The Gaia-ESO Survey: Homogenisation of stellar parameters and elemental abundances


(Affiliations can be found after the references)

ABSTRACT

The Gaia-ESO Survey is a public spectroscopic survey that targeted $\gtrsim 10^5$ stars covering all major components of the Milky Way from the end of 2011 to 2018, delivering its final public release in May 2022. Unlike other spectroscopic surveys, Gaia-ESO is the only survey that observed stars across all spectral types with dedicated, specialised analyses: from O ($T_{\text{eff}} \sim 30,000 – 52,000$ K) all the way to K-M ($\lesssim 3,500$ K). The physics throughout these stellar regimes varies significantly, which has previously prohibited any detailed comparisons between stars of significantly different types. In the final data release (internal data release 6) of the Gaia-ESO Survey, we provide the final database containing a large number of products, such as radial velocities, stellar parameters and elemental abundances, rotational velocity, and also, for example, activity and accretion indicators in young stars and membership probability in star clusters for more than 114,000 stars. The spectral analysis is coordinated by a number of working groups (WGs) within the survey, each specialised in one or more of the various stellar samples. Common targets are analysed across WGs to allow for comparisons (and calibrations) amongst instrumental setups and spectral types. Here we describe the procedures employed to ensure all survey results are placed on a common scale in order to arrive at a single set of recommended results for use by all survey collaborators. We also present some general quality and consistency checks performed on the entirety of the survey results.

Key words. Stars: abundances – Stars: fundamental parameters – Methods: statistical

1. Introduction

The launch of the European Space Agency’s astrometric Gaia mission in 2013 (e.g. Perryman et al. 2001; Gaia Collaboration et al. 2016, 2018, 2021) prompted a new wave of Galactic studies. Gaia is delivering precise distances, kinematics, photometry, and spectrophotometry for more than 1 billion stars, as well as radial velocities and chemical abundances for the brighter stars in the sample (see further Gaia Collaboration et al. 2021; Recio-Blanco et al. 2022). A variety of ground-based spectroscopic surveys have been carried out since 2010 to collect complementary stellar parameters, elemental abundances, and radial velocities, which, when combined with Gaia astrometry, have the power to revolutionise our view of the Milky Way. Spectroscopy breaks the degeneracy between foreground extinction and stellar temperature to which the Gaia Blue Photometer and Red Photometer Prism (BP/RP) spectrophotometry data alone are susceptible (Bailer-Jones 2011). The last decade has seen the advent of such spectroscopic surveys in anticipation of Gaia’s exquisite astrometry. Surveys such as the Radial Velocity Experiment (RAVE Steinmetz et al. 2006), GALactic Archeology with HERMES (GALAH De Silva et al. 2015), the Apache Point Observatory Galactic Evolution Experiment (APOGEE Majewski et al. 2017), and the Gaia-ESO Survey (Randich et al. 2022; Gilmore et al. 2022) are complete or close to complete, and the next generation of multi-fibre spectroscopic surveys are underway, such as WEAVE (Jin et al. 2022; Dalton et al. 2012), which is already making observations, 4MOST (de Jong et al. 2014) and MOONS (Cirasuolo et al. 2020), which are dedicating significant effort to operations in advance of observations, and SDSS-V (Kollmeier et al. 2017), which is under development, all of which use multi-object spectroscopy to provide detailed chemical and kinematic information for statistically significant samples of Milky Way stars.

In order to be able to draw accurate conclusions about Galactic structure, formation, and evolution, we need a set of consistent measurements of stellar properties across the Hertzsprung-Russell (H-R) diagram, sampling all major components of the Milky Way. The Gaia-ESO Survey (hereafter, GES) has done this for the first time at high resolution on an 8m class telescope, from the southern hemisphere. GES is a large pan-European effort employing the VLT FLAMES instrument (Pasquini et al. 2002) to obtain high-quality spectra of $\sim 10^5$ stars across the H-R diagram. GES is producing stellar atmospheric parameters, elemental abundances, and radial velocities for all stellar populations, which span the Galaxy from the halo to star-forming regions, sampling the thin and thick discs, the bulge, and open and globular clusters.

Article number, page 1 of 35

©ESO 2023

Astronomy & Astrophysics manuscript no. output
June 23, 2023
The large number of spectra harvested by spectroscopic surveys in the current era requires an automated analysis procedure. Typically, one pipeline is developed and applied to all of the spectra within a survey. However, it is well known that the various spectral-analysis methods suffer from strong systematic uncertainties due in part to factors such as the choice of atmosphere model, or the atomic and molecular transitions employed. Within the Gaia-ESO Survey, considerable effort has been invested in improving the quality of the input line lists (Ruffoni et al. 2014; Heiter et al. 2015b, 2021).

The Gaia-ESO Survey has a unique analysis structure (Gilmore et al. 2022; Randich et al. 2022). While the model atmospheres and the atomic/molecular data are fixed, the data are analysed by a multitude of different analysis teams hereafter referred to as ‘nodes’. Each node runs a pipeline that generally employs a different method from the other nodes, and is executed by experienced spectroscopists that are familiar with the pipeline. In this regard, the Gaia-ESO Survey has a unique advantage over all other spectroscopic surveys in that almost every spectroscopic analysis method ever considered is included in the survey, allowing us to make the first objective comparison between analysis methods, characterise the level of systematic error present in stellar spectroscopy, characterise the random and systematic uncertainty contributions for all measurements, and provide a robust ensemble measurement of stellar parameters and elemental abundances for the survey.

Within GES, a major focus has been placed on producing stellar parameters that are both internally self-consistent and externally calibrated with respect to a well-determined calibration sample of benchmark stars (Jofré et al. 2014; Blanco-Cuaresma et al. 2014). The GES spectra not only cover a wide range of stellar populations (and therefore parameter space) and are analysed with a variety of pipelines, but they are also taken with a variety of instrumental configurations designed to cover the characteristic spectral features of each stellar spectral type.

The effort to transpose such an inhomogeneous set of data and results onto a single, self-consistent scale is not trivial. Essential to this process is the availability of a comprehensive set of calibrators across the H-R diagram. These calibrators include globular and open clusters spanning a wide range in metallicity, as well as the Gaia Benchmark Stars. The design of the observational calibration programme for GES is described in Pancino et al. (2017a). Additionally, to facilitate exploitation of all current and future spectroscopic surveys, we need a practical cross-survey calibration targets and the placing of the stellar parameter and abundance results on a consistent physical scale. GES takes an important step towards a cross-survey calibration by defining this scale.

In this paper, we present the strategy used to homogenise the GES stellar parameters, elemental abundances, and radial velocities and discuss the challenges faced in attempting to define a self-consistent, externally calibrated scale for such a broad parameter range and for such a wide variety of analysis pipelines and methods. The structure of the paper is as follows. Section 2 presents the observations carried out for the GES project and the distribution of the tasks among the different Working Groups. In Section 3, the multi-method, multi-pipeline design of the GES analysis is described along with the homogenisation workflow. Section 4 presents the set of quality checks and tools used to provide the users with a global set of stellar parameters that can be used to compute the elemental abundances. In Section 5, we describe how the abundance homogenisation was performed. Section 6 gathers the sequence of quality checks used to validate the homogenisation process element by element. Section 7 reports the results of the radial-velocity determination of the GES sample and the homogenisation process. In Section 8 we report the determination of the errors for the stellar parameters and the abundances. Section 9 provides a short report on the propagation of the sets of technical and peculiar flags defined by Working Group 14 (WG14) that can be used to trace the analysis of the spectra and their quality. The last section (Sect. 10) summarises the results of the GES survey and presents broad comparisons with other spectroscopic surveys.

2. Observations and spectral-analysis workflow

Gaia-ESO was awarded 300 nights as an ESO Public Spectroscopic Survey on the VLT, with an additional 40 nights subsequently granted to compensate for bad weather and technical downtime. Observations were made between December 2011 and January 2018 using the FLAMES spectrograph (Pasquini et al. 2002) in multi-object spectroscopy mode with the GIRAFFE (ρ ~ 140 fibres) and UVES ( Dekker et al. 2000) (8 fibres for US80, 6 for US20) instruments. The wavelength ranges of the instrumental setups used are listed in Table 1. The GIRAFFE spectra are reduced using the dedicated Cambridge Astronomical Survey Unit (CASU) pipeline (Gilmore et al. 2022). The UVES spectra are reduced using a modified version of the ESO pipeline (Sacco et al. 2014; Modigliani et al. 2004). GES observing blocks are split into two or more exposures and individual spectra are stacked to produce nightly stacked spectra for each field. When all observations from a particular field are complete (certain fields are repeated across nights with arbitrary separation), all spectra for an object are stacked to produce a final stacked spectrum known as a ‘singlespec’ for each object or ‘CNAME’ (the GES object name based on its coordinates, equivalent to the ESO ‘OBJECT’). Where spectra are available in the ESO archive for the GES calibrators and objects in cluster fields in the GES instrumental setups, these have been retrieved, reduced with the GES pipelines, and added to the GES dataset. Radial velocities (RVs) are determined for all individual and stacked spectra (Sacco et al. 2014; Gilmore et al. 2022).

GES internal Data Release (iDR) cycles consist of the following general procedure, which is illustrated in Figs. 1 and 2. In Fig. 1, the general flow is described: targets are selected under the three programmes (Open Clusters, Milky Way, Calibrations), which are observed as necessary using UVES and GIRAFFE (see Randich et al. 2022; Gilmore et al. 2022; Pancino et al. 2017a, for the target selection in the three categories, respectively); raw spectra from a selected time period are reduced and released to the spectral analysis teams from the operational database at CASU in a standard FITS format (Wells et al. 1981) with radial velocities and useful ancillary information such as observing parameters and photometry attached in FITS extensions (the spectral metadata). The teams analyse the data and return catalogues of their results, which are then homogenised to produce a final catalogue of recommended results. Six data analysis cycles (iDRs) were completed as part of GES (See Randich et al. 2022). In Fig. 2, the internal analysis and homogenisation steps are highlighted: the first phase of the determination of the stellar parameters, followed by a homogenisation per WG and a general homogenisation operated by WG15; and the second

---

2. One fibre was broken
There are four spectral-analysis WGs dedicated to the analysis of outliers: operation their specific pipelines on the data within the final WG15 validation. All steps are supported by the use of calibrators.

The structure of the analysis teams in GES is described briefly here. Further details are provided in [Gilmore et al. (2022)](https://www.eso.org/sci/facilities/paranal/3). There are four spectral-analysis WGs dedicated to the analysis of different samples of stars within the GES consortium. Multiple analysis nodes operate their specific pipelines on the data within each WG. A further WG is dedicated to the characterisation of outliers:

1. **WG10**: This WG analyses the GIRAFFE spectra of FGK stars both in the Milky Way (MW) and in open and globular clusters. Four different GIRAFFE setups are used by WG10: (i) HR10-HR21 for MW field stars; (ii) HR15N for FGKM stars in open clusters; (iii) HR9B for stars of earlier types in open clusters; and (iv) HR21 for MW bulge stars.

2. **WG11**: This WG analyses the UVES spectra of FGK stars both in the Milky Way and in open and globular clusters. Two setups are used: U580 for late-type stars, and U520 for early-type stars.

3. **WG12**: This WG analyses the UVES and GIRAFFE spectra of main- and pre-main sequence stars (PMS) in young open clusters with U580 for UVES and HR15N for GIRAFFE.

4. **WG13**: This WG analyses the UVES and GIRAFFE spectra of stars in young clusters containing OBA-type stars. For UVES, the setup is U520, while for GIRAFFE, HR3, HR4, HR5B, HR6, HR9B, and HR14A are used.

5. **WG14**: This WG works to identify and characterise outlier stars from the whole survey. These are stars presenting peculiarities that endanger the determination of their stellar parameters and abundances by standard routines. The stars are tagged using a specially developed library of flags to allow filtering of the dataset during the homogenisation and scientifc analyses.

The characteristics of each setup in UVES and GIRAFFE, in which targets were observed for Gaia-ESO and the working groups that analysed each setup are given in Table 1. Each WG lead collates the results from their nodes and performs an initial homogenisation to put the parameters on a consistent scale for the WG. The interim results including node results files and WG-recommended results are delivered to the CASU operational database pending final homogenisation and delivery to the internal GES Science Archive at WFAU. The final homogenised results form the basis of the public Gaia-ESO catalogues delivered to the ESO archive. A detailed description of the contents of each internal Data Release and each public data release via ESO is contained in [Randich et al. (2022)](https://www.eso.org/sci/facilities/paranal/3).

GES makes use of the set of calibrators selected by the Calibration and Standards Working Group (WGS), which is responsible for the observational calibration strategy for GES [Pancino et al. (2017a)](https://www.eso.org/sci/facilities/paranal/4). By using the calibrators in a uniform way, current precise knowledge for select samples is extended to much larger samples of stars, covering a wider parameter space.

The calibration programme cannot take a one-size-fits-all approach due to the different specialisations of the various WGs. To allow an inter-calibration of the work of the different WGs, GES was designed to have several samples in common amongst the analysis WGs, which are graphically represented in the Venn diagram in Fig. 3. In Fig. 4, we present a Venn diagram of the different setups used for the calibrator samples: in particular, benchmark stars are observed in all combinations of setups, to be analysed by all WGs, whereas other samples of calibrators are observed with the setups more suited for their analysis, for example young calibrator open clusters containing hot stars with U520 and a combination of GIRAFFE setups or calibrator globular clusters with U580 and HR10-HR21. The benchmarks are used to tie the GES results to a well-determined external scale.

The *Gaia* Benchmark Stars were selected to comply with a variety of restrictions such that they serve as a reference, as described in detail in [Heiter et al. (2015a)](https://www.eso.org/sci/facilities/paranal/3) and summarised here. Firstly, each star should have a measurement of its angular diameter, parallax, and bolometric flux. This allows the effective temperature to be determined using the Stefan-Boltzmann relation, which is independent of the assumptions of spectroscopy. Secondly, the stars should adequately sample the parameter space for stellar populations in the Milky Way. This means that the sample is built to include dwarfs and giants, and have a spread in metallicity. Thirdly, the stars need to be located near the equator, such that they can be observed from both hemispheres. The parameters are determined with the following procedure. First, the effective temperature is determined using fundamental relations ([Heiter et al. (2015a)](https://www.eso.org/sci/facilities/paranal/3)). Then, surface gravity is determined using temperature, bolometric flux, parallax, and mass from a stellar track. Finally, $T_{\text{eff}}$ and log $g$ are fixed to the values obtained above and chemical abundances, including metallicity, are derived from high-signal-to-noise-ratio (high-S/N) and high-resolution optical spectra ([Blanco-Cuaresma et al. (2014)](https://www.eso.org/sci/facilities/paranal/4)) by various spectral analysis methods ([Jofré et al. 2014, 2015](https://www.eso.org/sci/facilities/paranal/3)). This leads to a sample of 34 stars with accurate temperatures —with a precision of about 100 K—, surface gravities, and abundances

---

Fig. 2: Gaia-ESO data-processing diagram for iDR6 showing the complexity of the interfaces between the reduction, parameter, abundance, and homogenisation processes. The GRR is the GES Results Repository at CASU.
with precisions of 0.1 dex. The latest catalogue can be found in Jofré et al. (2018).

One of the main motivations to assemble the Gaia Benchmark Star dataset was the calibration of GES. To this end, substantial spectral analysis of the sample has been performed in the manner of GES – namely, using a combination of analysis methods (and adopting many of the GES WG11 methods), using the same line list and atmosphere models, and looking for new candidates to match the needs of GES.

The original set of Gaia benchmarks was expanded for the final release of GES to include stars at lower metallicity and at higher and lower temperatures (warm and cool benchmarks) to better cover the parameter space of the diﬀerent science samples (which are described in e.g. Stonkutė et al. 2016). A set of hot stars with well-studied stellar parameters was included for WG 13 and several cool M-dwarf stars were included for WG 12. A sample of metal-poor candidates was proposed by Hawkins et al. (2016), and a workshop to understand specific differences of the analysis methods in GES using the benchmarks was organised (Jofré et al. 2017).

The Kiel diagram of the final sample of benchmark stars available in the final release iDR6 is presented in Fig. 6. The sample of benchmark stars divided into warm benchmarks (GE_SD_BW), FGK benchmarks (GE_SD_BM or AR_SD_BM), and cool benchmarks (GE_SD_BC) covers the parameter space mapped by the various WGs of GES.

Another two main classes of calibrator used in GES are well-studied open and globular clusters. Calibration using clusters is especially important for the WGs operating on stars at the edges of or outside the FGK range and to test the method on groups of stars that have the same ages, distances, and metallicities but different masses and evolutionary phases. The calibration open clusters are observed in setups matching the Milky Way field (MW) setups and those of the globular clusters, in addition to the setups used for the open cluster science. The literature metallicities of the final set of calibrator clusters for iDR6 are listed in Table 2.

A relative newcomer to calibration sets for stellar surveys is samples for which asteroseismic measurements are available upon which log g is determined. GES included the observation of targets from two key asteroseismic missions in its calibration plan, namely CoRoT and K2 (Pancino et al. 2017a).

### Table 1: Table of instrumental set-ups used by the analysis working groups within GES. The GIRAFFE instrument was refocused in February 2015. The Rold values refer to pre-refocusing values and R gives the new values, post-refocusing.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Setup</th>
<th>( \lambda_{\text{min}} ) (Å)</th>
<th>( \lambda_{\text{max}} ) (Å)</th>
<th>Rold</th>
<th>R</th>
<th>WG</th>
</tr>
</thead>
<tbody>
<tr>
<td>UVES</td>
<td>520</td>
<td>4140</td>
<td>6210</td>
<td>47000</td>
<td>47000</td>
<td>11,12,13,14</td>
</tr>
<tr>
<td>UVES</td>
<td>580</td>
<td>4760</td>
<td>6840</td>
<td>47000</td>
<td>47000</td>
<td>11,12,13,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR3</td>
<td>4033</td>
<td>4201</td>
<td>24800</td>
<td>31400</td>
<td>13,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR4</td>
<td>4188</td>
<td>4392</td>
<td>32500</td>
<td>35550</td>
<td>13,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR5A</td>
<td>4340</td>
<td>4587</td>
<td>18470</td>
<td>20250</td>
<td>13,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR6</td>
<td>4538</td>
<td>4759</td>
<td>20350</td>
<td>24300</td>
<td>13,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR9B</td>
<td>5143</td>
<td>5356</td>
<td>25900</td>
<td>31750</td>
<td>10,13,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR10</td>
<td>5339</td>
<td>5619</td>
<td>19800</td>
<td>21500</td>
<td>10,12,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR14A</td>
<td>6308</td>
<td>6701</td>
<td>17740</td>
<td>18000</td>
<td>13,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR15N</td>
<td>6470</td>
<td>6790</td>
<td>17000</td>
<td>19200</td>
<td>10,12,14</td>
</tr>
<tr>
<td>GIRAFFE</td>
<td>HR21</td>
<td>8484</td>
<td>9001</td>
<td>16200</td>
<td>18000</td>
<td>10,12,14</td>
</tr>
</tbody>
</table>

### Table 2: Literature metallicities of globular and open clusters adopted as calibrators.

<table>
<thead>
<tr>
<th>Name</th>
<th>[Fe/H]</th>
<th>type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>M15</td>
<td>−2.37</td>
<td>GC</td>
<td>Harris (1996)</td>
</tr>
<tr>
<td>NGC4590</td>
<td>−2.23</td>
<td>GC</td>
<td>Harris (1996)</td>
</tr>
<tr>
<td>NGC4372</td>
<td>−2.17</td>
<td>GC</td>
<td>Harris (1996)</td>
</tr>
<tr>
<td>NGC4833</td>
<td>−1.92</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>M2</td>
<td>−1.47</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>NGC1904</td>
<td>−1.51</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>NGC6752</td>
<td>−1.48</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>M12</td>
<td>−1.37</td>
<td>GC</td>
<td>Harris (1996)</td>
</tr>
<tr>
<td>NGC1261</td>
<td>−1.27</td>
<td>GC</td>
<td>Harris (1996)</td>
</tr>
<tr>
<td>NGC362</td>
<td>−1.12</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>NGC1851</td>
<td>−1.07</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>NGC2808</td>
<td>−1.03</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>NGC104</td>
<td>−0.71</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>NGC5927</td>
<td>−0.39</td>
<td>GC</td>
<td>Pancino et al. (2017b)</td>
</tr>
<tr>
<td>NGC6553</td>
<td>−0.25</td>
<td>GC</td>
<td>Harris (1996)</td>
</tr>
<tr>
<td>NGC2243</td>
<td>−0.48</td>
<td>OC</td>
<td>Pancino et al. (2017a)</td>
</tr>
<tr>
<td>Berkeley32</td>
<td>−0.21</td>
<td>OC</td>
<td>Friel et al. (2010)</td>
</tr>
<tr>
<td>Melotte71</td>
<td>−0.27</td>
<td>OC</td>
<td>Pancino et al. (2017a)</td>
</tr>
<tr>
<td>NGC2420</td>
<td>−0.05</td>
<td>OC</td>
<td>Pancino et al. (2017a)</td>
</tr>
<tr>
<td>M67</td>
<td>+0.03</td>
<td>OC</td>
<td>Randich et al. (2006)</td>
</tr>
<tr>
<td>NGC3532</td>
<td>+0.00</td>
<td>OC</td>
<td>Pancino et al. (2017a)</td>
</tr>
<tr>
<td>NGC2477</td>
<td>+0.07</td>
<td>OC</td>
<td>Pancino et al. (2017a)</td>
</tr>
<tr>
<td>NGC6705</td>
<td>+0.12</td>
<td>OC</td>
<td>Pancino et al. (2017a)</td>
</tr>
<tr>
<td>NGC6253</td>
<td>+0.34</td>
<td>OC</td>
<td>Pancino et al. (2017a)</td>
</tr>
</tbody>
</table>

### 3. Homogenisation procedure

The multi-method, multi-pipeline design of the GES analysis, implemented through the analysis node and WG structure outlined above, means that multiple results are delivered for many GES stars. This includes both parallel analyses of the same stellar parameters, elemental abundances, radial and rotational velocities, other specific quantities, and flags per star (or per CNAME). A schematic view of the GES analysis approach is presented in Fig. 7: the analysis process starts from
A&A proofs: manuscript no. output

Fig. 3: Venn diagram describing the role of calibration samples in the interconnection between the WGs.

Fig. 4: Venn diagram describing the setups with which the calibration samples were observed (see also Table 1).

Fig. 5: Histogram of the metallicities covered by the calibrator clusters (globular clusters in blue, open clusters in orange).

the nodes, which transmit their results to the WG. The first step, indicated by red arrows, denotes the determination of stellar parameters. Once homogenised by WG15, the stellar parameters are transmitted back to those nodes that determine the abundances. These are then homogenised by the WGs, and finally combined, together with the stellar parameters, by WG15 in the final database.

WG15 was led by P. François (from DR1 to DR6) and over the course of the survey the working group members were A. Hourihane (DR1-DR6), C. Worley (DR1-DR6), L. Magrini (DR2-DR6), A. Gonneau (DR5-DR6), and A. Casey (DR3-DR5).

3.1. Quality and format checks

The first step of the data analysis is to perform sanity checks on the files provided to WG15 from the four analysis WGs. A first pass is performed by a dedicated automated tool (the FITSchecker which was developed and maintained over the lifetime of the survey by C. Worley, A. Casey, D. Murphy, and A. Gonneau). This tool flags issues in the file formats, data statistics, and data completeness in a report that is sent to the submitter. The submitter must resolve any issues and resubmit the file until it is accepted by the tool as adhering to the GES data model. The GES data model is described in two technical documents governing the spectral formats from the spectral reduction pipelines and the analysis results catalogues from the nodes and WGs. WG15 members also carried out a visual inspection of the data statistics summarised in the FITSchecker report to identify any remaining spurious values and outliers. These were raised with the relevant WG leads for resolution.

5 Gaia-ESO Spectral Data and Formats, July 2013
In this section, we summarise the homogenisation flow; the quality checks performed during each step are described in Secs. 4, 5, and 7.1. The homogenisation workflow starts with the application of an algorithm that defines a set of rules to obtain the best set of stellar parameters in the case of multiple observations with different setups. This choice is not only based on spectral resolution or S/N, but also on which type of observation is best suited to the type of star and which WG uses the most appropriate methods; for example WG13 for hot stars or WG12 for cool stars.

On the one hand, the internal analysis processes of WG10, WG11, and WG12 were fully consistent in the last data release and have provided stellar parameters on the same scale thanks to continuous interaction between the WG leads and the WG15 team (see Smiljanic et al. 2014; Lanzafame et al. 2015; Worley et al. 2023, for details on the analysis of each WG). In particular, in the last data release, the analysis of the U580 and U520 spectra assigned to WG12 was performed by all WG11 nodes and was included in the homogenisation workflow of WG11, ensuring consistent treatment of the data. Similarly, the WG12 GIRAFFE spectra (HR15N) were homogenised with the same code as the WG10 results.

Details about the mapping of the WG10 and WG12 results on to WG11 are given in Worley et al. (2023). On the other hand, the WG13 results are located in a different region of the parameter space, and obtained with different methods. Therefore, they are treated separately and not homogenised with the results of the other WGs.

For these reasons, the WG15 algorithm does not apply further corrections (offsets or linear relations) to stellar parameters coming from the various different WGs. In the end, the homogenisation algorithm allows us to have a single set of parameters for each CNAME, uniquely chosen following the flow represented in Figure 8.

4. Stellar parameter homogenisation: Quality checks

The aim of the homogenisation process is to provide a recommended set of stellar parameters that are consistent with each other regardless of the setup used (medium or high resolution and the covered spectral range). To this purpose, the GES strategy acts with several different tools, including: (i) a set of well-defined calibrators, that is, the benchmark stars described above, which cover the whole parameter space ($T_{\text{eff}}$, $\log g$, [Fe/H]); (ii) a sample of targets observed with different setups, and whose stellar parameters are derived by different WGs; (iii) the Kiel diagrams of stars in the Milky Way fields that have metallicity in a given, restricted metallicity interval to be compared with the corresponding theoretical isochrones (we use Kiel diagrams rather than H-R diagrams as we have an estimate of the $\log g$ rather than the luminosity of each object); (iv) member stars in open and globular clusters, which share the same age and metallicity, and can be considered simple stellar populations, at least to a first approximation, and thus their stellar parameters are directly comparable with the corresponding isochrones; (v) a sample of asteroseismic targets observed in the K2 and CoRot fields.

The various subsamples are used directly as calibrators to map the results onto the reference ones (in the internal WG procedures, see Worley et al. (2023)), or as final checks of the WG results and on the final set of global parameters by WG15.

4.1. Benchmark stars

The Gaia Benchmark Stars described above are used as a reference set during the WG homogenisation to define the parameter scale. The benchmark sample has been expanded from the initial set to better cover the parameter space needed for the global
homogenisation of the survey results. The sample available in iDR6 contains 42 stars in total, that is, 21 FGK stars, 16 warm benchmarks (OBA stars), and 5 cool benchmarks (M stars).

As part of the quality checks on the parameters, several diagnostic plots are used. For example, the difference between the $T_{\text{eff}}$, log $g$, and [Fe/H] determined by each of the Working Groups for the benchmarks and the reference value is plotted with the result selected by the WG15 algorithm highlighted in order to ensure the appropriate quality of the results selected according to the rules.

The homogenised WG15 results are shown in Fig. 9. The plots of Delta Parameter per benchmark are ordered according to benchmark reference metallicity. For better visualisation of the results, the x-axis contains the list of benchmarks. The results demonstrate good agreement across the parameter space of the GES results with the literature ones. Two stars ($\tau$ Sco and $\gamma$ Peg) show a temperature difference of greater than 500 K, but these two stars are warm stars with temperatures above 22,000 K. A gravity difference of $-0.68$ dex is found for 32 Gem, but this is within the error estimates. This star is an A9III-type star for
which the gravity is difficult to estimate. Meanwhile, the GES parameters for the Sun are determined from archival spectra contained in the FLAMES solar atlas; they are reduced with the GES pipelines and homogenised to values of $T_{\text{eff}}$ of 5751 ± 11 K, log $g$ of 4.35 ± 0.02 dex, and [Fe/H] of 0.02 ± 0.02 dex. These are consistent with [(Fe/H)] or close to ($T_{\text{eff}}$, log $g$) the literature values for the Sun from Heiter et al. (2015a).

4.2. Milky Way field

The MW fields contain both UVES (WG11) and GIRAFFE (WG10) spectra. Target selection is described in Stonkutė et al. (2016). With the assumption that the two sets of spectra sample the same stellar population, we can compare the distributions in the $T_{\text{eff}}$–log $g$ plane for the stars for both samples and check for offsets in $T_{\text{eff}}$ and/or log $g$ (see Figure 10 for the bin centered at solar metallicity and for the one at [Fe/H] = −0.5). The isochrones are only representative and do not correspond to a specific age–metallicity relation fitted to the sample. Their parameters are compatible with the metallicity bin used in each plot. They are plotted as representative of the shape and the location of the main sequence and the giant branch.

4.3. Calibration open and globular clusters

As described in Pancino et al. (2017a), clusters form an essential part of the calibration set and provide a way to compare simple stellar populations of known ages with the outputs of theoretical models. In addition, they allow us to map stars analysed by different WGs onto a common scale when no actual stars in common are present between WGs (which is an unavoidable issue when dealing with WGs, which work on e.g. massive hot cluster stars and cool PMS stars). Specific clusters were observed for calibration. In addition to the calibration clusters, we also make use of clusters observed as science targets, in particular those in which a large number of stars were observed. Clusters were used to validate the results in various ways. First, we verified the good agreement in the log $g$–$T_{\text{eff}}$ plane with the isochrones corresponding to the age and metallicity of the various clusters. In addition, using the membership published in Jackson et al. (2022), we selected members of open clusters with a probability of membership $> 0.99$. Membership analysis for Melotte 71 and Br32 is not available in Jackson et al. (2022), and so we plot the members selected on the basis of their radial velocity in the figure. We computed the average metallicity of their members examined by the different WGs and used these to identify possible offsets due to the analysis process. An example is shown in Figure 11, where we plot the stellar parameters of the members of NGC6705 —one of the nine calibration open clusters— obtained by three different WGs and homogenised by WG15. We also compare the metallicity obtained from the different samples of stars, with the mean metallicity of the cluster provided in Randich et al. (2022). In this figure, we show the results of WG10 obtained with the combination of the two setups HR10 and HR21, which are used to observe the MW fields and for calibration purposes in calibration clusters and in benchmark stars. Although the data obtained at lower resolution from WG10 and for hot stars from WG13 show a larger scatter than the data from WG11, there is good agreement with the isochrones for the stellar parameters, and with the average metallicity of the cluster as derived by different WGs. In Fig. 12, we show the Kiel diagrams of the other calibration open clusters, adding an illustrative isochrone for each cluster. The PARSEC isochrones are selected at the cluster metallicity and age, considering a typical uncertainty in the age determination as given in Cantat-Gaudin et al. (2020). Overall, the agreement in the parameters along the isochrone is very good for the results obtained by the different WGs, with an excellent tracking of the cluster sequence.

There are stars from 15 globular clusters in GES iDR6, which were selected such that the globular clusters span a wide range in metallicity (see Fig. 5 for their metallicity distribution). These were observed with HR10, HR21, and U580, the setups used for the Milky Way fields. Figure 13 shows the Kiel diagram for each GC.

The stars shown for each globular cluster are those defined as cluster members using Gaia DR3 proper motions and GES radial velocities as described in (Worley et al. 2023). An illustrative isochrone at the reference values for age and metallicity for each globular cluster is shown in Figure 13. For each cluster, the WG11 (black) and WG10 (blue) results are shown with the median metallicity calculated for each set of WG results, as well as the median of the iDR6 recommended metallicities for that cluster. The reference metallicity is also provided.

Overall there is very good agreement between the stellar parameters of GES globular cluster stars and the isochrones. By distinguishing between WG11 and WG10 by colour, the agreement between the two WGs along the stellar evolution sequences confirms the consistency between the two sets of results. However, the WG10 results for two globular clusters, M12 and NGC2808, show a non-trivial disagreement with the respective isochrone, indicating further detailed study is warranted, although this is outside the scope of this paper.

Figure 14 shows the offset between the WG10 and WG11 mean [Fe/H] for the calibrating globular and open clusters. For each cluster, the median difference between WG10 and WG11 is represented as blue circles. Orange dotted lines show offset limits at ±0.2 dex. The figure shows the quality of the agreement between the results of WG10 and WG11. M2 is the only cluster with a noticeable bias, which is of the order of 0.15 dex. We measured a mean bias of 0.04±0.07 dex between the WG10 and the WG11 [Fe/H] median values.

4.3.1. Asteroseismic calibrations

The agreement with the log $g$ of the asteroseismic samples of CoRoT and K2 as described was investigated by WG15. For WG10, 1,512 CNAMEs had both CoRoT and GES log $g$, while 28 CNAMEs had both K2 and GES log $g$. For WG11, 86 CNAMEs had both CoRoT and GES log $g$, while 62 CNAMEs had both K2 and GES log $g$. The differences between these values and the seismic log $g$ values are shown in Figure 15, and the median difference and standard deviation in each set are provided.

In three of the four datasets, there is good agreement between the seismic and GES log $g$ values. However, the WG10 log $g$ values of the K2 sample are overestimated by 0.28±12 dex. While the agreement between the WG10 values and CoRoT is good, there is large scatter ($−0.07 ± 0.51$). The nature of the WG10 spectra, which span a shorter wavelength range and have a lower resolution than the WG11 spectra most likely contributes to both effects as noted in Worley et al. (2020). More extensive calibration samples combining spectroscopy and asteroseismology are needed to explore and refine this approach, which is being pursued by upcoming surveys.
Fig. 9: WG15 parameters versus reference parameters as described in Pancino et al. (2017a). Warm benchmarks (OBA stars) are marked with blue squares, while the cool benchmarks (M stars) are marked with red stars. The stars are presented with ascending [Fe/H] along the abscissa.

4.3.2. Stars in common between Working Groups

The GES strategy consists of having a number of targets observed in several setups, and consequently analysed by several WGs. There are about 708 spectra in common between WG11 and WG10, with most of them being benchmark stars and stars in clusters. There is also a small number of stars in the MW field. Figure 16 shows the comparison between the difference WG10−WG11 versus WG10 results for $T_{\text{eff}}$, log $g$, and [Fe/H]. The results are in very good agreement, with a low median difference for the different parameters.

There are more than 600 spectra in common between WG11 and WG12. Most of them are stars in clusters analysed by both working groups. Also, there are the usual calibrators, mainly benchmark stars. In Figure 17, we show the comparison between WG12 and WG10 results for $T_{\text{eff}}$, log $g$, and [Fe/H]. The differences between WG12 and WG11 are small, indicating very good agreement. There is no systematic offset to apply given the very low median difference compared to the dispersion. However, there is a trend in the [Fe/H] difference with an increase in the difference when the metallicity decreases. A substantial difference is also found for the difference in gravity at low log $g$, below log $g \approx 1$ dex.

5. Abundance homogenisation

In the following section, we discuss the homogenisation of the abundances derived by WG10 and WG11. We do not include the abundances of WG12 and WG13 in the discussion for the following reasons: On the one hand, in the final release the abundances of WG12 obtained at high resolution with UVES were
treated consistently with those of WG 11, following the same analysis flow; they therefore became part of the WG11 sample of abundances. On the other hand, medium-resolution observations of WG12 with the HR15N setup allow us to measure only Li, and Li abundance is homogeneously determined for the entire survey by a single node independently, the analysis of which is described in detail in Franciosini et al. (2022).

Finally, the abundances of WG13 are obtained for stars in different regions of the parameter space: the derived abundances are for different elements and ionisation states with respect to those obtained in FGK stars, often strongly influenced by diffusion and non-LTE effects, thus not directly comparable with those of cooler stars, even if they belong to the same cluster. We refer to Blomme et al. (2022) for a complete description of the process of analysis and homogenisation of WG13 spectra.

Figure 19 illustrates the rules applied in order to homogenise the elemental abundances from WG10, WG11 (including the analysis of the WG12 UVES spectra), and WG13.

Here, we focus on the WG10 and WG11 abundances, whose elemental abundances for the different setups are computed starting from the homogenised stellar parameters. As shown in the analysis of the abundances derived from the different nodes in WG11 (Smiljanic et al. 2014), this does not guarantee that they are automatically perfectly consistent. The process of mapping the results of WG10 onto those of WG11, described in detail in Worley et al. (2023), alleviated any eventual discrepancy between the results of the two WGs. As the large wavelength range...
Calibration clusters

Fig. 12: Kiel diagram of the calibration open clusters with homogenised data from WG10 (blue), WG11 (black) and WG13 (magenta) and the PARSEC isochrones for the age and metallicity of the clusters. In the plot, we include only stars with \(E_{\log g} < 0.35\).

and the high resolution of the UVES spectrograph permits a more precise determination of the stellar parameters and abundances than the GIRAFFE spectra do, the abundances obtained from the UVES spectra are taken as a reference for both the process of homogenisation and for defining the final checks. The task of WG15 in this final data-release cycle is therefore limited to a final set of checks for consistency and homogeneity of the results using the various tools and calibrators available. In what follows, we describe the main quality checks performed on the elemental abundances.

The abundances of light elements Li, C, N, and O do not enter into the abundance homogenisation cycle because they are derived by single nodes: the Arcetri node for Li and the Vilnius node for CNO. The lithium abundance is measured from the doublet lines at 670.8 nm in the U580 and HR15N setups. At the HR15N resolution, the doublet is blended with the nearby FeI line at 670.74 nm, while the two components are separated in UVES. In the final release, the Li abundances were derived in a homogeneous way by means of the equivalent widths (EWs) using a set of curves of growth (Franciosini et al. 2022) specifically derived for GES.

In the case of GIRAFFE, where only the total blended Li + Fe EW can be measured, the Li-only EW was first computed by applying a correction for the Fe blend. When the line is not visible or barely visible, an upper limit to the EW is given that is equal to the uncertainty, or to the measured EW if higher.

The abundances of C, N, and O are derived from molecular bands and atomic lines, with a simultaneous fit of the three abundances, as described in Tautvaišienė et al. (2015). The \(^{12}\)C-\(^{14}\)N molecular bands 6470-6490 Å, the \(^{2}\)C Swan (1,0) band head at 5135Å, the \(^{2}\)C Swan (0,1) band head at 5635.5Å, and the forbidden [O I] line at 6300.31Å are used. The analyses are performed through spectral synthesis with the Turbospectrum code (Plez 2012). For the determination of the oxygen abundance, the oscillator strengths of the two lines of Ni are taken into account (Johansson et al. 2003).

The carbon abundance is also derived from atomic lines (C I). In this case, the abundance is derived by several nodes, and homogenised with the same procedure as for the other elements. In the final database, the abundances of C and N from molecular bands are indicated with C_C2 and C_CN.

6. Quality checks on the elemental abundances

In this section, we present some examples of the quality checks performed on the final abundance database. Figures 20 and 21 show the abundance ratios of all the elements analysed in GES (except Li) as a function of [Fe/H]. The different colours represent the results from the different Working Groups.

6.1. Elemental abundances in benchmark stars

Quality checks include a visual inspection of the elemental abundance results in a variety of views, such as the abundance against [Fe/H] or per benchmark, with one plot per elemental species and with the results colour-coded by WG where individual WG results are plotted. For the elements for which reference abundances are available (the ten elements, besides Fe, in Jofré et al. 2015), the plots of delta abundance with respect to the reference abundances are generated and viewed. Figure 22 shows the delta abundance of the WG15 values minus the reference values per benchmark. The 15 panels represent the ten elements with reference values; for some of these we report results for more than one ionisation species (such as Ca I and Ca II). The comparison to
iDR5 was additionally checked and showed good consistency or improvement in the delta abundance results for iDR6. Results for the Sun are included in the delta abundance plots. See Randich et al. (2022) for more details on the quality of the GES solar abundances.

By visually checking the WG results with the selected WG15 results highlighted, we were able to identify updates that could be made to the homogenisation rules; for example, in cases where an abundance result was not available for particular benchmark stars in the preferred WG according to the homogenisation rules we implemented (e.g. WG11), but a result from another appropriate source was available (e.g. WG10). Whereas results from different WGs are not mixed for the stellar parameters (i.e. the parameters $T_{\text{eff}}$, $\log g$, and [Fe/H]) will always come from the same WG for a particular star), the rules for the abundance results are relaxed to allow mixing of the WG from which the results originate.

6.2. Elemental abundances in field stars

Figure 23 displays the abundances in the [X/Fe] versus [Fe/H] planes of the elements in common between WG10 and WG11,
and the smaller wavelength range. The smaller wavelength range can be explained by the lower dispersion of the spectra used by WG10 compared to the WG11 results, which is naturally from the very high-resolution (40,000-110,000), very high-dispersion in the \([\alpha/Fe]\) and \([\alpha/Fe]/[Fe/H]\) ratios as a function of \([Fe/H]\) in member stars of four calibration open clusters and in four calibration globular clusters. We computed the median abundances separately for member stars observed with the GIRAFFE setups HR10/HR21 and analysed by WG10 and for those observed with U580 and analysed by WG11. We removed the stars with E\_LOGG > 0.25 and with an uncertainty on \([\alpha/Fe], [\alpha/Fe] > 0.15 from computations of the median abundances. The members are selected to have a probability MEM3D > 0.95. Although U580 and HR10/HR21 are very different setups in terms of resolution and spectral range, the figures show that the homogenisation process produces values in very good agreement, in some cases within 1\(\sigma\), and in slightly worse cases within 2\(\sigma\), such as for NGC 2420. The comparison indicates that WG10 results are on average accurate (similar median values to WG 11) but—as expected given their lower resolution and shorter spectral coverage—less precise.

### 6.3. Elemental abundances in the combined cluster sample

In Fig. 24 the results for the open and globular clusters are combined. For each cluster, the difference in the median abundances between WG10 and WG11 are computed and plotted. The orange line indicates the median offset computed for the sample of globular and open clusters. The dotted orange lines indicate the standard deviation around the median value. We note that the offsets versus metallicity are in good agreement for both open and globular clusters. Discrepancies are found in some globular clusters for some elements that have only a few absorption lines measurable given the moderate resolution of the WG10 spectra. The abundances of members of clusters derived by WG10 and WG11 can be used to estimate the final accuracy and precision of our homogenised abundances. Figures 25 and 26 display, as an example, the average \(\alpha\)-element abundance over iron, which is computed by averaging \([Mg/Fe], [Si/Fe], [Ca/Fe],\) and \([Ti/Fe]\) as a function of \([Fe/H]\) in member stars of four calibration open clusters and in four calibration globular clusters. We computed the median abundances separately for member stars observed with the GIRAFFE setups HR10/HR21 and analysed by WG10 and for those observed with U580 and analysed by WG11. We removed the stars with E\_LOGG > 0.25 and with an uncertainty on \([\alpha/Fe], [\alpha/Fe] > 0.15 from computations of the median abundances. The members are selected to have a probability MEM3D > 0.95. Although U580 and HR10/HR21 are very different setups in terms of resolution and spectral range, the figures show that the homogenisation process produces values in very good agreement, in some cases within 1\(\sigma\), and in slightly worse cases within 2\(\sigma\), such as for NGC 2420. The comparison indicates that WG10 results are on average accurate (similar median values to WG 11) but—as expected given their lower resolution and shorter spectral coverage—less precise.

### 7. Radial and rotational velocity homogenisation

#### 7.1. Radial velocities

The *Gaia* Radial Velocity Standard (GRVS; *Soubiran et al.* 2013) stars were used to investigate the robustness of the Gaia-ESO RVs. While part of the calibration strategy, the 29 GRVS stars were observed in only five of the GES setups: HR10, HR15N, HR21, HR9B, and U580. The RVs for all of the setups were calculated as described in *Gilmore et al.* (2022). The RVs were calculated on the stacked spectra upon which the stellar parameters and abundances were also determined. For investigation of the RV variations between individual and nightly stacked spectra, we refer to *Jackson et al.* (2015).

However, the remaining setups (HR3, HR4, HR5A, HR5B, HR6, HR14A, HR14B, HR15N, and HR9B) were part of the WG13 Hot Star analysis for which a radial velocity was calculated using some combination of these setups (Blomme et al. 2022). This set of radial velocities was treated as a separate ‘setup’ for the purposes of the calibration procedure below.

Figure 27 shows the difference between the GRVS radial velocity and the radial velocity derived within Gaia-ESO for each GRVS for each relevant setup. We note that often multiple spectra were obtained for each GRVS for each setup and so the mean of the radial velocities per GRVS was calculated per setup. If there was only one measurement, the error taken was that associated with the value. If there were multiple measurements, the standard deviation of the measurements was taken as the error on the mean value. These are shown as error bars in Figure 27.

Comparing the relative values between setups of the offset mean and standard deviation, and the standard deviations per...
GRVS per setup, HR10 shows the most robust agreement with the GRVS values. It was therefore selected as the baseline setup to which radial velocities for the other setups would be calibrated. The goal was not to then calibrate to the GRVS but rather to report the homogenised Gaia-ESO radial velocities.

As an internal calibration set, the GRVS were limited in usefulness as they were only observed for 5 of the 12 setups used across the WGs. It was then necessary to construct a bootstrapping procedure to maximise the samples in common between setups in order to calibrate them to HR10. Each setup was investigated against all other setups to see which had the most stars in common and also what the ‘in-common’ set contained. In some cases, this was simply the Sun (e.g. HR5B and HR14B), and so the calibrations for some setups are not particularly robust.

In general, there were four possible bootstrap procedures:
1. Offset calculated directly with HR10;
2. Offset calculated directly with HR15N then bootstrapped to HR10;
3. Offset calculated directly with HR9B then bootstrapped to HR15N then bootstrapped to HR10;
4. Offset calculated directly with U580 then bootstrapped to HR15N then bootstrapped to HR10.

Table 3 provides details of the bootstrap procedure for each setup and the resulting offset applied to calibrate the radial velocities of each setup to HR10.

Offset sets were calculated between each of the setups and the zero point of the GES RV scale, HR10. The offsets were then applied to put the other setups onto the HR10 scale. For WG13, RVs based on a combination of WG13 setups were calculated for particular clusters (NGC 3293, NGC 6705, Trumpler 14, NGC 6530, NGC 2244, NGC 3766, and NGC 6649) and an offset was calculated with respect to HR10.

Having assessed the baseline SETUP as HR10, and after calculating offsets to put all RVs per SETUP onto the HR10 RV scale, the next stage was to assign an RV to each CNAME based on a set of rules. Figure 28 illustrates the rules used to select an RV per CNAME. The offsets listed in Table 3 were applied when an RV other than one from HR10 was selected.

The error is that associated with the selected value, except for the case where the value is the mean of the two values of the upper and lower arms of the UVES SETUP. In that case, the error is calculated as the sum in quadrature of the errors on the two arms. The homogenised radial velocity is reported as VRAD in the final database.

7.2. Rotational velocities
Rotational velocities were determined as part of the Arcetri UVES pipeline, the CASU GIRAFFE pipeline, by WG13, and by the OACT node. However, there was a recalibration of the GIRAFFE instrument by ESO after the internal DR4, which
changed the resolution such that the GIRAFFE radial velocity pipeline, which also derived rotational velocity (see Gilmore et al. 2022), could not consistently determine $v\sin i$ from the stacked spectra. Hence, after the internal DR4, no $v\sin i$ were reported for the GIRAFFE spectra.

Therefore the final GES catalogue reports $v\sin i$ values only from the Arcetri UVES pipeline, WG13 or OACT. The rules governing the assignment of $v\sin i$ are illustrated in Figure 29.

The error on $v\sin i$ is determined in the same way as the error on VRAD (see Section 7.1).

### 7.4. Other parameters

In the final catalogue, there are some other parameters that do not enter into the homogenisation process. The photometric temperature from the infrared flux method (TEFF_IRFM, see González Hernández & Bonifacio 2009) with its error are provided for more than 20,000 stars. This is derived as the weighted average of the IRFM values calculated on the 2MASS J, H, and Ks bands.

Some specific parameters for young stars are provided by WG12, such as veiling (VEIL), parameters describing the mass-accretion rate, such as the EW of the Hα line (EW_HA_ACC) and HA10, that is, 10% Hα, and parameters for the chromospheric activity obtained from the EW and flux of the Hα emission line, EW_HA_CHR and FHA_HA, and from the EW and flux of the Hβ line, EW_HB_CHR and FHB_CHR. In a few cases, the mass-accretion rate is also provided (LOG_MDOT_ACC). All quantities are provided with their associated uncertainties, and are described in Lanzafame et al. (2015).

The GAMMA index is supplied for more than 20,000 stars. It is an alternative parameter to be used when it is not possible to
Fig. 18: Comparison of stellar parameters between WG12 and WG11 for stars in common.

Table 3: Bootstrap procedure applied per setup (X) for calibration of radial velocities to HR10. Here, we provide the offset, standard deviation, and number of stars in common for the bootstrap to the initial setup (Y), and for the overall bootstrap (BS) to HR10.

<table>
<thead>
<tr>
<th>SETUP</th>
<th>N</th>
<th>SETUP</th>
<th>X-Y</th>
<th>N</th>
<th>X-BS-HR10</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR15N</td>
<td>3161</td>
<td>HR10</td>
<td>0.09</td>
<td>0.21</td>
<td>3161</td>
</tr>
<tr>
<td>HR21</td>
<td>48133</td>
<td>HR10</td>
<td>0.59</td>
<td>0.46</td>
<td>48133</td>
</tr>
<tr>
<td>HR9B</td>
<td>253</td>
<td>HR15N</td>
<td>0.08</td>
<td>1.30</td>
<td>558</td>
</tr>
<tr>
<td>U580</td>
<td>154</td>
<td>HR15N</td>
<td>0.21</td>
<td>0.53</td>
<td>695</td>
</tr>
<tr>
<td>U520</td>
<td>46</td>
<td>U580</td>
<td>-0.10</td>
<td>0.92</td>
<td>54</td>
</tr>
<tr>
<td>HR3</td>
<td>123</td>
<td>HR9B</td>
<td>-0.13</td>
<td>2.75</td>
<td>334</td>
</tr>
<tr>
<td>HR4</td>
<td>116</td>
<td>HR9B</td>
<td>-0.43</td>
<td>5.03</td>
<td>211</td>
</tr>
<tr>
<td>HR5A</td>
<td>131</td>
<td>HR9B</td>
<td>-0.10</td>
<td>3.98</td>
<td>151</td>
</tr>
<tr>
<td>HR5B*</td>
<td>1</td>
<td>HR10</td>
<td>0.15</td>
<td>0.22</td>
<td>1</td>
</tr>
<tr>
<td>HR6</td>
<td>118</td>
<td>HR9B</td>
<td>0.00</td>
<td>7.03</td>
<td>123</td>
</tr>
<tr>
<td>HR14A</td>
<td>145</td>
<td>HR9B</td>
<td>0.16</td>
<td>2.37</td>
<td>146</td>
</tr>
<tr>
<td>HR14B*</td>
<td>1</td>
<td>HR10</td>
<td>0.29</td>
<td>0.22</td>
<td>1</td>
</tr>
<tr>
<td>WG13Comb</td>
<td>506</td>
<td>HR10</td>
<td>0.24</td>
<td>3.94</td>
<td>507</td>
</tr>
</tbody>
</table>

* Only Sun in common

properly estimate the surface gravity in stars observed with the GIRAFFE setup HR15N. For a complete description of the gravity and temperature indices, including GAMMA, for HR15N we refer to Damiani et al. (2014). In addition to Li abundance, which is described in Sec. 5, the final catalogue provides the EWs of Li lines: the measured equivalent width EWLI, with its associated error and an indication of its upper limit or measurement, and the EW corrected for the contamination of the nearby Fe line, EWCLi, again with its error and upper limit indication. We refer to Franciosini et al. (2022) for a full description of the Li...
EW measurement. Finally, the membership probability for stars in the field of several open and globular clusters is provided in the MEM3D column. The membership analysis is described in Jackson et al. (2022).

8. Stellar parameter and abundance error distributions

For each measurement reported in the final GES catalogue, an associated uncertainty (e.g. in associated column ‘E_’) is also reported. Figure 30 shows the reported errors on $T_{\text{eff}}$, log $g$, [Fe/H], $\xi$, $V_{\text{rad}}$, and $v$ $\sin i$ as a function of S/N. The provenances of the errors are indicated by colour. For $T_{\text{eff}}$, log $g$, [Fe/H], and $\xi$, the provenance is one of WG10, WG11, WG12, or WG13, according to the rules of the homogenisation algorithm represented schematically in Fig. 8, and the provenance of the associated error aligns with that of the parameter. For $V_{\text{rad}}$ the provenance is from the radial velocity pipelines, Arcetri for UVES and CASU for GIRAFFE, respectively, or from further analysis by WG13 (Blomme et al. 2022). See Section 7.1 for the provenance selection details. For $v$ $\sin i$, the provenance is one of WG13, Arcetri, or OACT. See Section 7.2 for the provenance selection details.

Similarly, Fig. 31 shows the error for the element abundances reported by WG10, WG11, and WG13 (excluding L1, which is discussed in Section 7.4).

The error model for each parameter and abundance is defined per provenance source, and no homogenisation of the errors occurs for the final catalogue. However, each provenance source, whether its is at the WG level or from either RV pipeline, provides an internally consistent error model reflective of the analysis at that level. See the descriptions in the associated papers for more details (Worley et al. 2020; Gilmore et al. 2022; Blomme et al. 2022). In general, as shown in Fig. 31, the error models show decreasing error with increasing S/N as expected.

These error models are typically based on the measurements, not the errors provided by the node analysis. Therefore, another column was provided in order for WGs to report an uncertainty based on the reported node analysis uncertainties, namely ‘ENN_’. Figure 32 shows errors based on the node errors for the element abundances for which these values are available.

9. Flags

9.1. Flag homogenisation

A sophisticated system of flags (detailed flags, hereafter) has been designed within the Gaia-ESO survey by WG14 (see Van Eck et al. 2023, for details) to report and keep track of issues occurring during the analysis (TECH) and also to indicate physical peculiarities for a given target (PECULI).

The homogenisation is based on the dictionary of flags produced by WG14. We compared the flags produced by the different WGs and searched for possible conflicts. In the case of differences in the confidence level flag, we took the highest confidence flag. All the flags from the WG_Recommended files are included with any duplicates removed.

All the flags included in the final database are described in the document accompanying the public release in the ESO archive.

9.2. WG15 additional flags

Additional rules are added at the WG15 level depending on the REC_WG provenance previously assigned to each CNAME. This new set of rules is designed to add WG15 flags or in some cases to remove stellar parameters. The detailed flow chart is shown in Fig. 33.

‘Parameters’ here refers to the columns $T_{\text{eff}}$, LOGG, [Fe/H], $\xi$, $V_{\text{rad}}$, and $v$ $\sin i$ of the final database. We note that the flag suffixes do change between WGs, and so they are not all identical, even though they may appear to be at first sight. If multiple flags are activated, the WG15 flags are concatenated using ‘|’. The WG15 flags are then concatenated with the existing TECH column.

9.3. Simplified flags

The TECH flags cover a broad range of topics (S/N, data reduction, determination, and quality of stellar parameters/abundances). The syntax of the flags allows us to quickly identify the issue (prefix), trace the working group (WG ID) and node (node ID) from which it originates, and, in some cases, obtain extra information (suffix). However, this system is too detailed for the end users who want to quickly use the Gaia-ESO data.

For iDR6, a system of simplified flags has therefore been designed for the Gaia-ESO Survey. These simplified flags must allow the end users to quickly filter the data in order to meet their science goals; they should allow the user to quickly reject objects...
Fig. 20: Abundance ratios $[X/H]$ versus $[Fe/H]$ for all the elements analysed in GES (except Li). The results of WG10 are represented by blue circles, those of WG11 and WG12 are shown as black circles, and WG13 results are represented by magenta circles. $C_C2$ are carbon abundances from C2 molecular bands and $N_CN$ are nitrogen abundances from CN molecular bands.

with non-physical or highly suspicious results, and they complement the information already carried by the error bars associated with the observables. Converting any flag to the simplified scheme will cause a loss of valuable information, and therefore it is important to also release the detailed flags. The simplified flags consist of a small acronym (three letters) whose meaning is easily recoverable or indeed easily discernable without looking at the documentation. They are coded with Booleans (False/True), each in an individual column, allowing the end users to easily use them to sort the data. The acronym and meaning of each flag are listed in the table below. A comment is also provided to specify when the flag is raised and to briefly illustrate the conversion from the detailed scheme to the simplified scheme. The default value of the simplified flag is False; in other words, only the value True carries information.

All TECH flags (except some ‘neutral’ flags that are dropped during the conversion) have been translated into simplified flags (see following paragraph). On the other hand, only two simplified flags are defined to summarise the information carried by the most-used PECULI flags in order to quickly identify: (a) whether or not the object is suspected to be a spectroscopic multiple (BIN), or (b) whether or not emission lines are observed (EML).

Three simplified flags (SNR, SRP, SDS) deal with the intrinsic quality of the reduced spectra. The simplified flags pertaining to the stellar parameters (IPA, SSP, PSC) only deal with the effective temperature, the surface gravity, the metallicity, and the microturbulence. Two simplified flags (NIA, SSA) give a general indication of the availability of abundance determinations (for any element but iron) for a given star. There is a dedicated simplified flag for the radial velocity (SRV), and for the rotational velocity (SRO).

It is not possible to have a limited set of simplified flags and at the same time to have a detailed assessment of each stellar parameter (and respectively, abundance). This means that the end users have to perform further checks (e.g. based on the detailed flags) to decide which abundances can be kept when an object has the flag ‘some suspicious abundances’ raised. During the process of reducing the detailed flags to the simplified flags, a conservative approach was adopted, meaning that the problems
might be less severe than indicated by the simplified flags. For example, the SSP (some suspicious parameters) or IPA (incomplete parameter) flags are sometimes raised when some, though not all, analysis nodes provide uncertain parameters or abundances, despite the fact that other nodes might well have provided reliable results. Similarly, the flag SSA provides a general indication of the quality of the abundance ratios attached to a given star. Given that for example up to 20 chemical species are investigated in UVES observations, it is impossible for a unique simplified flag to provide an accurate and exhaustive description. Therefore, we advise that the flag SSA be used in a second step when outliers remain in the user’s selection to identify objects for which consultation of the detailed flags may be necessary. On the other hand, the simplified flags SNR, SRP, and NIA may be used a priori to clean the user’s sample. The list of simplified flags and comments can be found in Appendix A.1.

10. Discussion and conclusion

In this section, we present some validation plots of the final recommended set of stellar parameters. The catalogue with all final astrophysical quantities is publicly available in the ESO archive.

10.1. Final Kiel diagram

Figure 34 shows the Kiel diagram of the entire latest release of GES (>114000 unique CNAMEs). The diagram shows the variety of spectral types analysed, which is a unique aspect of GES compared to other surveys; ranging from cool PMS stars to hot early-type stars and red giant branch (RGB) stars, covering a metallicity range from \(-2.5\) to \(0.5\) dex, from globular clusters to inner-disc open clusters. Comparison with two representative sets of isochrones at solar metallicity and \([\text{Fe}/\text{H}]\) = \(-2\) dex indicates very good agreement, with a shift of RGB stars towards higher temperatures for the more metal-poor stars. Also notice-

---

Fig. 21: Abundance ratios \([\text{X}/\text{H}]\) versus \([\text{Fe}/\text{H}]\) for all the elements analysed in GES (except Li, which is presented in detail in Franciosini et al. (2022)). The results of WG10 are represented as blue circles, those of WG11 and WG12 are shown as black circles, and WG13 results are represented as magenta circles.
10.2. The [Mg/Fe] versus [Fe/H] diagram

Figure 35 displays the density plot of [Mg/Fe] as a function of [Fe/H] for the MW field populations. This diagram is usually used to separate the thin-disc population from the thick-disc population based on the findings of Wallerstein (1962). The combined sample, including both UVES WG11 and GIRAFFE WG10 results, indicates a gap at [Mg/Fe] ~ 0.2 and [Fe/H] ~ −0.4, in a location similar to the one discussed in Recio-Blanco et al. (2014) with data from a previous GES data release, and also similar to that discussed by Hayden et al. (2015) with APOGEE data and that discussed by Buder et al. (2021) with GALAH data. In the illustrative Fig. 35, we plot the entire GES MW sample, making a cut only in S/N. The exact location of the transition between thin and thick disc is a function of $R_{GC}$ and also of $z$, the height above the plane, and therefore it might vary as a function of the selected sample.
10.3. Comparison with GALAH and APOGEE surveys

Figure 36 shows comparisons of GES data to both GALAH data and APOGEE data. We selected the GES stars that are in common with each survey. The plots show typical comparisons for the $\alpha$-elements, Mg and Ca, an iron-peak element, Cr, and a neutron-capture element, Ce. It may also be seen that the dispersion found in the GES results is generally smaller than that found in the results of the other surveys; the results for the element Cr provide a good example of this. Table 4 presents the abundance [X/Fe] median differences and the associated dispersion between GALAH, APOGEE, and GES WG15 for the sample of stars in common. In most cases, the median difference is below 0.1 dex, demonstrating the excellent agreement between the surveys.

10.4. Comparison with Gaia radial velocities and calibrated metallicities

Figure 37 shows the radial velocity difference (GES - Gaia DR3) as a function of the Gaia radial velocity. The plot has been made for two GES subsamples in two different instrumental configurations (HR10 and HR21 GIRAFFE spectra) and U580 UVES spectra). The median difference is close to zero and the dispersion is respectively 2.74 and 3.52 for the two setups. The agreement between the GES and the Gaia radial velocities is excellent for both setups.

Figure 38 shows the plot of the difference in metallicity (GES - Gaia DR3) as a function of GES [Fe/H]. For Gaia, we used the calibrated spectroscopic metallicities as described in Recio-Blanco et al. (2022). As for the radial velocity, we separated the sample observed with U580 and the sample observed with the GIRAFFE setups. The latter contains observations with HR10, HR15N, HR10:HR21, HR21, and HR9B. The median differences for both samples are close to zero, with a dispersion of 0.16 dex, indicating, on average, very good agreement in terms of the accuracy of Gaia compared to GES, and, as expected, a lower precision for Gaia compared to GES.
Fig. 24: Open and globular clusters. Each panel represents the results for a different element. [Fe/H] is in abscissa. The ordinates are the values ‘median WG11 abundances minus median WG10 abundances’. The median global difference is shown as an orange line. Dotted orange lines indicate ±1σ standard deviation from the median value.

Table 4: Median abundance [X/Fe] difference between GALAH, APOGEE, and WG15 for the sample of stars in common. For the majority of the elements, the median difference is below 0.1 dex.

<table>
<thead>
<tr>
<th>Element</th>
<th>GALAH − WG15</th>
<th>APOGEE − WG15</th>
</tr>
</thead>
<tbody>
<tr>
<td>C I</td>
<td>0.16</td>
<td>0.35</td>
</tr>
<tr>
<td>O I</td>
<td>−0.05</td>
<td>0.36</td>
</tr>
<tr>
<td>Na I</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Mg I</td>
<td>−0.01</td>
<td>0.18</td>
</tr>
<tr>
<td>Al I</td>
<td>0.04</td>
<td>0.22</td>
</tr>
<tr>
<td>Si I</td>
<td>0.04</td>
<td>0.24</td>
</tr>
<tr>
<td>S I</td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td>Ca I</td>
<td>0.11</td>
<td>0.23</td>
</tr>
<tr>
<td>Sc I</td>
<td>0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>Ti I</td>
<td>0.03</td>
<td>0.23</td>
</tr>
<tr>
<td>Ti II</td>
<td>−0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>V I</td>
<td>0.13</td>
<td>0.35</td>
</tr>
<tr>
<td>Cr I</td>
<td>0.02</td>
<td>0.24</td>
</tr>
<tr>
<td>Mn I</td>
<td>0.00</td>
<td>0.32</td>
</tr>
<tr>
<td>Co I</td>
<td>0.06</td>
<td>0.42</td>
</tr>
<tr>
<td>Ni I</td>
<td>0.04</td>
<td>0.22</td>
</tr>
<tr>
<td>Cu I</td>
<td>0.09</td>
<td>0.23</td>
</tr>
<tr>
<td>Zn I</td>
<td>0.04</td>
<td>0.29</td>
</tr>
<tr>
<td>Y II</td>
<td>−0.03</td>
<td>0.39</td>
</tr>
<tr>
<td>Zr I</td>
<td>0.12</td>
<td>0.52</td>
</tr>
<tr>
<td>Mo I</td>
<td>0.15</td>
<td>0.37</td>
</tr>
<tr>
<td>Ba II</td>
<td>0.10</td>
<td>0.36</td>
</tr>
<tr>
<td>La II</td>
<td>0.06</td>
<td>0.34</td>
</tr>
<tr>
<td>Ce II</td>
<td>0.07</td>
<td>0.32</td>
</tr>
<tr>
<td>Nd II</td>
<td>0.17</td>
<td>0.39</td>
</tr>
<tr>
<td>Sm II</td>
<td>−0.05</td>
<td>0.44</td>
</tr>
<tr>
<td>Eu II</td>
<td>−0.05</td>
<td>0.22</td>
</tr>
</tbody>
</table>

10.5. Conclusion

By design, the Gaia-ESO survey is based on a heterogeneous set of data, namely medium-resolution spectra with different wavelength ranges (GIRAFFE) and high-resolution spectra (UVES). One of its strengths is that these spectra were acquired with high-efficiency multiplex spectrographs attached to an 8m class telescope. In contrast to other large surveys, the sample is not limited to FGK stars; it includes cool PMS and hot stars (OBA). The originality of the GES survey is that it does not rely on a single pipeline for the analysis of the spectra. The multi-method, multi-pipeline design of the GES analysis, implemented through the analysis node and WG structure, means that multiple results are delivered for many GES stars. This includes both parallel analyses of the same stellar samples within a WG, and the analysis of common calibration samples across WGs. The homogenisation process is based on a set of benchmark stars and calibration open and globular clusters able to cover the vast range of stellar parameters of the whole sample of stars of this survey. To provide a final consistent set of results, a dedicated working group (WG15) was set up within GES to provide the recommended results from WGs 10 to 13 and to transpose these onto a common scale. With this set of homogenised stellar parameters, detailed abundances were computed by the nodes and merged for each WG. The resulting abundances were analysed by WG15 to set them, element by element, onto a common scale. This paper describes the numerous steps followed by the WG15 that led to the final homogenised set of stellar parameters, abundances, and ve-
Fig. 25: $\alpha$-element abundance over iron, computed by averaging [Mg/Fe], [Si/Fe], [Ca/Fe], and [Ti/Fe] as a function of [Fe/H] in member stars of four calibration open clusters. The abundances from WG11 are shown in black and those from WG10 in blue. The red circles mark the median WG11 abundances within 1σ, 2σ, and 3σ. The green circles indicate the same quantities but computed with WG10 results.

Fig. 26: [$\alpha$/Fe] as a function of [Fe/H] in member stars of four calibration globular clusters. Symbols and colours are as in Fig. 25.

...locations for more than 110000 stars. The numerous figures of this article give a good overview of the quality of the homogenisation process and of the final abundance results compared to literature data and large spectroscopic survey results.

Many of the surveys currently underway and in preparation have been inspired by the Gaia-ESO structure and approach. For example, the FITS format for data exchange is currently adopted by most Galactic spectroscopic surveys, which was not previously common in this field; the Gaia-ESO line list is widely used, and the use of the various categories of calibrators has been widely adopted, particularly that of star clusters. The experiment with multiple pipelines was helpful in understanding the inherent limitations of spectral analysis and in finding the origin of systematic errors in the various approaches. This study has been very useful and the community has learnt a lot from it; at the same time, it was necessarily very time consuming, and has not been repeated. This is a very unique test, and will inform many choices for subsequent surveys.

Acknowledgements. Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programmes 188.B-3002, 193.B-0936, and 197.B-1074. These data products have been processed by the Cambridge Astronomy Survey Unit (CASU) at the Institute of Astronomy, University of Cambridge, and by the FLAMES/UVES reduction team at INAF/Osservatorio Astrofisico di Arcetri. These data have been obtained from the Gaia-ESO Survey Data Archive, prepared and hosted by the Wide Field Astronomy Unit, Institute for Astronomy, University of Edinburgh, which is funded by the UK Science and Technology Facilities Council. This work was partly supported by the European Union FP7 programme through ERC grant number 320360 and by the Leverhulme Trust through grant RPG-2012-541. 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”. 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 Programme. L. Magrini and M. Van der Swaalmak consortium, and by the INAF Grant “Cliche”. A.J.Korn acknowledges support by the Swedish National Space Agency (SNSA). A.Lobel acknowledges support in part by the Belgian Federal Science Policy Office under contract No. BR/143/A2/BRASS and by the European Union Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Sklodowska-Curie grant Agreement No. 823754. D.K. Feuillet was partly supported by grant No. 2016-03412 from the Swedish Research Council. D. Montes acknowledges financial support from the Agencia Estatal de Investigacion del Ministerio de Ciencia, Innovacion through project PID2019-109522GB-C54 /AEI/10.13039/501100011033 E.Marfil acknowledges financial support from the European Regional Development Fund (ERDF) and the Governo di Cari- narias through project Prod2021/010128. J.I. Gonzalez Hernandez acknowledges financial support from the Spanish Ministry of Science and Innovation (MICINN) project PID2020-117493GB-I00. M. Bergemann is supported through the Lise Meitner grant from the Max Planck Society and acknowledges support by the Collaborative Research centre SFB 881 (projects A5, A10), Heidelberg University, of the Deutsche Forschungsgemeinschaft (DFG, German Re- search Foundation). This project has received funding from the European Re- search Council (ERC) under the European Union, Horizon 2020 research and innovation programme (Grant agreement No. 949173). P. Jordi acknowledges fi- nancial support of PON07CYT Regolar 1200703 as well as Nucleo Mileno ERIS NCN2021_017. R. Smiljanic acknowledges support from the National Sci- ence Centre, Poland (2014/15/B/ST/03981), S. R. Berlanas acknowledges support by MCIN/AEI/10.13039/501100011033 (contract FJC 2020-049758-I) and NextGeneration EU/PRTR and MIU (UNI/551/2021) through grant Margarita Salas-ULL. T. Bensby acknowledges financial support by grant No. 2018-04857 from the Swedish Research Council. T. Merle is supported by a grant from the Foundation ULB. T. Morel is grateful to Belgian F.R.S.-FNRS for support, and are also indebted for an ESA/PRODEX Belspo contract related to the Gaia Data Analysis and Analysis Consortium and for support through an ARC grant for Concerted Research Actions financed by the Federation Wallonie-Brussels. W. Santos acknowledges FAPERJ for a Ph.D. fellowship. H.M. Tabernero acknowl- edges financial support from the Agencia Estatal de Investigacion of the Ministerio de Ciencia, Innovacion through project PID2019-109522GB-C51 /AEI/10.13039/501100011033.
Fig. 27: Difference between GRVS radial velocity and the mean Gaia-ESO radial velocity calculated per GRVS per setup. The mean difference and standard deviation per setup are given.

References

Cirasuolo, M. et al. 2020, The Messenger, 180, 10
Jin, S. et al. 2022, in prep
Van Eck, S. et al. 2023, in prep
Worley, C. et al. 2023, in prep

1 Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom e-mail: aph@ast.cam.ac.uk
2 GEPI, Observatoire de Paris, PSL Research University, CNRS, Univ. Paris Diderot, Sorbonne Paris Cité, 61 avenue de l’Observatoire, 75014, Paris, France
3 School of Physical and Chemical Sciences- Te Kura Matū, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
Fig. 28: Schematic view of the GES WG15 radial velocity homogenisation workflow.

Fig. 29: Schematic view of the GES WG15 rotational velocity homogenisation workflow.
Fig. 30: Error on $T_{\text{eff}}$, $\log g$, [Fe/H], $\xi$, $V_{\text{rad}}$ and $v \sin i$ against S/N. The provenances of the errors are indicated by colour (blue – WG10, black – WG11, orange – WG12, magenta – WG13, green – Arcetri node, red – CASU, yellow – OACT node).

35 Departamento de Física Aplicada, Facultad de Ciencias, Universidad de Alicante, 03690 San Vicente del Raspeig, Alicante, Spain
Fig. 31: WG level error on each element abundance against SNR. The provenance of the errors are indicated by colour (blue – WG10, black – WG11, magenta – WG13).
Fig. 32: Errors based on the node analysis errors on abundances as a function of S/N. The provenance of the errors are indicated by colour (blue – WG10, magenta – WG13).
Fig. 33: WG15 additional flags flow chart.
Fig. 34: Kiel diagram using the final recommended values for iDR6 (all stars). Two sets of isochrones at solar metallicity and $0.01 < \text{Age}[\text{Gyr}] < 13$ (in red) and at $[\text{Fe/H}] = -2$ and $1 < \text{Age}[\text{Gyr}] < 13$ (in blue) are shown.
Fig. 35: [Mg/Fe] vs. [Fe/H] diagram for the MW field stars plotted using the sDR6 final recommended values (selecting GES_TYPE=GE_MW stars and S/N>50).
Fig. 36: Comparison between WG15, GALAH, and APOGEE abundance ratios for the sample of stars in common. WG15 data are represented as blue circles. Red symbols represent APOGEE data and black symbols are the data from the GALAH survey.
Fig. 37: Radial velocity difference between GES and *Gaia* DR3 for the stars in common. The GES samples HR10|HR21 and U580 are plotted in different colours.

Fig. 38: [Fe/H] difference between GES and *Gaia* DR3 for the stars in common. The GES samples GIRAFFE (all setups) and U580 are plotted in different colours. The horizontal red lines mark the median difference U580 - *Gaia* DR3 calibrated metallicity (continuous line) and ±1σ (dot-dashed lines). The green lines indicate the median difference GIRAFFE - *Gaia* DR3 calibrated metallicity (continuous line) and ±1σ (dot-dashed lines).
### Table A.1: Simplified flags

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Meaning</th>
<th>Conditions for raising the flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR</td>
<td>No or inaccurate results due to low SNR</td>
<td>This flag is raised if the S/N is lower than 50 and if the object has an incomplete set of parameters.</td>
</tr>
<tr>
<td>SRP</td>
<td>Spectral reduction problem</td>
<td>This flag is raised if there are no parameters or abundances</td>
</tr>
<tr>
<td>SDS</td>
<td>Some discarded spectra</td>
<td>This flag is raised if there are some parameters and abundances despite a reduced amount of usable data. For example, it is raised when spectral reduction problems affect some settings, preventing the entirety of the results from being obtained, but allowing some parameters and abundances to nevertheless be determined.</td>
</tr>
<tr>
<td>IPA</td>
<td>Incomplete parameters</td>
<td>This flag is typically raised when a key setup for a given parameter is missing, or when the node experienced an issue when converging to a consistent set of parameters, or, alternatively, when the parameters were out of the parameter grid of model atmospheres used by a specific node.</td>
</tr>
<tr>
<td>SSP</td>
<td>Some suspicious parameters</td>
<td>This flag is raised when some parameters, but not all, could be determined. This can occur when renormalisation failed, when the code did not converge to a consistent set of parameters, or, again, because the parameters fell out of the node’s grid. It also occurs when a parameter was derived outside the group of validated nodes for this parameter. It is also raised in cases of spectroscopic multiplicity with at least two visible components (SBn, n \geq 2).</td>
</tr>
<tr>
<td>NIA</td>
<td>No individual abundance (except Fe)</td>
<td>This flag is usually raised when there are too few available lines for abundance determinations (except Fe).</td>
</tr>
<tr>
<td>SSA</td>
<td>Some suspicious abundances</td>
<td>This flag can be raised for metallicity, e.g. when the Fe I and Fe II lines are discrepant, or for other elements. It is raised in cases of high v sini values or of SBn, n \geq 2, or when the node was uncertain about this abundance.</td>
</tr>
<tr>
<td>PSC</td>
<td>Parameter space coverage</td>
<td>This flag is typically raised when the parameters are not within the model atmosphere grid parameters of the node, or are on the node’s grid edge. Some abundances might then be missing.</td>
</tr>
<tr>
<td>SRV</td>
<td>Suspicious or unreliable radial velocity</td>
<td>This flag is raised in cases where the CCF was corrupted, or where the RV was discrepant between setups, or the object was identified as an SBn.</td>
</tr>
<tr>
<td>SRO</td>
<td>Suspicious rotational velocity</td>
<td>This flag is raised in cases where there is no rotational velocity determination, or an excessively high or revised rotational velocity. It is also raised in case of SBn, n \geq 2.</td>
</tr>
<tr>
<td>BIN</td>
<td>Detected binary: SB1 or SBn \geq 2.</td>
<td></td>
</tr>
<tr>
<td>EML</td>
<td>Emission line: any line, not only \text{H}\alpha</td>
<td></td>
</tr>
</tbody>
</table>

**Appendix A: Simplified flags**