

UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS
DEPARTMENT OF PHYSICS

Klemen Čotar

**Understanding of open stellar clusters and peculiar
spectra in sky surveys using machine learning**

DOCTORAL THESIS

ADVISER: prof. dr. Tomaž Zwitter

Ljubljana, 2020

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
ODDELEK ZA FIZIKO

Klemen Čotar

**Razumevanje razsutih zvezdnih kopic in posebnih
tipov spektrov v pregledih neba z uporabo
strojnega učenja**

DOKTORSKA DISERTACIJA

MENTOR: prof. dr. Tomaž Zwitter

Ljubljana, 2020

Zahvala

Na tem mestu zapišite, komu se zahvaljujete za pomoč pri nastanku doktorske diplome.

**Razumevanje razsutih zvezdnih kopic in posebnih tipov spektrov v
pregledih neba z uporabo strojnega učenja**

IZVLEČEK

Kratek izvleček v slovenskem jeziku, do 300 besed.

Ključne besede: metode: analiza podatkov – razsute kopice in asociacije: splošno – Galaksija: zvezdne populacije – zvezde: zastopanosti – zvezde: hitrosti in premikanje – dvojnice: splošno – zvezde: posebni tipi – zvezde: karbonske – zvezde: aktivnost – zvezde: emisijske črte – zvezde: podobne Soncu – črte: profili – katalogi

PACS: 95.10.Eg, 95.75.De, 95.75.Fg, 95.75.Mn, 95.80.+p, 97.10.-q, 97.10.Ri, 97.10.Tk, 97.10.Vm, 97.21.+a, 97.30.Eh, 97.30.Fi, 97.80.-d, 97.80.Fk, 98.58.H

Understanding of open stellar clusters and peculiar spectra in sky surveys using machine learning

ABSTRACT

Kratek izvleček v angleškem jeziku, do 300 besed.

Keywords: methods: data analysis – open clusters and associations: general – Galaxy: stellar content – stars: abundances – stars: kinematics and dynamics – binaries: general – stars: peculiar – stars: carbon – stars: activity – stars: emission-line – stars: solar-type – line: profiles – catalogues

PACS: 95.10.Eg, 95.75.De, 95.75.Fg, 95.75.Mn, 95.80.+p, 97.10.-q, 97.10.Ri, 97.10.Tk, 97.10.Vm, 97.21.+a, 97.30.Eh, 97.30.Fi, 97.80.-d, 97.80.Fk, 98.58.Hf

Contents

1	Introduction	15
1.1	Open clusters in the <i>Gaia</i> era	16
1.2	Chemical tagging	17
1.3	Peculiar stellar spectra	18
1.4	Machine learning in large sky surveys	18
1.5	Our exploration of observed data	20
2	Spectroscopic, photometric, and astrometric surveys	23
2.1	Gaia	23
2.1.1	Photometry and astrometry	25
2.1.2	Spectroscopy	25
2.2	<i>Gaia</i> DR2	25
2.3	GALAH	28
2.3.1	Acquired spectra and target selection	28
2.3.2	Spectral reduction and parameters determination	29
2.4	Asiago	30
3	Chemo-dynamic tracing of open cluster stars	33
4	Chemically peculiar stars	35
4.1	Abstract	35
4.2	Introduction	35
4.3	Data	36
4.4	Detection procedure	38
4.4.1	Supervised classification	38
4.4.2	Unsupervised classification	42
4.5	Candidate characteristics	45
4.5.1	Radial velocity variations	45
4.5.2	Stellar parameters	46
4.5.3	S-process elements	48
4.5.4	Lithium abundance	48
4.5.5	Sub-classes	49
4.5.6	Match with other catalogues	49
4.6	Metal-poor candidates	52
4.7	Follow-up observation	53
4.8	Conclusions	56
4.9	Table description	57
5	Peculiar emission stars	63

6 Peculiar solar-like multiple stars	65
6.1 Introduction	65
6.2 Selection of the best solar-like spectra	66
6.2.1 Reference Solar spectrum	66
6.2.2 Stellar spectra preprocessing	66
6.2.3 Candidate selection	67
6.2.4 Spectral similarity	68
6.2.5 Search for Solar twin spectra	72
6.3 Comparison with the literature	79
6.3.1 18 Sco	79
6.3.2 ELodie archive	80
6.3.3 Ramirez study	80
6.3.4 Datson study	80
6.3.5 Santos study	80
6.3.6 M67 solar twins	80
6.4 Ponovljena opazovanja	80
6.5 Follow-up observation	80
6.6 Results	81
6.6.1 Atmospheric parameters	81
6.6.2 Spatial distribution	82
6.7 Abstract	84
6.8 Introduction	84
6.9 Data	85
6.10 Solar-like spectra	88
6.10.1 Candidate multiple systems	89
6.11 Single star models	90
6.11.1 Spectroscopic model	92
6.11.2 Photometric model	92
6.11.3 Limitations in the parameter space	94
6.12 Characterization of multiple system candidates	94
6.12.1 Fitting procedure	94
6.12.2 Photometric fitting - first step	95
6.12.3 Spectroscopic fitting - second step	97
6.12.4 Final fit - third step	98
6.12.5 Number of stellar components - final classification	98
6.12.6 Quality flags	99
6.13 Characterization of single star candidates	101
6.14 Orbital period constraints	101
6.14.1 Outer pair and <i>Gaia</i> angular resolution	102
6.14.2 Inner binary pair and formation of double lines in a spectrum	104
6.14.3 Multi-epoch radial velocities	106
6.15 Simulations and tests	109
6.15.1 Radial velocity separation between components	109
6.15.2 Analysis of synthetic multiple systems	110
6.15.3 Triple stars across the H-R diagram	111
6.15.4 Observational bias - Galaxia model	114
6.16 Conclusions	115
6.17 Table description and summary	115

6.18 Additional figures	118
7 Conclusions	123
Bibliography	125
Appendix A Naslov prvega dodatka	159
Razširjeni povzetek v slovenskem jeziku	161
List of publications related to this doctoral thesis	163

Chapter 1

Introduction

Observational studies and simulations show that stars in our Galaxy were not formed at the same time, but at multiple epochs in different places of the Galaxy [1, 2]. One of the youngest building blocks of the Galaxy are open stellar clusters whose stars formed from the same molecular cloud of material [3] and therefore retain some properties of the original cloud they were made from. Stellar properties of an individual star can be separated into a kinematic and chemical component. The first is describing position and movement of a star in the Galaxy and the second its chemical composition and physical properties. During the evolution of the Galaxy, clusters that have lower number of members and are therefore gravitationally loosely bound, slowly evaporate because of different effect such as gravitational dynamical friction and ejection of stars during close intra-cluster stellar interactions. The latter is a result of a close gravitational interactions among members, where one of them can be ejected out of a cluster at high velocities [4, 5, 6]. Such past members can on the sky be found even several degrees away from their main cluster body [7, 8, 9]. The gravitational dynamical friction happens when a cluster moves through regions of the Galaxy with higher destiny of stars [10]. During such event, gravity of a cluster starts pulling seemingly stationary stars around it. Considering energy and momentum conservation law, we can conclude that a cluster will be slowed down for the same amount. This subsequently changes orbit of individual cluster members around the center of the Galaxy. Small orbital changes lead to a gradual expansion of a cluster volume that can be tracked until its individual stars blend with a general stellar field population, making them unrecognizable as past cluster members. The most prominent transitional features we can observe are compact cluster tidal tails [11, 12, 13, 14] and loose extended halos of evaporated stars. The halos can be observed as a slowly decreasing over-density [15, 16, 17] of stars far from a denser main cluster core.

Born from the same molecular cloud, open clusters are therefore ideal tracers of formation, assembly and evolutionary history of their host galaxies. Being influenced by external and internal processes, their lifetime is limited from about 100 Myr to a few Gyr for the densest structures [18, 19]. Studies have shown that the dissolution time of a cluster depends mainly on its total mass, its radius and its galactic environment (density of the stars around it and on its galactic path). Many papers have so far dealt with the question of determining precise lifetime of stellar clusters before they completely dissipate into field population. The problem has been tackled using direct N-body simulations [18] where movement of individual stars was traced, and

from direct observational data [20, 21, 22, 23] using isochrone fits. Age distribution of open clusters within a distance of 1 kpc from the Sun shows a cluster median lifetime of 200 Myr. Their broad range in ages gives us a possibility of observing them at different evolutionary stages [24, 25] before they blend [26] into a field stellar population. On the other hand, such a short lifetime (in comparison with a lifetime of a Sun-like star) gives us a limited number of bounded clusters in the sky that can be studied at a given time.

The processes described above heavily depend on the mass of a cluster, its internal structure and mass distribution of its components. When star formation ceases, we are left with a wide range of stellar masses, their distribution can be described by an initial mass function (IMF) [27, 28, 29] that appears invariant among clusters and even stars in the field [30]. The initial spatial distribution of stars in young clusters may reflect the structure of parental molecular cloud [31]. In many stellar clusters, the brightest and most massive stars are concentrated towards the center of a cluster, this state is usually attributed to mass segregation. Whether mass segregation occurs due to an evolutionary effect or it is of primordial origin is not yet entirely clear [32, 33, 34, 35]. In the first case, massive stars are formed all around the cluster volume and eventually sank to its center through the effect of two-body relaxations. In the second scenario, massive stars form preferentially in the central region of a cluster either by gas accretion due to their favorable location at the bottom of a gravitational potential well or through a coalescence process of less massive stars. The fact that mass segregation is also observed in young clusters might suggest that the second scenario is more likely, but the question is still under investigation [36].

Similarly to IMF, we can also define initial binary population (IBP) of observed sample of stars. Characterizing fraction of binaries in stellar clusters is of great importance for many fields of astrophysics. Since binaries are on average more massive than single stars they are thought to be good tracers of dynamical mass segregation [37]. Comparing observations with theoretical model for radial distribution of binary systems and their properties (mass and luminosity ratio, orbital period) distribution can be used to asses dynamical state of a cluster [38].

1.1 Open clusters in the *Gaia* era

Since *Gaia* Data Processing and Analysis Consortium (DPAC) centers responsible for analysis of the *Gaia* observations started publishing publically available ready-to-use stellar properties, many new discoveries and insights have been made. For the first time in history, we reliably know position, distance, complete spatial velocity (proper motion and radial velocity) and luminosity for millions of bright and dim stars all across the observable sky. One of fields in astronomy that heavily benefited from those new and improved measurements was research of open stellar clusters.

A traditional historical way to look for open stellar clusters was through counting stars as they are apparently seen on the sky from Earth's location and finding over-densities in those counts [39, 40]. To perform a more robust selection, members are additionally filtered based on their apparent distance from the cluster center and their motion vectors [41]. Using the latest second release of *Gaia* data [DR2, 42], we can go beyond that and build upon the results achieved by the aforementioned methods. Complete *Gaia* information on stellar distance, kinematics and photomet-

ric measurements enables us to go beyond simple methodologies, to unravel even the faintest and sparsest components of open clusters. So far, many works have been published trying to refine parameters and membership information of long known open clusters [41, 43, 44] and find new, less numerous or fainter clusters [45, 46, 47, 48, 49]. Such thorough and improved investigation uncovered that many of the clusters listed in modern catalogues, initially discovered as apparent stellar overdensities, are no more than chance alignments of stars and not true physically bound clusters [50, 51, 52, 53, 54].

Precisely determined open cluster membership also enables accurate age determination of a cluster and its stars [55]. When observing stars in any random stellar field it is notoriously hard to determine their ages as they remain unchanged for majority of its lifetime. While it is possible to obtain precise age estimates for some specific classes of stars [56], an uniform methodology does not exist for all. This is much easier to determine for stars in a cluster as they are all of a very similar age. Stellar evolution and its cycle is determined mainly by the initial mass of a star. Observing stars of different masses in a cluster reveals its evolutionary stage and consequently also its age by comparing them to theoretical evolutionary models. The latest *Gaia* data enabled researches to compute stellar absolute magnitudes and colours that are commonly used in those comparisons [22, 23, 57].

Another advantage of the newest *Gaia* dataset, with accurately determined distance to objects, is possibility of determining actual shape of a cluster, mass segregation of member stars and their gravitational potential from the first principles. Until recently that was possible only trough n-body simulations [36, 58, 59] that were compared with available observational datasets. For distant stellar associations, the distance error bars are still larger than a cluster itself, but improvements are expected when *Gaia* EDR3 will be released in the second half of the year 2020 or later. Internal cluster dynamics can in those cases be inferred using precisely determined radial velocities whose uncertainty is not affected by stellar distance, but their apparent brightness. Reliable determination of stellar mass distribution in a cluster also depends on accurate classification of binary or higher multiplicity stars and their masses. By using pure *Gaia* information, they can be identified trough uncertain radial velocity [60], its temporal variation [61], using precise absolute magnitudes [62, 63, 64] or exploring astrometric uncertainties [65] that indicate movement of a combined photocentre.

1.2 Chemical tagging

Majority of previously described works in Section 1.1 relies on a complete 6D positional and kinematics information to discover clusters and their sub-structures. Advances in observational techniques and data analysis enables us to go beyond kinematics information and include multidimensional chemical signature of stars – procedure known as chemical tagging [66, 67]. So far a blind chemical tagging (without kinematics) that would delineate between cluster and field stars has not yet been demonstrated with great success, unless observed structure has obviously different chemical composition [68]. A trait that it is not common to open stellar clusters formed at about the same time [69], but to galactic components formed at vastly different epochs [70] of Galaxy formation.

Latest research showed that many, even unexpected, observational and data

reduction issues still have to be thoroughly investigated and resolved, especially if data from different surveys are to be combined [71]. Studies suggest that open clusters might not be as homogeneous as thought before [8, 72], and abundances show traits of stellar evolution [73, 74, 75]. On top of that, main concerns of chemical tagging are abundance trends assumingly induced by spectral analysis [72, 76, 77]. Observed trends depend on determined stellar physical parameters (i.e. T_{eff} and $v \sin i$) and might be results of inadequate stellar models or actual stellar processes. To cope with this complexity and uncertainties, complex Bayesian models are being developed [72, 78] in order to uncover and cluster abundance patterns.

1.3 Peculiar stellar spectra

Observational difficulties and data reduction issues are not the only obstacles that have to be controlled in order to perform successful blind chemical tagging. In every larger observational sample of stars, we can find a small sub-sample of stars with properties that deviate from the majority of randomly selected stellar population. Such stars can be identified by having kinematic properties vastly different than their local volume of space [79, 80, 81], having distinctly unusual chemical abundances [82] or have spectral features that are seen in a small percentage of stars [83, 84]. Among them, we also count sources on the sky whose observables show that they are a composite of two or more stellar components [85, 86, 87]. In the opposite context, we commonly refer to majority of spectra as normal, but the delineation between classes is not strict and may depend on considered scientific question.

Every peculiarity in observed spectra could potentially influence derived stellar physical parameters and consequently also chemical abundances as they all depend on comparison between observed spectra and computed synthetic spectra that rely on physical models of stellar interior. As the exploration of galactic history, also known as the galactic archaeology, relies on normal stars for which their parameters can be reliably measured, we would like our set of analysed stars to be as clean as possible. The sets are usually delineated by the procedure known as classification. This refers to a procedure that sorts objects into categories and labels them according to their properties. Those labels can reflect actual physical properties and nature of stars or uncover unexpected features in observed data. Anyhow, both normal and peculiar stars are important for understanding large-scale galactic dynamics, internal structure of stars and their life cycle.

1.4 Machine learning in large sky surveys

In the past decades, we witnessed a major change in many areas of scientific research, including the field of astronomy. New observational methods and advance scientific experiments are serving us ever increasing amounts of data that have to be analysed by researchers. For example, astronomy have seen a shift from painstaking and time-consuming observations of a single star and its spectra to fiber-fed spectrographs that are able to simultaneously acquire spectra of hundreds to thousands stars in a comparable time. Among the studies dedicated to exploration of stellar objects we can find: Sloan Extension for Galactic Understanding and Exploration (SEGUE [88]), Apache Point Observatory Galactic Evolution Experiment

(APOGEE [89]), RAdial Velocity Experiment (RAVE [90]), Gaia-ESO Survey (GES [91]), *Gaia* spacecraft, GALactic Archaeology with Hermes (GALAH [92]), 4-metre Multi-Object Spectrograph Telescope (4MOST [93]), WHT Enhanced Area Velocity Explorer (WEAVE [94]), and Funnel Web.

All mentioned surveys differ in the target selection methodology (magnitude and/or colour thresholds), properties of observed spectra (resolution, wavelength coverage etc.), and sky coverage, but all produce normal and peculiar spectra that have to be identified and analysed. Because of a large number of observations, it is not feasible or justified to manually reduce and analyse all those spectra, therefore automatic pipelines have to be used. The first step in understanding observed spectroscopic data is a reduction pipeline, tailored to a specific telescopic and spectrograph setup [95, 96, 97], that will homogeneously prepare all spectra for further use and potentially add quality flags, warning a user that something might be wrong.

After the data have been prepared, machine learning approaches have been used in many different ways to extract further information from the observed spectra. The most important stellar properties in galactic archaeology are metallicity [Fe/H] and individual chemical abundances. When having an access to a small set of observations with correctly determined parameters, different approaches have been used to project them onto the whole dataset. Currently the most widely used approaches are *The Cannon* [98, 99] and Payne [?] which fall into the category of generative approaches. Meaning that they, in order to infer parameters of analysed spectrum, internally generate a model spectrum that is compared to observations. Another large group of parameter determination machine learning approaches are regression algorithms that directly infer parameters defined by the training set. Of many existing regression approaches, currently the most explored is a use of various neural network architectures [100, 101, 102?]. The usage ranges from deep fully connected networks and convolutional networks to autoencoder structures that first extract latent spectral features which are latter used in the regression procedure.

Moving to chemical tagging experiments, that build on top of previously determined stellar parameters and abundances, a shift from supervised to unsupervised machine learning algorithms is sensible as we almost never do not know the desired result of a conducted experiment. In the literature, chemical tagging experiments often rely on existing unsupervised clustering methods such as K-means [68, 73, 103], DBSCAN [104], t-SNE [8, 70], hierarchical clustering [105, 106], and other custom methodologies [78].

For some other surveys, such as *Gaia*, users do not have full access to raw observations, but have to trust to published stellar properties and their warning signs - such as uncertainties and quality flags - to filter out published data before performing any machine learning operations. Understanding of used data and proper treatment of warning flags and uncertain data is essential before blindly applying any machine learning algorithm to an unknown data set. Failing to do so, we can get stuck in a process that is in computer science commonly referred to as "garbage in, garbage out". A phrase that emphasizes use of clean input training data as it is directly transferred upon investigated test set. Many attempts have been made to effectively filter out *Gaia* kinematic and astrometric parameters, but so far the most commonly used and recommended parameter is Renormalised Unit Weight Error (RUWE) [107] that is used in numerous explorations of the latest dataset.

Various machine learning approaches have been so far applied to the latest *Gaia*

DR2 dataset, among other trying to classify variable stars [108], determine stellar effective temperature [109], catalogue young stellar objects [110], study streams and moving groups in the galactic halo [111, 112], identify accreted stars and structures [113, 114], track hypervelocity stars [115], delineate between stellar and extragalactic sources [116, 117], determine interstellar extinction rates [118] and define new open stellar clusters [119].

1.5 Our exploration of observed data

Our research into open stellar clusters and spectroscopically peculiar stars consists of four related subtopics that are further explained in the text below: analysis of possible runaway stars that were ejected from their birth clusters (based on *Gaia* kinematic and positional measurements and GALAH spectroscopic parameters and abundances), identification of chemically peculiar and emission stellar spectra (based on GALAH spectroscopic data) and investigation of spectroscopic solar twin stars (based on multiple photometric surveys, *Gaia* distances and GALAH spectroscopic data).

To study **ejected stars**, we built upon available research that already defined membership possibilities and cluster parameters for more than 1000 open clusters in the Galaxy [43] using the latest *Gaia* DR2 data. The refined membership probabilities of previously known open clusters [120] were redefined using improved positional and kinematic information. To further define the best cluster members, we additionally used radial velocities to sift out outlying members and multiple stars. With the initial members of a cluster in place, we can continue with the analysis on ejected stars located far away from their cluster center. Observed full 6D positional and kinematics vector for every star enables simulating their position in the Galaxy at different epochs. By integrating those properties, we can model evolution of stellar clusters and their dissolution [18]. It has already been shown that it is possible to retrace some of the nearest runaway stars back to its original cluster using older Hipparcos astrometric observations [121].

To perform a similar task, we used the latest *Gaia* astrometric measurements supplemented with radial velocities measured by the GALAH spectroscopic survey. Their observations were done for fainter sources that are unaccessible for the spectrograph on-board the spacecraft. We integrate movement of stars inside and outside a cluster backwards in time to determine potential points in time where their orbits around the center of the Galaxy intersect. This give us an list of candidates that were once members of a cluster.

The prime focus of this whole open cluster exploration was determination, if machine learning approaches can be readily used for blind chemical tagging of known stellar structures. With this task in mind, we first determined homogeneity and trends of individual chemical abundances for clusters observed among the GALAH spectra. As every cluster is surrounded with numerous unrelated field stars, we explored if chemical signature of a cluster is any different of its surrounding. Greater the difference among both chemical signatures, easier the chemical tagging is. In the case of a very similar chemical compositions, a kinematic information of stars gives us and additional information, and sometimes the only one, that can help with the delineation among field, cluster and ejected stars. In-depth explanation of our work and results is given in Chapter 3.

Every large unbiased spectroscopic observational set is prone to target some **peculiar stars** whose spectrum does not resemble majority of so called normal spectra. During our search through the GALAH spectral dataset, we confirmed that this is as well true also for our acquired data. For a more consistent search of those peculiarities, we employed multiple supervised and unsupervised machine learning techniques to broaden the search onto a complete set of spectra. In it, we browsed for chemically peculiar spectra and spectra with pronounced emission lines. All of those special spectral types need to be identified as thoroughly as possible because they can be interesting for further studies on one hand and might be difficult to correctly determine their physical parameters and abundances using automated pipelines on other hand.

The supervised search for peculiar stars was based on the generation of a normal reference spectrum without any peculiarities that was compared with observed spectra. Any mismatch at the predetermined expected wavelengths was thoroughly measured and analysed in order to extract some physically meaningful explanations about the observed peculiarity. A reference spectra were in our case constructed using two different methodologies. During the supervised construction we compared observed spectra towards median spectrum of stars with similar physical properties. The more complicated unsupervised methodology performed spectrum generation using neural network autoencoder that extracts the most important latent features from a spectrum and reconstructs it using those few latent feature. A result of this spectral compression and un-compression was a peculiarity free reference spectrum.

For the same task of classifying peculiar stars, we also applied unsupervised clustering technique t-SNE [122] to raw acquired spectra. Further explanations of our search for peculiar spectra and results are given in Chapters 4 and 5.

Among all, the best studied chemically peculiar star is our own Sun. When compared to spectroscopically and/or photometrically similar stars, also named **solar twins** (as defined by [123]), it shows signs of under-abundance of volatile chemical elements [124] that may hint to a formation of a solar system around the observed star. When searching for solar twins, we do not consider only chemical composition of stars, but also their physical parameters. For us, it was interesting to see where in the Galaxy we could find solar twins observed by the GALAH. Especially interesting are solar twins embedded in open clusters for which we can study their possible differences in composition towards their birth open cluster. Studies revealed that old open cluster M67 (also observed in the GALAH) seems to contain several stars than could be regarded as solar twins [125, 126].

Solar twins are interesting even for few other studies. From their observed luminosity and known luminosity of the Sun, we can accurately determine their distance. Underabundance of volatile chemical elements in solar twins can be correlated with the presence of planetary systems around selected stars. Fraction of studied stars in the GALAH survey had or will be studied by finished K2 [127] and future TESS [128] spaceborn missions that are searching for planets around other stars. Matching known stars with planetary systems with our GALAH sample might reveal abundance patterns that can hint at the presence of rocky or gaseous planets.

To search for solar twins, we used raw spectra and compared them to reference solar spectrum, that was constructed by averaging a multitude of acquired twilight flats. The selection of the best candidates was based on a similarity metrics, where we selected only a few percents of the best matching spectra. Close inspection of ab-

Chapter 1. Introduction

solute magnitudes (observed magnitude corrected for stellar distance) showed that some of stars in the selection looks too bright for their spectral type and distance. To investigated possible scenarios, we build a spectroscopic and a photometric model of a single star. Stack of multiple single stars raveled that observed objects might contain multiple stellar sources that are slowly revolving around their common mass centre. Further details about the search for solar twins and their multiplicity modeling is given in Chapter 6.

Chapter 2

Spectroscopic, photometric, and astrometric surveys

In the last few decades, we are witnessing a fast and numerous shifts from dedicated single object observations to massive all-sky surveys producing hundreds or even thousands of unbiased observations in a single telescope pointing. Along with the complexity of data acquisition and storage, new challenges and problems involving data reduction arose, requiring dedicated computer power to reduce acquired data. The reduction challenges span from timely, almost real-time reduction requirements, to complex, computationally demanding processes that try to take into account as many telescopically and observationally induced biases as possible. Some of those processes will be discussed in the following sections discussing a specific survey.

This thesis shows a few cases of synergies of such vastly different data sets and at the same time points to a necessity of having knowledge about the automatic processing pipelines that produced final products. Something that is quite often blindly overlooked by users.

The main source of data for our studies are the following three stellar surveys producing informations about the stars' brightness, composition, distance, kinematics and many additional parameters that can be inferred from the observed quantities.

2.1 Gaia

Gaia is the one-billion-star surveyor of the European Space Agency (ESA). It has been continuously scanning the sky since July 2014 from its designated location close to the second Lagrange point of the Sun-Earth/Moon system. *Gaia*'s aim is to map the entire sky, down to magnitude ~ 20.7 , and to collect micro-arcsecond-level astrometry and milli-magnitude-level photometry for the brightest 1,000+ million stars as well as medium-resolution spectroscopy for mainly radial-velocity determination of the brightest subset of ~ 150 million objects.

The *Gaia* scanning of the sky is composed of two independent, superimposed motions: a rotation around the spacecraft spin axis with a period of 6 hours plus a slow, 63 day period precession of the spin axis around the Solar direction at a fixed Solar-aspect angle of 45° . Over the nominal five-year mission, *Gaia* has completed 29 of these precession periods, leading to an optimally uniform sky coverage with, on average, ~ 70 astrometric and photometric transits across the focal plane (and ~ 40 for the spectroscopic instrument). In the extended mission phase that started

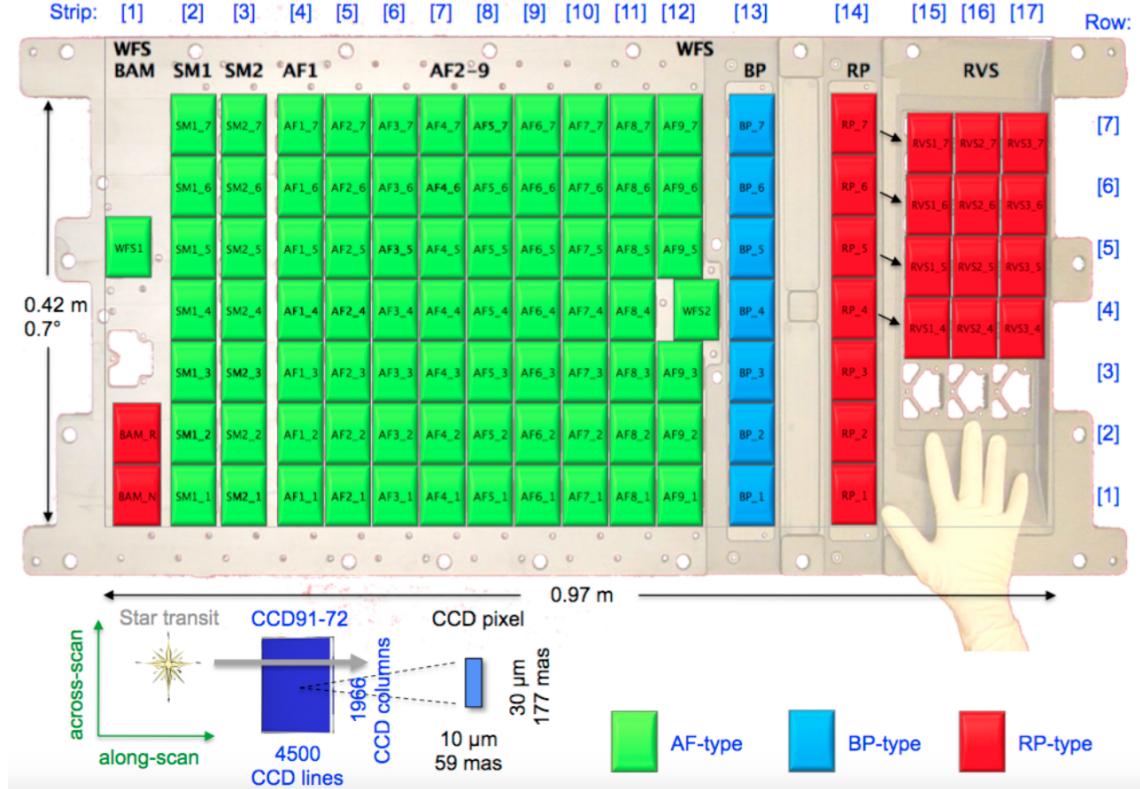


Figure 2.1: Size comparison and spatial arrangement of the CCD detectors in the *Gaia* focal plane. Image credit: Gaia Collaboration *et al.* [129].

Table 2.1: Past and future predicted release dates of *Gaia* data and products.

Release designation	Date
<i>Gaia</i> DR1	14 September 2016
<i>Gaia</i> DR2	25 April 2018
<i>Gaia</i> EDR3	2 nd half of 2020 or later
<i>Gaia</i> DR3	postponed into late 2021 or later
Final release	not yet determined

in July 2019, a similar scanning law is being employed but with a reversed precession direction during the first year. A passage of a star through the focal plane is called a field-of-view transit. During each transit, *Gaia* collects instantaneous, so-called epoch data of each object. Publication of all epoch data is scheduled (see Table 2.1 for the final data release).

As the spacecraft slowly rotates, observed stars traverse the *Gaia* focal plane equipped with 106 CCD detectors (show in Figure 2.1). Every star that gets observed therefore passes trough a sequence of those detectors who analyse a star in the given order:

2.1.1 Photometry and astrometry

The first array of CCDs that collects light from stars is a Sky Mapper (SM) that autonomously detect objects. Stars brighter than magnitude ~ 3 are too bright to be detected automatically. The faint detection threshold is set at 20.7 magnitude in the *Gaia* G band but is not infinitely sharp due to on-the-fly magnitude estimation errors of the on-board software.

After the source detection, stars pass into the largest array of CCDs that is attributed to the Astrometric Field (AF). It collects the instantaneous positions and fluxes of all objects detected by the Sky Mapper as they traverse along the field. Astrometric measurements are made in a white-light bandpass, covering the range from 3300 to 10500 Å, which is referred to as the *Gaia* G band.

The last spectro-photometric measurements are thereafter performed by two low dispersion detectors that are measuring precise fluxes in a number of narrow-pass sub-bands of previously mentioned wide-pass G band. The Blue Photometer (BP) collects low resolution spectra of all objects over the wavelength range from 3300 to 6800 Å. The integrated magnitude is referred to as the G_{BP} or BP magnitude. The Red Photometer (RP) collects low-resolution spectra of all objects over the wavelength range from 6300 to 10500 Å. The integrated magnitude is referred to as the G_{RP} or RP magnitude.

2.1.2 Spectroscopy

A final measurement performed by the spacecraft is spectroscopy over the whole observed field of the sky. The integral-field Radial Velocity Spectrometer (RVS) [130] collects medium-resolution spectra (spectral resolving power (R) $\sim 11,700$) over the wavelength range from 8450 to 8720 Å, for all objects brighter than magnitude ~ 16 in this bandpass. The location of the pass-band is selected to cover the ionised calcium (CaII) triplet with a prominent absorption features over a large temperature range of stars and can therefore be used to determine radial velocity of spectrally diverse stars. The integrated magnitude in the RVS bandpass is referred to as the G_{RVS} magnitude. The RVS has a reduced field of view orthogonal to the scan direction such that fewer observations up to the RVS limiting magnitude are collected compared to photometric fields in a ratio of 4:7.

2.2 *Gaia DR2*

The second data release of *Gaia* data (*Gaia* DR2) was heavily used during the production of this thesis, therefore it requires detailed description of provided tables, its use and potential problems. The release set is far from complete and similar to the final release, but on the other hand provides an unprecedented set of homogeneously acquired and reduced stellar informations newer seen before. Visual and numerical representation of the specific stellar products is given in Figure 2.2. *Gaia* DR2 is based on data collected by the spacecraft between 25 July 2014 and 23 May 2016, spanning a period of 22 months.

Along the calibrated raw measurements, Gaia Data Processing and Analysis Consortium (DPAC) provides numerous parameters and properties of stars, pre-selected Solar system bodies and quasars that can be used and explored by users.

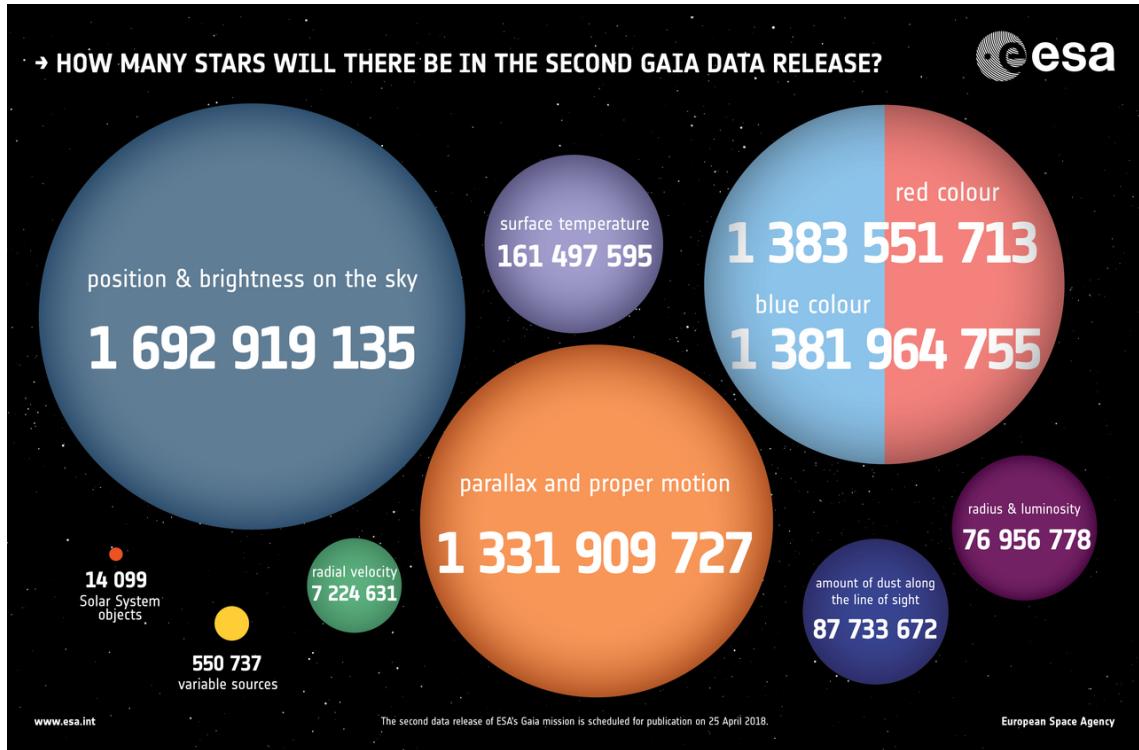


Figure 2.2: Visual and numerical representation of *Gaia* DR2 stellar content. Image credit: ESA, CC BY-SA 3.0 IGO.

Among them we can find:

- **Astrometric set** consisting of partial 2-parameter (limited to celestial positions α and δ) and full 5-parameter astrometric solution with addition of parallax, and proper motion in both coordinates. The 2-parameter sources are typically faint (with about half of them at magnitude $G > 20.6$), have very few observations (less than five as required for full solution), or very poorly fit the five-parameter astrometric model. All sources fainter than magnitude $G = 21$ have only positional information. In the current data release all stars are still treated as single during astrometric fit, which could significantly influence solutions in the case of multiple fast moving sources contributing to the position of an observed photocentre. The *Gaia* coordinate reference frame is aligned with the International Celestial Reference Frame (ICRF) using positions of extragalactic sources such as quasars [131]. After the initial release, it was determined that the quality of an astrometric fit is best described by the renormalised astrometric chi-square (RUWE) [107]. Further details about astrometric processing and validation are given in Lindegren *et al.* [132], Luri *et al.* [133].
- **Photometric data set** contains the broad band photometry in the G , G_{BP} , and G_{RP} bands, giving us colour information for *Gaia* DR2 sources that were observed at least twice. The mean value of the G -band fluxes is reported for all sources while colour information (BP and RP) is available for about 80% of them. The integrated colour information suffers from strong systematic effects at the faint end of the survey ($G > 19$), in crowded regions, and near bright stars. In the case when measured fluxes are inconsistent between the

G and the G_{BP} and G_{RP} bands (sum of the later two is significantly larger than G measurement) a warning is raised. A quantitative index of this effect is provided in the numerical form as *flux excess factor*. Further details about processing and photometry validation are given in Evans *et al.* [134], Riello *et al.* [135].

- **Radial velocity measurements** indicate stellar median radial velocity, averaged over the first 22 months of the observations. Therefore stellar multiplicity and their orbital motion can not be determined from current data. Velocities are provided for sources which are brighter than magnitude 12 in the G_{RVS} photometric band. Because of used set of spectral templates during radial velocity determination, velocities are reported only for stars with effective temperatures in the range between 3550 and 6900 K (referring to the effective temperature of a used template and not an actual effective temperature of a star). By the RVS pipeline design, determined absolute radial velocity are limited to 1000 km s^{-1} . The uncertainties of the radial velocities at the faint end depend on stellar effective temperature and range from 1.4 km s^{-1} for cooler to 3.6 km s^{-1} for hotter stars. The zero-point of the RVS velocities was determined using a comprehensive set of standard stars with numerous dedicated, precise and temporally spread radial velocity measurements [136]. Further processing details of RVS data are given in Sartoretti *et al.* [137].
- **Stellar variability data set** consists of sources that were firmly identified as variable (based on at least two photometric observations of the two *Gaia* telescopes). The final number still represents only a small subset of the total amount of variables expected in the *Gaia* survey. The sources were classified into the following nine categories based on their light curves: RR Lyrae (anomalous RRd, RRd, RRab, RRc), long period variables (Mira type and Semi-Regulars), Cepheids (anomalous Cepheids, classical Cepheids, type-II Cepheids), δ Scuti and SX Phoenicis stars. If a star had 12 or more observations its light curve was analysed in detail. They are designated as specific object studies (SOS) and consist of variables of the type Cepheid and RR Lyrae, long period variables, short time scale variables, and rotational modulation variables. Full details on the variable star processing, results and their validation are given in Holl *et al.* [138], Mowlavi *et al.* [139], Molnár *et al.* [140], Clementini *et al.* [141].
- **Astrophysical parameters** derived by the astrophysical parameter inference system in the *Gaia* data processing (Apsis) include estimates of T_{eff} extinction A_G and reddening $E(G_{BP} - G_{RP})$, radius, and luminosity for stars brighter than magnitude $G = 17$. Values of T_{eff} are reported only over the temperature range between 3000 and 10,000 K. Limits are induced by the training set of the algorithm responsible for the T_{eff} estimation. Estimates of the other astrophysical parameters are published for about half of the sources with determined T_{eff} . As the processing pipeline was performed individually for every object and with a limited set of input data (three *Gaia* photometric bands and parallax) some errors are expected because of high degeneracy between determined parameters. If a star is located far from expected isochrones used in the processing, extinction becomes overestimated. Full details of the astro-

physical parameter processing and result validation are described in Andrae *et al.* [142].

- **Solar system objects** (SSO) data set provides epoch astrometry and unfiltered G photometry for a pre-selected list of 14,099 known minor bodies in the solar system that are numbered in the Minor Planet Center repository. Each time a given SSO enters field of view of *Gaia* telescopes, celestial positions are recorded as seen from the spacecraft. The data set and its production are thoroughly described in Gaia Collaboration *et al.* [143].

The above sections are partially adapted and summarized from Gaia Collaboration *et al.* [42, 129], Gaia Helpdesk [144].

2.3 GALAH

The GALactic Archaeology with HERMES (GALAH) [145] is an ongoing spectroscopic survey that aims to unveil the Milky Way’s formation history by studying the detailed chemical composition of observed stars. Fossil remnants, which have been disrupted during the formation and are now dispersed around the Galaxy, are tough to have conserved the initial chemical signature of individual galactic components. It is essential to disentangle their formation location and migration history in order to explain current stellar populations. This can be achieved through the technique of chemical tagging [66] that promises identification of old dispersed fossil remnants based on their unique abundance patterns over numerous chemical elements. The GALAH aims to achieve this by measuring up to 31 elemental abundances (from 7 independent element groups with different physical origin) individually in every acquired spectrum.

The GALAH survey was the main driver for the construction of the High Efficiency and Resolution Multi-Element Spectrograph (HERMES) [146, 147], a multi-fibre spectrograph mounted on the 3.9-metre Anglo-Australian Telescope (AAT) situated at the Siding Spring Observatory, Australia. The spectrograph has a resolving power of $R \sim 28,000$ (or $R \sim 45,000$ when slit mask is used) and covers four separately acquired wavelength ranges ($4713 - 4903 \text{ \AA}$, $5648 - 5873 \text{ \AA}$, $6478 - 6737 \text{ \AA}$, and $7585 - 7887 \text{ \AA}$), together covering approximately 1000 \AA , including the $\text{H}\alpha$ and $\text{H}\beta$ lines. The ranges are frequently referred to as blue, green, red and near-infrared spectral arms. This configuration can simultaneously record spectra from up to 392 fibres distributed over a 2° diameter field of the night sky, with an additional 8 fibres used for the telescope guiding. The spectrograph can typically achieve a signal to noise ratio (SNR) ~ 100 per resolution element at magnitude $V=14$ in the red arm during a 1-hour long exposure.

2.3.1 Acquired spectra and target selection

The spectroscopic data used during the production of this thesis were taken from the pilot survey, the main GALAH survey [145], the K2-HERMES survey [148], the TESS-HERMES survey [149], and special dedicated the HEMRES open clusters (De Silva et al. in preparation) and the HERMES Orion star forming region (Kos et al. in preparation) surveys. Together they form a dataset of 669,845 successfully reduced stellar spectra, of which a small fraction belongs to a repeated observations.

All acquired spectra are homogeneously reduced to one dimensional spectrum, normalised and shifted to stellar reference frame (detailed description in Kos *et al.* [95]). Combination of those surveys produces increased number of spectra compared to the main GALAH survey, but at the same time breaks rule of a simple unbiased selection function (Sharma et al. in preparation) that is desired for population studies and easier comparison with synthetic galactic models.

The original selection function of the main GALAH survey is separated into two magnitude limited field selections - bright ($10 < V < 12$) and normal ($12 < V < 14$) fields whose target selection is colour independent. Used V magnitude is inferred from magnitudes measured by the Two Micron All-Sky Survey (2MASS) [150] whose photometric bands are shifted into infra-red spectral region. Because of that, some, especially peculiar and variable stars, might have erroneous estimation of V magnitude leading to an underexposure or excessive spectral crosstalk. Because of expected crowding problems (projected diameter of used optical fiber on the sky is equal to $2''$) observed stars are located at higher Galactic latitudes ($|b| > 10^\circ$) where density of stars is lower. Additional surveys sometimes break those rules by selecting fainter/dimmer stars, going closer to the Galactic plane, employ colour cuts, or favor interesting preselected stars such as K2 [127] targets, TESS [151] targets and cluster members. Therefore some care is needed when trying to infer global stellar or galactic properties based on such inhomogeneous selection criteria.

2.3.2 Spectral reduction and parameters determination

The first step after recording spectra in the form of a 2D image is their extraction to 1D spectrum. The procedure, extensively documented by Kos *et al.* [95], consist of the following steps: raw image cosmetic corrections, spectral tracing, optical aberrations correction, scattered light and apertures coss-talk removal, wavelength calibration, sky subtraction, and telluric absorption removal. After reduction, spectra are normalised and shifted into their rest frame by cross-correlating them with a set of 15 AMBRE model spectra [152]. Stellar atmospheric parameters and individual elemental abundances derived from normalised spectra, acquired by different surveys, are analysed with the same procedure that slowly evolved and improved during the course of the GALAH survey. The three most important milestones in an ongoing process are:

- The initial stellar parameters (T_{eff} , $\log g$, and $[\text{Fe}/\text{H}]$), accompanying the first GALAH data release (**GALAH DR1**), were derived as a global fit (all arms at the same time) of the observed spectra to a grid of 16,783 AMBRE spectra [152], which were convolved down to the average resolution of individual CCD. The aim of this procedure was to provide indicative stellar parameters, which could be used as a first initial guess to help speed up a more complex stellar parameters and abundance pipelines.
- The parameters (with extension to $v \sin i$, v_{mic} , and A_{K_S}) and up to 23 elemental abundances, released as part of the **GALAH DR2** [99], were produced using a multi-step data driven approach. The complete analysis depended on a set of 10,605 spectra that were selected in a such way to span a large portion of the parameter and did not contain any peculiar star, especially binary and emission line spectra. Selected spectra were analysed using a physics-driven

spectrum synthesis code Spectra Made Easy (*SME*) [153, 154] that performs spectrum synthesis for 1D stellar atmospheres models. In the case of DR2, it consist of MARCS theoretical 1D hydrostatic models [155] under the assumption of local thermodynamic equilibrium (LTE) for majority of elements. Only several key elements (Li, O, Na, Mg, Al, Si, and Fe) were analysed using non-LTE line formation.

To propagate the parameter and abundance results of the training set to the whole survey, *The Cannon* [98] generative data-driven approach was used. It adopts a simple quadratic model which uses stellar parameters to describes observed flux of a given spectrum. Independent model is build for every spectrum wavelength pixels. To train *The Cannon* model all spectra in the training set were interpolated to a common wavelength grid. After the training is performed, model was inverted to produce parameters and abundances for every observed spectrum by fitting it to internal generative spectrum produced by *The Cannon*. Further details of the described process are given in Buder *et al.* [99].

- Not relying only on the GALAH spectral information, but including *Gaia* parallax, colour, and absolute magnitude, additional constrains and priors can be used to infer stellar parameters. Adaption of *SME* software thoroughly described in (???????) used those additional information to produce **GALAH DR3** data set. Unlike in DR2, no data-driven methodology was used to produce stellar parameters and abundances as *SME* was run for every individual spectrum. Reference spectra were generated using MARCS 1D stellar models. For eleven abundances (out of 30) non-LTE computations were performed. The published catalogue contains several additional Value-Added-Catalogues defining stellar ages and stellar dynamics. To compute stellar masses and ages, the best matching PARSEC+COLIBRI isochrones were used.

2.4 Asiago

Vastly different from the previous two massive all-sky surveys, telescopes at the Asiago site are mainly used for dedicated observations or monitoring of previously selected targets, whose observational and astrophysical potential was identified from all-sky surveys. During our stay at the Asiago observatory, that usually lasted for four consecutive bright nights (close to full Moon) every month, we used 1.82 m Copernico telescope located on top of the nearby hill Mount Ekar (Asiago, Italy - altitude of 1,366 m).

All our observations were performed by the Echelle spectrograph that is mounted on the telescope on days around the full Moon when quality and deepness of photometric observations is heavily reduced. Design of the Echelle instrument and its slit length enables observer to observe only one star a time. Obtained spectra have a resolving power of $R \sim 20,000$ and a wide span of wavelengths between 3600 and 7400 Å. They are divided into 30 orders which partially overlap with succeeding and preceding order, providing an undisturbed coverage of observed wavelengths. Acquired spectra are recorded by Andor CCD with the size of 2048×2048 pixels. This setup enables us to capture spectra of stars with magnitudes $V < 10$ at high SNR with reasonable exposure time (less than 1 hour per spectrum). Because of

the mechanical limitation, observed stars must be positioned at least 15° above the local horizon. At those low altitudes, only the brightest stars are reasonable to be observed because of strong atmospheric attenuation.

Combining location of the observatory and above observational limitations with the fact that our interesting stars were selected from the GALAH survey, highly reduces the number of potentially observable objects. To reduce the atmospheric effect effects, we only observed stars which rose at least 30° above the local horizon. As described in more detail below (Chapters 4 and 6), we used additional Asiago observations to inspect spectroscopic features not accessible by the GALAH spectra and to prolong radial velocity time series of possible multiple stars who could show signs of radial velocity changes not detectable by a single epoch GALAH spectrum.

Additionally to our program observations, we also contributed spectroscopic observations that resulted in published astronomer's telegrams [156] and scientific papers [157].

Chapter 3

Chemo-dynamic tracing of open cluster stars

Chapter 4

Chemically peculiar stars

This chapter has been adapted from the published paper titled *The GALAH survey: a catalogue of carbon-enhanced stars and CEMP candidates* [158] that was prepared and written by the author of this Doctoral thesis. The used computer code was published on GitHub platform¹ and results of the analysis as a catalogue on VizieR service².

4.1 Abstract

Swan bands – characteristic molecular absorption features of the C₂ molecule – are a spectroscopic signature of carbon-enhanced stars. They can also be used to identify carbon-enhanced metal-poor (CEMP) stars. The GALAH (GALactic Archaeology with Hermes) is a magnitude-limited survey of stars producing high-resolution, high signal-to-noise spectra. We used 627,708 GALAH spectra to search for carbon-enhanced stars with a supervised and unsupervised classification algorithm, relying on the imprint of the Swan bands. We identified 918 carbon-enhanced stars, including 12 already described in the literature. An unbiased selection function of the GALAH survey allows us to perform a population study of carbon-enhanced stars. Most of them are giants, out of which we find 28 CEMP candidates. A large fraction of our carbon-enhanced stars with repeated observations show variation in radial velocity, hinting that there is a large fraction of variables among them. 32 of the detected stars also show strong Lithium enhancement in their spectra.

4.2 Introduction

Chemically peculiar stars whose spectra are dominated by carbon molecular bands were first identified by Secchi [159]. Their spectra are characterised by enhanced carbon absorption bands of CH, CN, SiC₂, and C₂ molecules, also known as Swan bands. Possible sources of enhancement are dredge-up events in evolved stars [160], enrichment by carbon-rich stellar winds from a pulsating asymptotic giant branch (AGB) star, which settles on a main sequence companion [161], or it can be the result of a primordial enrichment [162]. Historically, high latitude carbon stars, presumed to be giants, were used as probes to measure the Galactic rotation curve

¹<https://github.com/kcotar/GALAH-survey-Carbon-enhanced-stars>

²<http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=J/MNRAS/483/3196>

[163], velocity dispersion in the Galactic halo [164], and to trace the gravitational potential of the Galaxy.

Because of their strong spectral features, the most prominent candidates can easily be identified from large photometric surveys [165, 166]. Specific photometric systems [167, 168, 169] were defined in the past to discover and further classify stars with enhanced carbon features in their spectra. Specifics of those systems were catalogued, compared, and homogenised by Moro and Munari [170] and Fiorucci and Munari [171].

Other useful data come from low-resolution spectroscopic surveys, whose classification identified from a few hundred to a few thousand of those objects [172, 173, 174, 175, 176]. High-resolution spectroscopy is required to search for candidates with less pronounced molecular absorption features or to determine their stellar chemical composition. Multiple studies have been carried out to determine accurate abundances of metal-poor stars [177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188]. Such detailed abundance information is especially important for the analysis and classification of chemically peculiar objects [189].

Today, the most sought after, of all carbon-enhanced stars, are the carbon-enhanced metal-poor (CEMP) ones whose fraction, among metal-poor stars, increases with decreasing metallicity [M/H] [174, 177, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200]. Amongst these, those near the main-sequence turn-off are expected to be of particular importance, as they may have accreted enough material from their AGB companion to produce an observable change in their atmospheric chemical composition [187, 201, 202]. The accreted material could provide insight into the production efficiency of neutron capture elements in AGB stars [182]. Multiple studies show that a peculiar observed abundance pattern and carbon enrichment in a certain type of CEMP stars could be explained by the supernova explosions of first-generation stars that enriched the interstellar medium [203, 204, 205, 206]. The exact origin and underlying physical processes governing multiple classes of CEMP stars are not yet fully understood and are a topic of ongoing research [162, 207, 208]. Classification into multiple sub-classes is performed using the abundance information of neutron-capture elements [162, 209, 210, 211] that are thought to originate from different astrophysical phenomena responsible for the synthesis of those elements.

In this work, we propose a novel approach for the classification of carbon-enhanced stars using high-resolution stellar spectra covering parts of the visible domain. The goal is to identify a representative sample of carbon-enhanced stars, which can be used as an input to population studies. The paper is organised as follows; we start with a brief discussion of our spectroscopic observations and their reduction (Section 6.9), which is followed by the description of the used algorithms for the detection of carbon-enhanced stars in Section 4.4. Properties of the classified objects are investigated in Section 4.5, CEMP candidates are a focus of Section 4.6, with Section 4.7 describing a follow-up study for one of them. Final remarks are given in Section 4.8.

4.3 Data

The analysed set of stellar spectra was acquired by the High Efficiency and Resolution Multi-Element Spectrograph (HERMES), a fibre-fed multi-object spectrograph on the 3.9 m Anglo-Australian Telescope (AAT) of the Australian Astronomical Ob-

servatory. The spectrograph [146, 147] can simultaneously record spectra from up to 392 fibres distributed over a 2° field of the night sky, with an additional 8 fibres used for the telescope guiding. The spectrograph has a resolving power of $R \sim 28,000$ and consists of four spectral arms centred at 4800, 5761, 6610, and 7740 Å, together covering approximately 1000 Å, including the H α , and H β lines. Three dichroic beam splitters are used to separate incoming light into four separated colour beams that are analysed independently. The spectrograph can typically achieve a signal to noise ratio (SNR) ~ 100 per resolution element at magnitude V=14 in the red arm during a 1-hour long exposure.

Spectra used in this study have been taken from multiple different observing programmes using this spectrograph: the GALactic Archaeology with HERMES (GALAH) pilot survey [212], the main GALAH survey [145], the K2-HERMES survey [213], and the TESS-HERMES survey [149]. Most of those observing programmes exclude fields close to the Galactic plane (due to problems with high stellar density and Galactic extinction) or far away from it (not enough suitable targets to use all fibres), employ subtle different selection functions (position, limiting magnitude, crowding requirement, and photometric quality), but share the same observing procedures, reduction, and analysis pipeline [internal version 5.3, 95]. All programmes, except the pilot survey, are magnitude-limited, with no colour cuts. This leads to an unbiased sample of stars distributed mostly across the southern sky that can be used for different population studies. Additionally, all objects from different observing programmes are analysed with the same procedure named *The Cannon* [internal version 180325, 98, 99], so their stellar parameters are determined in a consistent manner and are hence comparable across the different programmes. *The Cannon* algorithm employs a data-driven interpolation approach trained on a set of high-quality benchmark stars spanning the majority of the stellar parameter space [for details, see 99].

The spectrum synthesis code Spectroscopy Made Easy [SME, 153, 154] was run on the spectra in the training set to determine their stellar parameters and atmospheric chemical abundances. These values were used to train *The Cannon* model, which was then run on every observed spectrum to determine its stellar parameters. Every determined parameter is accompanied by a quality flag identifying its usefulness and possible problems with grid interpolation. Therefore, we can easily remove parameters that were determined for stellar spectra far away from the training set or with possible problems in any of the HERMES spectral arms. An interpolation method proved beneficial in determining stellar physical parameters and up to 23 abundances (extendable up to 31 in future versions) for every analysed star, as described in Buder *et al.* [99]. The parameter [Fe/H] determined by *The Cannon* refers to an iron abundance and not overall metallicity [M/H] as often used in the literature. As both notations are used interchangeably in the literature to describe stellar metallicity, care must be taken when comparing those two measurements.

The Cannon approach has an advantage of treating stars with a peculiar composition, such as carbon-enhanced stars, in a consistent manner with common objects, thus avoiding arbitrary jumps or offsets that would depend on a degree of peculiarity or correctness of physics of the underlying model. As a data-driven approach, it only projects the learned training set onto the whole survey based on quadratic relations to stellar parameters.

Our data set consists of 627,708 successfully reduced spectra of 576,229 stars

observed between November 2013 and February 2018.

4.4 Detection procedure

To search for carbon-enhanced stars in the GALAH data set we focused on spectral features that can be clearly distinguished and are known markers of carbon enhancement. Instead of using one very weak atomic carbon absorption line (at 6587.61 Å), used by *The Cannon* to determine [C/Fe] abundance, we focused on a region between 4718 and 4760 Å observed in the blue arm that covers 4718–4903 Å in its rest-frame. In this range we can, depending on the radial velocity of the star, observe at least four Swan band features [214] with their band heads located at approximately 4715, 4722, 4737, and 4745 Å.

Carbon enhancement is observable in spectra as a strong additional absorption feature (Figure 4.1) that is the strongest at the wavelength of the band’s head. After that its power gradually decreases with decreasing wavelength. The most prominent and accessible for all of the spectra is a feature located at 4737 Å, produced by a $^{12}\text{C}^{12}\text{C}$ molecule. If other carbon features, like the one produced by a $^{13}\text{C}^{12}\text{C}$ molecule at 4745 Å (shown in Figure 4.16) are present in the spectrum, the carbon isotope ratio $^{12}\text{C}/^{13}\text{C}$ in a star can be determined. Its determination was not attempted in the scope of this paper.

Detection of spectral features was tackled using two different classification procedures. First, a supervised procedure was used to identify the most prominent spectra with carbon enhancement. It is based on the assumption that we know where in the spectra those features are located and how they behave. This was augmented with an unsupervised dimensionality reduction algorithm that had no prior knowledge about the desired outcome. The goal of a dimensionality reduction was to transform n-dimensional spectra onto a 2D plane where differences between them are easier to analyse. The unsupervised algorithm was able to discern the majority of carbon-enhanced spectra from the rest of the data set and enabled us to discover spectra with less prominent carbon enhancement features.

4.4.1 Supervised classification

To search for additional absorption features that are usually not found in spectra of chemically normal stars, we first built a spectral library of median spectra based on a rough estimates of stellar physical parameters derived by the automatic reduction pipeline, described in detail by Kos *et al.* [95]. The median spectrum for every observed spectrum in our data set was computed from physically similar spectra with stellar parameters in the range of $\Delta T_{\text{eff}} = \pm 75$ K, $\Delta \log g = \pm 0.125$ dex and $\Delta [\text{Fe}/\text{H}] = \pm 0.05$ dex around the stellar parameters of the investigated spectrum. The median spectrum was calculated only for observed targets with at least 5 similar spectra in the defined parameter range and with minimal SNR=15 per resolution element, as determined for the blue spectral arm. All considered spectra were resampled to a common wavelength grid with 0.04 Å wide bins and then median combined. The normalisation of the spectra along with the radial velocity determination and the corresponding shift to the rest frame was performed by the automatic reduction pipeline [95]. We checked that spectral normalisation and radial velocity determination are adequate also for carbon-enhanced stars. The normalisation procedure is

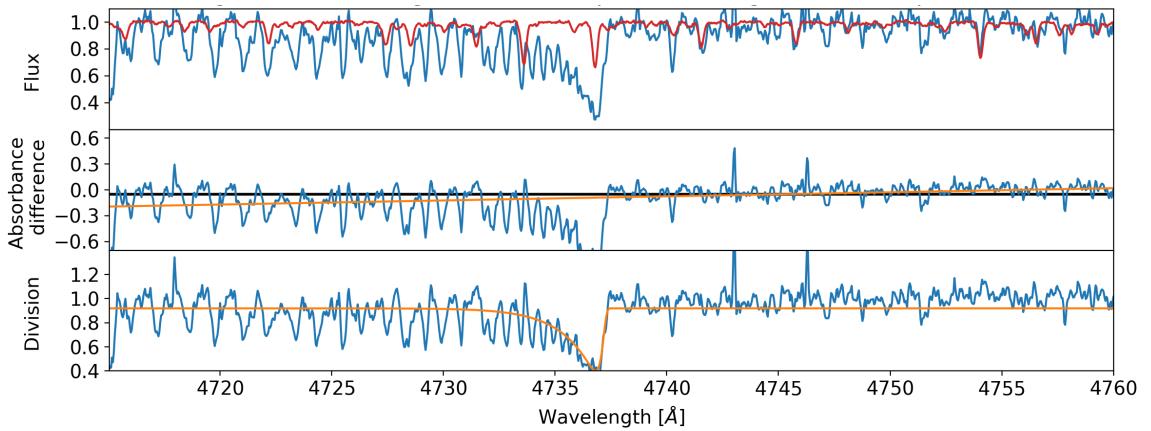


Figure 4.1: Example of a metal-poor carbon-enhanced candidate with strong Swan absorption feature at 4737 \AA , caused by the carbon C_2 molecules. The first panel shows the stellar spectrum in blue and a corresponding reference median spectrum in red. The reference median spectrum was computed as the per-pixel median flux value of spectra with similar stellar parameters as the spectrum shown in blue. The spectral difference (second panel) and division (third panel) were computed from those two spectra. The middle panel shows in orange a linear fit to the spectral difference that was used to identify spectra with the reduction problems on the blue border of the spectral range. The median value of the spectral difference is given by the black horizontal line. The orange curve in the last panel shows a fit that was used to determine the strength of the observed carbon feature. The shown spectrum belongs to a star with a Two Micron All-Sky Survey (2MASS) identifier J11494247+0051023 and iron abundance $[\text{Fe}/\text{H}]$ of -1.17 , as determined by *The Cannon*.

done using a polynomial of low-order that is not strongly affected by the Swan band features or similar spectral structures. The radial velocity of a star is determined as an average of radial velocities that were independently determined for the blue, green, and red spectral arm. If one of the arms has a radial velocity deviating for more than two times the difference between the other two, it is excluded from the average [further details in 95]. Therefore the final velocity should be correct even if one of those arms contains features that are not found in the set of reference spectra used in the cross-correlation procedure.

With the limitation of at least 5 spectra used for the computation of the median spectrum, some possibly carbon-enhanced stars, could be excluded from the supervised classification. The final number of spectra analysed by this method was 558,053.

Spectra, for which we were able to determine the median spectrum of physically similar objects, were analysed further. In the next step, we tried to determine possible carbon enhancement by calculating a flux difference and flux division between the observed stellar and median spectra, as shown in Figure 4.1.

In order to describe the position, shape, and amplitude of the Swan feature with its head at 4737 Å, we fitted a function that is based on a Log Gamma ($\log \Gamma$) distribution. The distribution, with three free parameters, was fitted to the division curve, where the Swan feature is most pronounced. Division curve, shown in the bottom panel of Figure 4.1, was computed by dividing observed spectrum with its corresponding median spectrum. The fitted function f can be written as:

$$f(\lambda) = f_0 - \log \Gamma(\lambda, S, \lambda_0, A). \quad (4.1)$$

The shape of the curve is defined by an offset f_0 , shape parameter S , centre wavelength λ_0 , and amplitude A of $\log \Gamma$ distribution, where λ represents rest wavelengths of the observed spectrum. This function was selected because of its sharp rise followed by the gradual descent that matches well with the shape of a residual absorption observed in the Swan regions. The steepness of the rising part is determined by the parameter S (lower value indicates steeper raise) and its vertical scaling by the parameter A . We are not aware of any other profile shapes used for fitting Swan bands in the literature.

To narrow down possible solutions for the best fitting curve, we used the following priors and limits. The initial value for the parameter f_0 was set to a median of all pixel values in the division curve and allowed to vary between 0.5 and 1.5. The limiting values are however never reached. The centre of the $\log \Gamma$ distribution λ_0 was set to 4737 Å and was allowed to vary by 2 Å. Wavelength limits were set to minimise the number of mis-fitted solutions, where the best fit would describe the nearby spectral absorption lines not present in the median spectra or problematic spectral feature caused by the spectral data reduction as shown by Figure 4.18. We did not set any limits on parameters A and S in order to catch fitted solutions describing a spectrum difference that is different from the expected shape of the molecular absorption band.

By integrating the surface between the offset f_0 and the fitted curve we calculated the strength of the Swan band. The integral (`swan_integ` in Table 6.8) is derived between 4730 and 4738 Å. It should not be used as a substitute for a carbon abundance measurement, but only to sort the detections of carbon-enhanced stars by their perceivable strength of the Swan band.

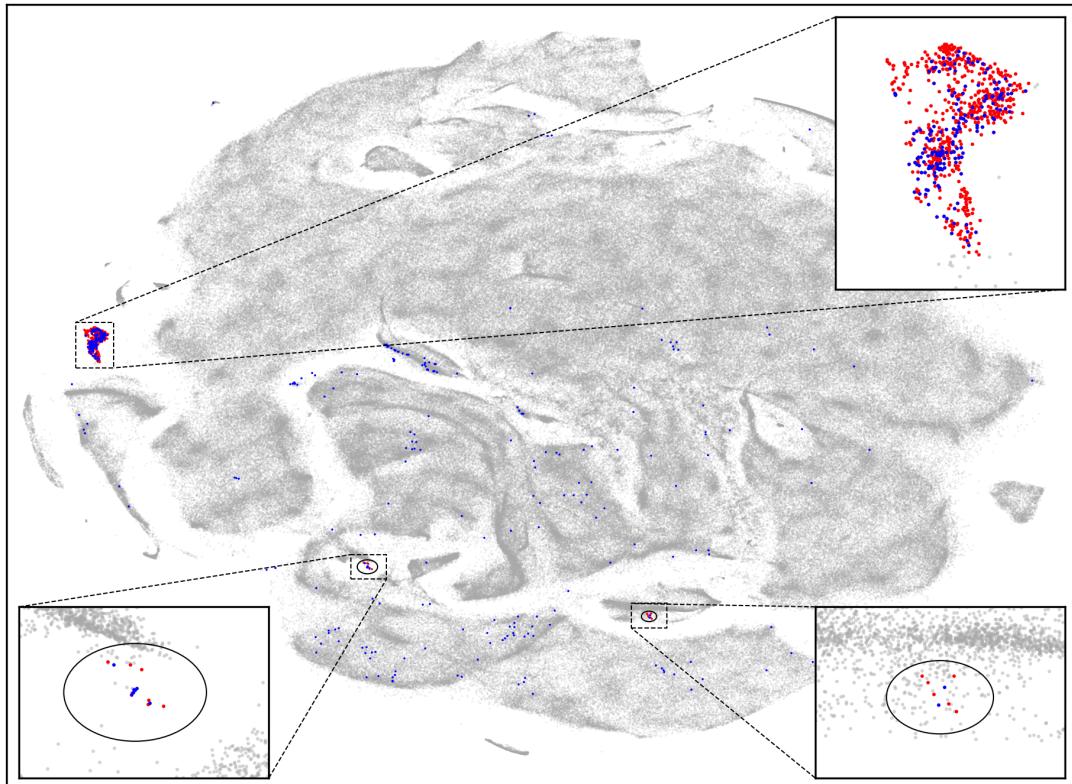


Figure 4.2: t-SNE projection of 588,681 observed spectra ranging between 4720 and 4890 Å. Red dots (756 spectra) mark a clump in the projection that was manually selected to contain carbon-enhanced spectra. Superimposed blue dots represent carbon-enhanced spectra determined by the supervised algorithm. Outside the t-SNE selected clump, we have 224 spectra that were determined to be carbon-enhanced only by the supervised method. All other analysed spectra are shown in grey shades, depending on their density in the 2D projection. Two ellipses indicate regions where the majority of CEMP candidates is located in the projection.

With so many spectra in our data set, unexpected reduction and analysis problems can hinder the selection of carbon-enhanced stars. In the first iteration, the results were ordered only by the value of the integrated Swan band region, but this proved to select too many spectra with reduction problems. Most of the problematic detections were caused by the incorrect normalisation of spectra with strong, non-carbon molecular bands. This is best observable at the border of a spectral range, where Swan bands are located in the case of HERMES spectra. There, normalisation can be poorly defined in the case of numerous nearby absorption lines. In order to prevent miss-detections, additional limits on the shape ($S \leq 1$) and amplitude ($A \leq 1$) of the $\log \Gamma$ distribution were used to filter out faulty fitting solutions. Figure 4.19 represents one such example where the function $f(\lambda)$ was fitted to the absorption lines of a double-lined spectroscopic binary, producing a shape of the function that is not characteristic for the analysed molecular band head. To remove spectra with reduction problems or peculiarity that would result in wrongly determined strength of the Swan band, we are also analysing the slope of the spectral difference and its integral in the limits of the Swan bands. One of the spectral trends that we are trying to catch with those indicators is shown in Figure 4.19, where spectral difference and its linear fit are steeply rising at the border of the spectrum.

By visual inspection of the algorithm diagnostic plots shown in Figure 4.1, we limited a final selection to 400 spectra with the strongest carbon enhancement that was still visually recognisable. The last selected spectrum is shown in the Figure 4.17. Selection of spectra with lower enhancement, would introduce possibly wrong classification of stars whose enhancement is driven by spectral noise levels, data reduction or any other process that has subtle effect on the spectral shape.

4.4.2 Unsupervised classification

With numerous spectra of different stellar types, chemical composition, and degree of carbon enhancement, some of them might show different carbon features or be insufficiently distinctive to be picked out by the above supervised algorithm.

Another analysis technique, which is becoming increasingly popular is a dimensionality reduction procedure named t-distributed Stochastic Neighbor Embedding [t-SNE, 215], that has already proved to be beneficial in comparison and sorting of unknown spectral features of the same data set [84]. This is done by projecting the complete spectra onto a 2D plane by computation of similarities between all pairs of investigated spectra. It has been shown that the algorithm can cluster and distinguish spectra with absorption or emission features. The algorithm arranges spectra in a 2D plane, such that it clusters similar spectra together based on their similarity measure. As the transformation is variable and non-linear, the actual distance between two objects in a final 2D plane does not linearly depend on the spectral similarity measure. This property of the t-SNE algorithm ensures more homogeneous coverage of the 2D plane in comparison to other dimensionality reduction methods.

The t-SNE projection shown in Figure 4.2 was computed from normalised spectra between 4720 and 4890 Å. To maximise the number of analysed spectra, no other limiting cuts than the validity of the wavelength solution [bit 1 in `red_flag` set to 0 by reduction pipeline, 95] in this arm was used. This resulted in 588,681 individual spectra being analysed by the automatic unsupervised algorithm. This is $\sim 30k$

more spectra than in the case of supervised classification, where we applied more strict criteria for the selection of analysed spectra (Section 4.4.1).

Without any prior knowledge about the location of objects of interest in the obtained projection, we would have to visually analyse every significant clump of stars in order to discover whether the carbon-enhanced population is located in one of them. This can be simplified by adding the results of the supervised classification into this new projection. In Figure 4.2, the stars identified by the supervised classification are shown as blue dots plotted over grey dots representing all spectra that went into the analysis. The majority of blue dots are located in a clump on the left side of the projection. A high concentration of objects detected by a supervised method leads us to believe, that this isolated clump represents carbon-enhanced objects in the t-SNE projection. To select stars inside the clump, we manually drew a polygon around it.

Inspection of other blue labelled spectra outside the main clump revealed that their slight carbon enhancement could not be identified by the t-SNE similarity metric as the spectra comparison might have been dominated by another spectral feature.

Additional exploration of the t-SNE projection revealed two smaller groups of metal-poor carbon-enhanced spectra located inside ellipses shown in Figure 4.2. A confirmation that those regions are populated with metal-poor stars can be found in Figure 4.14 where the dots representing spectra in the projection are colour coded by $[Fe/H]$ and T_{eff} . To maximise the number of those objects in the published catalogue, we manually checked all undetected spectra in the vicinity of the detected ones. This produced additional 13 CEMP detections.

t-SNE limitation

While checking the local neighbourhood of some of the blue dots in Figure 4.2 that are strewn across the t-SNE projection we identified a possible limitation of our approach for the automatic detection of specific peculiar spectra if their number is very small compared to the complete data set. Figure 4.3 shows a collection of a few carbon-enhanced spectra embedded between other normal spectra that were taken out of the right ellipsoidal region in Figure 4.2. As they are quite different from the others they were pushed against the edge of a larger cluster in the projection, but their number is not sufficient to form a distinctive group of points in the projection. Therefore any automatic algorithm that would try to distinguish those objects based solely on a local density of points would most probably fail.

Another specific of the t-SNE projection that we must be aware of is how it computes the similarity between analysed spectra. Combined similarity, which is computed as a sum of per pixel differences, has zero knowledge about the location where in the spectrum those differences occur. The red spectrum in Figure 4.4 with a slight signature of carbon enhancement in the range between 4734 and 4737 Å has been placed among spectra with similar physical properties. Its slight carbon enhancement and comparable spectral noise to other spectra in its vicinity are probably the reason why it was placed in such a region of the t-SNE projection. This could be solved by using a smaller portion of the spectrum in a dimensionality reduction, which could at the same time lead to a loss of other vital information about a star.

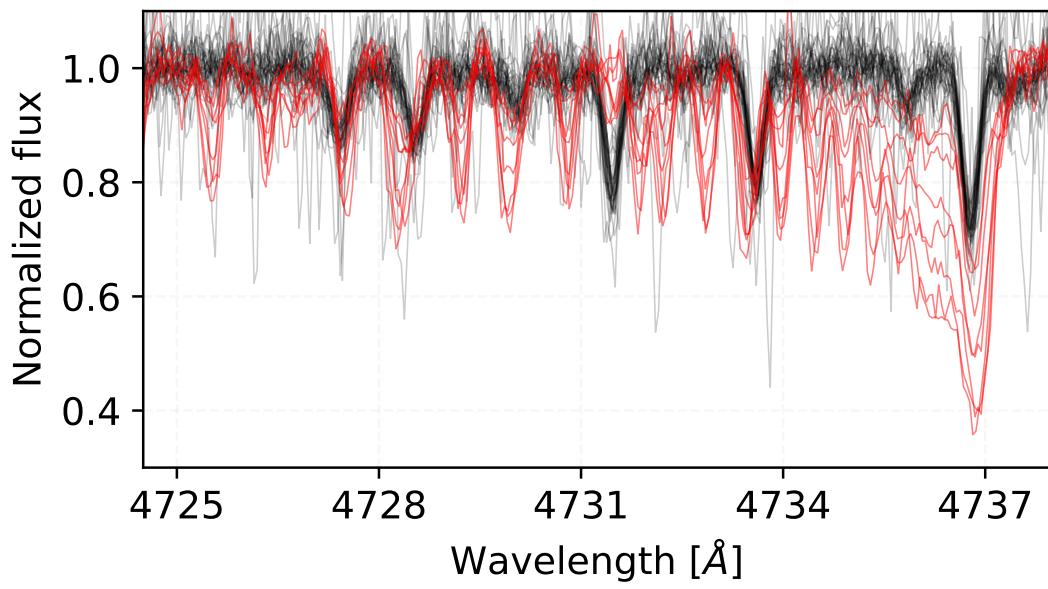


Figure 4.3: A collection of spectra that were determined to be mutually very similar by the t-SNE algorithm. Out of 46 spectra inside the right black ellipse in Figure 4.2 we identified 8 carbon-enhanced spectra with visually very different and distinctive spectrum in the region from 4734 to 4737 Å that is also depicted in this figure. For easier visual recognizability, they are coloured in red.

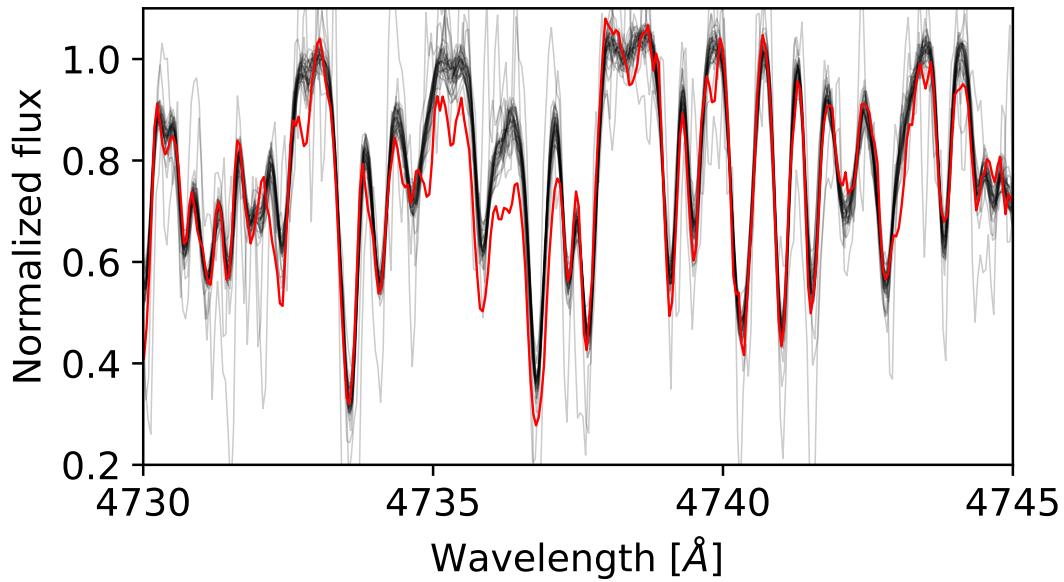


Figure 4.4: Spectral comparison between one of the detected carbon-enhanced stars in red and its 30 closest neighbours in the t-SNE projection shown as black curves. Enhancement in the spectrum was probably not sufficiently distinct and was dominated by the spectral noise. Therefore the spectrum was placed among other physically similar spectra without visible enhancement.

4.5 Candidate characteristics

The final list of detected carbon-enhanced stars consists of 918 stars, corresponding to 993 spectra detected by at least one of the described methods. Among them, 63 stars were observed and identified at least twice and up to a maximum of four times. Those identifications belong to repeated observations that were performed at different epochs. Because not all of the observed spectra were considered in the classification procedure (due to the limitations described in Section 4.4) this is not the final number of stars with repeated observations. By searching among the complete observational data set, the number of carbon-enhanced stars with repeated observations increases to 90.

Out of those 90 stars, every repeated observation of 56 stars was classified as being carbon-enhanced. In total, we detected 76.5 % of the carbon-enhanced spectra among repeated observations where at least one of the repeats have been classified as having enhanced carbon features in its spectrum. The unclassified instances usually have a low SNR value that could decrease their similarity value towards other carbon-enhanced stars in the t-SNE analysis or have incorrect stellar parameters and were therefore compared to an incorrect median spectra during the supervised analysis.

4.5.1 Radial velocity variations

With repeated observations in the complete observational data set, we can look into measured radial velocities and investigate a number of possible variables that should be high for certain types of carbon-enhanced objects [216]. Taking into account all of the repeated observations in our data set and not just the repeats among the identified spectra, 52 out of 90 stars show a minimum velocity change of 0.5 km s^{-1} (70 stars with minimum change of 0.25 km s^{-1}) and a maximum of 45 km s^{-1} in different time spans ranging from days to years. The detailed distribution is presented by Figure 4.5. That kind of change can hint at the presence of a secondary member or at intrinsic stellar pulsation [217, 218, 219], as carbon-enhanced stars are found among all long period variable classes [Mira, SRa, and SRb, 220, 221]. Follow-up observations are needed to determine their carbon sub-class and subsequently the reason behind variations of radial velocity.

Visual inspection of variable candidates revealed that none of them shows obvious multiplications of absorption spectral lines, a characteristic of a double-lined binary system. Therefore we can conclude that none of them is a binary member in which both components are of comparable luminosity and a difference between their projected radial velocities is high enough to form a double-lined spectrum. From our simulations with median spectra, such line splitting becomes visually evident at the velocity difference of $\sim 14 \text{ km s}^{-1}$. If the components do not contribute the same amount of flux, the minimal difference increases to $\sim 20 \text{ km s}^{-1}$.

Chemical peculiarity of a dwarf carbon-enhanced star (dC) that exhibits enhancement of C₂ in its spectra could be explained by its interaction with a primary star in a binary system [222]. Chemically enhanced material is thought to be accreted from the evolved AGB companion. Less than thirty of such systems, that show signs of the existence of an invisible evolved companion who might have enriched a dC by the carbon, have been identified spectroscopically to date [222, 223, 224], giving us the possibility to greatly increase the list with every additional confirmed

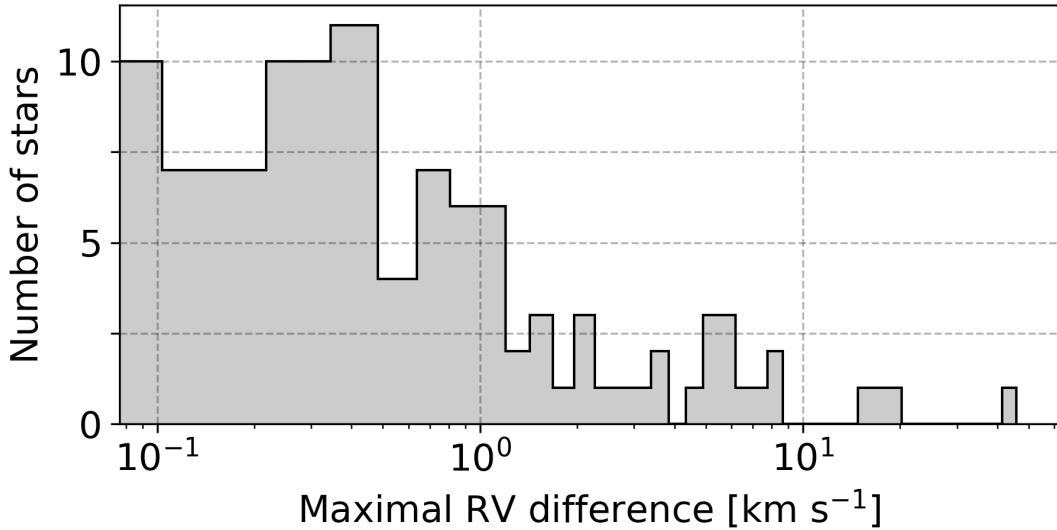


Figure 4.5: Distribution of maximal velocity change between repeated observations of the stars that were classified as carbon-enhanced.

object. The only detected dC star (for criteria see Section 4.5.2) with repeated observations shows that its radial velocity is unchanged on the order of 0.1 km s^{-1} during the 2 years between consecutive observations. Hence, it cannot be classified as a possible binary system from those two observations alone. The lack of a clear evidence for binarity among dC stars, especially among the most metal-poor, can also be explained by another enrichment mechanism. Farihi *et al.* [225] showed that a substantial fraction of those stars belongs to the halo population based on their kinematics information. Combined with the results of Yoon *et al.* [162] that classified the prototype dC star G 77-61 as a CEMP-no star, that are known to have intrinsically low binarity fraction [202, 226], their carbon-enhancement may be of a primordial origin.

4.5.2 Stellar parameters

For the analysis of stellar parameters, we used values determined by *The Cannon* data interpolation method that was trained on actual observed HERMES spectra. To exclude any potentially erroneous parameter, we applied a strict flagging rule of `flag_cannon=0` [an extensive description of flagging procedure can be found in 99], thus obtaining a set of 347 objects with trustworthy stellar parameters. Such a large percentage of flagged objects could be attributed to their nature as an additional elemental enhancement that we are looking for might not be a part of the training set. A raised quality flag would hint that the spectrum is different from any other in the training set or that the fit is uncertain and has a large χ^2 . Therefore flagged parameters have to be used with care, especially on the border of, and outside the training set.

The majority (338) of the unflagged detected objects are giants and only 9 are confirmed to be dwarf stars based on their spectroscopic stellar parameters (Figure 4.6).

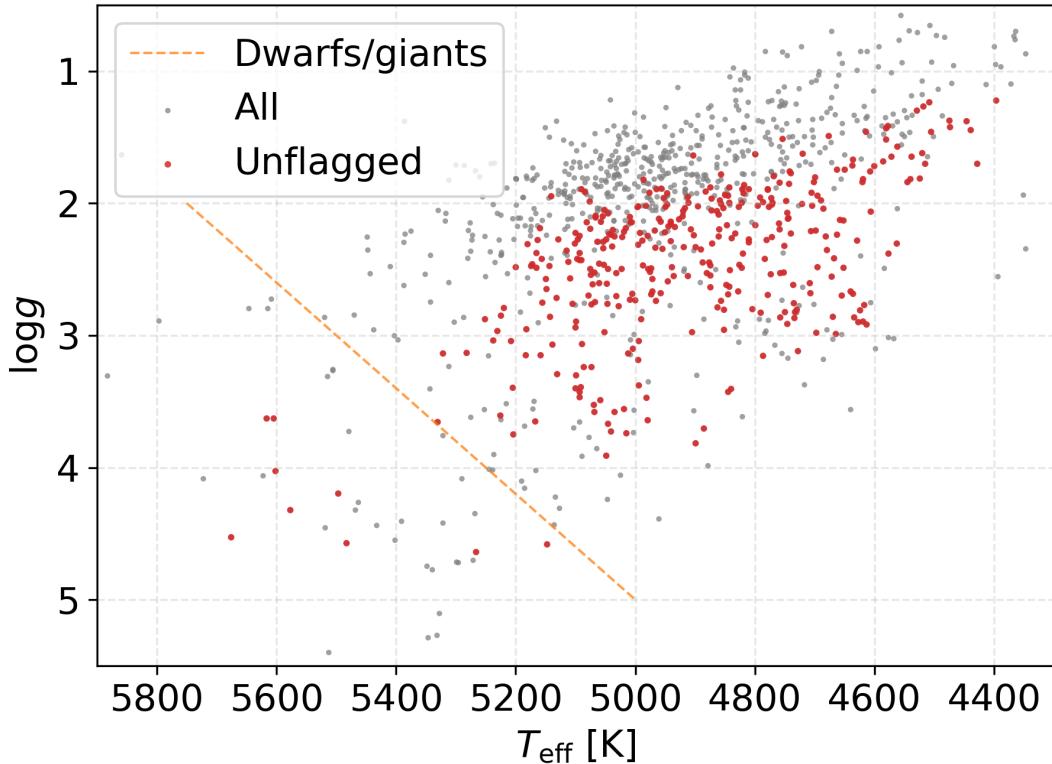


Figure 4.6: Kiel diagram for a subset of 338 detected carbon-enhanced stars with valid stellar parameters in red. Uncertain positions of flagged stars are shown with grey dots. Dashed orange line illustrates the border between giants and dwarfs.

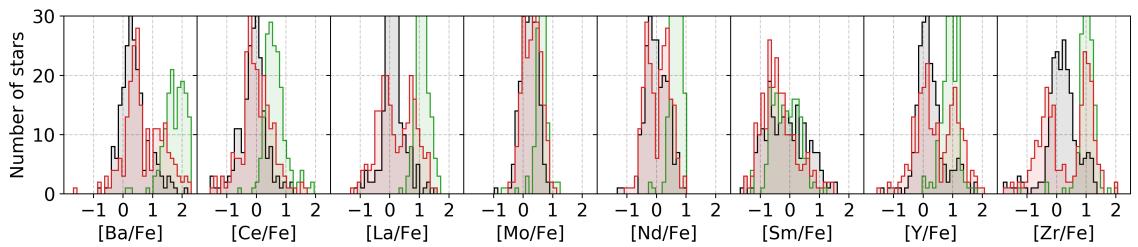


Figure 4.7: Distribution of s-process element abundances for stars in three different groups. The most enhanced group in green represent carbon-enhanced stars located in the t-SNE selected clump of stars. The red distribution presents all other detections that are placed around the projection, and outside the clump. As a control group, the same distribution in black is shown for their closest t-SNE neighbours, therefore the black and red distribution contain an equal number of objects. No abundance quality flags were used to limit abundance measurements.

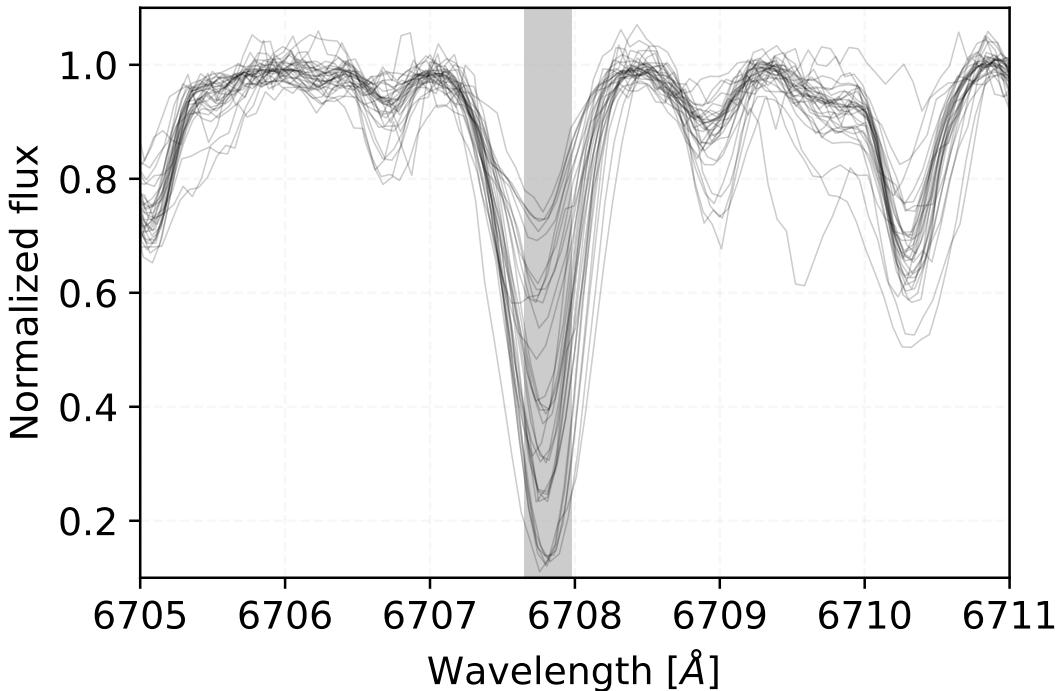


Figure 4.8: Spectral subset of 32 lithium-rich carbon-enhanced stars among the identified stars. The highlighted wavelength region is used by *The Cannon* to determine the lithium abundance of a star.

4.5.3 S-process elements

Focusing on a spectral signature of the detected objects inside and outside the t-SNE selected clump (Figure 4.2) we can further investigate which spectral feature might have contributed to their separation. The distributions of their abundances in Figure 4.7 and strength of spectral features corresponding to the same elements in Figure 4.20 hints to an enhancement of s-process elements among stars inside the selected clump. This additional enhancement might be another reason, besides the carbon enhancement, for the algorithm to cluster all of those stars as being different from the majority of spectra.

4.5.4 Lithium abundance

The derivation of elemental abundances for known carbon-enhanced stars has shown that some of them can exhibit strongly enhanced levels of Li in their atmosphere [227]. Lithium is thought to be produced by hot-bottom burning [228] and brought to the surface from the stellar interior. Investigation of the Li line at 6707 Å revealed 32 of such stars. Their spectra, centred around the Li feature, show a greatly varying degree of absorption in Figure 4.8.

4.5.5 Sub-classes

Following a revision of the original MK classification [229] introduced by Barnbaum *et al.* [230], carbon stars are separated into five different classes named C-H, C-R, C-J, C-N, and Barium stars. Of all the spectral indices proposed for the spectral classification, we are only able to measure a small part of Swan C_2 bands and Ba II line at 6496 Å. For a more detailed classification of detected objects into proposed classes, we would need to carry out additional observations with a different spectroscopic setup to cover all the significant features.

Additionally, the features caused by the $^{13}\text{C}^{12}\text{C}$ molecule are strongly enhanced only for a handful of spectra in our data set, therefore we did not perform any isotopic ratio analysis or identification of possible C-J objects, which are characterised by strong Swan bands produced by the heavier isotopes.

According to the abundance trends presented in Section 4.5.3 and the classification criteria defined by Barnbaum *et al.* [230], we could argue that the stars selected from the t-SNE projection belong to the C-N sub-class. Their s-process elements are clearly enhanced over Solar values (Figure 4.7), but the actual values should be treated with care as they are mostly flagged by *The Cannon*. This uncertainty might come from the fact that the training set does not cover carbon-enhanced stars and/or stars with such enhancement of s-process elements.

4.5.6 Match with other catalogues

In the literature we can find numerous published catalogues of carbon-enhanced (CH) stars [172, 175, 231] and CEMP stars [199, 232, 233, 234, 235, 236] observed by different telescopes and analysed in inhomogeneous ways. Most of those analyses were also performed on spectra of lower resolving power than the HERMES, therefore some visual differences are expected for wide molecular bands. By matching published catalogues with the GALAH observations that were analysed by our procedures, we identified 44 stars that matched with at least one of the catalogues. Of these, 28 were found in CH catalogues and 16 in CEMP catalogues.

From the stars recognised as CEMPs in the literature, we were able to recover only 1 of them. Visual assessment of the diagnostic plots provided by our analysis pipeline proved that the remaining 15 CEMP matches do not express any observable carbon enhancement in Swan bands and were therefore impossible to detect with the combination of our algorithms. The reason for this difference between our and literature results might be in the CEMP selection procedure employed by the aforementioned literature. Every considered study selects their set of interesting stars from one or multiple literature sources based on values of [M/H] and [C/Fe] that were measured from the atomic spectral lines and not molecular lines.

The match is larger in the case of CH matches, where we were able to confirm 11 out of 33 possible matched carbon-enhanced stars. As the observed molecular bands are prominent features in the spectra, we explored possible reasons for our low detection rate. Visual inspection of spectra for the remaining undetected matched stars proved that they also show no or barely noticeable carbon enhancement in the spectral region of Swan bands, therefore reason must lie in the detection procedures used in the cited literature. Christlieb *et al.* [172] used low-resolution spectra to evaluate enhancement of C_2 and CN bands. The results are also summarised in their electronic table. In here, all of our undetected stars are marked to contain

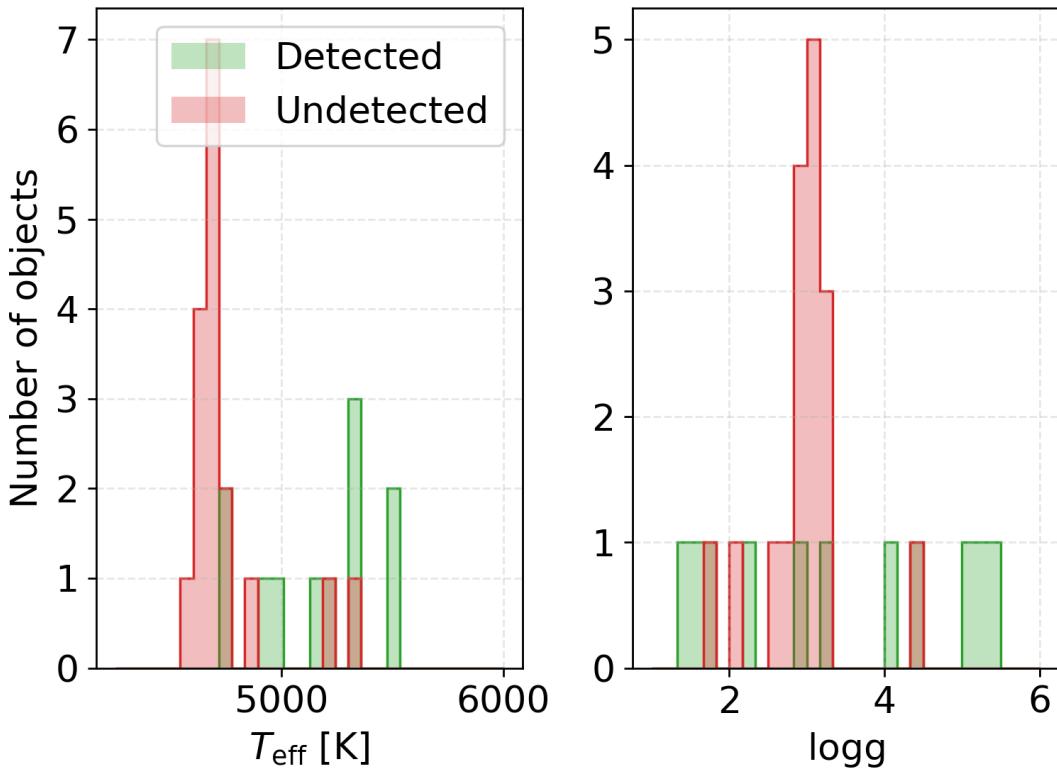


Figure 4.9: Comparison between the stellar parameters of detected (green histogram) and undetected (red histogram) carbon-enhanced stars found in literature.

enhanced CN bands but no C₂ bands. Combining this with Figure 4.9 we speculate that those stars occupy a narrow range of parameter space where C₂ is not expressed and therefore undetectable in the HERMES spectra.

Number of successfully detected stars matched between the surveys could also be influenced by different excitation temperatures of analysed carbon-rich molecules. Frequently studied photometric G-band, that is not present in our spectra, covers a spectral region rich in CH molecule features whose temperature dependence is different than for a C₂ molecule. Presence of those bands is identified by classifying a carbon-enhanced star into C-H sub-class (see Section 4.5.5). As we detected all C-H stars identified by Ji *et al.* [175], that are also present in the GALAH data set, we are unable to discuss about the selection effect in the T_{eff} range between ~ 5100 and ~ 5300 K where those three stars were found.

The position of all stars matched with the literature is also visualised on the t-SNE projection in Figure 4.15, where it can be clearly seen that they lie outside the selected clump with identified carbon enhancement and are strewn across the projection. Close inspection of spectra that are spatially near the aggregation of CEMP stars from the literature, revealed no visible carbon enhancement. The enhancement is present neither in form of molecular bands nor expressed as stronger atomic carbon line. They therefore are indistinguishable from other metal-poor stars with similar physical parameters.

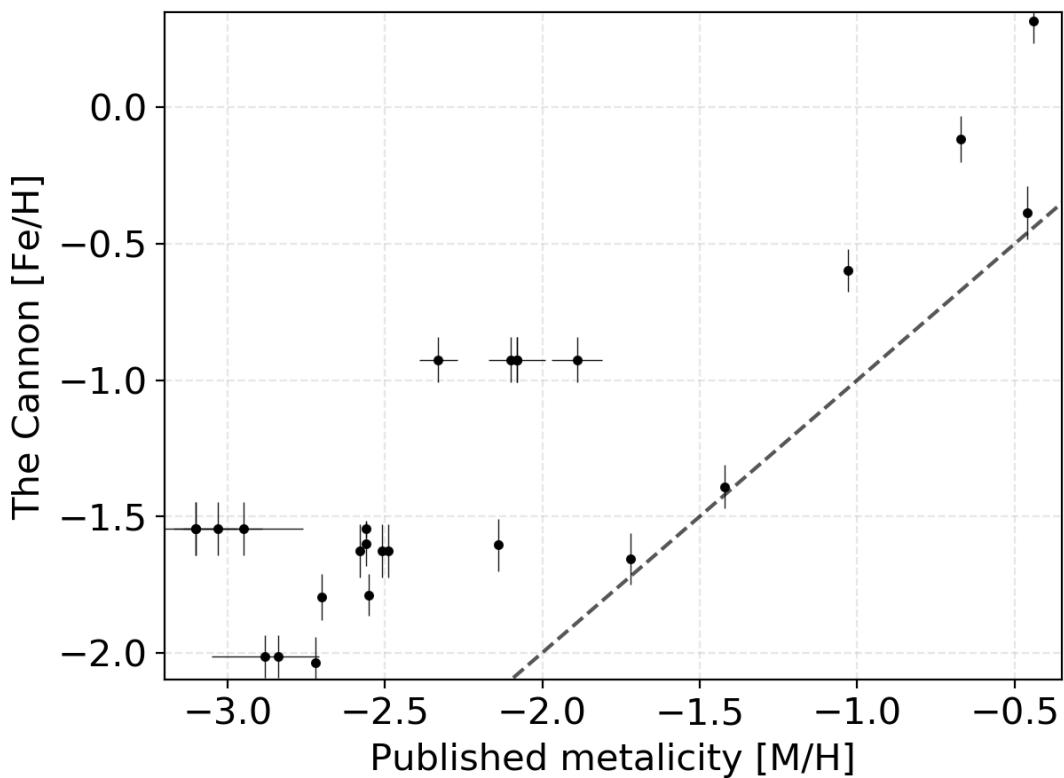


Figure 4.10: Correlation between published metallicities and *The Cannon* iron abundance for the stars that were classified as CEMPs in the literature. As some of those stars were taken from multiple literature sources, we also have multiple determinations of $[M/H]$ for them. This can be identified as horizontal clusters of dots at different $[M/H]$, but with the same $[Fe/H]$. Where available, uncertainties of parameters are shown. The dashed line follows a 1:1 relation.

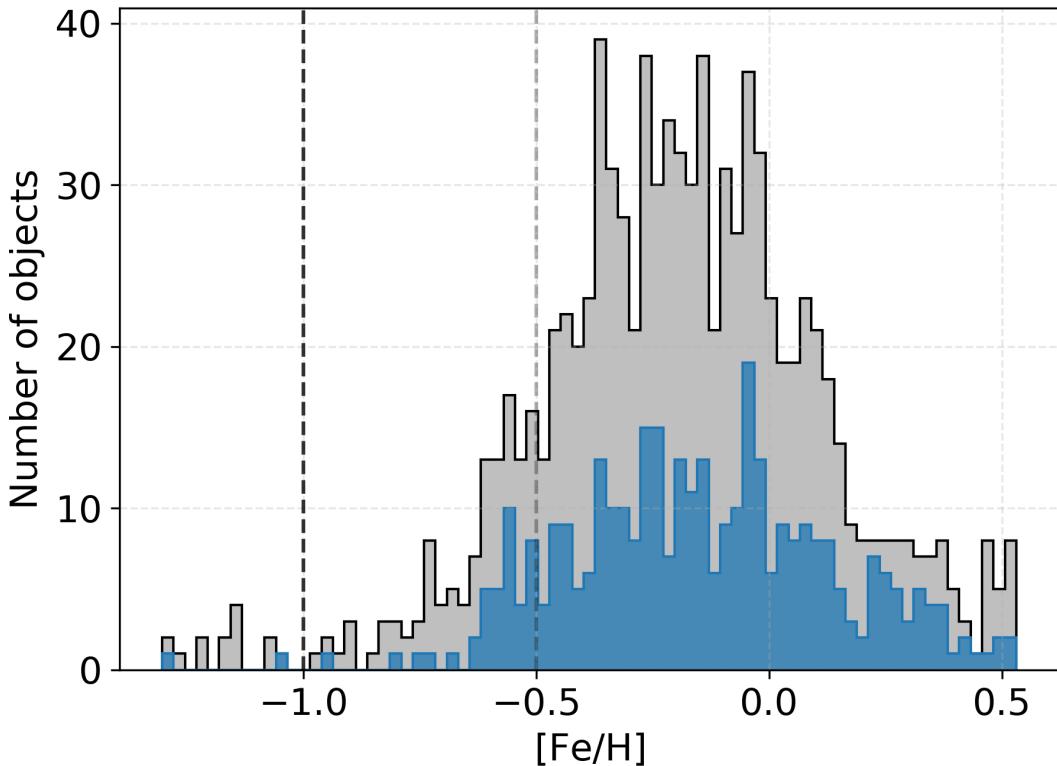


Figure 4.11: Histogram of $[\text{Fe}/\text{H}]$ for detected carbon-enhanced stars with valid *The Cannon* stellar parameters in blue and for every detected carbon-enhanced star in grey. Two vertical lines are located at iron abundances of -1.0 and -0.5 .

4.6 Metal-poor candidates

CEMP stars are defined in the literature as having low metallicity $[\text{M}/\text{H}] < -1$ and strong carbon enrichment $[\text{C}/\text{Fe}] > +1$. In the scope of this analysis, we assume that our measurement of $[\text{Fe}/\text{H}]$ is a good approximation for the metallicity. To be sure about this we compared $[\text{M}/\text{H}]$ values of CEMP stars found in the literature and $[\text{Fe}/\text{H}]$ derived by *The Cannon* for the same stars. The relation between them is shown in Figure 4.10. We see that our values start deviating from the published values at metallicities below -1.5 . Below that threshold the differences are in the range of ~ 1 dex, but the same trend is obvious for both data sets. The uncertainty of the published $[\text{M}/\text{H}]$, derived from multiple sources, can reach up to 0.5.

Taking unflagged *The Cannon* parameters and abundances of the detected objects we can determine possible CEMP candidates among our sample. As also shown by Figure 4.11 our set of carbon-enhanced stars consists of 41 objects with $[\text{Fe}/\text{H}] < -0.5$ and 2 objects with $[\text{Fe}/\text{H}] < -1.0$. If we also include potentially incorrect parameters, the number of objects with $[\text{Fe}/\text{H}] < -1.0$ increases to 28, which is equal to 2.8 % of detected carbon-enhanced spectra. In any case, none of them has a valid determination of carbon abundance. Analysing HERMES spectra in order to determine carbon abundance is difficult because the automatic analysis is based on only one very weak atomic absorption line that is believed to be free of any blended lines. Consequently, we are also not able to measure the $[\text{C}/\text{O}]$ abundance ratio, as a

majority of determined [C/Fe] abundances is flagged as unreliable. Complementary observations are needed to determine the abundance and confirm suggested CEMP candidates.

A low number of metal-poor candidates could also be explained by the specification of the HERMES spectrograph as its spectral bands were not selected in a way to search for and confirm most metal-poor stars. With the release of *Gaia* DR2 data [237], stars low/high-metallicity could also be compared with their Galactic orbits. To determine the distribution of detected stars among different Galactic components, we performed an orbital integration in `MWPotential2014` Galactic potential using the `galpy` package [238]. In order to construct a complete 6D kinematics information, *Gaia* parallax and proper motion measurements were supplemented with the GALAH radial velocities. Results shown in Figure 4.12 suggest that our CEMP candidates could belong to two different components of the Galaxy. Stars with maximal $z < 4$ kpc most probably belong to the thick disk and stars with $z > 5$ kpc to the halo population that is inherently metal-poor. This is also supported by their angular momentum in the same plot and their Galactic velocities shown in Figure 4.21.

When looking at the distribution of [Fe/H] for the complete set of observed stars, we find a comparable distribution as for carbon-enhanced stars. Similarly, about 1.8 % of stars are found to be metal-poor with $[\text{Fe}/\text{H}] < -1.0$.

4.7 Follow-up observation

To further classify and analyse one of the detected objects, a star with 2MASS identifier J11333341-0043060 was selected for a follow-up observation. We acquired its high-resolution Echelle spectrum (with the resolving power $R \sim 20,000$), using a spectrograph mounted on the 1.82 m Copernico telescope located at Cima Ekar (Asiago, Italy). Because only a few of our detected candidates are observable from the Asiago observatory, we selected the best observable CEMP candidate, whose [Fe/H] was determined by *The Cannon* to be -0.96 . The selected star, with $V = 12.79$, was on the dark limit of the used telescope, therefore low SNR was expected. The one-hour long exposure of the selected object was fully reduced, normalised order by order, and shifted to the rest frame.

Although the acquired spectrum covers a much wider and continuous spectral range (from 3900 to 7200 Å) than the HERMES spectra, only subsets, relevant for the classification of carbon-enhanced stars are presented in Figure 4.13. They were identified by visually matching our observed spectrum with the published moderate-resolution spectral atlas [230] of peculiar carbon stars. Where available, the GALAH spectrum is shown alongside the Asiago spectrum. Carbon enhancement is not expected to vary over a period of several years, therefore both spectra should show similar features. The second and fourth panel in Figure 4.13 confirm that both observations indicate a similar degree of carbon enhancement.

Following the classification criteria of carbon stars, we determined that the star belongs to the C-H sub-class. The definitive features for this class are strong molecular CH bands, prominent secondary P-branch head near 4342 Å (top panel in Figure 4.13), and noticeable Ba II lines at 4554 and 6496 Å [176], which are all present in the spectrum. The star definitely does not have a high ratio between ^{13}C and ^{12}C isotopes as the Swan features corresponding to ^{13}C are clearly not present, therefore

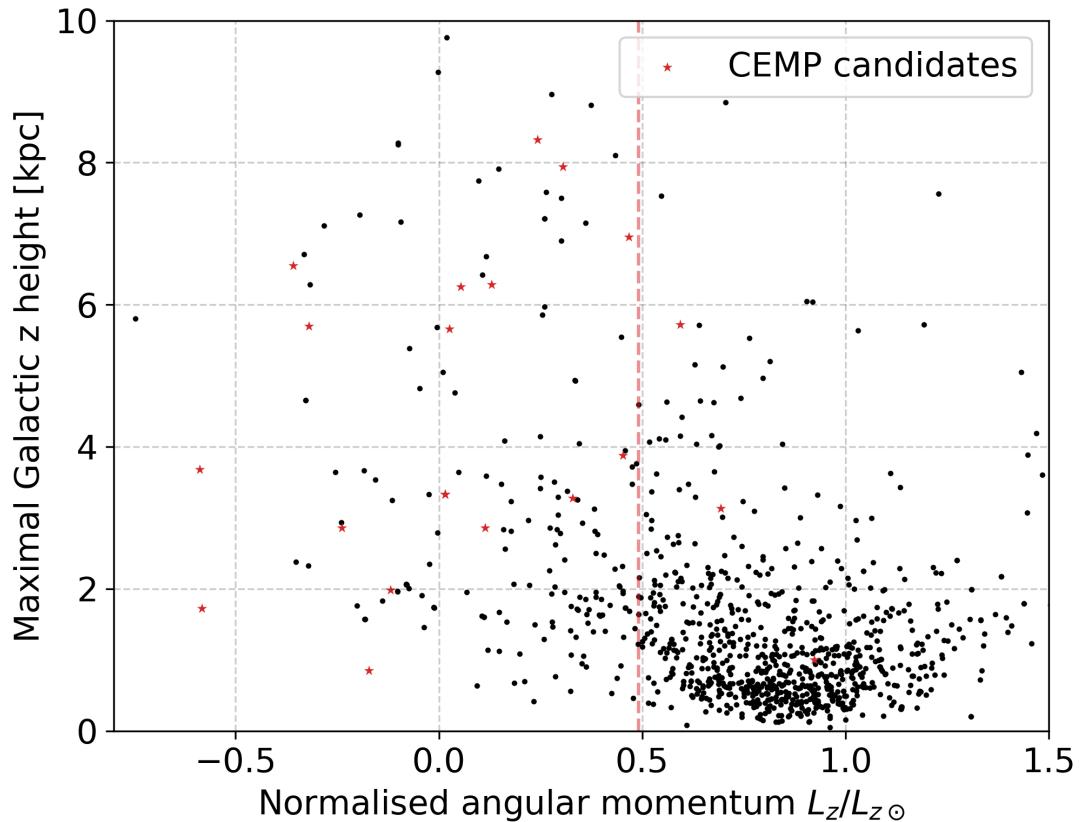
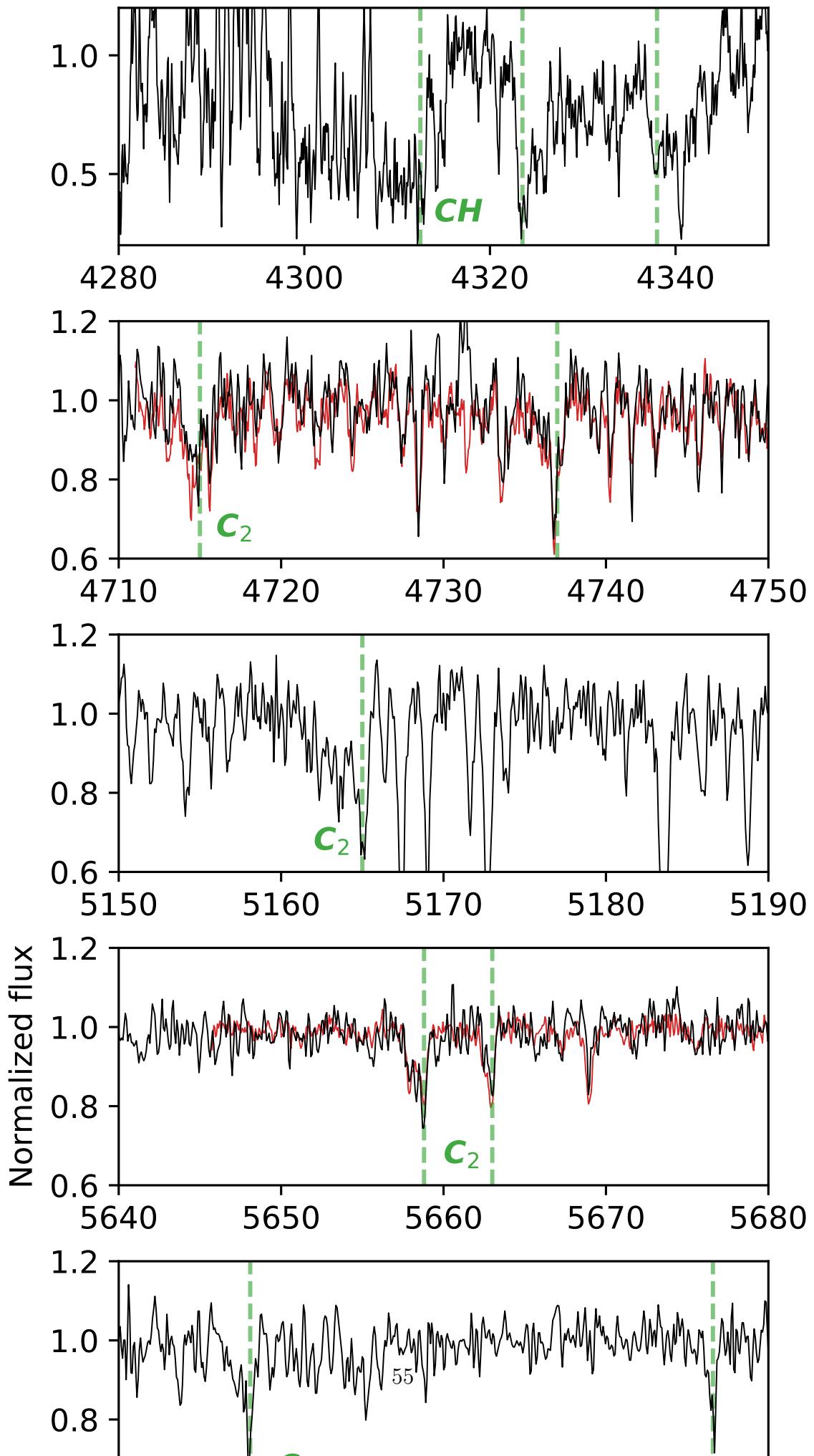


Figure 4.12: Distributions of maximal height above/below the Galactic plane reached by detected stars on their orbit around the centre of the Galaxy in comparison to their normalised angular momentum L_z . Vertical dashed line at $1000 \text{ km s}^{-1} \text{ kpc}$ highlights the transition from the halo to the disk population, where a majority of the halo stars is located below this threshold [the threshold was visually estimated from similar plots in 239]. CEMP candidates are marked with star symbols.



it can not be of a C-J sub-class.

Following the current state of knowledge [216, 240, 241] that most, if not all, C-H stars show clear evidence for binarity, we compared the radial velocity between both observations. They hint at the variability of the object as the follow-up radial velocity ($126.75 \pm 1.63 \text{ km s}^{-1}$) deviates by more than 3 km s^{-1} from the velocity ($123.43 \pm 0.08 \text{ km s}^{-1}$) observed as part of the GALAH survey. The time span between the two observations is more than 2.5 years, where the exact JD of the observation is 2458090.702 for the Asiago spectrum, and 2457122.095 for the GALAH spectrum. Further observations along the variability period would be needed to confirm whether it is a multiple stellar system.

4.8 Conclusions

This work explores stellar spectra acquired by the HERMES spectrograph in order to discover peculiar carbon-enhanced stars, which were observed in the scope of multiple observing programmes conducted with the same spectrograph.

We show that the spectra of such stars are sufficiently different from other stellar types to be recognisable in high-resolution spectra with limited wavelength ranges. This can be done using a supervised procedure, where some knowledge about the effects of carbon enhancement on the observed spectra is put into the algorithm, or using an unsupervised method. The latter was used to identify observed stars solely on the basis of acquired spectra. By combining both methodologies we identified 918 unique stars with evident signs of carbon enhancement of which 12 were already reported in the literature. Out of all matched objects from the literature, we were unable to detect and confirm 16 (57 %) CH and 15 (93 %) CEMP stars with our procedures. As some of those objects were proven to contain carbon enhancement detectable outside the HERMES wavelength ranges, this would have to be taken into account to say more about the underlying population of carbon-enhanced stars. In addition to a detection bias imposed by the analysis of C_2 bands and exclusion of CN, and CH molecular bands that might be excited in different temperature ranges, varying degree of carbon-enhancement also has to be accounted for accurate population studies. As shown by Yoon *et al.* [162], CEMP stars can be found within a wide range of absolute carbon abundances. When an object selection is performed with a pre-defined threshold, as in the case of our supervised methodology, this may reduce the number of objects in only one of the sub-classes. In the case of CEMP stars, this selection may influence a number CEMP-no stars that are known to have lower absolute carbon abundance [162].

The identified objects were separated into dwarf and giant populations using their stellar atmospheric parameters that were also used to select possible CEMP candidates. All of the detections, with multiple observations at different epochs, were investigated for signs of variability. More than half of the repeats show signs of variability in their measured radial velocities. This could be an indicator that we are looking at a pulsating object or a multiple stellar system.

With a follow-up observation of one of the identified stars, we were able to confirm the existence of carbon-rich molecules in its atmosphere in a wider wavelength range. The acquired spectrum was also used to determine its sub-class. Variation in radial velocity points to a possible variable nature of the star or binarity that is common for C-H stars.

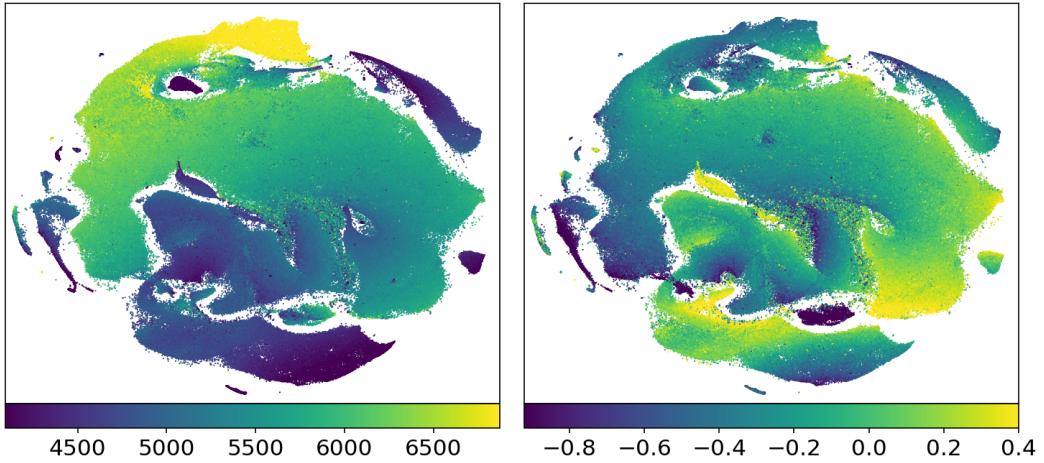


Figure 4.14: Spatial distribution of all available measurements of T_{eff} (left panel) and $[\text{Fe}/\text{H}]$ (right panel) as determined by *The Cannon*. Dots, representing analysed spectra in the t-SNE projection, are colour coded by their parameter values. Colours and their corresponding values are explained by a colourbar under the graph.

Follow-up observations are required to confirm variability of radial velocities observed for some of the detected carbon-enhanced stars and further investigate their nature. Careful spectral analysis, with the inclusion of carbon enhancement in models, is needed to confirm the metallicity levels of the metal-poor candidates.

The list of detected stars presented in this paper is accessible as electronic table through the CDS. Detailed structure is presented in Table 6.8. The list also includes stars from the literature, matched with our observations, for which we were unable to confirm their carbon enhancement. The list could be used to plan further observations, allowing a better understanding of these objects.

4.9 Table description

In the Table 6.8 we provide a list of metadata available for every object detected using the methodology described in this paper. The complete table of detected objects and its metadata is available only in electronic form at the CDS.

Table 4.1: List and description of the fields in the published catalogue of detected objects and objects matched with multiple literature sources.

Field	Unit	Description
source_id		<i>Gaia</i> DR2 source identifier
sobject_id		Unique internal per-observation star ID
ra	deg	Right ascension from 2MASS, J2000
dec	deg	Declination from 2MASS, J2000
det_sup	bool	Detected by supervised fitting method
det_usup	bool	Detected by t-SNE method
swan_integ		Swan band strength if determined
teff	K	<i>The Cannon</i> effective temperature T_{eff}
e_teff	K	Uncertainty of determined T_{eff}
logg		<i>The Cannon</i> surface gravity $\log g$
e_logg		Uncertainty of determined $\log g$
feh		<i>The Cannon</i> iron abundance [Fe/H]
e_feh		Uncertainty of determined [Fe/H]
flag_cannon	int	<i>The Cannon</i> flags in a bit mask format
type		G for giants and D for dwarfs
rv_var	bool	Is radial velocity variable
li_strong	bool	Shows strong lithium absorption
cemp_cand	bool	Is star CEMP candidate
bib_code		ADS bibcode of the literature match

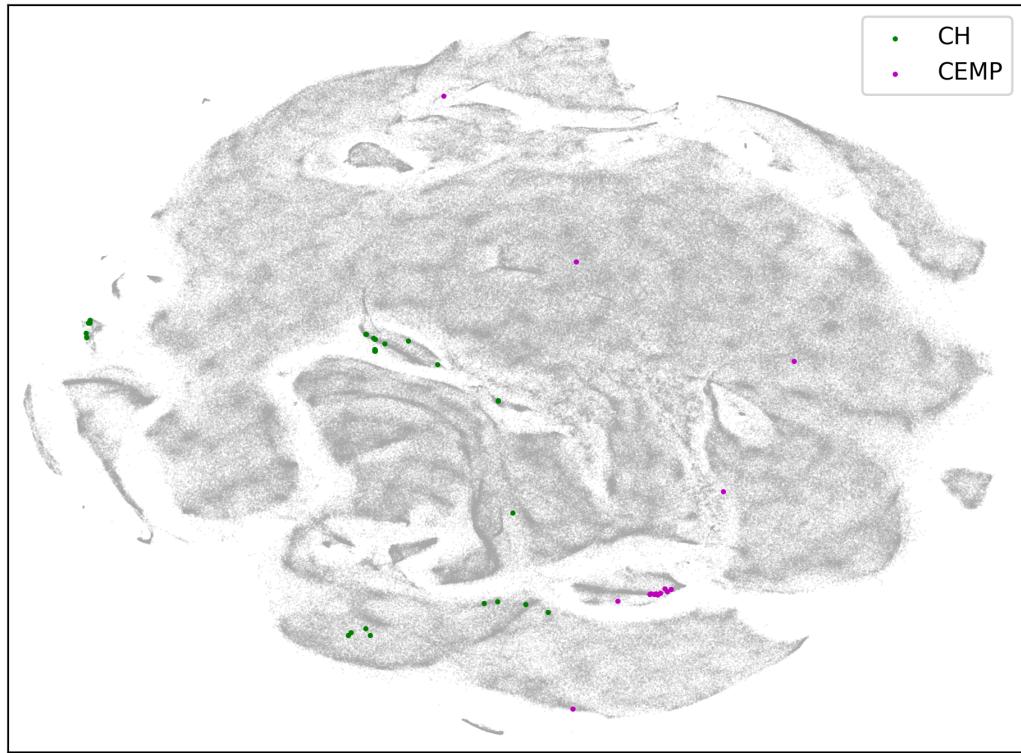


Figure 4.15: t-SNE projection with marked known carbon-enhanced and CEMP objects from multiple different catalogues found in the literature that are also part of our analysed set of spectra.

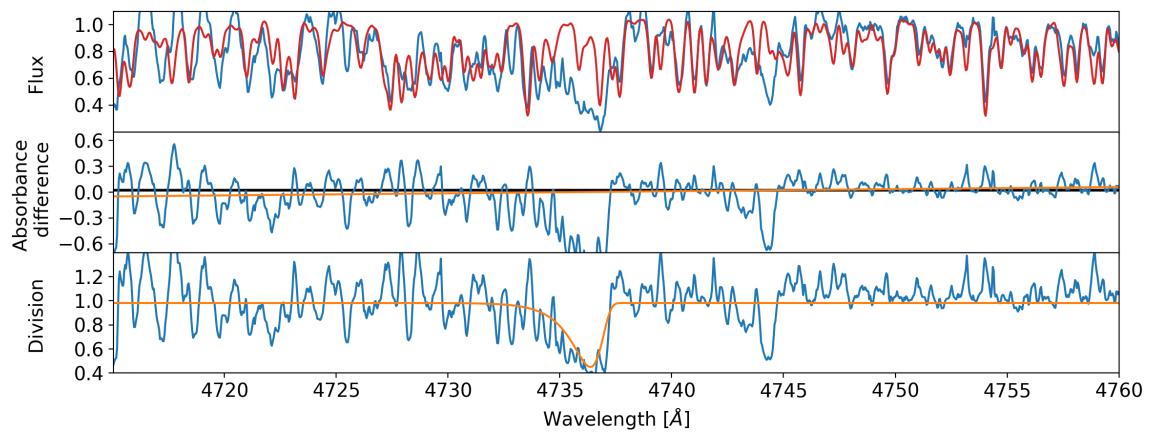


Figure 4.16: Equivalent plot as in the Figure 4.1 but presenting an example of a metal-rich star with multiple strong Swan features around 4737 and 4745 Å. Presented star has a 2MASS identifier J13121354-3533120 and is known Galactic carbon star [231].

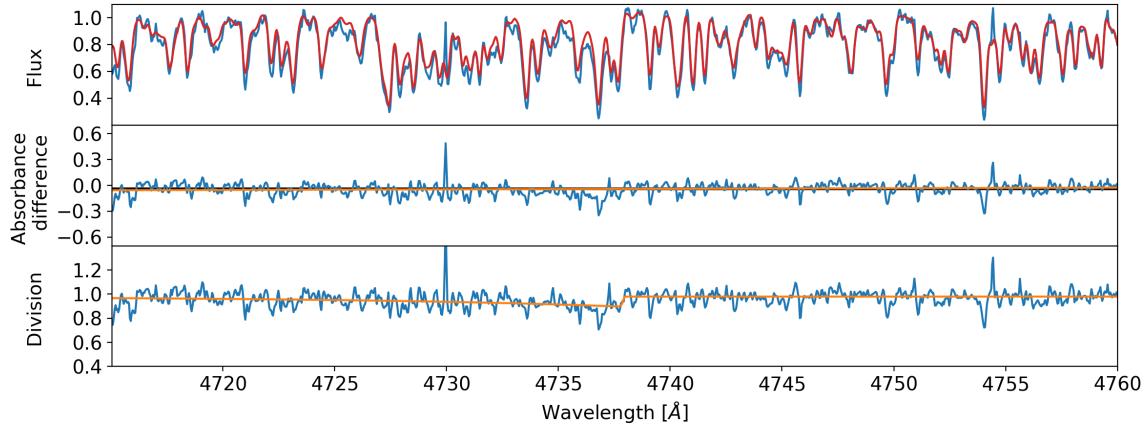


Figure 4.17: Equivalent plot as in the Figure 4.1 showing the last of 400 spectra, ordered by their degree of carbon enhancement, selected by the supervised methodology.

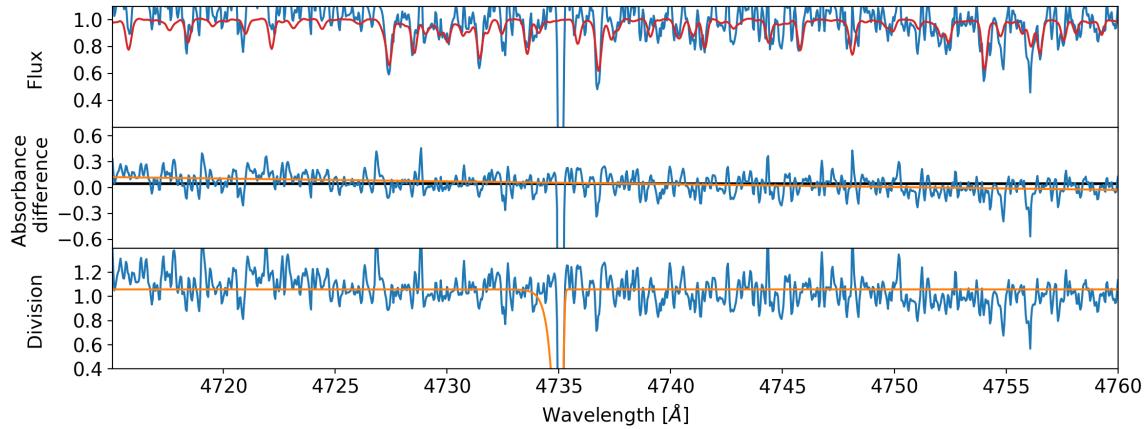


Figure 4.18: Equivalent plot as in the Figure 4.1 but representing grossly over exaggerated carbon enhancement by a fit that describes a reduction problem (a cosmic ray in a subtracted sky spectrum).

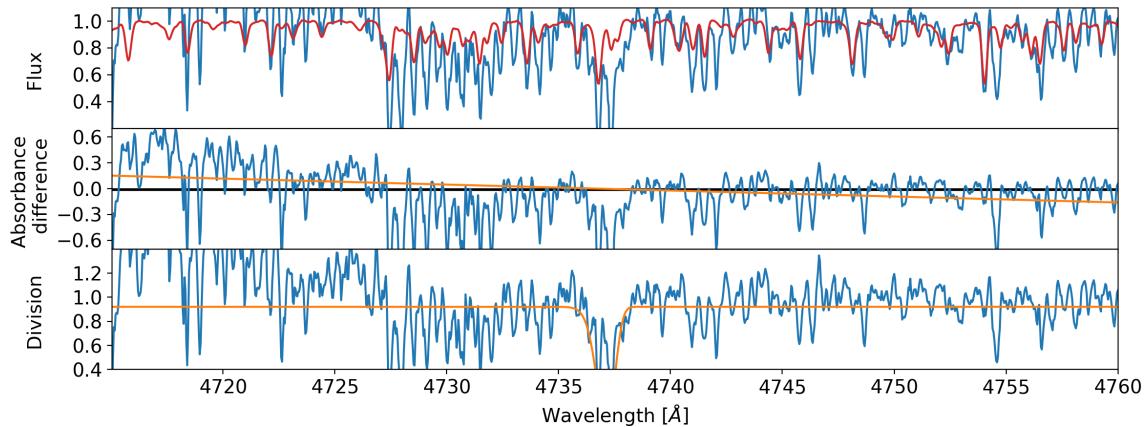


Figure 4.19: Equivalent plot as in the Figure 4.1 but representing a fit to absorption lines of a double-lined spectroscopic binary. Final fit is not skewed as would be expected in the case of carbon enhancement.

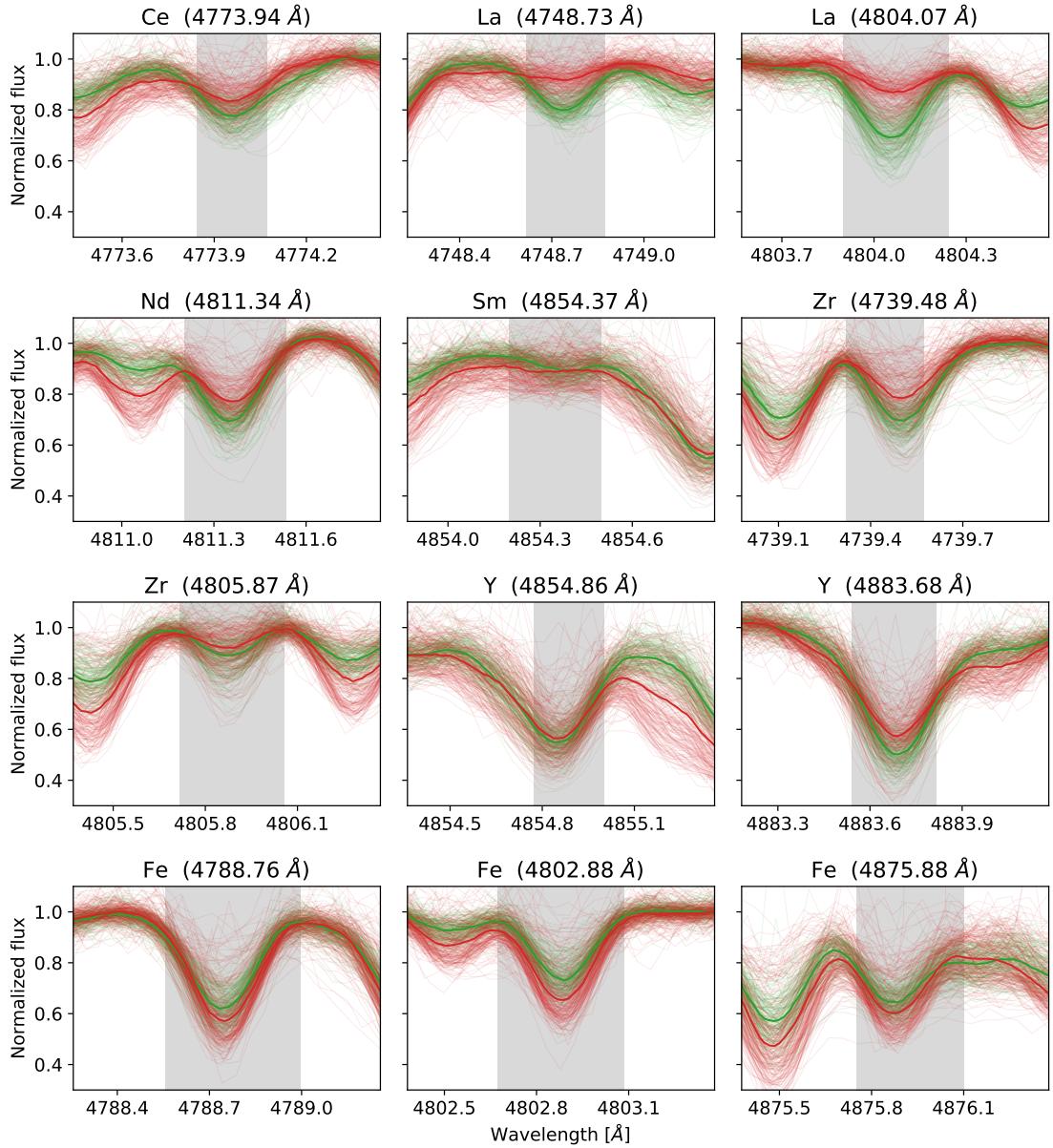


Figure 4.20: Spectral subset around the absorption features in the blue arm that were used to determine abundances of Fe and s-process elements. Same colour coding is used as in Figure 4.7. Spectra inside the t-SNE determined clump are shown in red, and outside it in green. Median of all spectra is shown with a bold line of the same colour. The shaded area gives the wavelength range considered in the computation of abundances.

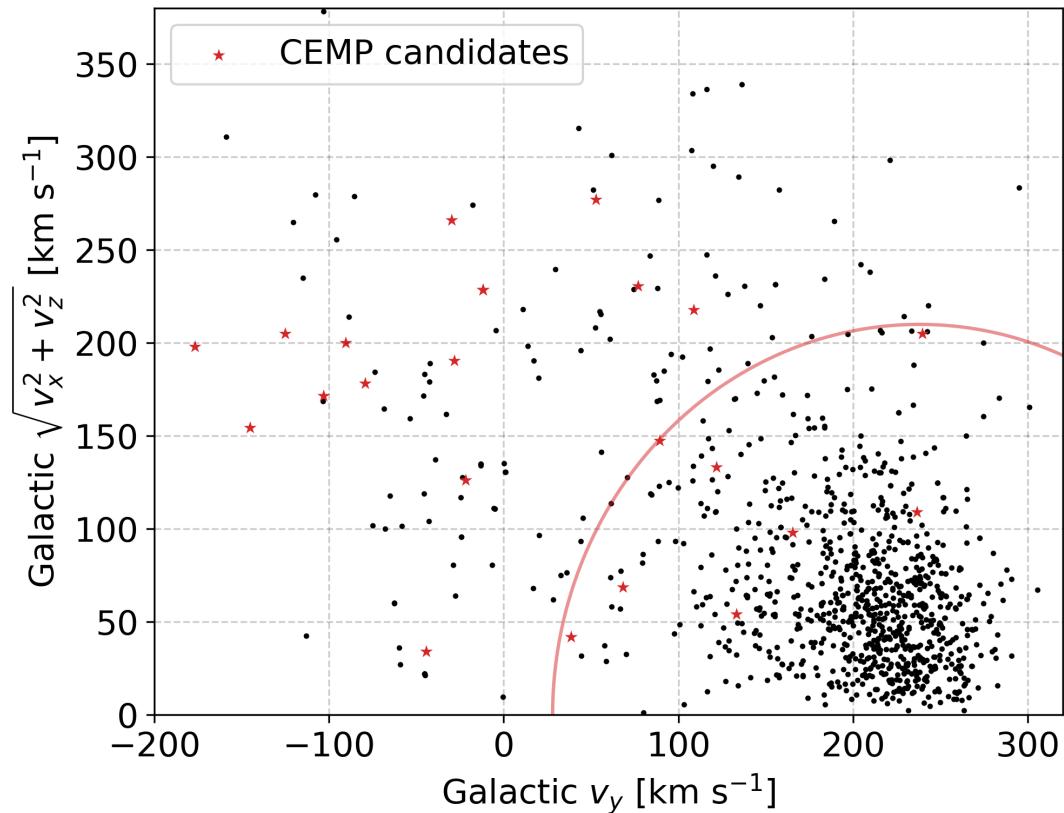


Figure 4.21: Toomre diagram used to identify possible local halo stars among our detected carbon-enhanced stars, especially CEMP candidates. Halo stars in this diagram are located above the red circular line, satisfying the velocity condition $|\mathbf{v} - \mathbf{v}_{\text{LSR}}| > 210 \text{ km s}^{-1}$ [the threshold taken from 239]. CEMP candidates are marked with star symbols.

Chapter 5

Peculiar emission stars

This chapter has been adapted from the published paper titled [] that was prepared and written by the author of this Doctoral thesis. The used computer code was published on GitHub platform¹ and results of the analysis as a catalogue on VizieR service².

¹

²

Chapter 6

Peculiar solar-like multiple stars

This chapter has been partially adapted from the published paper titled *The GALAH survey: unresolved triple Sun-like stars discovered by the Gaia mission* [64] that was prepared and written by the author of this Doctoral thesis. The used computer code was published on GitHub platform¹ and results of the analysis as a catalogue on VizieR service².

6.1 Introduction

The Sun, being the closest and most analyzed star is widely used as a reference for the calibration of many fundamental stellar astronomical parameters [? ?]. This implies a desire to find and catalogues as many stars similar to Sun as possible. They can be used for intra-calibration of observations done with the same setup or inter-calibration between multiple surveys that share observations of the same objects. Stars resembling the Sun in all physical parameters such as luminosity, mass, radius, rotation period, and chemical composition should also have an identical spectrum like our Sun. The definition of such stars, also known as solar twins, was first introduced by [?], but it is still far from being finalized as it follows evolution of astronomical instrumentation, observational techniques [123], and quality of determined stellar parameters and chemical abundances.

Before the release of accurate parallaxes, nowadays measured by the *Gaia* satellite, one of the approaches to determine distances of twin stars or its hosting stellar cluster was based on the assumption that solar twins should have the same intrinsic luminosity as the Sun. Therefore its apparent brightness on the sky is directly related to its distance [? ? ?].

Another widely debated topic is the chemical composition of the Sun and its evolution in time. [? ?] addressed the question if the composition of the Sun is peculiar when compared to other solar twins. Works like these benefit from using spectra with high signal to noise ratio and very high spectral resolution that are especially needed when trying to disentangle if the observed solar chemical composition is a consequence of the planetary system formation [?] or intrinsic galactic chemical composition during formation of the Sun [? ? ?]. However, even the best solar twin candidates (currently tough to be the star 18 Sco [?]), when observed with the highest possible precision, still show minute differences in the abundance

¹<https://github.com/kcotar/GALAH-survey-Solar-like-triple-stars>

²<http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=J/MNRAS/487/2474>

pattern which will make it even harder to disentangle the place where the Sun and its siblings were formed.

6.2 Selection of the best solar-like spectra

Our search for spectra that most resemble spectrum of the Sun uses data acquired by the GALAH survey and few other surveys that used the same telescope and spectrograph. The search is based on the direct comparison between a random acquired stellar spectrum and the reference solar spectrum, where both were observed with the same setup of the HERMES spectrograph.

6.2.1 Reference Solar spectrum

Before finding the best solar-like spectra in our data set, the reference spectrum had to be constructed. For its construction, we identified 3708 twilight flats with the SNR per resolution element in the green HERMES arm greater than 210. Selected spectra were first normalised using the 7th degree Chebyshev polynomial with an asymmetric sigma clipping (2σ low and 3σ high threshold) that was run for 11 steps. Normalisation procedure was tailored explicitly for solar-like objects as it excluded wavelength ranges with wide absorption features and regions with no near continuum levels that were determined from the high-resolution spectrum of the Sun created by [?]. Especially problematic for the normalisation process is the blue arm that is packed with absorption lines, topped by the H β line. Because of high SNR and presence of additional "wiggles" in observed twilight flats, we used a higher degree of polynomial function than in the case of a stellar spectrum normalisation (see Section 6.2.2). Source of identified additional spectra features is not known and might be contributed by a reflection in optics when observing bright targets or by a residual background flux that was not properly removed by the GALAH reduction pipeline [95].

The radial velocity of normalised twilight spectra was determined by correlating an observed spectrum with the high-resolution solar spectrum [?]. The peak of the resulting correlation function was determined by fitting the Gaussian function to it. Before correlation, the high-resolution solar spectrum was convolved with a Gaussian kernel to match its resolving power with the HERMES spectrograph and to visually look as close as observed spectra as possible.

Normalised twilight spectra, shifted to the rest-frame, were then median stacked to generate the final reference solar spectrum. Before it could be used, it was analyzed for possible residual flux offsets that are impossible to remove by the reduction pipeline as we can not acquire background spectrum at the same time as a twilight spectrum. Any residual flux would render too shallow absorption lines in the median reference Solar spectrum. To equalize strengths of absorption lines in our reference Solar spectrum and the high-resolution one, a small linearly changing offset was subtracted from the normalised spectrum that was re-normalised after subtraction.

6.2.2 Stellar spectra preprocessing

Preprocessing of observed stellar spectra began with combined and fluxed spectra that were prepared by the pipeline as an intermediate reduction result. Firstly

Element	Lines	Element	Lines
Al	4	Na	3
Ba	2	Nd*	5
C*	1	Ni	7
Ca	5	O	3
Co*	3	Rb*	1
Cr	9	Ru*	1
Cu	2	Sc	10
Eu*	2	Si	4
Fe	52	Sm*	2
K*	1	Sr*	1
La*	4	Ti	20
Li*	1	V	17
Mg	2	Y	4
Mn	4	Zn	2
Mo*	2	Zr*	4

Table 6.1: Number of absorption lines for different chemical elements that were used to measure similarity between observed and reference spectrum. Elements marked with * are problematic for determination in solar twins because of their shallow or rarely noticeable absorption feature. Total number of absorption lines is 180.

they were normalised using 5th degree Chebyshev polynomial and asymmetric sigma clipping (2σ low and 3σ high threshold) that was run for 15 steps. To determine the level of a continuum as accurately as possible, we masked spectral ranges without near continuum pixels ($H\alpha$ and $H\beta$) during the normalisation process. Normalised spectra were then shifted for the barycentric and radial velocity that was determined by the GALAH pipeline and linearly resampled to the same wavelength step as used by the generated reference solar spectrum. Performing both steps (radial velocity shift and resampling) at the same time retain a quality of the spectrum and reduces the number of resampling steps compared to if we would use already shifted and normalised spectra prepared by the pipeline.

6.2.3 Candidate selection

As the number of spectra in our set is quite large and some of them have completely different spectrum than the Sun (e.g. hot stars, metal poor giants) we perform initial selection of possible Solar twins purely on the basis of stellar physical parameters determined by *The Cannon* interpolation pipeline. For this we first had to determine possible offset of our parameters towards the well known stellar parameters

Object	T_{eff} [K]	$\log g$ [dex]	[Fe/H] [dex]
Sun	5771	4.44	0.02
Flats	5605 ± 40	4.21 ± 0.06	-0.14 ± 0.03
18 Sco	5810 ± 80	4.44 ± 0.03	0.03 ± 0.01
	5750 ± 20	4.37 ± 0.1	0.05 ± 0.02
β Hyi	5873 ± 45	3.98 ± 0.02	-0.04 ± 0.01
	5784 ± 5	3.93 ± 0.01	-0.11 ± 0.01
μ Ara	5902 ± 66	4.3 ± 0.03	0.35 ± 0.01
	5657 ± 20	4.15 ± 0.01	0.28 ± 0.01

Table 6.2: Stellar physical parameters for objects used to determine possible offsets between published values and values determined by *The Cannon*. **Most of the values are flagged in 180325, 18 Sco is problematic because of wrong wavelength solutions. Therefore this table is a bit questionable.**

for the Sun and benchmark Solar like stars. This can be done by comparing parameters of well studied stars that were also observed as part of The GALAH survey. Table 6.2 gives reference values of physical parameters for the Sun alongside median parameters of selected twilight flats and their standard deviations.

Table 6.2 also shows parameters for three benchmark stars with Solar like properties that were observed and reduced as any other object in our data set. Parameter values in the first line were taken from the published set of *Gaia* FGK benchmark stars ([? ?]) and the second line gives *The Cannon* parameters for the same object. From the comparison it is obvious that twilight flats have slightly more under-determined parameters as apposed to other stars used in the comparison. *The Cannon* trend, of stellar parameters being a bit too low, is observable for all stars and parameters given in the Table 6.2. Taking this into consideration, we initially selected stars within broader range of $T_{\text{eff}} \pm 250$ K, $\log g \pm 0.4$ dex and $[\text{Fe}/\text{H}] \pm 0.3$ dex around median parameters for twilight flats. Applying parameter cuts to our set, we were left with **92284** spectra that were analyzed further.

6.2.4 Spectral similarity

Similarity between observed spectra and generated reference Solar spectrum was calculated using multiple distance metrics, where identical spectra would be described by the similarity value of 0. Figure 6.1 shows that considered metrics are strongly linearly correlated with few outliers. Some degree of the correlation was expected as all of the shown metrics are based on differently weighted per pixel difference between two spectra. Few other similarity indices, like number of the flux values bellow/above the reference spectrum, were also computed but are not shown in the plot. For all the following comparisons we choose to use the Canberra similarity metric that is less sensitive to large outlying flux values caused by the cosmic particle hits than the Euclidean distance. Similarity value determined by the given function heavily depends on the SNR of the evaluated spectrum which effectively

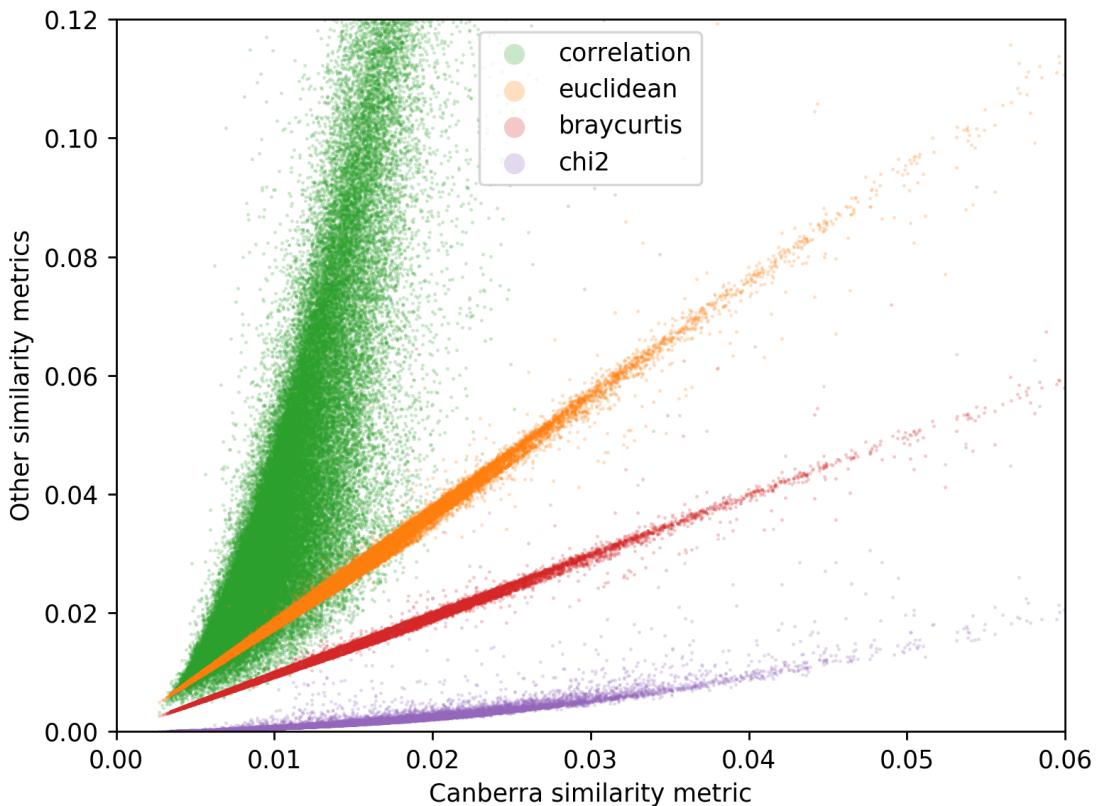


Figure 6.1: Correlation between different similarity metrics for the objects considered in the rough spectral comparison step.

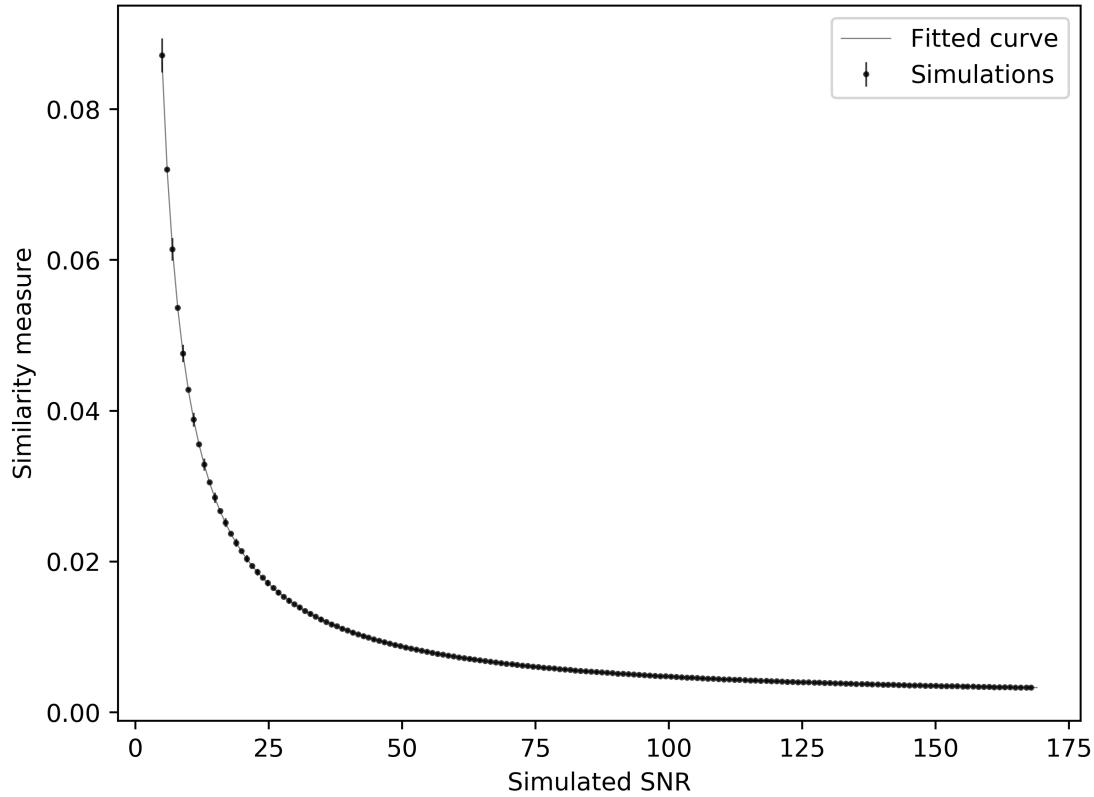


Figure 6.2: Function fitted to the data points describing dependence between spectra SNR and similarity of noisy Solar spectrum in comparison to the original noise free Solar spectrum.

decreases its spectral similarity at lower SNR values. To analyze this dependency, the same similarity metric was computed between original reference Solar spectrum and noisy reference that was generated by adding noise with Gaussian distribution to the original spectrum to represent spectra with different SNR values. To determine distribution of similarities, test was repeated 1500 times for every noise level. Smooth dependence between SNR and similarity metric can be described by the following composite of linear and power law function

$$sim(x) = A \left(\frac{x - x_{off}}{x_0} \right)^{-\alpha} + B(x - x_{off}) + C \quad (6.1)$$

that was fitted to the resulting points shown in the Figure 6.2. Parameters A , B , C , x_0 , x_{off} and α in Equation 6.1 are free and were fitted to the simulated similarity data that depend SNR value, here denoted by x .

Another test shown in the Figure ?? and 6.3 was performed to show how does this SNR relation change if the reference spectrum with simulated noise also has slightly different depths of the absorption lines. Absorption lines deviating from the normalized flux of 1 had its flux levels increased in the steps of 5 % from 0 to 20 % of the original normalized flux and then compared to the original reference spectrum in order to assess how slightly different chemical composition affect similarity measure.

For the described similarity computations we used the following wavelength ranges of the HERMES spectra: 4725 - 4895 Å in the blue arm, 5665 - 5865 Å

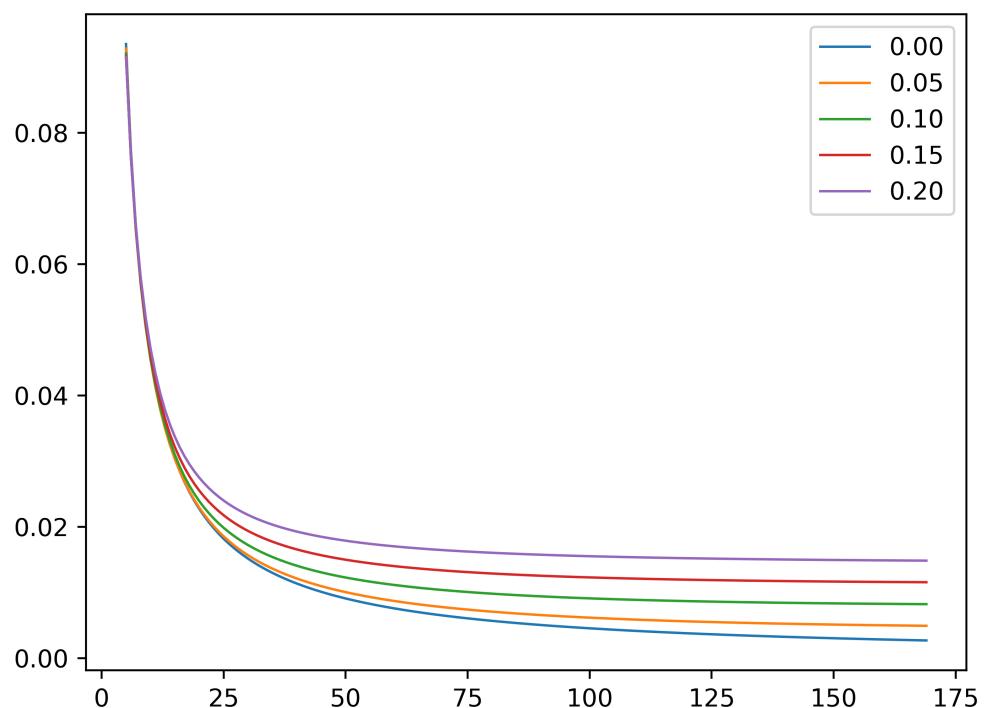


Figure 6.3: Multiple fitted functions representing relation between SNR and spectral similarity of noisy spectrum for different levels of flux added to the absorption lines.

in the green arm, 6485 - 6725 Å in the red arm and 7700 - 7875 Å in the near-infrared spectral arm. Selected ranges are valid for all observed spectra and do not pose any limitations on the number of usable spectra that might have undefined flux values at the range borders because of large radial velocities or construction specifics of the HERMES spectrograph. Similarity between spectra is not computed for the whole specified spectral range, but is limited to known and isolated, not blended spectral absorption lines, representing multiple different chemical elements. Their statistics is given in the Table ?? and is the same as used by *The Cannon* abundance determination procedure described in [99]. Around the central wavelength of each absorption line, pixels in the mean range of 0.47 Å are considered for the comparison.

6.2.5 Search for Solar twin spectra

To evaluate the similarity between selected subsets of the observed spectra and our created reference Solar spectrum we computed the same distance metric as described before for all comparisons. As initial number of candidate objects determined by the physical parameters was large, further comparison and selection was performed in two steps that greatly wary in the processing complexity.

Rough spectral comparison

In the first step, a fast and robust comparison method based on the similarity values was used to select only the most probable spectra and discard all other that do not match our reference. As the SNR value can greatly wary between spectral arms, especially true for the blue arm, similarity was independently calculated and modeled for every arm. Figure 6.4 shows distribution of similarities determined for candidate objects as a function of their SNR values. Overplotted are simulated relations with different percentages of flux offsets. Objects plotted bellow the lowest curve exhibit similarity better than the theoretical which most likely indicates underestimated SNR value. The selection of probable Solar twin objects in the first step consists of object with similarity measurement better than the theoretical curve representing 10 % flux offset, which must be true for all analyzed arms in order for the object to be considered Solar twin candidate. Consider every spectral arm individually, we effectively removed objects with the reduction problems in any of them. Applying the similarity cut on maximum difference in the flux levels we select **371** of the most probable Solar twin candidates that were evaluated more thoroughly in the next step.

Comparison with spectral noise modeling

The second comparison step involves modeling of the spectral noise through Gaussian process analysis, implemented in a software packed developed by [?]. Unlike assumed uncorrelated noised used for the generation of the theoretical similarity dependency curves in the first step, real observed spectral noise is correlated between consecutive wavelength pixels. The correlation can be described by a functions that depend only on the wavelength separation between observed pixel and its neighbors or number of steps between them in the case of a constant wavelength step between consecutive recorded fluxes.

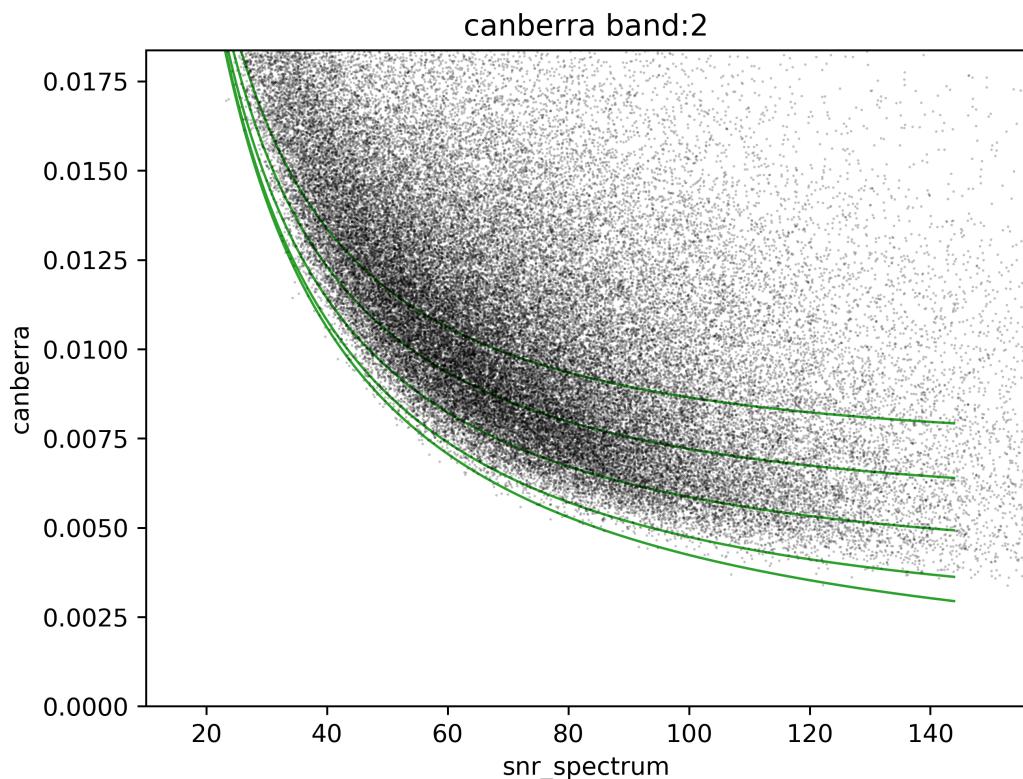


Figure 6.4: Distribution of the similarities between observed and reference spectrum in one of the HERMES spectrograph arms. Green curves represent theoretical relations of similarity for spectra with slightly different metallicities. Dots situated near the lowest green curve would suggest that the spectrum is identical to the Solar at given spectral noise level.

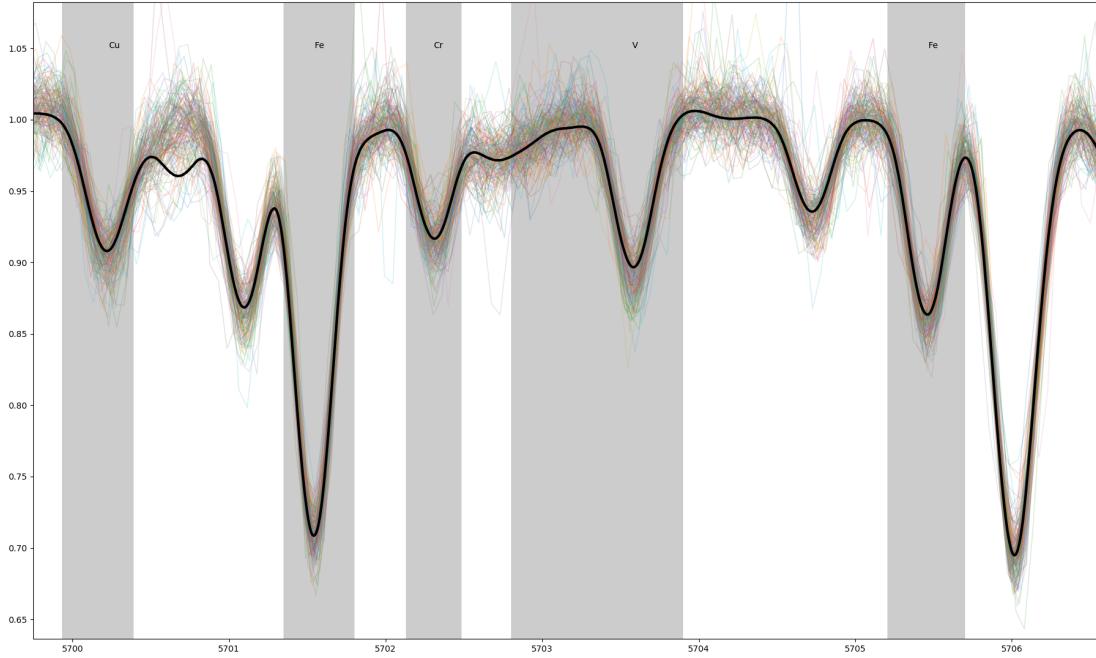


Figure 6.5: Visual comparison between a set of probable Solar twins spectra and the Solar reference spectrum in black. Grayed out regions represent wavelength ranges used in the similarity computation. Chemical element responsible for the observed line is also indicated above the spectra for every absorption line.

To model the observed spectral noise and possible minute dissimilarities in the shape of a normalized continuum levels of the compared spectra, we used two radial kernel functions. Matern 5/2 covariance function defined as

$$k_{52}(r^2) = (1 + \sqrt{5r^2} + \frac{5r^2}{3})\exp(-\sqrt{5r^2}) \quad (6.2)$$

representing fast and short-range pixel to pixel noise variations and Exponential-squared covariance function defined as

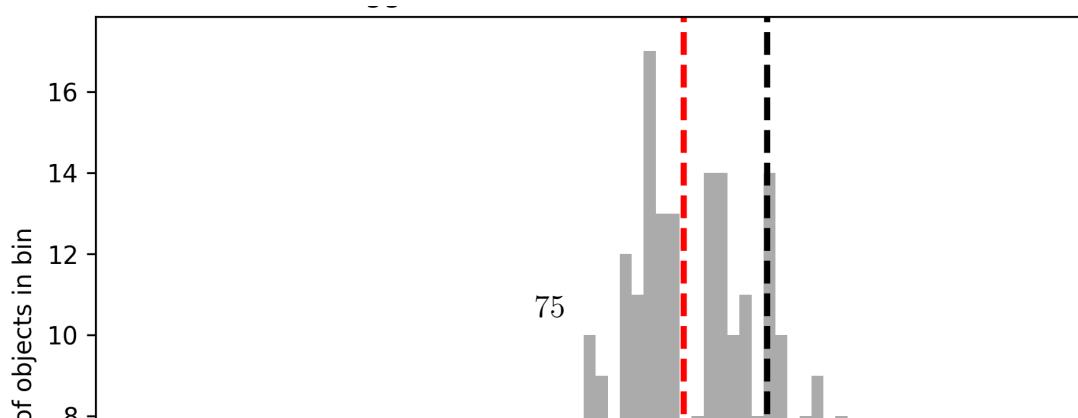
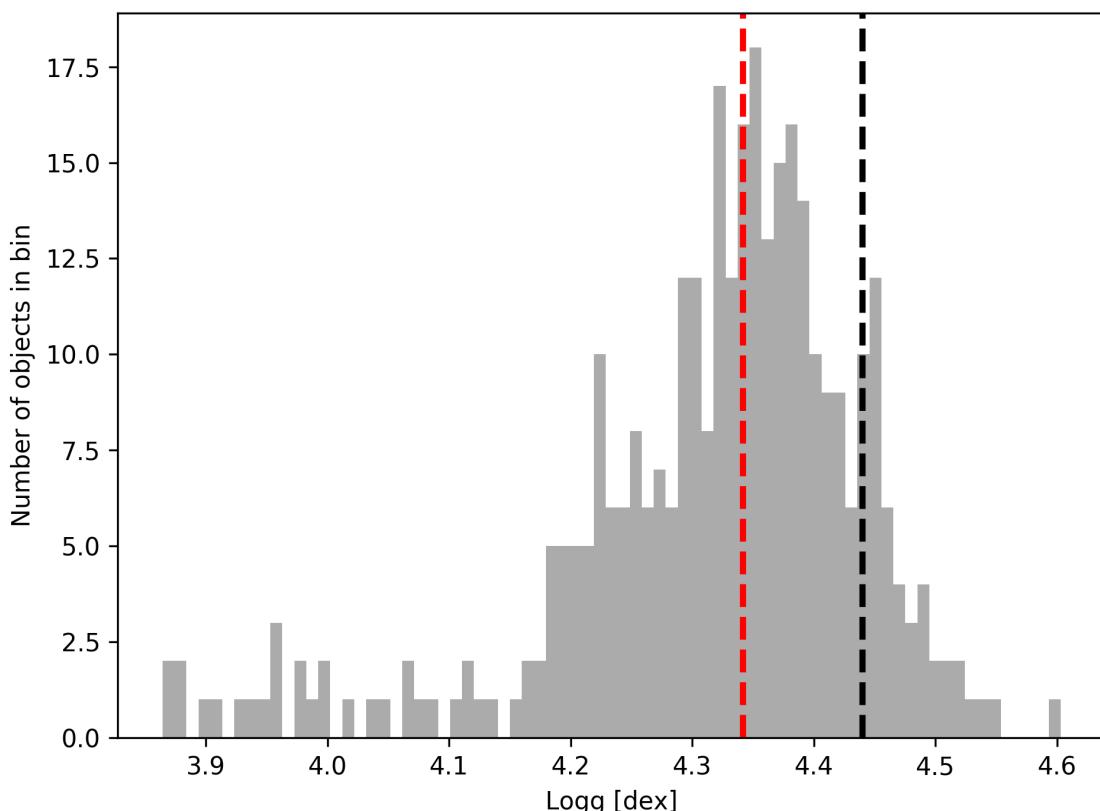
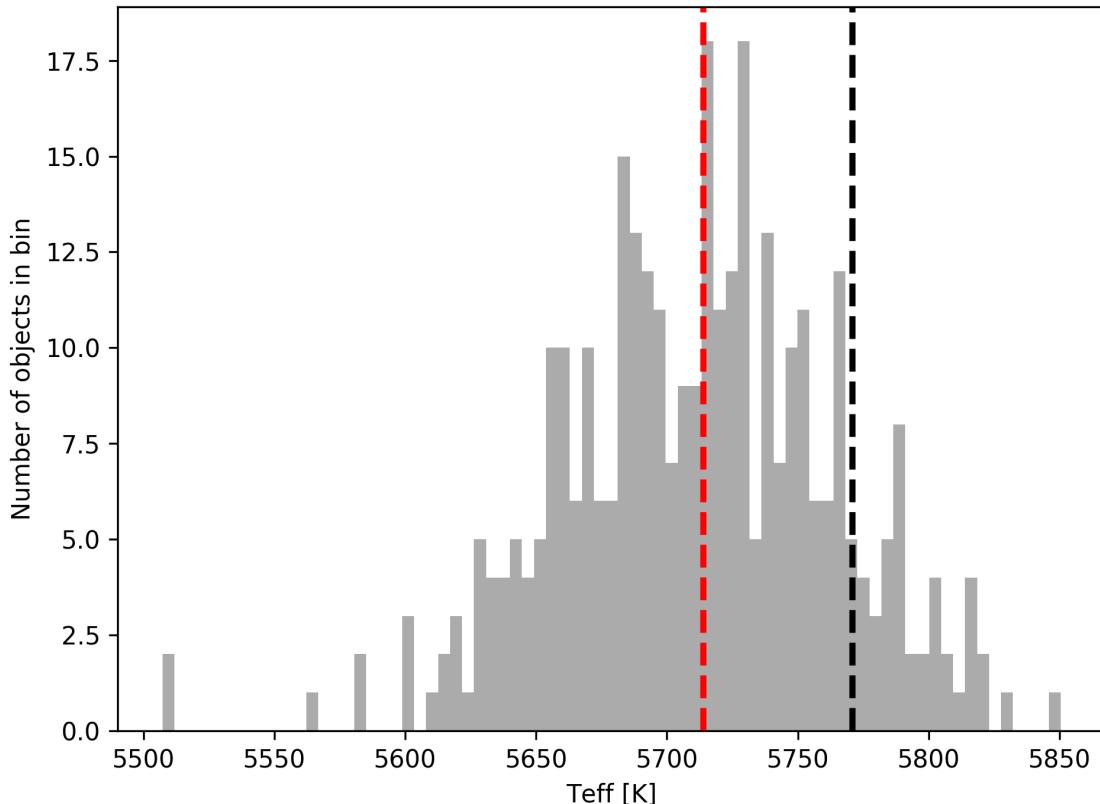
$$k_{exp}(r^2) = \exp(-\frac{r^2}{2}) \quad (6.3)$$

representing slowly changing differences in continuum levels. Both kernels were linearly added together to a relation used to describe the spectral noise and normalization discrepancies:

$$k(A_1, r_1, A_2, r_2) = A_1 k_{52}(r_1) + A_2 k_{exp}(r_2). \quad (6.4)$$

Free parameters A_1 and A_2 represent amplitudes of both kernels, r_1 and r_2 are kernel characteristic widths determining width of individual kernel and its impact on a neighboring pixels, effectively determining a degree of correlation between noise at neighboring pixels. Higher correlation between neighboring flux measurements is reflected in kernels with wider characteristics widths.

After combining the kernels, a Gaussian process is used to draw functions with correlated values (denoted as GP_{noise} from here on) complying with the specified



properties of the amplitude and kernel width. The generated function is of the same pixel length as investigated spectrum.

During precise investigation of the results in the first comparison step we noticed a minute, less than 1%, differences in continuum levels (observed at the wavelengths without any significant absorption features) between the normalized observed spectrum and reference Solar spectrum. The relation is shown in Figure ???. As it is clearly correlated with the degree of noise in the spectra, it seems be caused by the asymmetric sigma clipping employed during the normalization process. To correct for this effect, observed normalized spectrum was divided by an additional free normalization parameter A_{norm} that was fitted alongside other free parameters. Normalization correction applied to the spectrum is minute and determined A_{norm} deviates from 1.0 only for a few thousandths.

Employed fitting procedure tries to minimize the flux difference $\Delta flux$ between both spectra that are used in the comparison procedure and is defined as:

$$\Delta flux = (flux_{observed}/A_{norm}) - (flux_{Solar} + GP_{noise}). \quad (6.5)$$

Variable $flux_{observed}$ represents one of the ranges of the HERMES spectra and $flux_{Solar}$ its spectral equivalent observed for the Sun. A posterior distribution of such a fitting procedure is shown in the Figure 6.7. As the conditions differ from spectrum to spectrum, the free parameters defined in equations 6.4 and 6.5 are be defined dependently for every spectrum and its arm.

Free parameters were fitted with the implementation of a Bayesian MCMC scheme maintained by [287]. Initial parameter for the amplitude of the spectral noise is set to a spectral variance computed from a difference between observed and reference spectrum. As it was found out that all other parameters do not vary greatly between spectra, they are set to a predetermined fixed value that is arm depended. Initial value of the normalization parameter is set to 1 and is let to vary for less than 2 % in order to prevent over-fitting. Fit is initiated with 30 walkers, whose initial parameters are perturbed in a ball around the original values, so that every walker starts with different initial condition. Distribution of initial parameters in a ball is uniform in order to achieve good coverage of the parameter space that is explored by the MCMC walkers. Every walker is run for 125 steps to get enough samples for construction of posterior distribution shown in the Figure 6.7. The final parameter values are selected as a median value of the posterior distribution. The whole fitting procedure takes about an hour per spectrum on a Xeon cluster node with 48 cores.

After parameter fitting was done, we used determined functions to draw new noisy spectra from the predictive conditional distribution defined by the selected kernels of the Gaussian processes. Observed normalized spectrum and its per pixel uncertainties are used to condition the newly drawn spectra. We draw 1000 different spectra with the same noise and spectral properties as the observed spectrum. From those 1000 samples we computed a median spectrum that should be a close equivalent to the observed spectrum under ideal observing conditions without any noise present in the spectrum. Spectral similarity towards the reference Solar spectrum is computed for the medium spectrum produced by the Gaussian processes. Here we focus on two types of spectral similarities. The first one is combined similarity between all chemical elements and spectral arms. The second is computed on the element by element basis, enabling us to check for deviations in specific chemical

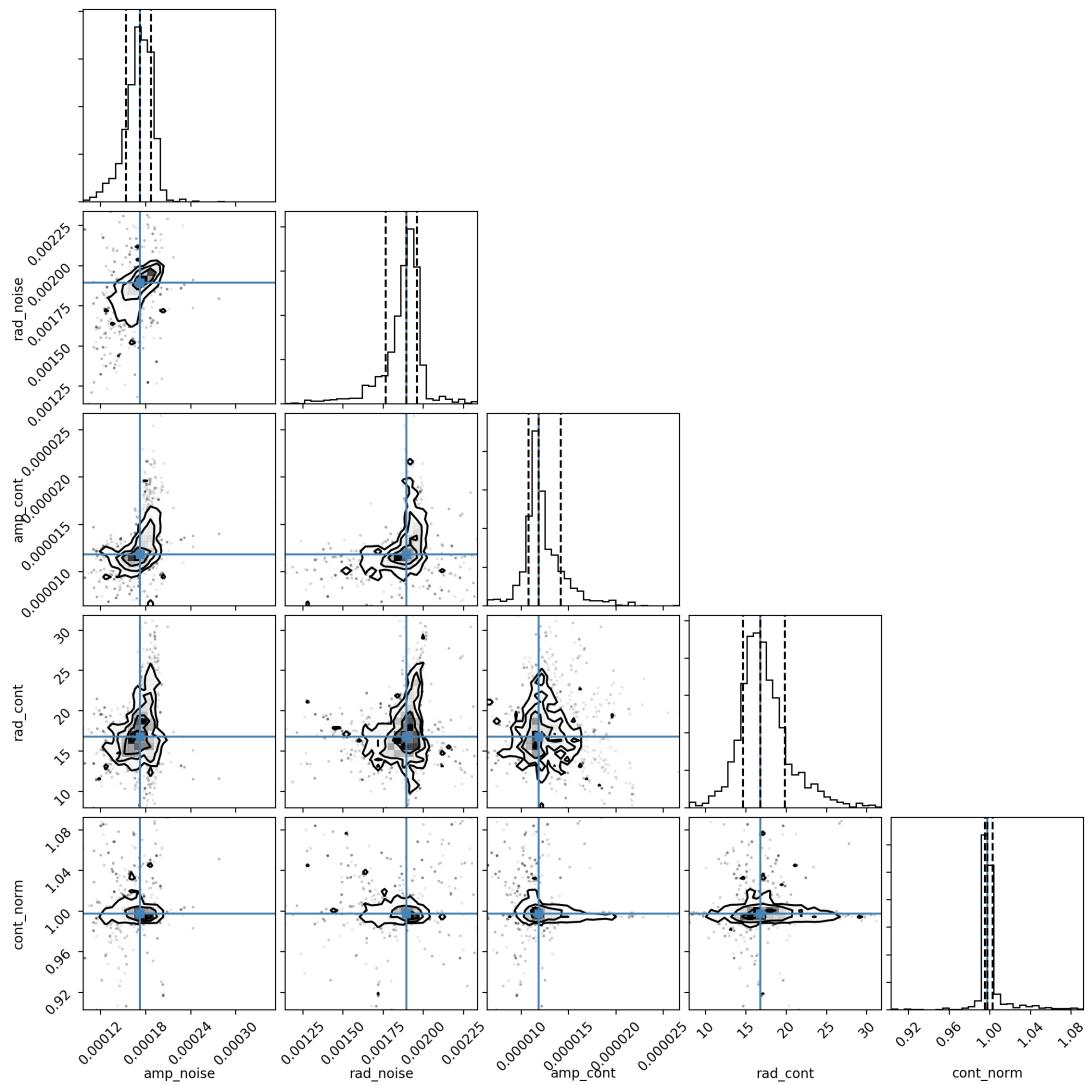


Figure 6.7: Distribution of posteriors considered by the MCMC walkers during the fitting procedure where we are determining free parameters of the model describing the observed spectral noise.

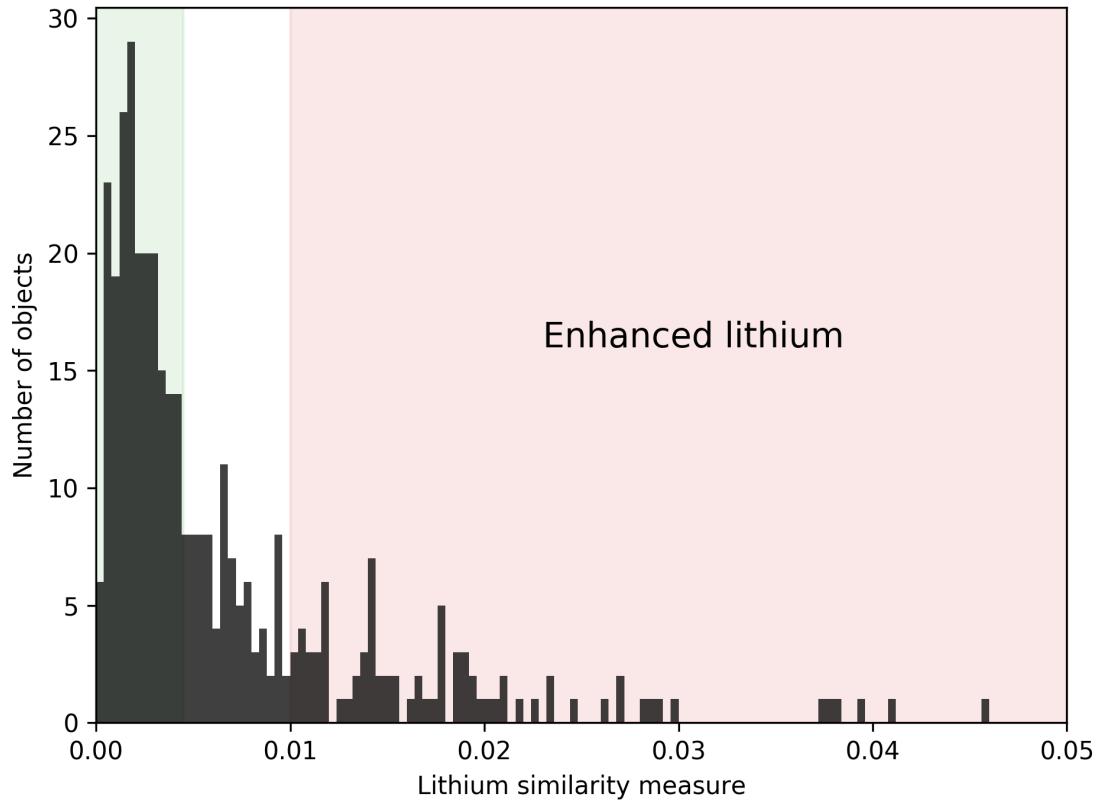


Figure 6.8: Distribution of lithium similarities as determined in the final step of the spectral comparison. Green shaded area represent object with almost exact lithium absorption as observed in Solar spectrum. In the red shaded area we find objects with strong lithium enhancement in their spectrum.

element that behaves differently as observed in the Solar spectrum. A condensed spectrum, where only considered absorption lines and its difference towards reference spectrum are shown, is presented in the Figure 6.9.

Final selection

To select the best Solar twins between our observations we considered spectral similarity that was computed for all evaluated chemical elements combined and for every individual element. This distinction is needed as we are evaluating elements with quite different number of absorption lines (see Table ?? for number of lines) which causes that the total similarity is driven by the absorption lines of Fe, Ti and V, blurring the traces of other elements with only few evaluated absorption lines.

In order to select stars with the age similar to the Sun, special attention was given to the lithium similarity that shows the greatest deviation. From different surveys and comprehensive study performed by [?] it has been demonstrated that the lithium abundance can show deviations in the order of 2 dex for Solar type stars. Following this we classified all Solar twin candidates into 3 distinct groups depending on their deviation from the Solar spectrum at the position of analyzed lithium 6707 Å line. The classes are show in the similarity distribution presented in the Figure 6.8, where we defined similarity region that does not show deviation

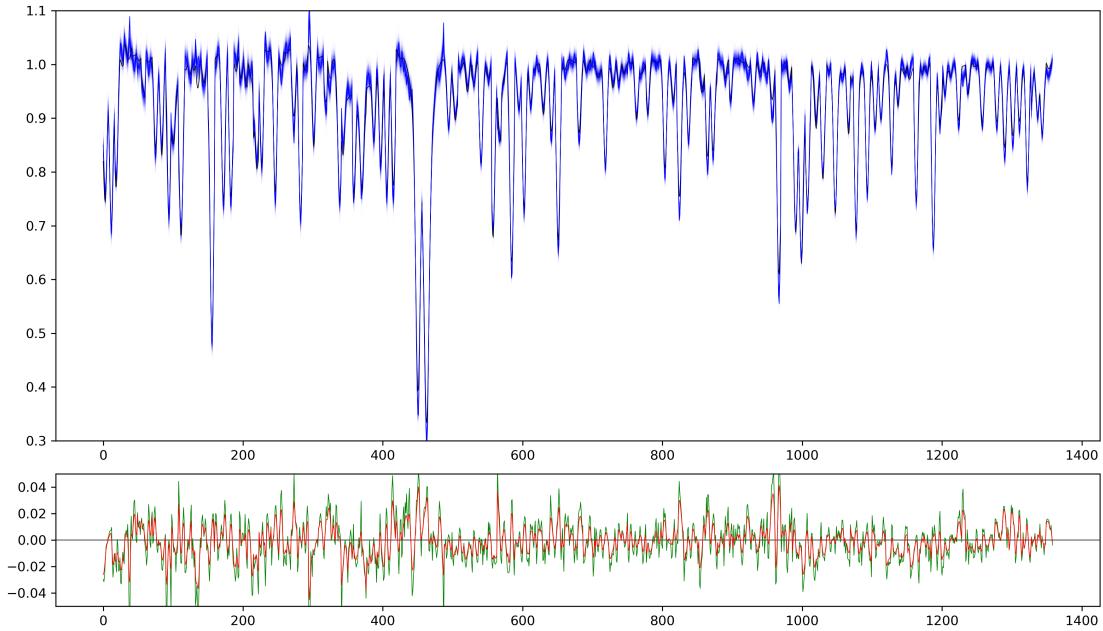


Figure 6.9: A plot in the upper panel shows condensed HERMES spectrum of spectral lines considered during spectral comparison. Blue curves represent few of the spectra that were drawn by the Gaussian processes. Under them a Solar spectrum is presented with black curve. Bottom panel indicates difference between considered spectra. Green curve represents flux difference between reference Solar spectrum and observed spectrum. The same difference, but towards the median of all spectra drawn by the Gaussian processes is given in by the red curve.

from the Solar spectrum and region with strong lithium enhancement. In between them their is a mixed region where slight lithium enhancement can be present in the spectra. After determining the similarity thresholds, spectra in every region were visually inspected to confirm the correctness of the defined similarity regions.

Out of the 371 Solar twin candidates we selected the best matching candidates based on the following criteria that are also visualized in the Figure ???. For a star to be selected as the best matching candidate, its overall similarity towards Solar spectrum must be among the 25 % best matching candidates and have at least 10 best matching chemical elements (same 25 % rule applies). For the elemental comparison we considered chemical elements with profound chemical enrichment in the Solar spectrum that are not marked by * in the Table ???. This selection also excludes lithium similarity that stands as independent check. By applying given selection criteria we are left with subset of 15 stars with most Solar like spectra. By additional consideration of lithium criteria presented in Figure 6.8 the selection narrows down to 8 stars with comparable lithium depletion as seen in the Sun.

6.3 Comparison with the literature

6.3.1 18 Sco

The best known Solar twin to date and one of the *Gaia* benchmark stars was also observed as part of the GALAH survey. It was observed at four different epoch, giv-

ing us excellent possibility to evaluate our pipeline for the detection of solar twins in the survey. All four observations have identified problems with the wavelength solution in one of the spectral arms. As this is the indirect prerequisite for the successful detection, all of them are eliminated in the first comparison step as unlikely candidates for Solar twin. **Skoda da je tako, kaj bi s tem? Po seznamih sem nasel da imamo 3 ali 4 opazovanja 18 Sco.**

6.3.2 ELODIE archive

[?] 18 Sco je edini skupni objekt. Cisto prevec svetle tarce za GALAH.

6.3.3 Ramirez study

[?] 18 Sco je edini skupni objekt. Cisto prevec svetle tarce za GALAH.

6.3.4 Datson study

[?] 18 Sco je edini skupni objekt. Cisto prevec svetle tarce za GALAH.

6.3.5 Santos study

[?] Cisto prevec svetle tarce za GALAH.

6.3.6 M67 solar twins

[?] imamo 2 ujemajoca objekta iz main run in nekaj vec (tudi ponovitve) iz commissioninga ko se je delalo kopice. Prva 2 ne zgledata nic podobna soncevemu spektru. Tudi ostali so vecinoma zelo dalec od njega, Li je vsepovsod. Primer slika 6.10.

6.4 Ponovljena opazovanja

Med najboljsimi zbranimi sta bili dve zvezdi opazovani 2-krat. Obe sta bili za 2. korak analize zbrani le v enem primru in nobena nima velikih variacij v rv.

Za prvo zgleda da je bila ponovitev vrzena zaradi redukcije. Polovica crt v rdecem kanalu je vidno mnogo prevec plitkih.

Za drugi primer gre pa za ponovitev v istem dnevu, najverjeneje zaradi razmer in slabega snr, vsaj sodec po parametrih in spektru.

Misljam da to sploh nima smisel vključevat.

6.5 Follow-up observation

In order to evaluate our selection of the best candidates we acquired high resolution Echelle spectrum for one and the only star TYC 502-985-1 that was observable from the Asiago observatory located in the northern hemisphere. Acquired spectrum that spans much wider wavelength range than the HERMES spectrum is presented in the Figure ???. Final spectrum consists of three 30 min exposures that were

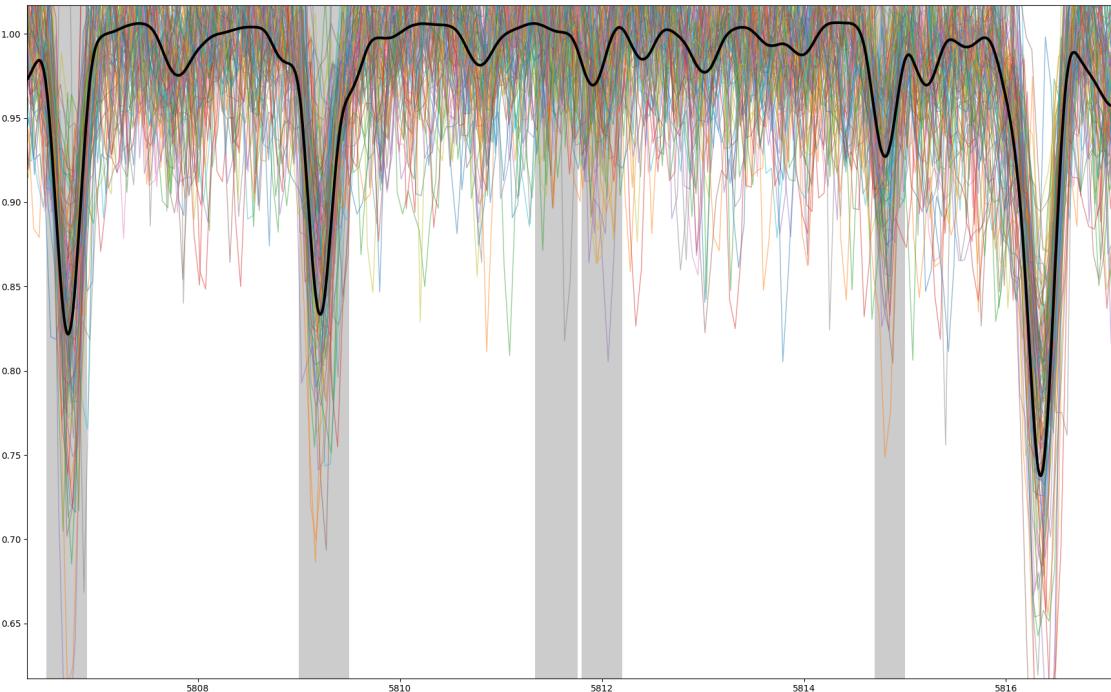


Figure 6.10: HERMES spektri za solar twins v m67. **Gre vn v koncni verziji, samo za primer.**

individually wavelength calibrated and had its background flux deducted. Individual calibrated spectra were then summed together and normalized order by order using 13th degree Chebyshev polynomial function. Normalized Echelle orders were at the end combined into one spectrum covering wavelengths from $\sim 3900 \text{ \AA}$ to $\sim 7300 \text{ \AA}$. Comparison in the Figure ?? shows remarkable spectral similarity between high resolution Solar spectrum and the spectrum of observed star. The only perceptible differences between them are caused by the telluric absorption and normalization procedure where no continuum pixels are present in wider wavelength range.

6.6 Results

Following the spectral analysis and comparison described in the previous sections we identified 371 possible Solar twin candidates, out of which we selected 15 best candidates based on the in-depth element by element similarity analysis. In this section we present properties of the stars identified during the direct spectral comparison.

6.6.1 Atmospheric parameters

Stellar parameters of the Sun and its chemical abundances have been studied thoroughly over the years and can be with certainty used as a reference in our comparison towards detected set of stars. Following their stellar parameter distribution in Figure 6.6 and abundance distribution in Figure ?? we can investigate possible offsets in parameters for Solar like objects determined by *The Cannon* interpolation procedure. Both figures present only parameter values that are deemed to be reliable - selected by using recommended flags `flag_cannon == 0` and `flag_X_abund_cannon == 0`,

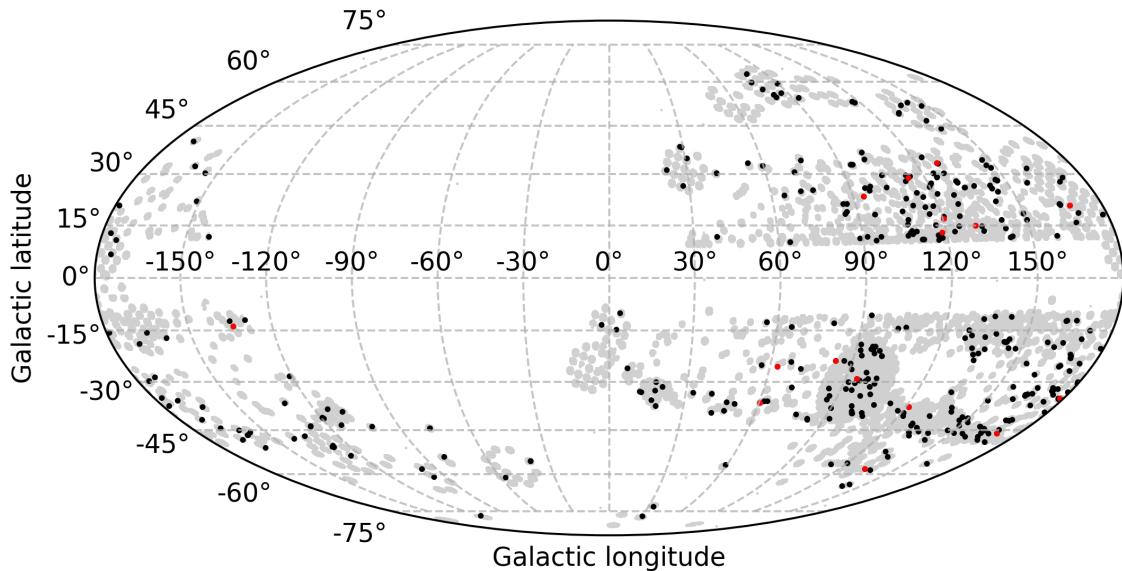


Figure 6.11: Distribution of Solar twins presented in the Galactic coordinates. Red dots indicates position of the best candidates for Solar twins among the analyzed HERMES spectra. Observed stellar fields, used in our analysis are shaded in gray.

where X represent measured chemical abundance.

Spectroscopic stellar parameters have been found to be underestimated when compared to Solar values. We show that the median of the T_{eff} distribution for possible Solar twins is too cold for 58 K, stars are too metal poor for 0.03 dex and their surface gravity $\log g$ is too weak for 0.1 dex. Similar offset were for Solar like objects derived by the works of [? ? ?].

As *The Cannon* provides abundances corrected for the Solar chemical abundance, every histogram presented in the Figure ?? should be centered around the value of 0.0 dex. We see that this is true for the majority of the elements except, for the Ba and V that deviate from that for ~ 0.1 dex. High deviations for the elements C, La, Li an Zr can easily be explained as their equivalent widths are extremely small and bellow the detection limit of the *The Cannon* algorithm. Therefore histograms show only values of highly enhanced members and the computed median of the distribution

6.6.2 Spatial distribution

As observable in the Figure 6.11 detections are evenly spread across the sky and among observed fields, without any strong grouping in any of the observable directions.

Combing them with the measurements derived from the observations performed by the *Gaia* satellite, we can analyze their position in the Galaxy. Matching Solar twin candidates with the *Gaia* DR2 sources ([42]) gives us 365 positive matches. One of the sources was subsequently removed because of its large parallax error and small parallax value, positioning it much further away than any other star. From the remaining data we can deduce that all of our candidates are located within the radius of 900 pc away from the Sun and most probably belong to the Galactic thin disk population. With few exceptions they are all located within 400 pc away from

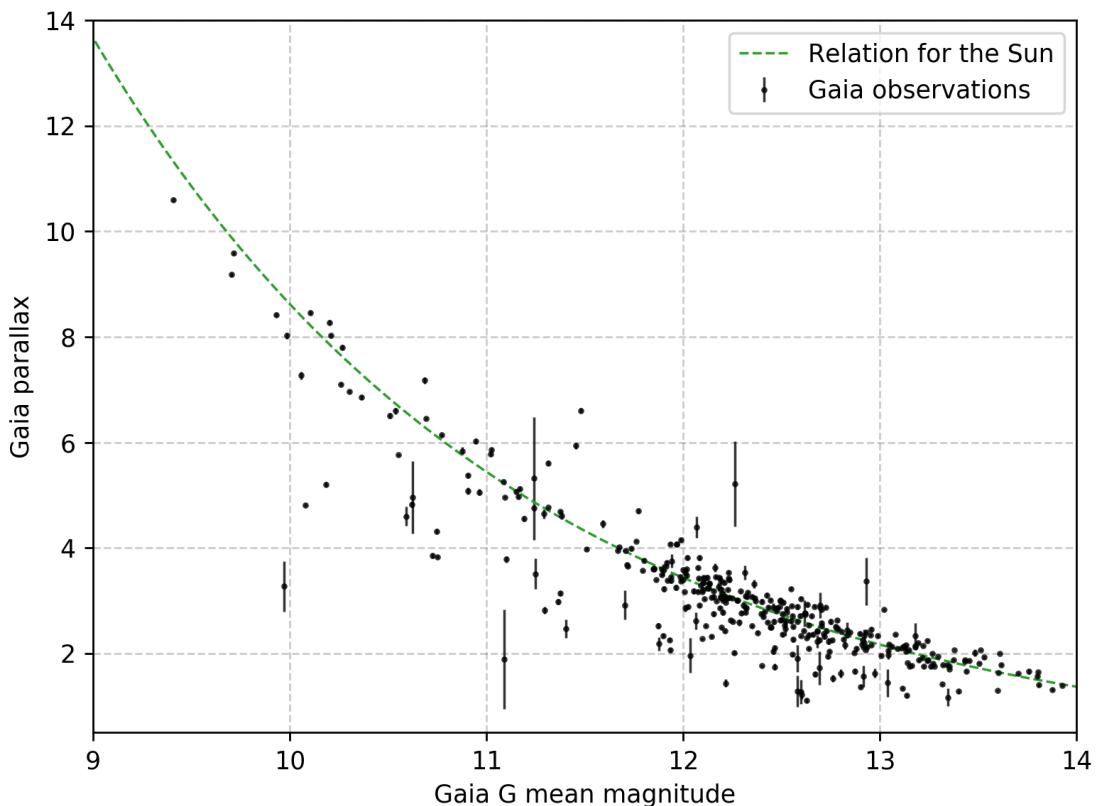


Figure 6.12: Following the statement that our twin candidates all have very similar absolute magnitude, we produced this plot showing connection between determined *Gaia* parallax and apparent G magnitude. Theoretical relation for the Sun with absolute magnitude $G = 4.68$ (computed from relations given by [279]) is shown with the green curve.

the Galactic plane.

6.7 Abstract

The latest *Gaia* data release enables us to accurately identify stars that are more luminous than would be expected on the basis of their spectral type and distance. During an investigation of the 329 best Solar twin candidates uncovered among the spectra acquired by the GALAH survey, we identified 64 such over-luminous stars. In order to investigate their exact composition, we developed a data-driven methodology that can generate a synthetic photometric signature and spectrum of a single star. By combining multiple such synthetic stars into an unresolved binary or triple system and comparing the results to the actual photometric and spectroscopic observations, we uncovered 6 definitive triple stellar system candidates and an additional 14 potential candidates whose combined spectrum mimics the Solar spectrum. Considering the volume correction factor for a magnitude limited survey, the fraction of probable unresolved triple stars with long orbital periods is $\sim 2\%$. Possible orbital configurations of the candidates were investigated using the selection and observational limits. To validate the discovered multiplicity fraction, the same procedure was used to evaluate the multiplicity fraction of other stellar types.

6.8 Introduction

The investigation of Solar-type stars in the Solar neighbourhood has revealed that around half of them are found in binary or more complex stellar systems [242, 243, 244]. Of all multiple systems, about 13 % are part of higher-order hierarchical systems [242, 245]. Beyond the Solar neighbourhood, the angular separation between members of such multiple star systems becomes too small for their components to be spatially resolvable in the sky. As a result, all stellar components contribute to the light observed by spectroscopic, photometric, and astrometric surveys. It has been suggested that the population of binaries in the field could be even higher than in the Solar neighbourhood [246], therefore a combination of multiple complementary approaches must be used to detect and analyse multiple stellar systems with different properties [244].

If the orbital period of such a system is relatively short, with high orbital velocities, it can be spectroscopically identified as a multiple system in two ways. When the components are of comparable luminosities, the effect of multiple absorption lines can be observed in the system's spectrum. Such an object is also known as a double-lined binary (SB2) system [85, 247, 248, 249]. By contrast, a single-lined binary (SB1) system does not show the same effect, as the secondary component is too faint to significantly contribute to the observables. Short period SB1 systems can be identified from the periodic radial velocity variations if multi-epoch spectroscopy is available [250, 251, 252, 253, 254]. Other extrema are very wide binaries [255, 256, 257, 258, 259, 260, 261], and co-moving pairs [262, 263] that can only be identified by their spatial proximity and common velocity vector.

Duchêne and Kraus [243] summarized that the majority of Solar-type stars are part of binary systems with periods of hundreds to thousands of years, whose period

distribution reaches a maximum at $\log(P) \approx 5$, for P measured in days. Because of the wide separation and long orbital periods of the components in such a scenario, the radial velocity variations will have both low amplitude and long period. This makes them challenging to detect in large-scale spectroscopic surveys, which typically last for less than a decade, and have a low number of revisits. A spectrum of such a binary or triple still contains a spectroscopic signature of all members, and those contributions can be disentangled into individual components [86, 87]. Such a decomposition is easier when a binary consists of spectrally different stars [264, 265, 266, 267, 268], but becomes much harder or near-impossible when the composite spectrum consists of contribution from near-identical stars whose individual radial velocities are almost identical [269]. In that case, additional photometric and distance measurement have to be used to constrain possible combinations [62]. If spectroscopic data are not available, determination of multiples can be attempted purely on the basis of photometric information [62, 270, 271, 272], but such approaches are limited to certain stellar types, and yield results that might be polluted with young pre-main-sequence stars [273].

In the scope of the GALactic Archaeology with HERMES (GALAH) survey, most stars are observed at only one epoch, limiting the number of available data points that can be used to identify and characterize unresolved spectroscopic binaries. For this study, GALAH observations were complemented by multiple photometric surveys presented in Section 6.9. Both spectroscopic (Section 6.11.1) and photometric (Section 6.11.2) observations were used to create data-driven models of a single star. Those models were used to analyse Solar twin candidates detected amongst the GALAH spectra (Section 6.10). Sections 6.12 and 6.13 describe the employed fitting procedure that was used to determine multiplicity and physical parameters of the twin systems. In Section 6.14 we determine constraints on orbital periods of the detected triple candidates. Constraints are imposed using the observational limits and simulations described in Section 6.15, where a population bias introduced by the magnitude-limited survey is assessed in Section 6.15.4. Concluding Section 6.16 summarises the results and introduces additional approaches that could be used to verify the results.

6.9 Data

In this work, we analyse a set of high-resolution stellar spectra that were acquired by the High Efficiency and Resolution Multi-Element Spectrograph [HERMES, 146, 147], a multi-fibre spectrograph mounted on the 3.9-metre Anglo-Australian Telescope (AAT). The spectrograph has a resolving power of $R \sim 28,000$ and covers four wavelength ranges ($4713 - 4903 \text{ \AA}$, $5648 - 5873 \text{ \AA}$, $6478 - 6737 \text{ \AA}$, and $7585 - 7887 \text{ \AA}$), frequently also referred to as spectral arms. Observations used in this study have been taken from multiple observing programmes that make use of HERMES spectrograph: the main GALAH survey [145], the K2-HERMES survey [148], and the TESS-HERMES survey [149]. Those observing programmes mostly observe stars at higher Galactic latitudes ($|b| > 10^\circ$), employ different selection functions, but share the same observing procedures, reduction, and analysis pipeline [internal version 5.3, 95]. All three programmes are magnitude-limited, with no colour cuts (except the K2-HERMES survey), and observations predominantly fall in the V magnitude range between 12 and 14. This leads to an unbiased sample

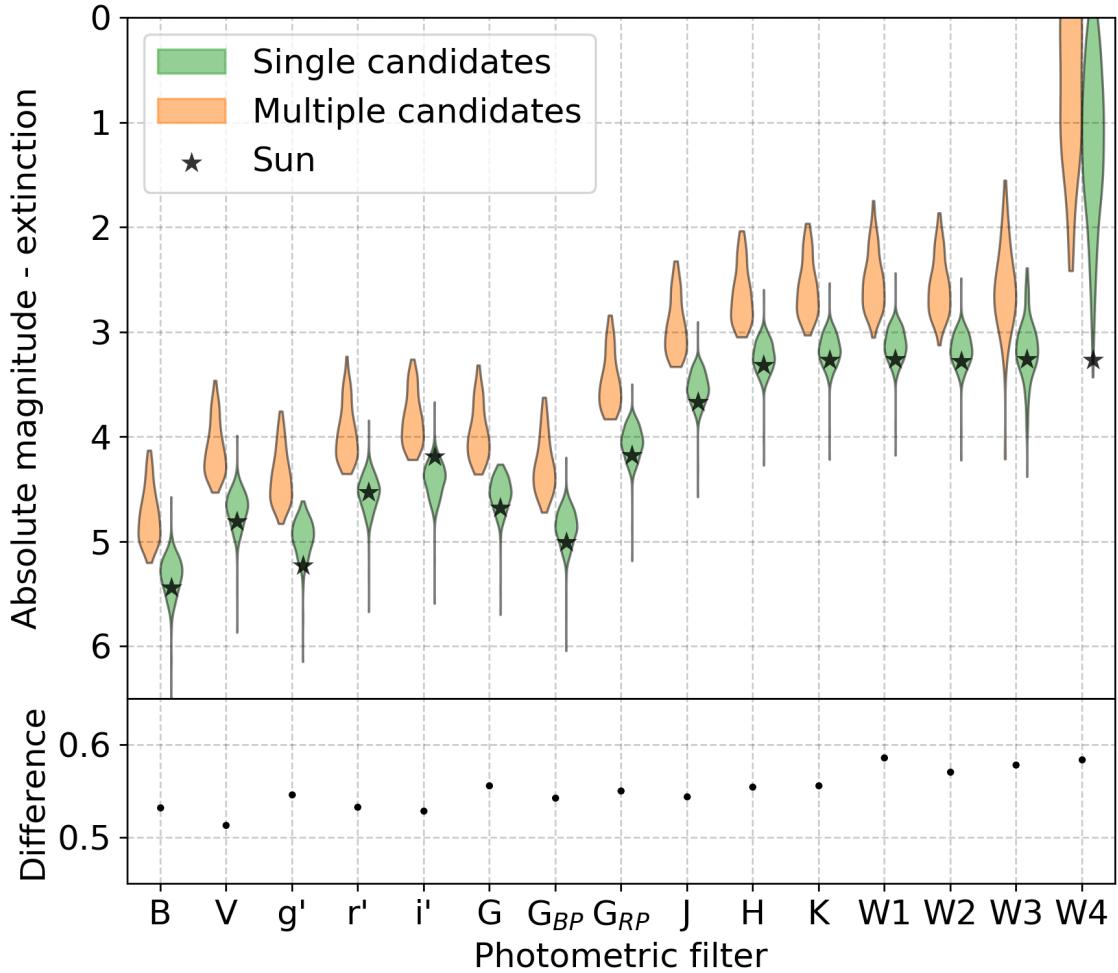


Figure 6.13: Violin density plots showing the distribution of extinction-corrected absolute magnitudes in multiple photometric systems. Separate distributions are given for stars that we considered as single and multiple in our analysis. Star symbols indicate absolute magnitude of the Sun [274]. The bottom panel shows a difference between the median magnitudes of both distributions.

of mostly southern stars that can be used for different population studies, such as multiple stellar systems in our case. Additionally, stellar atmospheric parameters and individual abundances derived from spectra acquired during different observing programmes are analysed with the same *The Cannon* procedure [internal version 182112 that uses parallax information to infer $\log g$ of a star, 98, 99], so they are inter-comparable. *The Cannon* is a data-driven interpolation approach trained on a set of stellar spectra that span the majority of the stellar parameter space [for details, see 99]. Whenever we refer to valid or unflagged stellar parameters in the text, only stars with the quality flag `flag_cannon` equal to 0 were selected.

With the latest *Gaia* DR2 release [42, 129], the determination of distance for stars within a few kpc away from the Sun becomes straightforward, and the derived distances are accurate to a few percent. Along with the measurements of parallax and proper motion, stellar magnitudes in up to three photometric bands (G , G_{BP} and G_{RP}) are provided. Coupling those two measurements together, we can determine the absolute magnitude and luminosity of a star. This can be done for the majority of stars observed in the scope of the GALAH survey as more than 99 % of them can be matched with sources in *Gaia* DR2. Although the uncertainty in *Gaia* mean radial velocities is much larger than those from GALAH [275], especially at its faint limit, they can still be used in the multi-epoch analysis (Section 6.14.3) to determine the lower boundary of its variability.

Because observations of both surveys were acquired at approximately the same epoch ranges, wide binaries with long orbital periods would show little variability in their projected velocities. Observational time-series were therefore extended into the past with the latest data release from The Radial Velocity Experiment [RAVE DR5, 90] that also surveyed the southern part of the sky. More recent spectra of accessible (stars' declination $> -25^\circ$ and G magnitude < 13.5) multiple candidates were acquired by the high-resolution Echelle spectrograph (with the resolving power $R \sim 20,000$) mounted on the 1.82-m Copernico telescope located at Cima Ekar in Asiago, Italy. As this telescope is located in a northern observatory, we were able to observe only a limited number of stars. Every observation consisted of two half-hour long exposures that were fully reduced, shifted to the heliocentric frame, combined together, and normalized. The system's radial velocity was determined by cross-correlating the observed spectrum with that of the Sun. Data from those two additional spectroscopic surveys enabled us to further constrain the variability of the analysed systems.

For a broader wavelength coverage, additional photometric data were taken from three large all-sky surveys. In the visual part of the spectrum we rely on the AAVSO Photometric All-Sky Survey [APASS, 276] B , V , g' , r' , i' bands that are supplemented by the Two Micron All-Sky Survey [2MASS, 150] J , H , K_S bands, and the Wide-field Infrared Survey Explorer [WISE, 277] $W1$, $W2$ bands. All of those surveys were matched with the GALAH targets, which resulted in up to 13 photometric observations per star. Photometric values and their uncertainties were taken as published in these catalogues, ignoring any specific quality flags. During the initial investigation of their usefulness, WISE $W3$, and $W4$ bands proved to be unreliable for our application and were therefore removed from further use. The main reason for their removal is a large scatter in magnitude measurements of similar stars and a strong overlap between single and multiple stars evident in Figure 6.13.

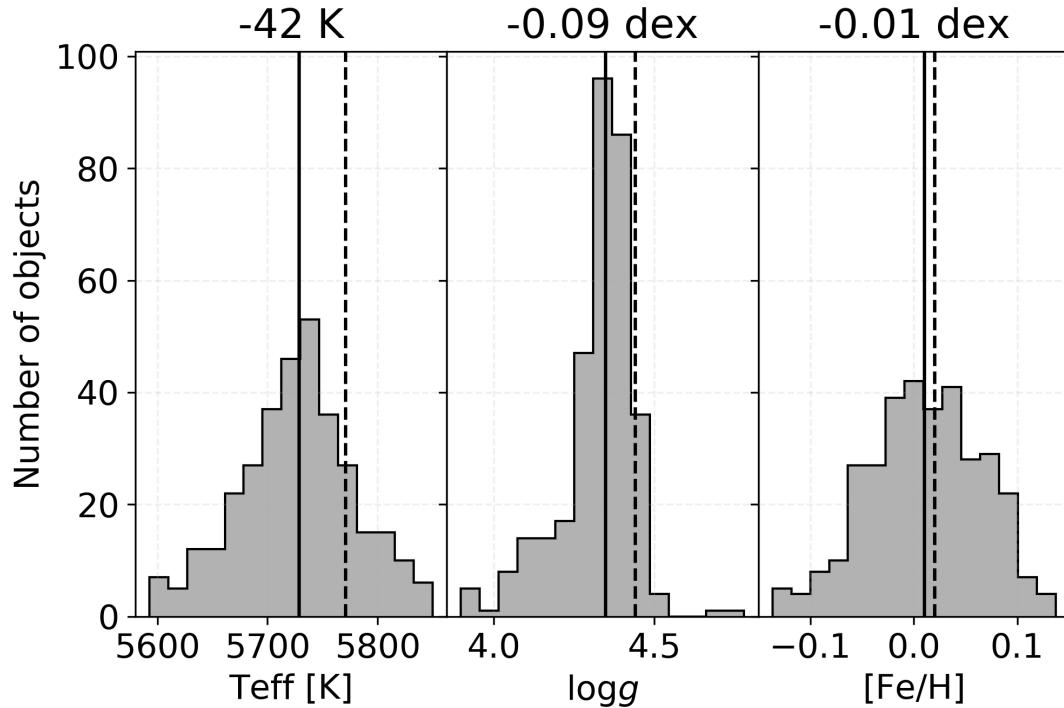


Figure 6.14: Distribution of physical parameters for discovered Solar twin candidates. Values above the plots represent difference between median of the distribution (solid vertical line) and actual Solar values (dashed vertical line).

6.10 Solar-like spectra

The selection of the most probable Solar twin candidates observed in the scope of the GALAH survey was performed purely on the basis of spectral comparison using the Canberra distance metric [278]. It is defined as

$$distance_{Canberra}(f_{\odot}, f_{obs}) = \sum_{\lambda} \frac{|f_{\odot,\lambda} - f_{obs,\lambda}|}{|f_{\odot,\lambda}| + |f_{obs,\lambda}|}, \quad (6.6)$$

where f_{\odot} is the reference Solar spectrum and f_{obs} observed spectrum. This avoided the need for prior knowledge about the physical parameters of the considered stars. Observed spectra were compared with reference twilight Solar spectrum that was acquired by the same spectrograph. The comparison was done in two steps. First, the observed spectra were compared arm by arm, where only 7 % of the best matching candidates were selected for every HERMES arm. The selection does not consist of only one hard threshold on spectral similarity, but it varies as a power law of spectral SNR. Larger discrepancies between spectra are therefore allowed for spectra with low SNR. The final selection consisted of a set of stars whose spectra were adequately similar to the Solar spectrum in every arm. Further processing consisted of modelling the spectral noise and the comparison of individual chemical abundances. For details about spectral comparison and Solar twin determination see Čotar et al. (in preparation). The parameter distribution for the selected set of objects is presented in Figure 6.14. As the computation of $\log g$ uses prior knowledge about

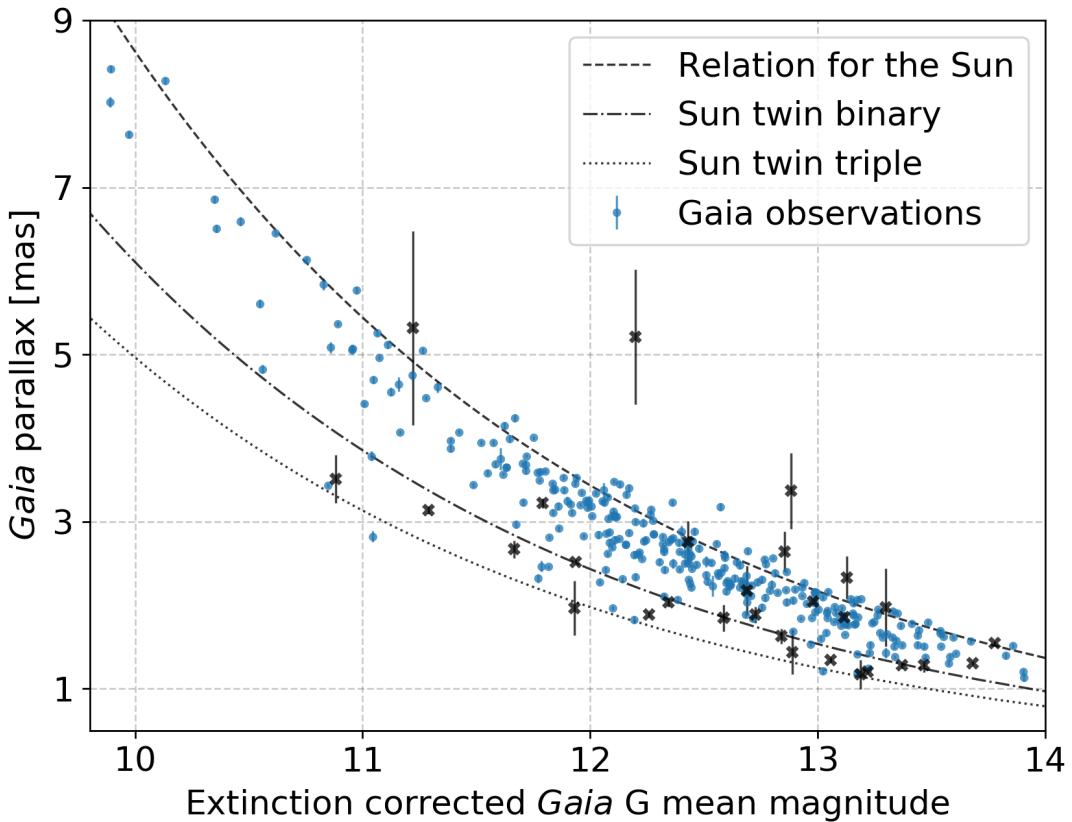


Figure 6.15: Parallax versus measured apparent *Gaia* G magnitude for our Solar twin candidates. Stars marked with black crosses have large normalised astrometric uncertainties ($\text{RUWE} > 1.4$) which may lead to wrongly determined distance and consequently multiplicity results. The dashed line represents a theoretical absolute magnitude of the Sun as it would be observed at different parallaxes. Similarly, the relation for a binary and a triple system composed of multiple Solar twins are plotted with the dash-dotted and dotted line.

the absolute magnitude of a star, lower $\log g$ values are attributed to objects that exhibit excess luminosity.

6.10.1 Candidate multiple systems

Among the determined Solar twin candidates, we noticed a photometric trend that is inconsistent with the distance of an object that resembles the Sun. Solar twins, mimicking the observed Solar spectrum, should also be similar to it in all other observables such as luminosity, effective temperature, surface gravity, chemical composition and absolute magnitude. Plotting their apparent magnitude against *Gaia* parallax measurement (Figure 6.15), all of the detected stars should lie near or on the theoretical line, describing the same relation for the Sun observed at different parallaxes. As the magnitude of the Sun is not directly measured by the *Gaia* or determined by the *Gaia* team, we computed its absolute magnitude using the relations published by Evans *et al.* [279] that connect the *Gaia* photometric system with other photometric systems. The reference Solar magnitudes (in multiple filters) that were

used in the computation were taken from Willmer [274]. The resulting absolute G magnitude of the Sun is 4.68 ± 0.02 , where the uncertainty comes from the use of multiple relations. This value also coincides with the synthetic *Gaia* photometry produced by Casagrande and VandenBerg [280], who determined magnitude of the Sun to be $M_{G,\odot} = 4.67$.

Within our sample of the probable Solar twins, we identified 64 stars that show signs of being too bright at a given parallax. In Figure 6.15 they are noticeable as a sequence of data points that lie below the theoretical line and are parallel to it. Another even more obvious indication of their excess luminosity is given by the colour-magnitude diagram in Figure 6.16. There, the same group of stars is brighter by ~ 0.7 magnitude when comparing stars with the same colour index. As both groups of stars are visually separable, the multiple stellar candidates can easily be isolated by selecting objects with extinction corrected absolute G magnitude above the binary limit line shown in Figure 6.16. To compute the absolute magnitudes, we used the distance to stars inferred by the Bayesian approach that takes into account the distribution of stars in the Galaxy [281]. As the reddening published along the *Gaia* DR2 [142] could be wrong for stars located away from the used set of isochrones, we took the information about the reddening at specific sky locations and distances from the all-sky three-dimensional dust map produced by Capitanio *et al.* [282]. To infer a band dependent extinction from the acquired reddening, a reddening coefficient (R) was used. The values of R , considering the extinction law $R_V = 3.1$, were taken from the tabulated results published in Schlafly and Finkbeiner [283].

To determine the limiting threshold between single and multiple candidates, we first fit a linear representation of the main sequence to the median of the absolute magnitudes distribution in the 0.02 mag wide colour bins. The lower limit for the binaries was placed 0.25 mag above the fitted line.

Confirmation that this extra flux could be contributed by the unseen companion also comes from other photometric systems, where the distribution of absolute magnitudes for both groups is shown in Figure 6.13. On average, multiple candidates are brighter by ~ 0.55 magnitude in every considered band. For an identical binary system, a measured magnitude excess would be 0.75, and 1.2 for a triple system. As the observed difference is not constant in every band, as would be expected for a system composed of identical stars, we expect some differences in parameters between the components of the system. This can be said under the assumption that all considered photometric measurements were performed at the same time. Of course, that is not exactly true in our case as the acquisition time between different photometric surveys can vary by a few to 10 years. As the Sun-like stars are normal, low activity stars, this effect is most probably negligible, but events like occultations between the stars in the system can still occur.

6.11 Single star models

Once the selection of interesting stars was performed, we began with the analysis of their possible multiplicity. Our procedure for the analysis of suspected multiple stellar systems is based on spectroscopic and photometric data-driven single star models that were constructed from observations taken from multiple large sky surveys. With this approach, we exclude assumptions about stellar properties and

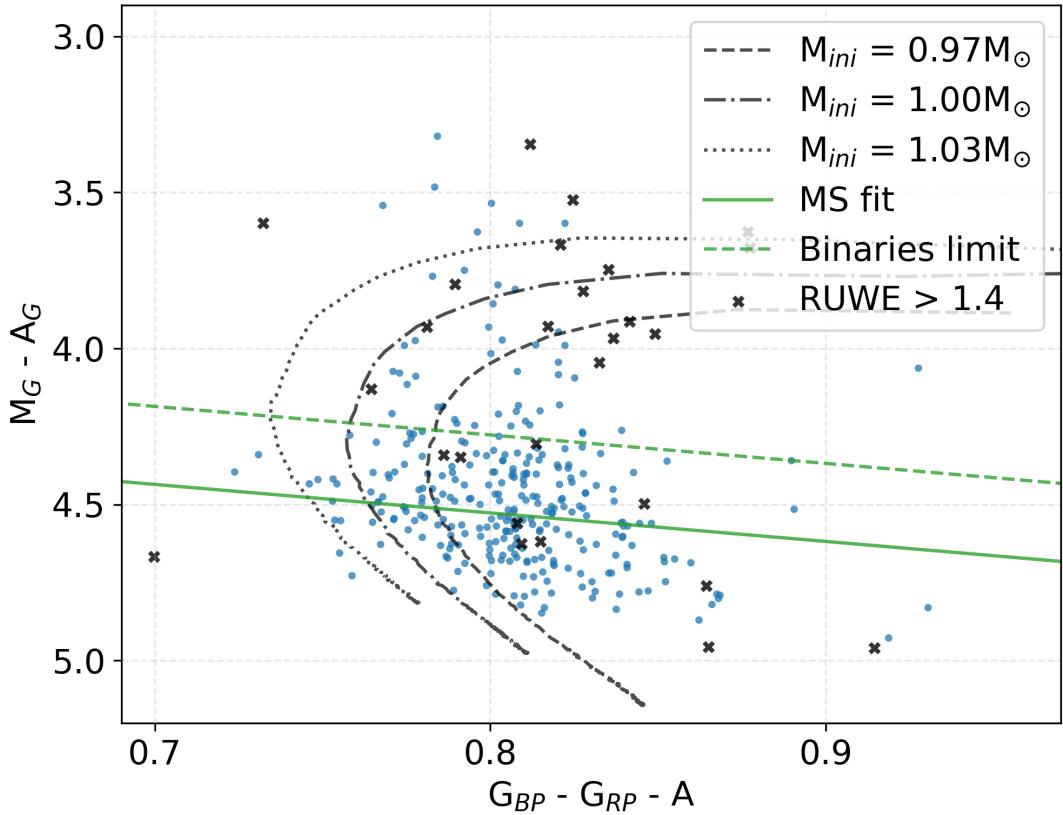


Figure 6.16: *Gaia* extinction corrected absolute G magnitude and colour index computed from G_{BP} , and G_{RP} bands. Stars marked with black crosses have large normalised astrometric uncertainty ($RUWE > 1.4$). The green dashed line represents a threshold that was used as a delimiter between objects treated as multiple and single stars. Overlaid evolutionary tracks, constructed from the PARSEC isochrones [284], represent an evolution of stars with Solar-like initial mass M_{ini} and metallicity $[M/H] = 0$ for stellar ages between 0.1 and 12 Gyr.

populations that are usually used to generate synthetic data. In this section, we describe approaches that were used to create those models.

6.11.1 Spectroscopic model

Every stellar spectrum can be largely described using four basic physical stellar parameters: T_{eff} , $\log g$, chemical composition, and $v \sin i$. To construct a model that would be able to recreate a spectrum corresponding to any conceivable combination of those parameters, we used a data-driven approach named *The Cannon*. The model was trained on a set of normalised GALAH spectra that meet the following criteria: the spectrum must not be flagged as peculiar [84], have a signal to noise ratio (SNR) per resolution element in the green arm > 20 , does not contain any monitored reduction problems and have valid *The Cannon* stellar parameters. Additionally, we limited our set to main sequence dwarf stars (below the arbitrarily defined line shown in Figure 6.17) as giants are not considered for our analysis. Additionally, the decision not to consider giants was taken as a result of the fact that accurate modelling of their spectra requires information about their luminosity. It should be noted that the application of these limits does not ensure that our training set is completely free from spectra of unresolved (or even clearly resolved SB2 binaries), as would be desired in the case of an ideal training set.

In order to train the model, all spectra were first shifted to the rest frame by the reduction pipeline [95], and then linearly interpolated onto a common wavelength grid. The training procedure consists of minimising a loss function between an internal model of *The Cannon* and observations for every pixel of a spectrum [98].

The result of this training procedure are quadratic relations that take desired stellar parameters T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and $v \sin i$ to reconstruct a target spectrum. Spectra generated in this manner are trustworthy only within the parameter space defined by the training set, where the main limitation is the effective temperature which ranges from ~ 4600 to ~ 6700 K on the main sequence. Spectra of hotter stars are easy to reproduce, but they lack elemental absorption lines that we would like to analyse. On the other hand, spectra of colder stars are packed with molecular absorption lines and therefore harder to reproduce and analyse. The model itself can be used to extrapolate spectra outside the initial training set, but as they can not be verified, they were not considered to be useful for the analysis.

6.11.2 Photometric model

With the use of a model that produces normalised spectra for every given set of stellar parameters, we lose all information about the stars' colour, luminosity, and spectral energy distribution. This can be overcome using another model that generates the photometric signature of a desired star. To create this kind of a model, we first collected up to 13 apparent magnitudes from the selected photometric surveys (*Gaia*, APASS, 2MASS, and WISE) for every star in the GALAH survey. Whenever possible, these values were converted to absolute magnitudes using the distance to stars inferred by the Bayesian approach [281]. Before using the pre-computed published distances, we removed all sources whose computed astrometric re-normalised unit-weight error [RUWE, 107] was greater than 1.4. The magnitudes of every individual star were also corrected for the reddening effect, except for the WISE

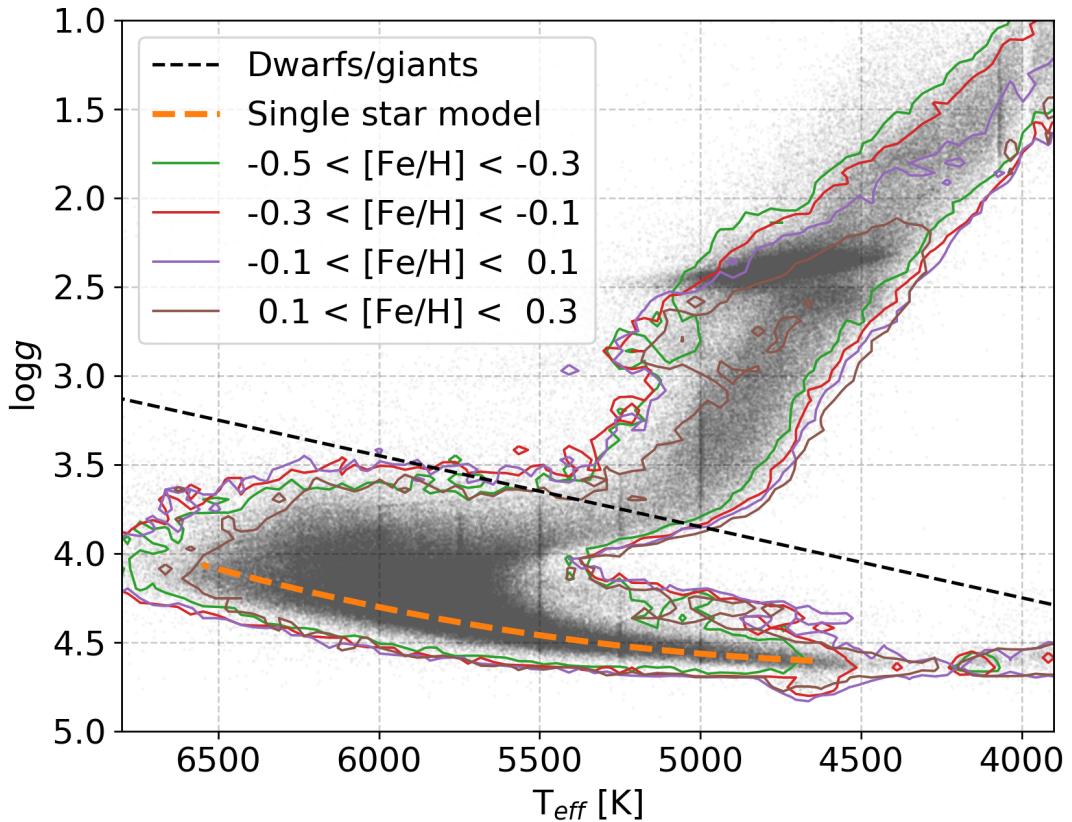


Figure 6.17: Complete observational set of valid stellar parameters shown as a varied density of grey dots. Contours around the diagram illustrate a coverage of parameter space in different $[Fe/H]$ bins. Overlaid orange dashed curve represents a relation on the main sequence from where single stars considered in the fit are drawn. Additionally, black dashed linear line represents arbitrarily defined limit between giant and dwarf stars that were used in the process of training a spectroscopic single star model.

photometric bands W1 and W2 that were considered to be extinction free.

Using the valid *The Cannon* stellar parameters, the inferred and corrected absolute magnitudes in multiple photometric bands were grouped into bins that contain stars within $\Delta T_{\text{eff}} = \pm 80$ K, $\Delta \log g = \pm 0.05$ dex, and $\Delta [Fe/H] = \pm 0.1$ dex around the bin centre. As spectroscopically unresolved multiple stellar systems could still be present in this bin, median photometric values are computed per bin to minimise their effect. Extrapolation outside this grid, that covers the complete observational stellar parameter space, is not desired nor implemented as it may produce erroneous values. When a photometric signature of a star with parameters between the grid points is requested, it is recovered by linear interpolation between the neighbouring grid points. The median photometric signature for the requested stellar parameters could also be computed on-the-fly using the same binning, but we found out that this produced insignificant difference and increased the processing time by more than a factor of 2.

6.11.3 Limitations in the parameter space

As already emphasized, spectroscopic and photometric models were built on real observations and are therefore limited by the training set coverage. The limitations are visually illustrated by Figure 6.17, where the dashed curve, from which single stars are drawn, is plotted over the observations. This arbitrary quadratic function is defined as:

$$\log g = 2.576 + 9.48 \cdot 10^{-4} T_{\text{eff}} - 1.10 \cdot 10^{-7} T_{\text{eff}}^2, \quad (6.7)$$

where values of T_{eff} and $\log g$ are given in units of K and cm s^{-2} respectively. The polynomial coefficients were determined by fitting a quadratic function to manually defined points that represent regions with the highest density of stars on the shown Kiel diagram. From Equation 6.7, it follows that the $\log g$ of a selected single star is not varied freely, but computed from the selected T_{eff} whose range is limited within the values $6550 > T_{\text{eff}} > 4650$ K.

Focusing on Solar-like stars gives us an advantage in their modelling as the whole observational diagram in Figure 6.17 is sufficiently populated with stars of Solar-like iron abundance. When going towards more extreme [Fe/H] values (high or low), coverage of the main sequence starts to decrease. For cooler stars, this happens at low iron abundance ($[\text{Fe}/\text{H}] < -0.3$) and for hot stars at high abundance ($[\text{Fe}/\text{H}] > 0.3$). Those limits pose no problems for our analysis, unless the wrong stellar configuration is used to describe the observations. At that point, the spectroscopic fitting procedure would try to compensate for too deep or too shallow spectral lines (effect of wrongly selected T_{eff}) by decreasing or increasing [Fe/H] beyond values reasonable for Solar-like objects.

6.12 Characterization of multiple system candidates

For the detailed characterization of Solar twin candidates that show excess luminosity, we used their complete available photometric and spectroscopic information. The excess luminosity can only be explained by the presence of an additional stellar component or a star that is hotter or larger than the Sun. Both of those cases can be investigated and confirmed by the data and models described in Sections 6.9 and 6.11. In the scope of our comparative methodology, we constructed a broad collection of synthetic single, double, and triple stellar systems that were compared and fitted to the observations.

As the measured *Gaia* DR2 parallaxes, and therefore inferred distances, of some objects are badly fitted or highly uncertain, the distance results provided by [281] actually yield three distinct distance estimates - the mode of an inferred distance distribution (`r_est`) and a near and distant distance (`r_lo`, and `r_hi`) estimation, between which 68 % of the distance estimations are distributed. As the actual shape of the distribution is not known and could be highly skewed, we did not draw multiple possible distances from the distribution, but only used its mode value.

6.12.1 Fitting procedure

A complete characterization and exploration fitting procedure for every stellar configuration (single, binary, and triple) consists of four consecutive steps that are detailed in the following sections. As we are investigating Solar twin spectra, the

initial assumption for the iron abundance of the system is set to $[Fe/H] = 0$. This also includes the assumption that stars in a system are of the same age, at similar evolutionary stages, and were formed from a similarly enriched material. If that is true, we can set iron abundance to be equal for all stars in the system. This notion is supported by the simulations [285] and studies [286] of field stars showing that close stars are very likely to be co-natal if their velocity separation is small.

The observed systems must be composed of multiple main sequence stars, otherwise the giant companion would dominate the observables, and the system would not be a spectroscopic match to the Sun. Therefore their parameters T_{eff} and $\log g$ are drawn from the middle of the main sequence determined by *The Cannon* parameters in the scope of the GALAH survey. The Kiel diagram of the stars with valid parameters and model of the main sequence isochrone used in the fitting procedure are shown in Figure 6.17.

6.12.2 Photometric fitting - first step

With those initial assumptions in mind, we begin with the construction of the photometric signature of the selected stellar configuration. To find the best model that describes the observations, we employ a Bayesian MCMC fitting approach [287], where the varied parameter is the effective temperature of the components. The selected T_{eff} values, and inferred $\log g$ (Equation 6.7), are fed to the photometric model (Section 6.11.2) to predict a photometric signature of an individual component. Multiple stellar signatures are combined together into a single unresolved stellar source using the following equation:

$$M_{\text{model}} = -2.5 \log_{10} \left(\sum_{i=1}^{n_s} 10^{-0.4M_i} \right); \quad n_s = [1, 2, 3], \quad (6.8)$$

where M_i denotes absolute magnitude of a star in one of the used photometric bands, and n_s number of components in a system. The newly constructed photometric signature at selected T_{eff} values is compared to the observations using the photometric log-likelihood function $\ln p_P$ defined as:

$$\ln p_P(T_{\text{eff}} | M, \sigma) = -\frac{1}{2} \sum_{i=1}^{n_p} \left[\frac{(M_i - M_{\text{model},i})^2}{\sigma_i^2} + \ln(2\pi\sigma_i^2) \right], \quad (6.9)$$

where M and σ represent extinction corrected absolute magnitudes, and their measured uncertainties that were taken for multiple published catalogues presented in Section 6.9. The constructed photometric model of a multiple system is represented by the variable M_{model} and the number of photometric bands by n_p . The maximum, and most common value for n_p is 13, but in some cases, it can drop to as low as 8. The MCMC procedure is employed to maximise this log-likelihood and find the best fitting stellar components.

To determine the best possible combination of T_{eff} values, we initiate the fit with 1200 uniformly distributed random combinations of initial temperature values that span the parameter space shown in Figure 6.17. The number of initial combinations is intentionally high in order to sufficiently explore the temperature space. Excessive or repeated variations of initial parameters are rearranged by a prior limitation that the temperatures of components must be decreasing, therefore $T_{\text{eff}1} >= T_{\text{eff}2} >= T_{\text{eff}3}$

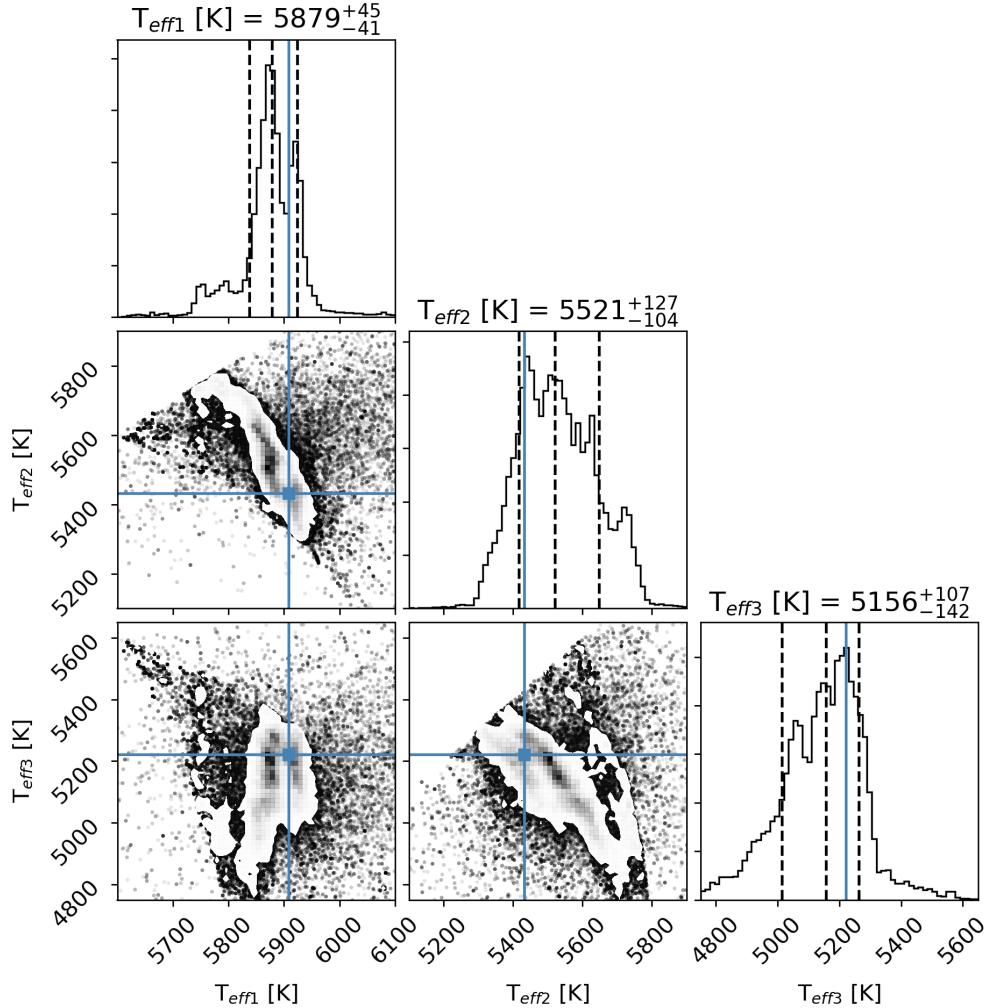


Figure 6.18: Distribution of considered posteriors during the initial MCMC photometric fit for one of the objects that was at the end classified as a triple candidate. As the plots show the first step in the fitting procedure that explores the complete parameter space, the distributions are not expected to be smooth because of possible multiple local minima. Values indicated on the scatter plots represent medians for the 10 % of the best fitting solutions.

in the case of a triple system (example of used initial walker parameters is shown in Figure 6.29). The initial conditions are run for 200 steps. The number of steps was selected in such a way to ensure a convergence of all considered cases (example of the walkers convergence is shown in Figure 6.30). The distribution of priors for such a run is shown as a corner plot in Figure 6.18. After that, only the best 150 walkers are kept, their values perturbed by 2 %, and run for another 200 steps to determine the posterior distribution of the parameters varied during the MCMC fit. This two-step run is needed to speed up the process and discard solutions with lower $\ln p_P$. During the initial tests, we found that the investigated parameter space can have multiple local minima which attract walkers, especially in the case of a triple system.

After the completion of all MCMC steps, the considered parameter combinations are ordered by their log-likelihood in descending order. Of those, only 10 % of

the best fitting combinations are used to compute final $T_{\text{eff[1-3]}}$ values. They are computed as the median of the selected best combinations.

6.12.3 Spectroscopic fitting - second step

After an effective temperature of the components has been determined, we proceed with the evaluation of how well they reproduce the observed spectrum. As a majority of the fitted systems do not consist of multiple components with a T_{eff} equal to Solar, the [Fe/H] of the system must be slightly changed to equalize absorption strength of the simulated and observed spectral lines.

To determine the [Fe/H] of the system, we first compute a simulated spectrum for every component using a spectroscopic model described in Section 6.11.1. Individual spectra are afterwards combined using Equation 6.10 in the case of a binary system or Equation 6.11 in the case of a triple system.

$$r_{12} = \frac{L_{2,\lambda}}{L_{1,\lambda}}; \quad \lambda = [1, 2, 3, 4] \quad (6.10)$$

$$f_{model,\lambda} = \frac{f_{1,\lambda}}{1 + r_{12}} + \frac{f_{2,\lambda}}{1 + 1/r_{12}}$$

$$r_{12} = \frac{L_{2,\lambda}}{L_{1,\lambda}}, \quad r_{13} = \frac{L_{3,\lambda}}{L_{1,\lambda}} \quad (6.11)$$

$$f_{model,\lambda} = \frac{f_{1,\lambda}}{1 + r_{12} + r_{13}} + \frac{r_{12}f_{2,\lambda}}{1 + r_{12} + r_{13}} + \frac{r_{13}f_{3,\lambda}}{1 + r_{12} + r_{13}}$$

Individual normalised spectra, denoted by $f_{n,\lambda}$ in Equations 6.10 and 6.11, are weighted by the luminosity ratios between the components (r_{xy}) and then summed together.

As the HERMES spectrum covers four spectral ranges, whose distribution of spectral energy depends on stellar T_{eff} , different luminosity ratios also have to be used for every spectral arm. Of all of the used photometric systems, APASS filters B, g', r', and i' have the best spectral match with blue, green, red, and infrared HERMES arm. The modelled APASS magnitudes of the same stars are used to compute luminosity ratios between them.

The described summation of the spectra introduces an additional assumption about the analysed object. With this step, we assume that components have a negligible internal spread of projected radial velocities that could otherwise introduce asymmetries in the shape of observed spectral lines. The assumption allows us to combine individual components without any wavelength corrections.

Similarly, as in the previous case, a Bayesian MCMC fitting procedure was used to maximise the spectroscopic log-likelihood $\ln p_S$ defined as:

$$\ln p_S([\text{Fe}/\text{H}]|f, \sigma) = -\frac{1}{2} \sum_{\lambda} \left[\frac{(f_{\lambda} - f_{model,\lambda})^2}{\sigma_{\lambda}^2} + \ln(2\pi\sigma_{\lambda}^2) \right], \quad (6.12)$$

where f and σ represent the observed spectrum and its per-pixel uncertainty, respectively. A modelled spectrum of the system, at selected [Fe/H], is represented by the variable f_{model} and the number of wavelength pixels in that model by λ . Combined, all four spectral bands consist of almost 16,000 pixels. T_{eff} values of the components are fixed for all considered cases.

The MCMC fit is initiated with 150 randomly selected [Fe/H] values, whose uniform distribution is centred at the initial [Fe/H] value of the system and has a span of 0.4 dex. All of the initiated walkers are run for 100 steps. At every [Fe/H] level, a new simulated spectrum composite is generated and compared to the observed spectrum by computing log-likelihood $\ln p_S$ of a selected [Fe/H] value. The range of possible [Fe/H] values considered in the fit is limited by a flat prior between -0.5 and 0.4 .

By the definition of [Fe/H] in the scope of GALAH *The Cannon* analysis, the parameter describes stellar iron abundance and not its metallicity as commonly used in the literature. Therefore only spectral absorption regions of un-blended Fe atomic lines are used to compute the spectral log-likelihood. Having to fit only one variable at a time, the solution is easily found and computed as a median value of all posteriors considered in the fit.

6.12.4 Final fit - third step

A changed value of [Fe/H] for the system will introduce subtle changes to its photometric signature, therefore we re-initiate the photometric fitting procedure. It is equivalent to the procedure described in Section 6.12.3, but with much narrower initial conditions. These new initial conditions are uniformly drawn from the distribution centred at T_{eff} values determined in the first step of the fitting procedure. The width of the uniform distribution is equal to 100 K. Drawn initial conditions are afterwards run through the same procedure as described before.

At this point, the second and third step in the fitting procedure can be repeatedly run several times to further pinpoint the best solution. We found out that further refinement was not needed in our case as it did not influence the determined number of stars in the system.

6.12.5 Number of stellar components - final classification

The fitting procedure described above was used to evaluate observations of every multiple stellar candidate to determine whether they belong to a single, binary or triple stellar system. This resulted in the following set of results for every configuration: predicted T_{eff} of the components, [Fe/H] of the system, simulated spectrum, and simulated photometric signature of the system.

As the photometric and spectroscopic fits do not always agree on the best configuration, the following set of steps and rules was applied to classify results in one of six classes presented in Table 6.3.

- Compute χ^2 between the simulated photometric signature of the modelled system and extinction corrected absolute photometric observations for every considered stellar configuration.
- Compute χ^2 between the simulated spectrum of the modelled system and the complete GALAH observed spectrum for every considered stellar configuration.
- Independently select the best fitting configuration with the lowest χ^2 for photometric and spectroscopic fit.

Table 6.3: Number of different systems discovered by the fitting procedure performed on possible multiple stars that exhibit excess luminosity.

Configuration classification	Number of systems
1 star	2
≥ 1 star	14
2 stars	27
≥ 2 stars	14
3 stars	6
Inconclusive	1
Total objects	64

- If the best photometric and spectroscopic fit point to the same configuration, then the system is classified as having a number of stars defined by both fitting procedures.
- If the best photometric and spectroscopic fit do not point to the same configuration, then the system is classified as having at least as many stars as determined by the prediction with a lower number of stars (e.g. ≥ 2 stars).
- If the difference between those two predictions is greater than 1 (e.g. photometric fit points to a single star and spectroscopic to a triple star), then the system is classified as inconclusive.

The classification produced using these rules is shown as colour coded *Gaia* colour-magnitude diagram in Figure 6.19.

6.12.6 Quality flags

In addition to our final classification, we also provide an additional quality checks that might help to identify cases for which our method might return questionable determination of a stellar configuration. Every of those checks, listed in Table 6.4, is represented by one bit of a parameter `flag` in the final published table (Table 6.8). The first bit gives us an indication of whether the object could have an uncertain astrometric solution, whereas the second and the third bits indicate if the final fitted solution has a worse match with the observations than the parameters produced by the *The Cannon* pipeline. To evaluate this, we used the original stellar parameters reported for the object to construct their photometric and spectroscopic synthetic model that was compared to the observations by computing their χ^2 similarity (`m_sim_p` and `m_sim_f` in Table 6.8). The resulting fitted spectrum or photometric signature is marked as deviating if its similarity towards observations is worse than for the reported one star parameters. This might not be the best indication of possible mismatch as it is common that *The Cannon* parameters of the analysed multiple candidates deviate from the main sequence in Kiel diagram (Figure 6.17) and therefore fall into less populated parameter space, where they can skew the single star models (Sections 6.11.1 and 6.11.2).

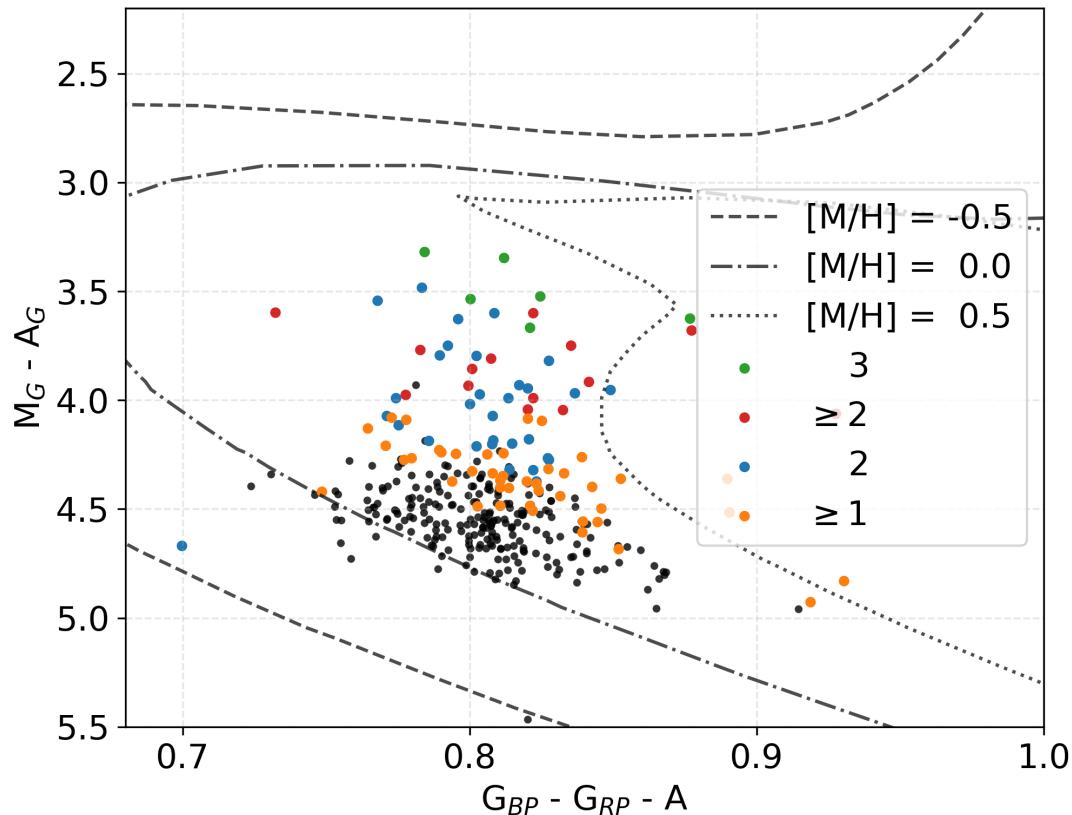


Figure 6.19: Showing the same data points as Figure 6.16 with indicated definite triple (green dots), possible triple (red dots), binary (blue dots), and possible binary stellar systems (orange dots). All other classes are shown with black dots. Overplotted are PARSEC isochrones [284] for stars with the age of 4.5 Gyr and different metallicities.

Table 6.4: Explanation of the used binary quality flags in the final classification of the stellar configuration. A raised bit could indicate possible problems or mismatches in the determined configuration. Symbol X in the last two descriptions represents the best fitting configuration, therefore $X = [1,2,3]$.

Raised bit	Description
0	None of the flags was raised
1 st bit	High astrometric uncertainty ($\text{RUWE} > 1.4$)
2 nd bit	Deviating photometric fit ($sX_{\text{sim_p}} > m_{\text{sim_p}}$)
3 rd bit	Deviating spectroscopic fit ($sX_{\text{sim_s}} > m_{\text{sim_s}}$)

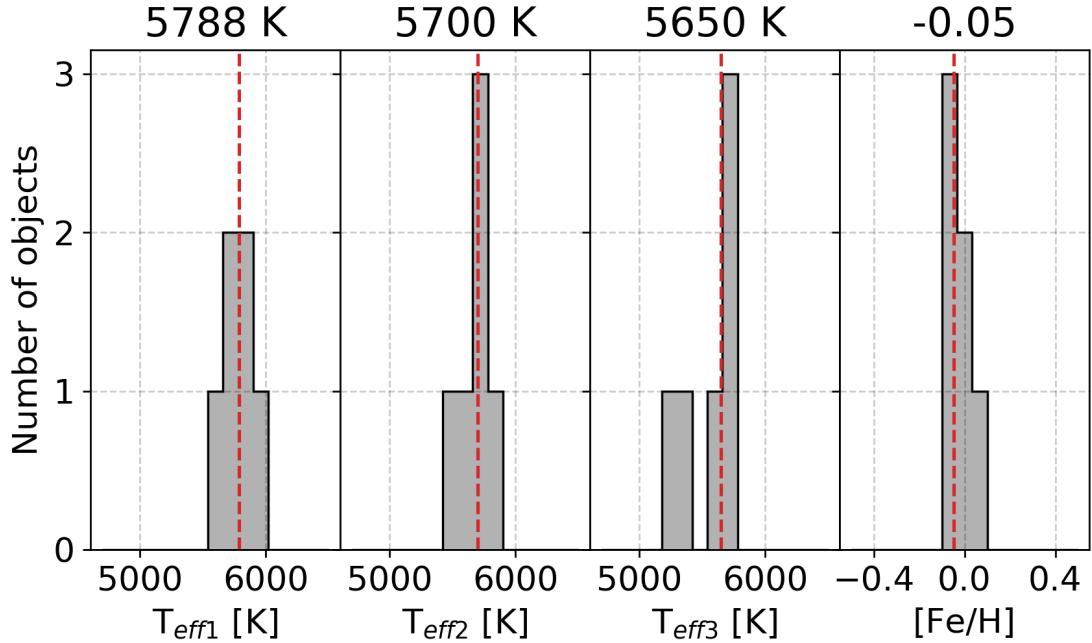


Figure 6.20: Parameters of triple stellar systems discovered and characterized by our analysis. Histograms represent distribution of all fitted results for 6 objects that were classified as triple stellar systems and observationally mimic Solar spectrum. Median values of the distributions are given above individual histogram and indicated with dashed vertical line.

6.13 Characterization of single star candidates

Once we concluded our analysis of the set of 64 objects that showed obvious signs of excess luminosity, we then proceeded to study the remaining 265 objects that are most probably not part of a complex stellar system. To explore their composition, they were analysed with the same procedure as multiple candidates (Section 6.12). Before running the procedure, we omitted the option to fit for a triple system as they clearly do not possess enough excess luminosity for that kind of a system.

The obtained results were analysed and classified using the same set of rules introduced in Section 6.12.5. Retrieved classes are summarized in Table 6.5 and in Figure 6.19. In the latter we see that the potential binaries are located on the top of the colour-magnitude diagram, which is consistent with the potential presence of an additional stellar source.

6.14 Orbital period constraints

The observational data we have gathered on possible multiple systems point to configurations that change slowly as a function of time. Using the observational constraints and models that describe the formation of the observed spectra, we try to set limiting values on the orbital parameters of the detected triple stellar system candidates. In order to have a greater sample size, we use both definite (class 3) and probable (class ≥ 2) triple stars.

With the limited set of observations, we have to set assumptions about the constitution of those systems. For a hierarchical $2 + 1$ system to be dynamically

Table 6.5: Number of different systems discovered by the fitting procedure performed on stars that do not exhibit excess luminosity. Classes and their description is the same as used in Table 6.3. In addition, number of stars without parallactic measurements is added for completeness.

Configuration classification	Number of systems
1 star	230
≥ 1 star	31
2 stars	4
≥ 2 stars	0
3 stars	0
Inconclusive	0
Unknown parallax	0
Total objects	265

stable on long timescales, its ratio between the orbital period of an inner pair P_S and outer pair P_L must be above a certain limit. Eggleton [288] showed that P_L/P_S must be higher than 5. The same lower limit is noticeable when the correlation between periods of the known triple stars is plotted [289, 290].

In order to estimate the periods, we have to know the masses and distances between the stars in a system. Without complete information about the projected velocity variation in a system, masses can also be inferred from the spectral type. As we are looking at the Solar twin triples, whose effective temperatures are all very similar and close to Solar values (see Figure 6.20), our rough estimate is that all stars also have a Solar-like mass $\sim M_\odot$. From this, we can set the inner mass ratio q_S to be close to 1 and the outer mass ratio $q_L \sim 0.5$. When we are dealing with the outermost star, the inner pair is combined into one object with twice the mass of the Sun. The likelihood of such a configuration is also supported by the observations [289] where a higher concentration of triple systems is present around those mass ratios. Contrary to our systems, twin binaries with equal masses usually have shorter orbital periods [291].

With the initial assumption about the configuration of the triple systems and masses of the stars, the periods of the inner and outer pair can be constrained to some degree as other orbital elements (inclination, ellipticity, phase ...) are not known.

6.14.1 Outer pair and *Gaia* angular resolution

Limited by its design, *Gaia* spacecraft and its on-ground data processing are in theory unable to resolve stars with the angular separation below ~ 0.1 arcsec. Above that limit, their separability is governed by the flux ratio of the pair. The validation report of the latest data release [292] shows that the currently achieved angular resolution is approximately 0.4 arcsec as none of the source pairs is found closer than this separation. We used this reported angular limit to assess possible or-

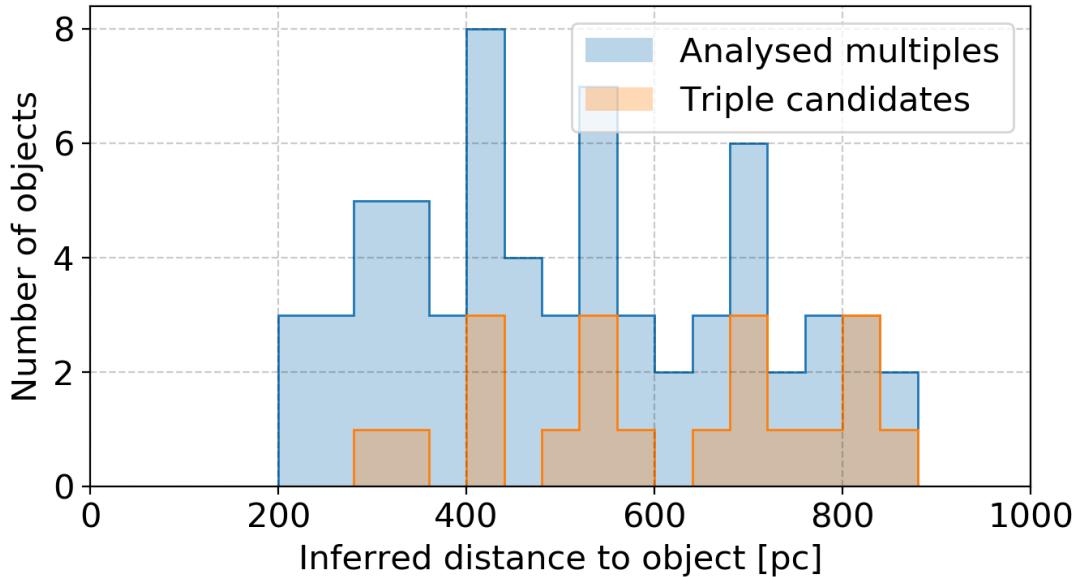


Figure 6.21: Histogram showing distribution of radial distances for the analysed set of stars. Distances were inferred by the Bayesian approach that takes into account the distribution of stars in the Galaxy [281].

bital configurations that are consistent with both the spectroscopic and astrometric observations.

Along with the final astrometric parameters, the *Gaia* data set also contains information about the goodness of the astrometric fit. Evans [60] used those parameters to confirm old and find new candidates for unresolved exoplanet hosting binaries in the data set. As we are looking at a system of multiple stars that orbit around their common centre of mass, we expect their photo-centre to slightly shift during such orbital motion. The observed wobble of the photo-centre also depends on the mass of stars in the observed system. In the case of a binary system containing two identical stars, the wobble would not be observed, but its prominence increase as the difference between their luminosities becomes more pronounced. This subtle change in the position of a photo-centre adds additional stellar movement to the astrometric fit, consequently degrading its quality. Improvements for such a motion will be added in latter *Gaia* data releases. If a system has an orbital period much longer than the time-span of *Gaia* DR2 measurements (22 months), its movement has not yet affected the astrometric solution. This puts an upper limit on an orbital period as it should not be longer than few years in order to already affect the astrometric fit results.

Setting limits to the parameters of astrometric fit quality [`astrometric_excess_noise > 5`, and `astrometric_gof_al > 20` as proposed by 60], none of our 329 stars meets those requirements. This suggests that all of them are most likely well below the *Gaia* separability limit and/or have long orbital periods. Another indicator for a lower-quality astrometric fit that we can use is RUWE. Figure 6.16 shows distribution of potentially problematic large RUWE among single and multiple candidates. The latter, on average, have a much poorer fit quality that might indicate a presence of an additional parameter that needs to be considered in the astrometric fit.

Distances to triple stars, shown by the histogram in Figure 6.21 range from around 0.3 to 0.9 kpc. From this we can assume the maximal allowable distance between components of an outer pair to be in the order of 100 – 350 AU, pointing to outer orbital periods larger than 500 years. To test if such systems would meet our detection constraints, we created 100,000 synthetic binary systems whose orbital parameters were uniformly distributed within the parameter ranges given in Table 6.6. Observable radial velocities of both components were computed using the following equations:

$$v_1 = \frac{2\pi \sin i}{P\sqrt{1-e^2}} \frac{aq}{1+q} (\cos(\theta + \omega) + e \cos \omega) \quad (6.13)$$

$$v_2 = \frac{2\pi a \sin i}{P\sqrt{1-e^2}} \frac{a}{1+q} (\cos(\theta + \omega + \pi) + e \cos(\omega + \pi)) \quad (6.14)$$

$$P = \sqrt{a^3 \frac{4\pi^2}{GM_1(1+q)}} \quad (6.15)$$

The distribution of velocity separations ($\Delta v = v_2 - v_1$) for synthetic systems is given by Figure 6.22, where we can see that more than 99.7 % of generated configurations would produce a spectrum that would still be considered as a Solar twin ($\Delta v < 6 \text{ km s}^{-1}$, see Section 6.15.1 and Figure 6.31 for further clarification). If we set the semi-major axis a (in Table 6.6) to single value in the same simulation, we can find the closest separation that would still meet the same observational criteria in at least 68 % of the cases. In our simulation this happens at the mutual separation of 10 AU (and at 50 AU for 95 % of the cases). The orbital period of a such outer pair is about 18 years (and 200 years for 50 AU) and is most probably way too long to had significant effect on the quality of the astrometric fit. Considering observationally favorable orbital configurations with face-on orbits, the actual system could be much more compact than estimated.

6.14.2 Inner binary pair and formation of double lines in a spectrum

An upper estimate of orbital sizes for an outer pair gives us a confirmation that almost every considered random orbital configurations would satisfy the observational and selection constraints. To ensure the long-term stability of such a system, the inner pair must have a period that is at least five times shorter [288]. At such short orbital periods, the inner stars could potentially move sufficiently rapidly in their orbits as to produce noticeable absorption line splitting for edge-on orbits. To be confident that none of the analysed objects produces such an SB2 spectrum, we visually checked all considered spectra and found no noticeable line splitting in any of the acquired GALAH or Asiago spectra. A subtle hint about a possible broadening of the spectral lines comes from the determined $v \sin i$. The median of its distribution in Figure 6.23 is higher for multiple candidates with excess projected velocity of $\sim 0.5 \text{ km s}^{-1}$.

Accounting for the GALAH resolving power and spectral sampling, we can estimate a minimal radial velocity separation between components of a spectrum to show clear visual signs of duplicated spectral lines. To determine a lower limit, we

Table 6.6: Ranges of the orbital parameters used for the prediction of observable radial velocity separation between stars in an outer binary pair. The range of the semi-major axis length a is set between the *Gaia* separability limit for the closest and farthest triple candidate. The uniform distribution of a is a good approximation of the real periodicity distribution published by Raghavan *et al.* [242] as we are sampling a narrow range of it. Use of the real observed distribution would in our case introduce insignificant changes in radial velocity separation as we are simulating wide, slowly rotating systems.

Parameter	Considered range
M_1	$2 M_{\odot}$
a	$100 \dots 350$ AU
q	$0.45 \dots 0.55$
$\sin i$	$0 \dots 1$ ($i = 0 \dots 90$ deg)
e	$0.1 \dots 0.8$
phase	$0 \dots 1$, used for calculation of θ
ω	$0 \dots 360$ deg

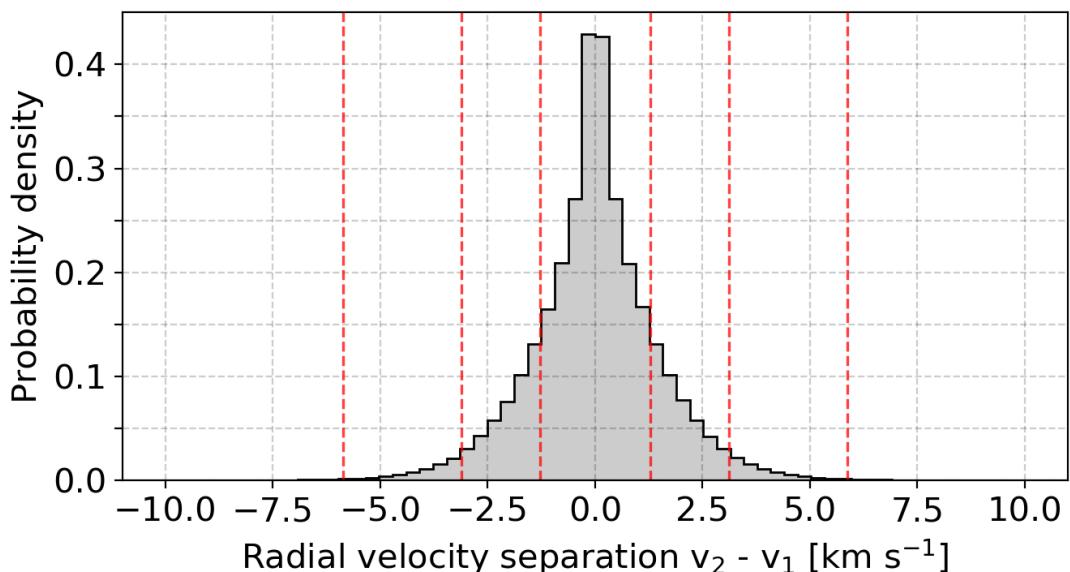


Figure 6.22: Distribution of computed radial velocity separations between a primary and secondary component of the simulated binary systems defined by the orbital parameters given in Table 6.6. Red vertical lines show 1, 2, and 3 σ probabilities of the given distribution.

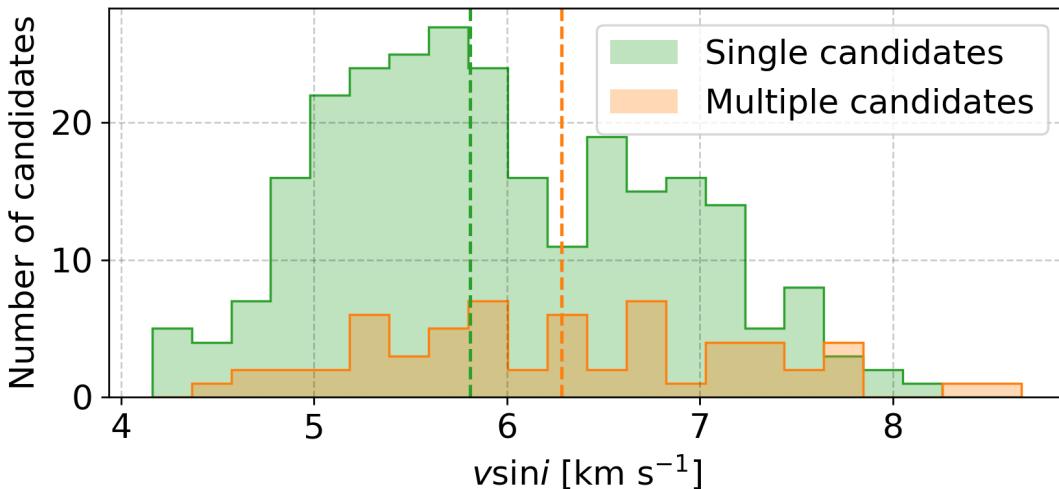


Figure 6.23: Distribution of determined $v \sin i$ for analysed single and multiple candidates. Median values of the distributions are marked with dashed vertical lines.

combined two Solar spectra of different flux ratios and visually evaluated when the line splitting becomes easily noticeable. With equally bright sources, this happens at the separation of $\sim 14 \text{ km s}^{-1}$. When the secondary component contributes $1/3$ of the total flux, minimal separation is increased to $\sim 20 \text{ km s}^{-1}$. A similar separation is needed when secondary contributes only 10% of the flux. As no line splitting is observed in our analysed spectra, we can be confident that the velocity separation between the binary components was lower than that during the acquisition.

Considering the minimal ratio between outer and inner binary period, we can deduce an expected radial velocity separation of an inner pair for the widest possible orbits. As explained in previous section, we used Equations 6.13-6.15 and possible ranges of inner orbital parameters (Table 6.7) to generate a set of synthetic binary systems. The distribution of their Δv is shown in Figure 6.24 and represent inner binaries with orbital periods from 100 to 700 years. At those orbital periods, more than 92 % of the considered configurations would satisfy the condition of $\Delta v < 4 \text{ km s}^{-1}$, ensuring that the observed composite of two equal Solar spectra would still be considered as a Solar twin (see Section 6.15.1 and Figure 6.26 for further clarification). If we limit our synthetic inner binaries to only one orbital period, we can estimate the most compact system that would still meet the observational criteria in at least 68 % of the cases. This would happen at the orbital period of 40 years with a semi-major axis of 14 AU.

6.14.3 Multi-epoch radial velocities

To support our claims about slowly changing low orbital speeds in detected triple candidates, we analysed changes in their measured radial velocities between the GALAH and other comparable all-sky surveys. Distributions of changes are presented by three histograms in Figure 6.25 where almost all velocity changes, except one, are within 5 km s^{-1} . Differences between the GALAH and *Gaia* radial velocities were expected to be small as the latter reports median velocities in the time-frame

Table 6.7: Same as Table 6.6, but for an inner binary of a hierarchical triple stellar system.

Parameter	Considered range
M_1	M_\odot
P_S	$P_L / 5$, used for calculation of a
q	0.9 … 1.0
$\sin i$	0 … 1 ($i = 0 \dots 90$ deg)
e	0.1 … 0.8
phase	0 … 1, used for calculation of θ
ω	0 … 360 deg

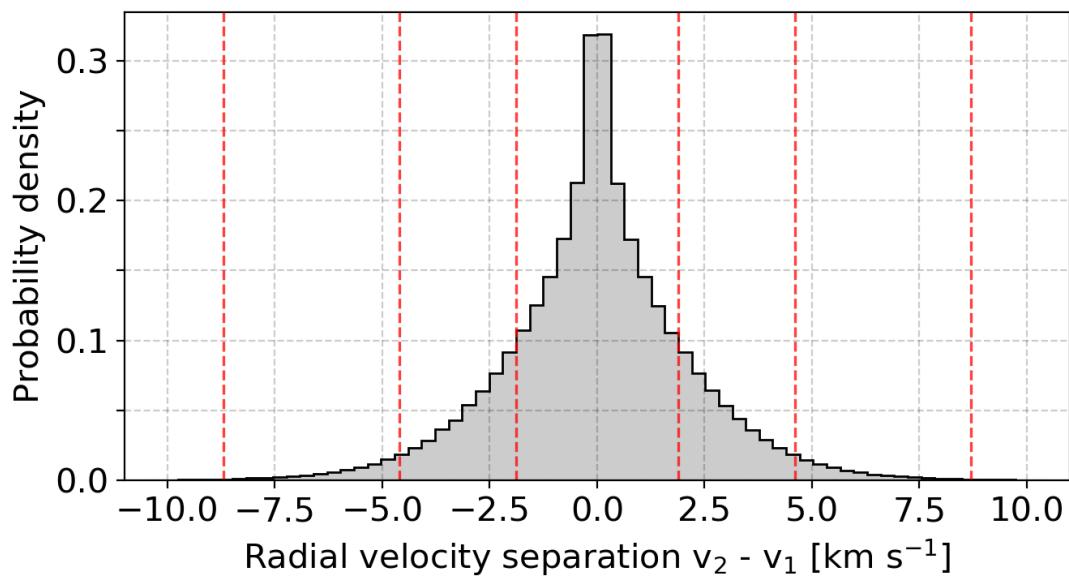


Figure 6.24: Same as Figure 6.22, but for an inner binary of a simulated triple stellar system.

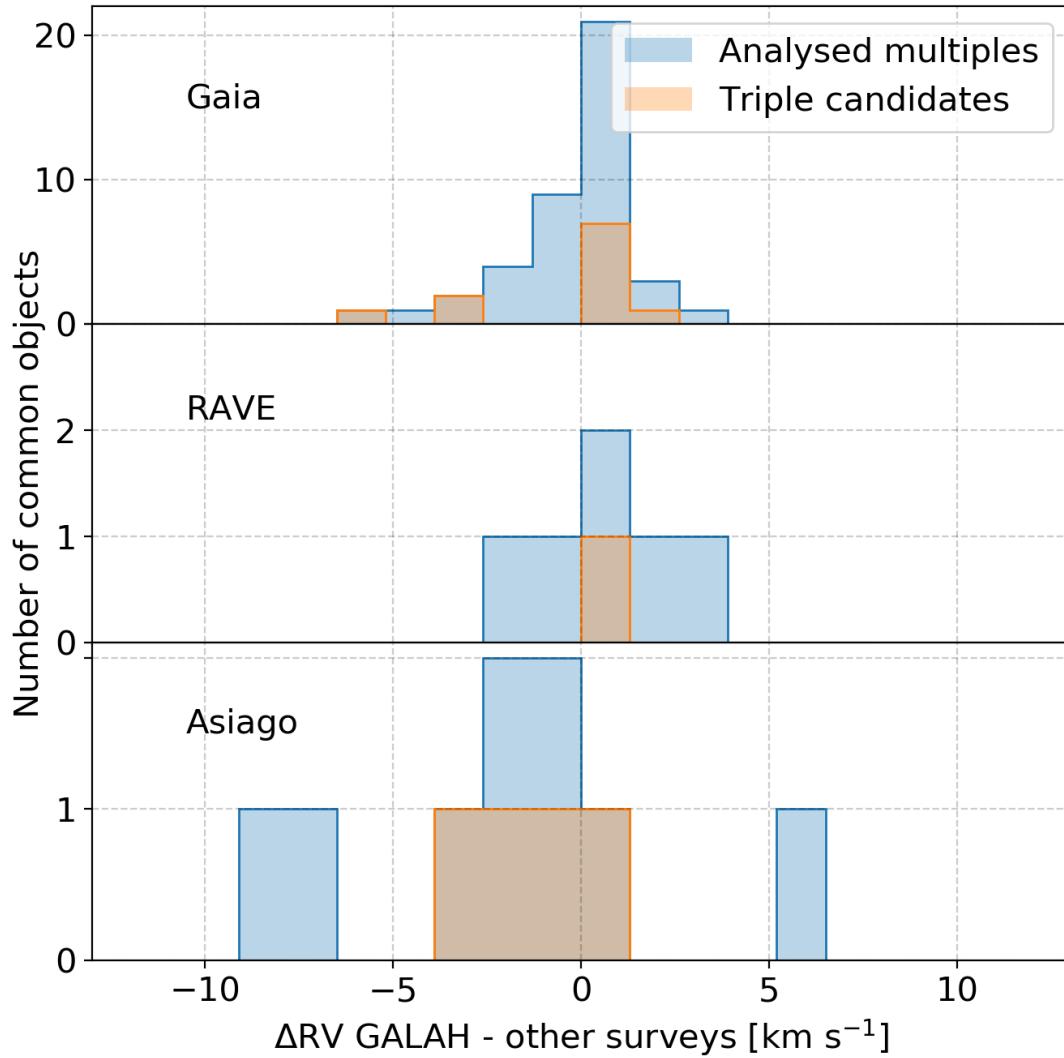


Figure 6.25: RV difference between GALAH and other large spectroscopic surveys. The name of an individual survey or observatory is given in the upper-left corner of every panel. Blue histograms represent velocity difference for all investigated multiple candidates and orange histograms for discovered triple systems only.

that is similar to the acquisition span of the GALAH spectra. Extending the GALAH observations in the past with RAVE and in the future with Asiago observations did not produce any extreme changes. For this comparison, we also have to consider the uncertainty of the measurements that are in the order of $\sim 2 \text{ km s}^{-1}$ for RAVE and $\sim 1 \text{ km s}^{-1}$ for Asiago spectra.

With the data synergies, we could produce only three observational time-series that have observations at more than two sufficiently separated times. With only three data points in each graph, not much can be said about actual orbits.

6.15 Simulations and tests

To determine the limitations of the employed algorithms, we used them to evaluate a set of synthetic photometric and spectroscopic sources. As the complete procedure depends on the criteria for the selection of Solar twin candidates, we first investigated if the selection criteria allows for any broadening of the spectral lines or their multiplicity as they are both signs of a multiple system with components at different projected radial velocities.

In the second part, we generated a set of ideal synthetic systems that were analysed by the same fitting procedure as the observed data set. The results of this analysis are used to determine what kind of systems could be recognized with the fitting procedure and how the results could be used to spot suspicious combinations of fitted parameters.

In the final part, we try to evaluate selection biases that might arise from the position of analysed stars on the *Gaia* colour-magnitude diagram and from the GALAH selection function that picks objects based on their apparent magnitude.

6.15.1 Radial velocity separation between components

To determine the minimal detectable radial velocity (RV) separation between components in a binary or a triple system, we constructed a synthetic spectrum resembling an observation of multiple Suns at a selected RV separation. The spectrum of a primary component was fixed at the rest wavelength with $\text{RV} = 0 \text{ km s}^{-1}$ and a secondary spectrum shifted to a selected velocity. After the shift, these two spectra were added together based on their assumed flux ratio.

The generated synthetic spectrum was compared to the Solar spectrum with exactly the same metric as described in Section 6.10. Computed spectral similarities at different separations were compared to the similarities of analyzed Solar twin candidates. The first RV separation that produces a spectrum that is more degraded than the majority of Solar twin candidates was determined to be a minimal RV at which the observed spectrum would be degraded enough that it would no longer be recognized as a Solar twin. The high SNR of the generated spectrum was not taken into account for this analysis in contrast to the algorithm that was used to pinpoint Solar twin candidates. Therefore we also omitted candidates with a lower similarity that in our case directly corresponds to their low SNR.

The result of this comparison is presented in Figure 6.26, where we can see that the minimal detectable separation of two equally bright stars resembling the Sun is $\sim 4 \text{ km s}^{-1}$. In the case where a primary star contributes $2/3$ of the total flux, the minimal RV increases to $\sim 6 \text{ km s}^{-1}$ (see Figure 6.31). Further increase in the ratio between their fluxes would also increase a minimal detectable separation, but only to a certain threshold from where on a secondary star would not contribute enough flux for it to be detectable, and its received flux would be comparable to the typical HERMES spectral noise. In our case, this happens when the secondary contributes less than 10 % of the total flux. These boundaries are only indicative as they also heavily depend on the quality of the acquired spectra. When a low level of noise with the Gaussian distribution ($\sigma = 0.01$) is added to a secondary component with a comparable luminosity, the minimal RV decreases because the similarity between spectra also decreases. In that case, the similarity for $\Delta v = 0$ is located near the

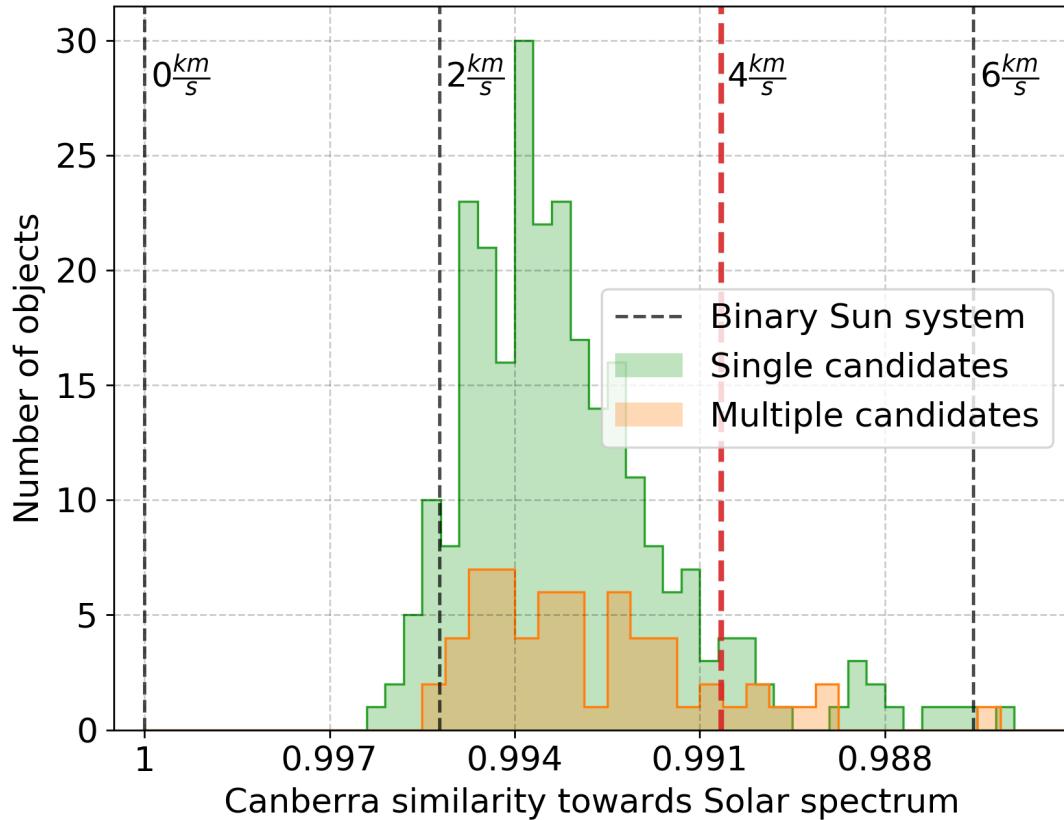


Figure 6.26: Histograms of Canberra similarities towards Solar spectrum for multiple stellar candidates (orange histogram) and single star candidates (green histogram) in the red HERMES arm. Vertical dashed lines represent the same similarity measure, but for a binary system comprising of two equally bright Suns, whose components are at different radial velocities. The separation between components is increasing in 2 km s^{-1} steps, where the leftmost vertical line represents the case where both stars are moving with the same projected velocity. The selected maximal velocity at which the composite spectrum would still be considered a Solar twin is marked with the thicker red vertical line. Distribution of histograms also shows that spectra of multiple system candidates are as (dis)similar as spectra of single stars.

mode value of similarity distribution in Figure 6.26.

6.15.2 Analysis of synthetic multiple systems

The limitations and borderline cases of the fitting procedure were tested by evaluating its performance on a set of synthetic binary and triple systems that were generated using the same models and equations as described in the fitting procedure. An ideal, virtually noiseless set comprised 95 binary and 445 triple systems whose components differ in temperature steps of 100 K. The hottest component in the system was set to values in the range between 5300 and 6200 K. The T_{eff} of the coldest component could go as low as 4800 K. The [Fe/H] of all synthetic systems was set to 0.0 to mimic Solar-like conditions. Condensed results are presented in Figures 6.27 and 6.32.

As expected, the fitting procedure (Section 6.12) did not have any problem determining the correct configuration of the synthesized system. Temperatures of the components were also correctly recovered with a median error of 0 ± 13 K for the hottest component and 0 ± 43 K for the coldest component in a triple system. A more detailed analysis with the distribution of prediction errors, where the temperature difference between components is taken into account, is presented in Figure 6.27. From that analysis, we can deduce that T_{eff} of the secondary component is successfully retrieved if it does not deviate by more than 1000 K from the primary. Results at such large temperature difference are inconclusive as the number of simulated systems drops rapidly. The same can be said for the tertiary component, but with the limitation that it should not be colder by more than 700 K when compared to the secondary star. Beyond that point, the uncertainty of the fitted result increases and a star is determined to be hotter as it really is.

Another application of such an analysis is to identify signs that could point to a possibly faulty solution when it is comparably likely that a multiple system is comprised of two or three components. The results of such an analysis are presented in Figure 6.32, where we tried to describe a binary system with a triple system fit. From the distribution of errors, we can observe that the effective temperature of a primary star with the largest flux is recovered with the smallest fit errors whose median is 20 ± 93 K. As we are using too many components in that fit, T_{eff} of a secondary is reduced in order to account for the redundant tertiary component in the fit. As imposed by the limit in the fitting procedure, the tertiary $T_{\text{eff}3}$ is set to as low as possible. If the same thing happens for a real observed system, this could be interpreted as a model over-fitting.

6.15.3 Triple stars across the H-R diagram

In the current stage of Galactic evolution, binary stars with a Solar-like T_{eff} are located near a region that is also occupied by main sequence turnoff stars in the Kiel and colour-magnitude diagram (Figures 6.17 and 6.19). This, combined with the fact that older stars with a comparable initial mass and metallicity would also pass a region occupied with binaries (Figure 6.16) poses an additional challenge for detection of unresolved multiples if their spectrum does not change sufficiently during the evolution.

To analyse the possible influence of the turnoff stars and older, more evolved stars, we ran the same detection procedure as described in Section 6.10 and analysis (Section 6.12) to determine the fraction of binaries and triples at different T_{eff} , ranging from 5100 to 6000 K, with a step of 100 K. A comparison median spectrum was computed from all spectra in the range $\Delta T_{\text{eff}} \pm 60$ K, $\Delta \log g \pm 0.05$ dex and $\Delta [\text{Fe}/\text{H}] \pm 0.05$ dex, where the main sequence $\log g$ was taken from the main sequence curve shown in Figure 6.17, and $[\text{Fe}/\text{H}]$ was set to 0.0. The limiting threshold for multiple candidates was automatically determined in the same way as for Solar twin candidates (Section 6.10.1).

Condensed results of the analysis are shown in Figure 6.28, where we can see that both the fraction of stars with excessive luminosity and triple candidates starts increasing above the $T_{\text{eff}} \sim 5600$ K. This might indicate that the underlying distribution of more evolved and/or hotter stars might have some effect on our selection function. On the other hand, this increasing binarity fraction coincides with other

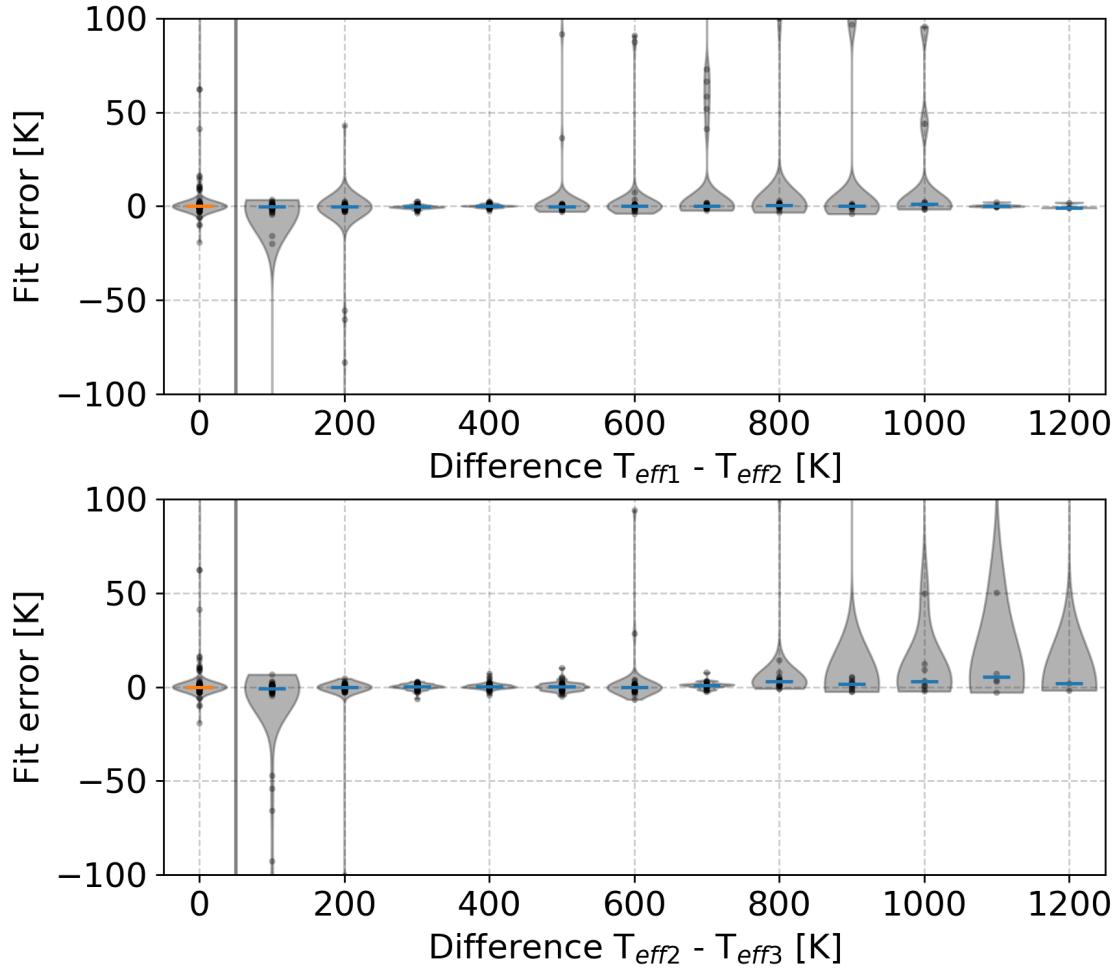


Figure 6.27: The accuracy of our analysis when the fitting procedure is applied to a synthetic triple system. Upper panel shows the distribution of T_{eff} prediction errors for a secondary star depending on the difference between selected temperatures of a primary and a secondary star. The lower panel shows the same relation but for a tertiary star in comparison to a secondary. As a reference, the prediction error of a primary star is shown on the left side of both panels. Labels $T_{\text{eff}1}$, $T_{\text{eff}2}$, and $T_{\text{eff}3}$ indicate decreasing effective temperatures of stars in a simulated system.

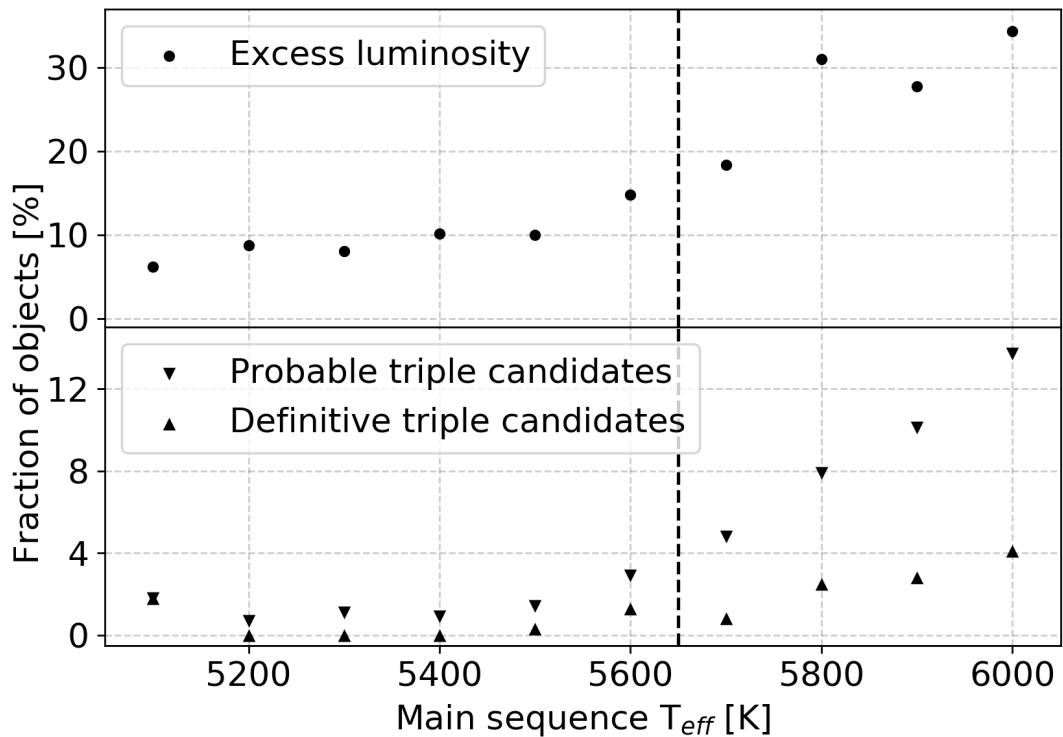


Figure 6.28: The upper panel shows the percentage of objects at different positions on the Kiel diagram shown in Figure 6.17 that show excessive luminosity and are spectroscopically similar to main sequence stars. Similarly, the lower panel shows upper and lower boundary on a percentage of triple system candidates at the same positions. For a definitive candidate, both fits must agree on a triple configuration. The strong dashed vertical line represents a point on Kiel diagram where the main sequence visually starts merging with the red-giant branch, the point where a region above the main sequence becomes polluted with more evolved stars.

surveys that show similar trends [243].

As the detected triple star candidates encompass a fairly small region in a colour-magnitude space, we were interested in the degree to which our analysed spectra are similar to those of other stars in the same region. For this purpose, the GALAH objects with absolute *Gaia* corrected magnitudes in the range $3.3 < M_G < 3.6$ and $0.77 < G_{BP}-G_{RP} < 0.84$ were selected and plotted in Figure 6.33, where spectra are arranged according to their similarity. In this 2D projection [details about the construction of which are given in 84, 99] it is obvious that spectra considered in this paper clearly exhibit a far greater degree of mutual similarity than other spectra with similar photometric signature. As expected, many of them lie inside the region of SB2 spectra. All of our Solar twin candidates were visually checked for the presence of a resolvable binary component. Nevertheless, this plot gives us additional proof of their absence as none of the analysed twins is located inside that region.

6.15.4 Observational bias - Galaxia model

Every magnitude limited survey, such as the GALAH, will introduce observational bias into the frequency of observed binary stars as their additional flux changes the volume of the Galaxy from where they are sampled. Their distances and occupied volume of space is located further from the Sun and therefore also greater than for comparable single stars with the same apparent magnitude and colour.

To evaluate this bias for our type of analysed stars, we created a synthetic Milky Way population of single stars using the Galaxia code [293]. First, the code was run to create a complete population of stars in the apparent V magnitude range between 10 and 16 without any colour cuts. In order for the distribution of synthesized stars to mimic the GALAH survey as closely as possible, only stars located inside the observed GALAH 2° fields were retained. After the spatial filtering, only stars with Sun-like absolute magnitude $M_V = 4.81 \pm 0.05$ and colour index $B - V = 0.63 \pm 0.05$ were kept in the set [reference magnitudes were taken from 274].

From the filtered synthetic data set we took three different sub-sets representing a pool of observable single Solar twins ($12.0 \leq V \leq 14.0$), binary twins ($12.75 \leq V \leq 14.75$), and triple twins ($13.2 \leq V \leq 15.2$) based on their apparent V magnitudes. Extinction and reddening were not used in this selection as their use did not significantly alter the relationship between the number of stars in sub-sets. Assuming that the frequency of multiple stars is constant and does not change in any of those sets, we can estimate the selection bias on the frequency of multiple stars. The number of stars in each sub-set was 15862, 35470, and 54567 respectively. According to those star counts, the derived frequency of binaries would be too high by a factor of 2.2 and a factor 3.4 for triple stars. Considering those factors, the fraction of unresolved triple candidates with Solar-like spectra is $\sim 2\%$ and $\sim 11\%$ for binary candidates.

Accidental visual binaries that lie along the same line-of-sight and have angular separation smaller than the field-of-view of the 2dF spectroscopic fiber ($2''$) or smaller than the *Gaia* end-of-mission angular resolution ($0.1''$) were not considered in this estimation.

6.16 Conclusions

Combining multiple photometric systems with spectroscopic data from the GALAH and astrometric measurements taken by *Gaia*, we showed the possible existence of triple stellar systems with long orbital periods whose combined spectrum mimics Solar spectrum. The average composition of such a system consists of three almost identical stars, where one of the stars is ~ 10 K warmer than the Sun and the coldest has an effective temperature ~ 120 K below the Sun. The derived percentage of such unresolved systems would be different for nearby/close stars as they become more/less spatially resolved. In the scope of our magnitude limited survey, we sampled only a fraction of possible distances to the systems.

Without any obvious signs of the orbital periodicity in the measured radial velocities of the systems, orbital periods were loosely constrained based on the observational limits and few assumptions. By the prior assumption that the *Gaia* spacecraft sees those systems as a single light source, we showed that they can be described by orbital periods where a difference between projected velocities of components does not sufficiently degrade an observed spectrum that it no longer recognized as Solar-like. The spectroscopic signature and radial velocity variations were further used to put a limit on the minimum orbital period of an inner pair to be at least 20 years. Shorter periods are not completely excluded as it could happen that the spectrum was acquired in a specific orbital phase where the difference between projected orbital speeds is negligible. From the fact that analysed objects are spatially unresolvable for the *Gaia* spacecraft, the orbital size of outer binary pair can extend up to 100-350 AU and therefore have orbital periods of order of a few hundred years.

To confirm their existence, detected systems are ideal candidates to be observed with precise interferometric measurements or high time-resolution photometers if they happen to be occulted by the Moon. Simulation of the lunar motion showed that four of the analysed multiple candidates lie in its path if the observations would be carried-out from the Asiago observatory that has suitable photon counting detectors.

The main drawback of the analysis was found to be its separate treatment of photometric and spectroscopic information in two independent fitting procedures. In future analyses, they should be combined to acquire even more precise results as different stellar physical parameters have a different degree of impact on these two types of measurements.

6.17 Table description and summary

In the Table 6.8 we provide a list of metadata available for every object detected using the methodology described in this paper. The complete table of detected objects and its metadata is available in electronic form at the CDS. An excerpt of the table, containing a subset of columns, for definitive and probable triple candidates is given in Table 6.9.

Table 6.8: List and description of the fields in the published catalogue of analysed objects.

Field	Unit	Description
source_id		<i>Gaia</i> DR2 source identifier
sobject_id		Unique internal per-observation star ID
ra	deg	Right ascension from <i>Gaia</i> DR2
dec	deg	Declination from <i>Gaia</i> DR2
ruwe		Value of re-normalized astrometric χ^2
m_sim_p		Photometric χ^2 for original parameters
m_sim_f		Spectroscopic χ^2 for original parameters
s1_teff1	K	$T_{\text{eff}1}$ in a fitted single system
s1_feh		[Fe/H] of a fitted single system
s1_sim_p		Photometric χ^2 of a fitted single system
s1_sim_f		Spectroscopic χ^2 of a fitted single system
s2_teff1	K	$T_{\text{eff}1}$ in a fitted binary system
s2_teff2	K	$T_{\text{eff}2}$ in a fitted binary system
s2_feh		[Fe/H] of a fitted binary system
s2_sim_p		Photometric χ^2 of a fitted binary system
s2_sim_f		Spectroscopic χ^2 of a fitted binary system
s3_teff1	K	$T_{\text{eff}1}$ in a fitted triple system
s3_teff2	K	$T_{\text{eff}2}$ in a fitted triple system
s3_teff3	K	$T_{\text{eff}3}$ in a fitted triple system
s3_feh		[Fe/H] of a fitted triple system
s3_sim_p		Photometric χ^2 of a fitted triple system
s3_sim_f		Spectroscopic χ^2 of a fitted triple system
n_stars_p		Best fitting photometric configuration
n_stars_f		Best fitting spectroscopic configuration
class		Final configuration classification
flag		Result quality flags

Table 6.9: Subset of results for definitive and probable Solar-like triple candidates detected by our selection and fitting procedure. The complete table is given as a supplementary material to this paper in a form of the textual CSV file. It is also available in the electronic form at the CDS portal.

source_id	ruwe	s2_teff1	s2_teff2	s2_feh	s3_teff1	s3_teff2	s3
6157059919188478720	11.1	6002	6000	0.24	5825	5787	
6777339306532222080	8.0	5875	5819	0.10	5642	5636	
6412815502155127808	1.1	5880	5316	-0.02	5922	4703	
6564302331580491904	1.1	5991	4701	0.13	5945	4702	
5386113598793714304	1.5	5882	5808	0.07	5878	5438	
2534579880633620992	1.2	5986	5876	0.15	5750	5739	
5484353352124904064	1.1	5876	5802	0.14	5844	5525	
5399712362903401984	1.3	5968	4703	0.08	5922	4703	
3688523450018482432	1.1	5820	5688	-0.015	5998	4717	
6220408320279116032	7.6	5923	5883	0.11	5724	5677	
6198738457927949440	0.9	6052	4798	0.15	5986	4704	
5822222074090352384	4.7	5651	5640	-0.04	5866	4702	
6355462192511955456	1.0	5700	5692	-0.11	5471	5449	
4678229218054642176	0.9	6028	4791	0.09	5968	4702	
4423111085550775680	1.6	5829	5774	-0.02	5645	5633	
6261736621608044416	4.7	5927	5890	0.07	6154	4702	
5947219160131475712	1.1	6071	5843	0.22	6007	5724	
6661888382295654656	1.5	5725	5678	-0.07	5956	4706	
6362136502970853120	1.8	5844	5826	0.013	5666	5658	
6645693508028329728	1.3	5936	4715	0.06	5833	4703	

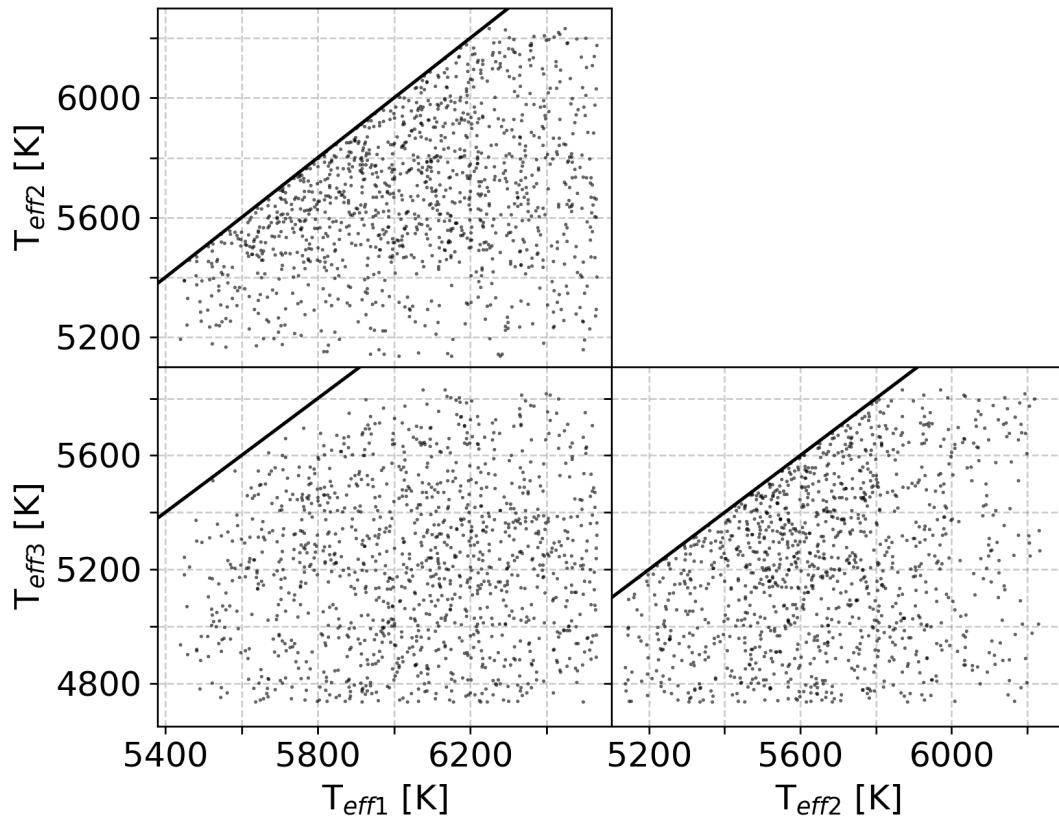


Figure 6.29: Initial distribution of walker parameters considered in the photometric fit. To ensure unique solutions with increasing T_{eff} , combinations above the linear line were not considered in the fit.

6.18 Additional figures

In order to increase the readability and transparency of the text, additional and repeated plots are supplied as appendices to the main text.

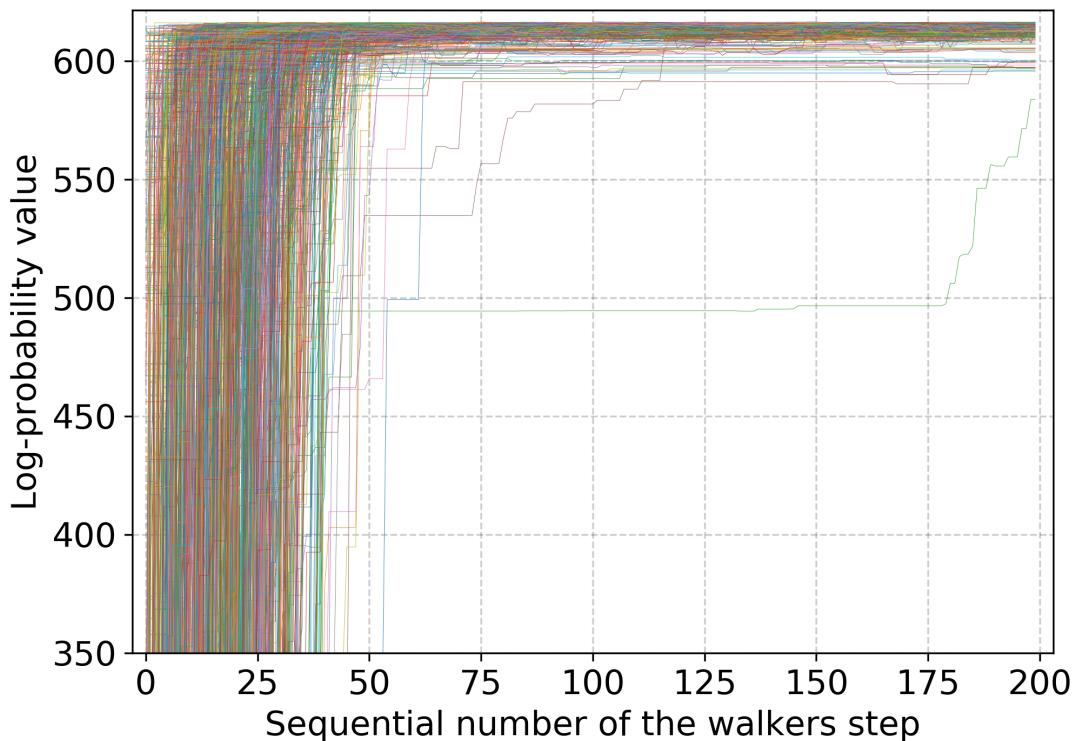


Figure 6.30: Convergence of walkers in the initial photometric fit. The plot shows a log-probability of the posteriors shown in the Figure 6.18. Walkers converge to the same value of log-probability after 50 steps. Every walker is plotted with different colour.

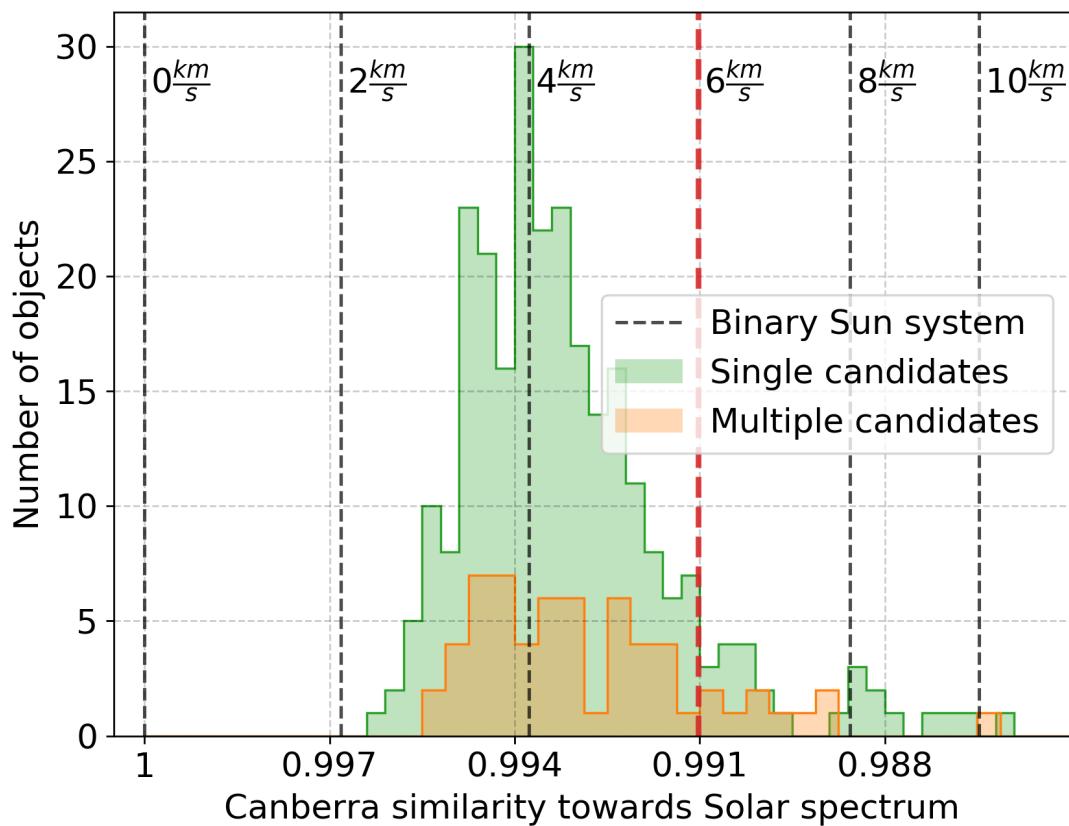


Figure 6.31: Same as Figure 6.26 but for the case of a triple system where only one component out of three has a radial velocity shift in comparison to other two.

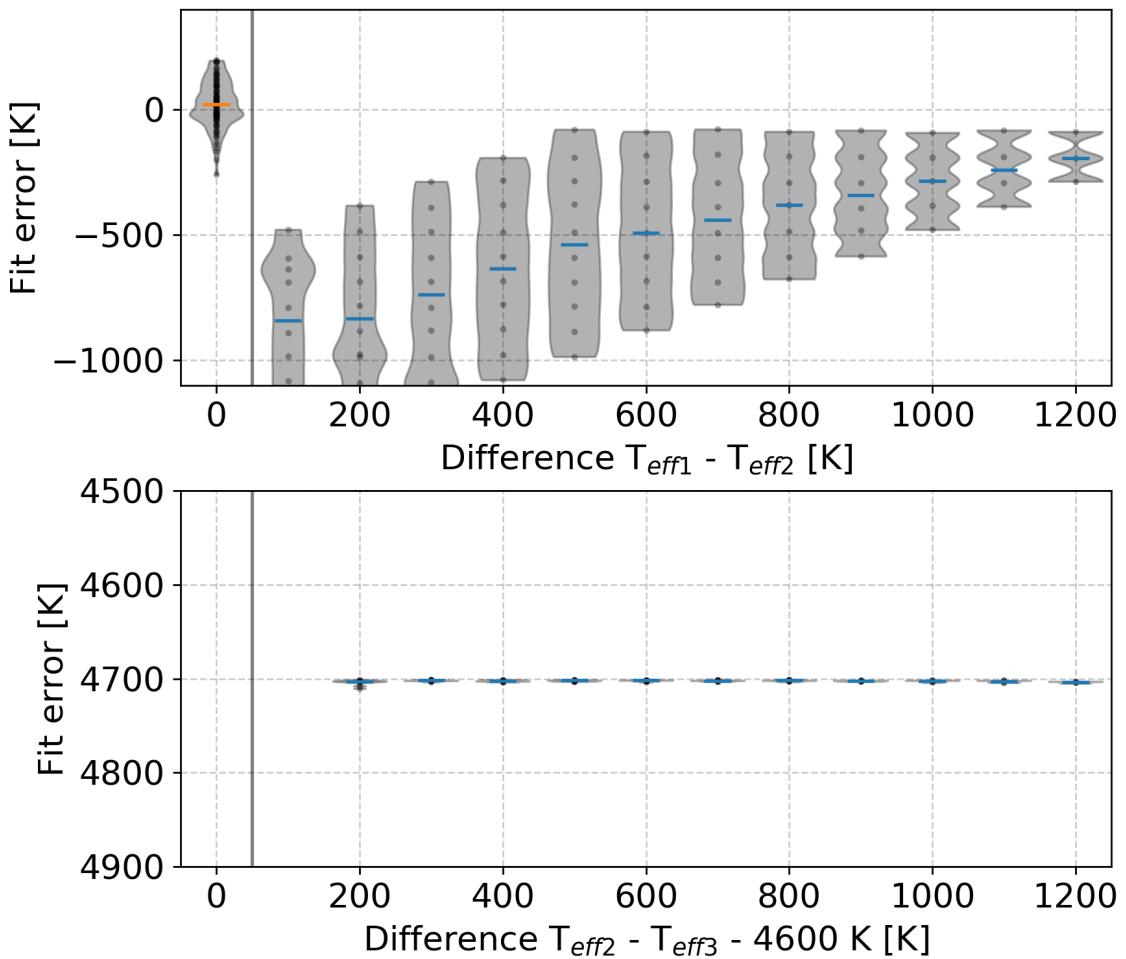


Figure 6.32: Similar to a plot in Figure 6.27, but representing a case when we try to describe a synthetic binary system with $T_{\text{eff}3} = 0$ using a triple star model.

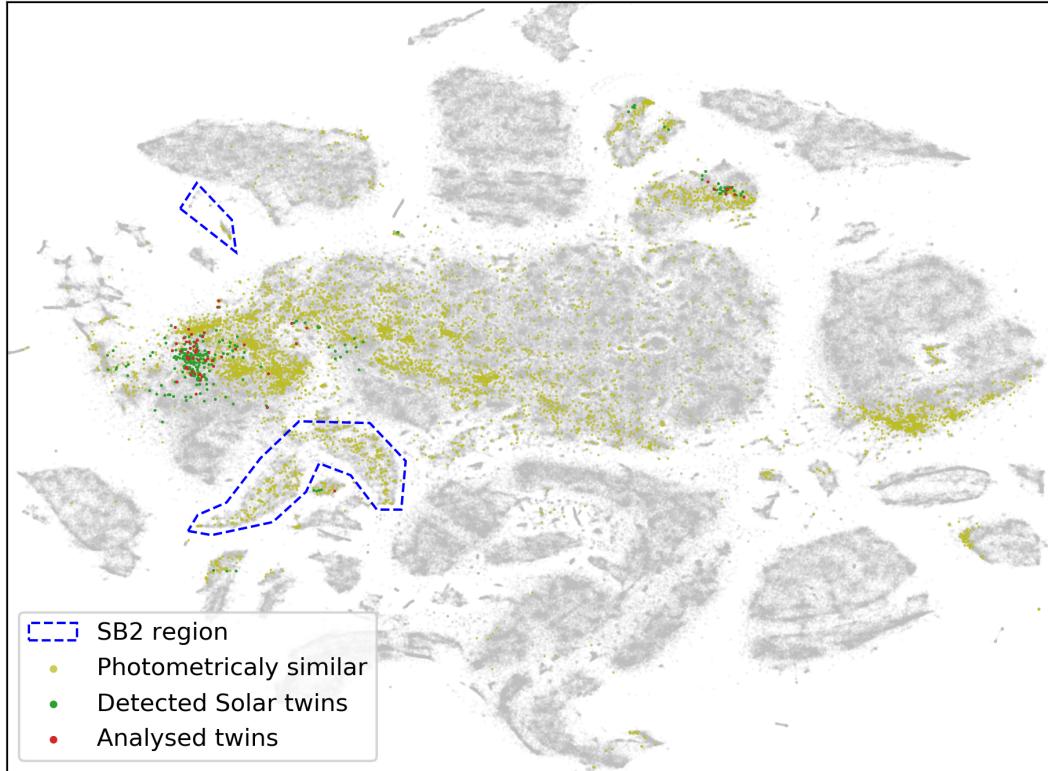


Figure 6.33: Visual representation of similarities between spectra using dimensionality reduction analysis. Clumps in this 2D projection represent morphologically similar spectra, whose features separate them from the rest of the data set. Blue-dashed polygons indicate over-densities, where SB2 spectra are located [projection and regions are taken from Figure 13 in 99]. Each grey dot represent one spectrum of GALAH survey. Green coloured dots represent objects considered in this paper, of which objects marked with red were analysed for higher order multiplicity. Yellowish dots represent objects whose *Gaia* magnitudes fall inside the range of determined triple star candidates.

Chapter 7

Conclusions

Bibliography

- [1] C. Chiappini, F. Matteucci and D. Romano, *Abundance Gradients and the Formation of the Milky Way*, **554**, 1044 (2001).
- [2] T. Naab and J. P. Ostriker, *Theoretical Challenges in Galaxy Formation*, **55**, 59 (2017).
- [3] C. J. Lada and E. A. Lada, *Embedded Clusters in Molecular Clouds*, **41**, 57 (2003).
- [4] V. V. Gvaramadze, A. Gualandris and S. Portegies Zwart, *On the origin of high-velocity runaway stars*, **396**, 570 (2009).
- [5] E. Gaburov, J. Lombardi, James C. and S. Portegies Zwart, *On the onset of runaway stellar collisions in dense star clusters - II. Hydrodynamics of three-body interactions*, **402**, 105 (2010).
- [6] T. Ryu, N. W. C. Leigh and R. Perna, *Formation of runaway stars in a star-cluster potential*, **470**, 3049 (2017).
- [7] A. Gualandris and S. Portegies Zwart, *A hypervelocity star from the Large Magellanic Cloud*, **376**, L29 (2007).
- [8] J. Kos, J. Bland-Hawthorn, K. Freeman, S. Buder, G. Traven, G. M. De Silva, S. Sharma, M. Asplund, L. Duong, J. Lin, K. Lind, S. Martell, J. D. Simpson, D. Stello, D. B. Zucker, T. Zwitter, B. Anguiano, G. Da Costa, V. D’Orazi, J. Horner, P. R. Kafle, G. Lewis, U. Munari, D. M. Nataf, M. Ness, W. Reid, K. Schlesinger, Y.-S. Ting and R. Wyse, *The GALAH survey: chemical tagging of star clusters and new members in the Pleiades*, **473**, 4612 (2018).
- [9] A. McBride and M. Kounkel, *Runaway Young Stars near the Orion Nebula*, **884**, 6 (2019).
- [10] K. Bekki, *Dynamical friction of star clusters against disc field stars in galaxies: implications on stellar nucleus formation and globular cluster luminosity functions*, **401**, 2753 (2010).
- [11] S. Röser and E. Schilbach, *Praesepe (NGC 2632) and its tidal tails*, **627**, A4 (2019).
- [12] F. C. Yeh, G. Carraro, M. Montalto and A. F. Seleznev, *Ruprecht 147: A Paradigm of Dissolving Star Cluster*, **157**, 115 (2019).

Bibliography

- [13] S. Meingast and J. Alves, *Extended stellar systems in the solar neighborhood. I. The tidal tails of the Hyades*, **621**, L3 (2019).
- [14] Y. Zhang, S.-Y. Tang, W. P. Chen, X. Pang and J. Z. Liu, *Diagnosing the Stellar Population and Tidal Structure of the Blanco 1 Star Cluster*, **889**, 99 (2020).
- [15] G. Carraro, *Photometry of dissolving star cluster candidates. The cases of NGC 7036 and NGC 7772*, **385**, 471 (2002).
- [16] C. Bonatto, E. Bica and D. B. Pavani, *NGC 2180: A disrupting open cluster?*, **427**, 485 (2004).
- [17] R. Carrera, M. Pasquato, A. Vallenari, L. Balaguer-Núñez, T. Cantat-Gaudin, M. Mapelli, A. Bragaglia, D. Bossini, C. Jordi, D. Galadí-Enríquez and E. Solano, *Extended halo of NGC 2682 (M 67) from Gaia DR2*, **627**, A119 (2019).
- [18] S. F. Portegies Zwart, P. Hut, J. Makino and S. L. W. McMillan, *On the dissolution of evolving star clusters*, **337**, 363 (1998).
- [19] G. R. I. Moyano Loyola and J. R. Hurley, *Stars on the run: escaping from stellar clusters*, **434**, 2509 (2013).
- [20] R. Wielen, *On the Lifetimes of Galactic Clusters*, **13**, 300 (1971).
- [21] R. Wielen, *Dissolution of star clusters in galaxies*, in *The Harlow-Shapley Symposium on Globular Cluster Systems in Galaxies*, IAU Symposium, Vol. 126, edited by J. E. Grindlay and A. G. D. Philip (1988) pp. 393–406.
- [22] H. Monteiro and W. S. Dias, *Distances and ages from isochrone fits of 150 open clusters using Gaia DR2 data*, **487**, 2385 (2019).
- [23] D. Bossini, A. Vallenari, A. Bragaglia, T. Cantat-Gaudin, R. Sordo, L. Balaguer-Núñez, C. Jordi, A. Moitinho, C. Soubiran, L. Casamiquela, R. Carrera and U. Heiter, *Age determination for 269 Gaia DR2 open clusters*, **623**, A108 (2019).
- [24] G. Carraro, *Open cluster remnants: an observational overview*, Bulletin of the Astronomical Society of India **34**, 153 (2006).
- [25] D. B. Pavani and E. Bica, *Characterization of open cluster remnants*, **468**, 139 (2007).
- [26] E. Bica, B. X. Santiago, C. M. Dutra, H. Dottori, M. R. de Oliveira and D. Pavani, *Dissolving star cluster candidates*, **366**, 827 (2001).
- [27] E. E. Salpeter, *The Luminosity Function and Stellar Evolution.*, **121**, 161 (1955).
- [28] J. M. Scalo, *The Stellar Initial Mass Function*, **11**, 1 (1986).
- [29] G. Chabrier, *Galactic Stellar and Substellar Initial Mass Function*, **115**, 763 (2003).

- [30] P. Kroupa, *On the variation of the initial mass function*, **322**, 231 (2001).
- [31] S. Hony, D. A. Gouliermis, F. Galliano, M. Galametz, D. Cormier, C.-H. R. Chen, S. Dib, A. Hughes, R. S. Klessen, J. Roman-Duval, L. Smith, J.-P. Bernard, C. Bot, L. Carlson, K. Gordon, R. Indebetouw, V. Lebouteiller, M.-Y. Lee, S. C. Madden, M. Meixner, J. Oliveira, M. Rubio, M. Sauvage and R. Wu, *Star formation rates from young-star counts and the structure of the ISM across the NGC 346/N66 complex in the SMC*, **448**, 1847 (2015).
- [32] D. Raboud and J. C. Mermilliod, *Evolution of mass segregation in open clusters: some observational evidences*, **333**, 897 (1998).
- [33] I. A. Bonnell and M. B. Davies, *Mass segregation in young stellar clusters*, **295**, 691 (1998).
- [34] R. de Grijs, G. F. Gilmore, R. A. Johnson and A. D. Mackey, *Mass segregation in young compact star clusters in the Large Magellanic Cloud - II. Mass functions*, **331**, 245 (2002).
- [35] C. Bonatto and E. Bica, *Mass segregation in M 67 with 2MASS*, **405**, 525 (2003).
- [36] S. Dib, S. Schmeja and R. J. Parker, *Structure and mass segregation in Galactic stellar clusters*, **473**, 849 (2018).
- [37] R. de Grijs, C. Li and A. M. Geller, *The dynamical importance of binary systems in young massive star clusters*, ArXiv e-prints (2015).
- [38] P. Kroupa, M. G. Petr and M. J. McCaughrean, *Binary stars in young clusters: models versus observations of the Trapezium Cluster*, **4**, 495 (1999).
- [39] P. J. T. Leonard, *Star counts in the open cluster NGC 2420*, **95**, 108 (1988).
- [40] S. Schmeja, N. V. Kharchenko, A. E. Piskunov, S. Röser, E. Schilbach, D. Froebrich and R.-D. Scholz, *Global survey of star clusters in the Milky Way. III. 139 new open clusters at high Galactic latitudes*, **568**, A51 (2014).
- [41] Gaia Collaboration, F. van Leeuwen, A. Vallenari, C. Jordi, L. Lindegren, U. Bastian, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, C. Babusiaux and et al., *Gaia Data Release 1. Open cluster astrometry: performance, limitations, and future prospects*, **601**, A19 (2017).
- [42] Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Eyer and et al., *Gaia Data Release 2. Summary of the contents and survey properties*, **616**, A1 (2018).
- [43] T. Cantat-Gaudin, C. Jordi, A. Vallenari, A. Bragaglia, L. Balaguer-Núñez, C. Soubiran, D. Bossini, A. Moitinho, A. Castro-Ginard, A. Krone-Martins, L. Casamiquela, R. Sordo and R. Carrera, *A Gaia DR2 view of the open cluster population in the Milky Way*, **618**, A93 (2018).

Bibliography

- [44] A. Castro-Ginard, C. Jordi, X. Luri, T. Cantat-Gaudin and L. Balaguer-Núñez, *Hunting for open clusters in Gaia DR2: the Galactic anticentre*, **627**, A35 (2019).
- [45] U. Bastian, *Gaia 8: Discovery of a star cluster containing beta Lyrae – and of a larger old (extinct) star formation complex surrounding it*, arXiv e-prints , arXiv:1909.04612 (2019).
- [46] L. Liu and X. Pang, *A Catalog of Newly Identified Star Clusters in Gaia DR2*, **245**, 32 (2019).
- [47] G. Sim, S. H. Lee, H. B. Ann and S. Kim, *207 New Open Star Clusters within 1 kpc from Gaia Data Release 2*, Journal of Korean Astronomical Society **52**, 145 (2019).
- [48] T. Cantat-Gaudin, A. Krone-Martins, N. Sedaghat, A. Farahi, R. S. de Souza, R. Skalidis, A. I. Malz, S. Macêdo, B. Moews, C. Jordi, A. Moitinho, A. Castro-Ginard, E. E. O. Ishida, C. Heneka, A. Boucaud and A. M. M. Trindade, *Gaia DR2 unravels incompleteness of nearby cluster population: new open clusters in the direction of Perseus*, **624**, A126 (2019).
- [49] A. Castro-Ginard, C. Jordi, X. Luri, J. Álvarez Cid-Fuentes, L. Casamiquela, F. Anders, T. Cantat-Gaudin, M. Monguió, L. Balaguer-Núñez, S. Solà and R. M. Badia, *Hunting for open clusters in {Gaia} DR2: 582 new OCs in the Galactic disc*, arXiv e-prints , arXiv:2001.07122 (2020).
- [50] H. Baumgardt, *The nature of some doubtful open clusters as revealed by HIP-PARCOS*, **340**, 402 (1998).
- [51] G. Carraro, *NGC 6994: An open cluster which is not an open cluster*, **357**, 145 (2000).
- [52] E. Han, J. L. Curtis and J. T. Wright, *The Putative Old, Nearby Cluster Lodén 1 Does Not Exist*, **152**, 7 (2016).
- [53] J. Kos, G. de Silva, S. Buder, J. Bland -Hawthorn, S. Sharma, M. Asplund, V. D’Orazi, L. Duong, K. Freeman, G. F. Lewis, J. Lin, K. Lind, S. L. Martell, K. J. Schlesinger, J. D. Simpson, D. B. Zucker, T. Zwitter, T. R. Bedding, K. Čotar, J. Horner, T. Nordlander, D. Stello, Y.-S. Ting and G. Traven, *The GALAH survey and Gaia DR2: (non-)existence of five sparse high-latitude open clusters*, **480**, 5242 (2018).
- [54] T. Cantat-Gaudin and F. Anders, *Clusters and mirages: cataloguing stellar aggregates in the Milky Way*, **633**, A99 (2020).
- [55] J. Choi, C. Conroy, Y.-S. Ting, P. A. Cargile, A. Dotter and B. D. Johnson, *Star Cluster Ages in the Gaia Era*, **863**, 65 (2018).
- [56] D. R. Soderblom, *The Ages of Stars*, **48**, 581 (2010).
- [57] J. Kos, J. Bland-Hawthorn, M. Asplund, S. Buder, G. F. Lewis, J. Lin, S. L. Martell, M. K. Ness, S. Sharma, G. M. De Silva, J. D. Simpson, D. B. Zucker, T. Zwitter, K. Čotar and L. Spina, *Discovery of a 21 Myr old stellar population in the Orion complex**, **631**, A166 (2019).

- [58] E. Terlevich, *Evolution of n-body open clusters*, **224**, 193 (1987).
- [59] A. F. Seleznev, *Open-cluster density profiles derived using a kernel estimator*, **456**, 3757 (2016).
- [60] D. F. Evans, *Evidence for Unresolved Exoplanet-hosting Binaries in Gaia DR2*, Research Notes of the American Astronomical Society **2**, 20 (2018).
- [61] D. Birko, T. Zwitter, E. K. Grebel, Q. A. Parker, G. Kordopatis, J. Bland - Hawthorn, K. Freeman, G. Guiglion, B. K. Gibson, J. Navarro, W. Reid, G. M. Seabroke, M. Steinmetz and F. Watson, *Single-lined Spectroscopic Binary Star Candidates from a Combination of the RAVE and Gaia DR2 Surveys*, **158**, 155 (2019).
- [62] A. Widmark, B. Leistedt and D. W. Hogg, *Inferring Binary and Trinary Stellar Populations in Photometric and Astrometric Surveys*, **857**, 114 (2018).
- [63] Gaia Collaboration, C. Babusiaux, F. van Leeuwen, M. A. Barstow, C. Jordi, A. Vallenari, D. Bossini, A. Bressan, T. Cantat-Gaudin, M. van Leeuwen, A. G. A. Brown, T. Prusti, J. H. J. de Bruijne, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Eyer, F. Jansen, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, H. I. Siddiqui, C. Soubiran, N. A. Walton, F. Arenou, U. Bastian, M. Cropper, R. Drimmel, D. Katz, M. G. Lattanzi, J. Bakker, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, F. De Angeli, C. Fabricius, R. Guerra, B. Holl, E. Masana, R. Messineo, N. Mowlavi, K. Nienartowicz, P. Panuzzo, J. Portell, M. Riello, G. M. Seabroke, P. Tanga, F. Thévenin, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. Burgess, G. Busso, B. Carry, A. Cellino, G. Clementini, M. Clotet, O. Creevey, M. Davidson, J. De Ridder, L. Delchambre, A. Dell’Oro, C. Ducourant, J. Fernández-Hernández, M. Fouesneau, Y. Frémat, L. Galluccio, M. García-Torres, J. González-Núñez, J. J. González-Vidal, E. Gosset, L. P. Guy, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, A. Hutton, G. Jasniewicz, A. Jean-Antoine-Piccolo, S. Jordan, A. J. Korn, A. Krone-Martins, A. C. Lanzafame, T. Lebzelter, W. Löffler, M. Manteiga, P. M. Marrese, J. M. Martín-Fleitas, A. Moitinho, A. Mora, K. Muinonen, J. Osinde, E. Pancino, T. Pauwels, J. M. Petit, A. Recio-Blanco, P. J. Richards, L. Rimoldini, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, M. Süveges, J. Torra, W. van Reeven, U. Abbas, A. Abreu Aramburu, S. Accart, C. Aerts, G. Altavilla, M. A. Álvarez, R. Alvarez, J. Alves, R. I. Anderson, A. H. Andrei, E. Anglada Varela, E. Antiche, T. Antoja, B. Arcay, T. L. Astraatmadja, N. Bach, S. G. Baker, L. Balaguer-Núñez, P. Balm, C. Barache, C. Barata, D. Barbato, F. Barblan, P. S. Barklem, D. Barrado, M. Barros, L. Bartholomé Muñoz, J. L. Bassilana, U. Becciani, M. Bellazzini, A. Berihuete, S. Bertone, L. Bianchi, O. Bienaymé, S. Blanco-Cuaresma, T. Boch, C. Boeche, A. Bombrun, R. Borrachero, S. Bouquillon, G. Bourda, A. Bragaglia, L. Bramante, M. A. Breddels, N. Brouillet, T. Brüsemeister, E. Brugaletta, B. Bucciarelli, A. Burlacu, D. Busonero, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, G. Cannizzaro, R. Carballo, T. Carlucci,

Bibliography

J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, P. Charlot, L. Chemin, A. Chiavassa, G. Cocozza, G. Costigan, S. Cowell, F. Crifo, M. Crosta, C. Crowley, J. Cuypers, C. Dafonte, Y. Damerdji, A. Dapergolas, P. David, M. David, P. de Laverny, F. De Luise, R. De March, D. de Martino, R. de Souza, A. de Torres, J. Debosscher, E. del Pozo, M. Delbo, A. Delgado, H. E. Delgado, S. Diakite, C. Diener, E. Distefano, C. Dolding, P. Drazinos, J. Durán, B. Edvardsson, H. Enke, K. Eriksson, P. Esquej, G. Eynard Bontemps, C. Fabre, M. Fabrizio, S. Faigler, A. J. Falcão, M. Farràs Casas, L. Federici, G. Fedorets, P. Fernique, F. Figueras, F. Filippi, K. Findeisen, A. Fonti, E. Fraile, M. Fraser, B. Frézouls, M. Gai, S. Galli, D. Garabato, F. García-Sedano, A. Garofalo, N. Garralda, A. Gavel, P. Gavras, J. Gerssen, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, F. Glass, M. Gomes, M. Granvik, A. Gueguen, A. Guerrier, J. Guiraud, R. Gutié, R. Haigron, D. Hatzidimitriou, M. Hauser, M. Haywood, U. Heiter, A. Helmi, J. Heu, T. Hilger, D. Hobbs, W. Hofmann, G. Holland, H. E. Huckle, A. Hypki, V. Icardi, K. Janßen, G. Jevardat de Fombelle, P. G. Jonker, Á. L. Juhász, F. Julbe, A. Karampelas, A. Kewley, J. Klar, A. Kochoska, R. Kohley, K. Kolenberg, M. Kontizas, E. Kontizas, S. E. Koposov, G. Kordopatis, Z. Kostrzewa-Rutkowska, P. Koubsky, S. Lambert, A. F. Lanza, Y. Lasne, J. B. Lavigne, Y. Le Fustec, C. Le Poncin-Lafitte, Y. Lebreton, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, H. Lenhardt, F. Leroux, S. Liao, E. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, M. López, S. Mangau, R. G. Mann, G. Mantelet, O. Marchal, J. M. Marchant, M. Marconi, S. Marinoni, G. Marschalkó, D. J. Marshall, M. Martino, G. Marton, N. Mary, D. Massari, G. Matijević, T. Mazeh, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, D. Molina, R. Molinaro, L. Molnár, P. Montegriffo, R. Mor, R. Morbidelli, T. Morel, D. Morris, A. F. Mulone, T. Muraveva, I. Musella, G. Nelemans, L. Nicastro, L. Noval, W. O'Mullane, C. Ordénovic, D. Ordóñez-Blanco, P. Osborne, C. Pagani, I. Pagano, F. Pailler, H. Palacin, L. Palaversa, A. Panahi, M. Pawlak, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, E. Poujoulet, A. Prša, L. Pulone, E. Racero, S. Ragaini, N. Rambaux, M. Ramos-Lerate, S. Regibo, C. Reylé, F. Riclet, V. Ripepi, A. Riva, A. Rivard, G. Rixon, T. Roegiers, M. Roelens, M. Romero-Gómez, N. Rowell, F. Royer, L. Ruiz-Dern, G. Sadowski, T. Sagristà Sellés, J. Sahlmann, J. Salgado, E. Salguero, N. Sanna, T. Santana-Ros, M. Sarasso, H. Savietto, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, I. C. Shih, L. Siltala, A. F. Silva, R. L. Smart, K. W. Smith, E. Solano, F. Solitro, R. Sordo, S. Soria Nieto, J. Souchay, A. Spagna, F. Spoto, U. Stampa, I. A. Steele, H. Steidelmüller, C. A. Stephenson, H. Stoev, F. F. Suess, J. Surdej, L. Szabados, E. Szegedi-Elek, D. Tapiador, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, D. Terrett, P. Teyssandier, W. Thuillot, A. Titarenko, F. Torra Clotet, C. Turon, A. Ulla, E. Utrilla, S. Uzzi, M. Vaillant, G. Valentini, V. Valette, A. van Elteren, E. Van Hemelryck, M. Vaschetto, A. Vecchiato, J. Veljanoski, Y. Viala, D. Vicente, S. Vogt, C. von Essen, H. Voss, V. Votruba, S. Voutsinas, G. Walmsley, M. Weiler, O. Wertz, T. Wevers, Ł. Wyrzykowski, A. Yoldas, M. Žerjal, H. Ziaeepour, J. Zorec, S. Zschocke, S. Zucker, C. Zurbach and T. Zwitter, *Gaia Data Release 2. Observational Hertzsprung-Russell diagrams*, **616**, A10 (2018).

- [64] K. Čotar, T. Zwitter, G. Traven, J. Kos, M. Asplund, J. Bland-Hawthorn, S. Buder, V. D’Orazi, G. M. de Silva, J. Lin, S. L. Martell, S. Sharma, J. D. Simpson, D. B. Zucker, J. Horner, G. F. Lewis, T. Nordlander, Y.-S. Ting, R. A. Wittenmyer and Galah Collaboration, *The GALAH survey: unresolved triple Sun-like stars discovered by the Gaia mission*, **487**, 2474 (2019).
- [65] V. Belokurov, Z. Penoyre, S. Oh, G. Iorio, S. Hodgkin, N. W. Evans, A. Everall, S. E. Koposov, C. A. Tout, R. Izzard, C. J. Clarke and A. G. A. Brown, *Unresolved stellar companions with Gaia DR2 astrometry*, arXiv e-prints , arXiv:2003.05467 (2020).
- [66] K. Freeman and J. Bland-Hawthorn, *The New Galaxy: Signatures of Its Formation*, **40**, 487 (2002).
- [67] J. Bland-Hawthorn, T. Karlsson, S. Sharma, M. Krumholz and J. Silk, *The Chemical Signatures of the First Star Clusters in the Universe*, **721**, 582 (2010).
- [68] D. W. Hogg, A. R. Casey, M. Ness, H.-W. Rix, D. Foreman-Mackey, S. Hasselquist, A. Y. Q. Ho, J. A. Holtzman, S. R. Majewski, S. L. Martell, S. Mészáros, D. L. Nidever and M. Shetrone, *Chemical Tagging Can Work: Identification of Stellar Phase-space Structures Purely by Chemical-abundance Similarity*, **833**, 262 (2016).
- [69] R. Garcia-Dias, C. Allende Prieto, J. Sánchez Almeida and P. Alonso Palicio, *Machine learning in APOGEE. Identification of stellar populations through chemical abundances*, **629**, A34 (2019).
- [70] F. Anders, C. Chiappini, B. X. Santiago, G. Matijević, A. B. Queiroz, M. Steinmetz and G. Guiglion, *Dissecting stellar chemical abundance space with t-SNE*, **619**, A125 (2018).
- [71] P. Jofré, U. Heiter and C. Soubiran, *Accuracy and Precision of Industrial Stellar Abundances*, **57**, 571 (2019).
- [72] J. Bovy, *The Chemical Homogeneity of Open Clusters*, **817**, 49 (2016).
- [73] S. Blanco-Cuaresma, C. Soubiran, U. Heiter, M. Asplund, G. Carraro, M. T. Costado, S. Feltzing, J. I. González-Hernández, F. Jiménez-Esteban, A. J. Korn, A. F. Marino, D. Montes, I. San Roman, H. M. Tabernero and G. Tautvaišienė, *Testing the chemical tagging technique with open clusters*, **577**, A47 (2015).
- [74] A. Dotter, C. Conroy, P. Cargile and M. Asplund, *The Influence of Atomic Diffusion on Stellar Ages and Chemical Tagging*, **840**, 99 (2017).
- [75] C. Bertelli Motta, A. Pasquali, J. Richer, G. Michaud, M. Salaris, A. Bragaglia, L. Magrini, S. Randich, E. K. Grebel, V. Adibekyan, S. Blanco-Cuaresma, A. Drazdauskas, X. Fu, S. Martell, G. Tautvaišienė, G. Gilmore, E. J. Alfaro, T. Bensby, E. Flaccomio, S. E. Koposov, A. J. Korn, A. C. Lanzafame, R. Smiljanic, A. Bayo, G. Carraro, A. R. Casey, M. T. Costado, F. Damiani, E. Franciosini, U. Heiter, A. Hourihane, P. Jofré, C. Lardo,

Bibliography

- J. Lewis, L. Monaco, L. Morbidelli, G. G. Sacco, S. G. Sousa, C. C. Worley and S. Zaggia, *The Gaia-ESO Survey: evidence of atomic diffusion in M67?*, **478**, 425 (2018).
- [76] L. Casamiquela, Y. Tarricq, C. Soubiran, S. Blanco-Cuaresma, P. Jofré, U. Heiter and M. Tucci Maia, *Differential abundances of open clusters and their tidal tails: chemical tagging and chemical homogeneity*, arXiv e-prints , arXiv:1912.08539 (2019).
- [77] M. Baratella, V. D’Orazi, G. Carraro, S. Desidera, S. Randich, L. Magrini, V. Adibekyan, R. Smiljanic, L. Spina, M. Tsantaki, G. Tautvaisiene, S. G. Sousa, P. Jofré, F. M. Jiménes-Esteban, E. Delgado-Mena, S. Martell, M. Van der Swaelmen, V. Roccatagliata, G. Gilmore, E. J. Alfaro, A. Bayo, T. Bensby, A. Bragaglia, E. Franciosini, A. Gonreau, U. Heiter, A. Hourihane, R. D. Jeffries, S. E. Koposov, L. Morbidelli, L. Prisinzano, G. Sacco, L. Sbordone, C. Worley, S. Zaggia and J. Lewis, *The Gaia-ESO Survey: a new approach to chemically characterising young open clusters*, arXiv e-prints , arXiv:2001.03179 (2020).
- [78] A. R. Casey, J. C. Lattanzio, A. Aleti, D. L. Dowe, J. Bland-Hawthorn, S. Buder, G. F. Lewis, S. L. Martell, T. Nordlander, J. D. Simpson, S. Sharma and D. B. Zucker, *A Data-driven Model of Nucleosynthesis with Chemical Tagging in a Lower-dimensional Latent Space*, **887**, 73 (2019).
- [79] A. H. W. Küpper, P. Kroupa, H. Baumgardt and D. C. Heggie, *Peculiarities in velocity dispersion and surface density profiles of star clusters*, **407**, 2241 (2010).
- [80] M. E. K. Williams, M. Steinmetz, S. Sharma, J. Bland-Hawthorn, R. S. de Jong, G. M. Seabroke, A. Helmi, K. C. Freeman, J. Binney, I. Minchev, O. Bienaymé, R. Campbell, J. P. Fulbright, B. K. Gibson, G. F. Gilmore, E. K. Grebel, U. Munari, J. F. Navarro, Q. A. Parker, W. Reid, A. Siebert, A. Siviero, F. G. Watson, R. F. G. Wyse and T. Zwitter, *The Dawning of the Stream of Aquarius in RAVE*, **728**, 102 (2011).
- [81] I. Carrillo, I. Minchev, G. Kordopatis, M. Steinmetz, J. Binney, F. Anders, O. Bienaymé, J. Bland-Hawthorn, B. Famaey, K. C. Freeman, G. Gilmore, B. K. Gibson, E. K. Grebel, A. Helmi, A. Just, A. Kunder, P. McMillan, G. Monari, U. Munari, J. Navarro, Q. A. Parker, W. Reid, G. Seabroke, S. Sharma, A. Siebert, F. Watson, R. F. G. Wyse and T. Zwitter, *Is the Milky Way still breathing? RAVE-Gaia streaming motions*, ArXiv e-prints (2017).
- [82] G. W. Preston, *The chemically peculiar stars of the upper main sequence*, **12**, 257 (1974).
- [83] L. Tomasella, U. Munari and T. Zwitter, *A High-resolution, Multi-epoch Spectral Atlas of Peculiar Stars Including RAVE, GAIA , and HERMES Wavelength Ranges*, **140**, 1758 (2010).
- [84] G. Traven, G. Matijević, T. Zwitter, M. Žerjal, J. Kos, M. Asplund, J. Bland-Hawthorn, A. R. Casey, G. De Silva, K. Freeman, J. Lin, S. L. Martell, K. J.

- Schlesinger, S. Sharma, J. D. Simpson, D. B. Zucker, B. Anguiano, G. Da Costa, L. Duong, J. Horner, E. A. Hyde, P. R. Kafle, U. Munari, D. Nataf, C. A. Navin, W. Reid and Y.-S. Ting, *The Galah Survey: Classification and Diagnostics with t-SNE Reduction of Spectral Information*, **228**, 24 (2017).
- [85] G. Matijević, T. Zwitter, U. Munari, O. Bienaymé, J. Binney, J. Bland-Hawthorn, C. Boeche, R. Campbell, K. C. Freeman, B. Gibson, G. Gilmore, E. K. Grebel, A. Helmi, J. F. Navarro, Q. A. Parker, G. M. Seabroke, A. Siebert, A. Siviero, M. Steinmetz, F. G. Watson, M. Williams and R. F. G. Wyse, *Double-lined Spectroscopic Binary Stars in the Radial Velocity Experiment Survey*, **140**, 184 (2010).
- [86] K. El-Badry, H.-W. Rix, Y.-S. Ting, D. R. Weisz, M. Bergemann, P. Cargile, C. Conroy and A.-C. Eilers, *Signatures of unresolved binaries in stellar spectra: implications for spectral fitting*, **473**, 5043 (2018).
- [87] K. El-Badry, Y.-S. Ting, H.-W. Rix, E. Quataert, D. R. Weisz, P. Cargile, C. Conroy, D. W. Hogg, M. Bergemann and C. Liu, *Discovery and characterization of 3000+ main-sequence binaries from APOGEE spectra*, **476**, 528 (2018).
- [88] B. Yanny, C. Rockosi, H. J. Newberg, G. R. Knapp, J. K. Adelman-McCarthy, B. Alcorn, S. Allam, C. Allende Prieto, D. An, K. S. J. Anderson, S. Anderson, C. A. L. Bailer-Jones, S. Bastian, T. C. Beers, E. Bell, V. Belokurov, D. Bizyaev, N. Blythe, J. J. Bochanski, W. N. Boroski, J. Brinchmann, J. Brinkmann, H. Brewington, L. Carey, K. M. Cudworth, M. Evans, N. W. Evans, E. Gates, B. T. Gänsicke, B. Gillespie, G. Gilmore, A. Nebot Gomez-Moran, E. K. Grebel, J. Greenwell, J. E. Gunn, C. Jordan, W. Jordan, P. Harding, H. Harris, J. S. Hendry, D. Holder, I. I. Ivans, Ž. Ivezić, S. Jester, J. A. Johnson, S. M. Kent, S. Kleinman, A. Kniazev, J. Krzesinski, R. Kron, N. Kuropatkin, S. Lebedeva, Y. S. Lee, R. French Leger, S. Lépine, S. Levine, H. Lin, D. C. Long, C. Loomis, R. Lupton, O. Malanushenko, V. Malanushenko, B. Margon, D. Martinez-Delgado, P. McGehee, D. Monet, H. L. Morrison, J. A. Munn, E. H. Neilsen, Jr., A. Nitta, J. E. Norris, D. Oravetz, R. Owen, N. Padmanabhan, K. Pan, R. S. Peterson, J. R. Pier, J. Platson, P. Re Fiorentin, G. T. Richards, H.-W. Rix, D. J. Schlegel, D. P. Schneider, M. R. Schreiber, A. Schwope, V. Sibley, A. Simmons, S. A. Snedden, J. Allyn Smith, L. Stark, F. Stauffer, M. Steinmetz, C. Stoughton, M. SubbaRao, A. Szalay, P. Szkody, A. R. Thakar, T. Sivarani, D. Tucker, A. Uomoto, D. Vanden Berk, S. Vidrih, Y. Wadadekar, S. Watters, R. Wilhelm, R. F. G. Wyse, J. Yarger and D. Zucker, *SEGUE: A Spectroscopic Survey of 240,000 Stars with $g = 14\text{--}20$* , **137**, 4377-4399 (2009).
- [89] S. R. Majewski, R. P. Schiavon, P. M. Frinchaboy, C. Allende Prieto, R. Barkhouser, D. Bizyaev, B. Blank, S. Brunner, A. Burton, R. Carrera, S. D. Chojnowski, K. Cunha, C. Epstein, G. Fitzgerald, A. E. García Pérez, F. R. Hearty, C. Henderson, J. A. Holtzman, J. A. Johnson, C. R. Lam, J. E. Lawler, P. Maseman, S. Mészáros, M. Nelson, D. C. Nguyen, D. L. Nidever, M. Pinsonneault, M. Shetrone, S. Smee, V. V. Smith, T. Stolberg, M. F. Skrutskie, E. Walker, J. C. Wilson, G. Zasowski, F. Anders, S. Basu, S. Beland, M. R.

Bibliography

- Blanton, J. Bovy, J. R. Brownstein, J. Carlberg, W. Chaplin, C. Chiappini, D. J. Eisenstein, Y. Elsworth, D. Feuillet, S. W. Fleming, J. Galbraith-Frew, R. A. García, D. A. García-Hernández, B. A. Gillespie, L. Girardi, J. E. Gunn, S. Hasselquist, M. R. Hayden, S. Hekker, I. Ivans, K. Kinemuchi, M. Klaene, S. Mahadevan, S. Mathur, B. Mosser, D. Muna, J. A. Munn, R. C. Nichol, R. W. O'Connell, J. K. Parejko, A. C. Robin, H. Rocha-Pinto, M. Schultheis, A. M. Serenelli, N. Shane, V. Silva Aguirre, J. S. Sobeck, B. Thompson, N. W. Troup, D. H. Weinberg and O. Zamora, *The Apache Point Observatory Galactic Evolution Experiment (APOGEE)*, **154**, 94 (2017).
- [90] A. Kunder, G. Kordopatis, M. Steinmetz, T. Zwitter, P. J. McMillan, L. Casagrande, H. Enke, J. Wojno, M. Valentini, C. Chiappini, G. Matijević, A. Siviero, P. de Laverny, A. Recio-Blanco, A. Bijaoui, R. F. G. Wyse, J. Binney, E. K. Grebel, A. Helmi, P. Jofre, T. Antoja, G. Gilmore, A. Siebert, B. Famaey, O. Bienaymé, B. K. Gibson, K. C. Freeman, J. F. Navarro, U. Munari, G. Seabroke, B. Anguiano, M. Žerjal, I. Minchev, W. Reid, J. Bland-Hawthorn, J. Kos, S. Sharma, F. Watson, Q. A. Parker, R.-D. Scholz, D. Burton, P. Cass, M. Hartley, K. Fiegert, M. Stupar, A. Ritter, K. Hawkins, O. Gerhard, W. J. Chaplin, G. R. Davies, Y. P. Elsworth, M. N. Lund, A. Miglio and B. Mosser, *The Radial Velocity Experiment (RAVE): Fifth Data Release*, **153**, 75 (2017).
- [91] G. Gilmore, S. Randich, M. Asplund, J. Binney, P. Bonifacio, J. Drew, S. Feltzing, A. Ferguson, R. Jeffries, G. Micela and et al., *The Gaia-ESO Public Spectroscopic Survey*, *The Messenger* **147**, 25 (2012).
- [92] S. L. Martell, S. Sharma, S. Buder, L. Duong, K. J. Schlesinger, J. Simpson, K. Lind, M. Ness, J. P. Marshall, M. Asplund, J. Bland-Hawthorn, A. R. Casey, G. De Silva, K. C. Freeman, J. Kos, J. Lin, D. B. Zucker, T. Zwitter, B. Anguiano, C. Bacigalupo, D. Carollo, L. Casagrande, G. S. Da Costa, J. Horner, D. Huber, E. A. Hyde, P. R. Kafle, G. F. Lewis, D. Nataf, C. A. Navin, D. Stello, C. G. Tinney, F. G. Watson and R. Wittenmyer, *The GALAH survey: observational overview and Gaia DR1 companion*, **465**, 3203 (2017).
- [93] R. S. de Jong, O. Bellido-Tirado, C. Chiappini, É. Depagne, R. Haynes, D. Johl, O. Schnurr, A. Schwabe, J. Walcher, F. Dionies, D. Haynes, A. Kelz, F. S. Kitaura, G. Lamer, I. Minchev, V. Müller, S. E. Nuza, J.-C. Olaya, T. Piffl, E. Popow, M. Steinmetz, U. Ural, M. Williams, R. Winkler, L. Wisotzki, W. R. Ansorge, M. Banerji, E. Gonzalez Solares, M. Irwin, R. C. Kennicutt, D. King, R. G. McMahon, S. Koposov, I. R. Parry, D. Sun, N. A. Walton, G. Finger, O. Iwert, M. Krumpe, J.-L. Lizon, M. Vincenzo, J.-P. Amans, P. Bonifacio, M. Cohen, P. Francois, P. Jagourel, S. B. Mignot, F. Royer, P. Sartoretti, R. Bender, F. Grupp, H.-J. Hess, F. Lang-Bardl, B. Muschielok, H. Böhringer, T. Boller, A. Bongiorno, M. Brusa, T. Dwelly, A. Merloni, K. Nandra, M. Salvato, J. H. Pragt, R. Navarro, G. Gerlofsma, R. Roelfsema, G. B. Dalton, K. F. Middleton, I. A. Tosh, C. Boeche, E. Cafau, N. Christlieb, E. K. Grebel, C. Hansen, A. Koch, H.-G. Ludwig, A. Quirrenbach, L. Sbordone, W. Seifert, G. Thimm, T. Trifonov, A. Helmi, S. C. Trager, S. Feltzing, A. Korn and W. Boland, *4MOST: 4-metre multi-object*

- spectroscopic telescope*, in , Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446 (2012) p. 84460T.
- [94] G. Dalton, S. C. Trager, D. C. Abrams, D. Carter, P. Bonifacio, J. A. L. Aguerri, M. MacIntosh, C. Evans, I. Lewis, R. Navarro, T. Agochs, K. Dee, S. Rousset, I. Tosh, K. Middleton, J. Pragt, D. Terrett, M. Brock, C. Benn, M. Verheijen, D. Cano Infantes, C. Bevil, I. Steele, C. Mottram, S. Bates, F. J. Gribbin, J. Rey, L. F. Rodriguez, J. M. Delgado, I. Guinouard, N. Walton, M. J. Irwin, P. Jagourel, R. Stuik, G. Gerlofsma, R. Roelfsma, I. Skillen, A. Ridings, M. Balcells, J.-B. Daban, C. Gouvret, L. Venema and P. Girard, *WEAVE: the next generation wide-field spectroscopy facility for the William Herschel Telescope*, in , Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446 (2012) p. 84460P.
- [95] J. Kos, J. Lin, T. Zwitter, M. Žerjal, S. Sharma, J. Bland-Hawthorn, M. Asplund, A. R. Casey, G. M. De Silva, K. C. Freeman, S. L. Martell, J. D. Simpson, K. J. Schlesinger, D. Zucker, B. Anguiano, C. Bacigalupo, T. R. Bedding, C. Betters, G. Da Costa, L. Duong, E. Hyde, M. Ireland, P. R. Kafle, S. Leon-Saval, G. F. Lewis, U. Munari, D. Nataf, D. Stello, C. G. Tinney, G. Traven, F. Watson and R. A. Wittenmyer, *The GALAH survey: the data reduction pipeline*, **464**, 1259 (2017).
- [96] R. Ahumada, C. Allende Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, R. Arcodia, E. Armengaud, M. Aubert, S. Avila, V. Avila-Reese, C. Badenes, C. Balland, K. Barger, J. K. Barrera-Ballesteros, S. Basu, J. Bautista, R. L. Beaton, T. C. Beers, B. I. T. Benavides, C. F. Bender, M. Bernardi, M. Bershadsky, F. Beutler, C. Moni Bidin, J. Bird, D. Bizyaev, G. A. Blanc, M. R. Blanton, M. Boquien, J. Borissova, J. Bovy, W. N. Brandt, J. Brinkmann, J. R. Brownstein, K. Bundy, M. Bureau, A. Burgasser, E. Burtin, M. Cano-Diaz, R. Capasso, M. Cappellari, R. Carrera, S. Chabrier, W. Chaplin, M. Chapman, B. Cherinka, C. Chiappini, P. D. Choi, S. D. Chojnowski, H. Chung, N. Clerc, D. Coffey, J. M. Comerford, J. Comparat, L. da Costa, M.-C. Cousinou, K. Covey, J. D. Crane, K. Cunha, G. da Silva Ilha, Y. S. Dai, S. B. Damsted, J. Darling, D. Horta Darrington, J. Davidson, James W., R. Davies, K. Dawson, N. De, A. de la Macorra, N. De Lee, A. B. d. A. Queiroz, A. Deconto Machado, S. de la Torre, F. Dell'Agli, H. du Mas des Bourboux, A. M. Diamond-Stanic, S. Dillon, J. Donor, N. Drory, C. Duckworth, T. Dwelly, G. Ebelke, S. Eftekharzadeh, A. Davis Eigenbrot, Y. P. Elsworth, M. Eracleous, G. Erfanianfar, S. Escoffier, X. Fan, E. Farr, J. G. Fernandez-Trincado, D. Feuillet, A. Finoguenov, P. Fofie, A. Fraser-Mckelvie, P. M. Frinchaboy, S. Fromenteau, H. Fu, L. Galbany, R. A. Garcia, D. A. Garcia-Hernandez, L. A. Garma Oehmichen, J. Ge, M. A. Geimba Maia, D. Geisler, J. Gelfand, J. Goddy, J.-M. Le Goff, V. Gonzalez-Perez, K. Grabowski, P. Green, C. J. Grier, H. Guo, J. Guy, P. Harding, S. Hasselquist, A. J. Hawken, C. R. Hayes, F. Hearty, S. Hekker, D. W. Hogg, J. Holtzman, J. Hou, B.-C. Hsieh, D. Huber, J. A. S. Hunt, J. Ider Chitham, J. Imig, M. Jaber, C. E. Jimenez Angel, J. A. Johnson, A. M. Jones, H. Jansson, E. Jullo, Y. Kim, K. Kinemuchi, I. Kirkpatrick, Charles C., G. W. Kite, M. Klaene, J.-P. Kneib, J. A. Kollmeier, H. Kong, M. Kounkel, D. Krish-

Bibliography

- narao, I. Lacerna, T.-W. Lan, R. R. Lane, D. R. Law, H. W. Leung, H. Lewis, C. Li, J. Lian, L. Lin, D. Long, P. Longa-Pena, B. Lundgren, B. W. Lyke, J. T. Mackereth, C. L. MacLeod, S. R. Majewski, A. Manchado, C. Maraston, P. Martini, T. Masseron, K. L. Masters, S. Mathur, R. M. McDermid, A. Merloni, M. Merrifield, S. Meszaros, A. Miglio, D. Minniti, R. Minsley, T. Miyaji, F. Gohar Mohammad, B. Mosser, E.-M. Mueller, D. Muna, A. Munoz-Gutierrez, A. D. Myers, S. Nadathur, P. Nair, J. Correa do Nascimento, R. J. Nevin, J. A. Newman, D. L. Nidever, C. Nitschelm, P. Noterdaeme, J. E. O’Connell, M. D. Olmstead, D. Oravetz, A. Oravetz, Y. Osorio, Z. J. Pace, N. Padilla, N. Palanque-Delabrouille, P. A. Palicio, H.-A. Pan, K. Pan, J. Parker, R. Paviot, S. Peirani, K. Pena Ramrez, S. Penny, W. J. Percival, I. Perez-Fournon, I. Perez-Rafols, P. Petitjean, M. M. Pieri, M. Pinsonneault, V. J. Poovelil, J. T. Povick, A. Prakash, A. M. Price-Whelan, M. J. Raddick, A. Raichoor, A. Ray, S. Barboza Rembold, M. Rezaie, R. A. Riffel, R. Riffel, H.-W. Rix, A. C. Robin, A. Roman-Lopes, C. Roman-Zuniga, B. Rose, A. J. Ross, G. Rossi, K. Rowlands, K. H. R. Rubin, M. Salvato, A. G. Sanchez, L. Sanchez-Menguiano, J. R. Sanchez-Gallego, C. Sayres, A. Schaefer, R. P. Schiavon, J. S. Schimoia, E. Schlaflay, D. Schlegel, D. P. Schneider, M. Schultheis, A. Schwone, H.-J. Seo, A. Serenelli, A. Shafieloo, S. J. Shamsi, Z. Shao, S. Shen, M. Shetrone, R. Shirley, V. Silva Aguirre, J. D. Simon, M. F. Skrutskie, A. Slosar, R. Smethurst, J. Sobeck, B. Cervantes Sodi, D. Souto, D. V. Stark, K. G. Stassun, M. Steinmetz, D. Stello, J. Stermer, T. Storchi-Bergmann, A. Streblyanska, G. S. Stringfellow, A. Stutz, G. Suarez, J. Sun, M. Taghizadeh-Popp, M. S. Talbot, J. Tayar, A. R. Thakar, R. Theriault, D. Thomas, Z. C. Thomas, J. Tinker, R. Tojeiro, H. Hernandez Toledo, C. A. Tremonti, N. W. Troup, S. Tuttle, E. Unda-Sanzana, M. Valentini, J. Vargas-Gonzalez, M. Vargas-Magana, J. A. Vazquez-Mata, M. Vivek, D. Wake, Y. Wang, B. A. Weaver, A.-M. Weijmans, V. Wild, J. C. Wilson, R. F. Wilson, N. Wolthuis, W. M. Wood-Vasey, R. Yan, M. Yang, C. Yeche, O. Zamora, P. Zarrouk, G. Zasowski, K. Zhang, C. Zhao, G. Zhao, Z. Zheng, Z. Zheng, G. Zhu and H. Zou, *The Sixteenth Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra*, arXiv e-prints , arXiv:1912.02905 (2019).
- [97] M. Steinmetz, G. Matijevic, H. Enke, T. Zwitter, G. Guiglion, P. J. McMillan, G. Kordopatis, M. Valentini, C. Chiappini, L. Casagrande, J. Wojno, B. Anguiano, O. Bienayme, A. Bijaoui, J. Binney, D. Burton, P. Cass, P. de Laverny, K. Fiegert, K. Freeman, J. P. Fulbright, B. K. Gibson, G. Gilmore, E. K. Grebel, A. Helmi, A. Kunder, U. Munari, J. F. Navarro, Q. Parker, G. R. Ruchti, A. Recio-Blanco, W. Reid, G. M. Seabroke, A. Siviero, A. Siebert, M. Stupar, F. Watson, M. E. K. Williams, R. F. G. Wyse, F. Anders, T. Antoja, D. Birko, J. Bland-Hawthorn, D. Bossini, R. A. Garcia, I. Carrillo, W. J. Chaplin, Y. Elsworth, B. Famaey, O. Gerhard, P. Jofre, A. Just, S. Mathur, A. Miglio, I. Minchev, G. Monari, B. Mosser, A. Ritter, T. S. Rodrigues, R.-D. Scholz, S. Sharma and K. Sysoliatina, *The Sixth Data Release of the Radial Velocity Experiment (RAVE) – I: Survey Description, Spectra and Radial Velocities*, arXiv e-prints , arXiv:2002.04377 (2020).
- [98] M. Ness, D. W. Hogg, H.-W. Rix, A. Y. Q. Ho and G. Zasowski, *The Cannon*:

- A data-driven approach to Stellar Label Determination*, **808**, 16 (2015).
- [99] S. Buder, M. Asplund, L. Duong, J. Kos, K. Lind, M. K. Ness, S. Sharma, J. Bland-Hawthorn, A. R. Casey, G. M. De Silva, V. D’Orazi, K. C. Freeman, G. F. Lewis, J. Lin, S. L. Martell, K. J. Schlesinger, J. D. Simpson, D. B. Zucker, T. Zwitter, A. M. Amarsi, B. Anguiano, D. Carollo, L. Casagrande, K. Čotar, P. L. Cottrell, G. Da Costa, X. D. Gao, M. R. Hayden, J. Horner, M. J. Ireland, P. R. Kafle, U. Munari, D. M. Nataf, T. Nordlander, D. Stello, Y.-S. Ting, G. Traven, F. Watson, R. A. Wittenmyer, R. F. G. Wyse, D. Yong, J. C. Zinn and M. Žerjal, *The GALAH Survey: second data release*, **478**, 4513 (2018).
- [100] H. W. Leung and J. Bovy, *Deep learning of multi-element abundances from high-resolution spectroscopic data*, **483**, 3255 (2019).
- [101] R. Wang, A. L. Luo, J.-J. Chen, W. Hou, S. Zhang, Y.-H. Zhao, X.-R. Li, Y.-H. Hou and LAMOST MRS Collaboration, *SPCANet: Stellar Parameters and Chemical Abundances Network for LAMOST-II Medium Resolution Survey*, **891**, 23 (2020).
- [102] R. Olney, M. Kounkel, C. Schillinger, M. T. Scoggins, Y. Yin, E. Howard, K. R. Covey, B. Hutchinson and K. G. Stassun, *APOGEE Net: Improving the derived spectral parameters for young stars through deep learning*, arXiv e-prints , arXiv:2002.08390 (2020).
- [103] B. Chen, E. D’Onghia, S. A. Pardy, A. Pasquali, C. Bertelli Motta, B. Hanlon and E. K. Grebel, *Chemodynamical Clustering Applied to APOGEE Data: Rediscovering Globular Clusters*, **860**, 70 (2018).
- [104] N. Price-Jones and J. Bovy, *Blind chemical tagging with DBSCAN: prospects for spectroscopic surveys*, **487**, 871 (2019).
- [105] P. Jofré, P. Das, J. Bertranpetti and R. Foley, *Cosmic phylogeny: reconstructing the chemical history of the solar neighbourhood with an evolutionary tree*, **467**, 1140 (2017).
- [106] S. Blanco-Cuaresma and D. Fraix-Burnet, *A phylogenetic approach to chemical tagging. Reassembling open cluster stars*, **618**, A65 (2018).
- [107] L. Lindegren, *"Re-normalising the astrometric chi-square in Gaia DR2"*, *Gaia Technical Note: GAIA-C3-TN-LU-LL-124-0* (2018).
- [108] I. Becker, K. Pichara, M. Catelan, P. Protopapas, C. Aguirre and F. Nikzat, *Scalable end-to-end recurrent neural network for variable star classification*, **493**, 2981 (2020).
- [109] Y. Bai, J. Liu, Z. Bai, S. Wang and D. Fan, *Machine-learning Regression of Stellar Effective Temperatures in the Second Gaia Data Release*, **158**, 93 (2019).

Bibliography

- [110] G. Marton, P. Ábrahám, E. Szegedi-Elek, J. Varga, M. Kun, Á. Kóspál, E. Varga-Verebélyi, S. Hodgkin, L. Szabados, R. Beck and C. Kiss, *Identification of Young Stellar Object candidates in the Gaia DR2 x AllWISE catalogue with machine learning methods*, **487**, 2522 (2019).
- [111] A. Helmi, J. Veljanoski, M. A. Breddels, H. Tian and L. V. Sales, *A box full of chocolates: The rich structure of the nearby stellar halo revealed by Gaia and RAVE*, **598**, A58 (2017).
- [112] N. W. Borsato, S. L. Martell and J. D. Simpson, *Identifying stellar streams in Gaia DR2 with data mining techniques*, **492**, 1370 (2020).
- [113] B. Ostdiek, L. Necib, T. Cohen, M. Freytsis, M. Lisanti, S. Garrison-Kimmel, A. Wetzel, R. E. Sanderson and P. F. Hopkins, *Cataloging Accreted Stars within Gaia DR2 using Deep Learning*, arXiv e-prints , arXiv:1907.06652 (2019).
- [114] L. Necib, B. Ostdiek, M. Lisanti, T. Cohen, M. Freytsis and S. Garrison-Kimmel, *Chasing Accreted Structures within Gaia DR2 using Deep Learning*, arXiv e-prints , arXiv:1907.07681 (2019).
- [115] T. Marchetti, E. M. Rossi, G. Kordopatis, A. G. A. Brown, A. Rimoldi, E. Starkenburg, K. Youakim and R. Ashley, *An artificial neural network to discover hypervelocity stars: candidates in Gaia DR1/TGAS*, **470**, 1388 (2017).
- [116] Y. Bai, J.-F. Liu and S. Wang, *Machine learning classification of Gaia Data Release 2*, Research in Astronomy and Astrophysics **18**, 118 (2018).
- [117] C. A. L. Bailer-Jones, M. Fouesneau and R. Andrae, *Quasar and galaxy classification in Gaia Data Release 2*, **490**, 5615 (2019).
- [118] Y. Bai, J. Liu, Y. Wang and S. Wang, *Machine-learning Regression of Extinction in the Second Gaia Data Release*, **159**, 84 (2020).
- [119] A. Castro-Ginard, C. Jordi, X. Luri, J. Álvarez Cid-Fuentes, L. Casamiquela, F. Anders, T. Cantat-Gaudin, M. Monguió, L. Balaguer-Núñez, S. Solà and R. M. Badia, *Hunting for open clusters in Gaia DR2: 582 new open clusters in the Galactic disc*, **635**, A45 (2020).
- [120] N. V. Kharchenko, A. E. Piskunov, E. Schilbach, S. Röser and R.-D. Scholz, *Global survey of star clusters in the Milky Way. II. The catalogue of basic parameters*, **558**, A53 (2013).
- [121] R. Hoogerwerf, J. H. J. de Bruijne and P. T. de Zeeuw, *The Origin of Runaway Stars*, **544**, L133 (2000).
- [122] L. van der Maaten, *Barnes-Hut-SNE*, ArXiv e-prints (2013).
- [123] V. Adibekyan, E. Delgado-Mena, S. Feltzing, J. I. González Hernández, N. R. Hinkel, A. J. Korn, M. Asplund, P. G. Beck, M. Deal, B. Gustafsson, S. Honda, K. Lind, P. E. Nissen and L. Spina, *Sun-like stars unlike the Sun: Clues for chemical anomalies of cool stars*, Astronomische Nachrichten **338**, 442 (2017).

- [124] J. Meléndez, M. Asplund, B. Gustafsson and D. Yong, *The Peculiar Solar Composition and Its Possible Relation to Planet Formation*, **704**, L66 (2009).
- [125] K. Biazzo, L. Pasquini, P. Bonifacio, S. Randich and L. R. Bedin, *True solar analogues in the open cluster M67*., **80**, 125 (2009).
- [126] F. Liu, M. Asplund, D. Yong, J. Meléndez, I. Ramírez, A. I. Karakas, M. Carlos and A. F. Marino, *The chemical compositions of solar twins in the open cluster M67*, **463**, 696 (2016).
- [127] S. B. Howell, C. Sobeck, M. Haas, M. Still, T. Barclay, F. Mullally, J. Troeltzsch, S. Aigrain, S. T. Bryson, D. Caldwell, W. J. Chaplin, W. D. Cochran, D. Huber, G. W. Marcy, A. Miglio, J. R. Najita, M. Smith, J. D. Twicken and J. J. Fortney, *The K2 Mission: Characterization and Early Results*, **126**, 398 (2014).
- [128] G. R. Ricker, J. N. Winn, R. Vanderspek, D. W. Latham, G. Á. Bakos, J. L. Bean, Z. K. Berta-Thompson, T. M. Brown, L. Buchhave, N. R. Butler, R. P. Butler, W. J. Chaplin, D. Charbonneau, J. Christensen-Dalsgaard, M. Clampin, D. Deming, J. Doty, N. De Lee, C. Dressing, E. W. Dunham, M. Endl, F. Fressin, J. Ge, T. Henning, M. J. Holman, A. W. Howard, S. Ida, J. Jenkins, G. Jernigan, J. A. Johnson, L. Kaltenegger, N. Kawai, H. Kjeldsen, G. Laughlin, A. M. Levine, D. Lin, J. J. Lissauer, P. MacQueen, G. Marcy, P. R. McCullough, T. D. Morton, N. Narita, M. Paegert, E. Palle, F. Pepe, J. Pepper, A. Quirrenbach, S. A. Rinehart, D. Sasselov, B. Sato, S. Seager, A. Sozzetti, K. G. Stassun, P. Sullivan, A. Szentgyorgyi, G. Torres, S. Udry and J. Villasenor, *Transiting Exoplanet Survey Satellite (TESS)*, in *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, , Vol. 9143 (2014) p. 914320, arXiv:1406.0151 [astro-ph.EP] .
- [129] Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, M. Biermann, D. W. Evans and et al., *The Gaia mission*, **595**, A1 (2016).
- [130] M. Cropper, D. Katz, P. Sartoretti, T. Prusti, J. H. J. de Bruijne, F. Chassat, P. Charvet, J. Boyadjian, M. Perryman, G. Sarri, P. Gare, M. Erdmann, U. Munari, T. Zwitter, M. Wilkinson, F. Arenou, A. Vallenari, A. Gómez, P. Panuzzo, G. Seabroke, C. Allende Prieto, K. Benson, O. Marchal, H. Huckle, M. Smith, C. Dolding, K. Janßen, Y. Viala, R. Blomme, S. Baker, S. Boudreault, F. Crifo, C. Soubiran, Y. Frémat, G. Jasniewicz, A. Guerrier, L. P. Guy, C. Turon, A. Jean-Antoine-Piccolo, F. Thévenin, M. David, E. Gosset and Y. Damerdji, *Gaia Data Release 2. Gaia Radial Velocity Spectrometer*, **616**, A5 (2018).
- [131] Gaia Collaboration, F. Mignard, S. A. Klioner, L. Lindegren, J. Hernández, U. Bastian, A. Bombrun, D. Hobbs, U. Lammers, D. Michalik, M. Ramos-Lerate, M. Biermann, J. Fernández-Hernández, R. Geyer, T. Hilger, H. I. Siddiqui, H. Steidelmüller, C. Babusiaux, C. Barache, S. Lambert, A. H. Andrei, G. Bourda, P. Charlot, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. A. L. Bailer-Jones, D. W. Evans, L. Eyer, F. Jansen,

Bibliography

C. Jordi, X. Luri, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, C. Soubiran, F. van Leeuwen, N. A. Walton, F. Arenou, M. Cropper, R. Drimmel, D. Katz, M. G. Lattanzi, J. Bakker, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, F. De Angeli, C. Fabricius, R. Guerra, B. Holl, E. Masana, R. Messineo, N. Mowlavi, K. Nienartowicz, P. Panuzzo, J. Portell, M. Riello, G. M. Seabroke, P. Tanga, F. Thévenin, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. Burgess, G. Busso, B. Carry, A. Cellino, G. Clementini, M. Clotet, O. Creevey, M. Davidson, J. De Ridder, L. Delchambre, A. Dell'Oro, C. Ducourant, M. Fouesneau, Y. Frémat, L. Galluccio, M. García-Torres, J. González-Núñez, J. J. González-Vidal, E. Gosset, L. P. Guy, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, D. Hestroffer, S. T. Hodgkin, A. Hutton, G. Jasniewicz, A. Jean-Antoine-Piccolo, S. Jordan, A. J. Korn, A. Krone-Martins, A. C. Lanzafame, T. Lebzelter, W. Löffler, M. Manteiga, P. M. Marrese, J. M. Martín-Fleitas, A. Moitinho, A. Mora, K. Muinonen, J. Osinde, E. Pancino, T. Pauwels, J. M. Petit, A. Recio-Blanco, P. J. Richards, L. Rimoldini, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, M. Süveges, J. Torra, W. van Reeven, U. Abbas, A. Abreu Aramburu, S. Accart, C. Aerts, G. Altavilla, M. A. Álvarez, R. Alvarez, J. Alves, R. I. Anderson, E. Anglada Varela, E. Antiche, T. Antoja, B. Arcay, T. L. Astraatmadja, N. Bach, S. G. Baker, L. Balaguer-Núñez, P. Balm, C. Barata, D. Barbato, F. Barblan, P. S. Barklem, D. Barrado, M. Barros, M. A. Barstow, L. Bartholomé Muñoz, J. L. Bassilana, U. Becciani, M. Bellazzini, A. Berihuete, S. Bertone, L. Bianchi, O. Bienaymé, S. Blanco-Cuaresma, T. Boch, C. Boeche, R. Borrachero, D. Bossini, S. Bouquillon, A. Bragaglia, L. Bramante, M. A. Breddels, A. Bressan, N. Brouillet, T. Brüsemeister, E. Brugaglia, B. Bucciarelli, A. Burlacu, D. Busonero, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, G. Cannizzaro, T. Cantat-Gaudin, R. Carballo, T. Carlucci, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, L. Chemin, A. Chiavassa, G. Cocozza, G. Costigan, S. Cowell, F. Crifo, M. Crosta, C. Crowley, J. Cuypers, C. Dafonte, Y. Damerdji, A. Dapergolas, P. David, M. David, P. de Laverny, F. De Luise, R. De March, R. de Souza, A. de Torres, J. Debosscher, E. del Pozo, M. Delbo, A. Delgado, H. E. Delgado, S. Diakite, C. Diener, E. Distefano, C. Dolding, P. Drazinos, J. Durán, B. Edvardsson, H. Enke, K. Eriksson, P. Esquej, G. Eynard Bon-temps, C. Fabre, M. Fabrizio, S. Faigler, A. J. Falcão, M. Farràs Casas, L. Federici, G. Fedorets, P. Fernique, F. Figueras, F. Filippi, K. Findeisen, A. Fonti, E. Fraile, M. Fraser, B. Frézouls, M. Gai, S. Galleti, D. Garabato, F. García-Sedano, A. Garofalo, N. Garralda, A. Gavel, P. Gavras, J. Gerssen, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, F. Glass, M. Gomes, M. Granvik, A. Gueguen, A. Guerrier, J. Guiraud, R. Gutié, R. Haigron, D. Hatzidimitriou, M. Hauser, M. Haywood, U. Heiter, A. Helmi, J. Heu, W. Hofmann, G. Holland, H. E. Huckle, A. Hypki, V. Icardi, K. Janßen, G. Jevardat de Fombelle, P. G. Jonker, A. L. Juhász, F. Julbe, A. Karampelas, A. Kewley, J. Klar, A. Kochoska, R. Kohley, K. Kolenberg, M. Kontizas, E. Kontizas, S. E. Koposov, G. Kordopatis, Z. Kostrzewa-Rutkowska, P. Koubsky, A. F. Lanza, Y. Lasne, J. B. Lavigne, Y. Le Fustec, C. Le Poncin-Lafitte, Y. Lebreton, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, H. Lenhardt, F. Leroux, S. Liao,

E. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, M. López, S. Managau, R. G. Mann, G. Mantelet, O. Marchal, J. M. Marchant, M. Marconi, S. Marinoni, G. Marschalkó, D. J. Marshall, M. Martino, G. Marton, N. Mary, D. Massari, G. Matijević, T. Mazeh, P. J. McMillan, S. Messina, N. R. Millar, D. Molina, R. Molinaro, L. Molnár, P. Montegriffo, R. Mor, R. Morbidelli, T. Morel, D. Morris, A. F. Mulone, T. Muraveva, I. Musella, G. Nelemans, L. Nicastro, L. Noval, W. O'Mullane, C. Ordénovic, D. Ordóñez-Blanco, P. Osborne, C. Pagani, I. Pagano, F. Pailler, H. Palacin, L. Palaversa, A. Panahi, M. Pawlak, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, E. Poujoulet, A. Prša, L. Pulone, E. Racero, S. Ragaini, N. Rambaux, S. Regibo, C. Reyillé, F. Riclet, V. Ripepi, A. Riva, A. Rivard, G. Rixon, T. Roegiers, M. Roelens, M. Romero-Gómez, N. Rowell, F. Royer, L. Ruiz-Dern, G. Sadowski, T. Sagristà Sellés, J. Sahlmann, J. Salgado, E. Salguero, N. Sanna, T. Santana-Ros, M. Sarasso, H. Savietto, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, I. C. Shih, L. Siltala, A. F. Silva, R. L. Smart, K. W. Smith, E. Solano, F. Solitro, R. Sordo, S. Soria Nieto, J. Souchay, A. Spagna, F. Spoto, U. Stampa, I. A. Steele, C. A. Stephenson, H. Stoev, F. F. Suess, J. Surdej, L. Szabados, E. Szegedi-Elek, D. Tapiador, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, D. Terrett, P. Teyssandier, W. Thuillot, A. Titarenko, F. Torra Clotet, C. Turon, A. Ulla, E. Utrilla, S. Uzzi, M. Vaillant, G. Valentini, V. Valette, A. van Elteren, E. Van Hemelryck, M. van Leeuwen, M. Vaschetto, A. Vecchiato, J. Veljanoski, Y. Viala, D. Vicente, S. Vogt, C. von Essen, H. Voss, V. Votruba, S. Voutsinas, G. Walsley, M. Weiler, O. Wertz, T. Wevers, Ł. Wyrzykowski, A. Yoldas, M. Žerjal, H. Ziaeepour, J. Zorec, S. Zschocke, S. Zucker, C. Zurbach and T. Zwitter, *Gaia Data Release 2. The celestial reference frame (Gaia-CRF2)*, **616**, A14 (2018).

- [132] L. Lindegren, J. Hernández, A. Bombrun, S. Klioner, U. Bastian, M. Ramos-Lerate, A. de Torres, H. Steidelmüller, C. Stephenson, D. Hobbs, U. Lammers, M. Biermann, R. Geyer, T. Hilger, D. Michalik, U. Stampa, P. J. McMillan, J. Castañeda, M. Clotet, G. Comoretto, M. Davidson, C. Fabricius, G. Gracia, N. C. Hambly, A. Hutton, A. Mora, J. Portell, F. van Leeuwen, U. Abbas, A. Abreu, M. Altmann, A. Andrei, E. Anglada, L. Balaguer-Núñez, C. Barache, U. Becciani, S. Bertone, L. Bianchi, S. Bouquillon, G. Bourda, T. Brüsemeister, B. Bucciarelli, D. Busonero, R. Buzzi, R. Cancelliere, T. Carlucci, P. Charlot, N. Cheek, M. Crosta, C. Crowley, J. de Bruijne, F. de Felice, R. Drimmel, P. Esquej, A. Fienga, E. Fraile, M. Gai, N. Garralda, J. J. González-Vidal, R. Guerra, M. Hauser, W. Hofmann, B. Holl, S. Jordan, M. G. Lattanzi, H. Lenhardt, S. Liao, E. Licata, T. Lister, W. Löffler, J. Marchant, J. M. Martin-Fleitas, R. Messineo, F. Mignard, R. Morbidelli, E. Poggio, A. Riva, N. Rowell, E. Salguero, M. Sarasso, E. Sciacca, H. Siddiqui, R. L. Smart, A. Spagna, I. Steele, F. Taris, J. Torra, A. van Elteren, W. van Reeven and A. Vecchiato, *Gaia Data Release 2. The astrometric solution*, **616**, A2 (2018).
- [133] X. Luri, A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, A. Castro-Ginard, J. de Bruijne, T. Prusti, C. Babusiaux and H. E. Delgado, *Gaia Data Release 2. Using Gaia parallaxes*, **616**, A9 (2018).

Bibliography

- [134] D. W. Evans, M. Riello, F. De Angeli, J. M. Carrasco, P. Montegriffo, C. Fabricius, C. Jordi, L. Palaversa, C. Diener, G. Busso, C. Cacciari, F. van Leeuwen, P. W. Burgess, M. Davidson, D. L. Harrison, S. T. Hodgkin, E. Pancino, P. J. Richards, G. Altavilla, L. Balaguer-Núñez, M. A. Barstow, M. Bellazzini, A. G. A. Brown, M. Castellani, G. Cocozza, F. De Luise, A. Delgado, C. Ducourant, S. Galleti, G. Gilmore, G. Giuffrida, B. Holl, A. Kewley, S. E. Koposov, S. Marinoni, P. M. Marrese, P. J. Osborne, A. Piersimoni, J. Portell, L. Pulone, S. Ragaini, N. Sanna, D. Terrett, N. A. Walton, T. Wevers and Ł. Wyrzykowski, *Gaia Data Release 2. Photometric content and validation*, **616**, A4 (2018).
- [135] M. Riello, F. De Angeli, D. W. Evans, G. Busso, N. C. Hambly, M. Davidson, P. W. Burgess, P. Montegriffo, P. J. Osborne, A. Kewley, J. M. Carrasco, C. Fabricius, C. Jordi, C. Cacciari, F. van Leeuwen and G. Holland, *Gaia Data Release 2. Processing of the photometric data*, **616**, A3 (2018).
- [136] C. Soubiran, G. Jasniewicz, L. Chemin, C. Zurbach, N. Brouillet, P. Panuzzo, P. Sartoretti, D. Katz, J. F. Le Campion, O. Marchal, D. Hestroffer, F. Thévenin, F. Crifo, S. Udry, M. Cropper, G. Seabroke, Y. Viala, K. Benson, R. Blomme, A. Jean-Antoine, H. Huckle, M. Smith, S. G. Baker, Y. Damerdji, C. Dolding, Y. Frémat, E. Gosset, A. Guerrier, L. P. Guy, R. Haigron, K. Janßen, G. Plum, C. Fabre, Y. Lasne, F. Pailler, C. Panem, F. Riclet, F. Royer, G. Tauran, T. Zwitter, A. Gueguen and C. Turon, *Gaia Data Release 2. The catalogue of radial velocity standard stars*, **616**, A7 (2018).
- [137] P. Sartoretti, D. Katz, M. Cropper, P. Panuzzo, G. M. Seabroke, Y. Viala, K. Benson, R. Blomme, G. Jasniewicz, A. Jean-Antoine, H. Huckle, M. Smith, S. Baker, F. Crifo, Y. Damerdji, M. David, C. Dolding, Y. Frémat, E. Gosset, A. Guerrier, L. P. Guy, R. Haigron, K. Janßen, O. Marchal, G. Plum, C. Soubiran, F. Thévenin, M. Ajaj, C. Allende Prieto, C. Babusiaux, S. Boudreault, L. Chemin, C. Delle Luche, C. Fabre, A. Gueguen, N. C. Hambly, Y. Lasne, F. Meynadier, F. Pailler, C. Panem, F. Riclet, F. Royer, G. Tauran, C. Zurbach, T. Zwitter, F. Arenou, A. Gomez, V. Lemaitre, N. Leclerc, T. Morel, U. Munari, C. Turon and M. Žerjal, *Gaia Data Release 2. Processing the spectroscopic data*, **616**, A6 (2018).
- [138] B. Holl, M. Audard, K. Nienartowicz, G. Jevardat de Fombelle, O. Marchal, N. Mowlavi, G. Clementini, J. De Ridder, D. W. Evans, L. P. Guy, A. C. Lanzafame, T. Lebzelter, L. Rimoldini, M. Roelens, S. Zucker, E. Distefano, A. Garofalo, I. Lecoeur-Taïbi, M. Lopez, R. Molinaro, T. Muraveva, A. Panahi, S. Regibo, V. Ripepi, L. M. Sarro, C. Aerts, R. I. Anderson, J. Charnas, F. Barblan, S. Blanco-Cuaresma, G. Busso, J. Cuypers, F. De Angeli, F. Glass, M. Grenon, Á. L. Juhász, A. Kochoska, P. Koubeky, A. F. Lanza, S. Leccia, D. Lorenz, M. Marconi, G. Marschalkó, T. Mazeh, S. Messina, F. Mignard, A. Moitinho, L. Molnár, S. Morgenthaler, I. Musella, C. Ordenovic, D. Ordóñez, I. Pagano, L. Palaversa, M. Pawlak, E. Plachy, A. Prša, M. Riello, M. Süveges, L. Szabados, E. Szegedi-Elek, V. Votruba and L. Eyer, *Gaia Data Release 2. Summary of the variability processing and analysis results*, **618**, A30 (2018).

- [139] N. Mowlavi, I. Lecoeur-Taïbi, T. Lebzelter, L. Rimoldini, D. Lorenz, M. Audard, J. De Ridder, L. Eyer, L. P. Guy, B. Holl, G. Jevardat de Fombelle, O. Marchal, K. Nienartowicz, S. Regibo, M. Roelens and L. M. Sarro, *Gaia Data Release 2. The first Gaia catalogue of long-period variable candidates*, **618**, A58 (2018).
- [140] L. Molnár, E. Plachy, Á. L. Juhász and L. Rimoldini, *Gaia Data Release 2. Validating the classification of RR Lyrae and Cepheid variables with the Kepler and K2 missions*, **620**, A127 (2018).
- [141] G. Clementini, V. Ripepi, R. Molinaro, A. Garofalo, T. Muraveva, L. Rimoldini, L. P. Guy, G. Jevardat de Fombelle, K. Nienartowicz, O. Marchal, M. Audard, B. Holl, S. Leccia, M. Marconi, I. Musella, N. Mowlavi, I. Lecoeur-Taibi, L. Eyer, J. De Ridder, S. Regibo, L. M. Sarro, L. Szabados, D. W. Evans and M. Riello, *Gaia Data Release 2. Specific characterisation and validation of all-sky Cepheids and RR Lyrae stars*, **622**, A60 (2019).
- [142] R. Andrae, M. Fouesneau, O. Creevey, C. Ordenovic, N. Mary, A. Burlacu, L. Chaoul, A. Jean-Antoine-Piccolo, G. Kordopatis, A. Korn, Y. Lebreton, C. Panem, B. Pichon, F. Thévenin, G. Walmsley and C. A. L. Bailer-Jones, *Gaia Data Release 2. First stellar parameters from Apsis*, **616**, A8 (2018).
- [143] Gaia Collaboration, F. Spoto, P. Tanga, F. Mignard, J. Berthier, B. Carry, A. Cellino, A. Dell’Oro, D. Hestroffer, K. Muinonen, T. Pauwels, J. M. Petit, P. David, F. De Angeli, M. Delbo, B. Frézouls, L. Galluccio, M. Granvik, J. Guiraud, J. Hernández, C. Ordénovic, J. Portell, E. Poujoulet, W. Thuillot, G. Walmsley, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Eyer, F. Jansen, C. Jordi, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, H. I. Siddiqui, C. Soubiran, F. van Leeuwen, N. A. Walton, F. Arenou, U. Bastian, M. Cropper, R. Drimmel, D. Katz, M. G. Lattanzi, J. Bakker, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, C. Fabricius, R. Guerra, B. Holl, E. Masana, R. Messineo, N. Mowlavi, K. Nienartowicz, P. Panuzzo, M. Riello, G. M. Seabroke, F. Thévenin, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, R. Blomme, P. Burgess, G. Busso, G. Clementini, M. Clotet, O. Creevey, M. Davidson, J. De Ridder, L. Delchambre, C. Ducourant, J. Fernández-Hernández, M. Fouesneau, Y. Frémat, M. García-Torres, J. González-Núñez, J. J. González-Vidal, E. Gosset, L. P. Guy, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, S. T. Hodgkin, A. Hutton, G. Jasniewicz, A. Jean-Antoine-Piccolo, S. Jordan, A. J. Korn, A. Krone-Martins, A. C. Lanzafame, T. Lebzelter, W. Lö, M. Manteiga, P. M. Marrese, J. M. Martín-Fleitas, A. Moitinho, A. Mora, J. Osinde, E. Pancino, A. Recio-Blanco, P. J. Richards, L. Rimoldini, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, M. Süveges, J. Torra, W. van Reeven, U. Abbas, A. Abreu Aramburu, S. Accart, C. Aerts, G. Altavilla, M. A. Álvarez, R. Alvarez, J. Alves, R. I. Anderson, A. H. Andrei, E. Anglada Varela, E. Antiche, T. Antoja, B. Arcay, T. L. Astraatmadja, N. Bach, S. G. Baker, L. Balaguer-Núñez, P. Balm, C. Barache, C. Barata, D. Barbato, F. Barblan, P. S. Barklem, D. Barrado, M. Barros, M. A. Barstow,

Bibliography

L. Bartholomé Muñoz, J. L. Bassilana, U. Becciani, M. Bellazzini, A. Berihuete, S. Bertone, L. Bianchi, O. Bienaymé, S. Blanco-Cuaresma, T. Boch, C. Boeche, A. Bombrun, R. Borrachero, D. Bossini, S. Bouquillon, G. Bourda, A. Bragaglia, L. Bramante, M. A. Breddels, A. Bressan, N. Brouillet, T. Brüsemeister, E. Brugaletta, B. Bucciarelli, A. Burlacu, D. Busonero, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, G. Cannizzaro, T. Cantat-Gaudin, R. Carballo, T. Carlucci, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, P. Charlot, L. Chemin, A. Chiavassa, G. Cocozza, G. Costigan, S. Cowell, F. Crifo, M. Crosta, C. Crowley, J. Cuypers, C. Dafonte, Y. Damerdji, A. Dapergolas, M. David, P. de Laverny, F. De Luise, R. De March, R. de Souza, A. de Torres, J. Deboscher, E. del Pozo, A. Delgado, H. E. Delgado, S. Diakite, C. Diener, E. Distefano, C. Dolding, P. Drazinos, J. Durán, B. Edvardsson, H. Enke, K. Eriksson, P. Esquej, G. Eynard Bon-temps, C. Fabre, M. Fabrizio, S. Faigler, A. J. Falcão, M. Farràs Casas, L. Federici, G. Fedorets, P. Fernique, F. Figueras, F. Filippi, K. Findeisen, A. Fonti, E. Fraile, M. Fraser, M. Gai, S. Galletti, D. Garabato, F. García-Sedano, A. Garofalo, N. Garralda, A. Gavel, P. Gavras, J. Gerssen, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, F. Glass, M. Gomes, A. Gueguen, A. Guerrier, R. Gutié, R. Haigron, D. Hatzidimitriou, M. Hauser, M. Haywood, U. Heiter, A. Helmi, J. Heu, T. Hilger, D. Hobbs, W. Hofmann, G. Holland, H. E. Huckle, A. Hypki, V. Icardi, K. Janßen, G. Jevardat de Fombelle, P. G. Jonker, Á. L. Juhász, F. Julbe, A. Karampelas, A. Kewley, J. Klar, A. Kochoska, R. Kohley, K. Kolenberg, M. Kontizas, E. Kontizas, S. E. Koposov, G. Kordopatis, Z. Kostrzewa-Rutkowska, P. Koubsky, S. Lambert, A. F. Lanza, Y. Lasne, J. B. Lavigne, Y. Le Fustec, C. Le Poncin-Lafitte, Y. Lebreton, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, H. Lenhardt, F. Leroux, S. Liao, E. Licata, H. E. P. Lindström, T. A. Lister, E. Livanou, A. Lobel, M. López, S. Managau, R. G. Mann, G. Mantelet, O. Marchal, J. M. Marchant, M. Marconi, S. Marinoni, G. Marschalkó, D. J. Marshall, M. Martino, G. Marton, N. Mary, D. Massari, G. Matijević, T. Mazeh, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, D. Molina, R. Molinaro, L. Molnár, P. Montegriffo, R. Mor, R. Morbidelli, T. Morel, D. Morris, A. F. Mulone, T. Muraveva, I. Musella, G. Nelemans, L. Nicastro, L. Noval, W. O'Mullane, D. Ordóñez-Blanco, P. Osborne, C. Pagani, I. Pagano, F. Pailler, H. Palacin, L. Palaversa, A. Panahi, M. Pawlak, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, A. Prša, L. Pulone, E. Racero, S. Ragaini, N. Rambaux, M. Ramos-Lerate, S. Regibo, C. Reylé, F. Riclet, V. Ripepi, A. Riva, A. Rivard, G. Rixon, T. Roegiers, M. Roelens, M. Romero-Gómez, N. Rowell, F. Royer, L. Ruiz-Dern, G. Sadowski, T. Sagristà Sellés, J. Sahlmann, J. Salgado, E. Salguero, N. Sanna, T. Santana-Ros, M. Sarasso, H. Savietto, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, I. C. Shih, L. Siltala, A. F. Silva, R. L. Smart, K. W. Smith, E. Solano, F. Solitro, R. Sordo, S. Soria Nieto, J. Souchay, A. Spagna, U. Stampa, I. A. Steele, H. Steidelmüller, C. A. Stephenson, H. Stoev, F. F. Suess, J. Surdej, L. Szabados, E. Szegedi-Elek, D. Tapiador, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, D. Terrlett, P. Teyssandier, A. Titarenko, F. Torra Clotet, C. Turon, A. Ulla, E. Utrilla, S. Uzzi, M. Vaillant, G. Valentini, V. Valette, A. van Elteren, E. Van Hemelryck, M. van Leeuwen, M. Vaschetto, A. Vecchiato, J. Veljanoski,

- Y. Viala, D. Vicente, S. Vogt, C. von Essen, H. Voss, V. Votruba, S. Voutsinas, M. Weiler, O. Wertz, T. Wevers, Ł. Wyrzykowski, A. Yoldas, M. Žerjal, H. Ziaeepour, J. Zorec, S. Zschocke, S. Zucker, C. Zurbach and T. Zwitter, *Gaia Data Release 2. Observations of solar system objects*, **616**, A13 (2018).
- [144] Gaia Helpdesk, *Gaia DR2 primer: Everything you wish you had known before you started working with Gaia Data Release 2, issue 1.0* (2019), [view 20. 12. 2019].
- [145] G. M. De Silva, K. C. Freeman, J. Bland-Hawthorn, S. Martell, E. W. de Boer, M. Asplund, S. Keller, S. Sharma, D. B. Zucker, T. Zwitter, B. Anguiano, C. Bacigalupo, D. Bayliss, M. A. Beavis, M. Bergemann, S. Campbell, R. Cannon, D. Carollo, L. Casagrande, A. R. Casey, G. Da Costa, V. D’Orazi, A. Dotter, L. Duong, A. Heger, M. J. Ireland, P. R. Kafle, J. Kos, J. Lattanzio, G. F. Lewis, J. Lin, K. Lind, U. Munari, D. M. Nataf, S. O’Toole, Q. Parker, W. Reid, K. J. Schlesinger, A. Sheinis, J. D. Simpson, D. Stello, Y.-S. Ting, G. Traven, F. Watson, R. Wittenmyer, D. Yong and M. Žerjal, *The GALAH survey: scientific motivation*, **449**, 2604 (2015).
- [146] S. C. Barden, D. J. Jones, S. I. Barnes, J. Heijmans, A. Heng, G. Knight, D. R. Orr, G. A. Smith, V. Churilov, J. Brzeski, L. G. Waller, K. Shortridge, A. J. Horton, D. Mayfield, R. Haynes, D. M. Haynes, D. Whittard, M. Goodwin, S. Smedley, I. Saunders, P. R. Gillingham, E. Penny, T. J. Farrell, M. Vuong, R. Heald, S. Lee, R. Muller, K. Freeman, J. Bland-Hawthorn, D. F. Zucker and G. De Silva, *HERMES: revisions in the design for a high-resolution multi-element spectrograph for the AAT*, in *Ground-based and Airborne Instrumentation for Astronomy III*, , Vol. 7735 (2010) p. 773509.
- [147] A. Sheinis, B. Anguiano, M. Asplund, C. Bacigalupo, S. Barden, M. Birchall, J. Bland-Hawthorn, J. Brzeski, R. Cannon, D. Carollo, S. Case, A. Casey, V. Churilov, C. Warrick, R. Dean, G. De Silva, V. D’Orazi, L. Duong, T. Farrell, K. Fiegert, K. Freeman, F. Gabriella, L. Gers, M. Goodwin, D. Gray, A. Green, R. Heald, J. Heijmans, M. Ireland, D. Jones, P. Kafle, S. Keller, U. Klauser, Y. Kondrat, J. Kos, J. Lawrence, S. Lee, S. Mali, S. Martell, D. Mathews, D. Mayfield, S. Miziarski, R. Muller, N. Pai, R. Patterson, E. Penny, D. Orr, K. Schlesinger, S. Sharma, K. Shortridge, J. Simpson, S. Smedley, G. Smith, D. Stafford, N. Staszak, M. Vuong, L. Waller, E. W. de Boer, P. Xavier, J. Zheng, R. Zhelem, D. Zucker and T. Zwitter, *First light results from the High Efficiency and Resolution Multi-Element Spectrograph at the Anglo-Australian Telescope*, *Journal of Astronomical Telescopes, Instruments, and Systems* **1**, 035002 (2015).
- [148] R. A. Wittenmyer, S. Sharma, D. Stello, S. Buder, J. Kos, M. Asplund, L. Duong, J. Lin, K. Lind, M. Ness, T. Zwitter, J. Horner, J. Clark, S. R. Kane, D. Huber, J. Bland-Hawthorn, A. R. Casey, G. M. De Silva, V. D’Orazi, K. Freeman, S. Martell, J. D. Simpson, D. B. Zucker, B. Anguiano, L. Casagrande, J. Esdaile, M. Hon, M. Ireland, P. R. Kafle, S. Khanna, J. P. Marshall, M. H. M. Saddon, G. Traven and D. Wright, *The K2-HERMES Survey. I. Planet-candidate Properties from K2 Campaigns 1-3*, **155**, 84 (2018).

Bibliography

- [149] S. Sharma, D. Stello, S. Buder, J. Kos, J. Bland-Hawthorn, M. Asplund, L. Duong, J. Lin, K. Lind, M. Ness, D. Huber, T. Zwitter, G. Traven, M. Hon, P. R. Kafle, S. Khanna, H. Saddon, B. Anguiano, A. R. Casey, K. Freeman, S. Martell, G. M. De Silva, J. D. Simpson, R. A. Wittenmyer and D. B. Zucker, *The TESS-HERMES survey data release 1: high-resolution spectroscopy of the TESS southern continuous viewing zone*, **473**, 2004 (2018).
- [150] M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk and S. Wheelock, *The Two Micron All Sky Survey (2MASS)*, **131**, 1163 (2006).
- [151] G. R. Ricker, J. N. Winn, R. Vanderspek, D. W. Latham, G. Á. Bakos, J. L. Bean, Z. K. Berta-Thompson, T. M. Brown, L. Buchhave, N. R. Butler, R. P. Butler, W. J. Chaplin, D. Charbonneau, J. Christensen-Dalsgaard, M. Clampin, D. Deming, J. Doty, N. De Lee, C. Dressing, E. W. Dunham, M. Endl, F. Fressin, J. Ge, T. Henning, M. J. Holman, A. W. Howard, S. Ida, J. M. Jenkins, G. Jernigan, J. A. Johnson, L. Kaltenegger, N. Kawai, H. Kjeldsen, G. Laughlin, A. M. Levine, D. Lin, J. J. Lissauer, P. MacQueen, G. Marcy, P. R. McCullough, T. D. Morton, N. Narita, M. Paegert, E. Palle, F. Pepe, J. Pepper, A. Quirrenbach, S. A. Rinehart, D. Sasselov, B. Sato, S. Seager, A. Sozzetti, K. G. Stassun, P. Sullivan, A. Szentgyorgyi, G. Torres, S. Udry and J. Villasenor, *Transiting Exoplanet Survey Satellite (TESS)*, Journal of Astronomical Telescopes, Instruments, and Systems **1**, 014003 (2015).
- [152] P. de Laverny, A. Recio-Blanco, C. C. Worley and B. Plez, *The AMBRE project: A new synthetic grid of high-resolution FGKM stellar spectra*, **544**, A126 (2012).
- [153] J. A. Valenti and N. Piskunov, *Spectroscopy made easy: A new tool for fitting observations with synthetic spectra.*, **118**, 595 (1996).
- [154] N. Piskunov and J. A. Valenti, *Spectroscopy Made Easy: Evolution*, **597**, A16 (2017).
- [155] B. Gustafsson, B. Edvardsson, K. Eriksson, U. G. Jørgensen, Å. Nordlund and B. Plez, *A grid of MARCS model atmospheres for late-type stars. I. Methods and general properties*, **486**, 951 (2008).
- [156] U. Munari, K. Cotar, V. Andreoli, S. Dallaporta and L. N. Yalyalieva, *Recent developments in optical spectra of nova AT 2019qwf (Nova V2891 Cyg)*, The Astronomer's Telegram **13340**, 1 (2019).
- [157] U. Munari, V. Joshi, D. P. K. Banerjee, K. Čotar, S. Y. Shugarov, Jurdana-Šepić, R. , R. Belligoli, A. Bergamini, M. Graziani, G. L. Righetti, A. Vagozzini and P. Valisa, *The 2018 eruption and long-term evolution of the new high-mass Herbig Ae/Be object Gaia-18azl = VES 263*, **488**, 5536 (2019).

- [158] K. Čotar, T. Zwitter, J. Kos, U. Munari, S. L. Martell, M. Asplund, J. Bland -Hawthorn, S. Buder, G. M. de Silva, K. C. Freeman, S. Sharma, B. Anguiano, D. Carollo, J. Horner, G. F. Lewis, D. M. Nataf, T. Nordlander, D. Stello, Y.-S. Ting, C. Tinney, G. Traven, R. A. Wittenmyer and Galah Collaboration, *The GALAH survey: a catalogue of carbon-enhanced stars and CEMP candidates*, **483**, 3196 (2019).
- [159] A. Secchi, *Schreiben des Herrn Professors Secchi an den Herausgeber*, Astronomische Nachrichten **73**, 129 (1869).
- [160] I. Iben, Jr., *Carbon star formation and neutron-rich isotope formation in low-mass asymptotic giant branch stars*, **275**, L65 (1983).
- [161] Z. Han, P. P. Eggleton, P. Podsiadlowski and C. A. Tout, *The formation of barium and CH stars and related objects*, **277**, 1443 (1995).
- [162] J. Yoon, T. C. Beers, V. M. Placco, K. C. Rasmussen, D. Carollo, S. He, T. T. Hansen, I. U. Roederer and J. Zeanah, *Observational Constraints on First-star Nucleosynthesis. I. Evidence for Multiple Progenitors of CEMP-No Stars*, **833**, 20 (2016).
- [163] P. Battinelli, S. Demers, C. Rossi and K. S. Gigoyan, *Extension of the C Star Rotation Curve of the Milky Way to 24 kpc*, Astrophysics **56**, 68 (2013).
- [164] G. Bothun, J. H. Elias, G. MacAlpine, K. Matthews, J. R. Mould, G. Neugebauer and I. N. Reid, *Carbon stars at high Galactic latitude*, **101**, 2220 (1991).
- [165] B. Margon, S. F. Anderson, H. C. Harris, M. A. Strauss, G. R. Knapp, X. Fan, D. P. Schneider, D. E. Vanden Berk, D. J. Schlegel, E. W. Deutsch, Ž. Ivezić, P. B. Hall, B. F. Williams, A. F. Davidsen, J. Brinkmann, I. Csabai, J. J. E. Hayes, G. Hennessy, E. K. Kinney, S. J. Kleinman, D. Q. Lamb, D. Long, E. H. Neilsen, R. Nichol, A. Nitta, S. A. Snedden and D. G. York, *Faint High-Latitude Carbon Stars Discovered by the Sloan Digital Sky Survey: Methods and Initial Results*, **124**, 1651 (2002).
- [166] R. A. Downes, B. Margon, S. F. Anderson, H. C. Harris, G. R. Knapp, J. Schroeder, D. P. Schneider, D. G. York, J. R. Pier and J. Brinkmann, *Faint High-Latitude Carbon Stars Discovered by the Sloan Digital Sky Survey: An Initial Catalog*, **127**, 2838 (2004).
- [167] R. F. Griffin and R. O. Redman, *Photoelectric measurements of the $\lambda 4200\text{ \AA}$ CN band and the G band in G8-K5 spectra*, **120**, 287 (1960).
- [168] R. D. McClure and S. van den Bergh, *Five-color intermediate-band photometry of stars, clusters, and galaxies.*, **73**, 313 (1968).
- [169] L. Häggkvist and T. Oja, *Narrow-band photometry of late-type stars*, **1**, 199 (1970).
- [170] D. Moro and U. Munari, *The Asiago Database on Photometric Systems (ADPS). I. Census parameters for 167 photometric systems*, **147**, 361 (2000).

Bibliography

- [171] M. Fiorucci and U. Munari, *The Asiago Database on Photometric Systems (ADPS). II. Band and reddening parameters*, **401**, 781 (2003).
- [172] N. Christlieb, P. J. Green, L. Wisotzki and D. Reimers, *The stellar content of the Hamburg/ESO survey II. A large, homogeneously-selected sample of high latitude carbon stars*, **375**, 366 (2001).
- [173] P. Green, *Innocent Bystanders: Carbon Stars from the Sloan Digital Sky Survey*, **765**, 12 (2013).
- [174] Y. S. Lee, T. C. Beers, T. Masseron, B. Plez, C. M. Rockosi, J. Sobeck, B. Yanny, S. Lucatello, T. Sivarani, V. M. Placco and D. Carollo, *Carbon-enhanced Metal-poor Stars in SDSS/SEGUE. I. Carbon Abundance Estimation and Frequency of CEMP Stars*, **146**, 132 (2013).
- [175] W. Ji, W. Cui, C. Liu, A. Luo, G. Zhao and B. Zhang, *Carbon Stars from LAMOST DR2 Data*, **226**, 1 (2016).
- [176] Y.-B. Li, A.-L. Luo, C.-D. Du, F. Zuo, M.-X. Wang, G. Zhao, B.-W. Jiang, H.-W. Zhang, C. Liu, L. Qin, R. Wang, B. Du, Y.-X. Guo, B. Wang, Z.-W. Han, M.-S. Xiang, Y. Huang, B.-Q. Chen, J.-J. Chen, X. Kong, W. Hou, Y.-H. Song, Y.-F. Wang, K.-F. Wu, J.-N. Zhang, Y. Zhang, Y.-F. Wang, Z.-H. Cao, Y.-H. Hou and Y.-H. Zhao, *Carbon Stars Identified from LAMOST DR4 Using Machine Learning*, **234**, 31 (2018).
- [177] J. E. Norris, S. G. Ryan and T. C. Beers, *Extremely Metal-poor Stars. IV. The Carbon-rich Objects*, **488**, 350 (1997).
- [178] W. Aoki, J. E. Norris, S. G. Ryan, T. C. Beers and H. Ando, *The Chemical Composition of Carbon-rich, Very Metal Poor Stars: A New Class of Mildly Carbon Rich Objects without Excess of Neutron-Capture Elements*, **567**, 1166 (2002).
- [179] R. Cayrel, E. Depagne, M. Spite, V. Hill, F. Spite, P. François, B. Plez, T. Beers, F. Primas, J. Andersen, B. Barbuy, P. Bonifacio, P. Molaro and B. Nordström, *First stars V - Abundance patterns from C to Zn and supernova yields in the early Galaxy*, **416**, 1117 (2004).
- [180] P. S. Barklem, N. Christlieb and T. C. Beers, *Metal-poor star abundances from the HERES project*, in *13th Cambridge Workshop on Cool Stars, Stellar Systems and the Sun*, ESA Special Publication, Vol. 560, edited by F. Favata, G. A. J. Hussain and B. Battrick (2005) p. 433.
- [181] J. G. Cohen, A. McWilliam, S. Shectman, I. Thompson, N. Christlieb, J. Melendez, S. Ramirez, A. Swensson and F.-J. Zickgraf, *Carbon Stars in the Hamburg/ESO Survey: Abundances*, **132**, 137 (2006).
- [182] W. Aoki, T. C. Beers, N. Christlieb, J. E. Norris, S. G. Ryan and S. Tsangarides, *Carbon-enhanced Metal-poor Stars. I. Chemical Compositions of 26 Stars*, **655**, 492 (2007).

- [183] J. E. Norris, N. Christlieb, A. J. Korn, K. Eriksson, M. S. Bessell, T. C. Beers, L. Wisotzki and D. Reimers, *HE 0557-4840: Ultra-Metal-Poor and Carbon-Rich*, **670**, 774 (2007).
- [184] J. K. Hollek, A. Frebel, I. U. Roederer, C. Sneden, M. Shetrone, T. C. Beers, S. Kang and C. Thom, *The Chemical Abundances of Stars in the Halo (CASH) Project. II. A Sample of 14 Extremely Metal-poor Stars*, **742**, 54 (2011).
- [185] D. Yong, J. E. Norris, M. S. Bessell, N. Christlieb, M. Asplund, T. C. Beers, P. S. Barklem, A. Frebel and S. G. Ryan, *The Most Metal-poor Stars. II. Chemical Abundances of 190 Metal-poor Stars Including 10 New Stars with [Fe/H] = -3.5*, **762**, 26 (2013).
- [186] I. U. Roederer, G. W. Preston, I. B. Thompson, S. A. Shectman, C. Sneden, G. S. Burley and D. D. Kelson, *A Search for Stars of Very Low Metal Abundance. VI. Detailed Abundances of 313 Metal-poor Stars*, **147**, 136 (2014).
- [187] T. Hansen, C. J. Hansen, N. Christlieb, T. C. Beers, D. Yong, M. S. Bessell, A. Frebel, A. E. García Pérez, V. M. Placco, J. E. Norris and M. Asplund, *An Elemental Assay of Very, Extremely, and Ultra-metal-poor Stars*, **807**, 173 (2015).
- [188] H. R. Jacobson, S. Keller, A. Frebel, A. R. Casey, M. Asplund, M. S. Bessell, G. S. Da Costa, K. Lind, A. F. Marino, J. E. Norris, J. M. Peña, B. P. Schmidt, P. Tisserand, J. M. Walsh, D. Yong and Q. Yu, *High-Resolution Spectroscopic Study of Extremely Metal-Poor Star Candidates from the SkyMapper Survey*, **807**, 171 (2015).
- [189] J. G. Cohen, N. Christlieb, I. Thompson, A. McWilliam, S. Shectman, D. Reimers, L. Wisotzki and E. Kirby, *Normal and Outlying Populations of the Milky Way Stellar Halo at [Fe/H] -2*, **778**, 56 (2013).
- [190] T. C. Beers, G. W. Preston and S. A. Shectman, *A search for stars of very low metal abundance. II*, **103**, 1987 (1992).
- [191] S. Rossi, T. C. Beers and C. Sneden, *Carbon Abundances for Metal-Poor Stars Based on Medium-Resolution Spectra*, in *The Third Stromlo Symposium: The Galactic Halo*, Astronomical Society of the Pacific Conference Series, Vol. 165, edited by B. K. Gibson, R. S. Axelrod and M. E. Putman (1999) p. 264.
- [192] J. G. Cohen, S. Shectman, I. Thompson, A. McWilliam, N. Christlieb, J. Melendez, F.-J. Zickgraf, S. Ramírez and A. Swenson, *The Frequency of Carbon Stars among Extremely Metal-poor Stars*, **633**, L109 (2005).
- [193] S. Lucatello, S. Tsangarides, T. C. Beers, E. Carretta, R. G. Gratton and S. G. Ryan, *The Binary Frequency Among Carbon-enhanced, s-Process-rich, Metal-poor Stars*, **625**, 825 (2005).
- [194] S. Rossi, T. C. Beers, C. Sneden, T. Sevastyanenko, J. Rhee and B. Marsteller, *Estimation of Carbon Abundances in Metal-Poor Stars. I. Application to the Strong G-Band Stars of Beers, Preston, and Shectman*, **130**, 2804 (2005).

Bibliography

- [195] A. Frebel, N. Christlieb, J. E. Norris, T. C. Beers, M. S. Bessell, J. Rhee, C. Fechner, B. Marsteller, S. Rossi, C. Thom, L. Wisotzki and D. Reimers, *Bright Metal-poor Stars from the Hamburg/ESO Survey. I. Selection and Follow-up Observations from 329 Fields*, **652**, 1585 (2006).
- [196] B. E. Marsteller, *The frequency of carbon-enhanced metal-poor stars and the origin of carbon in the universe*, Ph.D. thesis, Michigan State University (2007).
- [197] D. Carollo, T. C. Beers, J. Bovy, T. Sivarani, J. E. Norris, K. C. Freeman, W. Aoki, Y. S. Lee and C. R. Kennedy, *Carbon-enhanced Metal-poor Stars in the Inner and Outer Halo Components of the Milky Way*, **744**, 195 (2012).
- [198] D. Yong, J. E. Norris, M. S. Bessell, N. Christlieb, M. Asplund, T. C. Beers, P. S. Barklem, A. Frebel and S. G. Ryan, *The Most Metal-poor Stars. III. The Metallicity Distribution Function and Carbon-enhanced Metal-poor Fraction*, **762**, 27 (2013).
- [199] V. M. Placco, A. Frebel, T. C. Beers and R. J. Stancliffe, *Carbon-enhanced Metal-poor Star Frequencies in the Galaxy: Corrections for the Effect of Evolutionary Status on Carbon Abundances*, **797**, 21 (2014).
- [200] J. Yoon, T. C. Beers, S. Dietz, Y. S. Lee, V. M. Placco, G. Da Costa, S. Keller, C. I. Owen and M. Sharma, *Galactic Archeology with the AEGIS Survey: The Evolution of Carbon and Iron in the Galactic Halo*, **861**, 146 (2018).
- [201] T. Suda, M. Aikawa, M. N. Machida, M. Y. Fujimoto and I. Iben, Jr., *Is HE 0107-5240 A Primordial Star? The Characteristics of Extremely Metal-Poor Carbon-Rich Stars*, **611**, 476 (2004).
- [202] E. Starkenburg, M. D. Shetrone, A. W. McConnachie and K. A. Venn, *Binarity in carbon-enhanced metal-poor stars*, **441**, 1217 (2014).
- [203] H. Umeda and K. Nomoto, *First-generation black-hole-forming supernovae and the metal abundance pattern of a very iron-poor star*, **422**, 871 (2003).
- [204] H. Umeda and K. Nomoto, *Variations in the Abundance Pattern of Extremely Metal-Poor Stars and Nucleosynthesis in Population III Supernovae*, **619**, 427 (2005).
- [205] N. Tominaga, N. Iwamoto and K. Nomoto, *Abundance Profiling of Extremely Metal-poor Stars and Supernova Properties in the Early Universe*, **785**, 98 (2014).
- [206] P. Banerjee, Y.-Z. Qian and A. Heger, *s-Process in Massive Carbon-Enhanced Metal-Poor Stars*, [10.1093/mnras/sty2251](https://doi.org/10.1093/mnras/sty2251) (2018).
- [207] D. Carollo, K. Freeman, T. C. Beers, V. M. Placco, J. Tumlinson and S. L. Martell, *Carbon-enhanced Metal-poor Stars: CEMP-s and CEMP-no Subclasses in the Halo System of the Milky Way*, **788**, 180 (2014).

- [208] M. A. Cruz, H. Cogo-Moreira and S. Rossi, *Searching for chemical classes among metal-poor stars using medium-resolution spectroscopy*, **475**, 4781 (2018).
- [209] T. C. Beers and N. Christlieb, *The Discovery and Analysis of Very Metal-Poor Stars in the Galaxy*, **43**, 531 (2005).
- [210] M. Spite, E. Caffau, P. Bonifacio, F. Spite, H.-G. Ludwig, B. Plez and N. Christlieb, *Carbon-enhanced metal-poor stars: the most pristine objects?*, **552**, A107 (2013).
- [211] J. K. Hollek, A. Frebel, V. M. Placco, A. I. Karakas, M. Shetrone, C. Sneden and N. Christlieb, *The Chemical Abundances of Stars in the Halo (CASH) Project. III. A New Classification Scheme for Carbon-enhanced Metal-poor Stars with s-process Element Enhancement*, **814**, 121 (2015).
- [212] L. Duong, K. C. Freeman, M. Asplund, L. Casagrande, S. Buder, K. Lind, M. Ness, J. Bland-Hawthorn, G. M. De Silva, V. D’Orazi, J. Kos, G. F. Lewis, J. Lin, S. L. Martell, K. Schlesinger, S. Sharma, J. D. Simpson, D. B. Zucker, T. Zwitter, B. Anguiano, G. S. Da Costa, E. Hyde, J. Horner, P. R. Kafle, D. M. Nataf, W. Reid, D. Stello, Y.-S. Ting and R. F. G. Wyse, *The GALAH survey: properties of the Galactic disc(s) in the solar neighbourhood*, **476**, 5216 (2018).
- [213] R. A. Wittenmyer, S. L. Martell, J. Esdaile, S. Sharma and D. Stello, *The HERMES K2 Follow-up Program at the Anglo-Australian Telescope*, in *American Astronomical Society Meeting Abstracts #227*, American Astronomical Society Meeting Abstracts, Vol. 227 (2016) p. 138.16.
- [214] R. C. Johnson, *The Structure and Origin of the Swan Band Spectrum of Carbon*, Philosophical Transactions of the Royal Society of London Series A **226**, 157 (1927).
- [215] L. van der Maaten and G. Hinton, *Visualizing Data using t-SNE*, Journal of Machine Learning Research **9**, 2579 (2008).
- [216] J. Spurauskas, L. Začs, W. J. Schuster and V. Deveikis, *The Binary Nature of CH-Like Stars*, **826**, 85 (2016).
- [217] J. Bergeat, A. Knapik and B. Rutily, *The pulsation modes and masses of carbon-rich long period variables*, **390**, 987 (2002).
- [218] T. Lloyd Evans, *Carbon stars*, Journal of Astrophysics and Astronomy **31**, 177 (2010).
- [219] P. Battinelli and S. Demers, *Variability of halo carbon stars*, **544**, A10 (2012).
- [220] P. Battinelli and S. Demers, *The variability of carbon stars in the Sagittarius dwarf spheroidal galaxy*, **553**, A93 (2013).
- [221] P. Battinelli and S. Demers, *Miras among C stars*, **568**, A100 (2014).

Bibliography

- [222] B. Margon, T. Kupfer, K. Burdge, T. A. Prince, S. R. Kulkarni and D. L. Shupe, *The Binary Dwarf Carbon Star SDSS J125017.90+252427.6*, **856**, L2 (2018).
- [223] D. S. P. Dearborn, J. Liebert, M. Aaronson, C. C. Dahn, R. Harrington, J. Mould and J. L. Greenstein, *On the nature of the dwarf carbon star G77-61*, **300**, 314 (1986).
- [224] L. J. Whitehouse, J. Farihi, P. J. Green, T. G. Wilson and J. P. Subasavage, *Dwarf carbon stars are likely metal-poor binaries and unlikely hosts to carbon planets*, **479**, 3873 (2018).
- [225] J. Farihi, A. R. Arendt, H. S. Machado and L. J. Whitehouse, *Evidence for halo kinematics among cool carbon-rich dwarfs*, **477**, 3801 (2018).
- [226] T. T. Hansen, J. Andersen, B. Nordström, T. C. Beers, V. M. Placco, J. Yoon and L. A. Buchhave, *The role of binaries in the enrichment of the early Galactic halo. II. Carbon-enhanced metal-poor stars: CEMP-no stars*, **586**, A160 (2016).
- [227] C. Abia, H. M. J. Boffin, J. Isern and R. Rebolo, *IY Hya - A new super Li-rich carbon star*, **245**, L1 (1991).
- [228] I.-J. Sackmann, R. L. Smith and K. H. Despain, *Carbon and eruptive stars: surface enrichment of lithium, carbon, nitrogen, and ^{13}C by deep mixing.*, **187**, 555 (1974).
- [229] P. C. Keenan and W. W. Morgan, *The Classification of the Red Carbon Stars.*, **94**, 501 (1941).
- [230] C. Barnbaum, R. P. S. Stone and P. C. Keenan, *A Moderate-Resolution Spectral Atlas of Carbon Stars: R, J, N, CH, and Barium Stars*, **105**, 419 (1996).
- [231] A. Alksnis, A. Balklavs, U. Dzervitis, I. Eglitis, O. Paupers and I. Pundure, *General Catalog of Galactic Carbon Stars by C. B. Stephenson. Third Edition*, Baltic Astronomy **10**, 1 (2001).
- [232] Y. Komiya, T. Suda, H. Minaguchi, T. Shigeyama, W. Aoki and M. Y. Fujimoto, *The Origin of Carbon Enhancement and the Initial Mass Function of Extremely Metal-poor Stars in the Galactic Halo*, **658**, 367 (2007).
- [233] T. Masseron, J. A. Johnson, B. Plez, S. van Eck, F. Primas, S. Goriely and A. Jorissen, *A holistic approach to carbon-enhanced metal-poor stars*, **509**, A93 (2010).
- [234] V. M. Placco, C. R. Kennedy, S. Rossi, T. C. Beers, Y. S. Lee, N. Christlieb, T. Sivarani, D. Reimers and L. Wisotzki, *A Search for Unrecognized Carbon-Enhanced Metal-Poor Stars in the Galaxy*, **139**, 1051 (2010).
- [235] C. Abate, O. R. Pols, R. G. Izzard and A. I. Karakas, *Carbon-enhanced metal-poor stars: a window on AGB nucleosynthesis and binary evolution. II. Statistical analysis of a sample of 67 CEMP-s stars*, **581**, A22 (2015).

- [236] J. Yoon, T. C. Beers, V. M. Placco, K. C. Rasmussen, D. Carollo, S. He, T. T. Hansen, I. U. Roederer and J. Zeanah, *VizieR Online Data Catalog: Carbon-enhanced metal-poor (CEMP) star abundances (Yoon+, 2016)*, VizieR Online Data Catalog **183** (2017).
- [237] Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Eyer and et al., *Gaia Data Release 2. Summary of the contents and survey properties*, **616**, A1 (2018).
- [238] J. Bovy, *galpy: A python Library for Galactic Dynamics*, **216**, 29 (2015).
- [239] H. Koppelman, A. Helmi and J. Veljanoski, *One Large Blob and Many Streams Frosting the nearby Stellar Halo in GaiaDR2*, **860**, L11 (2018).
- [240] R. D. McClure and A. W. Woodsworth, *The binary nature of the barium and CH stars. III - Orbital parameters*, **352**, 709 (1990).
- [241] A. Jorissen, S. Van Eck, H. Van Winckel, T. Merle, H. M. J. Boffin, J. Andersen, B. Nordström, S. Udry, T. Masseron, L. Lenaerts and C. Waelkens, *Binary properties of CH and carbon-enhanced metal-poor stars*, **586**, A158 (2016).
- [242] D. Raghavan, H. A. McAlister, T. J. Henry, D. W. Latham, G. W. Marcy, B. D. Mason, D. R. Gies, R. J. White and T. A. ten Brummelaar, *A Survey of Stellar Families: Multiplicity of Solar-type Stars*, **190**, 1 (2010).
- [243] G. Duchêne and A. Kraus, *Stellar Multiplicity*, **51**, 269 (2013).
- [244] M. Moe and R. Di Stefano, *Mind Your Ps and Qs: The Interrelation between Period (P) and Mass-ratio (Q) Distributions of Binary Stars*, **230**, 15 (2017).
- [245] A. Tokovinin, *From Binaries to Multiples. II. Hierarchical Multiplicity of F and G Dwarfs*, **147**, 87 (2014).
- [246] C. F. Quist and L. Lindegren, *Statistics of Hipparcos binaries: probing the 1-10 AU separation range*, **361**, 770 (2000).
- [247] D. Pourbaix, A. A. Tokovinin, A. H. Batten, F. C. Fekel, W. I. Hartkopf, H. Levato, N. I. Morrell, G. Torres and S. Udry, *$S_{B⁹}$: The ninth catalogue of spectroscopic binary orbits*, **424**, 727 (2004).
- [248] M. A. Fernandez, K. R. Covey, N. De Lee, S. D. Chojnowski, D. Nidever, R. Ballantyne, M. Cottaar, N. Da Rio, J. B. Foster, S. R. Majewski, M. R. Meyer, A. M. Reyna, G. W. Roberts, J. Skinner, K. Stassun, J. C. Tan, N. Troup and G. Zasowski, *IN-SYNC VI. Identification and Radial Velocity Extraction for 100+ Double-Lined Spectroscopic Binaries in the APOGEE/IN-SYNC Fields*, **129**, 084201 (2017).
- [249] T. Merle, S. Van Eck, A. Jorissen, M. Van der Swaelmen, T. Masseron, T. Zwitter, D. Hatzidimitriou, A. Klutsch, D. Pourbaix, R. Blomme, C. C. Worley, G. Sacco, J. Lewis, C. Abia, G. Traven, R. Sordo, A. Bragaglia,

Bibliography

- R. Smiljanic, E. Pancino, F. Damiani, A. Hourihane, G. Gilmore, S. Randich, S. Koposov, A. Casey, L. Morbidelli, E. Franciosini, L. Magrini, P. Jofre, M. T. Costado, R. D. Jeffries, M. Bergemann, A. C. Lanzafame, A. Bayo, G. Carrao, E. Flaccomio, L. Monaco and S. Zaggia, *The Gaia-ESO Survey: double-, triple-, and quadruple-line spectroscopic binary candidates*, **608**, A95 (2017).
- [250] A. Duquennoy and M. Mayor, *Multiplicity among solar-type stars in the solar neighbourhood. II - Distribution of the orbital elements in an unbiased sample*, **248**, 485 (1991).
- [251] B. Nordström, M. Mayor, J. Andersen, J. Holmberg, F. Pont, B. R. Jørgensen, E. H. Olsen, S. Udry and N. Mowlavi, *The Geneva-Copenhagen survey of the Solar neighbourhood. Ages, metallicities, and kinematic properties of 14 000 F and G dwarfs*, **418**, 989 (2004).
- [252] G. Matijević, T. Zwitter, O. Bienaymé, J. Bland-Hawthorn, K. C. Freeman, G. Gilmore, E. K. Grebel, A. Helmi, U. Munari, J. F. Navarro, Q. A. Parker, W. Reid, G. M. Seabroke, A. Siebert, A. Siviero, M. Steinmetz, F. G. Watson, M. Williams and R. F. G. Wyse, *Single-lined Spectroscopic Binary Star Candidates in the RAVE Survey*, **141**, 200 (2011).
- [253] N. W. Troup, D. L. Nidever, N. De Lee, J. Carlberg, S. R. Majewski, M. Fernandez, K. Covey, S. D. Chojnowski, J. Pepper, D. T. Nguyen, K. Stassun, D. C. Nguyen, J. P. Wisniewski, S. W. Fleming, D. Bizyaev, P. M. Frinchaboy, D. A. García-Hernández, J. Ge, F. Hearty, S. Meszaros, K. Pan, C. Allende Prieto, D. P. Schneider, M. D. Shetrone, M. F. Skrutskie, J. Wilson and O. Zamora, *Companions to APOGEE Stars. I. A Milky Way-spanning Catalog of Stellar and Substellar Companion Candidates and Their Diverse Hosts*, **151**, 85 (2016).
- [254] C. Badenes, C. Mazzola, T. A. Thompson, K. Covey, P. E. Freeman, M. G. Walker, M. Moe, N. Troup, D. Nidever, C. Allende Prieto, B. Andrews, R. H. Barbá, T. C. Beers, J. Bovy, J. K. Carlberg, N. De Lee, J. Johnson, H. Lewis, S. R. Majewski, M. Pinsonneault, J. Sobeck, K. G. Stassun, G. S. Stringfellow and G. Zasowski, *Stellar Multiplicity Meets Stellar Evolution and Metallicity: The APOGEE View*, **854**, 147 (2018).
- [255] P. M. Garnavich, *Wide binary stars at the Galactic poles*, **335**, L47 (1988).
- [256] L. M. Close, H. B. Richer and D. R. Crabtree, *A complete sample of wide binaries in the solar neighborhood*, **100**, 1968 (1990).
- [257] A. Gould, J. N. Bahcall, D. Maoz and B. Yanny, *Star counts from the HST Snapshot Survey. 2: Wide binaries*, **441**, 200 (1995).
- [258] A. L. Kraus and L. A. Hillenbrand, *Unusually Wide Binaries: Are They Wide or Unusual?*, **703**, 1511 (2009).
- [259] E. J. Shaya and R. P. Olling, *Very Wide Binaries and Other Comoving Stellar Companions: A Bayesian Analysis of the Hipparcos Catalogue*, **192**, 2 (2011).

- [260] J. J. Andrews, J. Chanamé and M. A. Agüeros, *Wide binaries in Tycho-Gaia: search method and the distribution of orbital separations*, **472**, 675 (2017).
- [261] J. Coronado, M. P. Sepúlveda, A. Gould and J. Chanamé, *A distant sample of halo wide binaries from SDSS*, **480**, 4302 (2018).
- [262] S. Oh, A. M. Price-Whelan, D. W. Hogg, T. D. Morton and D. N. Spergel, *Comoving Stars in Gaia DR1: An Abundance of Very Wide Separation Comoving Pairs*, **153**, 257 (2017).
- [263] F. M. Jiménez-Esteban, E. Solano and C. Rodrigo, *A Catalog of Wide Binary and Multiple Systems of Bright Stars from Gaia-DR2 and the Virtual Observatory*, **157**, 78 (2019).
- [264] A. Skopal, *Disentangling the composite continuum of symbiotic binaries. I. S-type systems*, **440**, 995 (2005).
- [265] A. Rebassa-Mansergas, B. T. Gänsicke, P. Rodríguez-Gil, M. R. Schreiber and D. Koester, *Post-common-envelope binaries from SDSS - I. 101 white dwarf main-sequence binaries with multiple Sloan Digital Sky Survey spectroscopy*, **382**, 1377 (2007).
- [266] A. Rebassa-Mansergas, A. Nebot Gómez-Morán, M. R. Schreiber, B. T. Gänsicke, A. Schwone, J. Gallardo and D. Koester, *Post-common envelope binaries from SDSS - XIV. The DR7 white dwarf-main-sequence binary catalogue*, **419**, 806 (2012).
- [267] J. Ren, A. Luo, Y. Li, P. Wei, J. Zhao, Y. Zhao, Y. Song and G. Zhao, *White-dwarf-Main-sequence Binaries Identified from the LAMOST Pilot Survey*, **146**, 82 (2013).
- [268] A. Rebassa-Mansergas, J. J. Ren, S. G. Parsons, B. T. Gänsicke, M. R. Schreiber, E. García-Berro, X.-W. Liu and D. Koester, *The SDSS spectroscopic catalogue of white dwarf-main-sequence binaries: new identifications from DR 9-12*, **458**, 3808 (2016).
- [269] C. Bergmann, M. Endl, J. B. Hearnshaw, R. A. Wittenmyer and D. J. Wright, *Searching for Earth-mass planets around α Centauri: precise radial velocities from contaminated spectra*, International Journal of Astrobiology **14**, 173 (2015).
- [270] F. R. Ferraro, E. Carretta, F. Fusi Pecci and A. Zamboni, *Binary stars in globular clusters: detection of a binary sequence in NGC 2808?*, **327**, 598 (1997).
- [271] J. Hurley and C. A. Tout, *The binary second sequence in cluster colour-magnitude diagrams*, **300**, 977 (1998).
- [272] A. P. Milone, A. F. Marino, L. R. Bedin, A. Dotter, H. Jerjen, D. Kim, D. Nardiello, G. Piotto and J. Cong, *The binary populations of eight globular clusters in the outer halo of the Milky Way*, **455**, 3009 (2016).

Bibliography

- [273] M. Žerjal, M. J. Ireland, T. Nordlander, J. Lin, S. Buder, L. Casagrande, K. Čotar, G. De Silva, J. Horner, S. Martell, G. Traven and T. Zwitter, *The GALAH Survey: lithium-strong KM dwarfs*, **484**, 4591 (2019).
- [274] C. N. A. Willmer, *The Absolute Magnitude of the Sun in Several Filters*, **236**, 47 (2018).
- [275] T. Zwitter, J. Kos, A. Chiavassa, S. Buder, G. Traven, K. Čotar, J. Lin, M. Asplund, J. Bland-Hawthorn, A. R. Casey, G. De Silva, L. Duong, K. C. Freeman, K. Lind, S. Martell, V. D’Orazi, K. J. Schlesinger, J. D. Simpson, S. Sharma, D. B. Zucker, B. Anguiano, L. Casagrande, R. Collet, J. Horner, M. J. Ireland, P. R. Kafle, G. Lewis, U. Munari, D. M. Nataf, M. Ness, T. Nordlander, D. Stello, Y.-S. Ting, C. G. Tinney, F. Watson, R. A. Wittenmyer and M. Žerjal, *The GALAH survey: accurate radial velocities and library of observed stellar template spectra*, **481**, 645 (2018).
- [276] A. A. Henden, S. Levine, D. Terrell and D. L. Welch, *APASS - The Latest Data Release*, in *American Astronomical Society Meeting Abstracts #225*, American Astronomical Society Meeting Abstracts, Vol. 225 (2015) p. 336.16.
- [277] E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, S. A. Stanford, M. Cohen, R. G. Walker, J. C. Mather, D. Leisawitz, T. N. Gautier, III, I. McLean, D. Benford, C. J. Lonsdale, A. Blain, B. Mendez, W. R. Irace, V. Duval, F. Liu, D. Royer, I. Heinrichsen, J. Howard, M. Shannon, M. Kendall, A. L. Walsh, M. Larsen, J. G. Cardon, S. Schick, M. Schwalm, M. Abid, B. Fabinsky, L. Naes and C.-W. Tsai, *The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance*, **140**, 1868 (2010).
- [278] G. N. Lance and W. T. Williams, *Mixed-Data Classificatory Programs I - Agglomerative Systems.*, Australian Computer Journal **1**, 15 (1967).
- [279] D. W. Evans, M. Riello, F. De Angeli, J. M. Carrasco, P. Montegriffo, C. Fabricius, C. Jordi, L. Palaversa, C. Diener, G. Busso, C. Cacciari, F. van Leeuwen, P. W. Burgess, M. Davidson, D. L. Harrison, S. T. Hodgkin, E. Pancino, P. J. Richards, G. Altavilla, L. Balaguer-Núñez, M. A. Barstow, M. Bellazzini, A. G. A. Brown, M. Castellani, G. Cocozza, F. De Luise, A. Delgado, C. Ducourant, S. Galleti, G. Gilmore, G. Giuffrida, B. Holl, A. Kewley, S. E. Koposov, S. Marinoni, P. M. Marrese, P. J. Osborne, A. Piersimoni, J. Portell, L. Pulone, S. Ragagni, N. Sanna, D. Terrett, N. A. Walton, T. Wevers and Ł. Wyrzykowski, *Gaia Data Release 2. Photometric content and validation*, **616**, A4 (2018).
- [280] L. Casagrande and D. A. VandenBerg, *On the use of Gaia magnitudes and new tables of bolometric corrections*, **479**, L102 (2018).
- [281] C. A. L. Bailer-Jones, J. Rybizki, M. Fouesneau, G. Mantelet and R. Andrae, *Estimating Distance from Parallaxes. IV. Distances to 1.33 Billion Stars in Gaia Data Release 2*, **156**, 58 (2018).

- [282] L. Capitanio, R. Lallement, J. L. Vergely, M. Elyajouri and A. Monreal-Ibero, *Three-dimensional mapping of the local interstellar medium with composite data*, **606**, A65 (2017).
- [283] E. F. Schlafly and D. P. Finkbeiner, *Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD*, **737**, 103 (2011).
- [284] P. Marigo, L. Girardi, A. Bressan, P. Rosenfield, B. Aringer, Y. Chen, M. Dussin, A. Nanni, G. Pastorelli, T. S. Rodrigues, M. Trabucchi, S. Bladh, J. Dalcanton, M. A. T. Groenewegen, J. Montalbán and P. R. Wood, *A New Generation of PARSEC-COLIBRI Stellar Isochrones Including the TP-AGB Phase*, **835**, 77 (2017).
- [285] H. Kamdar, C. Conroy, Y.-S. Ting, A. Bonaca, B. Johnson and P. Cargile, *A Dynamical Model for Clustered Star Formation in the Galactic Disk*, arXiv e-prints , arXiv:1902.10719 (2019).
- [286] H. Kamdar, C. Conroy, Y.-S. Ting, A. Bonaca, M. Smith and A. G. A. Brown, *Stars that Move Together Were Born Together*, arXiv e-prints , arXiv:1904.02159 (2019).
- [287] D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, **125**, 306 (2013).
- [288] P. Eggleton, *Evolutionary Processes in Binary and Multiple Stars, by Peter Eggleton*, pp. . ISBN 0521855578. Cambridge, UK: Cambridge University Press, 2006. (2006).
- [289] A. Tokovinin, *Comparative statistics and origin of triple and quadruple stars*, **389**, 925 (2008).
- [290] A. Tokovinin, *The Updated Multiple Star Catalog*, **235**, 6 (2018).
- [291] A. A. Tokovinin, *On the origin of binaries with twin components*, **360**, 997 (2000).
- [292] F. Arenou, X. Luri, C. Babusiaux, C. Fabricius, A. Helmi, T. Muraveva, A. C. Robin, F. Spoto, A. Vallenari, T. Antoja, T. Cantat-Gaudin, C. Jordi, N. Leclerc, C. Reylé, M. Romero-Gómez, I.-C. Shih, S. Soria, C. Barache, D. Bossini, A. Bragaglia, M. A. Breddels, M. Fabrizio, S. Lambert, P. M. Marrese, D. Massari, A. Moitinho, N. Robichon, L. Ruiz-Dern, R. Sordo, J. Veljanoski, L. Eyer, G. Jasniewicz, E. Pancino, C. Soubiran, A. Spagna, P. Tanga, C. Turon and C. Zurbach, *Gaia Data Release 2. Catalogue validation*, **616**, A17 (2018).
- [293] S. Sharma, J. Bland-Hawthorn, K. V. Johnston and J. Binney, *Galaxia: A Code to Generate a Synthetic Survey of the Milky Way*, **730**, 3 (2011).

Bibliography

Appendix A

Naslov prvega dodatka

Appendix A. Naslov prvega dodatka

Razširjeni povzetek v slovenskem jeziku

Razširjeni povzetek v slovenskem jeziku naj bo dolg vsaj 10 strani. Vključuje naj tudi slike, tabele in enačbe, ki so nujne za razumevanje besedila povzetka.

List of publications related to this doctoral thesis

Scientific papers that were published by the author of this doctoral thesis:

1. **K. Čotar**, T. Zwitter, G. Traven, J. Kos, M. Asplund, J. Bland-Hawthorn, S. Buder, V. D’Orazi, G. M. de Silva, J. Lin, S. L. Martell, S. Sharma, J. D. Simpson, D. B. Zucker, J. Horner, G. F. Lewis, T. Nordlander, Y.-S. Ting, R. A. Wittenmyer and Galah Collaboration, The GALAH survey: unresolved triple Sun-like stars discovered by the Gaia mission, *Monthly notices of the Royal Astronomical Society*, 487, 2474 (2019).
2. **K. Čotar**, T. Zwitter, J. Kos, U. Munari, S. L. Martell, M. Asplund, J. Bland-Hawthorn, S. Buder, G. M. de Silva, K. C. Freeman, S. Sharma, B. Anguiano, D. Carollo, J. Horner, G. F. Lewis, D. M. Nataf, T. Nordlander, D. Stello, Y.-S. Ting, C. Tinney, G. Traven, R. A. Wittenmyer and Galah Collaboration, The GALAH survey: a catalogue of carbon-enhanced stars and CEMP candidates, *Monthly notices of the Royal Astronomical Society*, 483, 3196 (2019).

Scientific papers that were co-authored by the author of this doctoral thesis and were heavily used during the production of this thesis and above listed papers:

1. S. Buder, M. Asplund, L. Duong, J. Kos, K. Lind, M. K. Ness, S. Sharma, J. Bland-Hawthorn, A. R. Casey, G. M. De Silva, V. D’Orazi, K. C. Freeman, G. F. Lewis, J. Lin, S. L. Martell, K. J. Schlesinger, J. D. Simpson, D. B. Zucker, T. Zwitter, A. M. Amarsi, B. Anguiano, D. Carollo, L. Casagrande, **K. Čotar**, P. L. Cottrell, G. Da Costa, X. D. Gao, M. R. Hayden, J. Horner, M. J. Ireland, P. R. Kafle, U. Munari, D. M. Nataf, T. Nordlander, D. Stello, Y.-S. Ting, G. Traven, F. Watson, R. A. Wittenmyer, R. F. G. Wyse, D. Yong, J. C. Zinn and M. Žerjal, The GALAH Survey: second data release, *Monthly notices of the Royal Astronomical Society*, 478, 4513 (2018).