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

A. Hourihane, P. François, C. Worley, L. Magrini, A. Gonneau, A. Casey, G. Gilmore, S. Randich, G. Sacco, A. Recio-Blanco, et al.

Astronomy & Astrophysics

The *Gaia*-ESO Survey: Homogenisation of stellar parameters and elemental abundances[★]

A. Hourihane[1], P. François[2], C. C. Worley[1,3], L. Magrini[4], A. Gonneau[1], A. R. Casey[5], G. Gilmore[1],
S. Randich[4], G. G. Sacco[4], A. Recio-Blanco[6], A. J. Korn[7], C. Allende Prieto[8,9], R. Smiljanic[10], R. Blomme[11],
A. Bragaglia[12], N. A. Walton[1], S. Van Eck[13], T. Bensby[14], A. Lanzafame[15], A. Frasca[15], E. Franciosini[4], F. Damiani[16],
K. Lind[17], M. Bergemann[18,19], P. Bonifacio[20], V. Hill[6], A. Lobel[11], D. Montes[21], D. K. Feuillet[22],
G. Tautvaišienė[23], G. Guiglion[18], H. M. Tabernero[24], J. I. González Hernández[8], M. Gebran[25], M. Van der Swaelmen[4],
Š. Mikolaitis[26], S. Daflon[27], T. Merle[13], T. Morel[28], J. R. Lewis[1,†], E. A. González Solares[1], D. N. A. Murphy[1],
R. D. Jeffries[29], R. J. Jackson[29], S. Feltzing[22], T. Prusti[30], G. Carraro[31], K. Biazzo[32], L. Prisinzano[16], P. Jofré[33],
S. Zaggia[34], A. Drazdauskas[26], E. Stonkuté[26], E. Marfil[24], F. Jiménez-Esteban[24], L. Mahy[11],
M. L. Gutiérrez Albarrán[21], S. R. Berlanas[35,29], W. Santos[27], L. Morbidelli[4], L. Spina[34], and R. Minkevičiūtė[26]

(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 ($T_{\rm eff} \sim 30\,000$–$52\,000$ K) all the way to K-M ($\gtrsim 3500$ 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 – Galaxy: stellar content – methods: data analysis – methods: statistical – techniques: spectroscopic

1. Introduction

The launch of the European Space Agency's astrometric *Gaia* mission in 2013 (e.g. Perryman et al. 2001; Gaia Collaboration 2016, 2018, 2021) prompted a new wave of Galactic studies. *Gaia* is delivering precise distances, kinematics, photometry, and spectrophotometry for more than 1.5 billion stars, as well as radial velocities and chemical abundances for the brighter stars in the sample (see further Gaia Collaboration 2021; Recio-Blanco et al. 2023). 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. in prep.; 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.

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 strategy with other Southern and Northern surveys. This requires both the analysis of a common set of 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
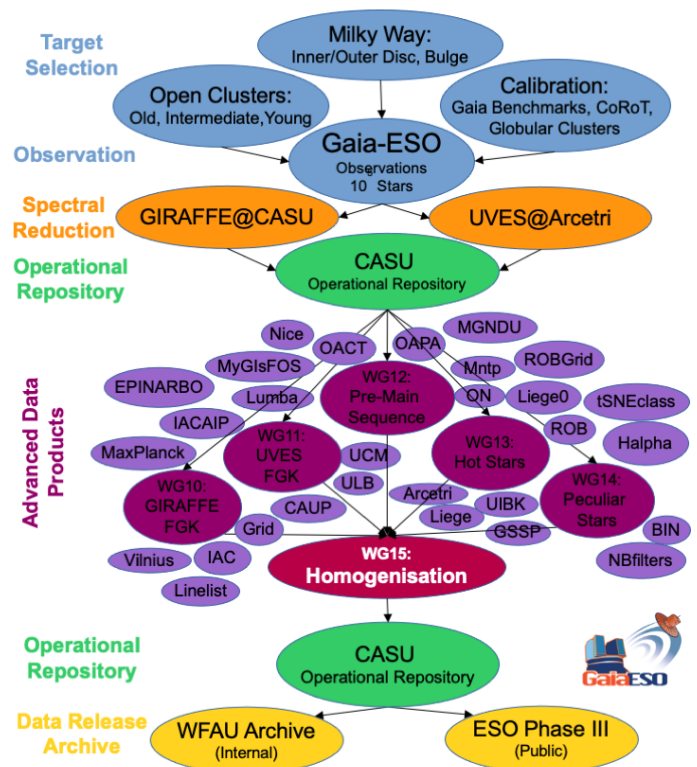
Fig. 1. *Gaia*-ESO data-flow diagram for iDR6 showing the key stages from target selection to data release via the archives.

distribution of the tasks among the different Working Groups. In Sect. 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 Sect. 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 Sect. 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[1] (~ 140 fibres) and UVES (Dekker et al. 2000; 8 fibres for U580[2],

---

[1] https://www.eso.org/sci/facilities/paranal/instruments/flames/inst/Giraffe.html
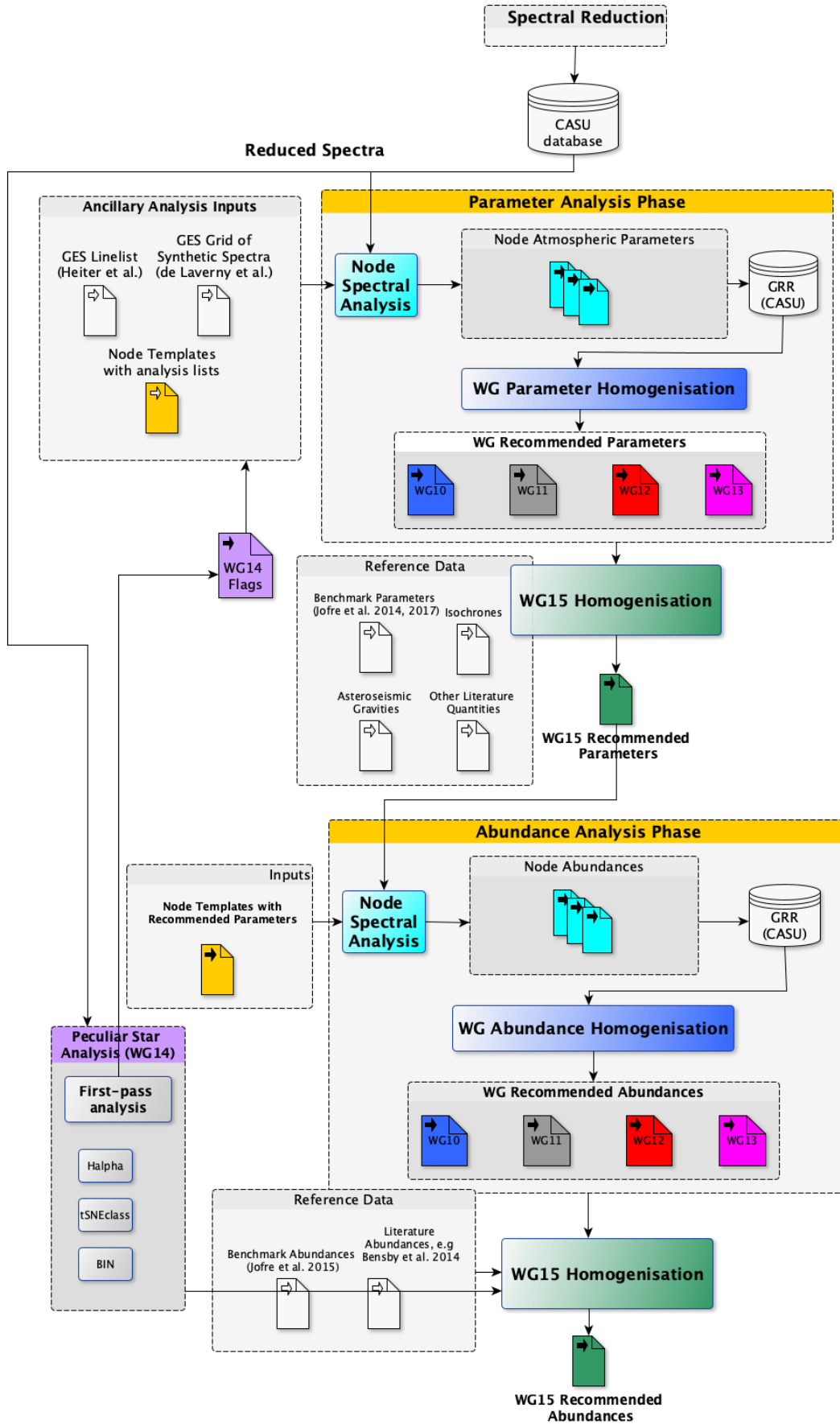
[2] One fibre was broken.

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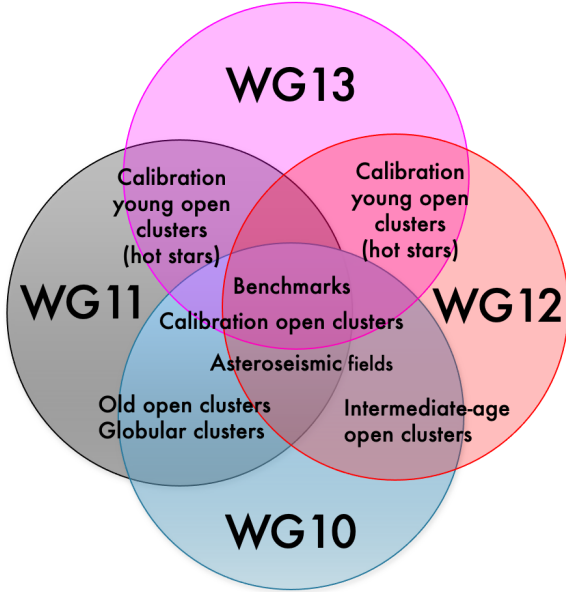
Fig. 3. Venn diagram describing the role of calibration samples in the interconnection between the WGs.

Fig. 5. Kiel diagram of the benchmark stars observed in GES. Stellar parameters are from the final data release. The symbols are colour-coded according to [Fe/H]. PARSEC isochrones at solar metallicity with ages from 0.2 to 13 Gyr are shown with grey curves. Warm benchmarks are marked with squares, while cool ones are marked with stars.
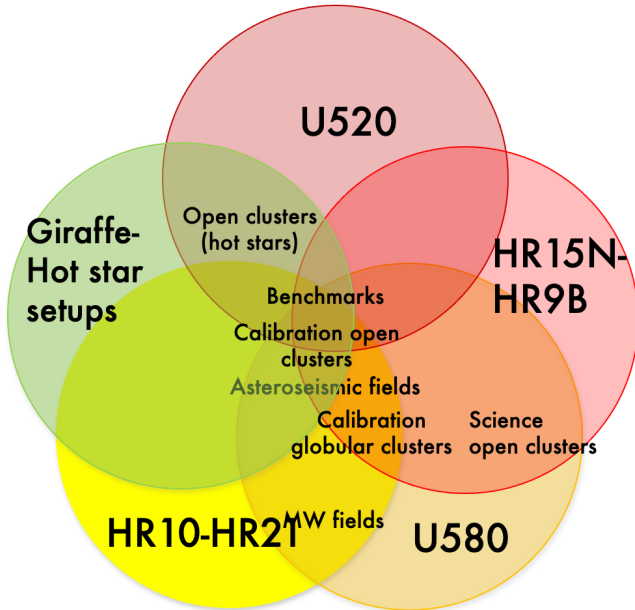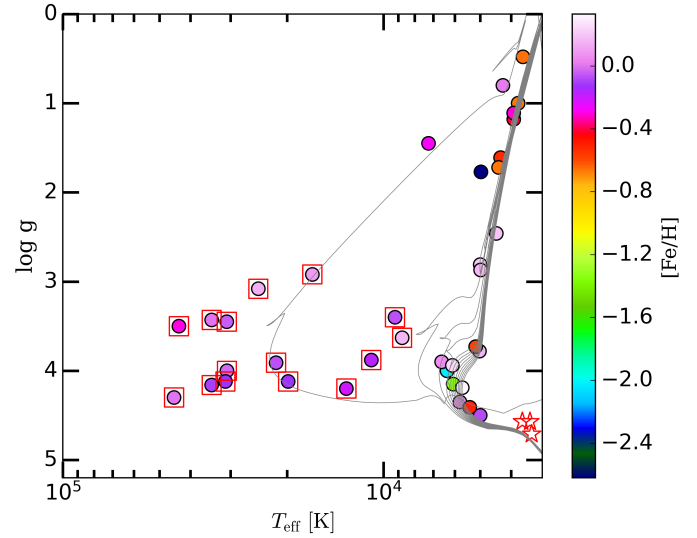

Fig. 4. Venn diagram describing the setups with which the calibration samples were observed (see also Table 1).

6 for U520) 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 phase of determination of abundances using the homogenised stellar parameters as input, and the definition of the final set of abundances passing through the WG homogenisation and 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). 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

Table 1. Instrumental set-ups used by the analysis working groups within GES.

Instrument	Setup	λ_{min} (Å)	λ_{max} (Å)	R_{old}	R	WG
UVES	520	4140	6210	47 000	47 000	11,12,13,14
UVES	580	4760	6840	47 000	47 000	11,12,13,14
GIRAFFE	HR3	4033	4201	24 800	31 400	13,14
GIRAFFE	HR4	4188	4392	32 500	35 550	13,14
GIRAFFE	HR5A	4340	4587	18 470	20 250	13,14
GIRAFFE	HR6	4538	4759	20 350	24 300	13,14
GIRAFFE	HR9B	5143	5356	25 900	31 750	10,13,14
GIRAFFE	HR10	5339	5619	19 800	21 500	10,12,14
GIRAFFE	HR14A	6308	6701	17 740	18 000	13,14
GIRAFFE	HR15N	6470	6790	17 000	19 200	10,12,14
GIRAFFE	HR21	8484	9001	16 200	18 000	10,12,14

Notes. The GIRAFFE instrument was refocused in February 2015. The R_{old} values refer to pre-refocusing values and R gives the new values, post-refocusing.

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 analysed 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 scientific analyses.

The characteristics of each setup in UVES[3] and GIRAFFE[4] 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).

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

Name	[Fe/H]	Type	Reference
M 15	−2.37	GC	Harris (1996)
NGC 4590	−2.23	GC	Harris (1996)
NGC 4372	−2.17	GC	Harris (1996)
NGC 4833	−1.92	GC	Pancino et al. (2017b)
M 2	−1.47	GC	Pancino et al. (2017b)
NGC 1904	−1.51	GC	Pancino et al. (2017b)
NGC 6752	−1.48	GC	Pancino et al. (2017b)
M 12	−1.37	GC	Harris (1996)
NGC 1261	−1.27	GC	Harris (1996)
NGC 362	−1.12	GC	Pancino et al. (2017b)
NGC 1851	−1.07	GC	Pancino et al. (2017b)
NGC 2808	−1.03	GC	Pancino et al. (2017b)
NGC 104	−0.71	GC	Pancino et al. (2017b)
NGC 5927	−0.39	GC	Pancino et al. (2017b)
NGC 6553	−0.25	GC	Harris (1996)
NGC 2243	−0.48	OC	Pancino et al. (2017a)
Berkeley 32	−0.21	OC	Friel et al. (2010)
Melotte 71	−0.27	OC	Pancino et al. (2017a)
NGC 2420	−0.05	OC	Pancino et al. (2017a)
M 67	+0.03	OC	Randich et al. (2006)
NGC 3532	+0.00	OC	Pancino et al. (2017a)
NGC 2477	+0.07	OC	Pancino et al. (2017a)
NGC 6705	+0.12	OC	Pancino et al. (2017a)
NGC 6253	+0.34	OC	Pancino et al. (2017a)

GES makes use of the set of calibrators selected by the Calibration and Standards Working Group (WG5), which is responsible for the observational calibration strategy for GES (Pancino et al. 2017a). 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,
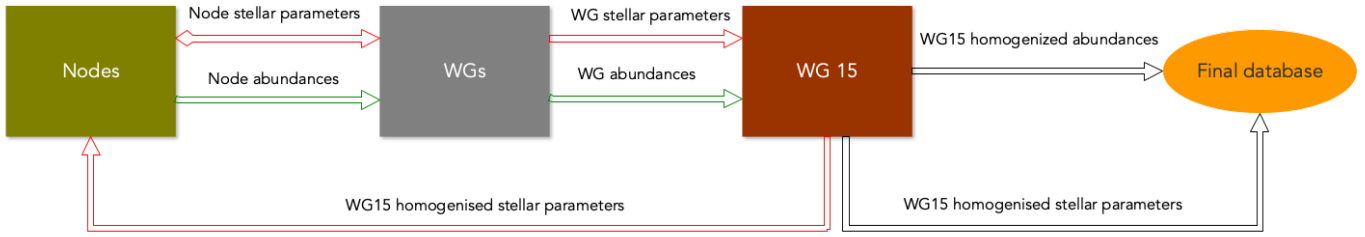
---

[3] https://www.eso.org/sci/facilities/paranal/instruments/uves/doc/ESO_411892_User_Manual_P109.pdf
[4] https://www.eso.org/sci/facilities/paranal/instruments/flames/inst/specs1.html

Fig. 6. Schematic view of the GES analysis approach. Arrows in red mark the first cycle of analysis dedicated to the determination of the stellar parameters; the arrows in green indicate the analysis of abundances (with homogenised parameters), while in black we show the process of homogenisation and preparation of the final catalogue.
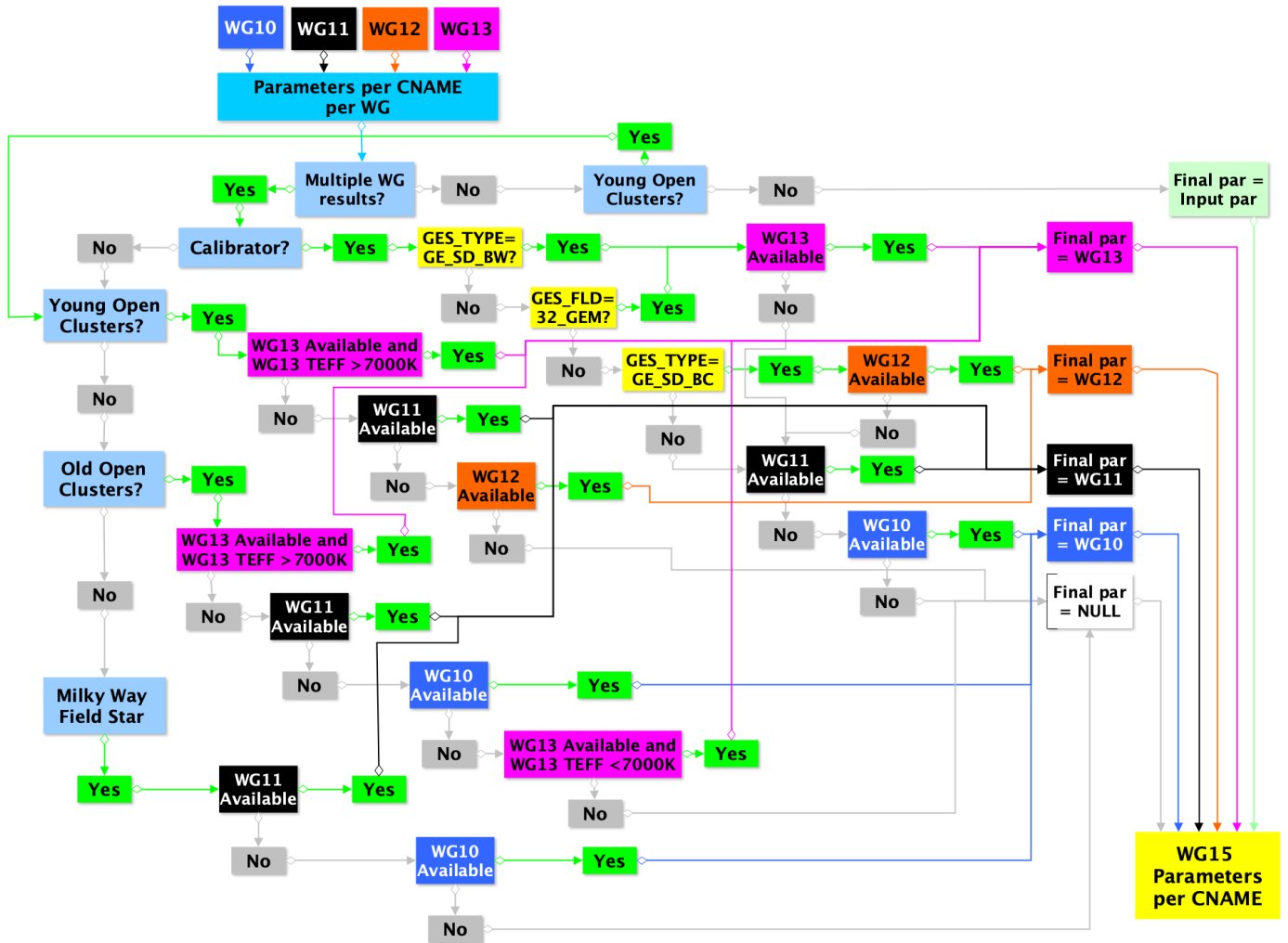

Fig. 7. Schematic view of the GES WG15 parameter homogenisation workflow.

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) 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). Then, surface gravity is determined using temperature, bolometric flux, parallax, and mass from a stellar track . Finally, $T_{\rm eff}$ and $\log g$ are fixed to the values
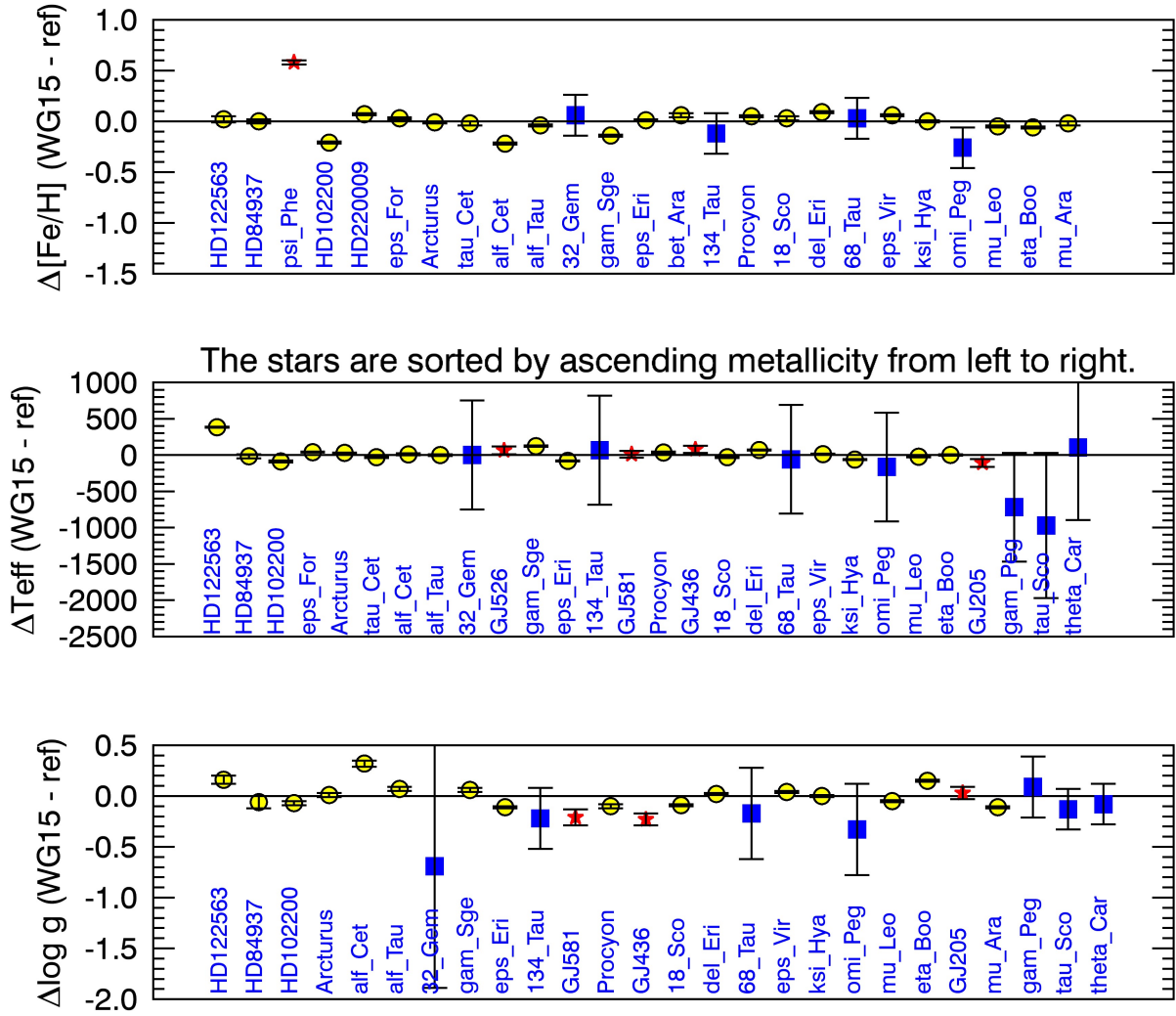
Fig. 8. WG15 parameters versus reference parameters as described in Pancino et al. (2017a). Warm benchmarks (OBA stars) are marked with blue squares, while the cool benchmarks (M stars) are marked with red stars. The stars are presented with ascending [Fe/H] along the abscissa.
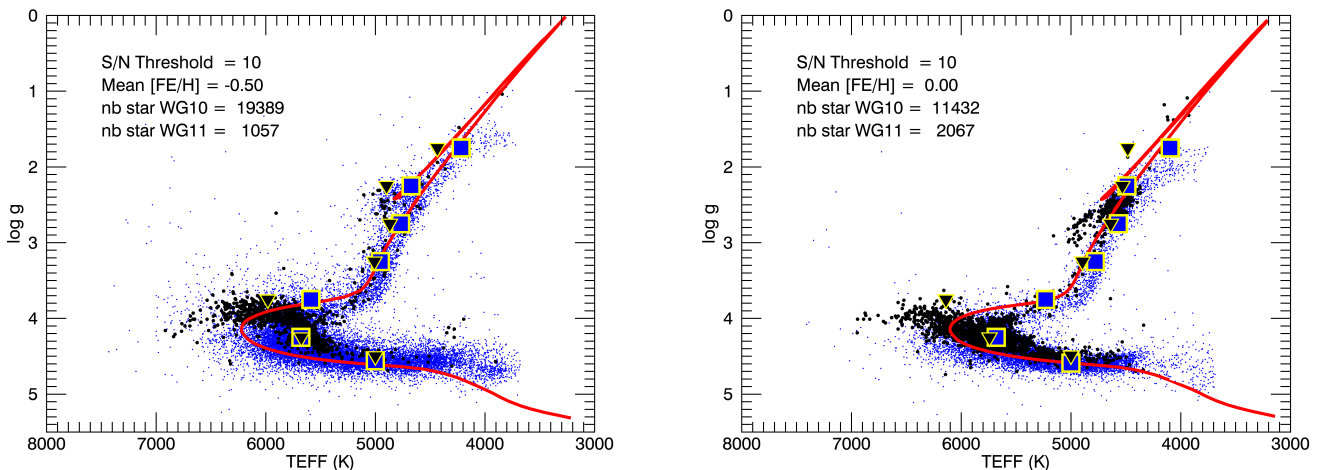

Fig. 9. GES iDR6 Kiel diagrams for MW stars for WG10 (the blue points) and WG11 (the black points) for two metallicity bins [−0.75 : −0.25], [−0.25 : +0.25] and S/N > 10. Isochrones are from PARSEC (Bressan et al. 2012) for $z = 0.01$, $z = 0.02$, and age = 5.7 Gyr. The black triangles and blue rectangles are median values for the WG10 and WG11 stars, respectively, in log g bins of 1 dex for the giant branch and in $T_{\rm eff}$ bins of 500 K for the main sequence stars.
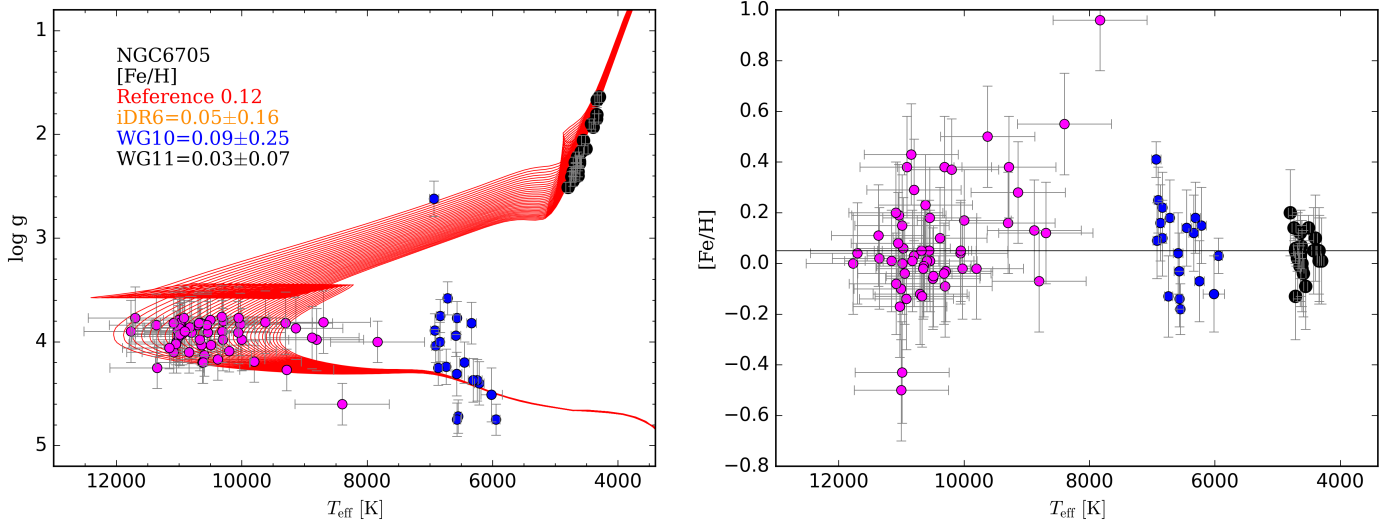
Fig. 10. Stellar parameters of the members of NGC 6705. Left: Kiel diagram showing the PARSEC isochrones in the age and metallicity range of NGC 6705 (solar metallicity, age from 200 to 500 Myr, steps of 50 Myr) and the data obtained by WG10 (blue), WG11 (black), and WG13 (magenta). Right: [Fe/H] vs. $T_{\rm eff}$ for the same stars. The black horizontal line is the mean metallicity for NGC 6705 in Randich et al. (2022). In the plot, we include only stars with E_LOGG < 0.35.
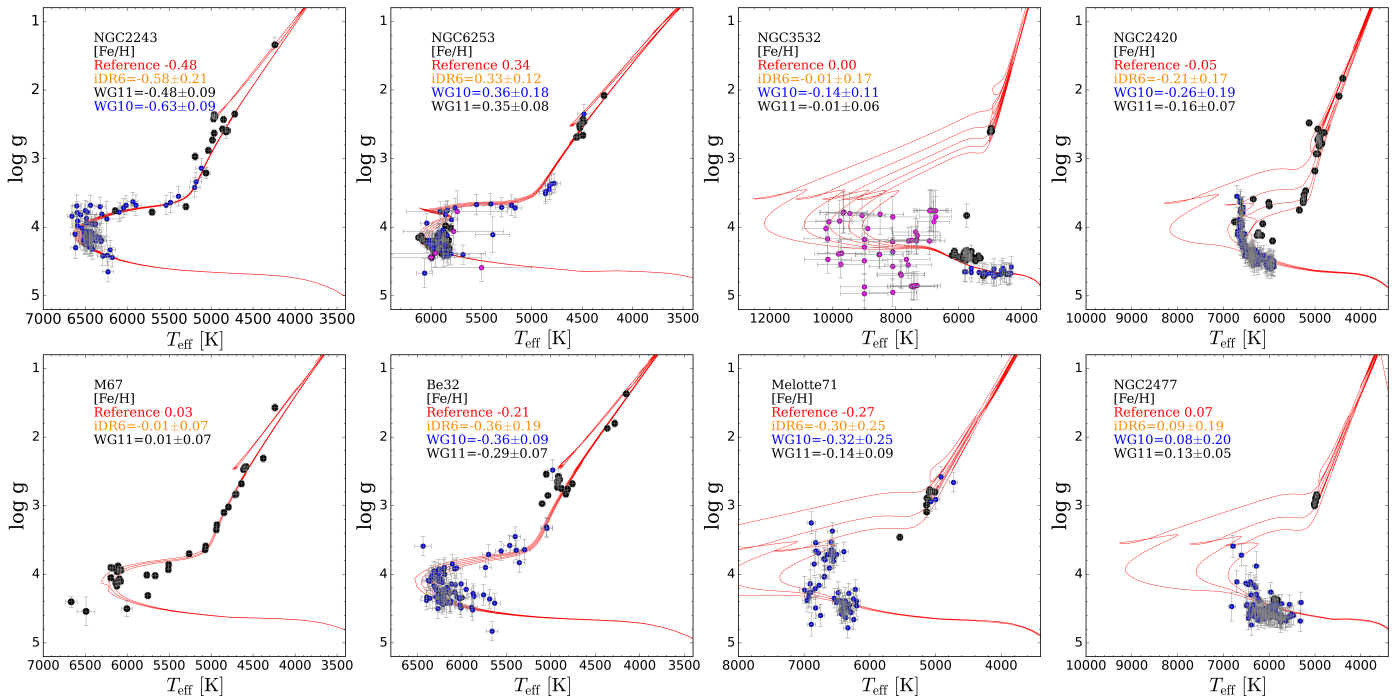
Calibration clusters

Fig. 11. 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_LOGG < 0.35.

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) by various spectral analysis methods (Jofré et al. 2014, 2015). This leads to a sample of 34 stars with accurate temperatures – with a precision of about 100 K –, surface gravities, and abundances 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
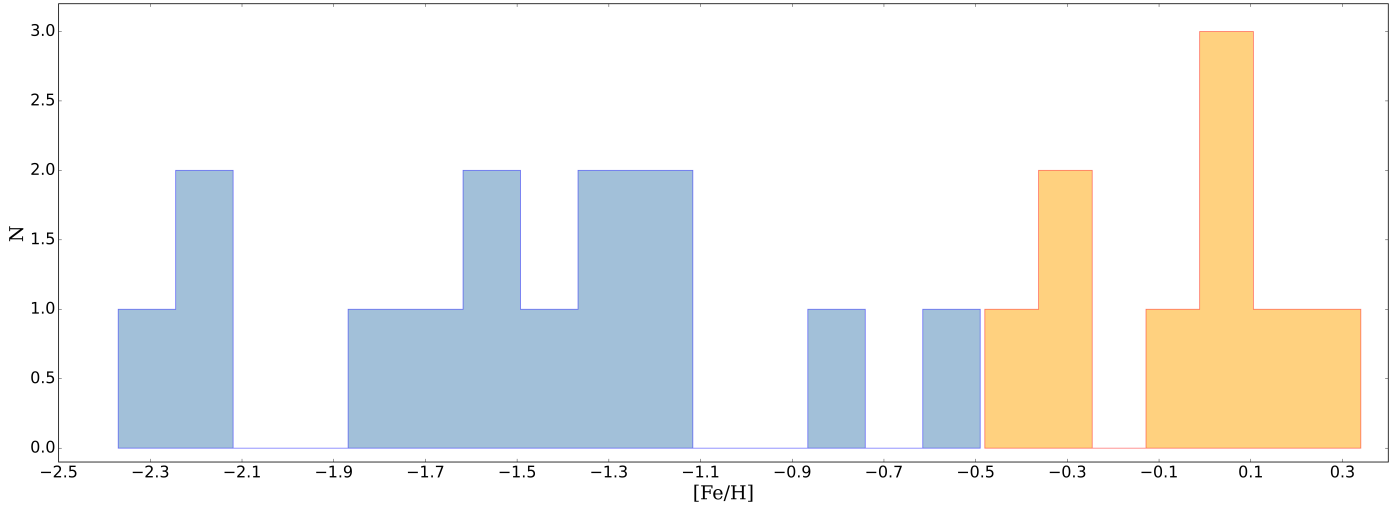
Fig. 12. Histogram of the metallicities covered by the calibrator clusters (globular clusters in blue, open clusters in orange).

stars with well-known 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. 5. 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).

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 samples within a WG, and the analysis of common calibration samples across WGs. To provide a final consistent set of results, the role of WG15 was to homogenise the recommended results from WGs 10–13, transposing them onto a common scale. The main product of WG15 is a catalogue of recommended astrophysical 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. 6: the analysis process starts from 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[5] and the analysis results catalogues from the nodes and WGs[6]. 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.

3.2. The homogenisation flow: From stellar parameters to elemental abundances

In this section, we summarise the homogenisation flow; the quality checks performed during each step are described in Sects. 4,
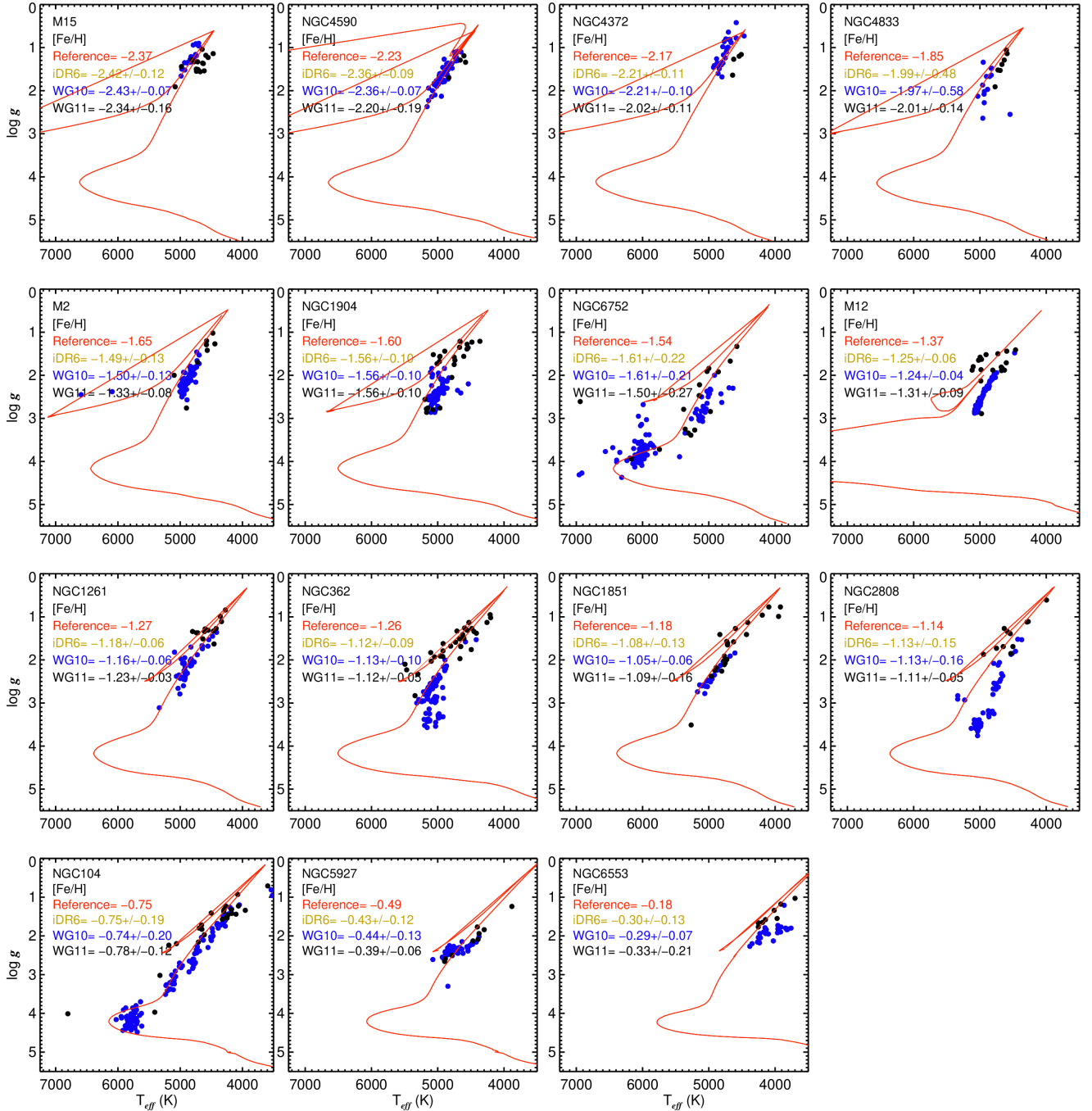

Fig. 13. Kiel diagram for each of the 15 globular clusters in iDR6 shown as increasing in metallicity from left to right. An isochrone at the reference values of age and metallicity for each globular cluster is displayed. The results are separated by WG: WG11 in black and WG10 in blue.

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, for details

on the analysis of each WG, Worley et al., in prep.). 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. (in prep.). 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.
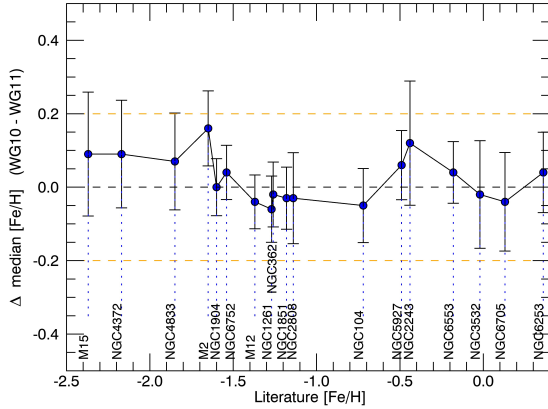
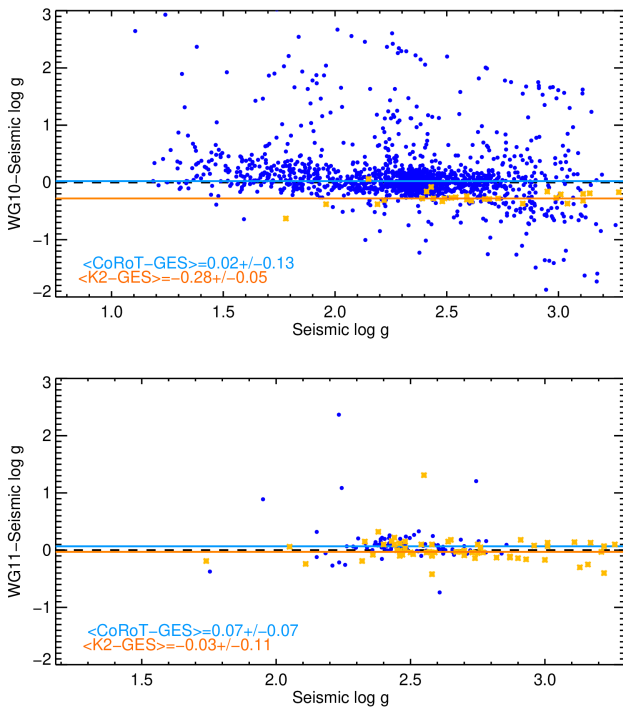**Fig. 14.** [Fe/H] bias amongst the GES GCs and calibrating OCs.

Fig. 15. Comparison between WG10 (top) and WG11 (bottom) $\log g$ values and the seismic $\log g$ from CoRoT and K2. The median and MAD are also given.

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 Fig. 7.

In the case where a given target is analysed by more than one WG, the results from the WG11 high-resolution analysis are taken with priority. In the case of hot stars, which means mostly young clusters (age < 100 Myr), the parameters of WG13 are chosen if available and if $T_{eff} > 7000$ K. Finally, in the case of cool stars ($T_{eff} < 7000$ K) in young clusters (age < 100 Myr), the results of WG12 are preferred.

Once WG15 has produced a single set of stellar parameters per CNAME, the homogenisation cycle returns them to the nodes in WGs 10–12, which carry out an abundance analysis. For both the parameter and abundance phases, nodes were

provided with a list of verifications to carry out on the key calibration sets for their own quality control prior to submission. With the relevant sets of node files, the WG leads performed their own internal homogenisation of the parameters in the parameter phase and the abundances in the abundance phase to provide a set of recommended results for each CNAME specific to the WG datasets.

At the end of each phase, WG15 performed a cross-WG homogenisation to produce the final recommended set of parameters for the parameter phase, and then abundances for the abundance phase. The homogenised WG14 flags are also included (see Van Eck et al., in prep., for details on the analysis of each WG). A first-pass analysis of the flags becomes available as a reference set during the parameter homogenisation.

The radial velocities (RVs) are homogenised by WG15 in parallel using stars analysed in common across different instrumental setups for calibration, as is done for parameter and abundance homogenisation. The procedure followed for radial and rotational velocity homogenisation is described in Sect. 7.

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_{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., in prep.), 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_{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.
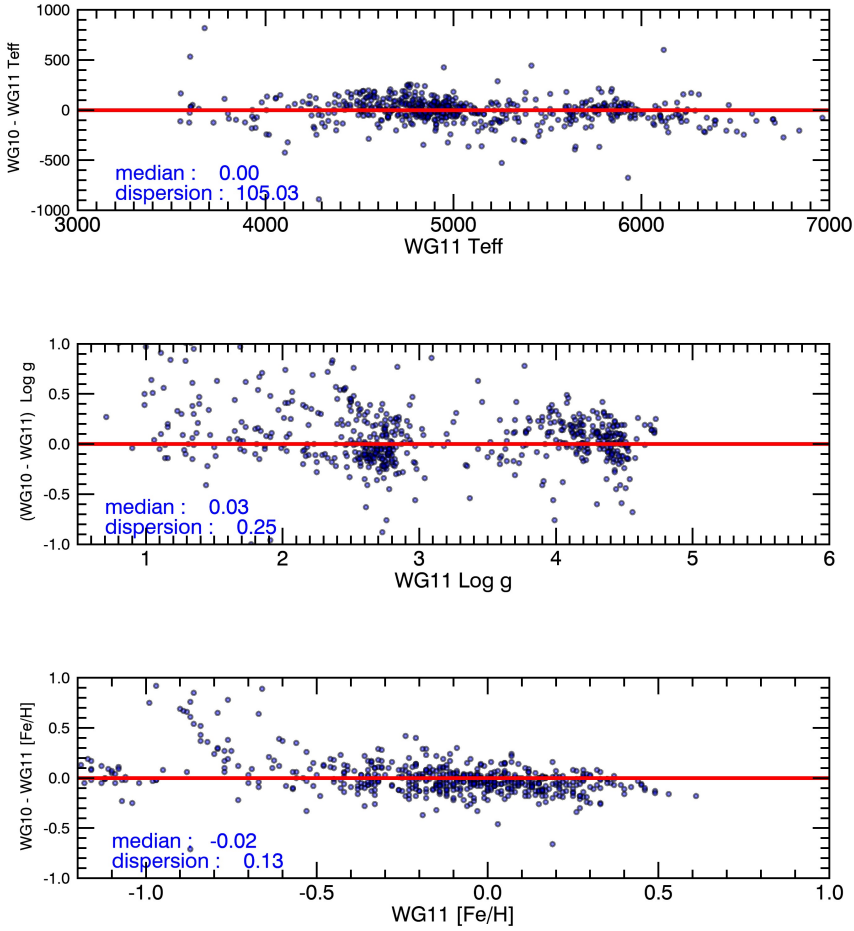
Fig. 16. Comparison of T_{eff}, $\log g$, and [Fe/H] between WG11 and WG10 for stars in common.

The homogenised WG15 results are shown in Fig. 8. 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 (τ Sco and γ 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_{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_{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_{eff}–$\log g$ plane for the stars for both samples and check for offsets in T_{eff} and/or $\log g$ (see Fig. 9 for the bin centred 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_{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 Br 32 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 Fig. 10, where we plot the stellar parameters of
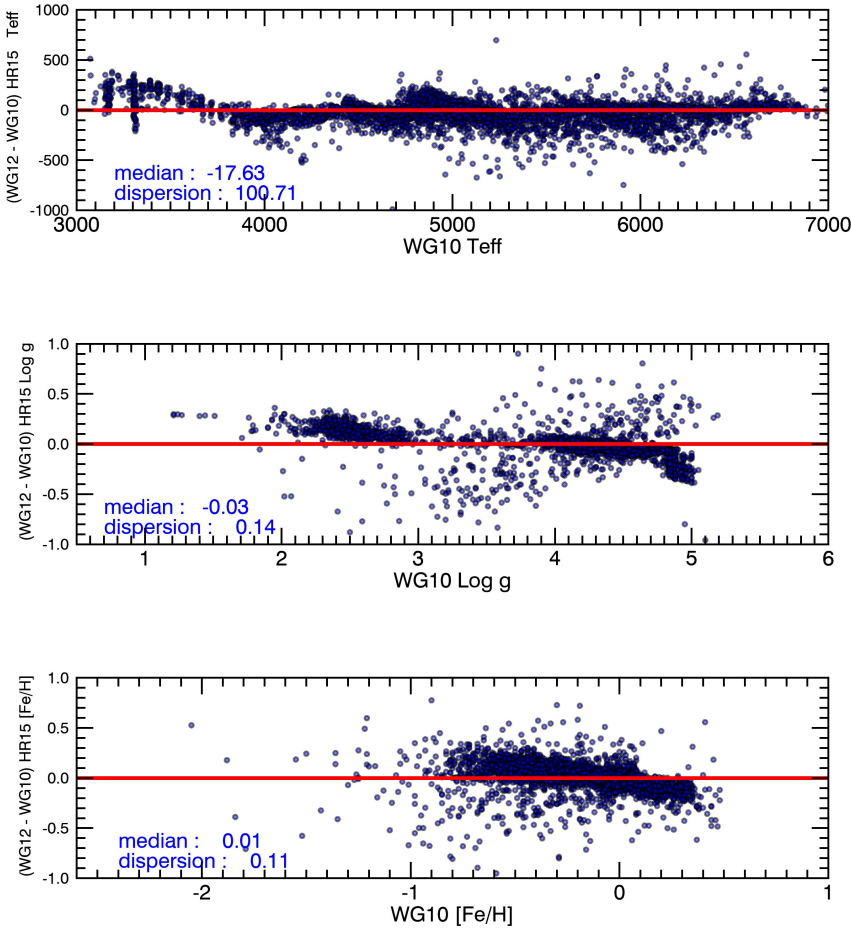
Fig. 17. Comparison of T_{eff}, $\log g$, and [Fe/H] between WG12 and WG10 for stars in common in the setup HR15. Median values and dispersion are given in each subpanel.

the members of NGC 6705 – 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. 11, 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. 12 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. (in prep.). An illustrative

isochrone at the reference values for age and metallicity for each globular cluster is shown in Fig. 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, 1512 CNAMEs had both CoRoT and
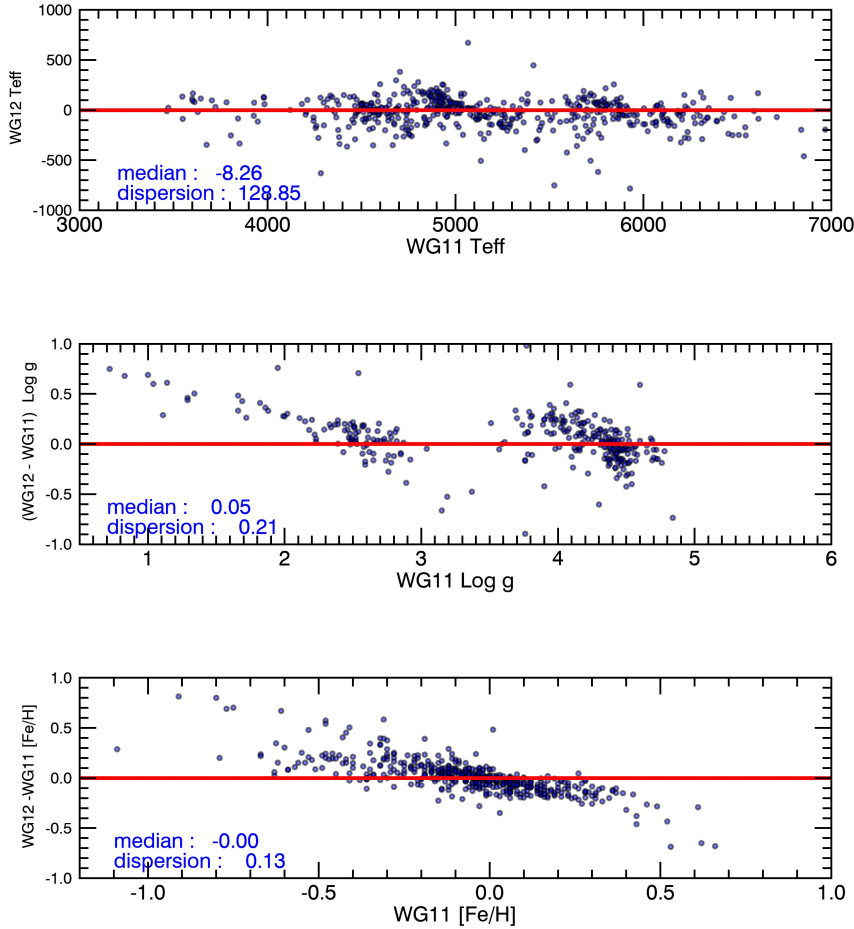
Fig. 18. Comparison of stellar parameters between WG12 and WG11 for stars in common.

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 Fig. 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.

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. These targets make it possible to verify the consistency of the results obtained by the various 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_{eff}, log g, and [Fe/H]. The results are in very

good agreement, with a low median difference for the different parameters.

There are 5171 spectra in common between WG12 and WG10 for the setup HR15. Most of these are stars in clusters analysed by both working groups, and then there are the usual calibrators, mainly benchmark stars. Figure 17 shows the comparison between the WG12 and WG10 results for T_{eff}, log g, and [Fe/H]. The results are in good agreement, as shown by the median difference and the dispersion. No bias correction is needed.

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 Fig. 18, we show the comparison between WG11 and WG12 results for T_{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 $g \simeq 1$ 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
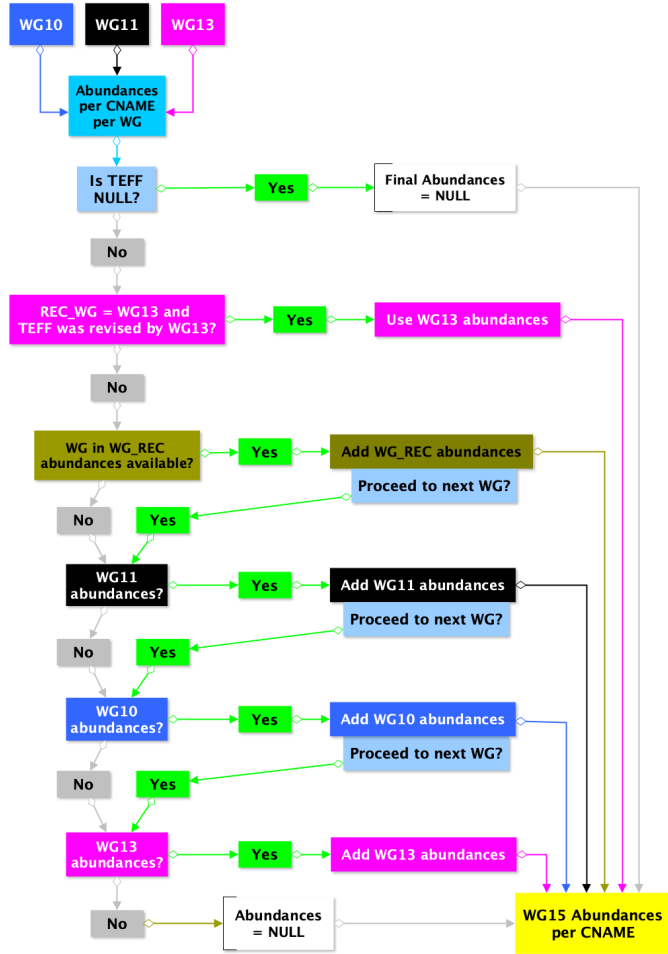
Fig. 19. Schematic view of the GES WG15 Abundance Homogenisation Workflow.

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. (in prep.), alleviated any eventual discrepancy

between the results of the two WGs. As the large wavelength range 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 C_2 Swan (1,0) band head at 5135 Å, the C_2 Swan (0,1) band head at 5635.5 Å, and the forbidden [O I] line at 6300.31 Å 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
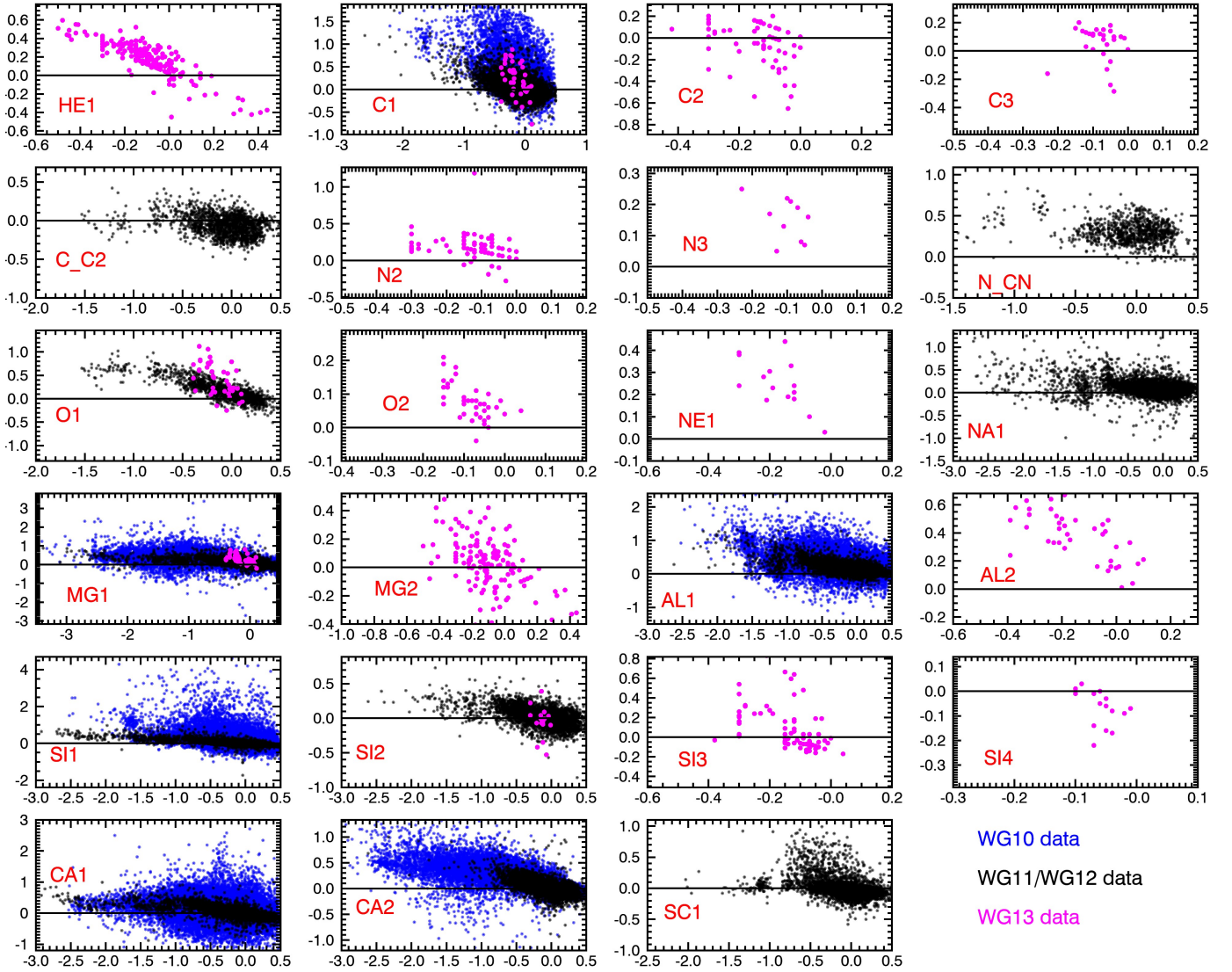
Fig. 20. Abundance ratios [X/H] versus [Fe/H] for all the elements analysed in GES (except Li). The results of WG10 are represented by blue circles, those of WG11 and WG12 are shown as black circles, and WG13 results are represented by magenta circles. C_C2 are carbon abundances from C2 molecular bands and N_CN are nitrogen abundances from CN molecular bands.

with reference values; for some of these we report results for more than one ionisation species (such as CaI and CaII). 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 $T_{\rm eff}$, $\log g$, 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 we compare them with a compilation of recent literature results. For WG10, we selected only the results obtained from spectra with S/N ≥ 10. The figures not only show the excellent agreement between WG10 and WG11 abundance trends, but also demonstrate a very good match with the literature data coming from the very high-resolution (40 000–110 000), very high-S/N (150–300)[7] spectral analyses from Bensby et al. (2014); Battistini & Bensby (2015, 2016). We also clearly see the larger dispersion in the [X/Fe] ratios as a function of [Fe/H] found in the WG10 results compared to the WG11 results, which is naturally explained by the lower dispersion of the spectra used by WG10 and the smaller wavelength range.

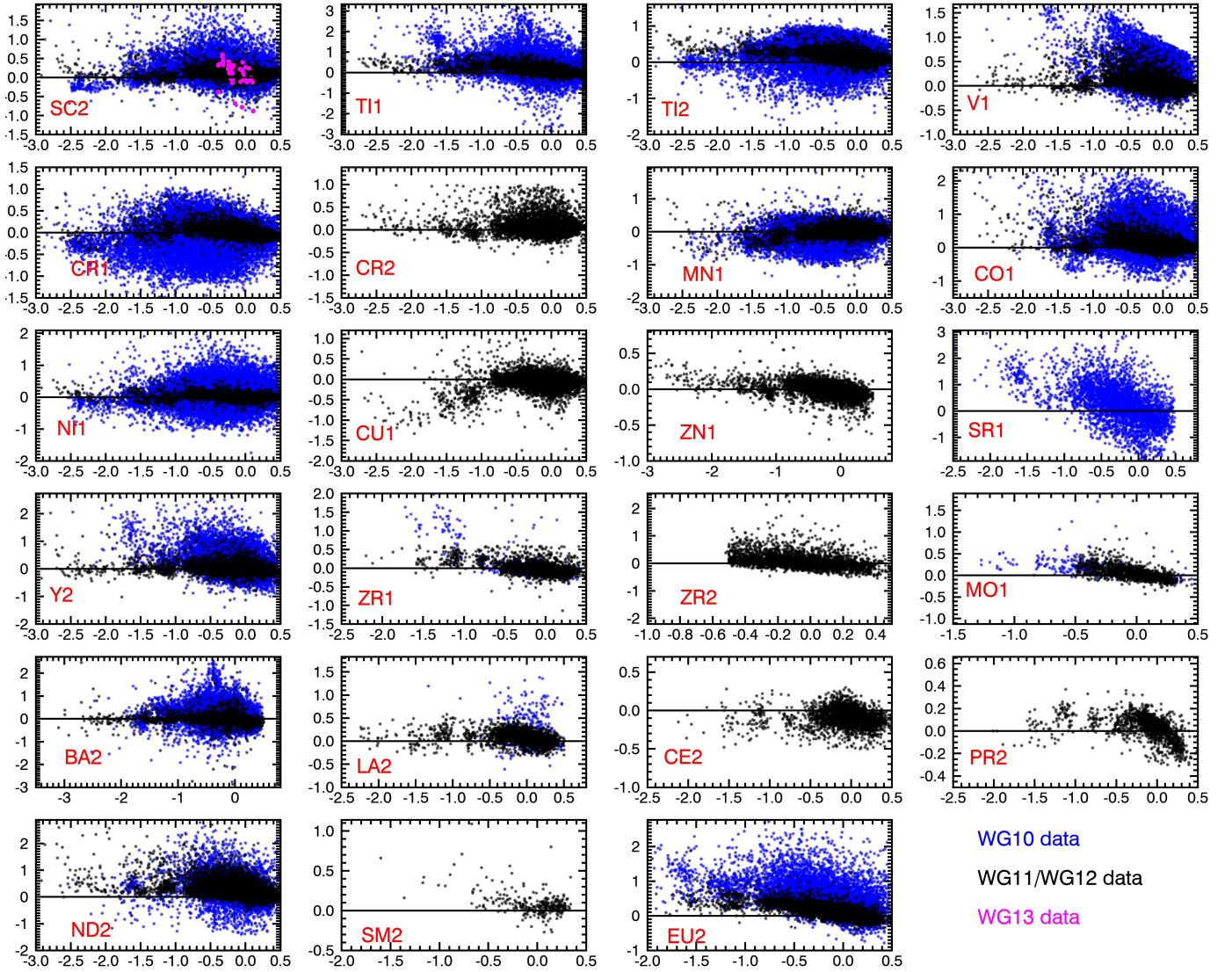
[7] Resolution and S/N values from Bensby et al. (2014).

Fig. 21. Abundance ratios [X/H] versus [Fe/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.

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 α-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 [α/Fe], e_[α/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 1σ, and in slightly worse cases within 2σ, 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,
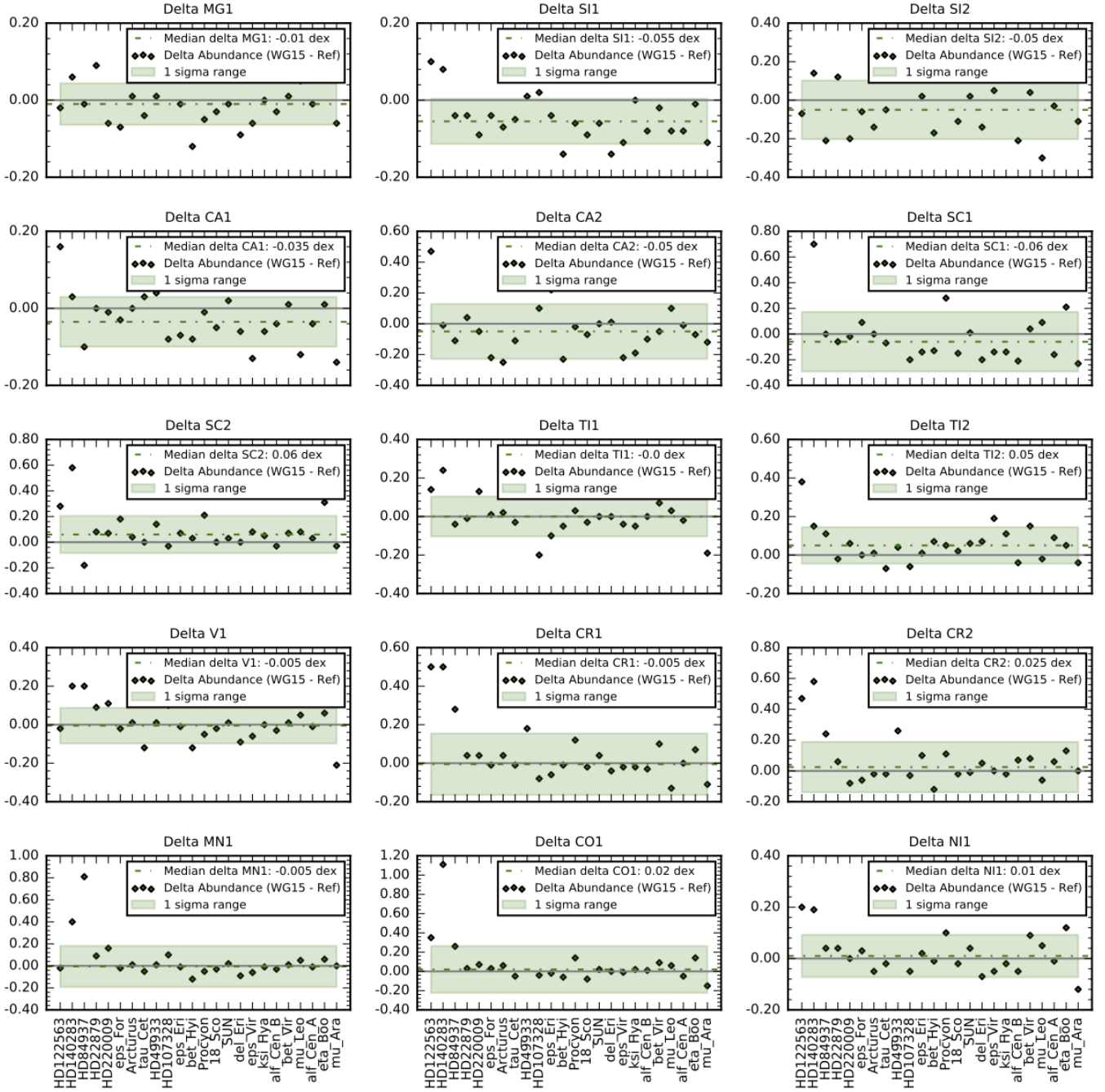
Fig. 22. Delta abundance ratios [X/Fe]: WG15 minus reference abundances for benchmark stars from Jofré et al. (2015).

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 Fig. 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
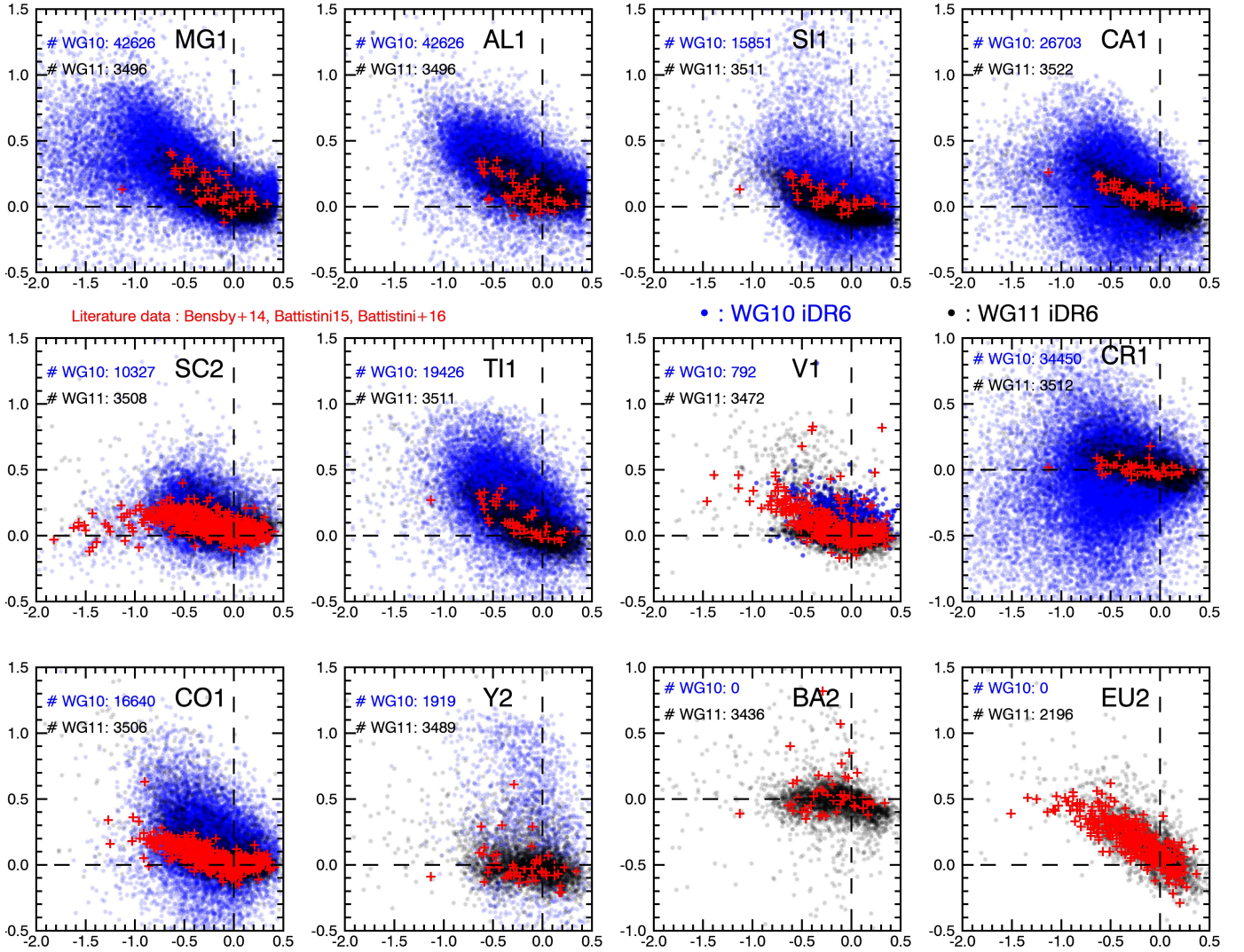
Fig. 23. WG10 and WG11 abundances ratios [X/Fe] vs. [Fe/H] for Milky Way stars: WG10 stars are represented by blue symbols and WG11 stars by black symbols, while the red symbols refer to literature data (Bensby et al. 2014; Battistini & Bensby 2015, 2016).

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.

Offsets 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 Fig. 29.

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

7.3. Signal-to-noise ratio

The S/N values reported in the final catalogue (in the column, 'SNR') were selected to match the selection of the radial velocities. When a GIRAFFE RV was selected for the homogenised VRAD, the S/N from the specific GIRAFFE SETUP was selected. Similarly, when a UVES RV was selected, the S/N from the specific UVES setup was selected. The sequence is as shown in Fig. 28.

However, often a combination of setups was used to calculate the VRAD. For instance, when combining UVES upper and lower arms, and the combination of setups used to calculate the WG13 VRADs. In these cases, the S/N values from the setups used in the calculated VRAD were summed in quadrature to provide the reported S/N.

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 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 Sect. 5, the final catalogue provides the EWs of Li lines: the measured equivalent width EW_LI, 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,
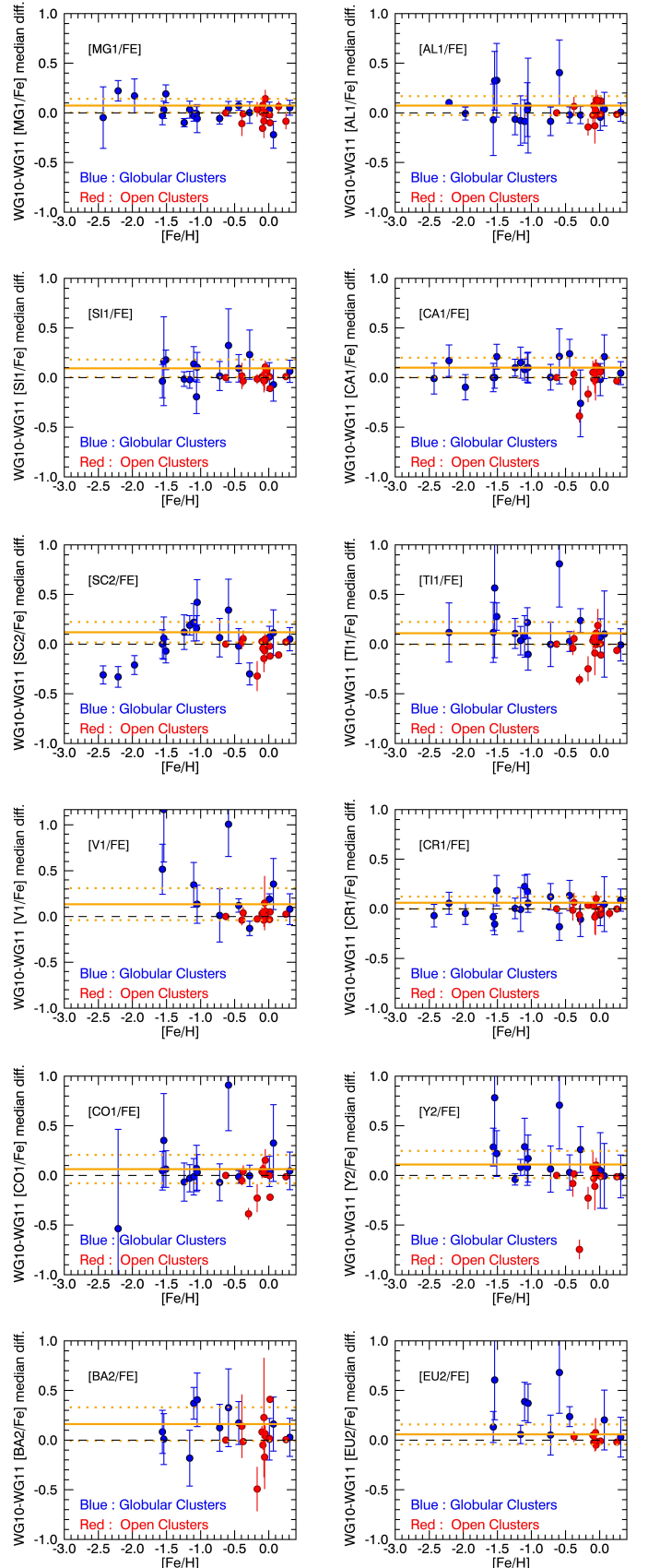

Fig. 24. Open and globular clusters. Each panel represents the results for a different element. [Fe/H] is in abscissa. The ordinates are the values 'median WG11 abundances minus median WG10 abundances'. The median global difference is shown as an orange line. Dotted orange lines indicate ±1σ standard deviation from the median value.

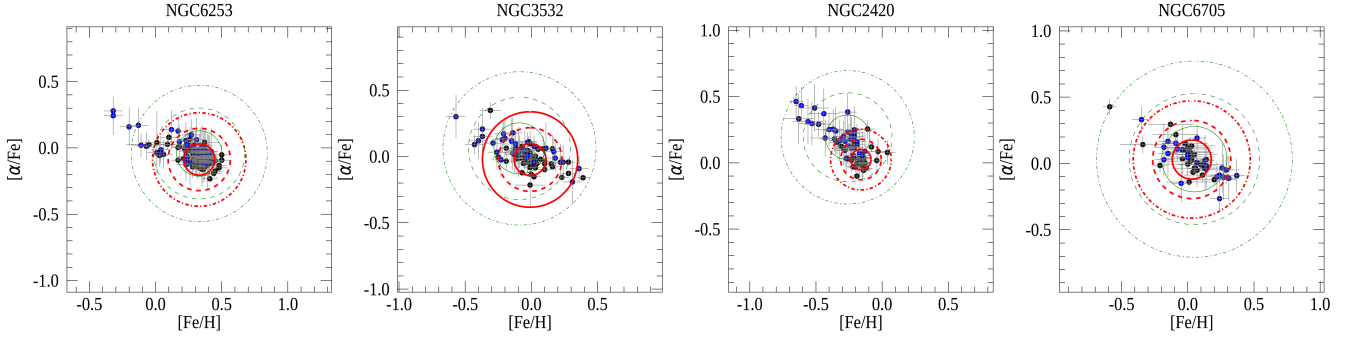**Fig. 25.** α-element abundance over iron, computed by averaging [Mg/Fe], [Si/Fe], [Ca/Fe], and [Ti/Fe] as a function of [Fe/H] in member stars of four calibration open clusters. The abundances from WG11 are shown in black and those from WG10 in blue. The red circles mark the median WG11 abundances within 1σ, 2σ, and 3σ. The green circles indicate the same quantities but computed with WG 10 results.
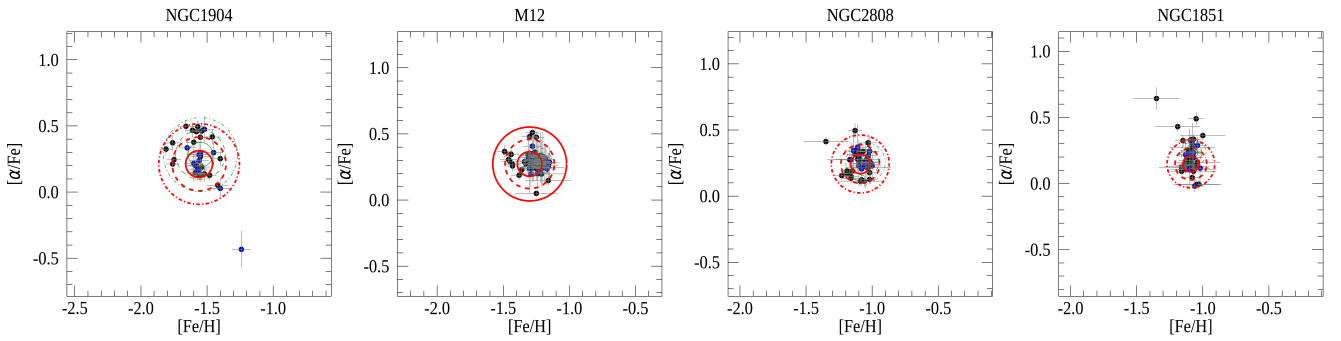
Fig. 26. [α/Fe] as a function of [Fe/H] in member stars of four calibration globular clusters. Symbols and colours are as in Fig. 25.
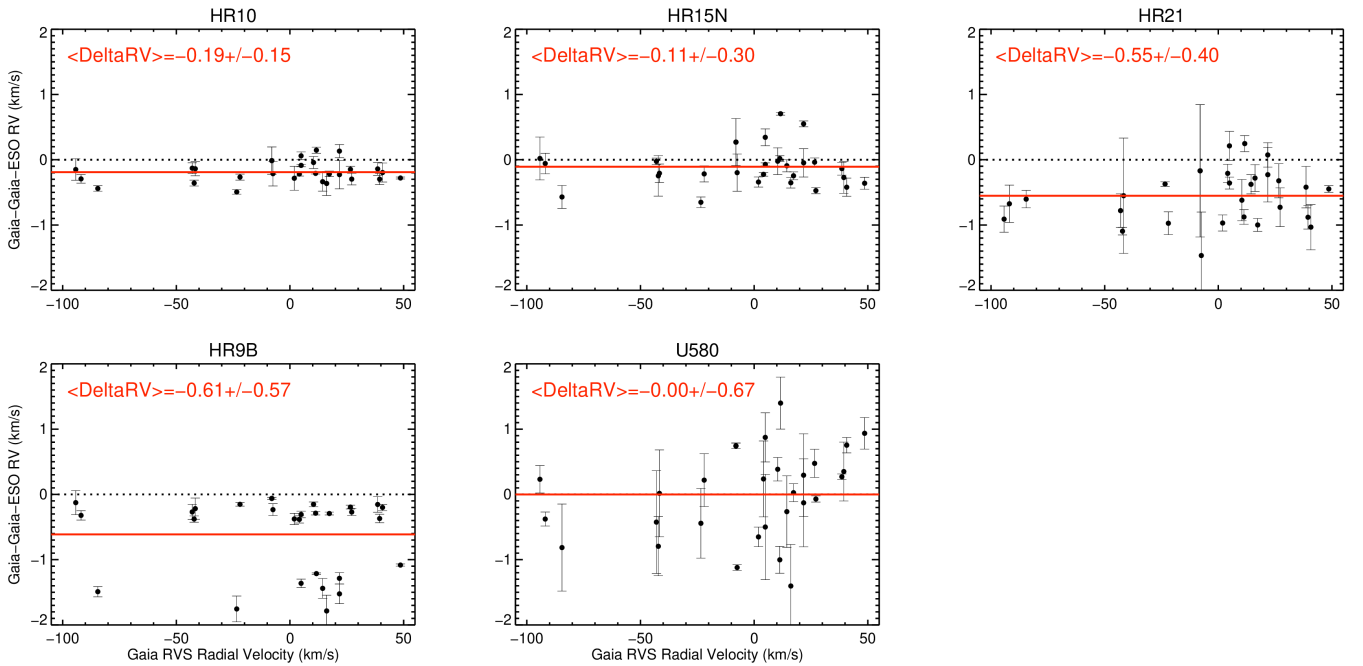
Fig. 27. Difference between GRVS radial velocity and the mean *Gaia*-ESO radial velocity calculated per GRVS per setup. The mean difference and standard deviation per setup are given.

EWC_LI, 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_{eff}, $\log g$,
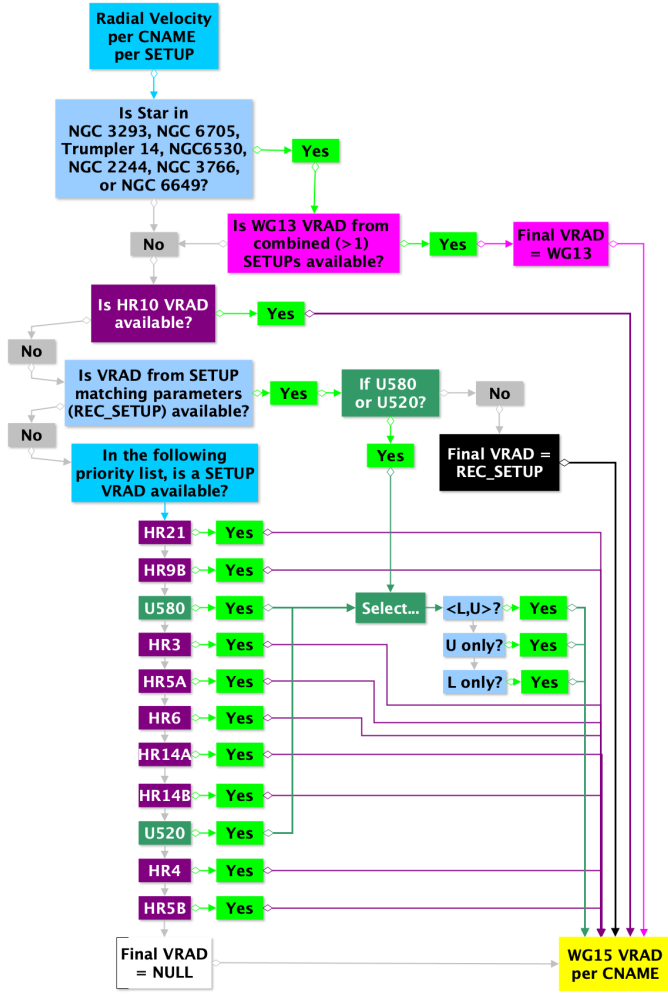
Fig. 28. Schematic view of the GES WG15 radial velocity homogenisation workflow.
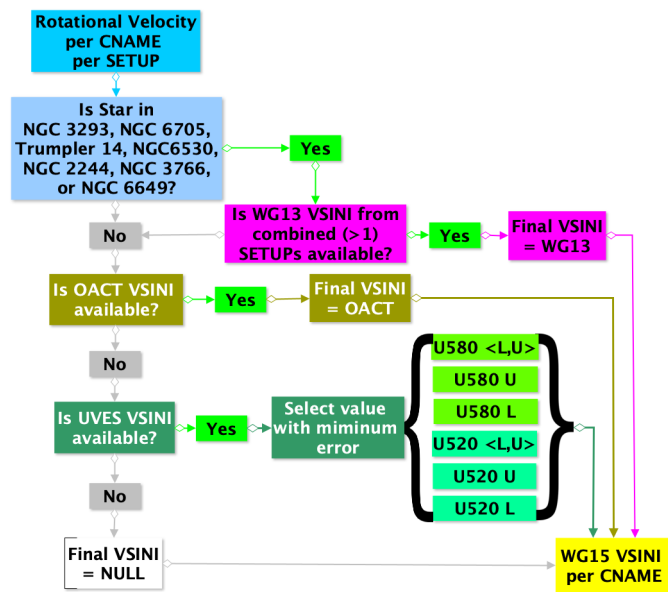

Fig. 29. Schematic view of the GES WG15 rotational velocity homogenisation workflow.

[Fe/H], ξ, $V_{\rm rad}$, and $v \sin i$ as a function of S/N. The provenances of the errors are indicated by colour. For $T_{\rm eff}$, $\log g$, [Fe/H], and ξ, the provenance is one of WG10, WG11, WG12, or WG13, according to the rules of the homogenisation algorithm represented schematically in Fig. 7, and the provenance of the associated error aligns with that of the parameter. For $V_{\rm rad}$ 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 Sect. 7.1 for the provenance selection details. For $v \sin i$, the provenance is one of WG13, Arcetri, or OACT. See Sect. 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 Sect. 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 it 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., in prep., 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 TEFF, LOGG, FEH, XI, MH, and ALPHA_FE and all associated number and error columns in 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.
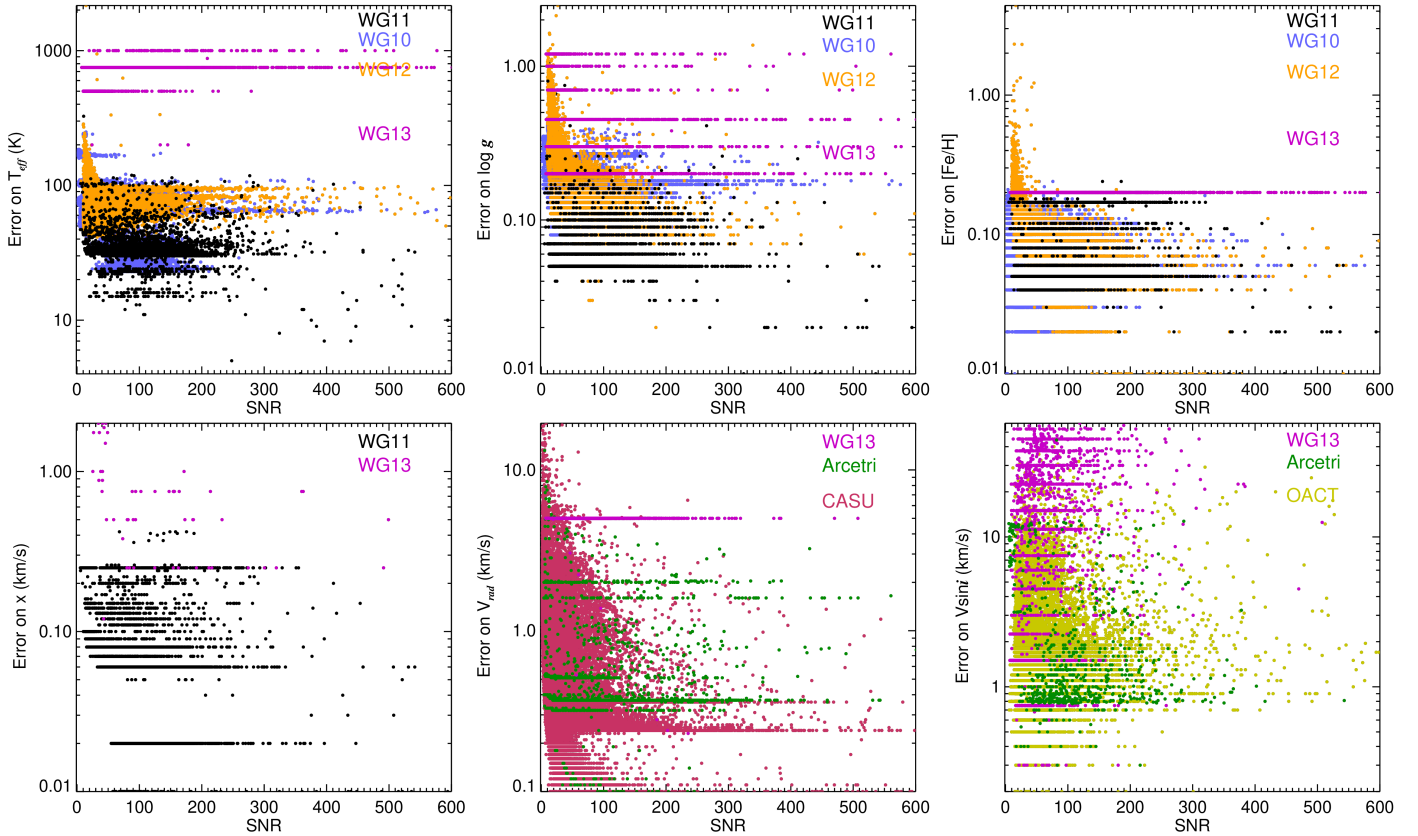
Fig. 30. Error on $T_{\rm eff}$, $\log g$, [Fe/H], ξ, $V_{\rm rad}$ and $v \sin i$ against S/N. The provenances of the errors are indicated by colour (blue – WG10, black – WG11, orange – WG12, magenta – WG13, green – Arcetri node, red – CASU, yellow – OACT node).

Table 3. Bootstrap procedure applied per setup (X) for calibration of radial velocities to HR10.

Setup X	N XxHR10	Setup Y	X-Y Offset	STD	N XxY	X-BS-HR10 Offset	STD
HR15N	3161	HR10	0.09	0.21	3161	0.09	0.21
HR21	48 133	HR10	0.59	0.46	48 133	0.59	0.46
HR9B	253	HR15N	0.08	1.30	558	0.17	1.32
U580	154	HR15N	0.21	0.53	695	0.30	0.57
U520	46	U580	–0.10	0.92	54	0.20	1.08
HR3	123	HR9B	–0.13	2.75	334	0.04	3.11
HR4	116	HR9B	–0.43	5.03	211	-0.26	5.20
HR5A	131	HR9B	–0.10	3.98	151	0.07	2.57
HR5B*	1	HR10	0.15	0.22	1	0.15	0.22
HR6	118	HR9B	0.00	7.03	123	0.17	2.74
HR14A	145	HR9B	0.16	2.37	146	0.33	3.97
HR14B*	1	HR10	0.29	0.22	1	0.29	0.22
WG13Comb	506	HR10	0.24	3.94	507	0.24	3.94
* Only Sun in common							

Notes. 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.

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 with non-physical or highly suspicious results, and they complement the information already carried by the error bars
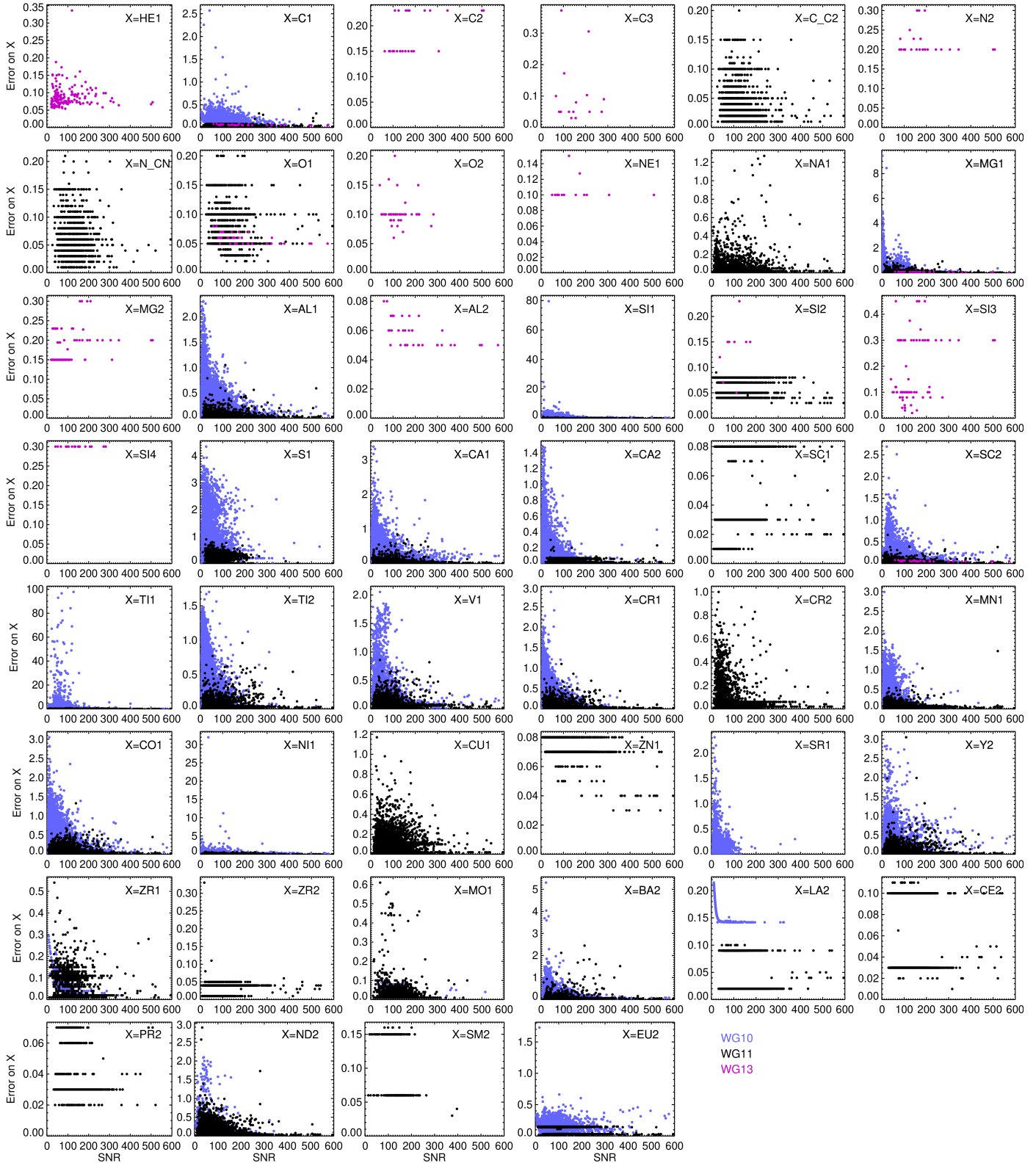
Fig. 31. WG level error on each element abundance against SNR. The provenance of the errors are indicated by colour (blue – WG10, black – WG11, magenta – WG13).

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
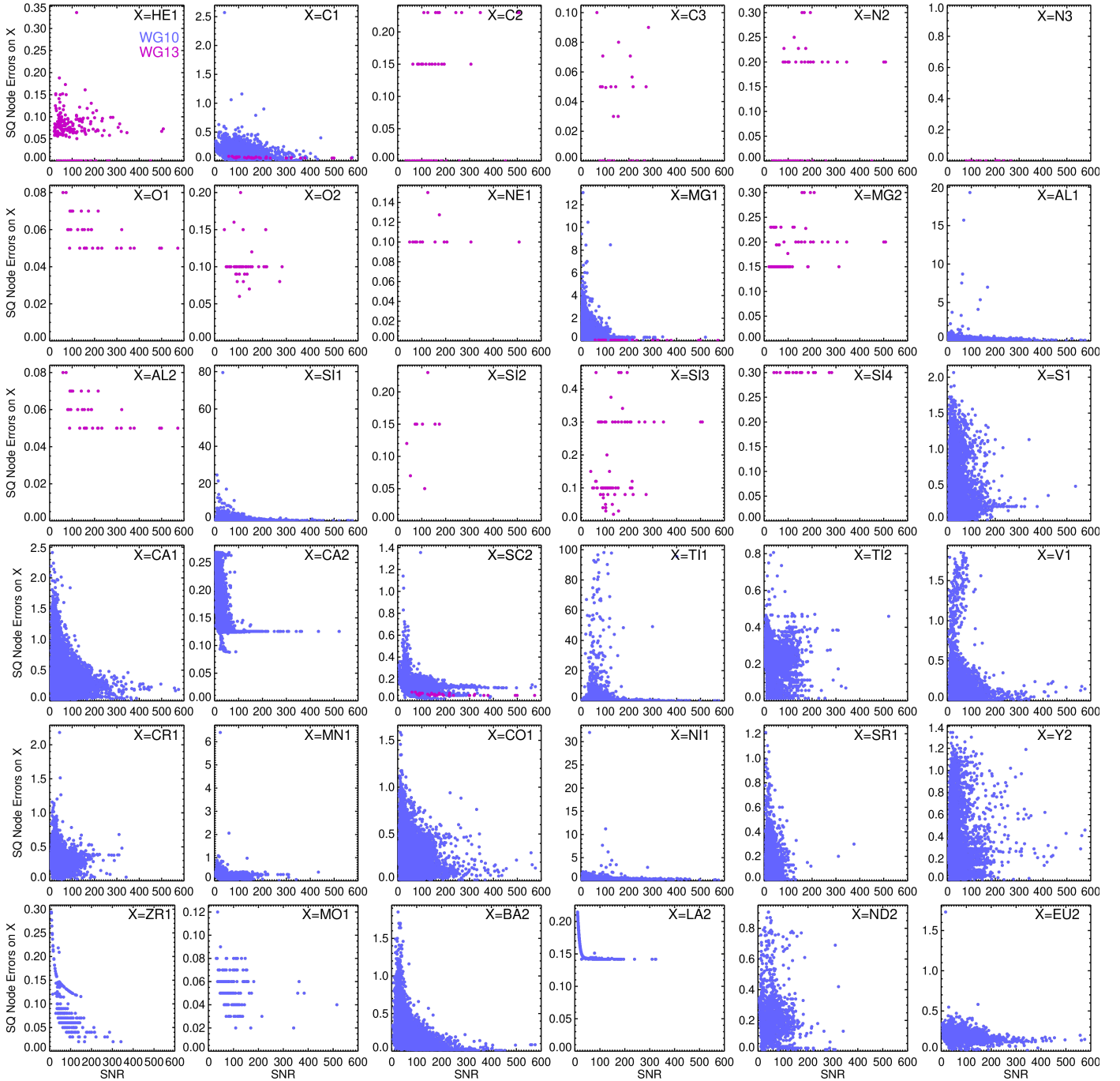
Fig. 32. Errors based on the node analysis errors on abundances as a function of S/N. The provenance of the errors are indicated by colour (blue – WG10, magenta – WG13).

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).
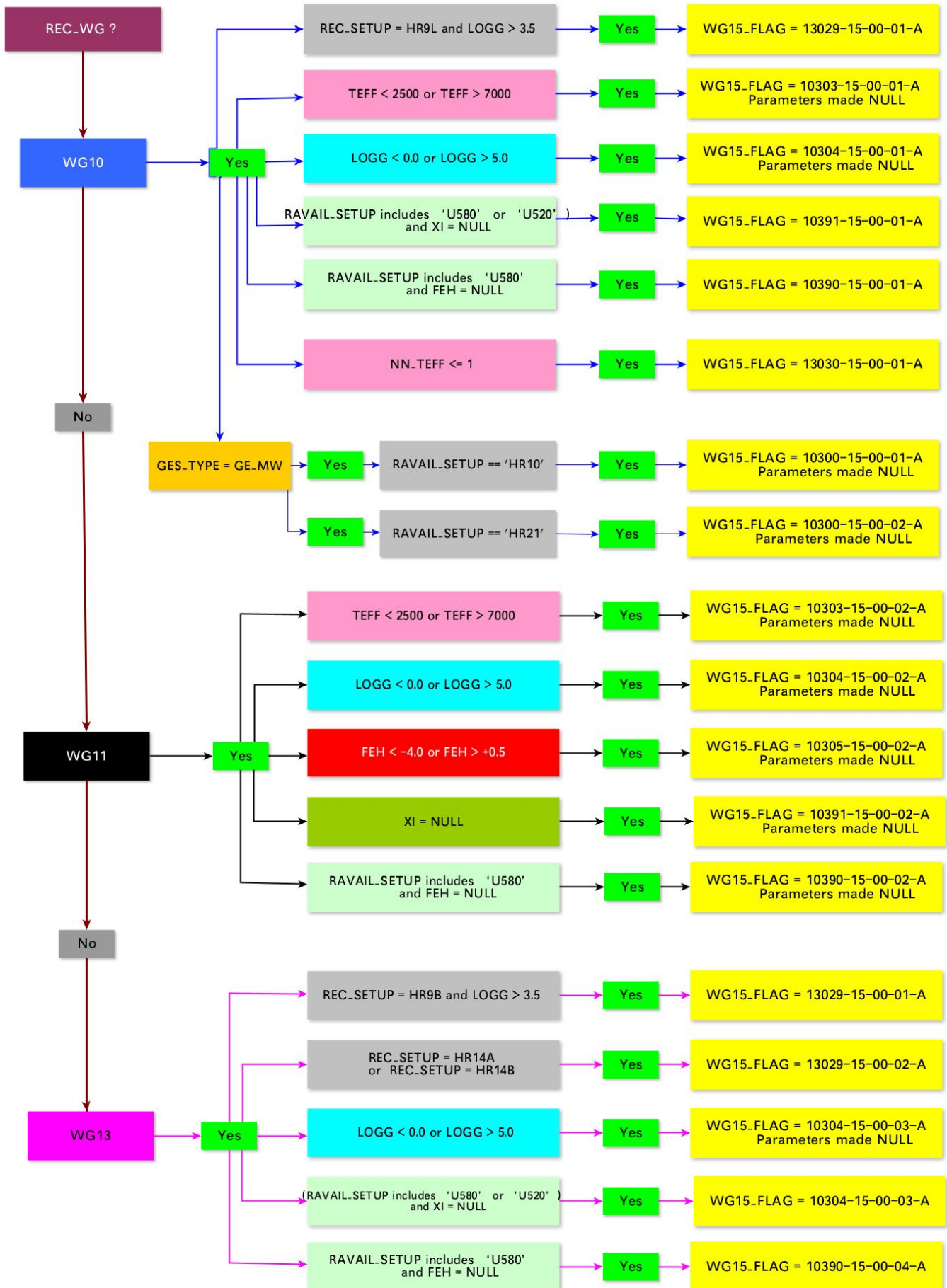
Fig. 33. WG15 additional flags flow chart.

A129, page 26 of 33
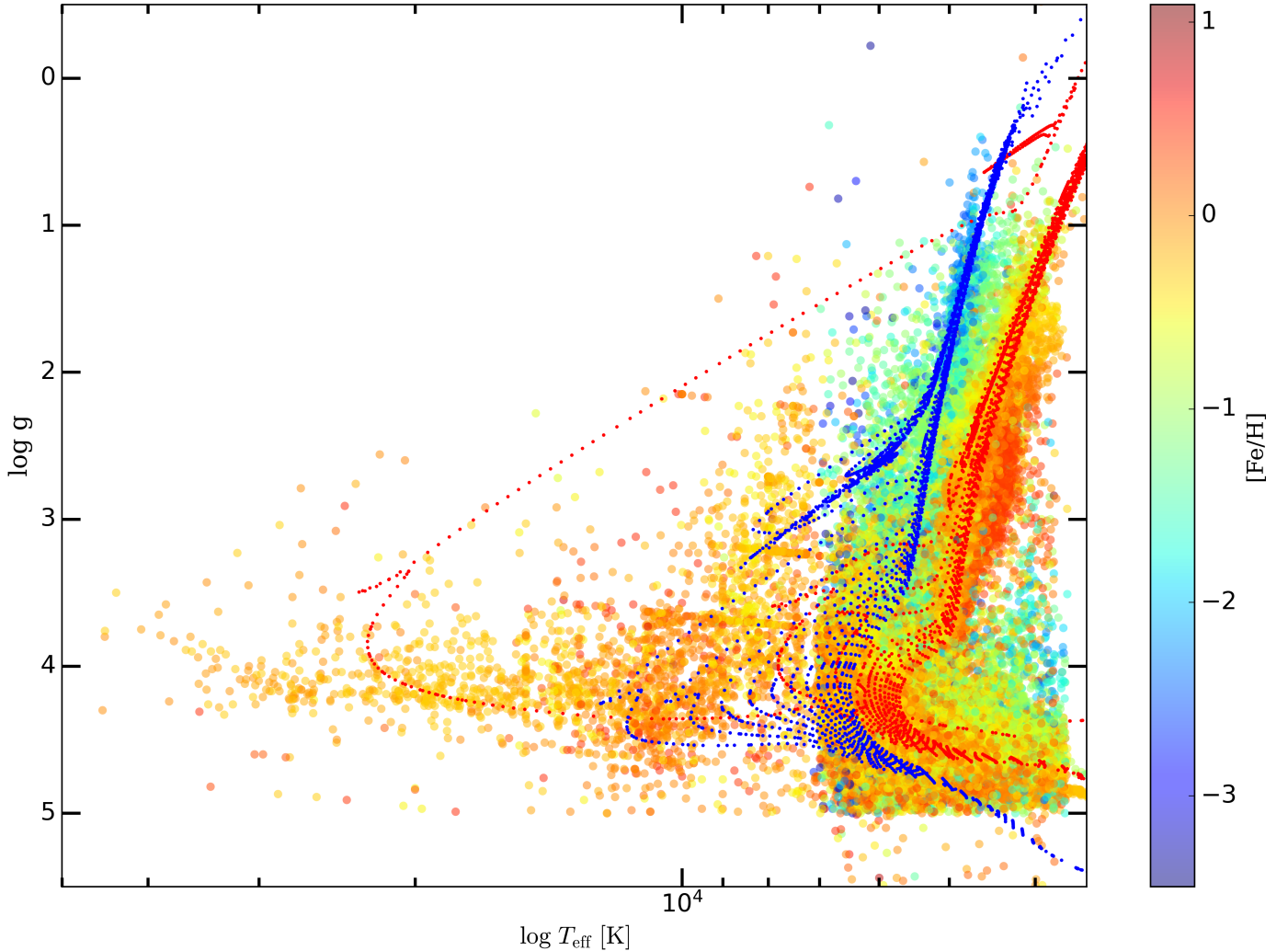
Fig. 34. Kiel diagram using the final recommended values for iDR6 (all stars). Two sets of isochrones at solar metallicity and 0.01<Age[Gyr]<13 (in red) and at [Fe/H]= −2 and 1<Age[Gyr]<13 (in blue) are shown.

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[8].

10.1. Final Kiel diagram

Figure 34 shows the Kiel diagram of the entire latest release of GES (>114 000 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 [Fe/H] = −2 dex indicates very good agreement, with a shift of RGB stars towards higher temperatures for the more metal-poor stars. Also noticeable from the figure is the intrinsic difficulty in measuring the metallicities of cool PMS stars, whose spectra are dominated by molecular bands.
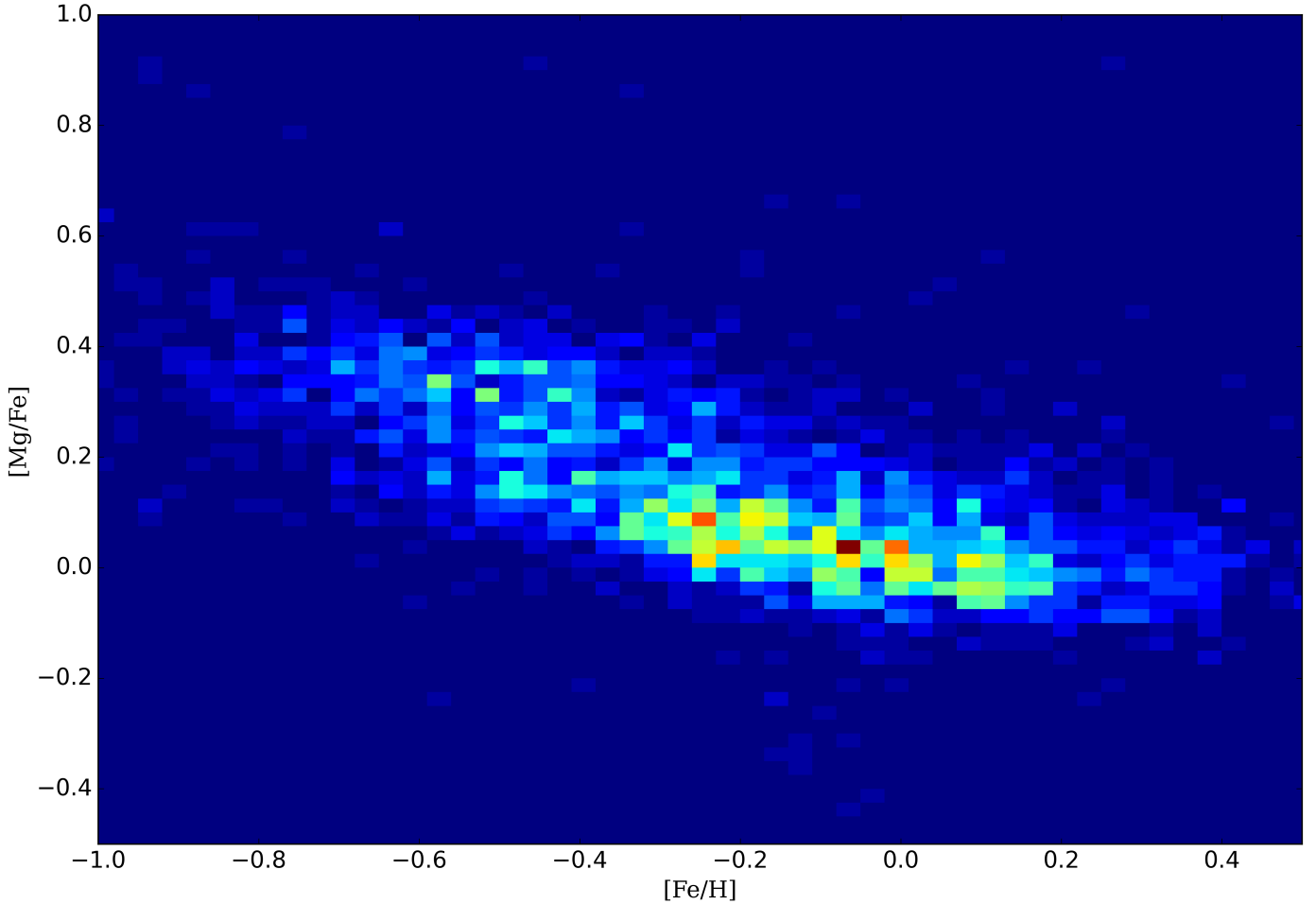
---

[8] https://www.eso.org/qi/catalogQuery/index/393

Fig. 35. [Mg/Fe] vs. [Fe/H] diagram for the MW field stars plotted using the iDR6 final recommended values (selecting GES_TYPE=GE_MW stars and $S/N > 50$).

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 α-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. (2023). 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
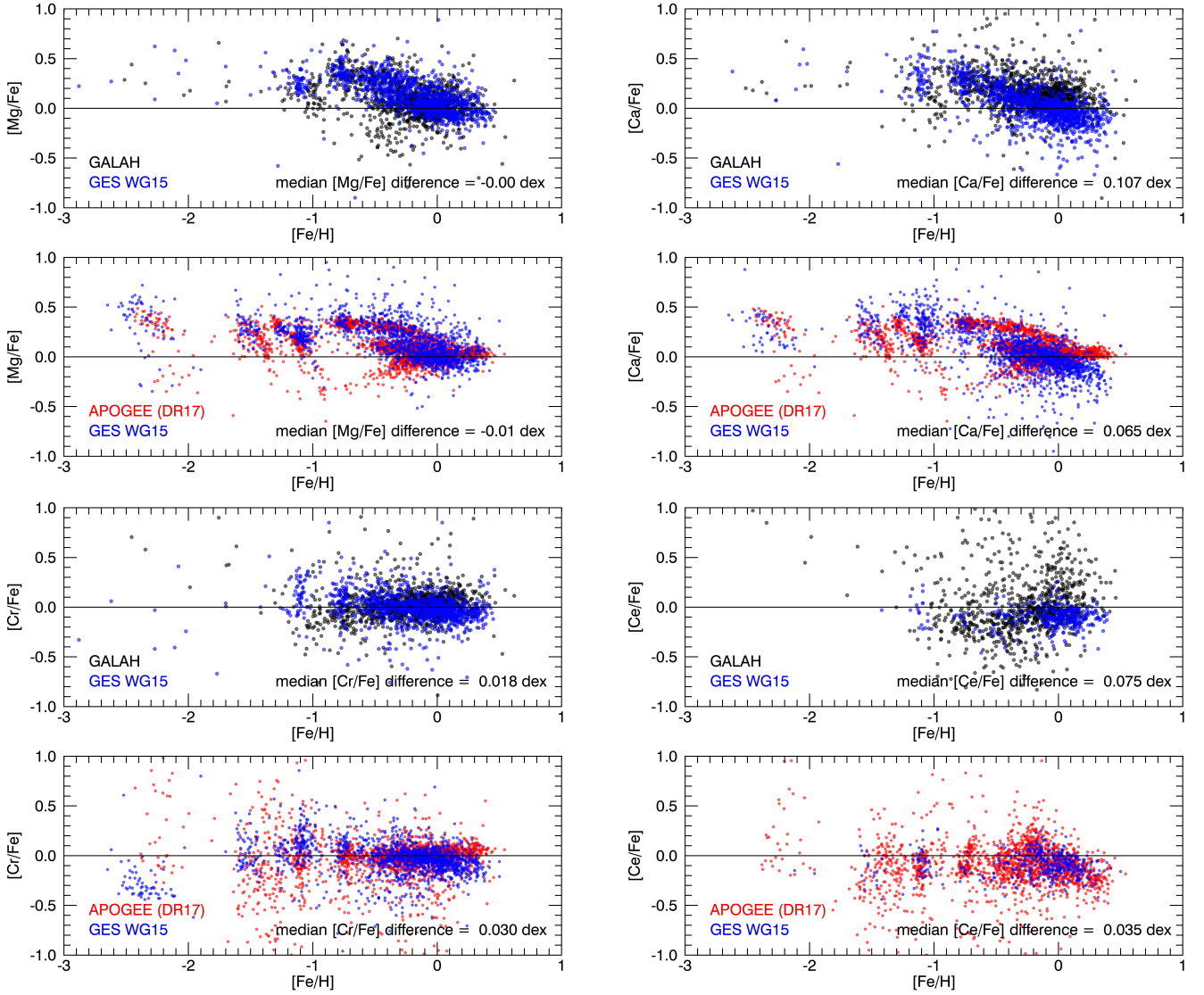
Fig. 36. Comparison between WG15, GALAH, and APOGEE abundance ratios for the sample of stars in common. WG15 data are represented as blue circles. Red symbols represent APOGEE data and black symbols are the data from the GALAH survey.
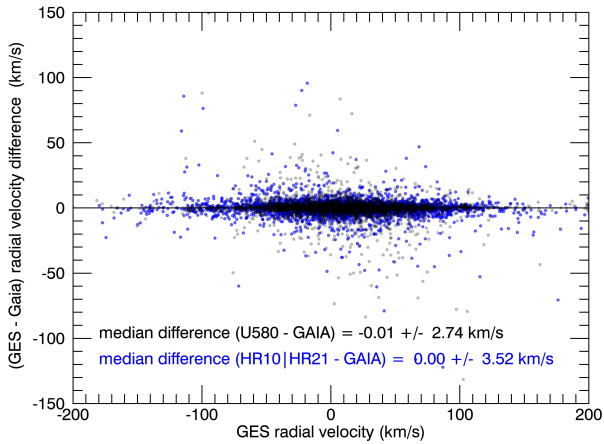

Fig. 37. Radial velocity difference between GES and *Gaia* DR3 for the stars in common. The GES samples HR10|HR21 and U580 are plotted in different colours.

accuracy of *Gaia* compared to GES, and, as expected, a lower precision for *Gaia* compared to GES.

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 8 m 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

Table 4. Median abundance [X/Fe] difference between GALAH, APOGEE, and WG15 for the sample of stars in common.

Element	GALAH − WG15		APOGEE − WG15	
	offset	std dev	offset	std dev
C I	0.16	0.35	0.01	0.31
O I	−0.05	0.36	−0.09	0.15
Na I	0.01	0.19	−0.08	0.34
Mg I	−0.01	0.18	−0.01	0.16
Al I	0.04	0.22	−0.12	0.17
Si I	0.04	0.24	−0.02	0.22
S I			0.33	0.76
Ca I	0.11	0.23	0.02	0.19
Sc I	0.10	0.17		
Ti I	0.03	0.23	−0.02	0.22
Ti II	−0.05	0.25		
V I	0.13	0.35		
Cr I	0.02	0.24	0.03	0.31
Mn I	0.00	0.32	0.04	0.22
Co I	0.06	0.42	−0.01	0.29
Ni I	0.04	0.22	0.02	0.16
Cu I	0.09	0.23		
Zn I	0.04	0.29		
Y II	−0.03	0.39		
Zr I	0.12	0.52		
Mo I	0.15	0.37		
Ba II	0.10	0.36		
La II	0.06	0.34		
Ce II	0.07	0.32	0.03	0.18
Nd II	0.17	0.39		
Sm II	−0.05	0.44		
Eu II	−0.05	0.22		

Notes. For the majority of the elements, the median difference is below 0.1 dex.
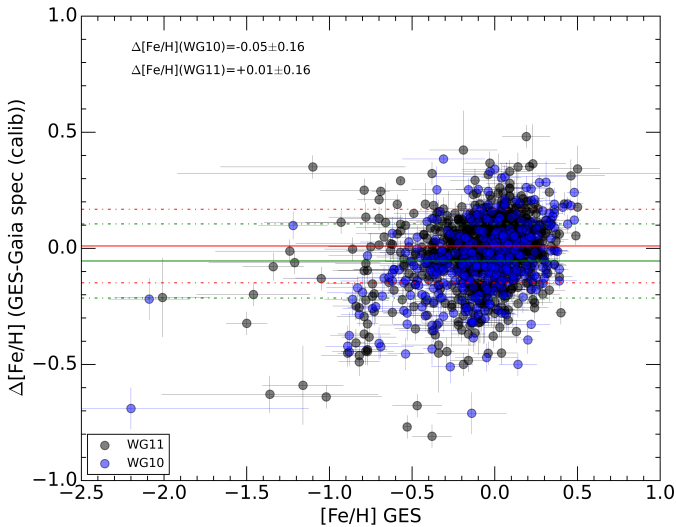
Fig. 38. [Fe/H] difference between GES and *Gaia* DR3 for the stars in common. The GES samples GIRAFFE (all setups) and U580 are plotted in different colours. The horizontal red lines mark the median difference U580 – *Gaia* DR3 calibrated metallicity (continuous line) and ±1σ (dot-dashed lines). The green lines indicate the median difference GIRAFFE – *Gaia* DR3 calibrated metallicity (continuous line) and ±1σ (dot-dashed lines).

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 velocities for more than 110 000 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.

References

Bailer-Jones, C. A. L. 2011, MNRAS, 411, 435
Battistini, C., & Bensby, T. 2015, A&A, 577, A9
Battistini, C., & Bensby, T. 2016, A&A, 586, A49
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Blanco-Cuaresma, S., Soubiran, C., Jofré, P., & Heiter, U. 2014, A&A, 566, A98
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Blomme, R., Daflon, S., Gebran, M., et al. 2022, A&A, 661, A120
Buder, S., Sharma, S., Kos, J., et al. 2021, MNRAS, 506, 150
Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1
Cirasuolo, M., et al. 2020, The Messenger, 180, 10
Dalton, G., et al. 2012, SPIE Conf. Ser., 8446, 84460P
Damiani, F., Prisinzano, L., Micela, G., et al. 2014, A&A, 566, A50
de Jong, R. S., et al. 2014, SPIE Conf. Ser., 9147, 91470M
De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, MNRAS, 449, 2604
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, Proc. SPIE, 4008, 534
Franciosini, E., Randich, S., de Laverny, P., et al. 2022, A&A, 668, A49
Friel, E. D., Jacobson, H. R., & Pilachowski, C. A. 2010, AJ, 139, 1942
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1
Gilmore, G., Randich, S., Worley, C. C., et al. 2022, A&A, 666, A120
González Hernández, J. I., & Bonifacio, P. 2009, A&A, 497, 497
Harris, W. E. 1996, AJ, 112, 1487
Hawkins, K., Jofré, P., Heiter, U., et al. 2016, A&A, 592, A70
Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, ApJ, 808, 132
Heiter, U., Jofré, P., Gustafsson, B., et al. 2015a, A&A, 582, A49
Heiter, U., Lind, K., Asplund, M., et al. 2015b, Phys. Scr., 90, 054010
Heiter, U., Lind, K., Bergemann, M., et al. 2021, A&A, 645, A106
Jackson, R. J., Jeffries, R. D., Lewis, J., et al. 2015, A&A, 580, A75
Jackson, R. J., Jeffries, R. D., Wright, N. J., et al. 2022, MNRAS, 509, 1664
Jofré, P., Heiter, U., Soubiran, C., et al. 2014, A&A, 564, A133
Jofré, P., Heiter, U., Soubiran, C., et al. 2015, A&A, 582, A81
Jofré, P., Heiter, U., & Buder, S. 2017, in Astronomical Society of India Conference Series, 14, 37
Jofré, P., Heiter, U., Tucci Maia, M., et al. 2018, RNAAS, 2, 152
Johansson, S., Litzén, U., Lundberg, H., & Zhang, Z. 2003, ApJ, 584, L107
Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv e-prints, [arXiv:1711.03234]
Lanzafame, A. C., Frasca, A., Damiani, F., et al. 2015, A&A, 576, A80
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Modigliani, A., Mulas, G., Porceddu, I., et al. 2004, The Messenger, 118, 8
Pancino, E., Lardo, C., Altavilla, G., et al. 2017a, A&A, 598, A5
Pancino, E., Romano, D., Tang, B., et al. 2017b, A&A, 601, A112
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, Astrophysics Source Code Library, [record ascl:1205.004]
Randich, S., Sestito, P., Primas, F., Pallavicini, R., & Pasquini, L. 2006, A&A, 450, 557
Randich, S., Gilmore, G., Magrini, L., et al. 2022, A&A, 666, A121
Recio-Blanco, A., de Laverny, P., Kordopatis, G., et al. 2014, A&A, 567, A5
Recio-Blanco, A., de Laverny, P., Palicio, P. A., et al. 2023, A&A, 674, A29
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
Smiljanic, R., Korn, A. J., Bergemann, M., et al. 2014, A&A, 570, A122
Soubiran, C., Jasniewicz, G., Chemin, L., et al. 2013, A&A, 552, A64
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Stonkutė, E., Koposov, S. E., Howes, L. M., et al. 2016, MNRAS, 460, 1131
Tautvaišienė, G., Drazdauskas, A., Mikolaitis, Š., et al. 2015, A&A, 573, A55
Wallerstein, G. 1962, ApJS, 6, 407

Wells, D. C., Greisen, E. W., & Harten, R. H. 1981, A&AS, 44, 363
Worley, C. C., Jofré, P., Rendle, B., et al. 2020, A&A, 643, A83

1. Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
   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
4. INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi, 5, 50125 Firenze, Italy
   e-mail: laura.magrini@inaf.it
5. School of Physics & Astronomy, Monash University, Wellington Road, Clayton 3800, Victoria, Australia
6. Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
7. Observational Astrophysics, Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden
8. Instituto de Astrofísica de Canarias, Vía Láctea s/n, 38205 La Laguna, Tenerife, Spain
9. Departamento de Astrofísica, Universidad de La Laguna, 38205 La Laguna, Tenerife, Spain
10. Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, ul. Bartycka 18, 00-716, Warsaw, Poland
11. ROB – Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
12. INAF – Osservatorio di Astrofisica e Scienza dello Spazio, via P. Gobetti 93/3, 40129 Bologna, Italy
13. Institut d'Astronomie et d'Astrophysique, Université Libre de Bruxelles, CP 226, Boulevard du Triomphe, 1050 Bruxelles, Belgium
14. Division of Astrophysics, Department of Physics, Lund University, Box 43, 22100 Lund, Sweden
15. INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, 95123 Catania, Italy
16. INAF – Osservatorio Asronomico di Palermo, Piazza del Parlamento, 1 90134 Palermo, Italy
17. Department of Astronomy, Stockholm University, AlbaNova University Center, 106 91 Stockholm, Sweden
18. Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
19. Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark
20. GEPI, Observatoire de Paris, Université PSL, CNRS, 5 place Jules Janssen, 92190 Meudon, France
21. Departamento de Física de la Tierra y Astrofísica & IPARCOS-UCM (Instituto de Física de Partículas y del Cosmos de la UCM), Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 28040 Madrid, Spain
22. Lund Observatory, Department of Astronomy and Theoretical Physics, Box 43, 22100 Lund, Sweden
23. Institute of Theoretical Physics and Astronomy, Vilnius University, Sauletekio av. 3, 10257 Vilnius, Lithuania
24. Centro de Astrobiología (CSIC-INTA), Departamento de Astrofísica, campus ESAC. Camino bajo del castillo s/n, 28 692 Villanueva de la Cañada, Madrid, Spain
25. Department of Chemistry and Physics, Saint Mary's College, Notre Dame, IN 46556, USA
26. Astronomical Observatory, Institute of Theoretical Physics and Astronomy, Vilnius University, Sauletekio av. 3, 10257 Vilnius, Lithuania
27. Observatório Nacional – MCTI (ON), Rua Gal. José Cristino 77, São Cristóvão, 20921-400, Rio de Janeiro, Brazil

[28] Space Sciences, Technologies, and Astrophysics Research (STAR) Institute, Université de Liège, Quartier Agora, Bât B5c, Allée du 6 août, 19c, 4000 Liège, Belgium

[29] Astrophysics Group, Keele University, Keele, Staffordshire ST5 5BG, UK

[30] European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands

[31] Department of Physics and Astronomy, University of Padova, v. dell'Osservatorio 2, 35122 Padova, Italy

[32] INAF – Osservatorio Astronomico di Roma, Via Frascati 33, 00040 Monte Porzio Catone (Roma), Italy

[33] Nucleo Milenio ERIS & Instituto de Estudios Astrofisicos, Universidad Diego Portales, Ejército 441, Santiago, Chile

[34] INAF – Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy

[35] Departamento de Física Aplicada, Facultad de Ciencias, Universidad de Alicante, 03690 San Vicente del Raspeig, Alicante, Spain

Appendix A: Simplified flags

Table A.1: Simplified flags

Acronym	Meaning	Conditions for raising the flag
SNR	No or inaccurate results due to low SNR	This flag is raised if the S/N is lower than 50 and if the object has an incomplete set of parameters.
SRP	Spectral reduction problem	This flag is raised if there are no parameters or abundances
SDS	Some discarded spectra	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.
IPA	Incomplete parameters	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.
SSP	Some suspicious parameters	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$2).
NIA	No individual abundance (except Fe)	This flag is usually raised when there are too few available lines for abundance determinations (except Fe).
SSA	Some suspicious abundances	This flag can be raised for metallicity, e.g. when the Fe I and Fe II lines are discrepant, or for other elements. It is raised in cases of high v sini values or of SBn, n$\geq$2, or when the node was uncertain about this abundance.
PSC	Parameter space coverage	This flag is typically raised when the parameters are not within the model atmosphere grid parameters of the node, or are on the node's grid edge. Some abundances might then be missing.
SRV	Suspicious or unreliable radial velocity	This flag is raised in cases where the CCF was corrupted, or where the RV was discrepant between setups, or the object was identified as an SBn.
SRO	Suspicious rotational velocity	This flag is raised in cases where there is no rotational velocity determination, or an excessively high or revised rotational velocity. It is also raised in case of SBn, n$\geq$2.
BIN	Detected binary : SB1 or SBn$\geq$ 2.	
EML	Emission line: any line, not only Hα	