Analysis of stellar spectra with machine learning

Álvaro Segovia Otero

Lund Observatory
Lund University

2020-EXA159

Degree project of 60 higher education credits (for a degree of Master)
May 2020

Supervisor: Ross Church

Lund Observatory
Box 43
SE-221 00 Lund
Sweden

Abstract

Researchers in the field of Galactic Archaeology have entered the era of industrial revolution. Upcoming surveys are planning on observing tens of millions of stars and high precision and accuracy must be ensured when deriving their stellar parameters and elemental abundances. Unconventional data-driven techniques hold the promise of efficiently dealing with these vast collections of data while still rendering results of astrophysical value.

The Cannon is a supervised machine learning algorithm implemented to transfer stellar properties or *labels* from a dataset of reference to any desired collection of stars. In this thesis, *The Cannon* is trained on a set of synthetic spectra generated *ab initio* and applied to a sub-set of 1410 FGK-type stars from the *Gaia*-ESO Survey for a label space of high dimensionality ($T_{\rm eff}$, $\log g$, ξ, $v \sin i$ and 16 [X/H] abundances, where X is Mg, Na, Ca, Sc, Si, V, Ti, Mn, Fe, Ni, Cr, Co, Ba, Eu, O and Al). The aforementioned synthetic training set does not represent a grid of synthetic spectra or a sub-sample of stars with well studied properties. Instead, we have designed a sophisticated training set predominantly based on the Bensby catalogue of 714 stars with well measured stellar parameters and elemental abundances.

The Cannon is indeed very fast, taking an average time of 15 seconds to simultaneously fit 20 labels on one single spectrum after having trained on the model. It succeeds in recovering the $T_{\rm eff}$, $\log g$ and [Fe/H] stellar parameters with typical deviations of $\sigma_{\rm [Fe/H]} = 0.08$ dex, $\sigma_{\rm Teff} = 88$ K and $\sigma_{\log g} = 0.14$ dex in the label offsets with respect to the GES values, as well as determine 15 elemental abundances within a SNR range spanning from 10 to 300.

Acknowledgements

I would like to thank my supervisor Ross Church for his help throughout this year-and-a-half long project. I am especially grateful for the unconditional help of Gregor Traven and Dominic Ford in guiding me through the nitty-gritty of the pipeline. Special thanks to Massteo Prgomet, my office mate, with whom I pushed through the never ending study sessions, Rebecca Forsberg for sharing her knowledge about chemical abundances and the inhabitants of Winstrups Kollectiv for managing to distract me with board games and lovely meals. Finally, I would like to express my deepest gratitude to Anna Llambrich for her endless encouragement, support and for having dealt with me during these last two years of degree.

Popular science

Astronomers are time travellers and stars are their time capsules. Such a statement is particularly symbolic in the field of Galactic Archaeology, where the ultimate goal is to reconstruct the history of our galaxy by studying in great detail the light radiated from its stars.

The Milky Way, our galaxy, is a mixture of interacting gas and stars bound together by gravity, and so, its evolution goes hand in hand with the evolution of these two components. A key point here is the idea of chemical enrichment. Primordial gas is mainly composed of hydrogen and helium, but as stars form, heavier chemical elements are produced in their interiors and injected into the surrounding gas at death. This metal-rich gas will further clog up into giant clouds, collapse and cool down to produce a new generation of stars re-initiating the cycle. In addition to this, during its lifetime, the Milky Way has collided with other galaxies in its vicinity. This makes different enrichment channels even more complex to unravel, as gas and stars get violently mixed and scattered around. The result is a plethora of stellar populations with distinct dynamical and chemical imprints, some of which are preserved in stellar atmospheres and accessible by state-of-the-art observatories.

Our current understanding of the Milky Way classifies it as a spiral galaxy, with a distinguishable disk in which the Solar System is embedded. It is in this Galactic disk where most of the gas and stars are located, allowing us to pursue extensive and precise surveys of the properties of millions of these stars to disentangle the formation and evolution of the Milky Way. Among these observable properties we find the temperature of the stellar atmosphere, related to how bright its surface is, the surface gravity, which gives us an idea of its pressure structure, and the metallicity of the star, representing the abundance of heavier elements blocking the radiation emitted from the stellar atmosphere towards our telescope. These three main stellar parameters along with other element abundances fix the shape of stellar spectra, i.e. "stellar IDs" containing information about the chemistry and dynamics of stars.

At this point one might imagine the computational and human efforts required to analyse data from these large stellar surveys characterising millions of objects. Designed methodologies must therefore be effectively automatised. They need to ensure both efficiency when examining stellar spectra for such large numbers of observed objects, but also high accuracy in order to obtain precise enough results to distinguish among the various chemical enrichment channels. Another significant concern in the study of these massive catalogues are the theoretical assumptions made when modelling the transferred radiation through stellar atmospheres. Stellar properties change at different depths of the atmospheres in very convoluted ways. Only a simplified model of the real astrophysical phenomena can be constructed, sometimes in strong disagreement with one another even when the same stars are examined.

Here, an algorithm called *The Cannon* is implemented to cast some light upon these issues. *The Cannon* is a machine-learning code, meaning that it "learns" to relate spectra with stellar properties by optimizing purely mathematical functions within a dataset of reference and then applies such functions to any test set of interest. This is a data-driven approach as it does not contain any astrophysical assumptions and is extremely fast as the optimized mathematical expressions can be computationally cheap to calculate. In this work, we prove the scientific value of this method when substituting the reference set by artificially generated stellar spectra with known stellar properties and applied to a test set of 1410 stars observed with a high resolution instrument at the Very Large Telescope (VLT) of the European Southern Observatory (ESO, Chile).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The ultimate goal in Galactic Archaeology is to unveil the star formation history and chemical evolution of the Milky Way (MW) by studying the kinematics, ages and elemental abundances of stars in their current arrangement. The underlying idea is that analysing spectra of stellar populations with different parameters and abundances can probe different aspects of the Galaxy.

To reconstruct the chemical history of the MW using elemental abundances, several assumptions have to be taken into account. Chemical enrichment is a natural consequence of stellar evolution, hence it gradually increases with time. Hydrogen, helium and a few other light elements are produced in the Big Bang nucleosynthesis, but in general, elements heavier than hydrogen are synthesised inside stars via nuclear fusion (Burbidge et al. 1957). Exothermic reactions of the sort generate elements all the way up to ^{56}Fe and are injected into the interstellar medium (ISM) by winds and supernovae (SNe) at later stages of stellar evolution. New born stars are then born from this enriched medium, producing new elements that will be further spread into the Galaxy again in a repeated cycle. As a result, each star-forming region has a unique chemical imprint that can be retained by stellar atmospheres and passed along to future generations of stars (Freeman & Bland-Hawthorn 2002).

Such is the case of FGK type stars, whose lifetime in the main-sequence can exceed that of the age of the Milky Way. These stars are still burning hydrogen to helium in their cores and for this reason, few of them have begun climbing the Red Giant Branch (RGB), where they experience the first dredge-up mixing. At this stage shallow convective zones grow and cause substantial mixing of their material, erasing any preserved chemical fossils. Hence, dwarf stars conserve the chemical make-up of earlier epochs of the protogalactic cloud. However, identifying these patterns in stellar populations is a much more complex exercise. One that involves coupling large-scale with small scale astrophysics.

Disentangling the formation of the Milky Way is indeed anchored to the present-day Λ-Cold Dark Matter cosmological paradigm (ΛCDM). In this picture, baryons form galaxies

inside CDM over-densities that collapse under the force of gravity into dark matter halos. Structure grows and continues evolving driven by gravitational interactions to create the vast cosmic web observed nowadays. A MW-like galaxy has hierarchically aggregated smaller objects along the way, impacting the chemical and thermodynamical state of the gas, dynamics of its stars and even modifying its size and mass which in turn influence the star formation rate and its ability to keep the enriched material bound.

Events at large scales are therefore connected to small scale phenomena and vice versa. For example, the environment and chemistry of dense molecular clouds regulate the stellar mass of new born stars, which might end up as SNe enriching the ISM. Yet what sets the span in resulting stellar masses, thus lifetimes, or which specific physical processes are favoured leading to relevant nucleosynthesis channels is still a mystery. Furthermore, apart from being subject to a galactic potential, stars suffer from radial and vertical migration with respect to their cloud of birth. This is due to the presence of permanent sub-structures like the bar, transient density waves like spiral arms and gravitational encounters by giant molecular clouds all of which exert a torque on stars and modify their angular momentum. If the internal distribution of angular momentum is not conserved, much of the dynamical information, and the fossil chemical gradients are diluted irreversibly (Freeman & Bland-Hawthorn 2002).

In order to find preserved signatures in such a plethora of stellar populations it is imperative to ensure high accuracy and precision in the calculation of abundances, stellar parameters, distances, positions and proper motions for a statistically significant sample of stars. Only then can constraints be set on the various chemo-dynamical models. The measurement of stellar ages also plays an important role now accessible by asteroseismic surveys, but via isochrone fitting or the [C/N] abundance ratio in giants too. All in all, for a one to one mapping of the stars to their specific formation sites in the protocloud of the Galaxy, at least the *when* (age), *where* (positions in the sky and distances) and *how* (abundances and dynamics) are required.

In this sense, the field has experienced a major improvement with the release of *Gaia* DR2 (Gaia Collaboration et al. 2018) and further follow-up spectroscopic surveys. Studies can now complement the five-parameter astrometric solution of unprecedented accuracy with high resolution ground-based observations providing elemental abundances and radial velocities. That is the case of recent publications by Helmi et al. (2017), Hayes et al. (2018) and Haywood et al. (2018), who were able to identify metal-poor halo stars in retrograde orbits with distinct patterns in chemical and dynamical space. These stellar populations carry potential information on the mass and time of the merger events or "building blocks" interacting with the Galaxy at its early epochs, i.e. the *Gaia*-Enceladus or Sequoia accretion events.

The prospect has utterly changed within the last decade where the lack of good measurements is no longer that big of an issue. Nevertheless, piecing together chemo-dynamical

substructures into one common framework remains challenging. Large spectroscopic surveys indeed contain a huge amount of information, yet they do not come without tremendous effort and time invested in their planning (Jofré et al. 2019; Ruchti et al. 2016; Lindegren & Feltzing 2013). One of these surveys is known as the *Gaia*-ESO Public Spectroscopic Survey (GES, Randich et al. 2013; Gilmore et al. 2012) and was put together with the aim of providing one of the first homogeneous catalogues of approximately a hundred thousand stars with known kinematics and elemental abundances.

Galactic Archaeology has thereupon entered an era of industrial scales. Large spectroscopic surveys are regarded as the best ready-to-use products but also face formidable data reduction and spectral modelling challenges. Considering the great number of stellar parameters and abundances that have to be derived, pipelines have been developed along with their corresponding surveys. They are in charge of automatising the most redundant steps in the process, e.g. background subtraction, barycentre correction, quality selection, etc. Nevertheless, more efficient pipelines are still required for upcoming surveys, which hold the promise of characterising tens of millions of stars and therefore need to drastically reduce the number of synthetic spectra used for spectroscopic analyses (Ness et al. 2018; Ting et al. 2017). Moreover, many spectroscopists have their own automatised pipeline tailored on their more traditional experience and optimised to perform in a particular spectral region, with a custom-made wavelength mask and for certain types of stars. Unsurprisingly, different research groups can obtain contradictory results for the same stars stemming from the choice of different assumptions and methodologies. Among most of the known issues one finds: 1D, LTE model atmospheres, incomplete atomic data and restricted spectral grids extremely sensitive to the *signal-to-noise ratio* (SNR) of the data.

An alternative to these physics-based pipelines is presented here. *The Cannon* (Ness et al. 2015) is a type of supervised machine learning algorithm that analyses stellar spectra without any input of astrophysically motivated spectral model. It relies, however, on the existence of a reference dataset with well-known stellar attributes from which information is transferred to a second dataset by optimising a purely mathematical model. In doing so, it is exceptionally fast while still meeting the uncertainties in stellar properties quoted in physics-based approaches. *The Cannon* differs from other data-driven techniques because it contains an intrinsic probabilistic spectral model with its own noise prescription (Casey et al. 2016). This provides a universal scale to, among others, consistently cross-calibrate spectroscopic surveys with different wavelength coverages, SNRs and resolving powers given a sub-set of common stars (Ho et al. 2017).

In this thesis, both stellar parameters and elemental abundances are computed for 1410 GES stars, although the approach has been adjusted slightly. Here, the reference set in *The Cannon* is replaced by a set of spectra synthesised with *Turbospectrum* (Plez 2012) and based on trustworthy stellar parameters and abundances derived in modern publications (Bensby & Feltzing 2006; Bensby et al. 2014; Battistini & Bensby 2015, 2016; Delgado Mena et al. 2017; Mishenina et al. 2019; Casali et al. 2019).

The thesis is structured as follows:

- In Chapter 2 we briefly lay out our current understanding of the structure of the Milky Way and what can we learn from it by studying the stellar parameters and elemental abundances of stars in the solar neighbourhood.
- In Chapter 3 we summarize how our pipeline works and inspect some preliminary results on a sub-set of 39 GES stars.
- In Chapter 4 we discuss the results after validating the pipeline and applying it to the full 1410 GES set of stars.
- In Chapter 5 we sum up the whole thesis and highlight the main conclusions of the project.

Chapter 2

Scientific background

This chapter deals in more depth with the Milky Way in the context of Galactic Archaeology. Its current, coupled, multi-component structure provides fossil information of its assembly history and evolution. Particularly helpful is understanding the mechanisms that influenced the formation of the stellar disk, as it appears to be ubiquitous in the local Universe and the major contributor of baryons in the Galaxy. Investigations addressing other constituents like the halo and the outermost bulge are favoured by our location in the MW, i.e. the thinness of the galactic disk. Deprojecting the latter, on the other hand, is fraught with uncertainty because of interstellar extinction and source confusion. Despite these disadvantages, being immersed in the galactic disk implies granted access to an extensive sample of stars in the solar neighbourhood that can be examined with unparalleled detail. The upcoming sections will follow the scheme listed below:

- Sketch of the different parts of the Milky Way that are observationally distinguishable: the thick disk, the thin disk, the *in-situ* halo and the accreted halo. One can probe this canonical picture using chemistry, kinematics/dynamics and the spatial distribution of stars in the Galaxy.
- These observations favour solar neighbourhood stars because of their proximity, but roughly, two classes of stars are found: dwarfs and giants. The next section is aimed towards understanding why it is more convenient for us to use dwarfs.
- After agreeing over the objects of interest targeted in this project, a discussion about the main stellar parameters needed to catalogue stellar spectra is conducted.
- Finally, a brief review of five hand-picked elements will be given in order to understand how we can interpret abundance plots.

2.1 The canonical Milky Way

Within the classical picture of the Milky Way, four constituents are identified which probably formed at different stages of its evolution: the thin disk, the thick disk, the bulge and the halo. We are of course ignoring the dark matter halo for the purpose of this project. In this work we have a more observational perspective of the field of Galactic Archaeology and therefore we tend to define these Galactic components in three phase spaces: spatially, kinematically and chemically.

It is an ongoing debate whether all four regions can be traced as independent entities or belong to a continuous chain of events within the context of an evolving galaxy. Indeed, none of these existing frameworks allows to uniquely separate all substructures as there is significant overlap among the distribution functions in the respective phase spaces. With that, a quest to find the best tracer that distinguished among them began, pursuing high resolution studies of a wide range of stellar populations (APOGEE, Majewski et al. 2017, GALAH, De Silva et al. 2015, GES, Smiljanic et al. 2014, 4MOST, de Jong et al. 2014).

Recent publications involving these vast surveys identify, broadly speaking, four separate regions with unique chemistry and kinematics (Bland-Hawthorn & Gerhard 2016; Freeman & Bland-Hawthorn 2002). Figure 2.1 pieces these four regions together in a Toomre diagram (LHS) and an $[\alpha/\mathrm{Fe}] - [\mathrm{Fe/H}]$ abundance plot (RHS). The power of the Toomre diagram is based on the assumption that each Galactic structure has increasing velocity vectors. Regarding the abundance of α elements, its conclusive power is built on the idea that since it is a relatively easy parameter to measure, there is direct access to the hopefully frozen-in-time properties of the Galaxy.

The first and most prominent in terms of number of stars is the Galactic thin disk. The classical picture of the thin disk explains that it emerged around $z \sim 1$ (Mo et al. 1998), when most of the baryons dissipated into the midplane of the Galaxy and ended up distributed in a decreasing exponential profile with small vertical scale height of 300 pc and large radial scale length around 15 kpc (Gilmore & Reid 1983; Bovy & Rix 2013). Thin disk stars are younger on average, and located within the metallicity range of $-0.7 < [\mathrm{Fe/H}] < 1$. If placed in the Toomre diagram they appear kinematically cold, with weak radial and vertical components resulting in a total velocity of below 70 km s^{-1} with respect to the LSR. They practically describe near-circular orbits and co-rotate with the disk of the Galaxy with low velocity dispersion. The thin disk sequence is termed as low-α in modern literature due to the fact that its stars have lower α-abundances over a wide range of metallicities (RHS of Figure 2.1).

Closely related to the low-α sequence is the high-α sequence, more traditionally referred to as thick disk (Bensby et al. 2014). Despite its unclear formation channel (merger-driven heating of the thin disk, secular heating, radial migration, etc. see Bovy & Rix 2013), the scientific community acknowledges the existence of a thick disk with larger scale height ($\sim$

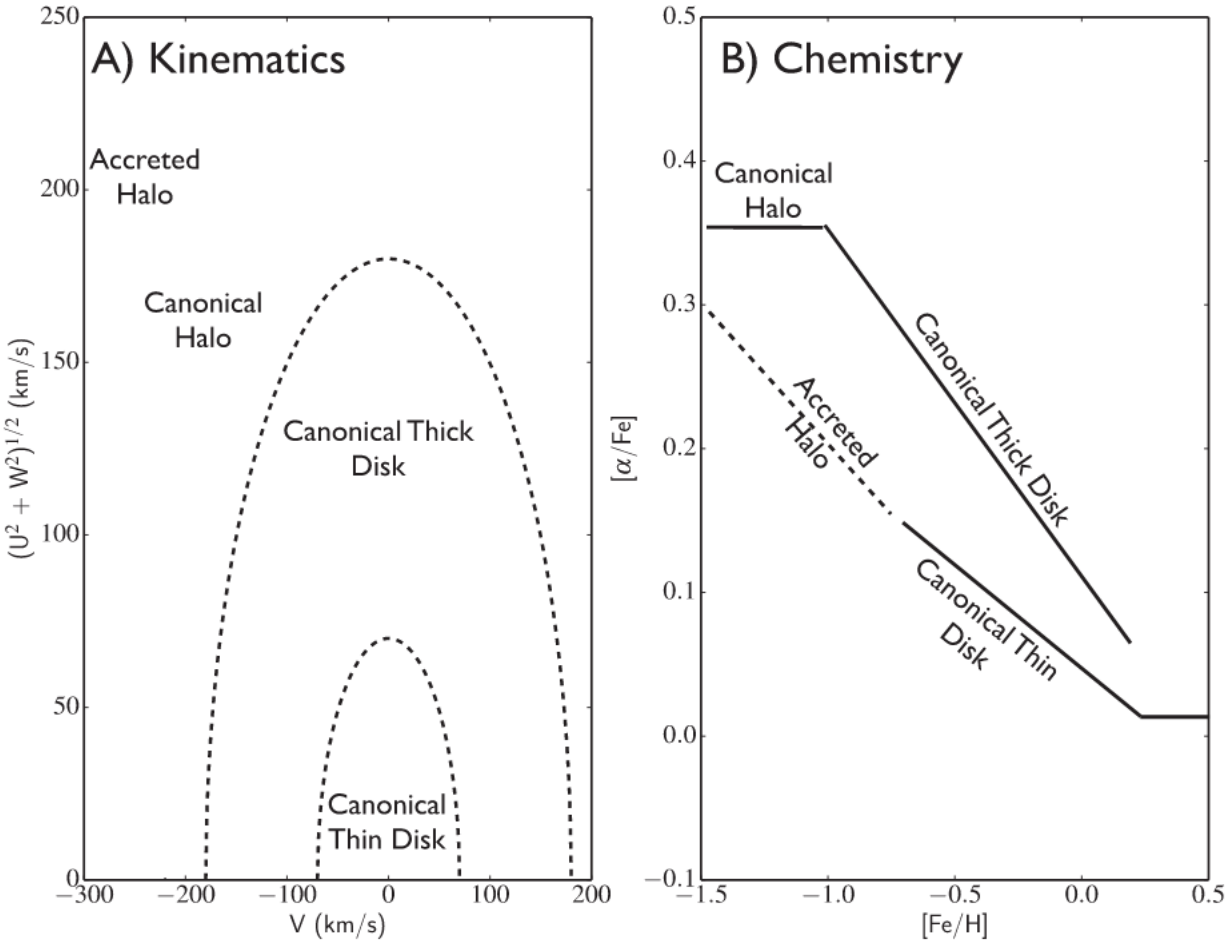


Figure 2.1: Canonical overview of the Milky Way and its main components in (a) a Toomre diagram and (b) [α/Fe]-[Fe/H] plane. (a) This schematic plot shows how stars are arranged according to their combined vertical (W) and radial (U) velocities against their rotational velocity (V) with respect to the LSR. The dotted lines represent lines of constant total velocity. (b) Sketch of the [α/Fe] − [Fe/H] space. Credit: Hawkins et al. (2015).

1 kpc). Early photometric studies supported this claim further suggesting that the thick disk is less dense, with only 12% of the surface brightness of its thinner sibling, hosting older stellar populations with a shorter scale length. Thick disk stars are more metal poor on average, α-enhanced and are governed by hotter kinematics (between orbital velocities of $100 < V < 200$ km s^{-1}).

Finally, we have the halo stars, sparsely spread at predominantly low metallicities ([Fe/H] ≤ -1, RHS of Figure 2.1). Stellar populations in the halo are split in two (Nissen & Schuster 2010). Metal-poor high-α stars are thought to have dissipatively collapsed early on along with the rest of the Galactic components. The metal-poor low-α sequence is associated to stars born in an environment separate from the halo and perhaps incorporated later on.

Both in-situ and ex-situ populations stand out as having the hottest kinematics (above 200 km s^{-1}, LHS of Figure 2.1), with out-weighting U and W velocity components.

It must be mentioned that even though this is the age of data-driven techniques and extensive stellar surveys, our homogeneous characterization of stellar kinematics and chemistry targeting various patches of sky and different wave bands is still biased towards solar neighbourhood stars. It is by closely observing them that the scientific community has reached the conclusions above. Challenges to this general description are mainly caused by perturbing the original distribution functions leading to significant overlap in space, kinematics, and chemistry (Bovy & Rix 2013; Minchev et al. 2015). Specially in the $-1 < [\mathrm{Fe/H}] < 0.5$ metallicity window, where halo, thin and thick disk stars are present. Bensby et al. (2014) find that the low velocity tail of thick disk stars blends with the high velocity tail of thin disk stars. To convincingly split their sample between thin and thick disk stars, a combination of stellar ages (Haywood et al. 2018) and α abundances (Adibekyan et al. 2012; Reddy et al. 2006) is applied. Still, there are old stars that are kinematically cold and α-enhanced as well as young, kinematically hot, α-depleted ones. The metal-poor end of the $[\alpha/\mathrm{Fe}] - [\mathrm{Fe/H}]$ diagram is also a good example of mixture between the disk population smoothly transitioning to the halo population of stars. Even if there are reasons to believe that low metallicity, α-rich stars are born in-situ and are responsible for the early chemical enrichment of the protogalactic thick disk (Hayes et al. 2018), they could in fact be relics of the last major merger with the Milky Way, the *Gaia*-Enceladus event (Haywood et al. 2018; Koppelman et al. 2019; Myeong et al. 2019).

2.2 Probing the Milky Way with FGK type stars

The stellar spectroscopy community has often preferred targeting FGK-type stars to tackle the questions posed in Galactic Archaeology. The reader must be warned that for the purpose of this work, FGK stars will also be referred to as dwarfs or main-sequence stars (though catalogued K stars can also be giants) and used indistinctly throughout. But this has not always been the case as giants offer many advantages. Jönsson et al. (2017a,b), Lomaeva et al. (2019) and Forsberg et al. (2019) perform high resolution spectroscopy on a large sample of bulge giants with a broad wavelength coverage in the optical. Giants are brighter and therefore can be observed at higher SNRs for the same resolving power and exposure times compared to dwarf stars. They are also less affected by attenuation, so they open up the window towards dusty environments closer to the Galactic centre.

Spectra of giant stars are hard to deal with though. In what follows, a comparison between these two sorts of stars will be made, emphasising on the convenience of using solar-type dwarfs.

- FGK type stars have larger effective temperatures, unbinding molecules in their atmosphere. Molecular transitions are problematic in stellar spectra because they increase

the number of lines and create prominent absorption bands. For particularly cool, metal-rich giants, gathering a proper selection of absorption lines gets complicated avoiding both blending and a proper selection of the continuum (even with high resolution spectrographs). Moreover, a growing number of molecular lines favours spectral lines to be blended with lines with unknown or uncertain atomic data.

- Unlike giants, dwarfs are located in the main-sequence of the Hertzsprung - Russell diagram (HRD), still burning hydrogen to helium in their cores surrounded by a shallow convective zone. Consequently they have not experienced heavy mixing of the inner material processed by thermonuclear burning and the surface material, and thus faithfully preserve the chemical composition of their birth environments.

- Dwarf stars are characterised by a higher surface gravity ($\log g > 3.5$ dex), reinforcing the continuum opacity and therefore decreasing the strength of the absorption line at stake. This keeps weak lines in their linear regime and prevents some deeper lines from saturating. Saturation is something spectroscopists avoid because it complicates the calculation of the chemical abundances from the profile of absorption lines. In terms of equivalent widths, the curve of growth mapping is less sensitive to abundance values. In the context of synthetic spectra, saturation increases the non-linear behaviour of spectra, making more difficult to model.

- Elemental abundance determination for dwarf stars is differential with respect to the Sun to reduce systematic errors. Either spectral lines can be compared to the solar spectrum or one can rely on the accuracy with which abundances and stellar parameters are determined as well as the completeness of the line list.

- Unfortunately, dwarfs are dimmer than giants. This requires longer exposure times to reach a high enough SNR for high resolution spectrographs and limits observations to stars in the vicinity of the Sun.

- If filtered suitably, dwarfs can be placed on the turn-off of the HRD. This, together with a reliable stellar parameters, enables to derive the age of stars via isochrone fitting. For giants, spectroscopists rely on the [C/N] abundance ratio if there are no available asteroseismic measurements.

2.3 Stellar parameters

In this section an explanation of the relevant stellar parameters for our analysis and how they affect absorption lines is given. Unless specified otherwise, all of the succeeding section applies to weak metal lines in FGK-type stars. That is, the strength of a weak metal line is proportional to the ratio between the line and continuous opacity if the temperature is more or less constant with optical depth for a given layer of the stellar atmosphere. In more physical terms, the strength of a line depends on its absorption coefficient and the

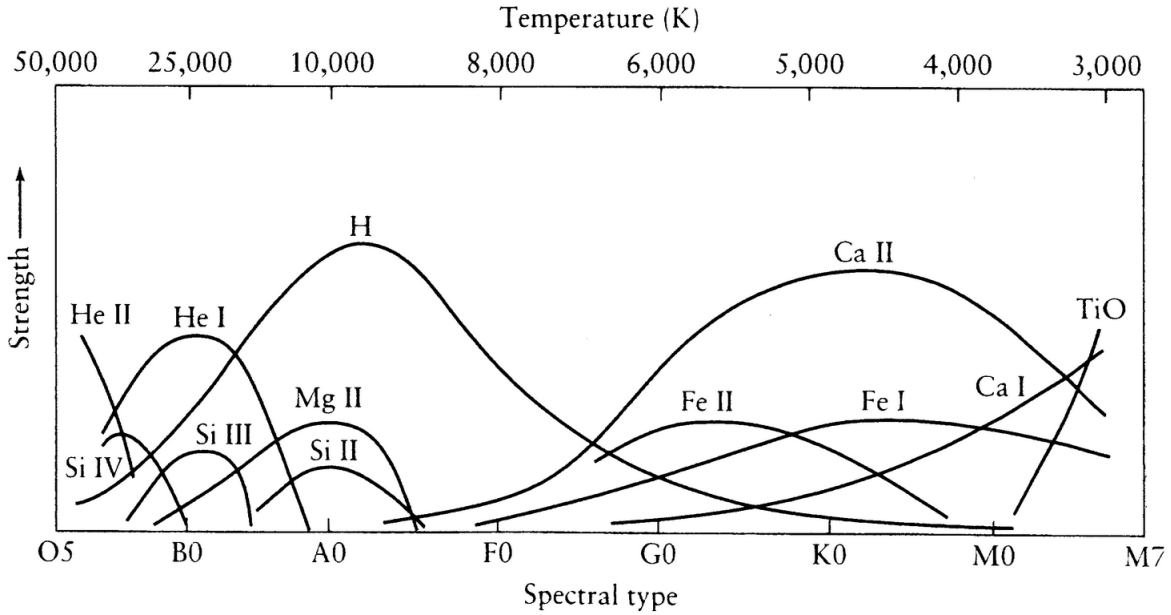


Figure 2.2: Strength of lines as a function of effective temperature and spectral type for a few chemical species. Credit Kaler (2011).

number of absorbers derived from excitation and ionisation equations, themselves dependent on temperature and gas and electron pressure.

In short, the relation between line strength or *equivalent width* (EW) of a line and abundances depends on the stellar parameters.

Effective temperature

The effective temperature of a star, $T_{\rm eff}$, is defined as the temperature of a blackbody radiatior that would produce the same luminosity as the star does. If stars are considered blackbody radiators, their luminosity $L_\star$ is related to their effective temperature by the Stefan-Boltzmann equation:

$$T_{\rm eff} = \left(\frac{L_\star}{4\pi\sigma_{\rm SB} R_\star^2} \right)^{1/4} \tag{2.1}$$

where $T_{\rm eff}$ is the effective temperature at the surface of the star or, for the sake of argument, in the stellar photosphere, $R_\star$ is its radius and $\sigma_{\rm SB}$ the Stefan-Boltzmann constant.

The effective temperature heavily controls the placement of the continuum and the strength of absorption lines from the exponential and power dependencies in the radiative transfer equations regulating the ionisation and excitation equilibrium as well as the continuum opacity.

An increase of EW goes hand in hand with that of temperature because of an increase in excitation. After reaching its maximum in strength (Figure 2.2), the line weakens as the continuous opacity of the negative hydrogen ion (H^-) starts increasing due to high electron pressures. A decrease in EW can also be due to the frequent ionisation of absorbers at high temperatures, specially in the optical where the majority of lines come from neutral species. Strong lines like the Na D lines depend on the collisional damping constant, which is a function of the temperature, but are also affected by NLTE effects. The wings of hydrogen lines are affected by the Stark effect (Gray 2005).

Measurement uncertainties in $T_{\rm eff}$ are typically found to be around 200 - 300 K. If different methods are available, the errors can be brought down to 50 K (Jofré et al. 2019). It is also reported that a change of 100 K affects abundances obtained from neutral lines up to 0.1 dex, yet those derived from ionised lines change very little (Jofré et al. 2019).

There are multiple approaches to obtain $T_{\rm eff}$ such as the infrared flux method, excitation balance using Fe I lines or by fitting the wings of neutral hydrogen lines in the Balmer series.

Surface gravity

The surface gravity, commonly found in the literature in logarithmic scale as $\log g$, is the Newtonian gravitational acceleration of the star at its surface assuming it is perfectly spherical,

$$g = \mathrm{G}\frac{\mathrm{M}_\star}{\mathrm{R}_\star^2} \tag{2.2}$$

where $M_\star$ is the mass enclosed by $R_\star$, the stellar radius, and G is the gravitational constant. Since stars are astrophysical systems in equilibrium, the gravitational force that binds the stellar atmosphere with the star is balanced out by the pressure gradient and therefore $\log g$ is directly related to the pressure structure of the stellar atmosphere.

Spectra are not as $\log g$-sensitive as they are $T_{\rm eff}$-sensitive, so it is a more challenging parameter to measure spectroscopically. There are three ways to visualise pressure effects. The first comes from the abundance of absorbers with respect to the continuous opacity. The second is via the damping coefficient of strong lines and the third is due to pressure dependencies of the Stark effect. For the purpose of this thesis, only the first one is explained. More information can be found in Gray (2005). The key factor here is the ionisation state of the elements involved. Not only does the number of absorbers increase with pressure, but also the number of free electrons provided by the ionised species further affecting the continuous opacity (H^- free-free transitions). If ion or neutral atom lines are in the same or higher ionisation levels compared to the majority of the elements of the same species, their lines strengthen when the pressure, thus continuous opacity, lowers down.

Reported errors of $\log g$ do not get better than 0.1 dex, though asteroseismic surveys might improve these uncertainties by a factor of 10 (Jofré et al. 2019). As expected, surface gravity influences the chemical abundances derived from weak ionised lines more than it does neutral lines.

There are multiple approaches to obtain $\log g$ such as the ionisation balance of the Fe II to Fe I line ratio, through the wings of strong lines and with asteroseismology.

Metallicity

Metallicity is the third stellar parameter even though it just represents the chemical abundance ratio between iron and hydrogen scaled with respect to the Sun:

$$[\mathrm{Fe/H}] = \log\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\star} - \log\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\odot} \tag{2.3}$$

where $N_{\mathrm{Fe,H}}$ is the number density of either iron or hydrogen.

Iron is used as a proxy to represent metals in general, as the number of lines present in the optical are much more abundant than any other element. However, as we will see later on, abundances of other elements do not necessarily have to correlate with Fe.

Metallicity is so important because it shapes the overall structure of the stellar atmosphere, influencing the strength of absorption lines and continuous opacities. High metallicity implies larger abundances of metallic electron donors resulting in a change in the continuous absorption through H^- (Jofré et al. 2019; Gray 2005).

Microturbulence

It is considered the fourth stellar parameter as it directly affects radiative transfer in stellar atmospheres and therefore elemental abundance determination. Denoted in the literature as v_{mic} or ξ, microturbulence is an empirical stellar parameter that refers to turbulent motion, where the dimensions of the eddies are smaller than the local mean free path of the photons in the atmosphere.

Microturbulence is the most notable source of line broadening in one dimensional spectral modelling apart from the thermal Doppler shifts. Both phenomena are analogous in nature, so it is, effectively, a calibration parameter needed to correct for any excess broadening with a larger effect the stronger the line is (Jofré et al. 2019; Gray 2005).

Others

Two more stellar parameters or broadening terms are worth mentioning in this thesis. They are the macroturbulence (v_{mac}) and the projected rotational velocity ($v \sin i$).

Likewise, v_{mac} describes velocity fields in stellar atmospheres. Yet the turbulence scales are now large enough to retain the same photon from its creation to its escape from the star. This parameter wants to account for the simultaneous radial and tangential motions of macro-cells of gas, but it is very complicated to disentangle it from the effect of stellar rotation directly from observed spectra. Henceforth, it is common practice to set both parameters to zero and determine a global broadening term which encompasses them (Jofré et al. 2019; Gray 2005).

2.4 Chemical abundances

In this section I will divide the periodic table in five groups: light, α, iron-peak, s-process and r-process elements. The reason for this classification goes back to the diverse nucleosynthesis channels and observational features in abundance space. A general picture will be sketched, paying special attention to the five representative elements chosen in Figure 2.3.

2.4.1 Light elements

Ignoring hydrogen and helium, the most abundant elements in the Universe can be traced all the way back to the Big Bang nucleosynthesis (e.g. Li, Be, B), but can in fact be produced by other means such as cosmic ray spallation, dying low mass stars and exploding massive stars (e.g. C, Na, Al). The element we chose to be a representative light elements is sodium.

Sodium, Na

Na (first panel of Figure 2.3) is a light odd-Z element mainly produced in the carbon-burning process. It then gets injected into the ISM predominantly via core-collapse SNe or SNe Type II (SNII), but also Asymptotic Giant Branch (AGB) stars (Clayton 2003).

As opposed to Al, another odd-Z element which depicts α-like behaviour in the [X/Fe]-[Fe/H] plane, the [Na/Fe] has an upwards trend at super-solar [Fe/H]. It then takes solar values at solar metallicities right before slowly increasing at sub-solar metallicities with a maximum of [Na/Fe] ≈ 0.15 dex at [Fe/H] ≈ -0.6 dex (Adibekyan et al. 2012; Bensby et al. 2014; Reddy et al. 2006). Metal-poor stars after the maximum are depleted in [Na/Fe]. The same trends are found after correcting for NLTE effects.

From Figure 2.3 we observe that a pure chemical analysis shows no hints of the thin-thick disk dichotomy whatsoever.

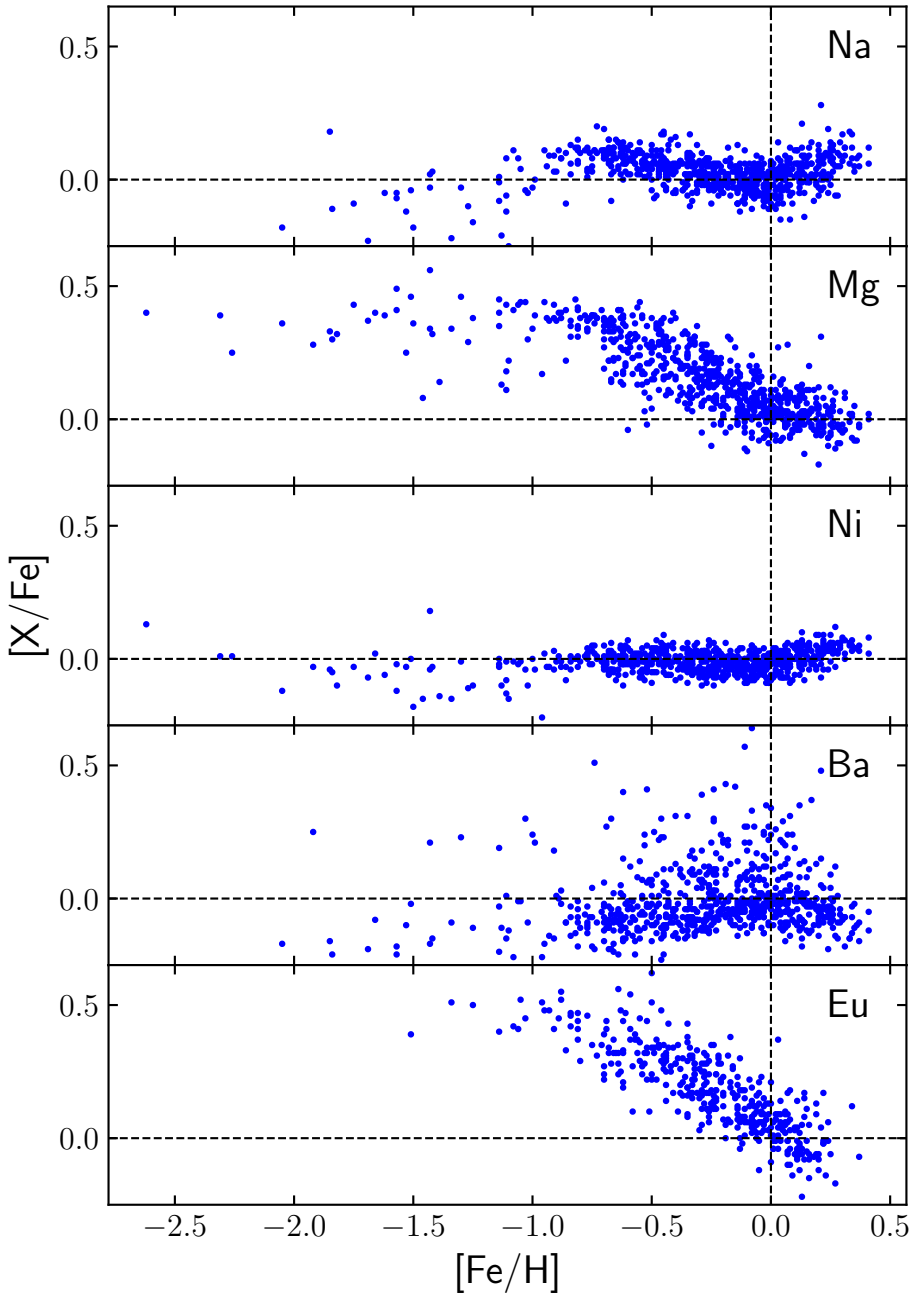


Figure 2.3: [X/Fe]-[Fe/H] plot for the five reference elements. From top to bottom, the catalogues providing with the data are: Bensby et al. (2014) for Na, Mg, Ni and Ba and Battistini & Bensby (2016) for Eu. The horizontal and vertical dashed lines indicate solar values of metallicity and [X/Fe].

2.4.2 α-elements

The α-elements are those elements whose nuclei are composed of integer multiples of helium nuclei, also known as α-particles, e.g. O, Mg, Si, Ca, Ti. Observationally these elements share a similar behaviour, viewed in Figure 2.4. A plateau extends from very low metallicities up to [Fe/H] $\sim$ -0.5 dex, where a gradual decrease or *knee* towards solar values at solar metallicities is seen.

The initial flat trend depends on the Initial Mass Function (IMF), which sets the number of massive stars that will undergo SNII. They are the predominant synthesis channel for α-elements and their progenitors are short lived (1-10 Myr). As time goes by modulated by the star formation rate and efficiency of the host galaxy, the aforementioned knee kicks in with an earlier or more delayed onset of SNe Type Ia (SNIa), which have a lifetime of a few Gyr and generate mostly iron peak elements, decreasing the [α/Fe] abundance.

The α-trait in Figure 2.4 is better appreciated in the very close solar neighbourhood, at higher temperatures and in the metallicity bin -0.7 dex $\leq$ [Fe/H] ≤ -0.35 dex (Fuhrmann 2011; Reddy et al. 2006; Adibekyan et al. 2012; Bensby et al. 2014). The potential gap suggested by Figure 2.4, however, can be easily washed out by uncertainties. Our reference light element is magnesium.

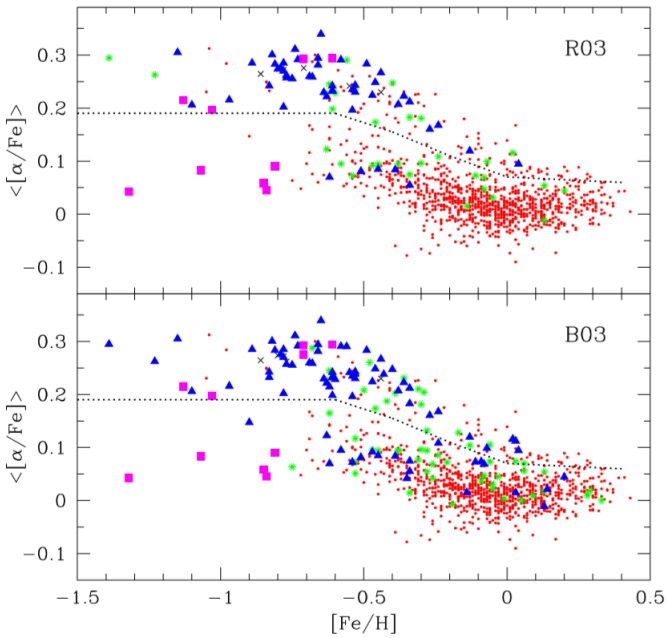


Figure 2.4: Abundance plot of [α/Fe] versus [Fe/H]. Top and bottom subplots correspond to different underlying kinematic criteria used by Robin et al. (2003) and Bensby et al. (2003) respectively. Blue triangles stand for thick disk stars, red circles for thin disk, magenta squares belong to the halo and green asterisks and black crosses to the thin-thick disk and thick disk-halo transitions. Credit: Adibekyan et al. (2012).

Magnesium, Mg

Mg is a primary alpha element also synthesised in the carbon-burning process of massive stars ending their lives as a SNII (Clayton 2003).

As a well behaved α element, [Mg/Fe] is reasonably constant up to [Fe/H] ≈ -0.5 dex. After the knee, there is a general downwards trend to almost solar, continuing practically flat at [Fe/H] > 0.

Although high [α/Fe] and low [α/Fe] stars are potentially identified as two split populations, no strict chemical evidence hints towards a separation in the Galactic disk by looking at Figure 2.3 (Adibekyan et al. 2012; Bensby et al. 2014; Reddy et al. 2006). Moreover, Adibekyan et al. (2012) and Bensby et al. (2014) claim they find a group of High α Metal Rich stars dubbed as HAMR around [α/Fe] ≈ 0.17 dex and after a gap at a metallicity of [Fe/H] ≈ -0.2 dex. On average, they are old stars, but have thin disk kinematics. No clear gap in metallicity is observed in the second panel of Figure 2.3 either.

2.4.3 Iron-peak elements

Iron-peak elements are thought to follow Fe in lock-step, however, their production sites are often different than Fe. On that note, they display complex abundance trends evoking numerous channels of enrichment, not necessarily produced by SNIa (Battistini & Bensby 2015). Elements like Sc, V, Cr, Mn, Co, Ni are iron-peak, among which Ni is plotted in the third panel of Figure 2.3.

Nickel, Ni

This element is produced in neutron rich environments like SNIa, i.e. via explosive nuclear burning (Clayton 2003).

Ni goes hand in hand with Fe, though the most exciting feature about the [Ni/Fe] against metallicity trend is its similarity to [Na/Fe] (with comparatively fainter trends). Both Na and Ni are controlled by the excess of neutrons in the cores of stars, so it is not surprising to find a correlation in the [Na/Fe]-[Ni/Fe] plane. In fact, this ratio has been a decisive tracer to distinguish between high-α and low-α stars in the halo (total space velocity larger than 180 km s^{-1}, Bensby et al. 2014; Nissen & Schuster 2010).

Other than that, and by strict inspection of the Ni panel in Figure 2.3, Ni appears unperturbed by the thin and thick disk division (Reddy et al. 2006).

2.4.4 Neutron capture elements

Nuclear fusion reactions are energetically favorable up to ^{56}Fe. Heavier elements (A > 56) form from neutron capture reactions (Burbidge et al. 1957), regulated by the density of

neutrons in the surrounding environments and the cross-section of the isotopes at stake. Therefore, determining the abundance of these sort of elements is crucial for understanding their source of neutrons.

Depending on the stability of the isotope, the neutron capture rate can be compared to β^--decay rates inside atomic nuclei,

$$n \rightarrow p + e^- + \nu_e \tag{2.4}$$

where a neutron n is captured by a nucleus and turned into a proton p by emitting an electron neutrino and a β-particle (electron in the case of β^--decay). Thus, an interplay between neutron capture and β^--decay reactions is set and dependent on the neutron budget, absorption cross-section and stability of nuclei. At this point two main sub-processes have to be defined: the s-process, where the isotope β^--decays due to *slow* neutron capture rates; the r-process characterized by a *rapid* build up of neutrons with respect to decaying reactions. No clear distinction between slow and rapid neutron capture processes can be made as both mechanisms contribute to the overall abundance of heavier elements. The last two panels in Figure 2.3 represent barium and europium, two neutron-capture elements primarily produced by the s- and r-processes respectively:

s-process: Barium, Ba

Produced by low- and intermediate-mass stars in the AGB phase of stellar evolution. It dominates for lighter elements above mass number $A > 60$. The magic numbers N = 50, 82, 126 cause a pile up in the abundance of stable nuclei at A = 90, 138, 208 (elements in the valley of stability) as their neutron capture cross-sections are small, bottlenecking the production of heavier elements.

Ba is a neutron capture element 83% of which is produced by s-processes (Delgado Mena et al. 2017 and references therein), thus it is often used as reference in abundance determination of neutron capture elements. From the Ba panel in Figure 2.3, an overall under-abundance throughout [Fe/H] is observed. Perhaps a gradual rising trend with metallicity towards solar values is appreciated, only to fall very smoothly at higher metallicities (Bensby et al. 2014). Delgado Mena et al. (2017) find more pronounced features, with steeper declines at super-solar [Fe/H]. The scatter at large [Ba/Fe] abundances is concentrated around [Fe/H] ≈ 0. These stars have generally hotter $T_{\rm eff}$, and since Ba is known to be sensitive to NLTE corrections, Bensby et al. (2014) claim the abundances are overestimated. Delgado Mena et al. (2017) on the other hand, include these stars in their analysis as they are well mixed with the rest of stars in the sample.

r-process: Europium, Eu

The r-process is the predominant channel for elements heavier than $A > 130$ as long as there is a fast enough intake of neutrons that keep the decay timescales large. Owing to the

discovery of Eu-abundant, metal-poor stars (Sneden et al. 2008) and recent detections of compact mergers with their electromagnetic counterparts (Rosswog et al. 2018), the study of r-process elements has flourished.

The bottom panel in Figure 2.3 suggests that [Eu/Fe] might be connected to SNII explosions as it clearly shows α-like behaviour, i.e. it quickly increases from solar values up to [Eu/Fe] ≈ 0.5 dex at [Fe/H] ≈ -0.8 dex. For this reason, Eu was thought to be produced by the same nucleosynthesis channels as α elements (Battistini & Bensby 2016; Delgado Mena et al. 2017). However, recent work by Rosswog et al. (2018) matching the electromagentic follow-up of the GW170817 neutron star merger to r-process yields indicated that if not dominant, these events are a major contributor of rapid neutron capture elements.

Furthermore, Eu is the r-process prototype element as only 7% of it is synthesised by slow neutron capture. Thus, the [Eu/Ba] abundance ratio can be used as a diagnostic of the neutron-capture process (Battistini & Bensby 2016).

Chapter 3

Methodology

This project explores the applicability of a machine learning technique to determine stellar labels for Galactic Archaeology studies. The fundamental interest in this field resides in deriving elemental abundances from stellar spectra and linking them to a particular nucleosynthesis channel. However, this is a process dependent on the coupling between large- and small-scale physics, difficult to unravel unless accurate and precise measurements of stellar labels are provided. One way of ensuring high quality investigations is by obtaining high resolution and high SNR spectra, more easily achieved with a reduced number of stars following a more specific selection function. Here, we implement our data-driven algorithm, *The Cannon*, in the context of large spectroscopic surveys. It presents itself as a solution that provides results that meet the quality standards even with spectra of lower resolution or SNR. At the same time it has also proven to be extremely efficient when dealing with datasets of the size of these spectroscopic surveys. In a nutshell this algorithm transfers information from a reference sample of stars, or *training set*, to any dataset of interest, referred to as *test set*, by optimizing purely mathematical expressions. Such methodology is conceptually different to those followed in more traditional investigations in Galactic Archaeology, so this chapter is aimed towards explaining in detail the decision making in developing this unconventional pipeline:

- We start off by defining the type of objects targeted in this study, namely, FGK stars from GES. The so-called test set will therefore be composed of continuum-normalised and radial-velocity corrected spectra corresponding to these stars.

- The second section is dedicated to the design and synthesis of a grid of stellar spectra, acting as the training set. This spectral model is generated from first principles. That is, integrating the radiative transfer equations with the stellar labels as inputs, which are in fact based on spectroscopic studies of solar neighbourhood dwarf stars.

- Finally, an explanation on how *The Cannon* operates and deals with the input of both training and test sets will close up the chapter and lead the way towards the final implementation of the pipeline.

3.1 Test sample: FGK stars

We briefly described some of the main motivations for choosing FGK type stars in the previous chapter. This section is now aimed towards motivating our choice of large spectroscopic survey and laying out the criteria followed in the selection of stars with such spectral types. Understanding which region in the stellar label space corresponds to these stars will help us design a representative and hopefully successful spectral model.

3.1.1 *Gaia*-ESO Survey

The Gaia-ESO Public Spectroscopic Survey is a multi-purpose collaboration lead by two Co-PIs, Gerry Gilmore and Sofia Randich, using the facilities at the European Southern Observatory (ESO) that wishes to homogeneously characterise around 10^5 stars of various stellar populations and clusters in the Milky Way. The survey serves as a follow-up effort of the European Space Agency (ESA) Gaia mission by integrating its astrometric solution

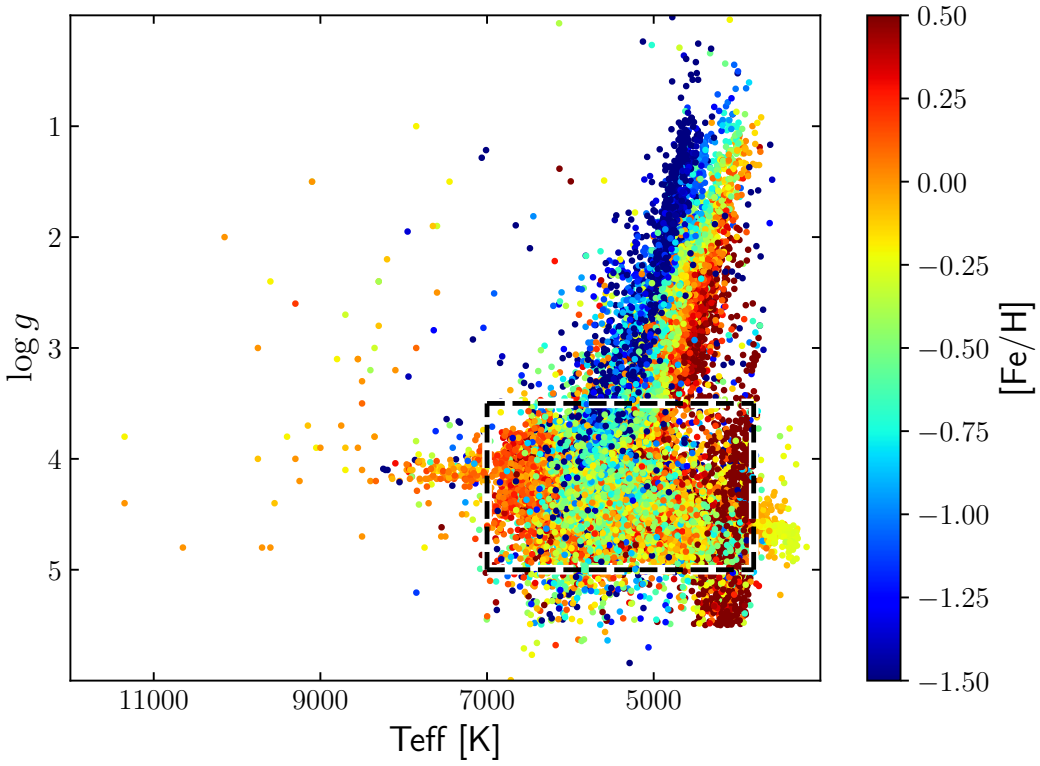


Figure 3.1: Kiel diagram of all 55000 stars belonging to GES iDR4, observed with FLAMES (both GIRAFFE and UVES). The dashed black box encompasses all GES dwarfs, limited by a hard cut at $\log g = 3.5$ dex (Smiljanic et al. 2014), an upper limit at $\log g = 5$ dex and effective temperatures ranging from 3800 to 7000 K.

to chemical abundances of multiple elements from bulge, halo, thin and thick disk stars of the MW.

Spectra from giant stars are predominantly observed with the medium-resolution instrument, GIRAFFE, part of the Fiber Large Array Multi Element Spectrograph (FLAMES) at the Very Large Telescope (VLT) in Chile. About 5000 dwarfs within 2 kpc from the Sun are targeted by the high-resolution Ultraviolet and Visual Echelle Spectrograph (UVES, Dekker et al. 2000).

The goal in this project is to analyse all FGK-type dwarfs from GES observed with the UVES spectrograph. The HRD in Figure 3.1 contains GES stars whose parameters belong to the the Internal Data Release 4 (iDR4). Limited by a dashed black box are both GIRAFFE and UVES stars with dwarf-like stellar parameters. From all stars inside the black box, a sample of 1580 dwarf stars have been observed with the UVES instrument. We then took their positions in equatorial coordinates and utilised them to carry out a 1 arcsec cone search to extract their spectra from the ESO Phase3 Archive Interface [1].

A final set of 1410 spectra was collected, processed by the GES pipeline in charge of background subtracting, calibration, order merging, SNR calculation, quality checking etc. For more detailed information about the data treatment see Sacco et al. (2014) or the ESO Phase 3 Data Release Description document [2]. The extracted *science* spectra are characterized by a resolving power of $R = 47000$, belong to the U580 UVES setting centered at 580 nm and covering the 480-680 nm region and correspond to the GES iDR5 release, which is probably reason why a slightly smaller number of spectra was returned using the iDR4 equatorial coordinates in the cone search. These final 1410 U580 spectra define our test set, with associated radial velocities (RV), projected rotational velocities ($v \sin i$) and stellar parameters that we are going to take as ground truth to Doppler shift and downgrade our spectra as well as further compare them to the results of our pipeline. In other words, these spectra belong to stars whose parameters have already been determined by GES and used in this project to validate *The Cannon* results for these same stars.

3.1.2 Continuum normalization

Unfortunately, among the features included in the GES data reduction pipeline, continuum normalisation is not one of them. Hence, in order to run our machine-learning algorithm, we developed our own normalisation procedure.

The choice of a proper continuum level is a long-lasting issue in the field of spectroscopic astrophysics. The continuum level sets the baseline of the stellar flux (or intensity) from which the strength of a spectral line is evaluated. This spectral strength or EW is pro-

[1] http://archive.eso.org/wdb/wdb/adp/phase3_main/form
[2] http://www.eso.org/rm/api/v1/public/releaseDescriptions/92

portional to the column density of a particular species given the atomic (or molecular) parameters related to the transition under study. Elemental abundances are derived by measuring the contrast between the line and continuous opacities. This means that unreliable results are obtained if the continuum is misplaced.

Continuum normalisation was left as a post-processing task for the end user in order to preserve as much information as possible from the UVES observations. Only the merged spectrum for each star is available. With the suggested guidelines in Sacco et al. (2014), we developed an algorithm to normalise these object spectra into our almost-final dataset. The procedure runs as follows:

(a) Extract the fluxes and intrinsic flux scatter (flux errors) for every pixel.

(b) Mask out the strong lines that affect the shape of the continuum, i.e. Na D doublet, Mg B triplet, Hα and Hβ.

(c) Re-size wavelength raster if needed to remove noisier spectrum edges.

(d) Split the spectrum in 60 fixed spectral bins.

(e) Calculate the median, $f_{\rm med}$, of the fluxes and errors per spectral bin.

(f) Asymmetric σ-clipping per bin, centered around $f_{\rm med}$, excluding pixels outside the $(f_{\rm med})^{+5\sigma}_{-\sigma}$ window.

(g) Calculate the median of the clipped fluxes and errors.

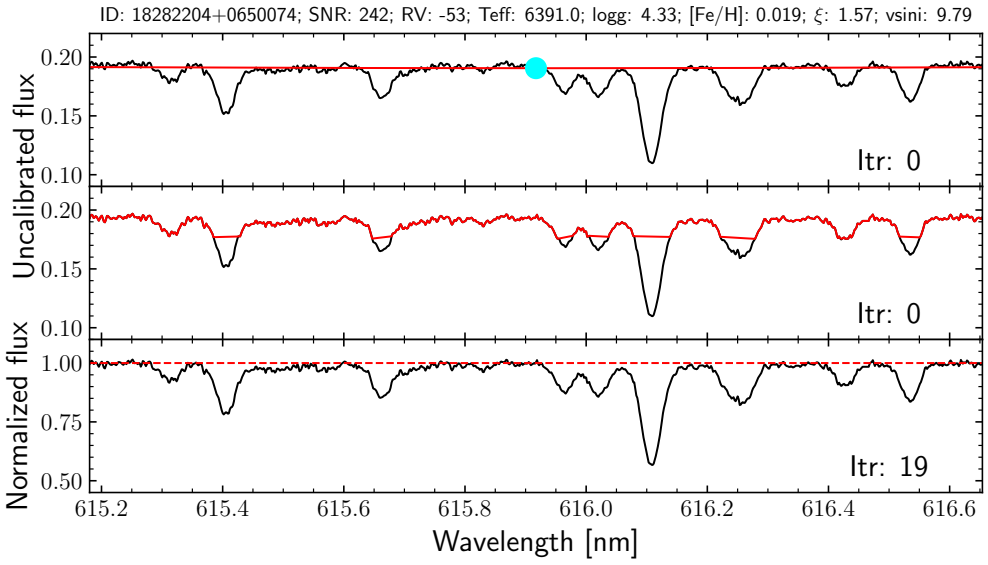
(h) Fit a 3rd degree spline through the median fluxes of the 60 spectral bins.

(i) Use the spline fit to normalise the input spectrum, $f_{out} = \dfrac{f_{in}}{f_{fit}}$.

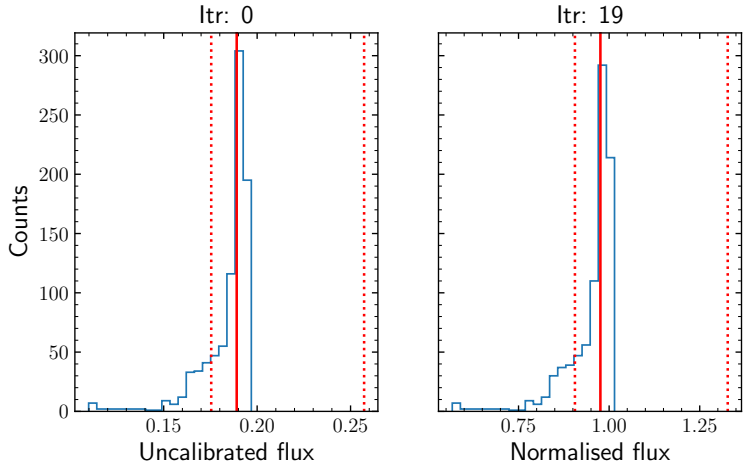
(j) Overwrite the input spectrum with the newly normalised spectrum and go back to (e) again.

(k) Exit the loop once we reached the desired goodness of the spline fit.

Figure A.1 in Appendix A and Figure 3.2a summarise the continuum normalisation implementation for an example spectrum and spectral bin respectively. The top panel in Figure 3.2a shows the initial shape of the spectral bin, further normalised by the spline to render the first normalised spectrum estimation. In every iteration, the recently normalised spectrum is fed into the algorithm as a non-normalised spectrum in order to re-normalize over and over again until the output spectrum in the bottom panel is obtained. For every step we are essentially approximating the continuum level to a spline function, an assumption that is as valid as the σ-clipping technique is successful at capturing the true continuum pixels. Both the intermediate panel and Figure 3.2b help explain how this works. Per iteration and bin, we calculate $f_{\rm med}$ and the upper (5σ) and lower (σ) rejection conditions. The



(a)



(b)

Figure 3.2: (a) Spectral bin illustrating the continuum normalization of a GES star observed with the U580 UVES setting whose labels, SNR, RV and identifier are placed in the header of the figure. The top panel shows the spectrum before normalisation (zeroth iteration) and the fitting spline function in solid red with one of its nodes in cyan; the second panel shows the same spectrum, with pixels in red indicating that they remain after the σ-clipping; the third panel contains the final version of the normalised spectrum, after iteratively scaling the input spectrum by the spline fit nineteen times. A horizontal red dashed line guides the eye of the reader towards the expected continuum level. (b) Histogram of the fluxes corresponding to the same bin showing the median in a solid red line and the upper 5σ and lower σ rejection criteria in dotted red lines for the zeroth and last iteration of the algorithm.

clipped pixels are then located between the red dotted lines in Figure 3.2b corresponding to these criteria and also coloured in red in the second panel of Figure 3.2a. Hopefully all true continuum pixels land inside this window, yet we are also capturing pixels that belong

to other spectral features. If these spectral traits are very prominent, the continuum level will be severely influenced translated into an over or underestimation of $f_{\rm med}$. Finally, since we fix the continuum level to unity, $f_{\rm med} \approx 1$, we say the algorithm has converged when the deviation of the fluxes around the median is smaller than 10^{-3}. The histogram on the RHS of Figure 3.2b indicates that our selection of continuum seems to be underestimated by 1-5%. It is visible from the solid vertical red line, our continuum placement, that we do not quite reach unity. However, the standard deviation of the pixels around this median is small enough.

Some caveats are worth mentioning in this normalisation program. First, the presence of more absorption lines will drag the continuum level downwards causing over-fitting. In this sense, we would require mobile rejection criteria to shrink the window as we loop over the same spectrum object to exclude false continuum pixels. On a different note, we tested our script with spectra of SNR > 200, where the deviation from $f_{\rm med}$ is smaller and prone to gradually reduce for increasing number of iterations. Low SNR spectra will, in principle, suffer from this as they have a systematic broad spread around $f_{\rm med}$. All in all, as much as it would have been desirable to analyse both L580 (480-580 nm) and U580 (580-680 nm) arms of the 580 setting, the density of lines in the blue spectra severely under-placed the continuum level. We therefore limited ourselves to the U580 setting.

3.1.3 Radial velocity correction

The wavelength shift of a spectrum belonging to a star in radial motion with respect to the observer is proportional to its velocity in such direction. In this survey (and data-release), the RVs are calculated by comparing the targeted spectra to a synthetic spectral grid. These templates cover effective temperatures between 3100 K $< T_{\rm eff} <$ 8000 K, surface gravities of 2.5 dex $< \log g <$ 5 dex and metallicities of [Fe/H] = {-1, 0} dex, compatible with the stellar label space of the GES dwarf stars we are interested in. For more information, see the ESO Phase 3 Data Release Description document.

To correct for this Doppler effect we used the RVs provided by GES and the script developed by Dominic Ford available in his Github page [3]. This Doppler shift is calculated using the known RVs for each star. A positive radial velocity means that the object is moving away from the observer:

$$\lambda_{\rm new} = \lambda_{\rm old} \sqrt{\frac{1 - {\rm RV}/c}{1 + {\rm RV}/c}} \tag{3.1}$$

where $\lambda_{\rm new}$ is the corrected wavelength of the spectrum, $\lambda_{\rm old}$ is the wavelength grid of the observed spectrum and c is the speed of light. It is worth mentioning that we detected some stars with incorrect RVs in GES and had to exclude them from our analysis (they

[3] https://github.com/dcf21

were already accounted for in our data selection scheme).

With continuum normalised and RV-corrected spectra composing our ready-to-use test set of 1410 FGK stars, we are ready to move on.

3.2 Training sample: spectral model

Previous work done by Ness et al. (2015), Casey et al. (2016) and Ho et al. (2017) define their *training set* as a sub-sample of reference objects whose stellar parameters are precisely measured or have well-defined chemical patterns. Good object candidates might be the Sun, the *Gaia* benchmark stars or open and globular clusters stars since their abundances should be the same overall. In this thesis we take our *training set* to be a synthetic set of spectra. Putting it in other terms, for each modelled spectrum, its flux as a function of wavelength is generated from a given configuration of its stellar labels. In principle, this model can extrapolate stellar attributes outside its label space, but in practice the results are not trustworthy. Thus, the quality of the investigation is tightly bound to the choice of label space in the same way as the quality of observational studies can only be as good as the results obtained for the reference stars. Hence, for our synthetic training set to properly emulate the targets of this study, the aforementioned high-dimensional grid has to be compatible with the stellar label space covered by the FGK type stars in GES. Only then will *The Cannon* be able to use the training set effectively as an interpolator to fit the stellar labels of the stars of interest.

3.2.1 Designing a spectral model

A statistically significant, thus successful, sub-set of stars or spectral model has to densely and broadly populate the stellar label space we wish to consider. Luckily, synthesising this model *a priori* allows the user to tweak and tune it in advance. The underlying point of the training set is to synthesise a spectral grid with the ability of modelling all possible spectral features that we expect to appear in the test spectra.

The Bensby catalogue

It is then imperative for a proper training sample to cover both the stellar parameter and chemical abundance space. To first order, the stellar parameters are the labels that fix the shape of stellar spectra; chemical abundances can be interpreted as a second order correction to the profile of absorption lines. Figure 3.3 compares two catalogues of stars, arranged in a Kiel diagram. The grey dots represent FGK type stars in GES while the color-coded ones correspond to stars from Bensby et al. (2014) (hereafter addressed as the Bensby catalogue). Main-sequence and turn-off stars (dwarfs) are separated from giants by a hard cut in $\log g$ according to Smiljanic et al. (2014). The pertinent sample of stars is therefore contained below the horizontal dashed line, at higher $\log g$ values. As we can

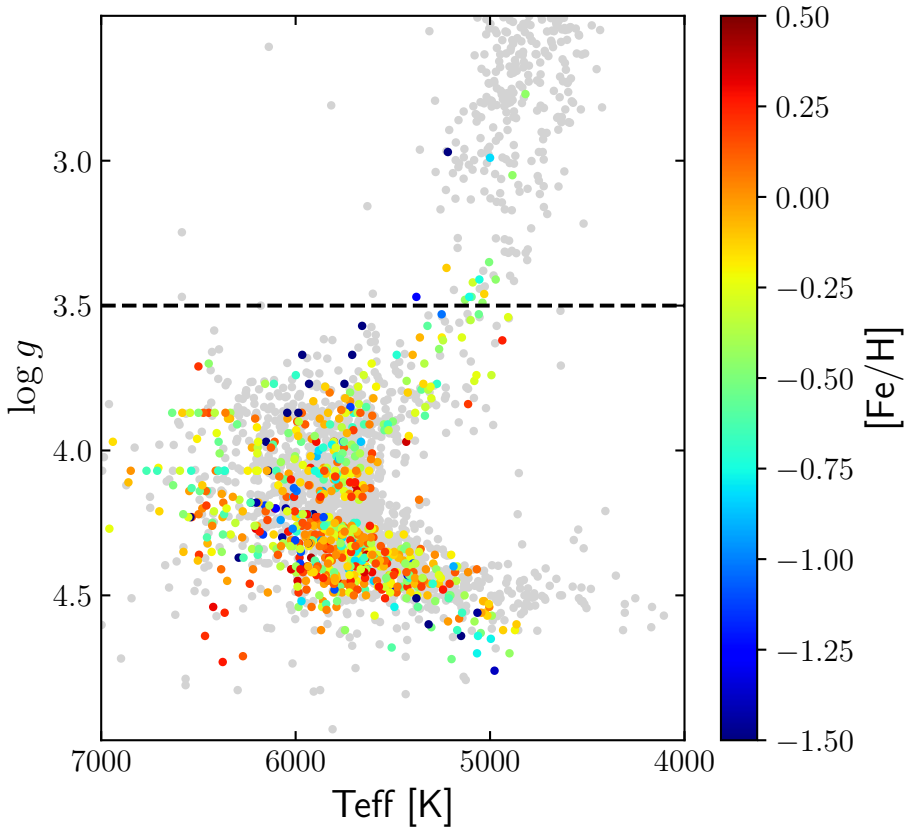


Figure 3.3: Kiel diagram comparing the Bensby catalogue of 714 FGK stars (colour-coded) to over 2000 UVES stars in GES (grey). The horizontal dashed line at $\log g = 3.5$ dex separates the relevant dwarf sample from the giants.

clearly observe from the scatter plot, the Bensby catalogue encompasses stars with appropriate effective temperatures and surface gravities.

As stars in the Bensby catalogue belong to the solar neighbourhood and have been selected according to their kinematics to focus on the metal poor and rich limits of the thick disk, the metal poor limit of the thin disk and the metal-rich limit of the halo as well as the Arcturus moving group and Hercules stream. Moreover, the sample contains a significant number of turn-off stars, which are known to make age determination more attainable. Furthermore, Bensby et al. (2014) determine abundances of of 13 elements: O, Na, Mg, Al, Si, Ca, Ti, Cr, Fe, Ni, Zn, Y, and Ba, indeed studied in GES. All in all, the Bensby catalogue provides a good stellar sample to establish the groundings for our modelled spectra.

Catalogue	Chemical elements	X-matched stars
Bensby et al. (2014)	O, Na, Mg, Al, Si, Ca, Ti, Cr, Fe, Ni, Zn, Y, Ba	714
Battistini & Bensby (2016)	Sr, Zr, La, Ce, Nd, Sm, Eu	593
Battistini & Bensby (2015)	Sc, V, Mn, Co	662
Delgado Mena et al. (2017)	Cu	167
Bensby & Feltzing (2006)	C	50
Casali et al. (2019)	N	0
Mishenina et al. (2019)	Mo, Ru	-

Table 3.1: Catalogues cross-matched with the Bensby catalogue. The two following columns display, from left to right, the elements whose abundances we are interested in deriving and the number of stars in common with the Bensby catalogue.

Complementing the Bensby catalogue

We are carrying out a spectroscopic analysis of dwarf stars in the optical region of the electromagnetic spectrum. Thus, a model is perfectly complete for us when it recreates all absorption lines of all the relevant atomic species in such wavelength range. Of course, there are computed abundances of other elements included in GES that are not found in the Bensby catalogue. One option would be to assign solar values of the abundances to all our synthetic stars, yet a more realistic and sophisticated chemical labelling scheme was developed in this project.

First of all, a literature search was carried out to cross-match the Bensby catalogue stars with other catalogues containing spectroscopic studies of different groups of elements. Apart from the Bensby catalogue, 6 additional papers are listed in Table 3.1 together with the chemical elements they investigate and the number of cross-matched stars with the Bensby catalogue. A special treatment was developed for the two last catalogues:

- Casali et al. (2019): In this paper, the [C/N] abundance ratio is used as a proxy for stellar ages of giant stars experiencing first dredge-up mixing. Carbon and nitrogen are synthesised in the CNO-cycle and brought up to the surface when the convective zone grows and penetrates deep down into the interior of a star. This mixing washes away its original chemical imprint at birth and changes their observable C and N abundances. In particular, the nitrogen abundance is boosted with respect to carbon. The [C/N] abundance ratio is then a sensitive tracer dependent on how much helium is produced by the CNO-cycle, thus dependent on stellar masses and therefore stellar ages. We intend to examine dwarf stars, not giants, but given that we do have [C/Fe] abundances from Bensby & Feltzing (2006) we can obtain [N/Fe] by:

$$[\mathrm{N/Fe}] = [\mathrm{C/Fe}] - [\mathrm{C/N}] \tag{3.2}$$

  where [C/N] values are sampled from a uniform distribution such that $\mathrm{U} \sim (-0.9, 0.1)$.

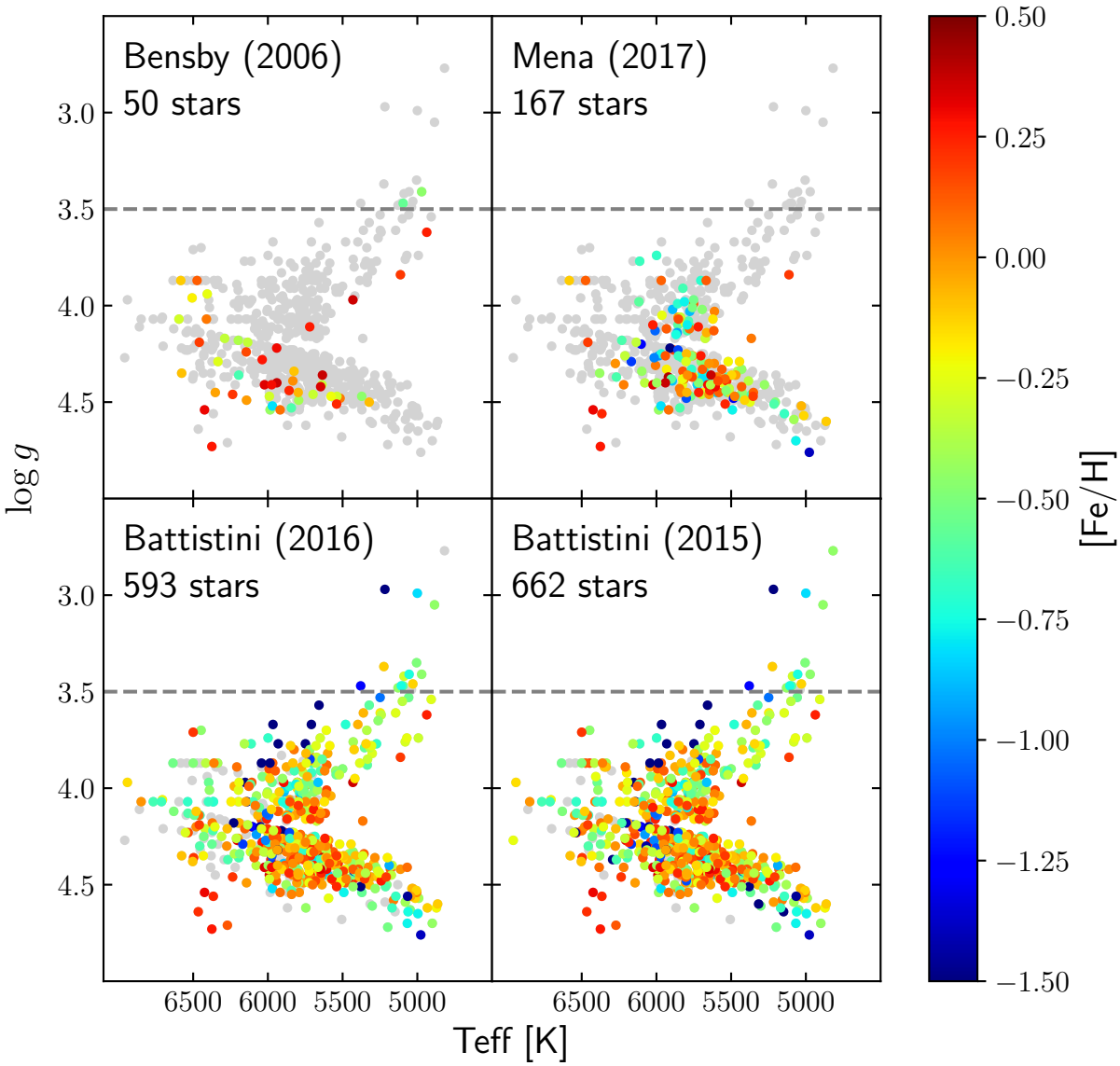


Figure 3.4: Kiel diagrams comparing stars in the Bensby catalogue (grey dots) with the cross-matched stars from the publications noted in the top left corner of each subplot. The horizontal dashed line shows the same hard cut at $\log g = 3.5$, above which GES stars are considered to be giants (Smiljanic et al. 2014).

The limits of the distribution have been drawn from Figure 9 in Casali et al. (2019) utilizing the Web Plot Digitizer tool. N is enhanced due to the CNO-cycle in giant

stars, but it should be kept fairly constant in main-sequence stars.

- Mishenina et al. (2019): This paper analyses the [Mo/Fe] and [Ru/Fe] abundance ratios for a sample of FGK type stars whose main stellar parameters are compatible with the HRD of the Bensby catalogue. However, no public online table was uploaded by the authors and therefore the desired abundance ratios and corresponding metallicities were also extracted with the Web Plot Digitizer tool [4].

The Kiel diagrams for the remaining 4 cross-matched catalogues are displayed in Figure 3.4. Stars in the Bensby catalogue are plotted in grey while the cross-matched stars in common are colour-coded by metallicity. It is immediately visible from all panels that many stars from the Bensby catalogue are missing in other catalogues and therefore cannot be assigned an abundance value. Once again, instead of setting solar values by default to all missing stars we will follow the next recipe, using C as an example (Figure 3.5):

(a) Cross-match the Bensby catalogue with any desired catalogue to identify which stars they have in common. The LHS plot in Figure 3.5 shows such cross-match with the stellar sample in Bensby & Feltzing (2006), where stars in common (blue dots) contain [C/H] abundances.

(b) Place the cross-matched stars in a [X/H]-[Fe/H] plane and fit a 1st degree polynomial through them (RHS panel in Figure 3.5).

(c) Extrapolate the fit to all other stars in the Bensby catalogue by relating their [Fe/H] to [C/H] relying on the empirical fit. The exact value is drawn from a Gaussian distribution such that $\mathrm{N} \sim \left([\mathrm{C/H}]_{fit}, \sigma_{fit}\right)$, i.e. the curve is centered around the [C/H] value directly provided by the empirical relation and spread according to the standard deviation of the observed data to the fit.

Figures B.1, B.2 and B.3 in Appendix B display [X/H]-[Fe/H], [X/Fe]-[Fe/H] and histogrammed [X/Fe] plots respectively for all 28 elements in Table 3.1. All α elements, O, Mg, Si, Ca, Ti, show a more or less clear double peak in their histograms related to the low- and high-α sequence. Since most of the stars in the Bensby catalogue have α abundances the addition of synthetic values does not wash away the thin and thick disk features, hopefully recuperated in our final analysis. Also worth mentioning is the dramatic depletion in number of stars at [Fe/H] < -1.5 dex, which makes the interpretation of abundance trends difficult in this region. Looking at this low metallicity end in Figure B.1 for elements like O, Eu or Cu, we suspect that a change of slope towards low metallicities might be happening suggesting that our fit could be improved by including a broken linear polynomial fit for these metal-poor stars.

Motivated by the same completeness arguments, we decided to include all other elements in the periodic table with at least one absorption line within the wavelength range covered by

[4] https://apps.automeris.io/wpd/

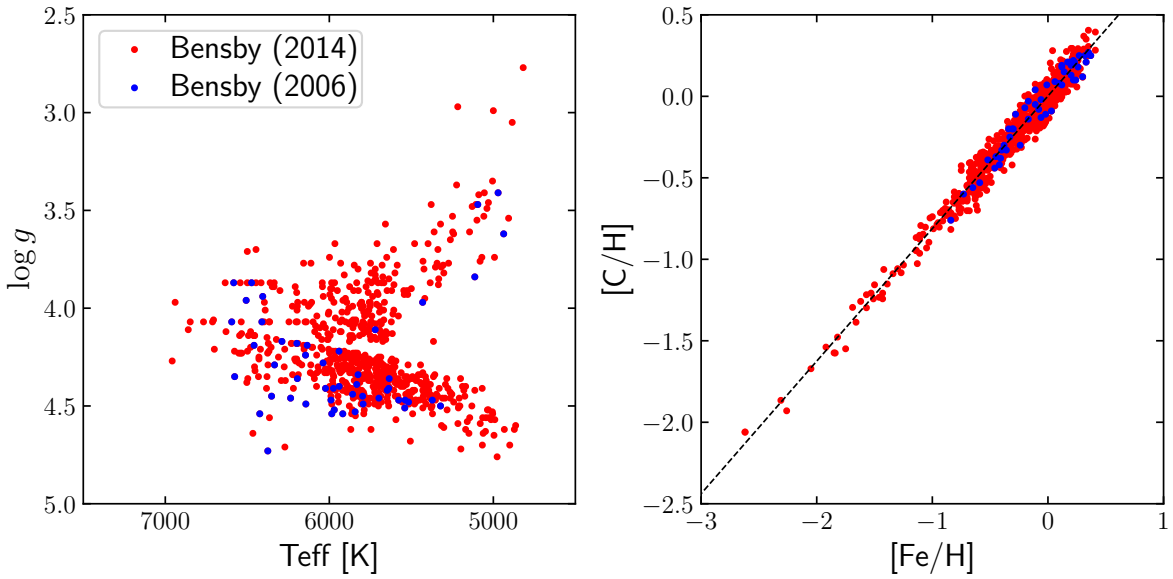


Figure 3.5: Missing abundances recipe using C as an example. Left: Cross-match between Bensby & Feltzing (2006) and Bensby et al. (2014) to identify which stars have carbon abundances. Coloured in red are stars exclusively catalogued in the Bensby catalogue whereas blue stars can be found in both publications. Right: Extrapolation of [C/H] abundances using the metallicities of the stars in both catalogues.

our test spectra. However, for most of these elements, no previous work was found in the literature utilizing dwarf stars. Nevertheless, we still applied the same missing abundances recipe, but as there are no stars to cross-match with the Bensby catalogue, the linear fit extrapolation was performed using 4 reference elements. Depending on which chemical group the element at stake belonged to, one of the five elements in Figure 3.6 is chosen: Mg for α, Ni for iron-peak, Ba for s-process and Eu for r-process elements. We decide to ignore Na and instead assign solar values such that $[\mathrm{X/Fe}] = 0 \pm N \sim (0.2, 1)$.

It might seem unnecessary to develop such an elaborate scheme to artificially tag stars with chemical abundances that will probably not be studied in the final implementation of the machine-learning pipeline. Nevertheless, early tests with *The Cannon* showed that when setting solar abundances by default it biased the results by correlating the abundances with the strength of Fe lines.

3.2.2 Synthesising a spectral model

Generating synthetic spectra from first principles means directly accessing the equations of radiative transfer (RT) and solving them in the context of modeled stellar atmospheres (Gray 2005):

$$I_\nu^+(\tau_\nu, \mu) = \int_{\tau_\nu}^{\infty} S_\nu(\tau'_\nu)\ e^{-(\tau'_\nu - \tau_\nu)}\ \mathrm{d}\tau'_\nu \tag{3.3}$$

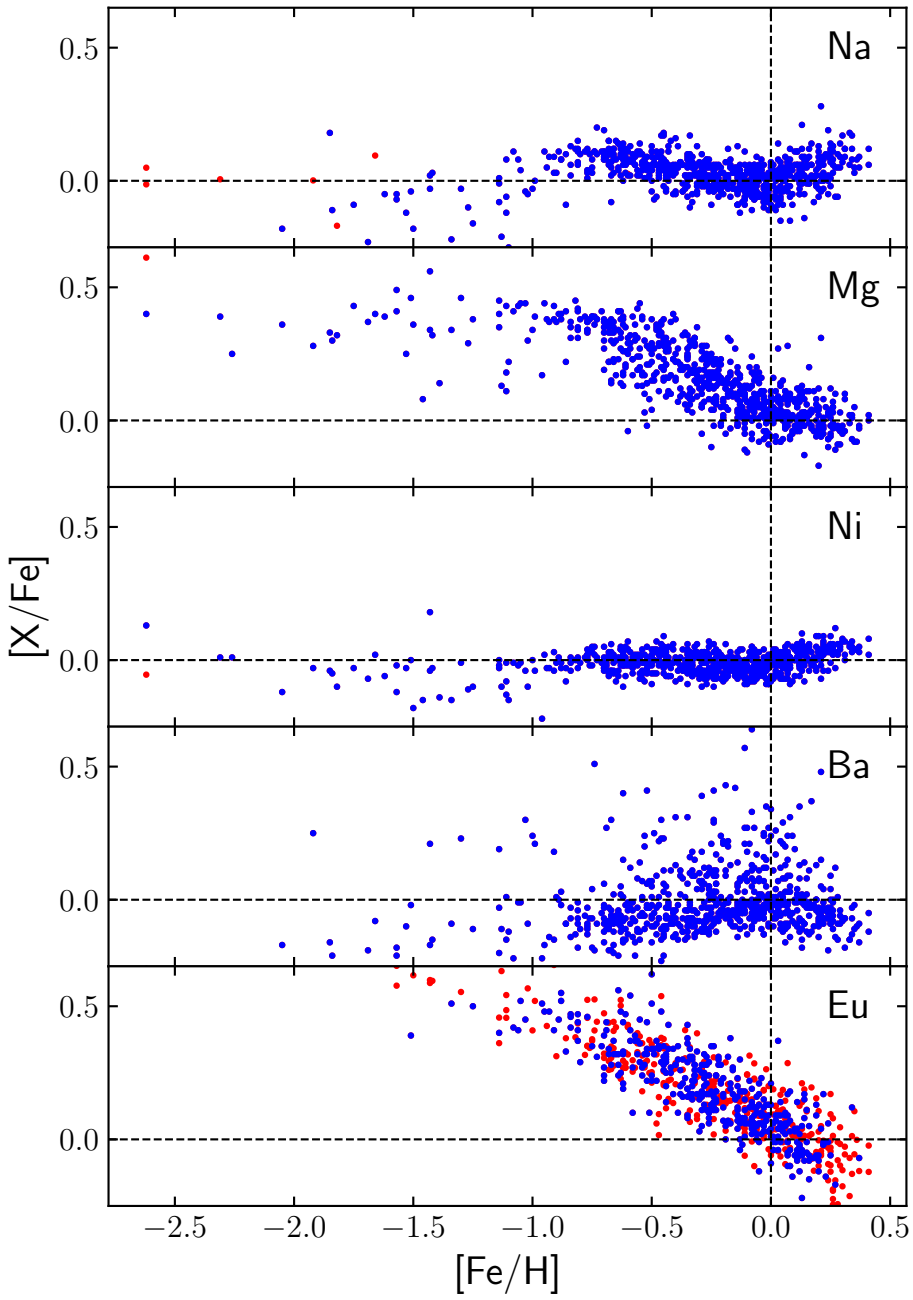


Figure 3.6: Abundance plots similar to those in Figure 2.3. Here, red dots have been added in the background to identify the Bensby catalogue stars with synthetic [X/Fe] abundances.

where I_ν^+ is the outward directed monochromatic intensity (a similar equation can be written for the flux), S_ν is the so-called source function of the model atmosphere and τ_ν is the optical depth defined as the geometrical length along the stellar radius weighted by the continuum opacity, α_c, and line opacity, α_l:

$$\tau_\nu = \int_\infty^r -(\alpha_c + \alpha_l)\ \mathrm{d}r' \tag{3.4}$$

Integrating the above mathematical expressions means understanding how the energy coming from the stellar interior is transferred through different depths of the atmospheric layers, reprocessed by the particles in them and finally re-emitted in the line of sight. Broadly speaking, two fundamental parameters matter: the source function S_ν and the extinction coefficient $\alpha_\nu = \alpha_c + \alpha_l$. The former describes the thermodynamical state of the atmosphere as a statistical collection of atoms, molecules, ions and electrons. The latter determines the optical thickness of the medium and gathers information about the composition of the gas in the atmosphere, ionization and excitation states, transition probabilities, broadening effects and so on. In short, much is to be borne in mind to solve Equations 3.3 and 3.4, where S_ν and α_ν might depend on each other in a rather complex fashion. In fact, the only way we can currently go about it is by making fairly strong assumptions:

- Hydrostatic equilibrium: Stellar atmospheres are physical systems in equilibrium, where the force of gravity keeping the layers bound together is balanced by the pressure gradient inside the star. No mass loss is considered.

- Local Thermodynamic Equilibrium (LTE): Equilibrium between matter and radiation is established in local parcels of gas in the stellar atmosphere. This is achieved due to the high collision rate among particles, responsible for most of the atomic and molecular transitions as well as re-distributing the energy inside the parcel. The mean free path of the photon is much smaller than the radius of the star and the timescale between collisions is much shorter than the timescale of change of observable stellar properties.

  In LTE, the excitation state of particles follows a Boltzmann distribution, their ionisation state can be described with the Saha equation, the velocity distribution of particles is Maxwellian and the source function, S_ν, turns into the Planck function. This simplifies equations greatly, where the most important idea is that under LTE conditions we can fully characterise our gas by knowing the temperature at each parcel.

- Homogeneity: The stellar atmosphere is homogeneous and therefore one can treat S_ν independently of the optical depth τ_ν.

- Plane-parallel atmosphere: The model atmosphere can be simplified by the plane-parallel approximation. The atmospheric layers are assumed to be parallel with properties conveniently expressed as a function of one dimension, the optical depth. Plane-parallel atmospheres work for dwarf stars, whose atmospheres are thin compared to the more extended ones in giant stars.

Integrating Equations 3.3 and 3.4 means theoretically deriving an intensity (or flux) as a function of frequency (or wavelength) of light exiting the stellar atmosphere. *Turbospectrum* (Plez 2012; de Laverny et al. 2012; Alvarez & Plez 1998) is the RT software implemented in this project to synthesise spectra from the library of stellar labels designed for our training set.

Three main input parameters are needed to run *Turbospectrum*, the effective temperature, surface gravity and metallicity of a star. The software takes these parameters to interpolate among the MARCS grid of model atmospheres (Gustafsson et al. 2008) to get the properties of the stellar photosphere at every opacity-weighted stellar radius (optical depth). These are 1D, plane-parallel, LTE model atmospheres containing the temperature profile, gas and electron pressure structures, opacity and microturbulence velocity as a function of an optical depth of reference, e.g. τ_{500}, the optical depth at 500 nm.

The output file from MARCS is a so-called opacity file, containing the atmospheric properties for every optical depth needed to roughly outline the shape of the spectrum of the star. This opacity file refers to the continuum opacity, in charge of accounting for continuous processes such as scattering phenomena or collisional broadening with neutral hydrogen (Barklem & O'Mara 1998). The opacity file is then combined with the third version of the Vienna Atomic Line Database (VALD3, Piskunov et al. 1995; Kupka et al. 1999; Ryabchikova et al. 2015) to re-calculate the spectral energy distribution and include the extra opacity coming from the presence of chemical species in the atmosphere. Information on the central wavelength of their transitions, their Einstein coefficients or $\log gf$ values, excitation potentials and broadening terms are provided by these line lists.

Synthetic spectra produced by *Turbospectrum* are infinitely resolved, that is, absorption lines are only naturally broadened. Such a model is not useful at all when other factors apart from more intrinsic mechanisms like microturbulence modify the profile of absorption lines. Fortunately enough *Turbospectrum* outputs continuum normalised spectra, yet the next section goes through the process of degrading our infinitely resolved modelled spectra to be more comparable to observational spectra [5].

Spectra for almost all stars in the Bensby catalogue were generated using their corresponding stellar parameters and chemical abundances for all relevant elements in the VALD3 line list (synthesised as explained in the previous section when necessary). Table 3.2 lists a sub-sample of 9 stars that happen to break the synthesis routine. Out of these, 8 have a surface gravity of $\log g = 4.50$ dex. Apparently, these stars sampled a problematic region of the MARCS grid for that $\log g$ value. This issue could have been solved by simply changing our value to $\log g = 4.50001$ dex, but we just removed the stars from the training

[5]Dominic Ford developed the bases of the pipeline utilized through out this project: both Python wrappers for *Turbospectrum* and *The Cannon* as well as the scripts to downgrade spectra. They were initially designed for the 4MOST Facility Simulator (4FS), so I adapted them to work for the UVES spectra from GES.

HIP	T_{eff} [K]	$\log g$ [dex]	[Fe/H] [dex]
2194	5211	4.50	-0.60
17970	5204	4.50	-0.49
28369	6940	3.97	-0.15
55210	5374	4.50	-0.25
56336	5878	4.50	-0.42
56452	5131	4.50	-0.40
82588	5322	4.50	-0.08
97358	5638	4.50	-0.03
106701	5701	4.50	0.08

Table 3.2: Main stellar parameters of stars from the Bensby catalogue that were not synthesised. Stars are identified by their *Hipparcos* ID number.

set. The remaining star was much more troublesome because it lead to the MARCS model having a negative convective velocity, and hence flux, at the most interior point. Again, we could have fixed this by setting these two variables to zero, but that could have turned into unwanted effects on the spectrum [6].

3.2.3 Downgrading a spectral model

Downgrading the synthetic set of spectra implies reproducing the observed line profiles by adjusting the modelled spectrum. We do it in three steps: convolving spectra to match the UVES resolving power and the projected rotational velocities, resampling the spectral fluxes to the wavelength grid of the U580 setting and last, adding noise to reduce the SNR. Figure 3.7 will help visualize this procedure with two example spectra observed with UVES: a solar spectrum in the top panel and a spectrum in the bottom panel for a star with a substantial rotational velocity.

Convolution

Parameters like the resolving power of the spectrograph, the stellar rotation or the macro-turbulence are equal or similar to changing the resolution of the observed spectral features. The difference between the two effects lies in the fact that a change in resolution broadens or shrinks the width of absorption lines, but the integrated flux or area under the line profile must be kept constant.

In this work we account for two sources of broadening, instrumental resolving power and projected stellar rotation. After some trial and error we decided to mimic the broadening effect by convolving our spectral lines with a Gaussian convolution kernel:

[6] The specific model atmosphere was `p7000 g+3.5 m0.0 t02 st z+0.00 a+0.00 c+0.00 n+0.00 o+0.00 r+0.00 s+0.00.mod`. See the MARCS website, https://marcs.astro.uu.se/.

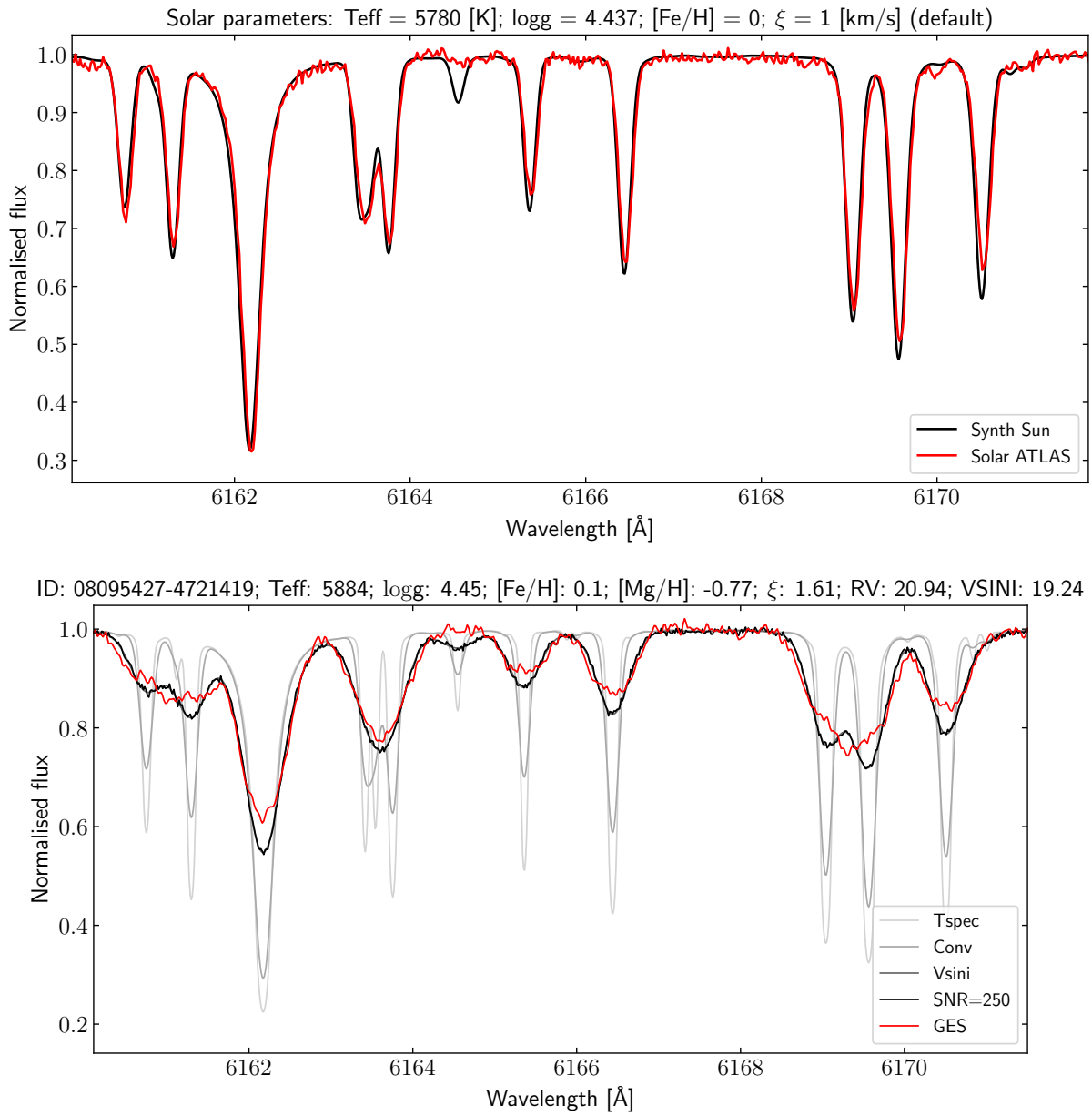


Figure 3.7: Top: Comparison between the normalised solar ATLAS (red) and the *Turbospectrum* rendering of the Sun (black) given the stellar labels in the header. The synthetic solar spectrum has been further downgraded to a resolving power of $R = 47000$, but has not been RV shifted nor broadened due to its rotational velocity. Bottom: Downgraded synthetic spectrum compared to the GES observation. Several curves are observed in the graph. The red spectrum corresponds to the GES observation while the black curve represents the fully downgraded synthetic version for the same stellar labels. The remaining grey lines represent the intermediate steps in the downgrading process: *Tspec* stands for the infinitely resolved synthetic spectrum; *Conv* represents the *Tspec* curve broadened to match the UVES resolving power; *Vsini* adds the projected stellar rotation into the broadening and *SNR=250* includes enough noise to match the specified SNR. Both *Vsini* and *SNR=250* are almost identical as the only difference is the addition of a small percentage of Gaussian noise with respect to the signal.

$$g = \exp\left\{-\frac{\lambda_{ker}^2}{2\sigma_{ker}^2}\right\}, \quad \text{with} \quad \sigma_{ker} = \frac{\text{FWHM}}{2\sqrt{2\ln 2}} \tag{3.5}$$

Here, λ_{ker} is the pixel grid of the convolution kernel and σ_{ker} is the width of the Gaussian. This kernel, g, is therefore passed along every pixel to widen spectral features according to its σ_{ker}, different for each broadening effect:

- Resolving power: We know that UVES spectra from GES have a resolution of $R = 47000$, which we can relate to the FWHM of the Gaussian as:

$$\text{FWHM}_{\text{UVES}} = \frac{\delta\lambda_{\text{UVES}}}{\Delta\lambda} = \frac{\langle\lambda\rangle}{R\Delta\lambda} \tag{3.6}$$

  where $\delta\lambda_{\text{UVES}}$ is the spectral element width, $\Delta\lambda$ the wavelength step of the input spectrum and $\langle\lambda\rangle$ its average wavelength. Caution must be taken when analysing the units of these new variables introduced. We are interested in a σ in pixel units, that is why we scale $\delta\lambda_{\text{UVES}}$ with the wavelength or pixel step. The top panel in Figure 3.7 shows the outstanding agreement between the solar ATLAS (available in the ESO website) and the convolved solar synthetic spectrum.

- Projected stellar rotation or $v \sin i$: The FWHM of the Gaussian kernel that would mimic the broadening due to $v \sin i$ is

$$\text{FWHM}_{rot} = \frac{v \sin i}{c} \frac{\langle\lambda\rangle}{\Delta\lambda} \tag{3.7}$$

  where c is the speed of light in km s^{-1}. The motivation for including broadening of spectral lines due to stellar rotation is summarized in the bottom panel of Figure 3.7. When downgrading the infinite resolution *Tspec* spectrum to the UVES resolving power *Conv*, we realised that the lines were still deeper and narrower. After adding $v \sin i$ to the convolution the synthetic model and observed spectrum looked much more alike.

Since we are dealing with statistically independent Gaussian broadenings, we can add our widths in quadrature, $\sigma_{ker}^2 = \sigma_{\text{UVES}}^2 + \sigma_{rot}^2$. Nonetheless it is worth commenting on the fact that neither the instrumental broadening or that due to stellar rotation have to be Gaussian necessarily. The first depends on the response function of the instrument and the second is supposed to follow a half ellipse profile (Gray 2005). The reason why we chose to convolve our lines with a Gaussian kernel was, again, by trial and error trying to recreate observed GES spectra with their stellar labels as inputs. There is still some mismatch between the black and red spectra in the bottom panel in Figure 3.7. We concluded that since macroturbulence is another parameter we did not account for (with Gaussian-like broadening), we might have been trying to compensate for both widening effects by forcing our $v \sin i$ term to be Gaussian.

Resampling

Resampling our synthetic spectra means re-expressing the synthetic flux on the same wavelength grid used for the U580 setting. Since one could, in principle, synthesise an infinitely sampled spectrum, it is common practice to start with a sampling 10 times more refined than that of UVES spectra. Once again, we used Dominic's routine to do so.

This routine essentially calculates the integrated flux sampled on the synthetic wavelength raster and linearly interpolates it onto the wavelength raster of choice with the `np.interp` command from Numpy. To ensure that the integrated flux is kept constant, the cumulative sum of the integrated flux is interpolated onto the U580 raster, divided by the width of the pixel.

Finally, resampling must come after convolution to avoid loss of information.

Addition of noise

Numerous SNR definitions are found in the literature. For the sake of argument, we will limit ourselves to a purely qualitative description.

We understand the SNR as the amount of noise compared to signal that contaminates each pixel of the observed spectra. Roughly, a SNR = 250 indicates that 1 out of 250 flux counts at a given wavelength is noise. We will then introduce noise in our spectral model by shifting the flux values such that:

$$f_i' = f_i + noise \tag{3.8}$$

Here, f_i is the flux of the ith pixel in the convolved and resampled spectrum and f_i' is the shifted flux after adding *noise*, a random number sampled from a Gaussian distribution with $\mu = 0$ and $\sigma = N$. Moreover, N is the noise level inversely proportional to the SNR value, $N = S/\text{SNR}$, where S is the signal level, i.e. $S = f_i$. A more detailed explanation can be found in Dominic's code.

3.3 The Cannon

Two datasets have been constructed so far. First, a set of spectra synthesised from observationally derived stellar labels. We name this spectral grid the *training set*. The second one, a sub-sample of FGK-type GES dwarfs whose stellar labels we want to recover. We name it the *test set*. The goal is to transfer the labels (stellar parameters and chemical abundances) from one dataset to the other without any additional physical modelling along the process.

To do this we make our machine, *The Cannon*, "learn" about the known stellar labels in the training set. We call this stage the *training step*, consisting of the creation of a

mathematical model which allows us to predict fluxes at all wavelengths as a function of stellar labels. The second stage, the *test step*, employs this predictive power to measure stellar labels for the test set.

The Cannon is built on the following assumptions:

- Stars whose stellar labels look alike have similar spectra. This is true if the picked labels suffice to fully characterise the star to be analysed.
- The stellar spectrum is a smoothly varying function of the stellar labels such that it can be approximated with a second order polynomial of the stellar labels.
- Training and test sets have been "observed" with the same instrument. The resolution of the spectra from both datasets is identical and are both sampled in the same wavelength raster.
- Training and test spectra are continuum-normalised.
- The intrinsic flux scatter at each wavelength pixel is known, normally distributed and statistically independent from the other flux variances of the other wavelength pixels.
- The stellar labels of the training set are well-known and homogeneously cover the relevant label space. We treat them as the basis to propagate our labels to the test set.

3.3.1 Training stage

The Cannon is a *generative model*. That is, it generates the probability density function of the flux at all wavelengths for every object spectrum as a function of the stellar labels. This generative model is a"purely mathematical relation" between spectra and labels and is what *The Cannon* optimizes or trains on by looking at the training set.

The predictive power of the model then relies on its flexibility, i.e. on how well the model replicates the label-spectral flux relations. It has been reported in the literature that a quadratic functional form of the model is sufficiently specific whilst flexible:

$$f_{n\lambda} = \boldsymbol{\theta}_\lambda^T \cdot \boldsymbol{\ell}_n + e_{n\lambda} \tag{3.9}$$

with

$$\boldsymbol{\ell}_n \equiv \begin{array}{l} [1,\ \ell_{n1},\ \ell_{n2},\ \ell_{n3},\ \cdots\ ,\ \ell_{nK}, \\ \ell_{n1}\cdot\ell_{n1},\ \ell_{n1}\cdot\ell_{n2},\ \ell_{n1}\cdot\ell_{n3},\ \cdots\ , \\ \ell_{nK}\cdot\ell_{nK}] \end{array} \tag{3.10}$$

and

$$e_{n\lambda} = \xi_{n\lambda}\sqrt{\sigma_{n\lambda}^2 + s_\lambda^2} \tag{3.11}$$

Here, the subscript n stands for every single object spectrum and λ represents each pixel (or wavelength element) within an object spectrum. In this sense, $f_{n\lambda}$ is the spectral flux per pixel per object spectrum and $\boldsymbol{\ell}_n$ is our quadratic-in-labels vector. This "vectorizing function" of the labels defines the freedom of the generative model and can, in principle, be any arbitrarily complex function of the stellar labels. For a quadratic polynomial Equation 3.10 encompasses a total of K linear ($\ell_{n1} = \mathrm{T_{eff}}$, $\ell_{n2} = \log g$, $\ell_{n3} = [\mathrm{Fe/H}]$), quadratic ($\ell_{n1}^2 = {\mathrm{T_{eff}}}^2$) and cross terms ($\ell_{n1} \cdot \ell_{n2} = \mathrm{T_{eff}} \cdot \log g$) per object spectrum. The first element of $\boldsymbol{\ell}_n$ sets the continuum at unity and forms the constant term of the polynomial. The polynomial coefficients at each wavelength are symbolized by $\boldsymbol{\theta}_\lambda^T$ and $e_{n\lambda}$ represents the flux error at each wavelength element in each object spectrum. As long as the model is good, $e_{n\lambda}$ can be sampled from a Gaussian distribution with zero mean and variance $\sigma_{n\lambda}^2 + s_\lambda^2$. Equation 3.11 shows the mathematical form of the noise. The $\xi_{n\lambda}$ term is just a random number drawn from a normal distribution, $\sigma_{n\lambda}$ is the uncertainty of the flux at each pixel of the training or test spectra and s_λ, the intrinsic flux scatter of The Cannon model linked to the pixel at stake. This last can also be interpreted as the general ability (or inability) of The Cannon model to faithfully reproduce the flux at each pixel

Inasmuch as the flux errors are drawn from a probability density function, *The Cannon* is a probabilistic model, thus it estimates the mean and the variance of the flux at all wavelengths and object spectra. We optimize this model by writing out the single-pixel log-likelihood function:

$$\ln p(f_{n\lambda} \mid \boldsymbol{\theta}_\lambda^T, \boldsymbol{\ell}_n, s_\lambda^2) = -\frac{1}{2}\frac{\left[f_{n\lambda} - \boldsymbol{\theta}_\lambda^T \cdot \boldsymbol{\ell}_n\right]^2}{\sigma_{n\lambda}^2 + s_\lambda^2} - \frac{1}{2}\ln\left(\sigma_{n\lambda}^2 + s_\lambda^2\right) \tag{3.12}$$

This formula is a log-likelihood function for both polynomial coefficients and the labels, which essentially looks like a χ^2 distribution. That is to say, maximizing the likelihood of either $\boldsymbol{\theta}_\lambda$ or $\boldsymbol{\ell}_n$ is the same as minimizing the χ^2 function.

In the training step we are interested in the Maximum Likelihood Estimate (MLE) of the $\boldsymbol{\theta}_\lambda$ coefficients. We fix the K-lists of labels for all training set stars taking advantage of the fact that both fluxes and labels are known and solve Equation 3.12:

$$\{\hat{\boldsymbol{\theta}}_\lambda, \hat{s}_\lambda\} \leftarrow \mathrm{argmax} \sum_{n=1}^{N_{ref}} \ln\left(f_{n\lambda} \mid \boldsymbol{\theta}_\lambda^T, \boldsymbol{\ell}_n, s_\lambda^2\right) \tag{3.13}$$

It must be emphasised that the likelihood function is optimized in such a way that we compute the MLE of the polynomial coefficients for each pixel individually but considering all $N_{ref} = 705$ object spectra at once. Our generative model is then composed by this library of coefficients $\hat{\theta}_{\lambda k}$:

$$\Lambda \text{ number of pixels} \downarrow \quad \xrightarrow{K\text{-list of labels}} \quad \begin{pmatrix} \hat{\theta}_{11} & \hat{\theta}_{12} & \cdots & \hat{\theta}_{1K} \\ \hat{\theta}_{21} & \hat{\theta}_{22} & \cdots & \hat{\theta}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}_{\Lambda 1} & \hat{\theta}_{\Lambda 2} & \cdots & \hat{\theta}_{\Lambda K} \end{pmatrix}$$

3.3.2 Testing stage

The Cannon has now been trained to estimate the probability density function of the flux, given a certain configuration of stellar labels. We then assume that our data-driven probabilistic generative model is also representative of the test set and solve for the stellar labels:

$$\{\hat{\ell}_{mk}\} \leftarrow \text{argmax} \sum_{\lambda=1}^{\Lambda} \ln\left(f_{m\lambda} \mid \boldsymbol{\theta}_{\lambda}^{T}, \boldsymbol{\ell}_{m}, s_{\lambda}^{2}\right) \tag{3.14}$$

Now, the observed flux is known, so we exploit the recently calculated library of $\hat{\boldsymbol{\theta}}_{\lambda}$ coefficients to extract the labels of the test spectra. From the mathematical expression above we realise that even if the χ^2 is fundamentally similar, we are now fixing $\hat{\boldsymbol{\theta}}_{\lambda}$ and $\hat{s}_{\lambda}$ to infer $\hat{\ell}_{mk}$ by considering one single test spectrum m (out of the total $M = 1410$ GES dwarfs) and optimizing for all pixels in such spectrum at the same time. The final output of *The Cannon* will look like the following array of labels:

$$M \text{ test stars} \downarrow \quad \xrightarrow{\text{Labels}} \quad \begin{pmatrix} \hat{\ell}_{11} & \hat{\ell}_{12} & \cdots & \hat{\ell}_{1K} \\ \hat{\ell}_{21} & \hat{\ell}_{22} & \cdots & \hat{\ell}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\ell}_{M1} & \hat{\ell}_{M2} & \cdots & \hat{\ell}_{MK} \end{pmatrix}$$

This whole process of computing stellar properties is called *label transfer*.

3.3.3 Fine tuning

In the previous sections we outlined what is considered a standard *The Cannon* (hereafter abbreviated as TC) run. Early work with the standard TC (Ness et al. 2015; Ho et al. 2017) involved small number of labels, e.g. T_{eff}, $\log g$, [Fe/H], [α/Fe]. It turned out that, against expectations, one could easily extend the number of labels without requiring an exponentially growing training set to properly cover the label space. The reason behind this behaviour is that TC acts as an interpolator transferring the labels by using the quadratic polynomial as an interpolating function. Furthermore they realised that when optimising TC model to fit chemical abundances, the polynomial coefficients corresponding to a large

number of wavelength pixels had a non-zero contribution. Physically this makes no sense as the [Mg/H] abundance, for instance, must only be affected by the flux in pixels belonging to Mg absorption lines and nothing else. This lead to the idea of *regularization*.

Regularization methods incorporate an extra term when optimizing the log-likelihood functions in the training step: the *penalizing function*. From a more qualitative perspective, this function encourages θ coefficients to become very small, i.e. it increases the number of null components in the generative model with the sole purpose of restricting the wavelength dependencies for some desired labels. Not only did this make TC much faster, but also avoided overfitting of the polynomial coefficients. Discussion about the different regularization techniques is out of the scope of this project, but we motivate the curious reader to check the theory about LASSO regularization in Tibshirani (1996), which is the kind of regularization integrated in this TC implementation.

Censoring mask

Another useful concept is the *censoring mask*. The user may decide to run TC passing along a list of important absorption lines related to transitions of elements whose abundance are to be estimated. If so, TC will train and test only looking at narrow spectral windows around these lines, which might potentially change its overall performance. In this final section, we will discuss and motivate our choice of censoring mask by briefly

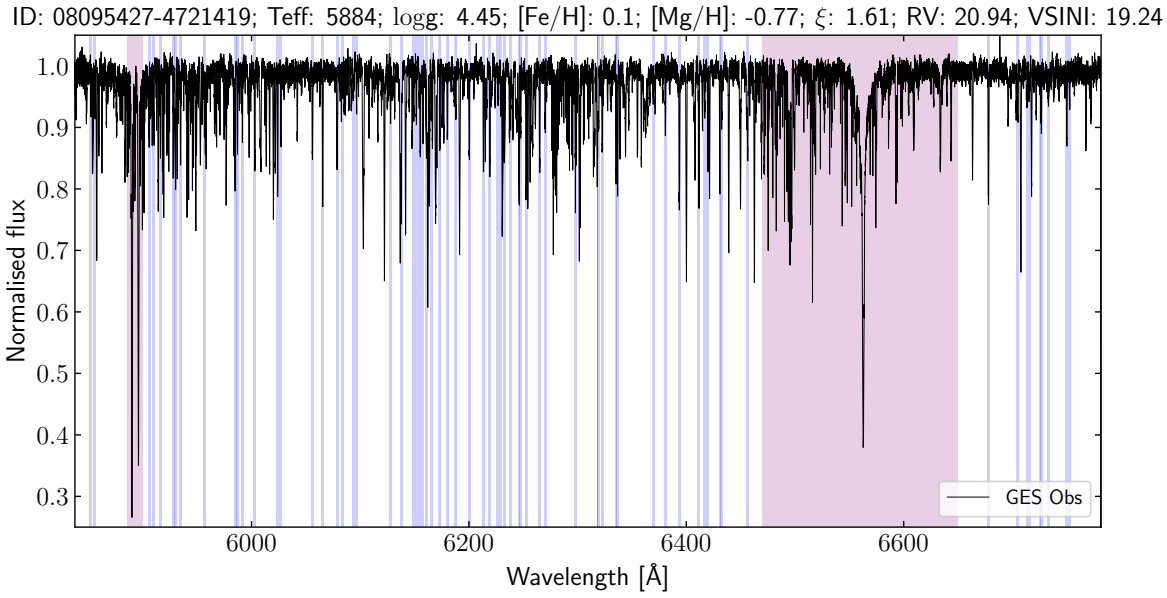


Figure 3.8: Masked U580 GES spectrum according to the line list published in Ruchti et al. (2016). The original mask was modified slightly to exclude the NaD lines (5886-5900 Å) and the Hα region (6470-6650 Å). Pixels in white and red are ignored by TC while pixels in blue are accounted for during the optimization.

analysing some preliminary results of a sub-set of 39 GES dwarfs (Table C.1 in Appendix C).

Figure 3.8 represents an example spectrum with a censoring mask based on the line list from Ruchti et al. (2016). White pixels are ignored while blue-shaded ones are "active", meaning that they would be included in TC analysis. One thing to note is that a large fraction of the spectrum is removed from the run. Two red (and also ignored) regions stand out: the Hα 6470-6650 Å window and the Na D lines at 5889.95 Å and 5895.92 Å. They have been highlighted in red because they are exceptionally problematic parts of the spectrum. In fact, it is while pondering on how to handle these two conflictive regions that we adopt a successful (enough) censoring mask based on the original Ruchti list.

Figure 3.9 zooms into these two problematic parts of the spectrum for two different stars and versions of the Ruchti mask. Looking at the top panel, the Hα panel, we see a mismatch in both core and wings when modelling the Hα line. We anticipated the cores of strong lines not to be properly profiled as *Turbospectrum* does not correct for non-LTE behaviour. On the other hand, we expected to get the wings right. The issue causing our spectrum to deviate from the model is most probably related to the continuum-normalisation stage. As a reminder, the Hα region was completely masked out from the normalisation process to avoid overfitting the continuum. The full extent of the Hα region of influence was decided upon by eye, inspecting the resulting shape of the line in the normalised output spectra for all stars. It turned out that its profile was extremely sensitive to the actual selection of the limits of this influence region. This lead to the idea that we could either mask out the whole 6470-6650 Å region as in Figure 3.8 or exclude the Hα core (as well as other lines in the vicinity) but keep some pixels of its wings to help constrain the effective temperature and surface gravity (shaded regions in the top panel of Figure 3.9).

The bottom panel in Figure 3.9, compares observed and modelled spectrum focusing on the NaD region. The plot is a good example of how ambiguous the shape of these two lines can be. What is more, Munari & Zwitter (1997) computed the equivalent widths of this doublet to estimate the interstellar reddening. Thus, it is not surprising that they often present blends with interstellar Na lines whose central wavelengths have been shifted due to the RV of the cloud in the line of sight. Driven by this chain of thoughts, we carried out a detailed by-eye comparison between the normalised observed spectra and the *Turbospectrum* rendering using the same labels. We found that our synthetic model failed to represent some spectral components, not only in the vicinity of the NaD doublet, but also around other absorption lines throughout the spectral range. They were close enough to the central wavelength of the transition to fall inside the spectral window of the censoring mask. Consequently, apart from trying out the idea of excluding the cores of the NaD lines, we also tested different widths of the spectral window: 0.5 Å for the cores, 1 Å to cover part of the wings and 2 Å, the default setting of the line mask.

To sum up, two characteristics of the censoring mask will be modified: the presence of

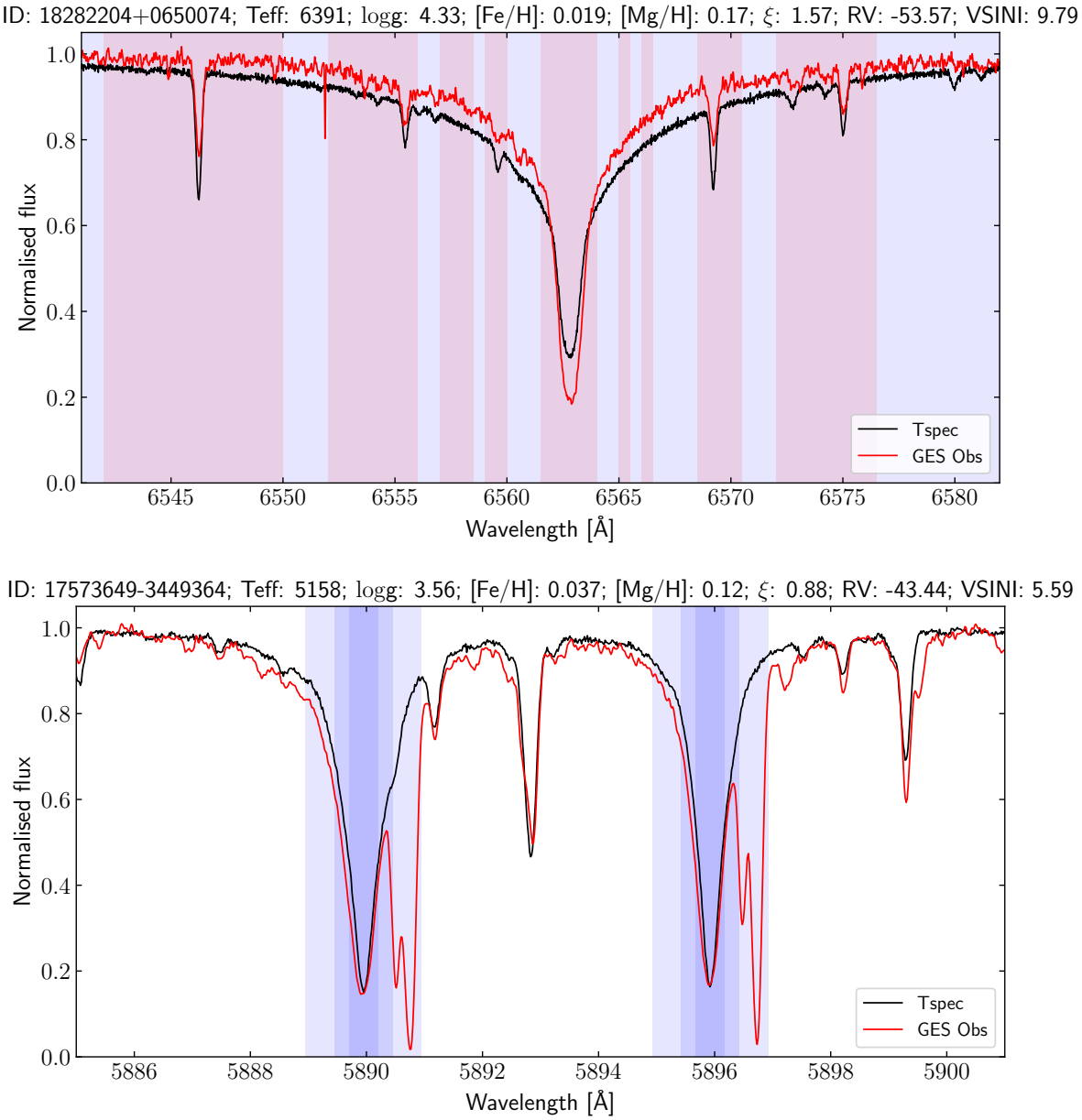


Figure 3.9: Top: Zoomed-in plot of the Hα spectral region. Two spectra are compared: the synthetic spectrum in black and the observed spectrum in red for the same stellar labels. As in Figure 3.8, the blue-shaded pixels are "active" pixels and the red-shaded ones remain excluded. Bottom: Zoomed-in plot of the Na D lines for a different star. Spectral model and GES observation are compared again, but now, three spectral windows around each line are displayed. From the narrowest and darkest window to the broadest and most transparent one the window widths are: 0.5 Å, 1 Å and 2 Å.

Hα and NaD lines and the size of the fixed width of the spectral window for all lines in the censoring mask. To decide which configuration works best in terms of TC performance, we will run our pipeline for 39 GES stars. Table C.1 contains the main stellar labels

Mask	ΔT_{eff} [K]	$\Delta \log g$ [dex]	Δ[Fe/H] [dex]
Full	-42 ± 238	-0.05 ± 0.33	-0.05 ± 0.12
No cores	-215 ± 188	-0.01 ± 0.30	-0.09 ± 0.09
No Hα/NaD	8 ± 154	-0.007 ± 0.219	-0.01 ± 0.08

Table 3.3: Mean and standard deviation of TC offsets with respect to GES for the three main stellar parameters and for the three types of masks represented in Figure 3.10.

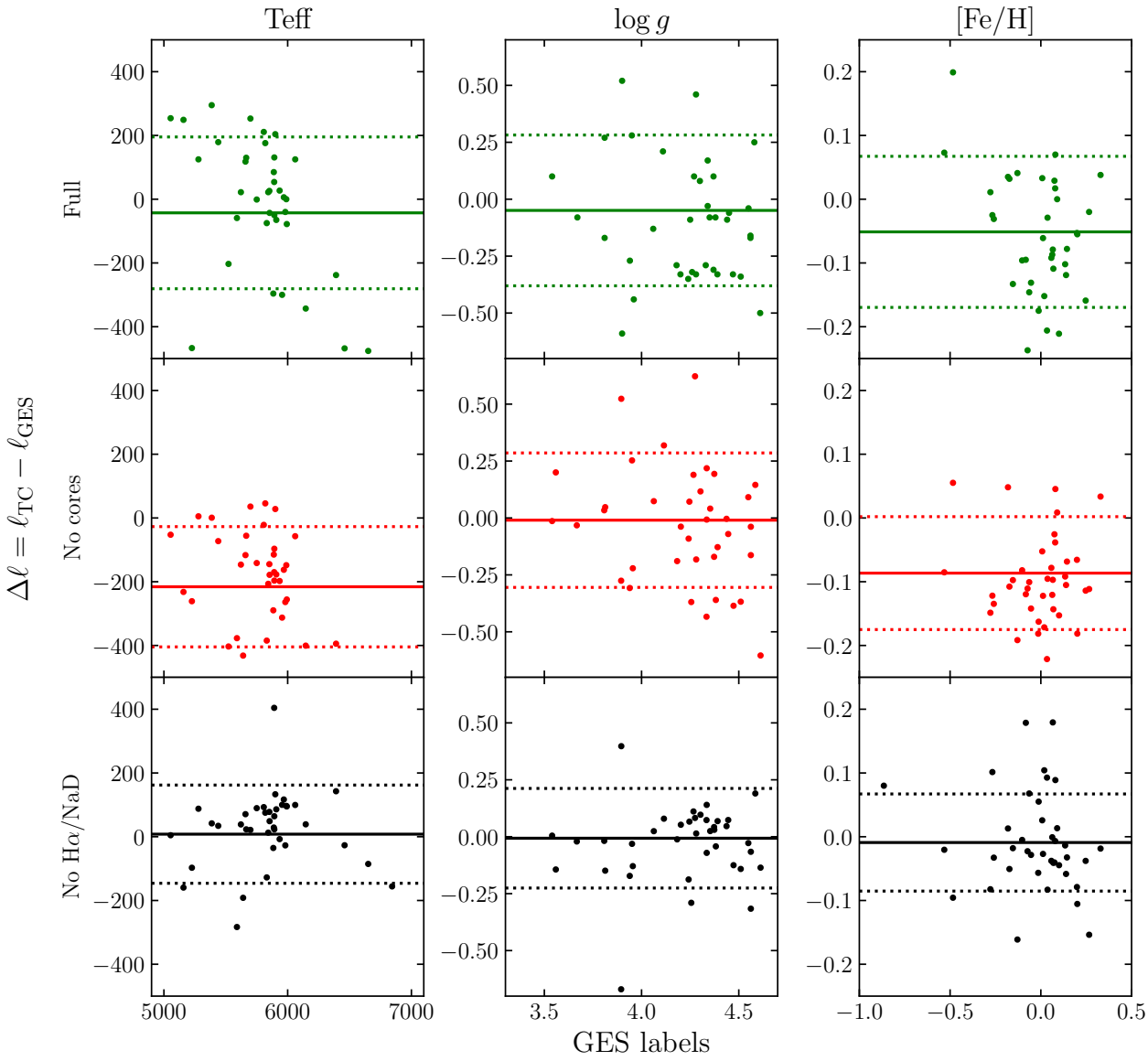
of these 39 stars, which were selected to have relatively solar-like parameters and high SNR.

To decide on how to manage the two strong line regions we ran TC varying the settings in three ways: first, without a censoring mask at all, i.e. looking at the whole spectrum; second, passing a mask that excludes the core of Hα (and some lines in its wings as seen in the top panel of Figure 3.9) and the Na D (bottom panel of Figure 3.9, where the narrowest window is turned into an excluding mask) lines; third and last, fully excluding the Hα and Na D regions. For these three different configurations, we kept the size of the spectral window at 2 Å. Figure 3.10 and Table 3.3 display the offsets between TC and GES values, their mean and standard deviation for these three configurations when training on 7 stellar labels: $\{T_{eff}, \log g$, [Fe/H], ξ, $v \sin i$, [Mg/H] and [Na/H]$\}$. However, results are only shown for the three main stellar parameters as they are the ones for which the improvement is most noticeable.

In general, TC labels experience a noticeable reduction in their offset with respect to the GES values by applying censoring masks that exclude a larger number of pixels. The dotted lines in Figure 3.10 and the uncertainties of TC results in Table 3.3 show the plus/minus standard deviations (SDs) of these offsets centered around their mean. As we select a more conservative mask, they decrease in magnitude indicating that the scatter of the offsets is reduced. The mean offset is also an important variable to check as it gives us an idea of how biased our results are with respect to the GES values. It is then obvious from the average offsets of Figure 3.10 and Table 3.10 that the mask that completely excludes the strong line regions is favoured with small bias and SDs of $\sigma_{\rm Teff} = 154$ K, $\sigma_{\log g} = 0.22$ dex, $\sigma_{\rm [Fe/H]} = 0.08$ dex.

Now, lets move on to varying the size of our spectral window. Now we will keep the mask that fully ignores the Hα and NaD regions and test three spectral window widths: 0.5 Å, 1 Å and 2 Å. Likewise, we produced the 3-panel figure for the three different window widths, however, no significant preference of size was visible upon inspection. We tried something new. Figure 3.11 shows TC averaged abundance offsets for 15 elements with at least one absorption line in the modified Ruchti mask. The three colours are linked to the three window sizes as indicated in the graph. We observe that our TC implementation seems to be systematically underestimating the abundances for all elements, with emphasis on Sc,

Figure 3.10: TC run for our subset of 39 GES stars. The dots represent stars in a plane with GES labels on the x-axis and the offsets between TC and GES labels on the y-axis. Three different runs are represented in each row of this plot corresponding to the three different colors: green for the run with no mask, red for the run where the mask excludes the cores of strong lines and black for the mask that excludes these regions. The horizontal solid lines represent the averaged offsets and the dotted lines, also colour-coded, represent the SD of the offsets.



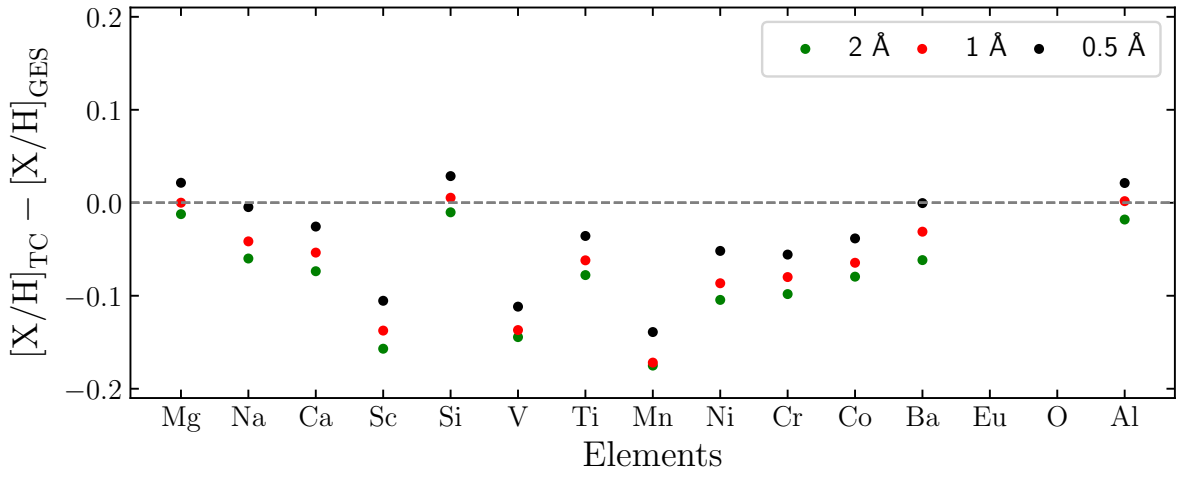


Figure 3.11: Offsets for 15 elements averaged over the 39 GES validation stars. Three TC runs are compared to each other, colour-coded by window sizes of 0.5 Å, 1 Å and 2 Å. A horizontal dashed line at zero was placed to guide the eye. Eu and O have no abundance values in GES and therefore no offset can be calculated for these two elements, yet TC does produce a chemical abundance for these elements.

V and Mn. The narrower the spectral window gets the more we rectify for this bias.

To conclude, from Figures 3.10 and 3.11 we learn that our current pipeline is improved by implementing a censoring mask based on the line list in Ruchti et al. (2016), arranged to exclude the Hα and NaD spectral regions and with spectral windows of 0.5 Å of width.

Chapter 4

Results & Discussion

To recap, we wish to transfer stellar labels (stellar parameters and elemental abundances) from a reference set of stars to a larger set in a spectroscopic survey applying a data-driven technique. The latter set of stars, addressed to as *test set*, consists of 1410 GES FGK-type dwarf stars with normalised RV-corrected UVES spectra covering a spectral region of 5800 Å-6800 Å with a resolving power of $R = 47000$ and a wide range of SNRs (10-300). The former, denoted the *training set*, represents a synthetic spectral model of 705 stars whose stellar labels are based on spectroscopic studies of solar neighbourhood dwarf stars. This model was further convolved to account for instrumental and stellar rotation broadening, resampled onto the U580 raster and degraded to a SNR of 250. A preliminary validation step was carried out to refine *The Cannon*, our machine learning algorithm. We realised that masking out certain chunks of spectra improved its performance. Indeed, we decided upon the use of a censoring mask based on the line list of Ruchti et al. (2016), slightly modified to exclude the Hα region of influence as well as the NaD lines and only include pixels in the vicinity of lines within a window of 0.5 Å around the central wavelength.

In this chapter we will analyse the final results of running TC on the full set of 1410 test stars training on a total of 20 labels: $\{T_{\rm eff}$, $\log g$, [Fe/H], ξ, $v \sin i$, [Mg/H], [Na/H], [Ca/H], [Sc/H], [Si/H], [V/H], [Ti/H], [Mn/H], [Ni/H], [Cr/H], [Co/H] [Ba/H], [Eu/H], [O/H], [Al/H]$\}$. The results for 25 randomly picked stars for the three first labels are listed in Table D.1 in Appendix D.

The first check to make sure that we are properly recovering the stellar parameters is to plot a Kiel diagram (or HRD). Figure 4.1 illustrates the recovery power of our pipeline. As much as the metallicity of stars appears to have an exceptional match, the shape of the main-sequence-to-turn-off transition is less smooth. We recover a less scattered Kiel diagram concentrated in the main-sequence, except for some outliers at $\log g < 3.5$ dex. It is worth mentioning that our pipeline has the ability to extrapolate stellar labels, so nothing stops it from placing stars outside the expected locations in the Kiel diagram. Moreover, if we scroll all the way back to Figure 3.3, we will realise that even if the test set was confined under a hard cut at $\log g = 3.5$ dex, a few of our synthetic stars were kept

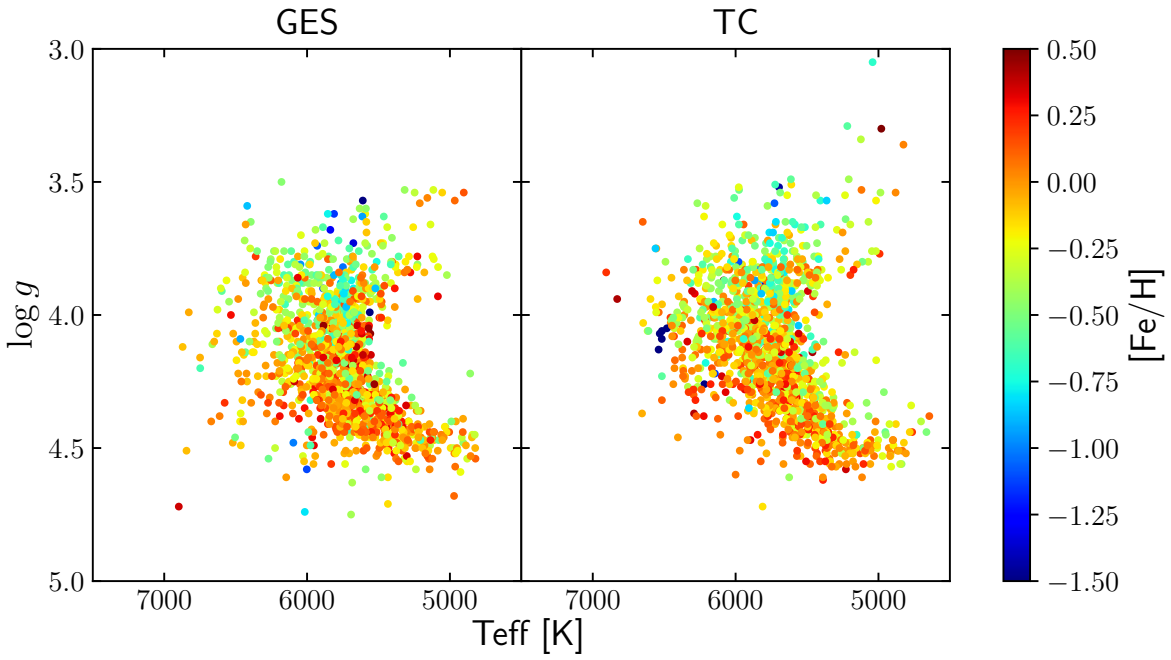


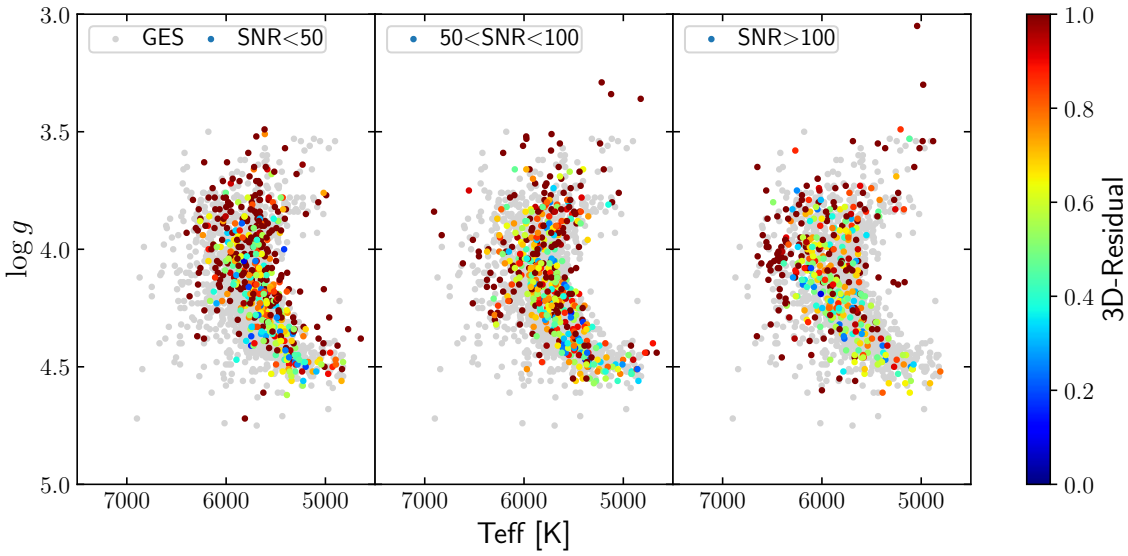
Figure 4.1: Comparison figure of two Kiel diagrams: the GES-reported stellar parameters on the LHS and the TC-reported values on the RHS.

beyond this boundary. These stars extend the label space of the training set making it more flexible and capable of extrapolating stellar properties towards that region. However, it also increases the tendency of TC to find a local minimum in the χ^2 fitting corresponding to stellar labels in such region. A gathering of very metal-poor stars was detected in the turn-off too, we will come back to these later on.

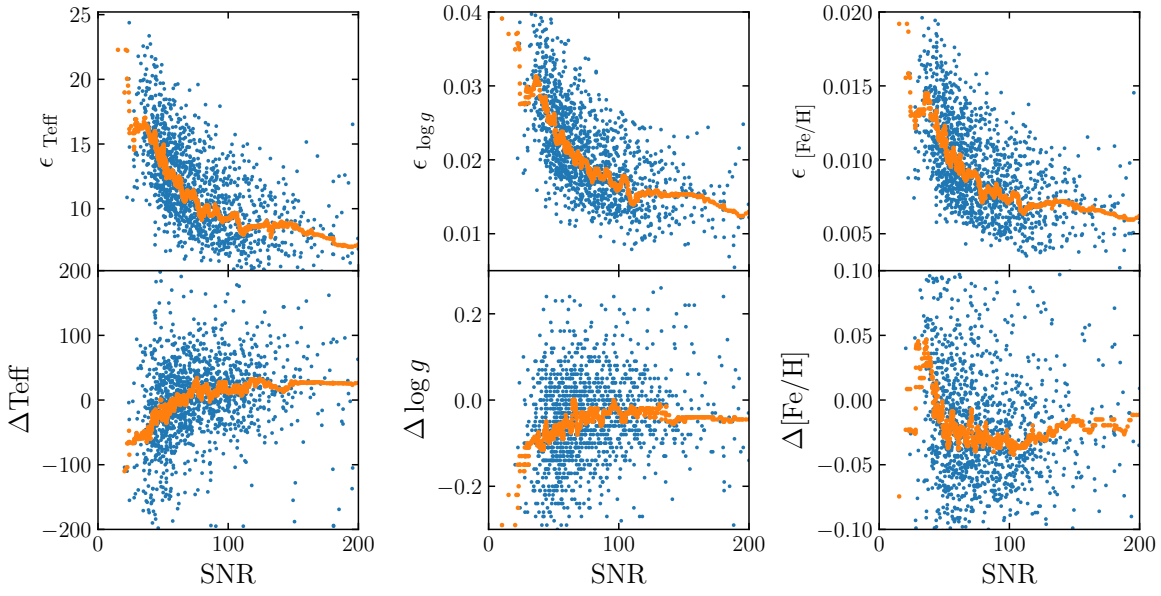
Given that our analysis encloses a wide spread of SNRs, we decided to divide the test set in 3 bins: the first 500 stars have SNRs between 10 and 61, referred to as the low SNR stars or SNR $<$ 50 bin; the next 500 stars have a SNR in between 61 and 94, referred to as medium SNR stars or 50 $<$ SNR $<$ 100 bin; the last 410 stars span from SNRs of 94 to 319 and are named high SNR stars or SNR $>$ 100 bin. Figure 4.2a compares three Kiel diagrams owing to the three different cuts in SNR. The GES test stars correspond to the background grey stars, but the coloured ones are coded depending on a 3D-residual calculated as:

$$3\text{D}{-}\sigma = \sqrt{\left(\frac{\Delta \text{T}_{\text{eff}}}{\sigma_{\text{Teff}}}\right)^2 + \left(\frac{\Delta \log g}{\sigma_{\log g}}\right)^2 + \left(\frac{\Delta[\text{Fe/H}]}{\sigma_{[\text{Fe/H}]}}\right)^2} \tag{4.1}$$

where the offsets of the three main stellar parameters are, once again, the difference between the TC and GES values, $\Delta\ell = \ell_{\text{TC}} - \ell_{\text{GES}}$. These offsets are weighted by σ_ℓ, which stands as the reported SDs in the bottom row of Figure 3.10 or Table 3.3 (TC run using



(a)



(b)

Figure 4.2: (a) Three HRDs for, from left to right, low, medium and high SNR. The GES stellar parameters render the stars in grey in the background while the color coded ones are TC results. (b) Scatter plot as a function of SNR of TC-reported uncertainties (top row) and the offsets to the GES parameters (bottom row). Orange dots represent a running mean with window size of 100 stars, performed on the scattered data.

a mask with window size of 2 Å and excluding the problematic strong line regions).

From Figure 4.2a one realises that the transition from low SNR to high SNR takes us through a slow decrease in the number of present stars with high 3D-σ. This indicates that TC provides smaller offsets at high SNRs, which is not unexpected as the improvement of the performance of our pipeline and the quality of the data go hand in hand. Perhaps more worth commenting on is the fact that although TC performs slightly better in the SNR > 100 bin, we still do a good job at analysing low SNR stars. Another common characteristic of the Kiel diagrams in Figure 4.2a is that the stars with smallest offset are located on the main sequence, where the training set is most dense. Specially in the high SNR bin, we see that most red stars are scattered around these main-sequence and turn-off features. It is also reassuring to observe that our "sub-giant" renderings correlate with large offsets.

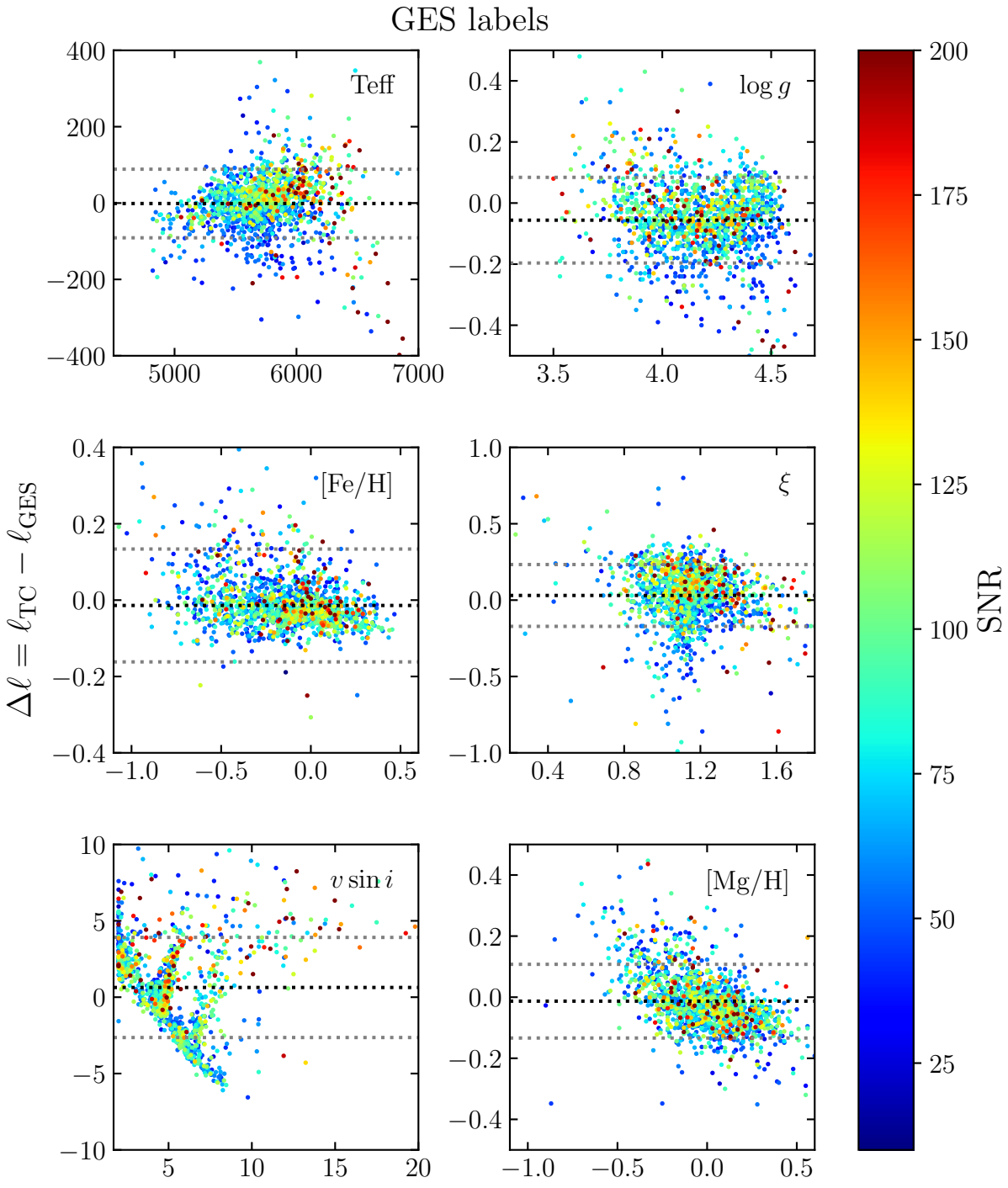
To complement Figure 4.2a we have Figure 4.2b. This array of subplots lay out the performance of TC as a function of SNR. We observe that the bottom row of this figure shows a deviating trend from zero as we move towards lower SNRs. It is at SNR < 75 that the curve bends towards underestimations of T_{eff} and $\log g$, but over-estimations of the metallicity. Nevertheless, offsets are generally well constrained. Regarding the top row of this figure, it is not entirely sure whether one can attribute it physical meaning. TC uncertainties do not reflect physical errors and therefore they are underestimated. In some sense, the TC observational flux uncertainties might be correct, but there are additional systematic uncertainties that come into play here when determining other stellar labels.

From this point on, and to a certain extent, we will refer to the GES labels as ground truth. In this sense, the purpose of Table D.1, Figure 4.3 and Figures D.1, D.2, D.3, D.4, D.5, D.6, D.7 and D.8 from Appendix D is to evaluate how well we reproduce the stellar parameters in such survey as a function of SNR. In Figure 4.3 the label offsets are plotted versus the GES values (colour-coded by SNR) along with their average and SD. Table 4.1 lists the mean labels offsets and SDs averaged over the number of stars in each SNR bin.

Label offset	Low SNR	Medium SNR	High SNR	All SNR
ΔT_{eff} [K]	-23 ± 90	10 ± 75	13 ± 100	-1 ± 90
$\Delta \log g$ [dex]	-0.09 ± 0.15	-0.04 ± 0.12	-0.04 ± 0.14	-0.06 ± 0.14
Δ[Fe/H] [dex]	0.003 ± 0.084	-0.01 ± 0.07	-0.04 ± 0.24	-0.01 ± 0.15
$\Delta\xi$ [km/s]	0.003 ± 0.202	0.03 ± 0.20	0.07 ± 0.20	0.03 ± 0.20
$\Delta v \sin i$ [km/s]	0.7 ± 3.6	0.2 ± 2.6	1.1 ± 3.5	0.6 ± 3.3
Δ[Mg/H] [dex]	-0.013 ± 0.144	-0.01 ± 0.1	-0.01 ± 0.11	-0.01 ± 0.12

Table 4.1: Averaged offsets of the 6 stellar labels represented in Figure 4.3 for the low, medium and high SNR bins as well as for the total 1410 stars.

Figure 4.3: 6-panel figure where the label offsets $\Delta\ell$ are plotted against the GES labels we pursue to recover. Stars are colour-coded by their SNR and are scattered around two dotted grey lines representing their ±SD: $\{\sigma_{\mathrm{Teff}} = 90$ K, $\sigma_{\log g} = 0.14$ and $\sigma_{[\mathrm{Fe/H}]} = 0.15\}$.



The first thing to notice from Table D.1 and Figure 4.3 is that the overall recovery of the GES labels is unbiased, with label offsets compared to the GES values in good terms with their corresponding uncertainties quoted in the literature (Jofré et al. 2019). Nevertheless, in strong disagreement with what has been discussed so far in this section, there does not seem to be a distinct improvement in the performance of TC as the SNR is increased. It is true that stars placed in the medium SNR bin are less biased and produce lower offsets with respect to the GES values when compared to the low SNR bin. However, high SNR stars do not seem to further correct the small biases in our labels and display dramatically large SDs. While low SNR stars could potentially appear as more volatile than high SNR stars, the outliers in the former lie closer to the zero offset mark compared to those in the latter, enhancing their offset uncertainties in Table D.1 (Figures D.1, D.3, D.4 and D.6). Lets take a closer look at Figure 4.3:

- Effective temperature, T_{eff}: The overall spread in effective temperature offset is $\sigma_{Teff} = 90$ K. Fitting T_{eff} works best for stars with solar-like temperature. At higher temperatures the number of stars both in the testing and training set are depleted. As a result, TC fails at predicting stellar effective temperatures above 6250 K approximately. A quick glance at Figure 4.1 suggests that artificially filling the high temperature sections of the HRD of our training set will most likely improve the measurements of T_{eff}. A population of low SNR stars appear to be systematically underestimated (see Figures D.1 and D.4).

- Surface gravity, $\log g$: The offset in $\log g$ has a typical deviation of $\sigma_{\log g} = 0.14$ dex for all SNR stars. At first sight it is perceived as more volatile than T_{eff}, perhaps owing to the fact that we completely removed the wings of strong lines from our analysis. Moreover, there is a population of stars that are systematically biased towards the bottom of the $\log g$ panel, trend that becomes more visible at medium and high SNR bins.

- Metallicity, [Fe/H]: The offset in [Fe/H] is the one whose SD suffers the most from outliers. This is the case for the high SNR bin with a SD of $\sigma_{[Fe/H]} = 0.24$ dex (see Figures D.3 and D.6), more than 3 times larger than the previously quoted for

CNAME	T_{eff}^{GES}	T_{eff}^{TC}	$\log g$ GES	$\log g$ TC	[Fe/H] GES	[Fe/H] TC	$v \sin i$ GES	$v \sin i$ TC
06403749+0954579	5992	6531	4.16	4.07	0.06	-2.14	-	3
06412224+0907372	6749	6539	4.16	4.13	-0.06	-2.16	-	3
06420663+0956026	6871	6516	4.12	4.06	-0.09	-2.14	35	2
08090850-4701407	6650	6517	4.10	4.09	-0.06	-2.07	-	3
10423830-6408320	6305	6481	3.97	4.05	0.01	-2.08	-	3

Table 4.2: Three main stellar parameters plus projected rotational velocities of five problematic stars. Values for TC and GES are quoted with the corresponding super-indices.

medium SNR stars ($\sigma_{\rm [Fe/H]} = 0.07$ dex, see Figures D.2 and D.5). The full extent of these outliers is best seen in Figure D.7, where TC completely disagrees with GES when fitting the [Fe/H] label for the 5 stars found in Table 4.2. These stars have a large effective temperature and, upon inspection of their spectra, we concluded that they also have a high projected rotational velocity as almost all spectral features were practically smoothed out. In other words, it is not surprising that a poor fit is obtained. Removing these stars from the analysis results in overall offsets of $\Delta T_{\rm eff} = -1 \pm 88$ K, $\Delta \log g = -0.06 \pm 0.140$ dex, Δ[Fe/H] $= 0.007 \pm 0.081$ dex, $\Delta\xi = 0.03 \pm 0.20$ km s^{-1}, $\Delta v \sin i = 0.7 \pm 3.2$ km s^{-1}, Δ[Mg/H] $= -0.01 \pm 0.12$ dex.

- Two more panels attract our attention, that of the microturbulence and projected rotational velocity. Both ξ and $v \sin i$ are effectively regarded as nuisance parameters concerning the broadening of the spectrum, but it becomes clearer from the panels that not much science can be extracted from them.

Regarding element abundances, we will move onto Figures 4.4 and 4.5 (complemented by Figures D.9, D.10 and D.11 in Appendix D for all elements). The first of these graphs displays the abundance offsets and deviations averaged over all our test set stars. As already hinted in Figure 3.11, TC consistently reports lower abundances for almost all elements. On the brighter side, the overall bias is minor, compatible with a 1σ spread of 0.05 dex for most species. Again, Sc, V, and Mn are especially underestimated, with striking error bars in the case of Sc. This would come as a surprise if it were not for the GES abundance panel for Sc in Figure D.9. The dark grey dots, corresponding to [Sc/Fe] abundances derived from Sc I lines have a huge spread in the abundance plane with respect to TC results. For V the bias is mainly due to TC analysing stars of solar metallicity

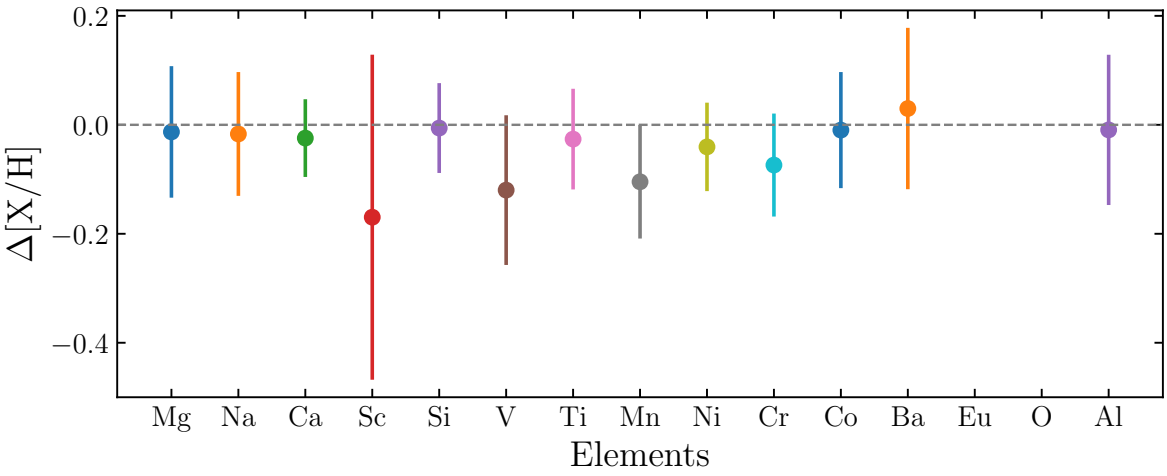


Figure 4.4: Averaged abundance offset and SD for all 15 elements studied. The horizontal dashed line has just been placed at zero for reference. Eu and O have no abundance values in GES and therefore no offset can be calculated for these two elements.

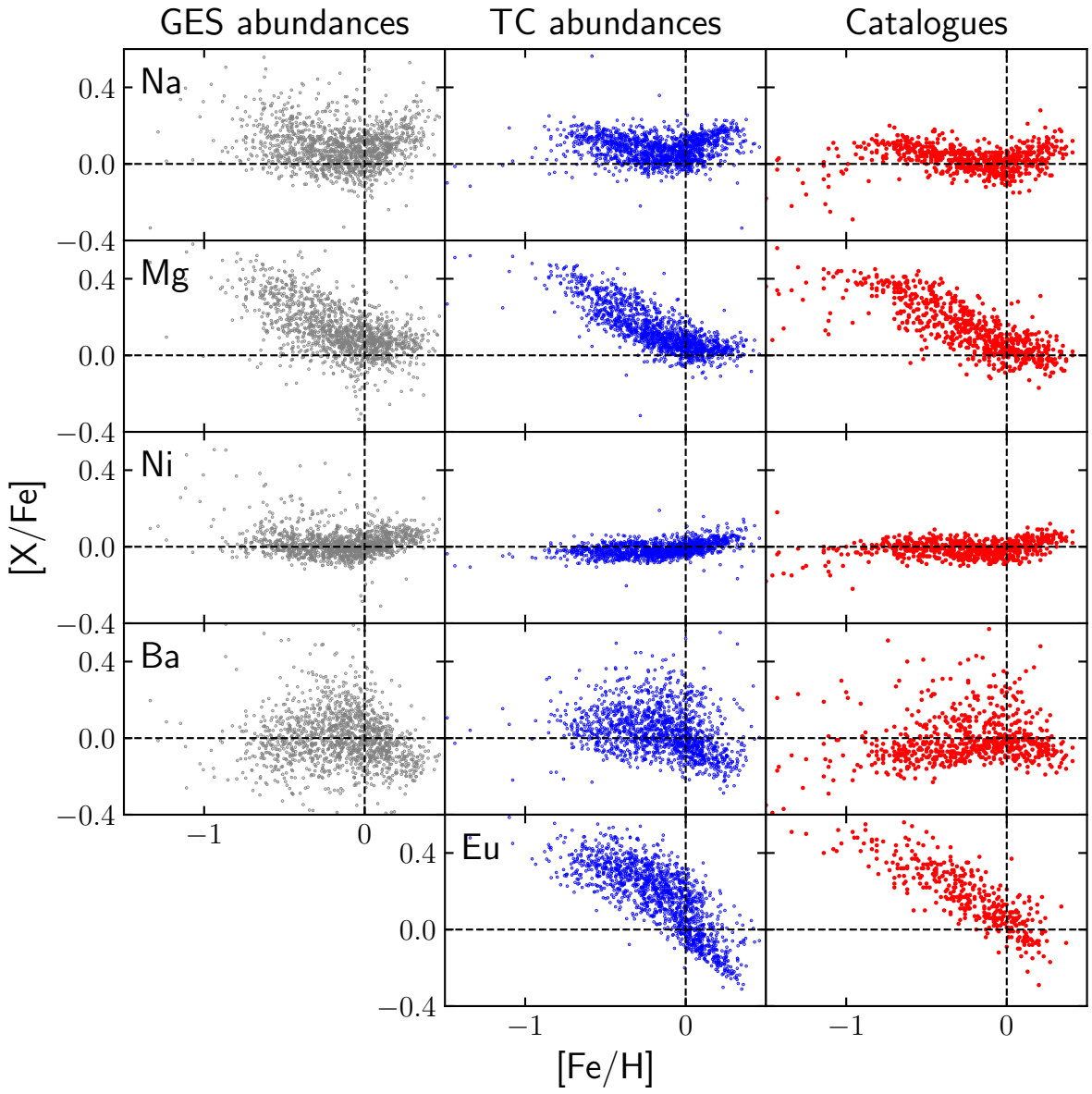


Figure 4.5: [X/Fe]-[Fe/H] abundance plots for our reference elements, comparing the results from GES, TC and the relevant catalogue of the element at stake: Bensby et al. (2014) for Na, Mg, Ni and Ba and Battistini & Bensby (2016) for Eu. Roughly speaking, the abundance plots in red represent the training set. Just as a reminder, Eu has no GES panel owing to the absence of Eu abundances for our test stars.

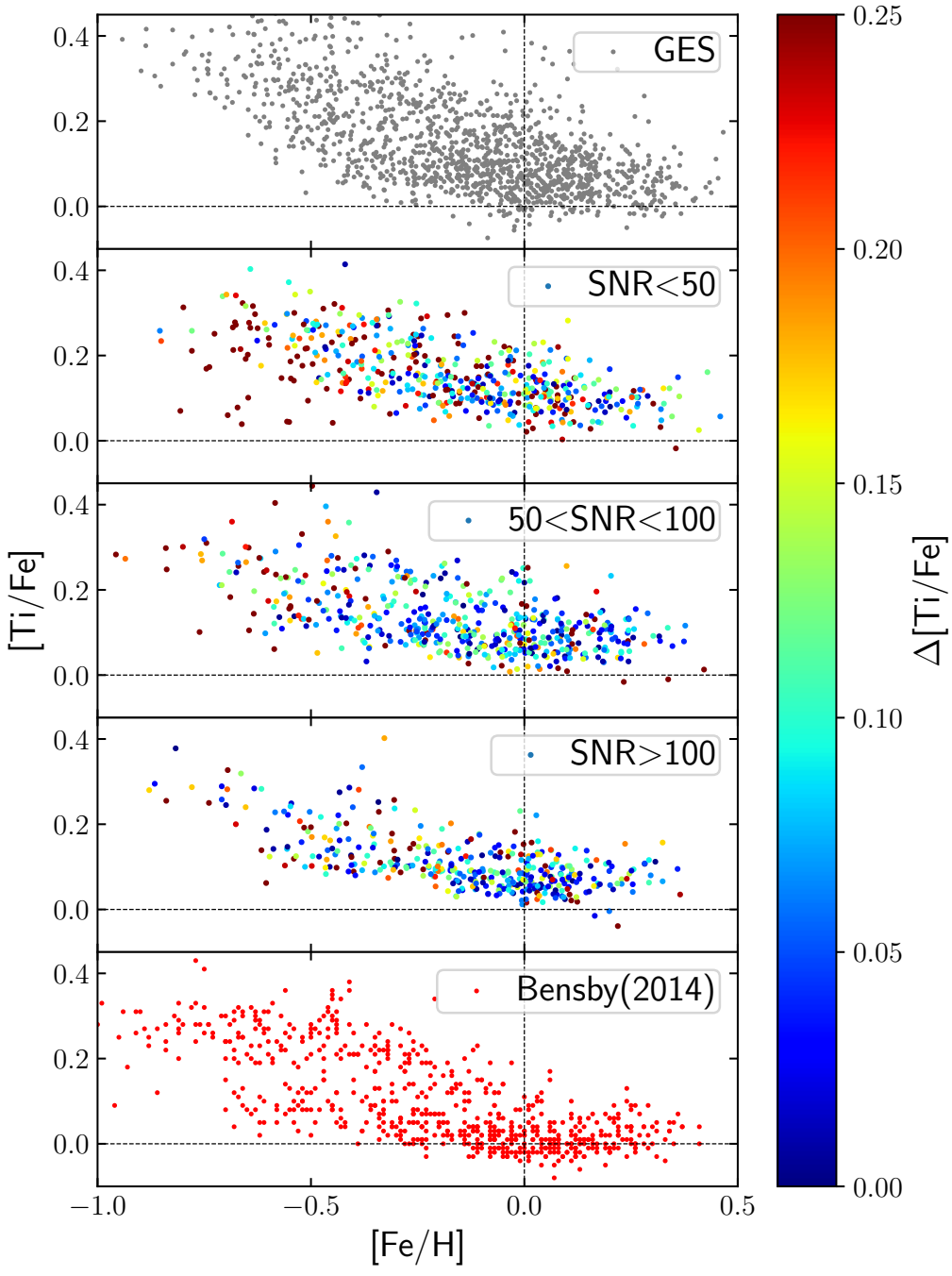
with an under-abundance of V with respect to GES and Battistini & Bensby (2015). The issue in [Mn/Fe] addresses low metallicity stars, which seem to contain constant [Mn/Fe] ratios at [Fe/H] < -0.5 dex in GES and Battistini & Bensby (2015). Compared to these results, TC values look like they reproduce an ever decreasing trend for this metal-poor stars.

Figure 4.5 is a convenient collection of plots to visualise the performance of TC when emulating abundance trends. On the face of it, TC abundance trends have substantially less scatter than the GES trends. One could interpret this as an improved performance over the GES results, with tighter trends being recovered. However, it is a general concern that this might come from an insufficient coverage of the abundance label space in the training set. Linking features to actual astrophysical mechanisms or systematics arising from the choice of training set is a slippery slope. If we draw our attention towards the Ni row, we observe a clear trend in TC and Bensby et al. (2014) panels towards sub-solar [Ni/Fe] as we approach metallicities bellow -0.7 dex. This is in strong disagreement with the GES [Ni/Fe] abundance ratios, with metal-poor stars having an enhanced Ni abundance. Either this is caused by not having high enough [Ni/Fe] stars in our training set (red panel) or perhaps TC is picking up a correlation between Ni abundance and Fe lines since we are not forcing it to train on Ni lines only (this statement can be claimed for every element). This possible biased Ni trend towards our selection of training set is worrying. On the brighter side, we can check the Ba row. A convex banana-like shape is distinguishable in all three panels, yet this "banana" curve is much more flattened and spread out in Bensby et al. (2014). These results suggest that our training set selection might not be seriously biasing our abundances. A sanity check would be to artificially extend the training set to astrophysically unlikely values to see whether the trends are severely affected or not. For instance, if we manage to pin-point these missing high [Ni/Fe] stars.

The case of Eu is rather special. There is only one single Eu absorption line in the line list from Ruchti et al. (2016), but it lies within the excluded Hα region of influence. Yet this does not stop TC from actually computing [Eu/H] abundances. Given that we have purposely removed all pixels which should in principle directly affect the value of this label, we think that TC is picking up spurious changes in flux that correlate with the Eu abundance.

One of the most important plots in Bensby et al. (2014) is Figure 22 in the paper, where the thin and thick disk are easily recognized as the low- and high-α branches in [Ti/Fe]-[Fe/H] space. Moreover, this distinction is also visible in the ages of stars. Even though we did not include stellar ages in our analysis we attempted to identify these thin and thick disk substructures with Figures 4.6 and 4.7. The selection function in Bensby et al. (2014) was orchestrated to target more kinematically hot stars in the thick disk, hence the more noticeable gap between the above-named low- and high-α sequences. Regarding TC results, a common feature among all three panels is that [Ti/Fe] is shifted to higher abundances. Even though we do see both thin and thick disk stars, the split of points is not nearly as visible in the rest of the panels perhaps due to this offset or simply caused by an increase in scatter with respect to Bensby et al. (2014). If any, the thin-thick-disk

Figure 4.6: [Ti/Fe] as a function of metallicity in GES (top grey panel), Bensby et al. (2014) (bottom red panel, basically our training set) and TC (three middle panels). These last have been separated into low, medium and high SNR bins as the plots are read from top to bottom and have been colour-coded by [Ti/Fe] offset.



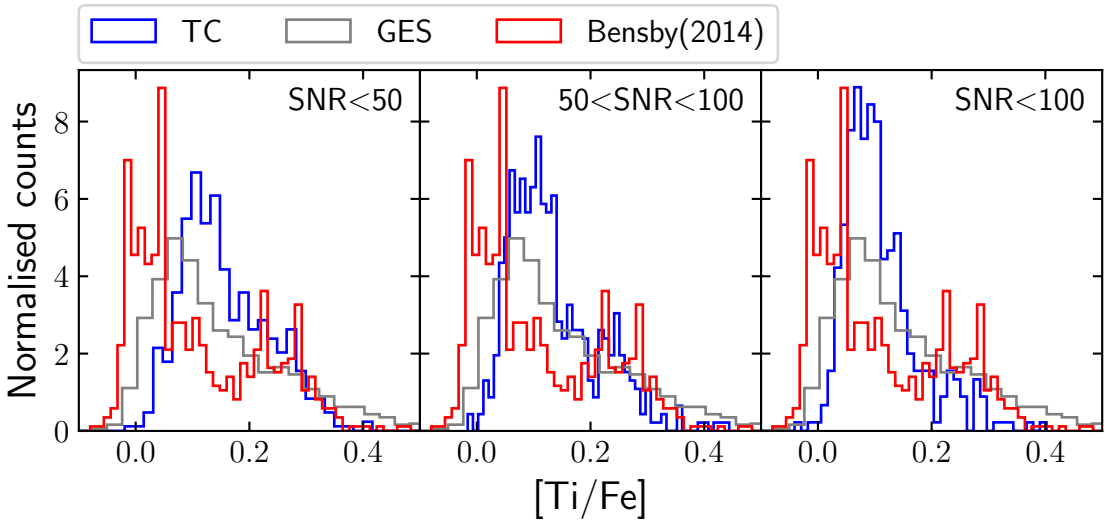


Figure 4.7: [Ti/Fe] histograms for the three SNR bins. Every panel contains the GES (grey), Bensby et al. (2014) (red) and TC abundances for each SNR bin (blue) for comparison purposes.

division is visible at medium SNR, which contains more stars than the high SNR bin. This is somewhat confirmed by Figure 4.7. Histograms of all 15 elements are found in Figure D.12 in Appendix D.

The last plot in this chapter is Figure 4.8. With this plot we want to compare the titanium versus iron abundance plots for all the datasets we have been dealing with and those obtained in Buder et al. (2019). None of the datasets seem to clearly distinguish between thin and thick disk, yet more concerning is the fact that the [Ti/Fe] in GALAH are under abundant with respect to TC results by 0.2 dex. Since abundances in Buder et al. (2019) are computed using Spectroscopy Made Easy (SME, Valenti & Piskunov 1996), we might be noticing some systematic biases between both codes, however, this distinct shift might also be due to a very different selection of stars in the survey.

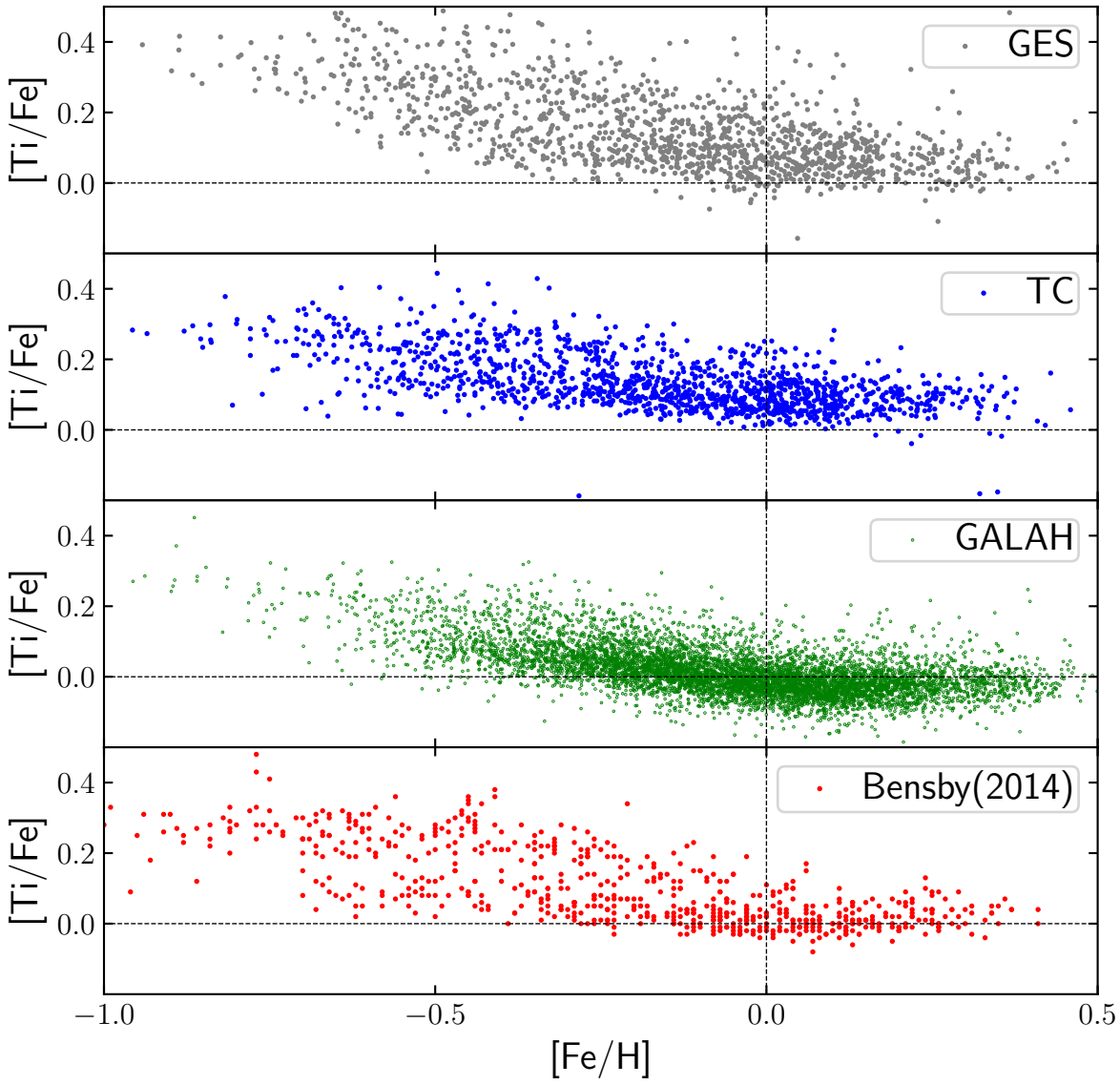


Figure 4.8: [Ti/Fe]-[Fe/H] abundance plots comparing our results to other datasets used in this work and those measured by SME for 8000 dwarf stars in GALAH (Buder et al. 2019).

Chapter 5

Conclusion

The purpose of this thesis is to explore an alternative data-driven approach in the analysis of stellar spectra. *The Cannon* is our machine-learning algorithm, designed to efficiently and accurately transfer stellar properties or *labels* from a reference dataset to any dataset of interest. The targets of this study, referred to as the *test set*, are then FGK-type dwarf stars from the *Gaia*-ESO Survey, among which we selected 1410 after a series of quality cuts and filtering of stars outside the relevant stellar parameter space. The dataset of reference, referred to as the *training set*, was substituted by a spectral model. *Turbospectrum* was the radiative transfer software used to synthesise the model based on stellar parameters of 705 stars from Bensby et al. (2014) and complemented with element abundances of modern publications in the field of observational Galactic Archaeology (Battistini & Bensby 2015, 2016; Bensby & Feltzing 2006; Delgado Mena et al. 2017; Mishenina et al. 2019; Casali et al. 2019). These two datasets required massaging before being fed to *The Cannon*. We continuum-normalised and RV-corrected the test set and downgraded the training set to match the resolution and wavelength raster of the UVES spectrograph. Some preliminary runs were carried out to evaluate the performance of the pipeline under different versions of censoring masks. These initial analyses trained on a validation sample of 39 GES stars by fitting only 7 labels: $\{T_{\rm eff}$, $\log g$, [Fe/H], ξ, $v \sin i$, [Mg/H] and [Na/H]$\}$. In the end, after fine tuning *The Cannon*, it was configured to fit 20 labels, $\{T_{\rm eff}$, $\log g$, [Fe/H], ξ, $v \sin i$, [Mg/H], [Na/H], [Ca/H], [Sc/H], [Si/H], [V/H], [Ti/H], [Mn/H], [Ni/H], [Cr/H], [Co/H] [Ba/H], [Eu/H], [O/H] and [Al/H]$\}$, on the full test set.

In summary, the main conclusions of this work are:

- We have successfully tagged 1410 GES dwarf stars by transferring information on 20 stellar labels from our synthetic training sample. Among all 20, valuable scientific results can be inferred from 18 of them: the effective temperature $T_{\rm eff}$, surface gravity $\log g$, metallicity [Fe/H] and 15 element abundances. The remaining 2 labels, ξ and $v \sin i$, are regarded as nuisance parameters due to their lack of scientific meaning and poor theoretical understanding in terms of how they affect absorption lines. As much as microturbulence is modelled in *Turbospectrum*, nothing else can be said from

its panel in Figure 4.3. The projected stellar rotation felt like a necessary broadening parameter to properly emulate the observed spectra. Nonetheless, we treated it as a fudge parameter in an attempt to account for other sources of broadening such as macroturbulence. Some finger-shaped features are observed in the $v \sin i$ panel of Figure 4.3. We assume these artifacts arise because of how we assigned $v \sin i$ values to the training set, i.e. we convolved each synthetic spectrum of our training set by randomly picking a from the GES $v \sin i$ values of the test set. An immediate improvement concerning this parameter would be to sample the $v \sin i$ values from a smooth distribution.

- Preliminary results strongly encouraged the use of a censoring mask (Figure 3.10). Since we are not forcing TC to look at a particular set of pixels depending on the label we wish to fit, the pipeline renders more scientifically valuable results excluding "unnecessary" or "problematic" pixels. Unnecessary in the sense of avoiding pixels other than those covering, say, Mg transition lines when fitting [Mg/H]. Problematic in the sense that our 1D, LTE spectral model does not fully cover all spectral features. One must also note that such improvement in TC performance is due to the careful selection of the censoring mask, which is based on the line list published in Ruchti et al. (2016). The spectral lines in this paper have been carefully selected to fit Galactic Archaeology studies using dwarf stars in the optical range.

- Two spectral regions were excluded as detrimental: the Hα and the Na D doublet. Concerning the first, we expected not to model the core of Hα because all cores in strong lines suffer from NLTE effects. *Turbospectrum* is a 1D, LTE radiative transfer code and therefore does not account for NLTE effects when producing the training set spectra. The wings of Hα are also poorly fit, and it is most probably due to our continuum normalisation scheme not being able to model non-linear spectral features such as the full Hα region of influence. Regarding the Na D lines, we realised it is common practice to always mask them out as they blend with Na lines in the ISM (Munari & Zwitter 1997).

- Another characteristic of the mask we tested was the width of the spectral window covering absorption lines of interest. It was not until we extended the number of labels to fit a larger number of chemical abundances that we realised TC systematically underestimated all of them. Narrowing the size of such window from 2 Å to 0.5 Å corrected this bias slightly (Figure 3.11). We are still not certain about what could be going on, yet one thing is clear, very few pixels from the full spectrum are required in the actual analysis.

- Continuum-normalisation is difficult to pin down. More accurately speaking, coding a functional continuum-normalisation for a general use is a hard task. The presence of numerous absorption lines, spectral features arising from the merging of orders in an echelle spectrograph and a wide range of SNRs can comfortably hide the true continuum pixels and completely shift the baseline. Given that TC shows better

results after a deliberate removal of pixels, the possibility of avoiding continuum-normalisation of the full spectrum is tempting. Since TC is effectively looking at narrow windows around the relevant absorption lines, normalising the individual absorption lines would likely suffice.

- The recovery of the three main stellar parameters is summarised in Table 4.1. However, we realised that it was worth excluding a handful of stars which TC classified as extremely metal-poor ([Fe/H] < -2). After removing these stars (Table 4.2), the SD of the label offsets for the three main labels are $\sigma_{\mathrm{Teff}} = 88$ K, $\sigma_{\log g} = 0.14$ dex and $\sigma_{[\mathrm{Fe/H}]} = 0.08$ dex. These uncertainties are comparable to those reported in the literature and hold (even improve) throughout a wide interval of SNRs. Measurements at low SNR are favoured by the statistics, though there is a population of stars with a systematically underestimated $\mathrm{T_{eff}}$ and $\log g$ probably due to bad performance of the continuum normalisation routine. At high SNRs the scatter in label offsets is not significantly improved even after the removal of the hot, fast rotators. This indicates that a more in depth examination of the stellar spectra involved is needed to identify potential reasons for large label offsets. Finally, medium SNR stars report the best results with practically unbiased results and SDs of $\sigma_{\mathrm{Teff}} = 75$ K, $\sigma_{\log g} = 0.12$ dex and $\sigma_{[\mathrm{Fe/H}]} = 0.07$ dex for the three main stellar parameters.

- Perhaps, a price to pay for the conscious removal of pixels from both our training and test spectra could be the loss of important information to help constrain our labels. In this sense, special emphasis must be made regarding our inability to replicate the wings of the Hα line. These are strongly dependent on the pressure and temperature of the atmospheric layers where they form and should be good tracers of $\mathrm{T_{eff}}$ and $\log g$ if modeled properly.

- The scientific value of our results is best represented by the 15 abundance plots from Figures D.9, D.10 and D.11. One would still need to carry out some extra tests to analyse the impact of the label space covered by the training set on the final abundance plots, e.g. whether we recover stars with high Ni abundance or not by artificially extending the label space of the training set in that direction. Nevertheless, we manage to reproduce similar trends with a significant reduced scatter in all 15 elements. It is in examining these similarities and differences with both training and test sets that we would be able to gain a better understanding of TC, the GES and the different catalogues we compared with.

Bibliography

Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012, A&A, 545, A32

Alvarez, R. & Plez, B. 1998, A&A, 330, 1109

Barklem, P. S. & O'Mara, B. J. 1998, MNRAS, 300, 863

Battistini, C. & Bensby, T. 2015, A&A, 577, A9

Battistini, C. & Bensby, T. 2016, A&A, 586, A49

Bensby, T. & Feltzing, S. 2006, MNRAS, 367, 1181

Bensby, T., Feltzing, S., & Lundström, I. 2003, A&A, 410, 527

Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71

Bland-Hawthorn, J. & Gerhard, O. 2016, ARA&A, 54, 529

Bovy, J. & Rix, H.-W. 2013, ApJ, 779, 115

Buder, S., Lind, K., Ness, M. K., et al. 2019, A&A, 624, A19

Burbidge, E. M., Burbidge, G. R., Fowler, W. A., & Hoyle, F. 1957, Rev. Mod. Phys., 29, 547

Casali, G., Magrini, L., Tognelli, E., et al. 2019, A&A, 629, A62

Casey, A. R., Hogg, D. W., Ness, M., et al. 2016, arXiv e-prints, arXiv:1603.03040

Clayton, D. 2003, Handbook of Isotopes in the Cosmos

de Jong, R. S., Barden, S., Bellido-Tirado, O., et al. 2014, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9147, 4MOST: 4-metre Multi-Object Spectroscopic Telescope, 91470M

de Laverny, P., Recio-Blanco, A., Worley, C. C., & Plez, B. 2012, A&A, 544, A126

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, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4008, Design, construction, and performance of UVES, the echelle spectrograph for the UT2 Kueyen Telescope at the ESO Paranal Observatory, ed. M. Iye & A. F. Moorwood, 534–545

Delgado Mena, E., Tsantaki, M., Adibekyan, V. Z., et al. 2017, A&A, 606, A94

Forsberg, R., Jönsson, H., Ryde, N., & Matteucci, F. 2019, A&A, 631, A113

Freeman, K. & Bland-Hawthorn, J. 2002, ARA&A, 40, 487

Fuhrmann, K. 2011, MNRAS, 414, 2893

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25

Gilmore, G. & Reid, N. 1983, MNRAS, 202, 1025

Gray, D. F. 2005, The Observation and Analysis of Stellar Photospheres

Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951

Hawkins, K., Jofré, P., Masseron, T., & Gilmore, G. 2015, MNRAS, 453, 758

Hayes, C. R., Majewski, S. R., Shetrone, M., et al. 2018, ApJ, 852, 49

Haywood, M., Di Matteo, P., Lehnert, M. D., et al. 2018, ApJ, 863, 113

Helmi, A., Veljanoski, J., Breddels, M. A., Tian, H., & Sales, L. V. 2017, A&A, 598, A58

Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2017, ApJ, 836, 5

Jofré, P., Heiter, U., & Soubiran, C. 2019, ARA&A, 57, 571

Jönsson, H., Ryde, N., Nordlander, T., et al. 2017a, A&A, 598, A100

Jönsson, H., Ryde, N., Schultheis, M., & Zoccali, M. 2017b, A&A, 598, A101

Kaler, J. B. 2011, Stars and their Spectra

Koppelman, H. H., Helmi, A., Massari, D., Price-Whelan, A. M., & Starkenburg, T. K. 2019, A&A, 631, L9

Kupka, F., Piskunov, N., Ryabchikova, T. A., Stempels, H. C., & Weiss, W. W. 1999, A&AS, 138, 119

Lindegren, L. & Feltzing, S. 2013, A&A, 553, A94

Lomaeva, M., Jönsson, H., Ryde, N., Schultheis, M., & Thorsbro, B. 2019, A&A, 625, A141

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94

Minchev, I., Martig, M., Streich, D., et al. 2015, ApJ, 804, L9

Mishenina, T., Pignatari, M., Gorbaneva, T., et al. 2019, MNRAS, 489, 1697

Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319

Munari, U. & Zwitter, T. 1997, A&A, 318, 269

Myeong, G. C., Vasiliev, E., Iorio, G., Evans, N. W., & Belokurov, V. 2019, MNRAS, 488, 1235

Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16

Ness, M., Rix, H. W., Hogg, D. W., et al. 2018, ApJ, 853, 198

Nissen, P. E. & Schuster, W. J. 2010, A&A, 511, L10

Piskunov, N. E., Kupka, F., Ryabchikova, T. A., Weiss, W. W., & Jeffery, C. S. 1995, A&AS, 112, 525

Plez, B. 2012, Turbospectrum: Code for spectral synthesis

Randich, S., Gilmore, G., & Gaia-ESO Consortium. 2013, The Messenger, 154, 47

Reddy, B. E., Lambert, D. L., & Allende Prieto, C. 2006, MNRAS, 367, 1329

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Rosswog, S., Sollerman, J., Feindt, U., et al. 2018, A&A, 615, A132

Ruchti, G. R., Feltzing, S., Lind, K., et al. 2016, MNRAS, 461, 2174

Ryabchikova, T., Piskunov, N., Kurucz, R. L., et al. 2015, Phys. Scr., 90, 054005

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

Sneden, C., Cowan, J. J., & Gallino, R. 2008, ARA&A, 46, 241

Tibshirani, R. 1996, Journal of the Royal Statistical Society: Series B (Methodological), 58, 267

Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2017, ApJ, 843, 32

Valenti, J. A. & Piskunov, N. 1996, A&AS, 118, 595

Appendix A

Continuum normalisation

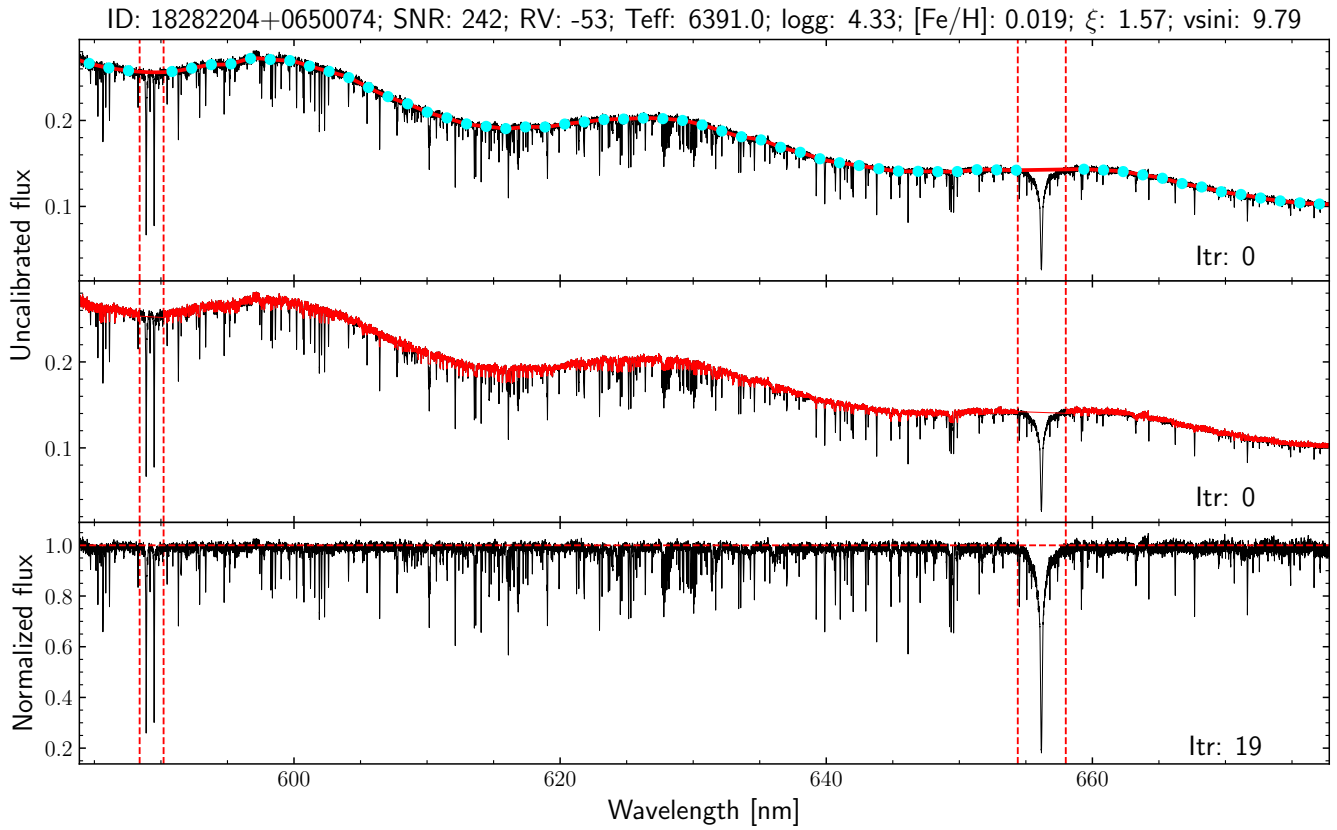
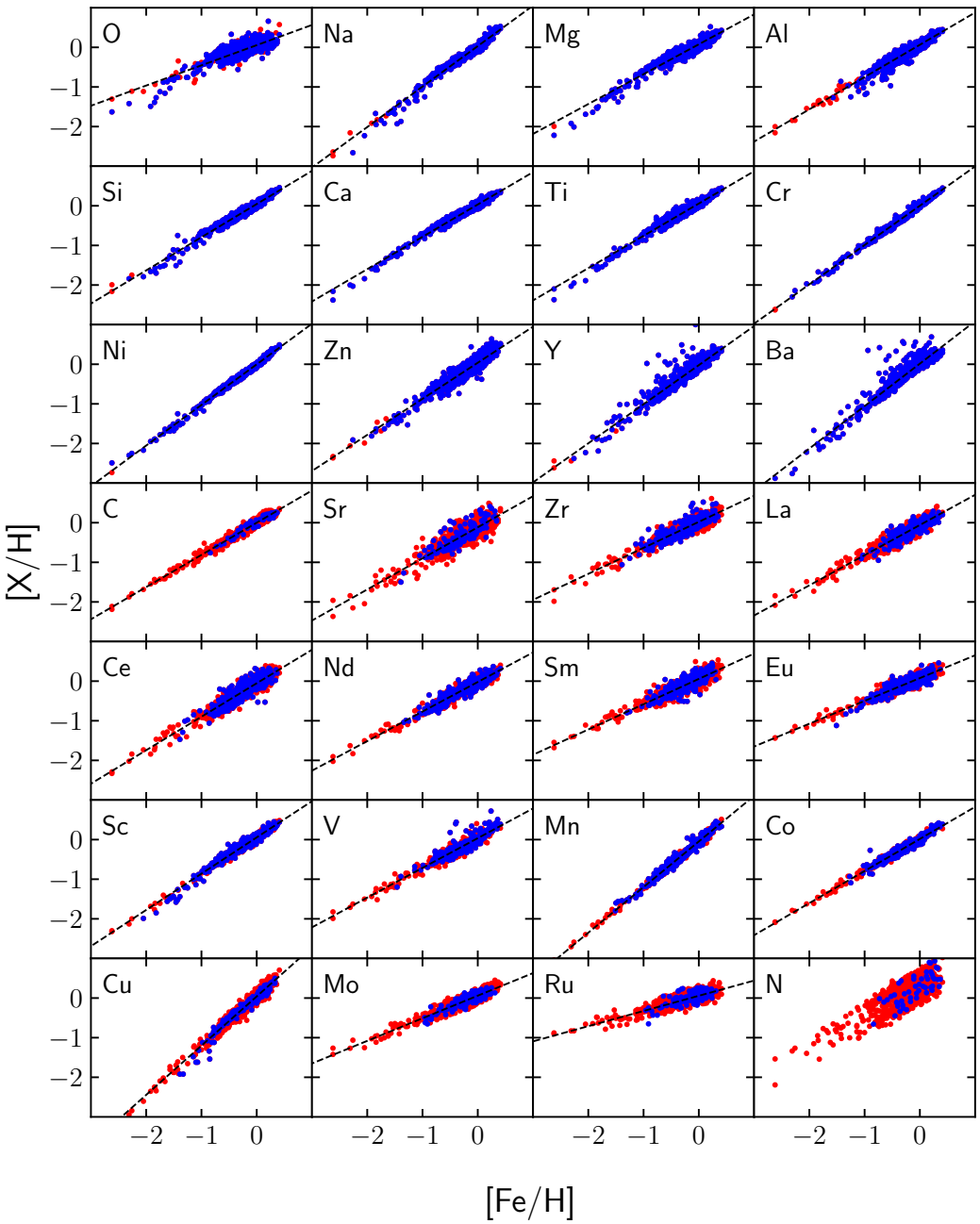


Figure A.1: Continuum normalization of GES star observed with the U580 UVES setting whose labels, SNR, RV and identifier are shown in the title of the figure. The top panel shows the spectrum before normalisation (zeroth iteration) and the fitting spline function in solid red with its nodes (cyan dots); the panel in between shows the same spectrum, with pixels in red indicating that they remain after the σ-clipping; the bottom panel contains the final version of the normalised spectrum, iteratively scaling the input spectrum by the spline fit. The dashed vertical red lines mask out the broadest lines in the piece of spectrum, namely, H_α and Na doublet.

Appendix B

Abundance plots in the training set

Figure B.1: [X/H] as a function of metallicity for all 28 elements with available observational data for the FGK type stars. Dots in blue represent stars with observed elemental abundances, cross-matched between the Bensby catalogue and the catalogue of stars corresponding to the element on the top left corner of each panel. Stars in red belong to the Bensby catalogue too, but are not picked in the cross-match and therefore have a synthetic abundance value. The dashed line shows the first order polynomial fit.



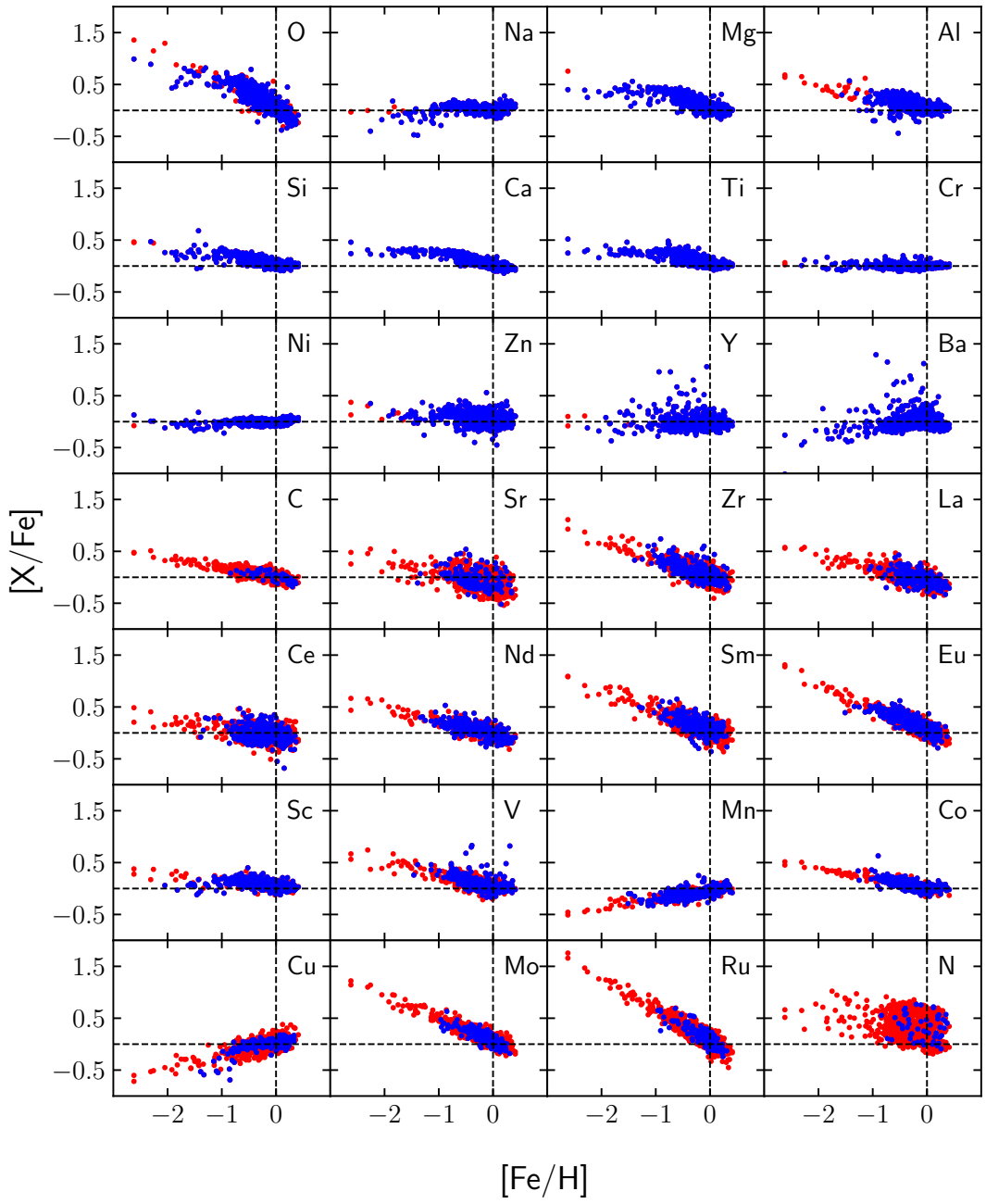


Figure B.2: [X/Fe]-[Fe/H] plots for all 28 elements. The same colour scheme as Figure B.1 is employed. Here, the dashed vertical and horizontal lines show the default solar metallicity and abundance value.

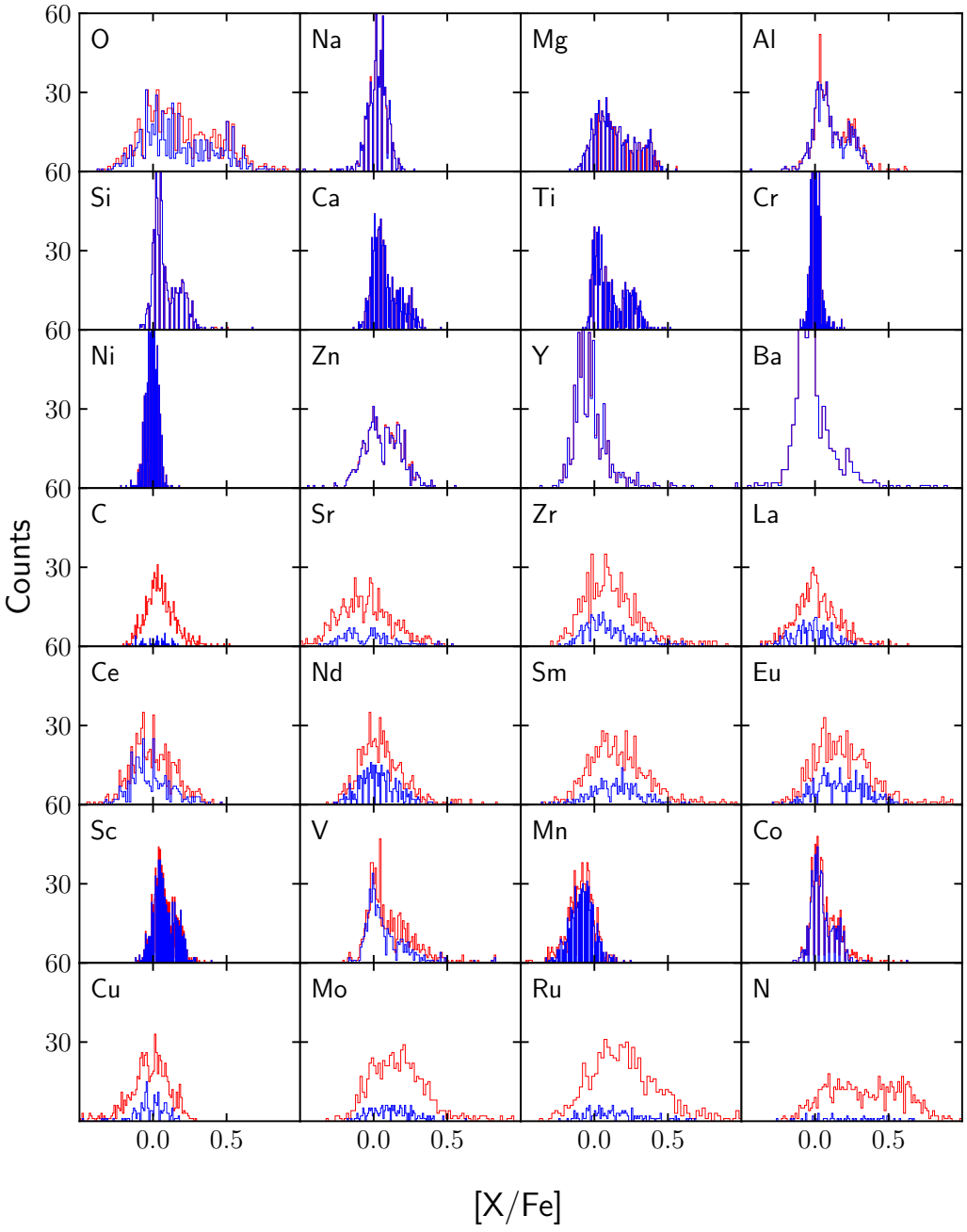


Figure B.3: Histogramed [X/Fe] ratios in 100 bins following the same colour scheme as in Figure B.1.

Appendix C

Fine tuning

ID	T_{eff}	$\log g$	[Fe/H]	ξ	$v \sin i$	[Mg/H]
00240628-7204370	5522.0	3.9	-0.87	0.81	9.58	-0.24
01402870-0059453	5699.0	3.95	0.08	1.18	2.8	0.18
03394563+0008017	5819.0	4.28	0.33	1.14	3.04	0.42
04394032-4603107	5989.0	4.37	-0.05	1.15	4.83	-0.19
04591551-5202392	5807.0	4.3	0.08	1.03	2.7	0.15
04595520-5159408	5892.0	4.34	0.01	1.16	2.0	0.09
05000360-5152370	6651.0	4.47	0.03	1.64	15.06	0.21
05202229-4653038	5854.0	4.18	0.07	1.03	4.78	0.13
06403720+0926513	6061.0	4.2	0.01	1.21	5.73	0.04
06410157+0915215	5386.0	3.9	0.27	1.06	2.89	0.44
06413715+0920599	5900.0	4.34	0.08	1.12	5.11	0.22
06440226-0101566	5622.0	4.44	0.14	1.08	4.0	0.21
07033617-4229583	6460.0	4.25	0.07	1.93	14.98	-0.09
07211345-0034280	5439.0	3.67	-0.28	1.12	4.64	-0.15
07444655-4235508	5891.0	3.77	-0.27	0.34	13.78	-0.07
08003075-0029292	5981.0	3.81	-0.15	1.43	5.22	-0.11
08093304-4737066	5640.0	4.26	-0.01	1.82	14.1	-0.18
08095427-4721419	5884.0	4.45	0.10	1.61	19.24	-0.77
09272162-0026042	5843.0	4.38	0.06	1.02	2.0	0.13
10484720-6403081	6844.0	4.51	-0.08	2.38	12.69	-0.05
10592101-4106265	5665.0	4.37	0.20	0.96	2.0	0.25
11002390-4101447	5279.0	4.55	-0.18	1.05	5.94	-0.1
11072859-5847592	5908.0	4.35	0.25	1.23	4.48	0.26
11073982-5842055	6146.0	4.61	-0.01	1.56	12.46	0.13
11083921-5845051	5955.0	4.56	-0.06	1.34	12.07	0.06
11124268-7722230	5226.0	3.94	-0.07	2.02	10.07	0.13
11214703-0957090	5750.0	4.39	-0.10	1.14	6.08	0.03
13011600-4101507	5935.0	4.24	-0.17	1.12	2.0	-0.08

13592093-4058230	5831.0	4.56	0.20	0.86	9.96	0.17
14002595-1003372	5659.0	3.81	-0.53	1.14	4.96	-0.27
14003318-4054310	5888.0	3.96	-0.26	1.15	3.33	-0.16
14595971-1955054	5852.0	4.27	0.14	1.01	4.73	0.18
15540366-4052141	5969.0	4.11	0.09	1.13	2.24	0.14
17345383-4300502	5993.0	4.06	0.13	1.24	4.41	0.17
17573649-3449364	5158.0	3.56	0.04	0.88	5.59	0.12
18182202-3252027	5054.0	3.54	-0.13	1.08	3.05	-0.05
18281038+0647407	5590.0	4.58	-0.48	0.69	11.91	-0.18
18282204+0650074	6391.0	4.33	0.02	1.57	9.79	0.17
22504769-5242498	5892.0	4.28	0.06	1.06	4.2	0.1

Table C.1: Main stellar labels for the sub-sample of 39 GES stars.

Appendix D

Final TC run

ID	SNR	$\mathrm{Teff}_{\mathrm{GES}}$	$\mathrm{Teff}_{\mathrm{TC}}$	$\log g_{\mathrm{GES}}$	$\log g_{\mathrm{TC}}$	$[\mathrm{Fe/H}]_{\mathrm{GES}}$	$[\mathrm{Fe/H}]_{\mathrm{TC}}$
03595585-2952287	73	5360	5366	4.4	4.49	-0.032	0.029
13395534-3412246	37	5579	5655	4.09	4.02	-0.43	-0.401
05121384-6155151	128	5405	5320	4.48	4.4	-0.09	-0.15
00295076-0334166	82	5639	5678	4.09	4.15	0.114	0.089
14002595-1003372	55	5918	5983	4.03	4.06	-0.487	-0.38
01505298-5007283	157	5267	5281	4.44	4.49	-0.031	-0.038
04302280-5008330	73	5553	5588	4.32	4.31	-0.172	-0.087
21200837-4804424	115	5857	5918	4.1	4.14	0.001	0.003
18492377-4228283	90	5913	6063	4.13	4.16	-0.466	-0.416
03200852-5558261	110	5644	5634	4.04	3.94	-0.003	0.021
10235288-3545538	200	5306	5341	4.46	4.56	-0.021	-0.054
02294307-0328342	161	5527	5448	4.42	4.37	0.171	0.103
04200948-0015371	104	5496	5505	4.31	4.29	-0.051	-0.089
10592101-4106265	112	5430	5430	4.33	4.4	-0.151	-0.15
15295442-1957254	60	5809	5803	4.11	3.9	-0.228	-0.326
10403923-1807097	135	5990	5976	3.88	3.53	-0.467	-0.528
14193707-3703162	225	5061	5055	4.45	4.51	0.04	-0.003
20590947-5459144	43	5907	5909	4.12	4.02	0.171	0.096
05102002-6152039	160	5419	5507	4.37	4.55	0.302	0.302
07452778-4227198	44	5681	5632	4.28	4.24	0.305	0.234
23201331-5451382	46	5514	5497	4.44	4.41	0.158	0.1
10230126-3535321	69	5356	5341	4.36	4.41	0.259	0.191
09202790-1758277	98	5659	5634	4.42	4.35	-0.135	-0.177
08393608-0030520	64	5752	5752	4.04	3.86	-0.439	-0.526
08093304-4737066	119	6034	6059	3.94	3.82	-0.149	-0.154

Table D.1: Main stellar labels plus ID and SNR for 25 randomly picked stars from the full test set of 1410 GES stars. The GES and TC values are listed for comparison purposes.

Figure D.1: Adaptation of Figure 4.3 only including stars in the low SNR bin.

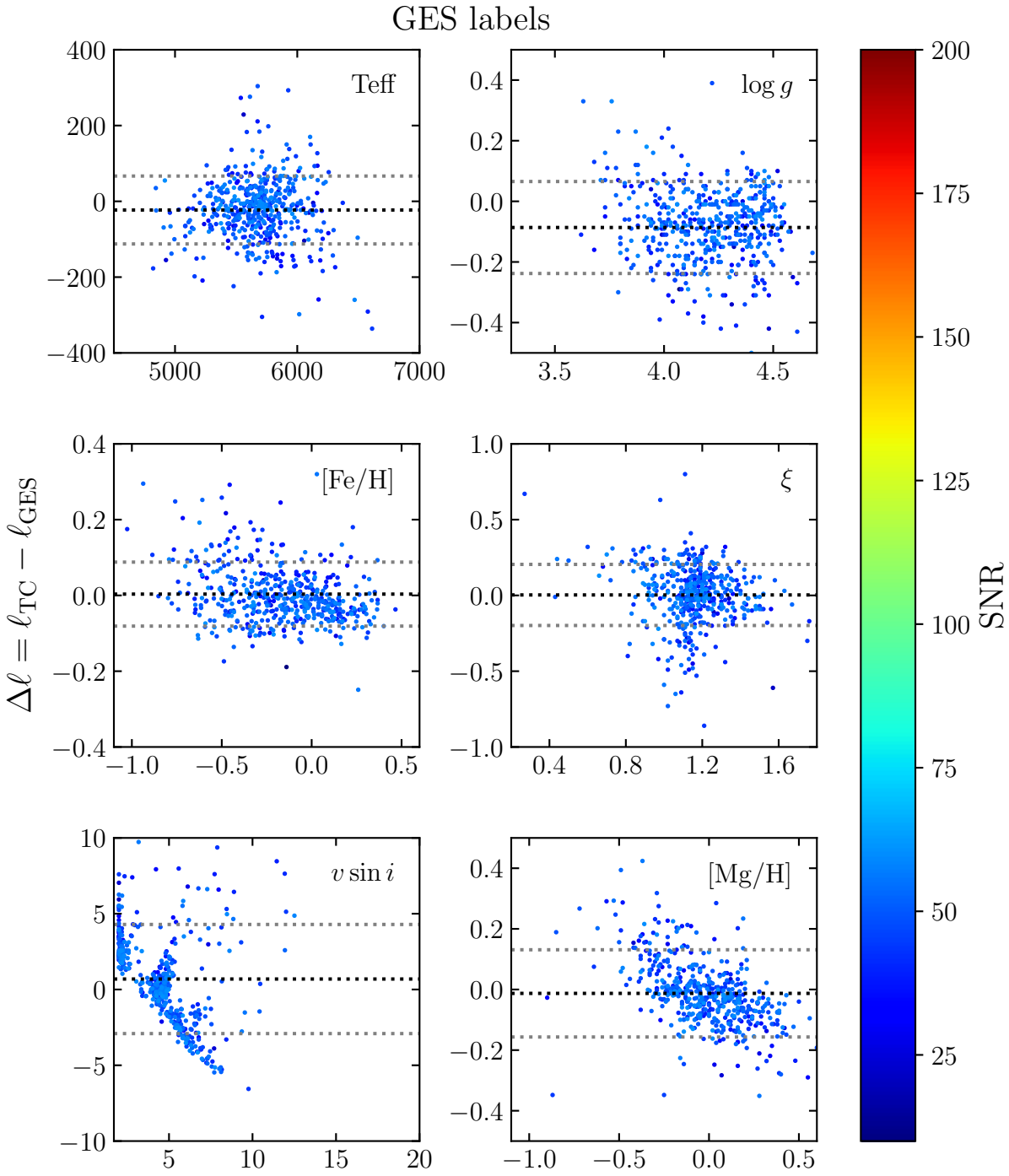


Figure D.2: Adaptation of Figure 4.3 only including stars in the medium SNR bin.

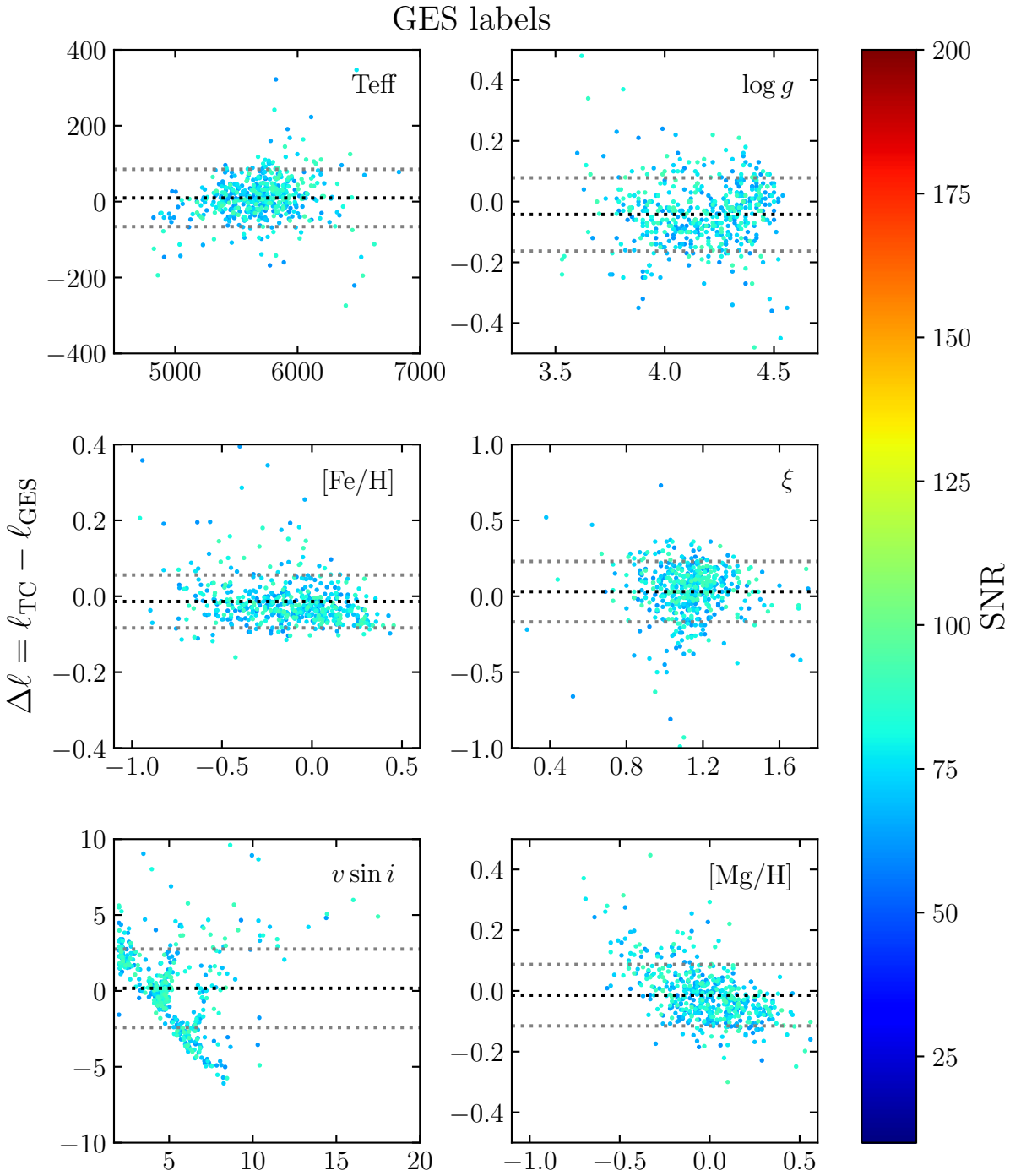


Figure D.3: Adaptation of Figure 4.3 only including stars in the high SNR bin.

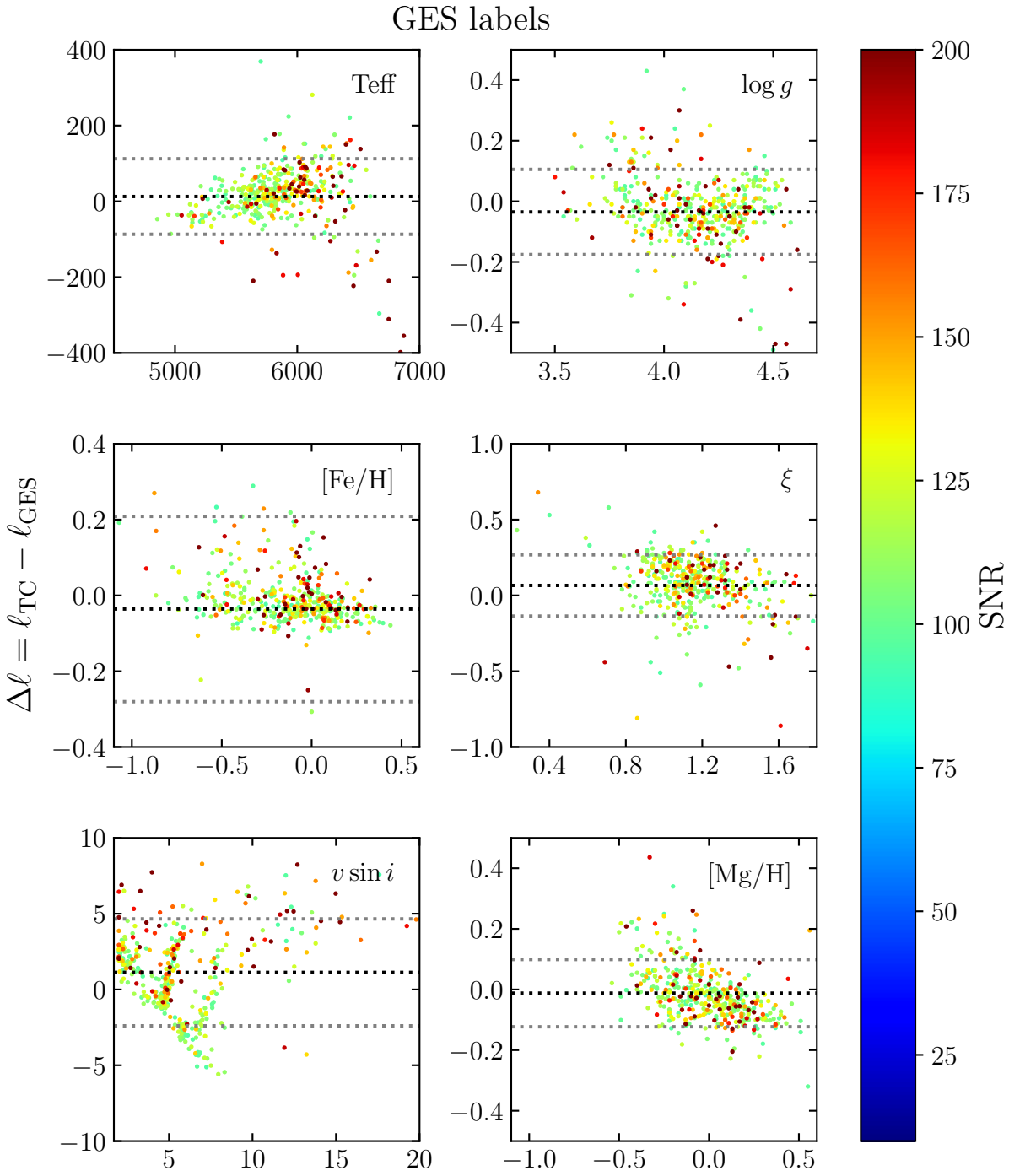


Figure D.4: Histograms of the label offsets in Figure D.1, with their mean and SD.

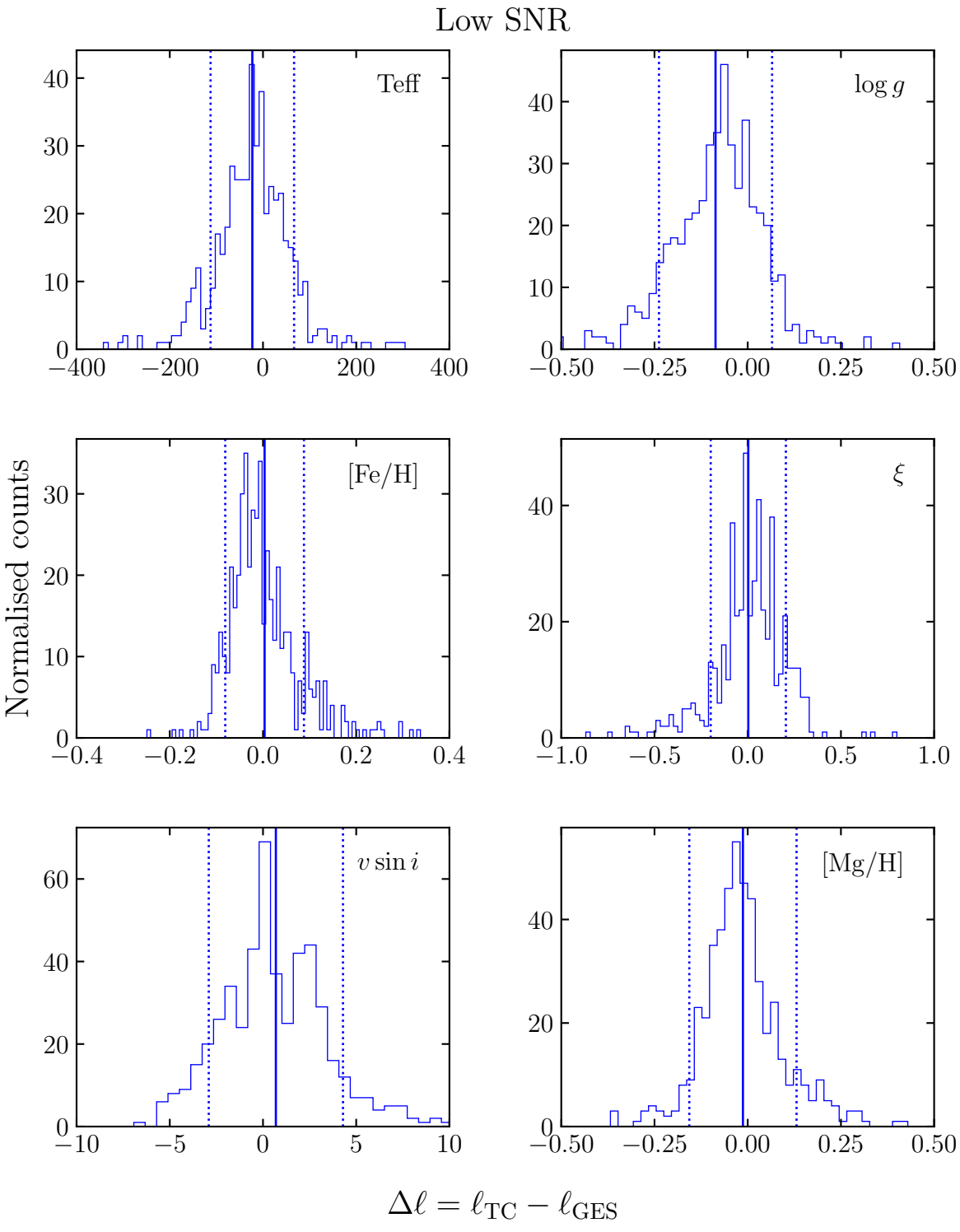


Figure D.5: Histograms of the label offsets in Figure D.2, with their mean and SD.

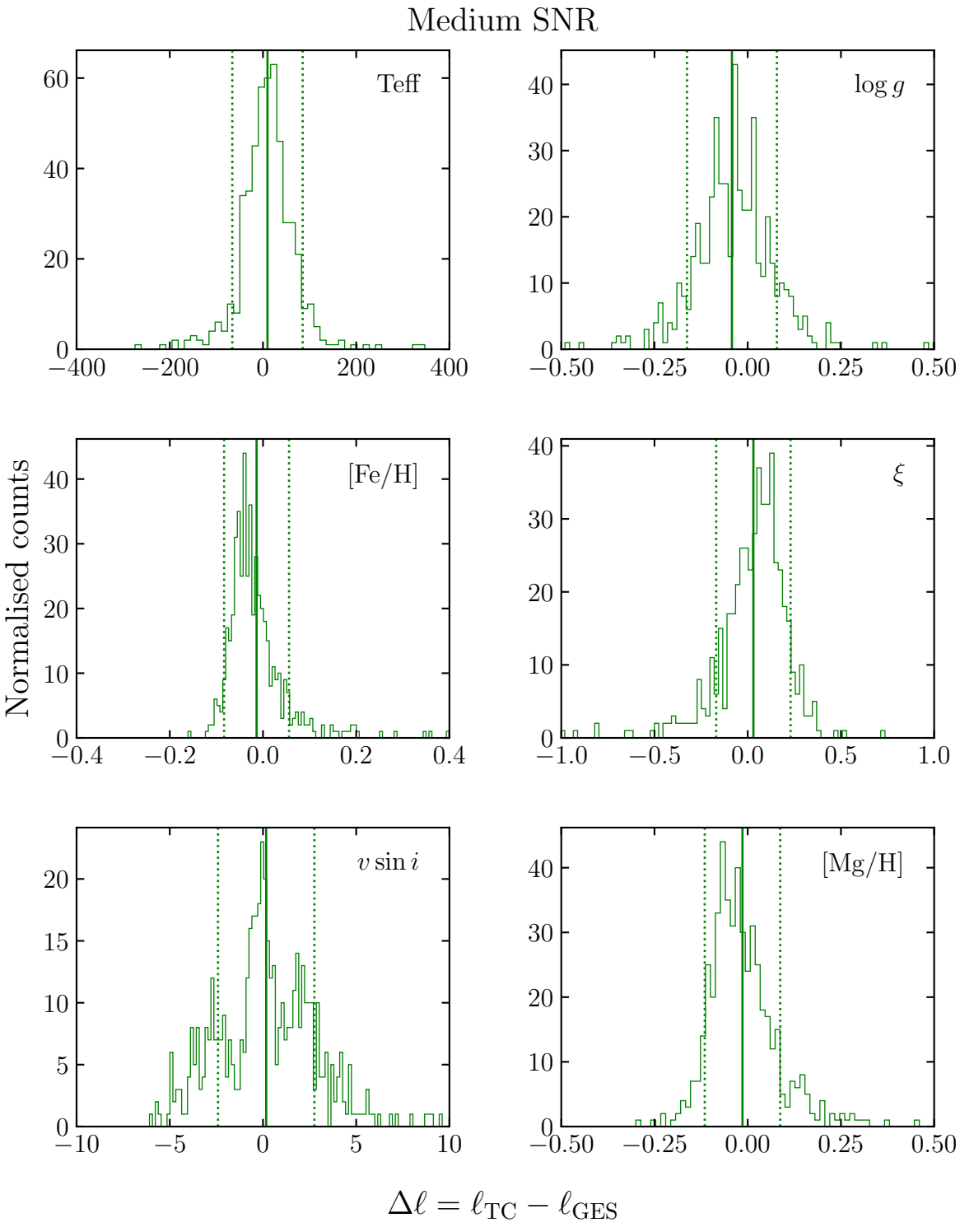


Figure D.6: Histograms of the label offsets in Figure D.3, with their mean and SD.

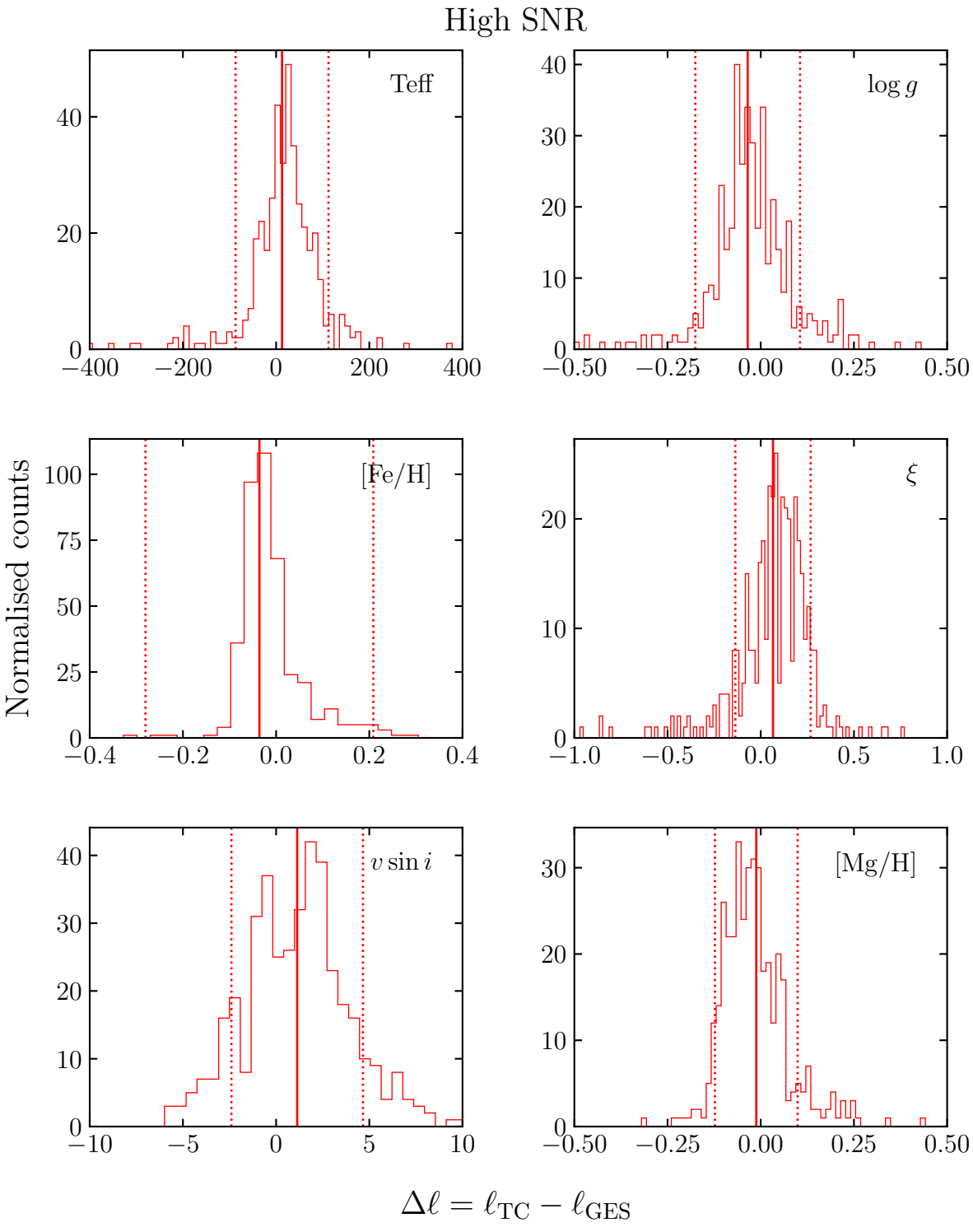


Figure D.7: 6-panel scatter plot where TC labels are represented versus the GES labels for all 1410 test stars, colour-coded by SNR.

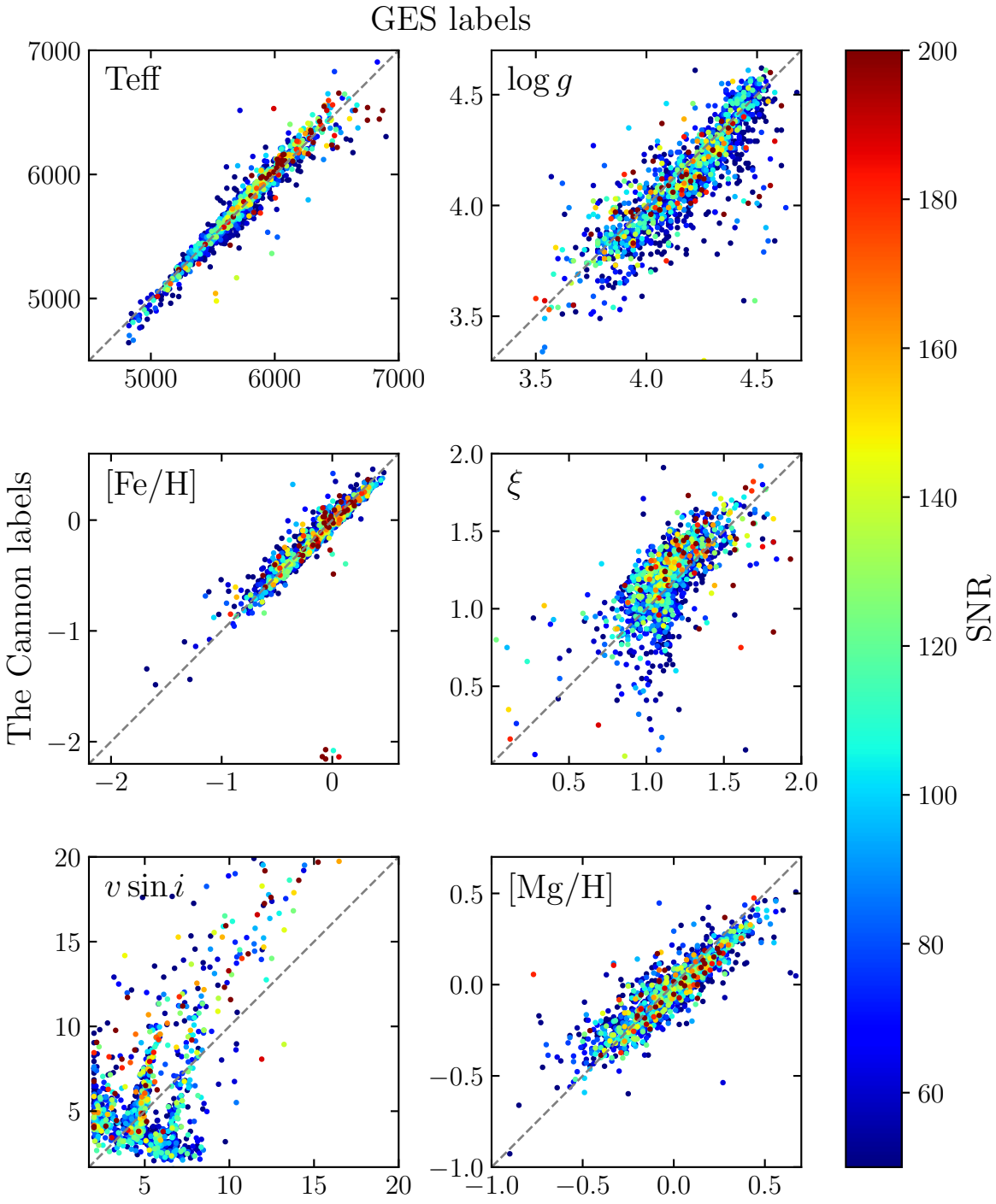
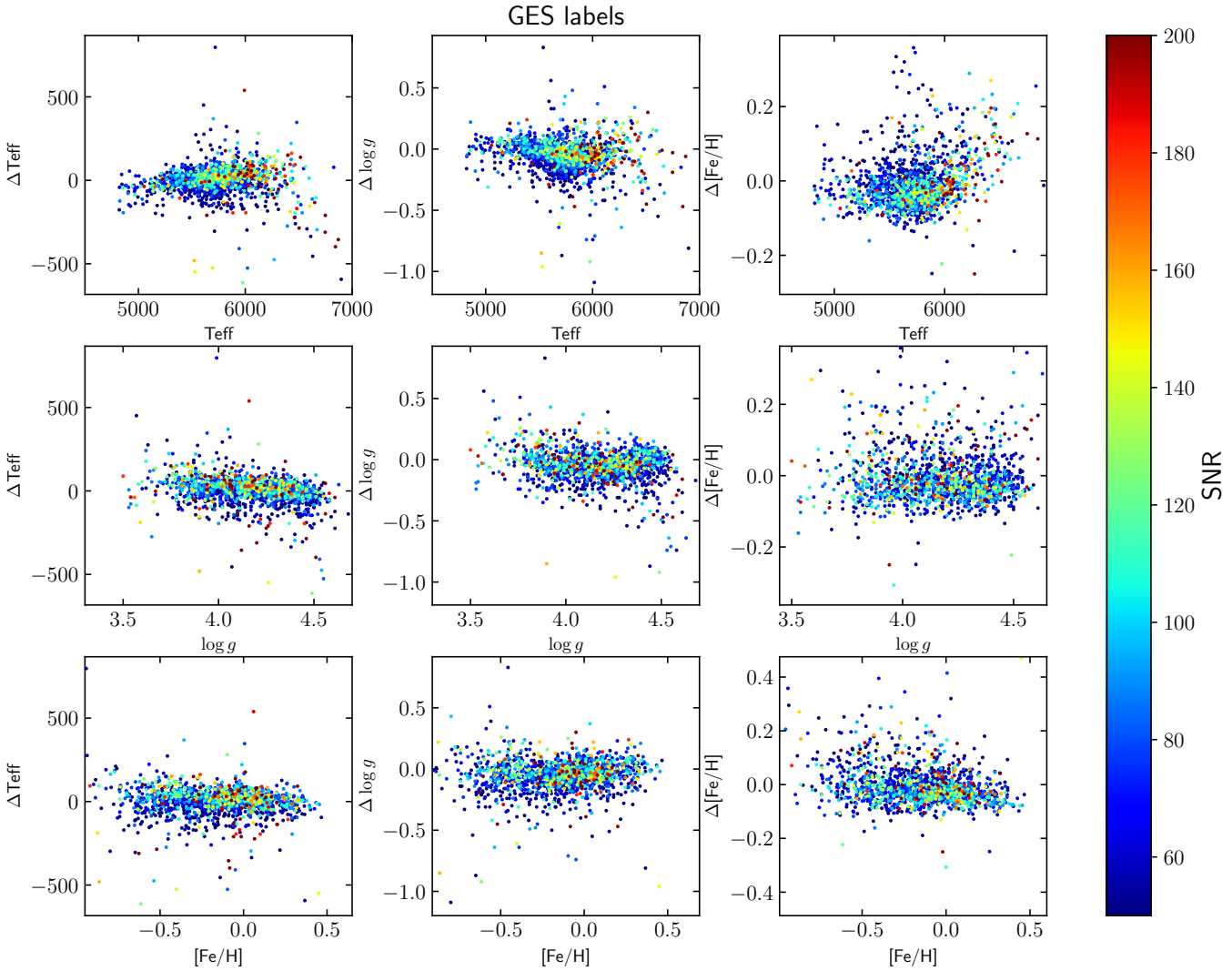


Figure D.8: Correlation between labels offsets and GES values for T_{eff}, $\log g$ and metallicity in the test set. Again, dots are coloured in accordance with their SNR.



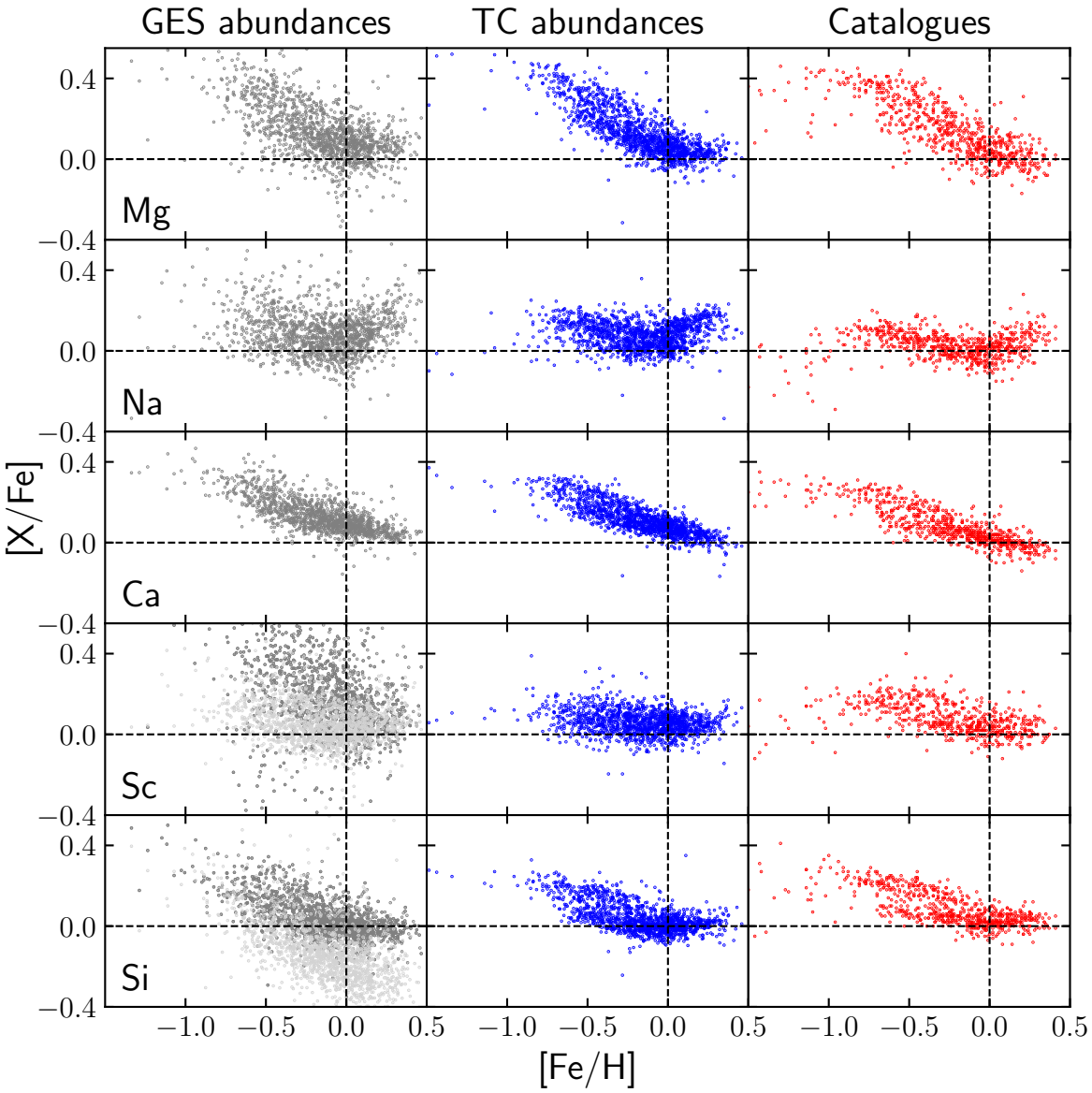


Figure D.9: [X/Fe]-[Fe/H] comparison figure among GES and TC abundances for the test stars and the reported abundances from the catalogues in Table 3.1 for the relevant elements. Mg, Na, Ca, Si, Ti, Ni, Cr, Ba, O and Al are studied in Bensby et al. (2014), Sc, V, Mn and Co in Battistini & Bensby (2015) and Eu in Battistini & Bensby (2016). Sc, Si and Ti have a second set of light grey dots added in their GES abundance panels. In these three cases, dark grey dots correspond to abundances calculated from transitions within the neutral species, whereas light grey dots stand for abundances determined from transitions from the ionised sibling. Missing GES abundance panels for Eu and O just represent the absence of [Eu/Fe] and [O/Fe] abundances for all these 1410 stars.

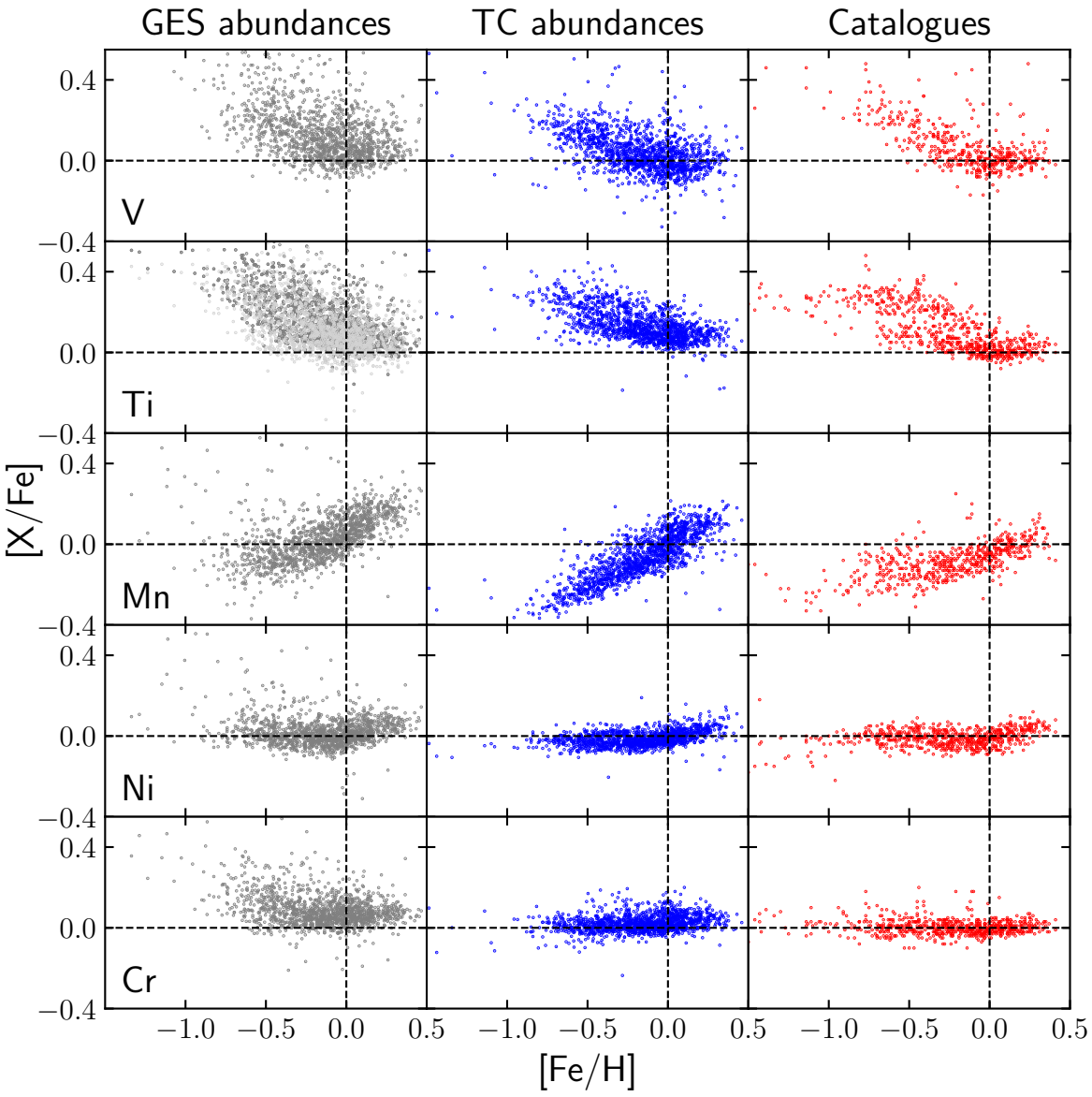


Figure D.10: Continuation of Figure D.9.

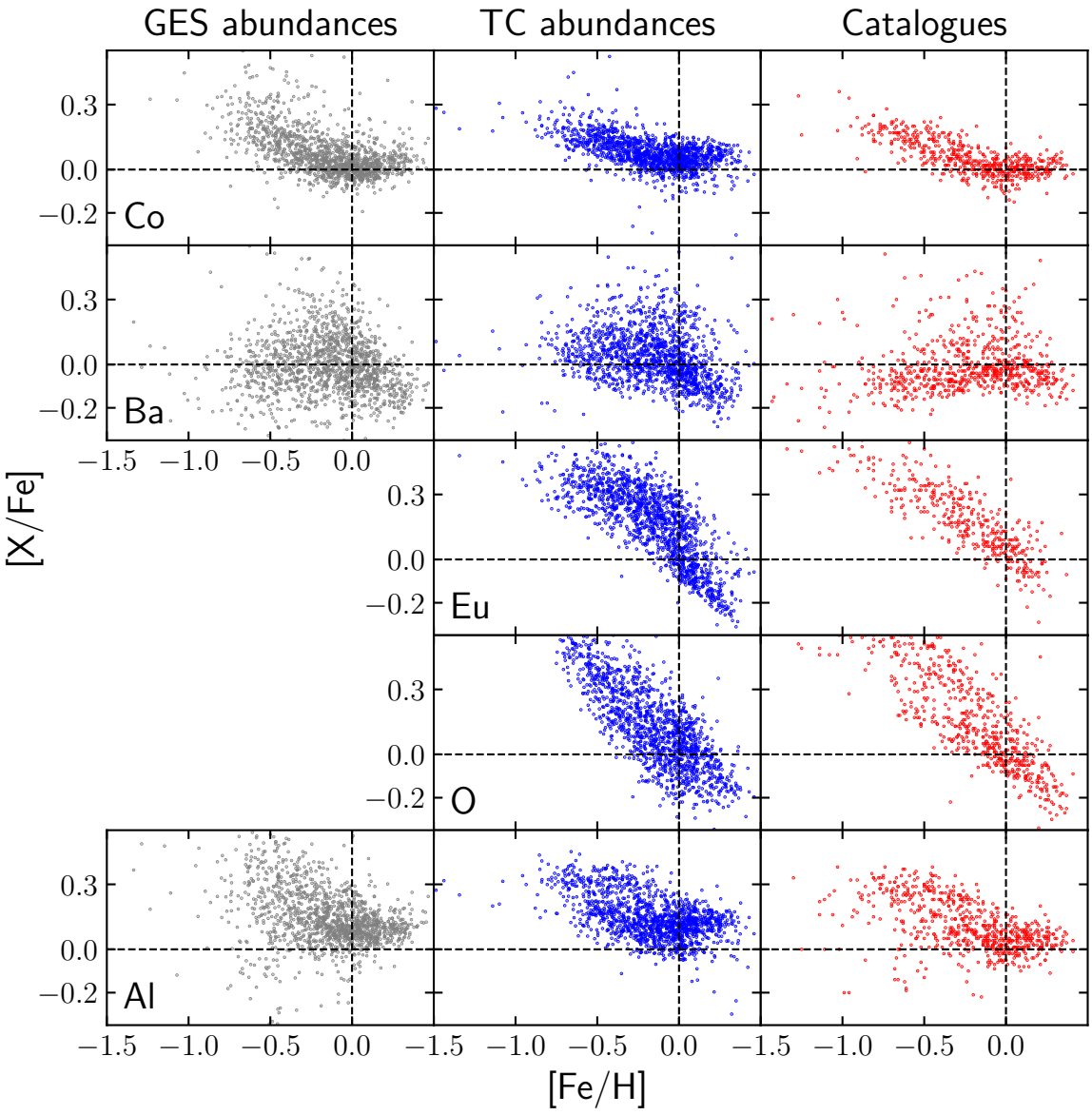


Figure D.11: Continuation of Figure D.9.

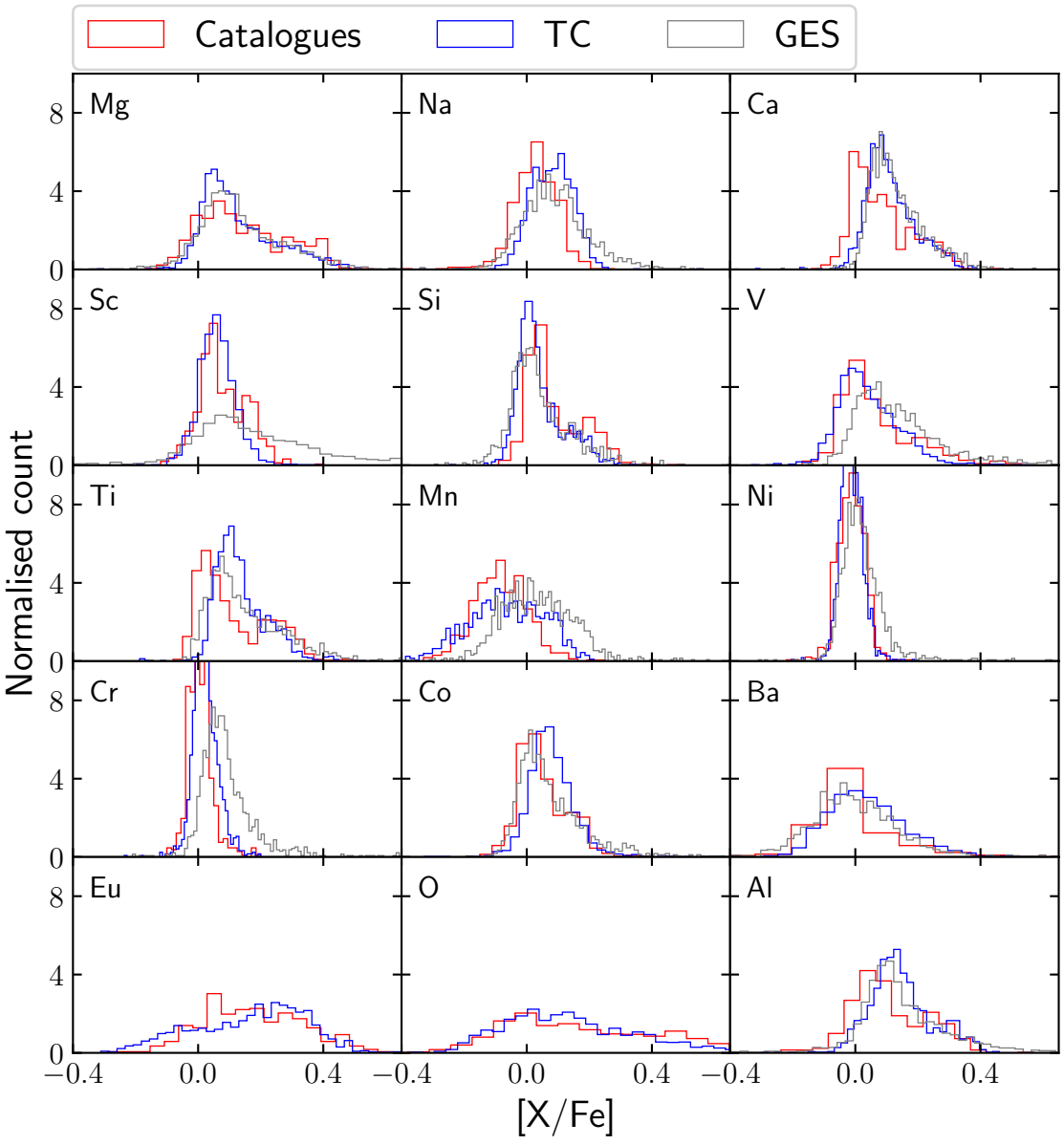


Figure D.12: Normalised histograms of TC, GES and catalogued results for all 15 elements in Figures D.9, D.10 and D.11