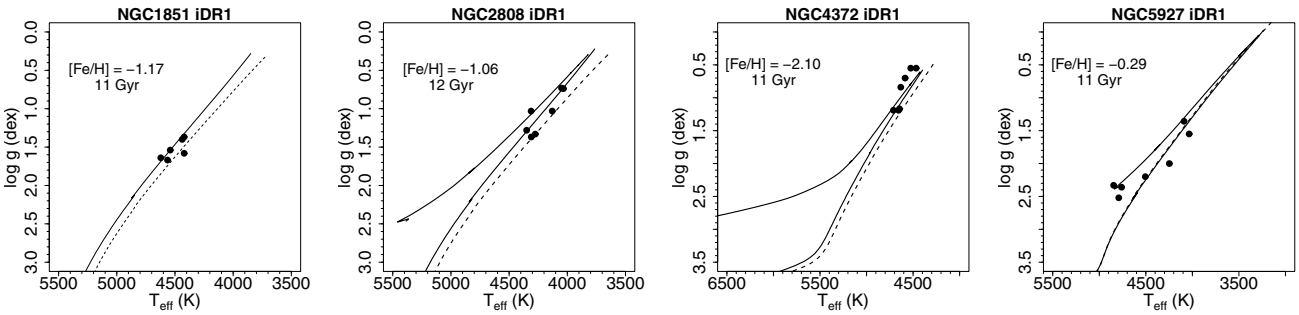

Fig. B.7. Recommended parameters of the stars in the calibration clusters of iDR1 in the $T_{\rm eff}-\log g$ plane. No attempt was made to identify nonmember stars. The plots include all stars observed in the field of the clusters. The ages, metallicities, and isochrones are the same as in Fig. 9.

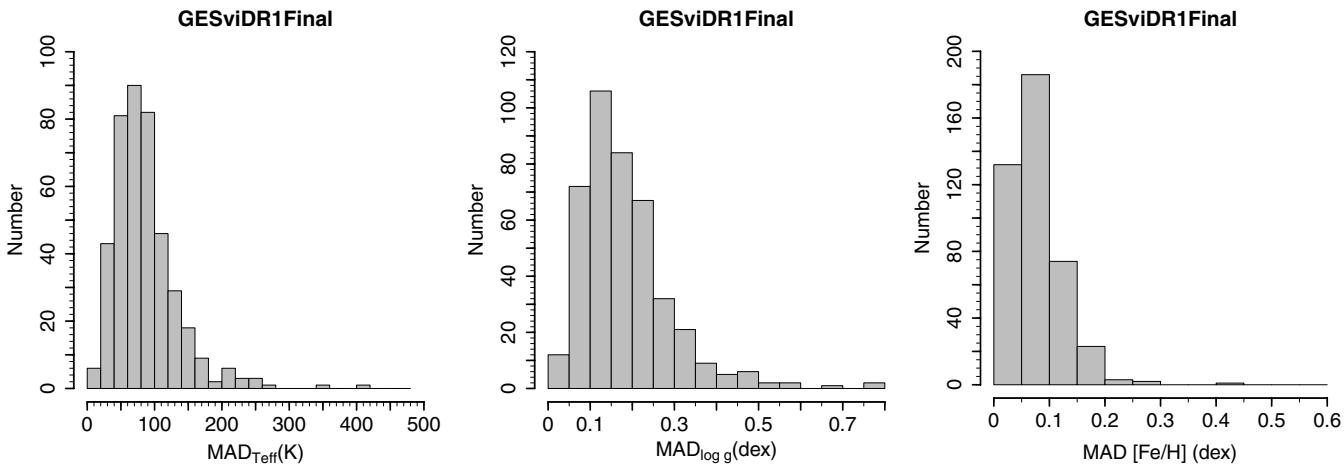

Fig. B.8. Histograms showing the distribution of the method-to-method dispersion of the atmospheric parameters of the 421 stars that are part of GESviDR1Final. *Left*: the dispersion of $T_{\rm eff}$. *Center*: the dispersion of $\log g$. *Right*: the dispersion of [Fe/H].

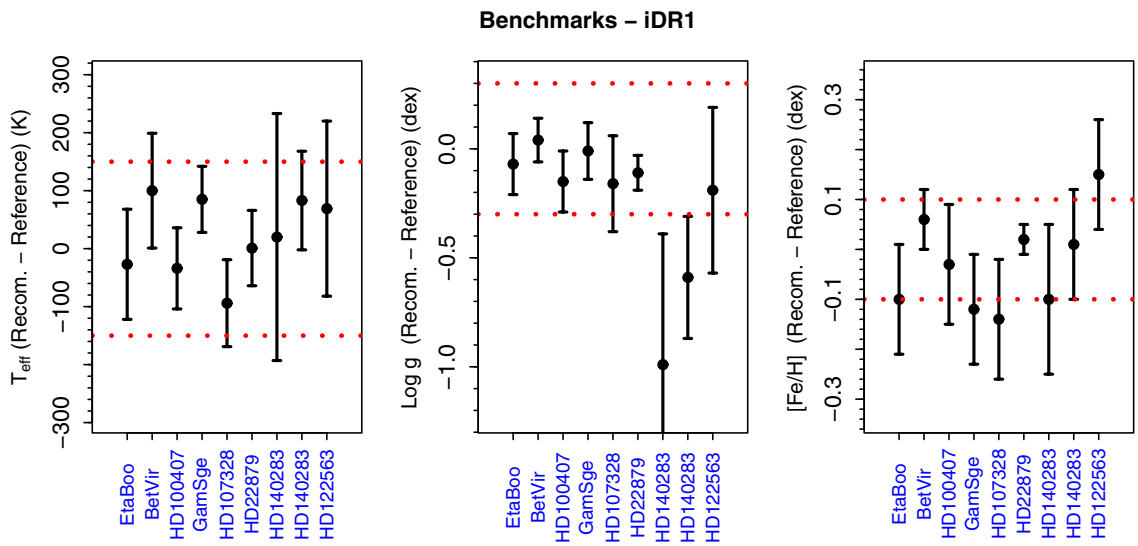

Fig. B.9. Difference between the recommended values of $T_{\rm eff}$, $\log g$, and [Fe/H] for the benchmark stars of GESviDR1Final and the reference values. The error bars are the method-to-method dispersions. The stars are sorted in order of decreasing [Fe/H] (*left to right*). Dotted red lines indicate limits of ±150 K for $T_{\rm eff}$, of ±0.30 dex for $\log g$, and of ±0.10 dex for [Fe/H]. Star HD 140283 appears twice because two different spectra of this star (based on different exposures) were produced and analyzed separately.

We then computed the median of the multiple determinations for each of these two cases. For iDR1, we did not homogenize the line-by-line abundances, but only the final values of each element in each star. In Fig. B.10, we compare the two sets of abundances for a few elements in stars of globular and open clusters. It is clear from this plot that there is no significant difference between the final abundances computed with the two sets of atmospheric parameters.

In addition, it is apparent that the star-to-star scatter of the abundances does not seem to increase when using one or the other set of atmospheric parameters. This lends confidence to the approach adopted here, of having multiple abundances determined by different groups and adopting the median values as the recommended best values.

For the final set of recommended abundances included in GESviDR1Final, we decided to adopt the median of the results

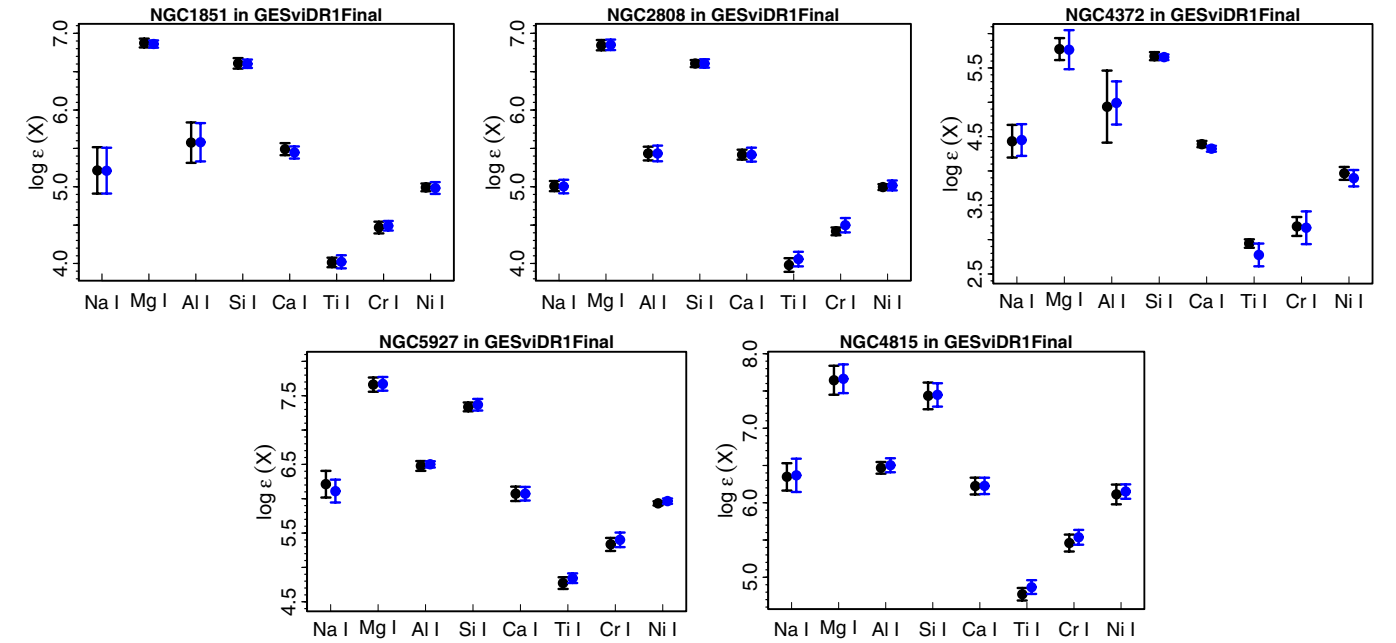

Fig. B.10. Comparison of two sets of abundances for the stars of calibrating globular and open clusters included in iDR1. Points are the averages for all stars observed in the field of a given cluster. No attempt was made to identify nonmembers. Symbols in black are the abundances computed with the recommended atmospheric parameters. Symbols in blue are the abundances computed with the atmospheric parameters of the Nodes. Error bars are the standard error of the mean.

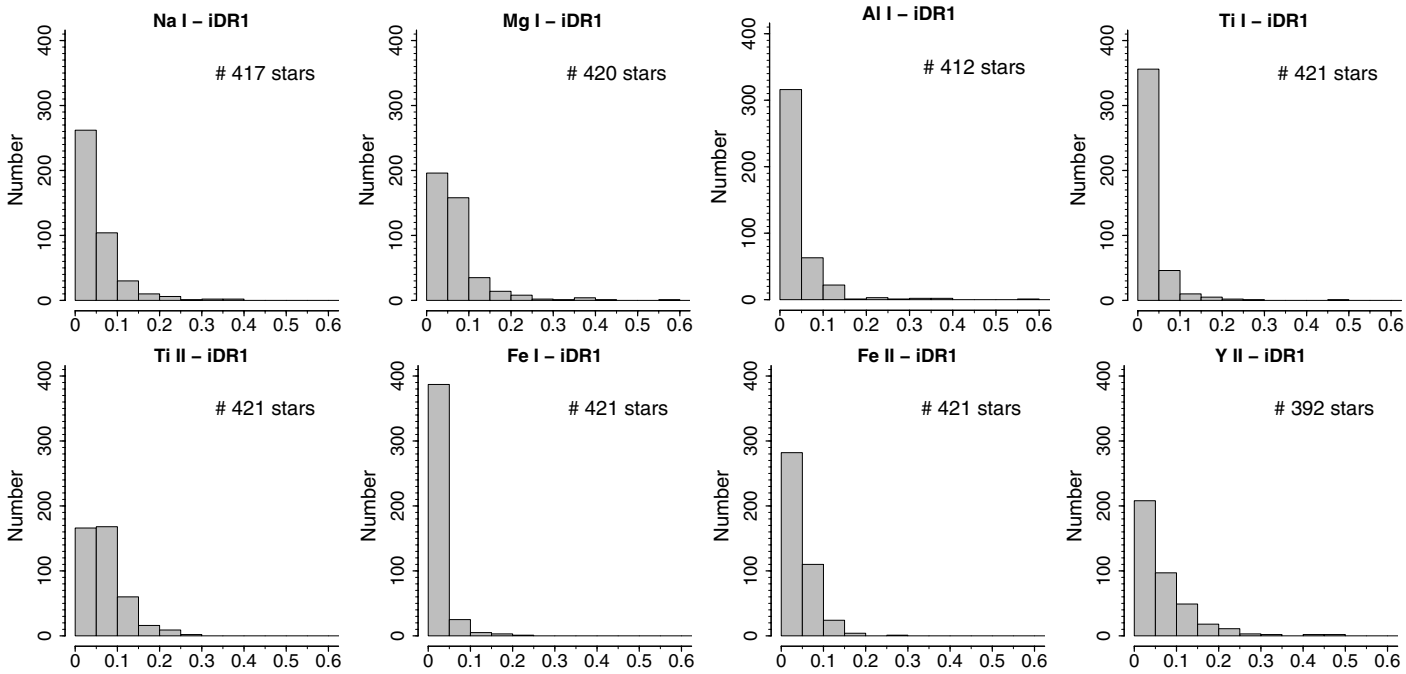

Fig. B.11. Histograms with the method-to-method dispersion of selected species included in the iDR1 results.

calculated using as input the recommended values of $T_{\rm eff}$, $\log g$, [Fe/H], and ξ. The MAD was again adopted as an indicator of the uncertainties (as it is a measurement of the precision with which multiple methods agree). The following 16 elements were analyzed and abundances for at least a handful of stars are included in GESviDR1Final: Li, O, Na, Mg, Al, Si, S, Ca, Ti, Cr, Fe, Ni, Zn, Y, Zr, and Ce. Except for Li, O, S, Zn, Zr, and Ce, all the abundances have been determined by at least three different Nodes. Elements that have important hyperfine structure were not included, as this kind of data were not part of the *Gaia*-ESO line list (version 3.0) when the abundances were calculated.

B.3.1. Method-to-method dispersion

The method-to-method dispersion of the abundances can be used as an indicator of the precision with which the results were derived. In Fig. B.11, we show the histogram of the MADs of a few selected elements. The third quartile of the method-to-method dispersion distribution is equal to or below 0.05 dex for the elements: Al I, Ti I, Fe I, and Ni I. The third quartile of the dispersion is between 0.06 and 0.10 dex for the other elements with multiple measurements: Na I, Mg I, Si I, Ca I, Ti II, Cr I, Cr II, and Fe II. The MADs were adopted as the typical uncertainties.