Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

Master's Thesis

in Physics

Hot Subdwarf Stars as Light Bulbs to Trace the Interstellar Medium and Search for Circumstellar Material

Sebastian Weich

Supervisors: Prof. Dr. Ulrich Heber and Dr. Matti Dorsch

Dr. Karl-Remeis-Sternwarte, Astronomical Institute
Friedrich-Alexander-Universität Erlangen-Nürnberg

Submission date: September 30, 2025

Abstract

The interstellar medium (ISM) plays a central role for the evolution of galaxies, in particular for that of our home galaxy, the Milky Way. While its structure and distribution is well studied within the Galactic disk and the solar neighborhood, its properties at high Galactic latitudes and within the halo are poorly constrained. A reason for that is the lack of bright, distant background sources. This thesis presents hot subdwarf stars as excellent light bulbs to shine through the ISM, especially at high Galactic latitudes. These stars represent late stages of stellar evolution and are formed in binary interaction, for example in common envelope evolution or merger events. They are observable out to large distances of several kpc and due to their high temperatures, they do not intrinsically show the same absorption lines as the ISM in their spectra. This makes them to ideal background sources. Therefore, the main aim of this work is to probe the interstellar medium, preferentially at high Galactic latitudes, by using the light of such hot subdwarf stars. The second aim concerns the properties of the hot subdwarfs themselves. Residual common envelope material or merger debris is expected to remain bound around hot subdwarfs after their formation. It might be detectable as circumstellar material (CSM) the same way as the ISM. Hence, in addition this work investigates the potential presence of CSM around hot subdwarfs, which would provide insights into their formation channels.

The main constituents of the ISM are interstellar gas and dust. Interstellar gas can be probed by absorption of starlight visible in their spectra. In the optical, the most prominent absorption features are the Ca II H&K as well as Na I D1&D2 resonance lines. In the spectra of cooler stars these absorption lines are strong and separating the ISM components from the photospheric blends might be difficult. An advantage of hot subdwarf stars are their high effective temperatures of $T_{\rm eff} > 20,000$ K. Calcium and sodium are ionized to higher ionization stages such that the Ca II H&K as well as Na I D1&D2 lines do not intrinsically appear. This facilitates the distinction between the photospheric and interstellar component. The strengths of the interstellar lines represent a proxy for the amount of interstellar gas along the line of sight. On the other hand, interstellar dust affects the starlight by extinction and reddening. By modeling stellar spectral energy distributions (SEDs) the reddening can be quantified as proxy for the amount of interstellar dust along the line of sight.

In this thesis, a comprehensive dataset of initially approximately 17,500 optical spectra was collected from large-scale low resolution surveys (LAMOST, $R \sim 1,800$; SDSS, $R \sim 2,000$; DESI, $R \sim 3,000$) and archival higher resolution observations (X-Shooter, $R \sim 9,800$; SPY/UVES, $R \sim 20,000$; FEROS, $R \sim 48,000$). By applying quality criteria the sample reduced to 6,400 spectra for about 3,500 unique objects. Due to the footprints of the large all-sky surveys, predominantly the Northern sky is covered by sight lines, mostly excluding the Galactic disk.

To carry out quantitative spectral analyses, a set of stellar Python scripts was newly developed. Spectral fits are necessary to use the photospheric spectrum as continuum for the ISM, which resolves possible blends. Furthermore, it includes the reading and processing of various kinds of spectra, an interpolation routine to linearly interpolate within model grids, and a script for the comprehensive analysis of interstellar absorption lines. The ISM line analysis comprises the determination of equivalent widths, column densities from Voigt profile fits, and radial velocities. The methods were tested on simulated data in order to quantify the effect of low spectral resolution and noise on the measured parameters. This lead to strict quality cuts on the real data and an adopted generic systematic uncertainty on derived column densities of $\Delta \log N_{\rm sys} = 0.5$.

As complementary part to the spectral analyses, the derived atmospheric parameters from the spectral fits are taken to model stellar SEDs. By including the reddening law of the Milky Way in the SED fitting routine, the amount of reddening from interstellar dust along the line of sight can be quantified. Of particular importance are well constrained effective temperatures due to a strong correlation between temperature and derived reddening. Moreover, if additionally *Gaia* parallaxes are provided for the stars, the SEDs allow to constrain the stellar parameters like mass, radius, and luminosity.

Therefore, as valuable side-effect the analyses of this work yield atmospheric and stellar parameters for all of the background sources. These parameters were checked for consistency and provide a huge dataset for the study of hot subdwarf stars themselves. However, this is not the purpose of this work and is therefore left as an outlook.

The investigation of the interstellar medium revealed several meaningful findings. On average the calcium and sodium column densities are $\log \bar{N}_{\mathrm{Ca\,II}} = 12.7^{+0.6}_{-0.5}$ and $\log \bar{N}_{\mathrm{Na\,I}} = 12.4^{+0.9}_{-0.5}$, respectively, while the average reddening is $\bar{E}(44-55) = 0.05^{+0.09}_{-0.04}$ mag. The confidence intervals rather represent the dispersion within the ISM than uncertainties. The average calcium to sodium column density ratio of $N_{\mathrm{Ca\,II}}/N_{\mathrm{Na\,I}} \approx 0.4$ indicates that in the covered volume of about 4 kpc around the Sun predominantly warm (partially) ionized medium with temperatures between about 2,500-13,000 K is present. This includes parts of the Galactic halo. Furthermore, compiled two dimensional interstellar gas and dust maps indicate that Na I and dust is much more closely confined to the Galactic disk than Ca II, which extends more uniformly up to higher Galactic latitudes. The precises parallaxes of the background sources provided by the *Gaia* mission allow to probe the ISM as a function of distance as well. The variations of the interstellar parameters with distance, direction, and height from the Galactic plane are consistent with the results from the sky maps. Finally, the connection between interstellar gas and dust, which are expected to coexist in interstellar clouds, was analyzed in the form of gas to dust ratios. Based on former attempts, a linear relation was fitted between the logarithm of reddening parameters and interstellar absorption line equivalent widths. This general relation can be, for instance, used to guess the expected reddening along certain sight lines, if no other reddening estimate is available. Moreover, the measured gas to dust ratios suggest that Na I is more closely connected with interstellar dust than Ca II. Na I and dust represent a major constituent of cold neutral medium, while Ca II represents one of warm (partially) ionized medium.

The second, equally important aim of this thesis is to search for CSM around hot subdwarf stars. The test runs of the scripts on simulated data have shown that measurements from low resolution spectra suffer from large uncertainties. For example, the radial velocities of the ISM lines are highly unconstrained with uncertainties of about $\Delta v_{\mathrm{rad,ISM}} \sim 30\,\mathrm{km\,s^{-1}}$. Still, a recent study claims the detection of ubiquitous circumstellar material around 145 hot subdwarfs based on low resolution LAMOST spectroscopy. Similar methods have been used for their detections as in this work and the analyzed spectra are also part of this sample. Therefore, the study was revisited and the results of both analyses were compared. It became apparent that many sources of uncertainty were formerly neglected. A more careful treatment reduced the sample to 49 candidates. However, they can still not be claimed as detection and high resolution follow-up is required to confirm or falsify them.

A dedicated search for CSM candidates based on 918 higher resolution spectra for 170 unique objects identified three promising candidates due to the appearance of Ca II H&K lines in their spectra, for which an interstellar or photospheric origin can be most likely excluded: The two helium-poor subdwarf B stars HD 205805 and EC 03591-3232 as well as the intermediate helium-rich OB subdwarf UVO 0825+15. All of them exhibit weak Ca II lines exactly at the radial velocity of the stars. Na I lines are completely absent, which is expected for material close to hot stars, where sodium is probably ionized. Assuming CSM, the measured calcium column densities of $\log N_{\mathrm{Ca\,II}} \lesssim 12$ indicate rather thin and diffuse material. The spectra of all candidates show many metal lines, which could indicate their recent formation. Despite possibly formed via merger events, they do not rotate. But the mostly slow rotation rates of hot subdwarfs are a puzzle for their formation scenarios in general. The atmospheric and stellar parameters of the candidates show no peculiarities. However, it is known that EC 03591-3232 shows a ^{3}He isotope anomaly and UVO 0825+15 is enriched with heavy elements and pulsates. In diagnostic diagrams, like the Kiel or Hertzsprung-Russell diagram, they are located on the extreme horizontal branch (HD 205805 and EC 03591-3232) or on the helium main sequence (UVO 0825+15). Final confirmation of the presence of CSM will require follow-up by even higher resolution spectroscopy, ultraviolet observations, or radio studies of neutral hydrogen.

In conclusion, this thesis provides insights into the properties and structure of the ISM up to

high Galactic latitudes. Future spectroscopic surveys will help to increase the coverage of sight lines on the sky. Furthermore, the search for circumstellar material from low resolution spectra failed to detect any evidence. Previous claims for detections around hot subdwarfs from LAMOST data are shown to be spurious. It is demonstrated that high resolution spectroscopy is a necessity for CSM detection and three candidates were found among 170 stars observed with high resolution spectrographs. Follow-up observations are required to confirm them.

Contents

Chapter 1

Introduction and Theoretical Background

1.1 Hot Subdwarfs as Light Bulbs to Trace the Interstellar Medium

All direct information we receive about the universe and its ingredients, like stars and galaxies, reach us in the form of messengers that traveled long distances until they arrive at the solar system and are recorded by us. Such messengers can be of various kind, for example small particles like protons, ions, electrons, neutrinos, etc. or electromagnetic radiation, i.e. light. In the optical wavelength regime, most of this light is emitted by stars, which we can observe in a clear night even with the naked eye. However, also the space between the stars is not empty. It is filled with the so-called interstellar medium (ISM), which primarily consists of gas and dust. These gas and dust particles are able to scatter and absorb light as it passes through. Therefore, they play a crucial role for many astronomical observations as they often heavily affect starlight before it reaches us as observer. But by no means the ISM is just an annoying factor disturbing our observations. Actually, it is arguably the most important component of galaxies, because it it responsible for the formation of stars that are the dominant sources of energy. As stars evolve, they expel material back into the ISM, for instance by stellar winds or supernova explosions. From this material a new generation of stars can be formed. This is called the cosmic cycle of matter. Therefore, the ISM is usually one of the major constituents of a galaxy, in particular of our home galaxy, the Milky Way, and determines the structure and evolution of the galaxy. Overall, it is thus desirable to understand the properties and the distribution of the ISM in great detail, not only to correct observations for interstellar absorption and extinction, but also to gain insights into the history and evolution of the Galaxy.

In the recent decades a lot of effort has been put into the compilation of precise maps tracing the ISM. For example, radio observations allow to map the distribution of hydrogen in the Galaxy (e.g., Kalberla & Kerp, 2009), infrared observations provide us with the distribution of dust (e.g., Schlegel et al., 1998; Schlafly & Finkbeiner, 2011; Green et al., 2019), CO observations with that of molecular gas (e.g., Dame et al., 2001), diffuse interstellar bands with that of (yet unknown) large molecules (Lan et al., 2014; Baron et al., 2015; Zasowski et al., 2015), and interstellar absorption with that of atomic gas apart from hydrogen[1] (e.g., Murga et al., 2015). An example dust map is shown in Figure 1.1. The most prominent feature is the Galactic disk at low Galactic latitudes, where most of the interstellar dust is concentrated. In contrast, almost no dust is found at higher Galactic latitudes. However, from radio observations of neutral hydrogen (Figure 1.2), we know that also the space between stars in the Galactic halo is not entirely empty. Still, it is poorly mapped and analyzed. A reason for that is the low number of available background sources in the Galactic halo that can act as light bulbs to shine through the ISM. By now, no active star formation is going on in the halo anymore and hence, all young and bright stars that burn their fuel rapidly have already ended their lives. On the other hand, older and fainter stars, which might still exist in the halo, can not be observed out to such great distances.

Here, the population of hot subdwarf stars comes into play. These hot ($T_{\rm eff} > 20,000\,{\rm K}$) but

[1] Surprisingly, even though metals are a fundamental component of the ISM, their distribution in the ISM is still poorly mapped. The gas maps of Murga et al. (2015) also cover only parts of the sky.

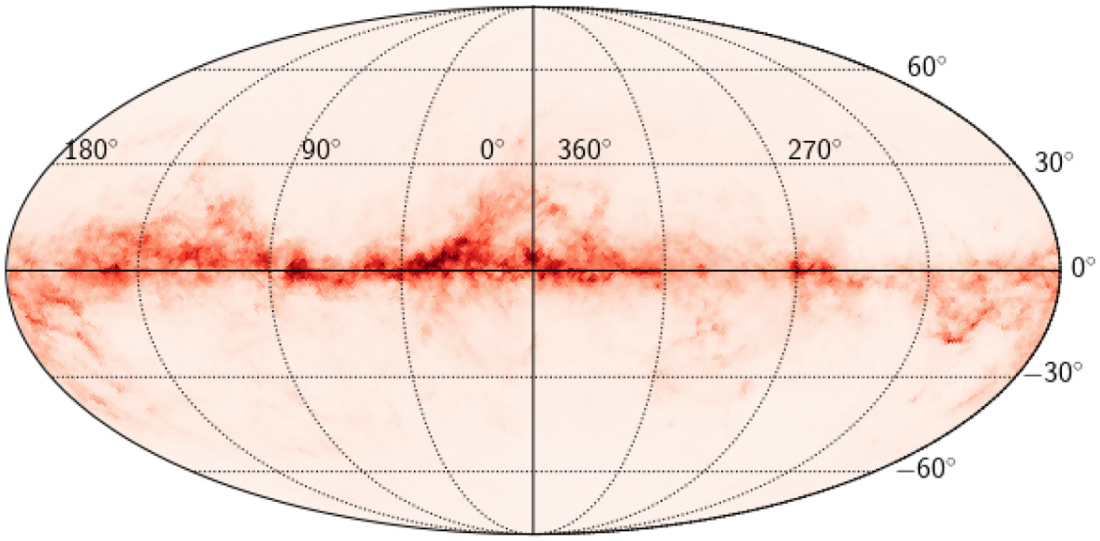


Figure 1.1: Example dust map in Galactic coordinates taken from Anders et al. (2022). It clearly shows that most of the interstellar dust is closely confined to the Galactic disk at low Galactic latitudes. Almost no features are visible towards higher Galactic latitudes.

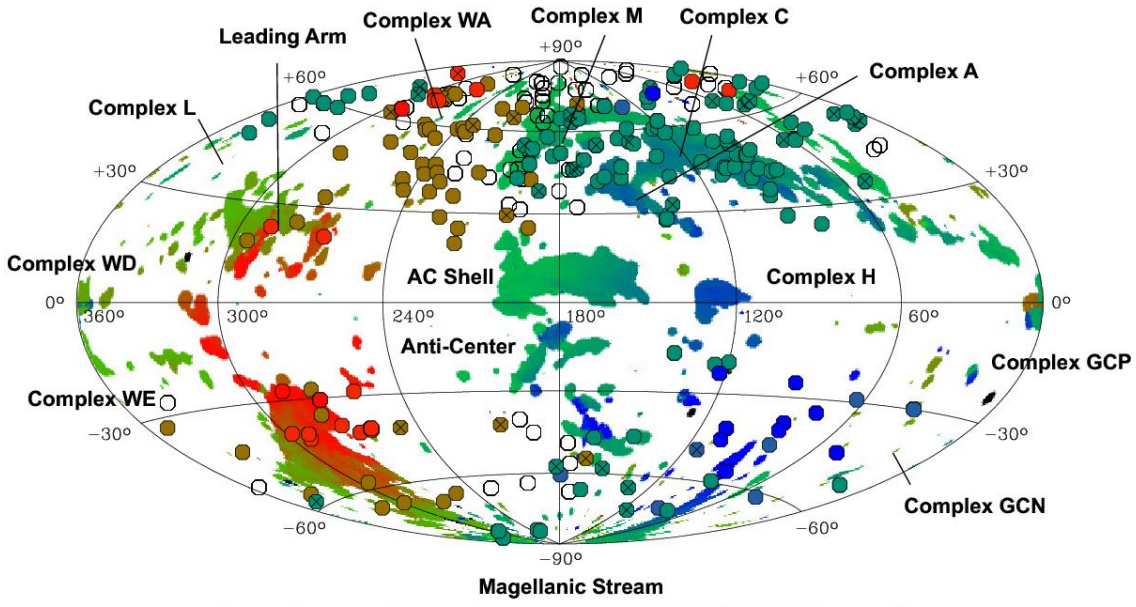


Figure 1.2: Map of neutral hydrogen derived from radio observations by Richter et al. (2017). Also higher Galactic latitudes are covered. However, due to a lack of distance estimates to the clouds, not all of them are necessarily of Galactic origin. While for example the complexes A, C, and M are associated with the Milky Way halo at an approximate distance of 5 kpc, the Magellanic stream is about 50 kpc away from us. For more details see Richter et al. (2017).

subluminous objects are typically less massive and smaller than the Sun. Their formation is not yet fully understood, but most likely includes close binary interaction, since they are not predicted by single star evolution. They are core helium burning stars on the extreme end of the horizontal branch, but due to their low masses they are less wasteful with their fuel compared to their main sequence counterparts of spectral types O and B. Therefore, they are still existent in the Galactic halo and are bright enough to be observed out to high distances of several kpc. Since the scale height of the Galactic disk is only a few hundred pc (Irrgang et al., 2013), these distances are sufficient to place high Galactic latitude objects in the halo. Thus, hot subdwarfs can be ideally utilized to act as light bulbs for the investigation of the ISM at high Galactic latitudes. Interstellar absorption and extinction, which usually hinders stellar observations, can be measured to quantify the amount of intervening interstellar material. This is one aim of this work.

To this end, the spectra of many hot subdwarf stars are analyzed. Apart from spectral lines originating from their atmosphere, they often show interstellar absorption lines as well, whose properties serve as proxy for the amount of intervening interstellar gas. Another advantage of the stars are their high temperatures. The same elements within the ISM that cause the absorption are also often present in stellar atmospheres, but at high temperatures ionized to higher ionization stages. Hence, hot subdwarf stars do not intrinsically contribute with those lines in their spectra unlike cooler stars and generally show mostly hydrogen and helium lines. This facilitates the identification of the interstellar absorption lines and the distinction from stellar photospheric lines.

As complementary analysis the spectral energy distributions (SEDs) of the stars can be modeled with the help of photometric data. This gives conclusion about interstellar extinction caused by dust particles in the ISM and quantifies the amount of intervening interstellar dust. With the results of both analyses, the properties of interstellar gas and dust can be traced and investigated for a large volume within the Milky Way.

A major side-effect of analyzing the spectra and modeling the SEDs of the hot subdwarf stars is that a lot of parameters and properties of the background sources themselves are measured. Their formation is thought to include stages of binary interaction and mass transfer. Some formation channels predict matter ejected from the system that might persist gravitationally bound as circumstellar material. After ejection and expansion it is expected to mimic interstellar conditions and to cause the same absorption lines as the ISM. Therefore, as second, equally important aim of this thesis it is explicitly searched for such material using the same methods as before. Together with the measured atmospheric and stellar parameters, detections could shed light on the formation history of the respective objects.

The following sections provide a brief introduction to the involved topics in this work. First, general properties of the ISM as well as its effects on starlight as it passes through are discussed in Section 1.2. Then in Section 1.3, hot subdwarf stars are introduced including their formation scenarios and observational properties. Finally, in Section 1.4 the functionality of spectrographs, with which stellar spectra are recorded, is outlined and the necessary theoretical background about spectroscopy, e.g. the formation and shape of spectral lines, is summarized. Section 1.5 briefly outlines the analyses carried out as part of this thesis.

1.2 The Interstellar Medium

This section provides an overview about the interstellar medium (ISM), in particular about interstellar gas and dust as well as how it is affecting our observations of starlight. If not stated otherwise, it follows standard literature about the ISM and astrophysics in general, like Draine (2011) or Carroll & Ostlie (2017).

1.2.1 Constituents and Characteristic Phases

Up to 10% of the baryonic matter in the Milky Way is found in the ISM. However, not only the baryonic matter (atoms, ions, molecules, etc.) is subject to the ISM, but literally everything in the Galaxy that is between the stars. This includes the following constituents:

- **Interstellar gas:** Atoms, ions, and molecules in the gas phase with (mostly) thermal velocity distributions (Equation 1.13).
- **Interstellar dust:** Small solid particles often coexisting and mixed with the gas in interstellar clouds.
- **Cosmic rays:** Charged particles like protons, electrons, ion, pions, etc. with non-thermal and often extremely relativistic kinetic energies.
- **Electromagnetic radiation:** Photons from various sources. This includes, for example, the cosmic microwave background (CMB), gamma rays, bremsstrahlung[2], synchrotron radiation[3], (thermal) emission from interstellar gas and dust heated by starlight, and in principle also starlight itself.
- **Magnetic fields:** Interstellar magnetic fields result from electric currents in the ISM. They are, for instance, responsible for deflecting charged particles like that of cosmic rays.
- **The gravitational field:** All matter in the Milky Way like that in stars, stellar remnants, dark matter, and also in the ISM, causes gravitation. Some interstellar clouds even collapse under their own gravitation, which leads to the formation of stars.
- **Dark matter:** Dark matter seems to be existent in most galaxies, is apparent due to its gravitational influence, and often even dominates the total mass of a galaxy. However, its nature is still unclear.

This work exclusively focuses on the distribution and properties of the interstellar gas and dust component. It is found in a wide range of different temperatures and densities. In principle, all conditions in that range can be found somewhere in the Milky Way, but it is observed that several distinct characteristic phases are more common than others. Some of these phases are:

- **Dark clouds:** Dark clouds consist of dense[4] molecular gas as well as dust. They typically have extremely low temperatures of $T \sim 10 - 50\,\mathrm{K}$ and high hydrogen densities of $n_\mathrm{H} \sim 10^3 - 10^6\,\mathrm{cm}^{-3}$. They are called "dark" due to their high extinction of optical light. Dark clouds are often self-gravitating, collapsing, and forming stars. They can be observed via molecular CO emission or dust emission in the far infrared.
- **Cold neutral medium:** Cold neutral medium predominantly consists of atomic gas as well as dust. Temperatures and densities are typically up to $T \sim 100\,\mathrm{K}$ and $n_\mathrm{H} \sim 30\,\mathrm{cm}^{-3}$, respectively. It makes up about 1% of the volume of the nearby ISM and can be observed via the hyper fine structure 21 cm line of neutral hydrogen or absorption of light in the optical and ultraviolet (UV).
- **Warm neutral (or partially ionized) medium:** Warm neutral (or sometimes partially ionized) medium predominantly consists of atomic gas (or ionized gas in low ionization stages) with temperatures of $T \sim 2,000 - 15,000\,\mathrm{K}$ and densities of $n_\mathrm{H} \sim 0.5\,\mathrm{cm}^{-3}$. This diffuse material fills a large fraction of the local volume of the Galactic disk (up to 40%) and can be observed the same way as cold neutral medium.
- **Hot ionized medium:** The most extreme temperatures shows shock-heated hot ionized medium with $T \sim 10^6\,\mathrm{K}$ that was, e.g., hit by the blast waves from near supernova explosions. The densities are $n_\mathrm{H} \sim 0.004\,\mathrm{cm}^{-3}$ and the particles are highly ionized. It makes up approximately half of the volume of the Galactic disk and cools very slowly. Sometimes it is also called coronal gas, because of its similarity to the solar corona. It can be observed via UV and X-ray emission as well as radio synchrotron emission.

[2] Emission by decelerated charged particles, e.g. in matter.
[3] Emission by accelerated charged particles, e.g. due to magnetic fields.
[4] "Dense" in the context of ISM would still qualify it for ultrahigh vacuum on Earth.

1.2.2 Absorption by Interstellar Gas

Primordial gas from the time directly after the Big Bang primarily consisted of hydrogen (H) and helium (He). From this material the first stars and galaxies have formed, while the original interstellar medium was still of primordial composition. Then, mainly due to nuclear fusion in the cores of stars, heavier elements (generically called "metals" in astrophysics) have formed. At the end of their lives, stars expel parts of their material, which also contains these heavier elements, back into the ISM, from what in turn new stars are formed. Each generation of stars produces more metals. This cosmic cycle of matter successively enriches the interstellar medium with heavier elements (atomic number $Z \geq 3$), which either exist as atomic/ionized gas or form molecules or dust grains. Today, H and He still account for most of the ISM and metals make up only about 1% of the total mass. Nevertheless, these metals often determine the conditions in the ISM like chemistry, ionization state, and temperature. Furthermore, the interaction with light by these metals provides the most valuable way of investigating the structure and properties of the ISM.

For example, gas particles are able to absorb photons. Their electrons occupy different energy states. Now, if the energy of a photon exactly matches the energy difference between two states of an electron, it might absorb this photon and jump into the higher (excited) level. The photon disappears, resulting into less flux at the wavelength corresponding to the transition. If many absorbers are available to remove photons at this wavelength from a beam of light, a spectral absorption line is formed in the spectrum. A prominent example are the hydrogen Balmer lines in the optical, often present in the spectra of stars. They correspond to a transition from the first excited state to a higher one and are caused by atomic hydrogen in stellar atmospheres pervaded by light emitted from the interior. The interstellar medium, however, is often much cooler than stellar atmospheres and the electrons in hydrogen rather occupy the ground state than the first excited state. The ground state lies energetically much deeper and hence, the energy differences between the levels are much larger and higher energetic UV photons are necessary to excite the hydrogen atoms and to form the so-called Lyman line series[5]. The same is true for most of the transitions in metals. Therefore, the vast majority of absorption lines from the ISM, especially that from neutral hydrogen, are found in UV spectra. Unfortunately, Earth' atmosphere blocks most of the UV light, which does not allow for ground-based observations of such spectra and hence, the by far highest number of stellar spectra is available for the optical wavelength regime.

In the optical, essentially only four interstellar absorption lines are found: One line doublet from singly ionized calcium (Ca II) and one from neutral sodium (Na I). In the historic Fraunhofer notation they are called the Ca II H&K and Na I D1&D2 lines. They were first detected in the spectrum of the Sun, where they are among the strongest spectral lines. However, this means they can be produced by stellar atmospheres as well, which complicates the distinction between a possible interstellar and photospheric[6] origin. Fortunately, the photospheric lines disappear for hotter O- and B-type stars as background sources, where calcium and sodium in the atmosphere are both ionized to higher ionization stages. Since these four lines play a central role for this work, their origin as well as the corresponding atomic properties are briefly discussed in the following two paragraphs.

The Ca II H&K Absorption Lines

The Ca II H&K lines are both resonance lines, i.e. originating from transitions between the ground state and the first excited level. Their relevant atomic data is listed in Table 1.1, which was taken from the NIST data base (Kramida et al., 2024). In general, the transition probability A_{lu} quantifies the likeliness for an electron to jump between two quantum states (lower level l and upper level u) when interacting with light. It can be considered as the inverse of the lifetime of the state and

[5] Spectral lines originating from transitions from the ground state to a higher one are generally called "resonance lines".

[6] A stellar atmosphere typically consists of the photosphere, the chromosphere, and the corona. The deepest layer is the photosphere that can be considered as the "surface" of a star from which the photon flux emerges. Since it is not dealt with stellar chromospheres or coronae in this work, the words "atmosphere" and "photosphere" are used interchangeably.

Table 1.1: Initial and final configurations, lower (l) and upper (u) energy states, air wavelengths (λ), transition probabilities (A_{lu}), and oscillator strengths of the Ca II H&K as well as Na I D1&D2 spectral lines. The atomic data is taken from the NIST data base (Kramida et al., 2024).

Line	Configurations	l	u	λ [Å]	A_{lu} [$10^8\,\mathrm{s}^{-1}$]	f_{lu}
Ca II K	$3p^6 4s - 3p^6 4p$	$^2\mathrm{S}_{1/2}$	$^2\mathrm{P}^\mathrm{o}_{3/2}$	3933.663	1.470	0.682
Ca II H	$3p^6 4s - 3p^6 4p$	$^2\mathrm{S}_{1/2}$	$^2\mathrm{P}^\mathrm{o}_{1/2}$	3968.469	1.400	0.330
Na I D2	$2p^6 3s - 2p^6 3p$	$^2\mathrm{S}_{1/2}$	$^2\mathrm{P}^\mathrm{o}_{3/2}$	5889.951	0.616	0.641
Na I D1	$2p^6 3s - 2p^6 3p$	$^2\mathrm{S}_{1/2}$	$^2\mathrm{P}^\mathrm{o}_{1/2}$	5895.924	0.614	0.320

is sometimes also called an Einstein coefficient. The oscillator strength f_{lu} is a dimensionless parameter, derived from that probability and representing how strongly a particular transition couples to the radiation field. This means the higher the oscillator strength, the stronger is the respective line on average.

The Ca II H&K lines are particularly strong under ISM conditions, because most of the ions reside in the ground state. This is mainly because of the low densities in the ISM, since then the likelihood is low that higher states are excited by collisions between the particles, which is a common way to excite atoms or ions. Furthermore, the lines have high transition probabilities, which leads to strong absorption even for relatively low density gas. Hence, they are an important tool to trace interstellar gas, especially with optical data. The line doublet is located at wavelengths[7] of 3933.663 Å (K) and 3968.469 Å (H), which places it in the blue regime of an optical spectrum, and it corresponds to transitions from the $4s\ ^2\mathrm{S}_{1/2}$ ground state to the $4p\ ^2\mathrm{P}$ excited states[8]. Since the oscillator strength of the H line is much lower than that of the K line, also the spectral line is much weaker.

Calcium is one out of a series of refractory elements that are completely depleted onto dust grains in cold interstellar medium. At higher temperatures, where dust is more and more dissociated, neutral calcium becomes ionized because of its low ionization potential of $\chi = 6.1\,\mathrm{eV}$ ($\mathrm{Ca\,I} \rightarrow \mathrm{Ca\,II}$). Therefore, Ca II is the dominant ionization stage of calcium in the warm interstellar medium, from which the Ca II H&K lines mainly arise. On the other hand, in the cold medium calcium is almost not observable.

The Na I D1&D2 Absorption Lines

The Na I D1&D2 lines are also both resonance lines and their atomic data is listed in Table 1.1 as well. Just like for calcium, at typical ISM conditions most of the sodium atoms reside in the ground state, making the lines particularly strong. Even though their transition probabilities are lower than that of the calcium lines, they are high enough to cause strong absorption. With wavelengths of 5889.951 Å (D2) and 5895.924 Å (D1), this line doublet is much closer to each other and located in the redder part of an optical spectrum. It corresponds to transitions from the $3s\ ^2\mathrm{S}_{1/2}$ ground state to the $3p\ ^2\mathrm{P}$ excited states. Again one oscillator strength is much lower than the other, such that the D2 is in general stronger.

The depletion of sodium on dust grains is not well-determined, but there is circumstantial evidence that it does not strongly deplete (Weingartner & Draine, 2001). Therefore, a lot of Na I is found in the cold ISM coexisting with interstellar dust. At higher temperatures, however, it is ionized to Na II due to its low ionization potential $\chi = 5.1\,\mathrm{eV}$ ($\mathrm{Na\,I} \rightarrow \mathrm{Na\,II}$) and no longer observed.

Overall, the Ca II H&K lines are a well suited tracer for the diffuse warm neutral or partially ionized interstellar medium, while the Na I D1&D2 lines are better suited for the dense cold neutral interstellar medium. Together, both line doublets comprise the full picture of the cool to moderately

[7] In the course of this work all wavelengths are given in air. In vacuum they would be slightly shifted.

[8] For a review of the spectroscopic notation, see for example Chapter 4 of Draine (2011).

warm phases of the interstellar medium. More details on the information that can be derived from the appearance and shape of the interstellar absorption lines is provided in Subsection 1.4.4. The developed analysis methods are extensively discussed in Section 3.3, especially in Subsection 3.3.4.

1.2.3 Extinction and Reddening by Interstellar Dust

A complementary constituent of the ISM to interstellar gas is the interstellar dust. It consists of small solid particles with a broad size distribution between about $0.01 - 0.2\,\mu$m. In general, those are produced by condensation in the outflows of evolved stars, in the ejecta of supernovae, and by accretion of particles onto pre-existing cores in dense molecular clouds. The dust particles are, for example, silicate grains (Mg, Si, O, Fe), carbonaceous grains (amorphous carbon, graphite), hydrocarbons (e.g., polycyclic aromatic hydrocarbons called PAHs), (silicon) carbides (SiC), metallic Fe, or mixed material grains with ice mantles (in cold, dense regions). The properties and effects of interstellar dust are versatile and a highly topical field of current research. Some of the many facets are:

- Heating and cooling of the ISM controlled by dust grains.
- Polarization of light by elongated dust grains aligned with Galactic magnetic fields.
- Scattering of light in reflection nebulae by dust.
- Depletion of certain elements from the interstellar gas onto dust grains.
- Formation of molecular hydrogen (H_2) with dust grains as catalyst.
- The influence of dust on the chemistry of the ISM.
- Thermal emission from dust in the infrared.
- Wavelength dependent extinction (absorption plus scattering) of light by interstellar dust.

Despite the wealth of intriguing characteristics of interstellar dust, it is not in the scope of this work to address all of them. The main topics dealt with are the two latter ones: While the extinction of starlight is directly utilized to quantify the amount of dust in the line of sight, the thermal emission in the infrared is only used indirectly, by comparing observations to comprehensive dust emission maps (Schlegel et al., 1998; Schlafly & Finkbeiner, 2011; Green et al., 2019).

The extinction of starlight comprises both, absorption and scattering. While absorption affects only certain wavelength intervals, the scattering of light by dust grains is omnipresent from the UV to the infrared. In fact, this scattering is highly wavelength dependent and is much stronger in the UV/blue regime than at longer wavelength. This means, if a beam of starlight passes through a dust cloud, most of the red light is transmitted, while blue light is scattered and removed from the line of sight. This effect is the reason why stars often appear redder than they actually are and is called "selective extinction" or simply "reddening".

The overall wavelength dependent extinction A_λ is measured in magnitudes (mag), just as the brightness of stars. For example, an extinction of $A_B = 0.2$ mag corresponds to the dimming of a star by 0.2 mag in the blue regime (B). In general, the extinction is defined as[9]

$$A_\lambda = 2.5 \log \frac{F_\lambda^0}{F_\lambda}\,, \tag{1.1}$$

where F_λ is the observed flux of a star and F_λ^0 is the true flux that would have been observed without extinction. The slope of this "extinction curve" at visible wavelengths (V) is typically characterized by the dimensionless ratio

$$R_V = \frac{A_V}{A_B - A_V} = \frac{A_V}{E(B-V)}\,, \tag{1.2}$$

[9] In general for this thesis, log always represents the logarithm to the base of 10, while ln is the natural logarithm.

where A_B and A_V are the extinctions measured in the blue and visual photometric wavelength bands centered at about 4,400 Å and 5,500 Å, respectively. The parameter $E(B-V) = A_B - A_V$ is called the color excess or reddening parameter. Often it is just called "the reddening". It can be considered as the color difference $E(B-V) = (B-V)_{\text{observed}} - (B-V)_{\text{intrinsic}}$ between the observed color of a star and its true one. The slope R_V has been measured along many sight lines and an average value of $R_V = 3.1$ has been established for the Milky Way. It is commonly considered as constant and then, the overall extinction is purely determined by the reddening $E(B-V)$. However, recent studies show that R_V can vary drastically depending on the observed interstellar environment (Zhang et al., 2023; Zhang & Green, 2025). Green et al. (2025) even claims that extinction curves contain more than one degree of freedom and that it is not sufficient to parametrize it with the single variable R_V. Nevertheless, in the course of this work it is stick to the conventional formalism with a constant R_V.

A slightly modified formalism, which is later on used for the reddening determination, is that of Fitzpatrick et al. (2019). There, the average Milky Way extinction curve is given by

$$k(\lambda - 55) = \frac{A(\lambda) - A(55)}{A(44) - A(55)}\,, \tag{1.3}$$

where 44 and 55 are short for 4,400 Å and 5,500 Å, respectively. The reddening is defined as the monochromatic color excess $E(44-55)$, which parametrizes the extinction model

$$R(55) = \frac{A(55)}{E(44-55)} \tag{1.4}$$

and again purely determines the strength of the extinction. These two equations can be combined to the normalized extinction curve[10]

$$A_{\text{norm}}(\lambda) = E(44-55) \cdot (k(\lambda - 55) + R(55))\,. \tag{1.5}$$

Example extinction curves for different reddening and $R(55)$ values are depicted in Figure 1.3. $R_V = 3.1$ converts to a monochromatic value of $R(55) = 3.02$, which is adopted for this work. A general procedure for converting broad band parameters to monochromatic ones and vice versa, like $E(44-55)$ to $E(B-V)$, is outlined in Appendix B of Fitzpatrick et al. (2019). However, the deviation is in general small.

These extinction curves are included in the modeling of stellar spectral energy distributions (SEDs) as described in Section 3.4. The effect of wavelength dependent extinction on stellar SEDs is illustrated in Figure 1.4. It can be clearly seen that the flux at shorter wavelengths is much more suppressed by reddening than at longer wavelength, where the true and observed SEDs converge. This is exactly the effect, which allows for the determination of the reddening parameter as proxy for the amount of interstellar dust in the line of sight. By applying this method to a large sample of stars across the sky, the distribution of interstellar dust can be traced. It represents a complementary analysis to the investigation of the interstellar absorption lines probing the interstellar gas.

[10] For a more detailed view on this formalism see Fitzpatrick et al. (2019).

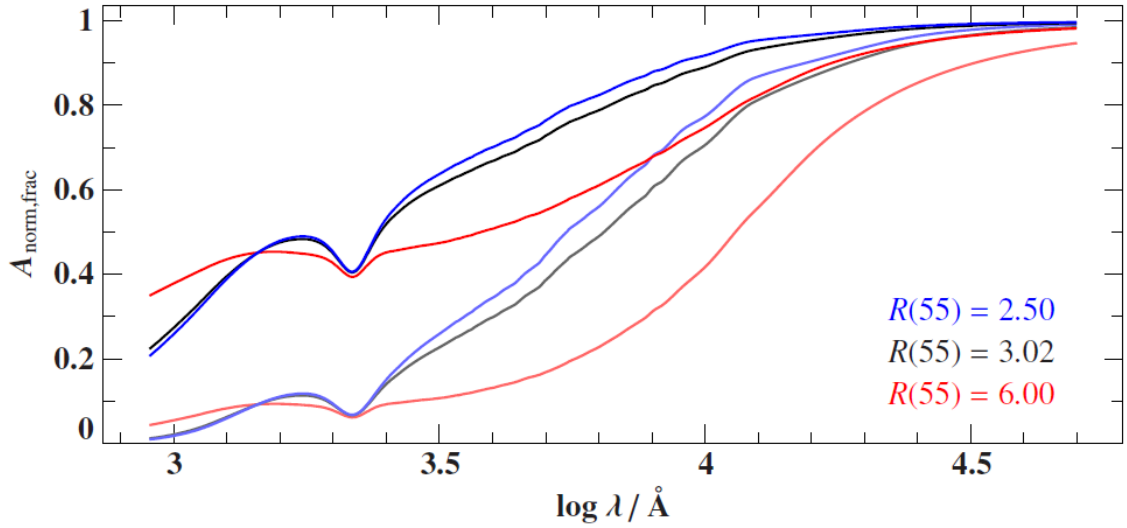


Figure 1.3: Empirical fractional normalized extinction curves from Fitzpatrick et al. (2019) for different values of $R(55)$ in a logarithmic wavelength scale. The upper three curves are for a reddening of $E(44-55) = 0.1$ mag and the lower three curves for a reddening of $E(44-55) = 0.3$ mag. The exact origin of the extinction bump at ~2,175 Å in the UV is still not clarified, but likely caused by PAHs or a mixture of molecules (Ma et al., 2020). The figure is taken from Dorsch (2023).

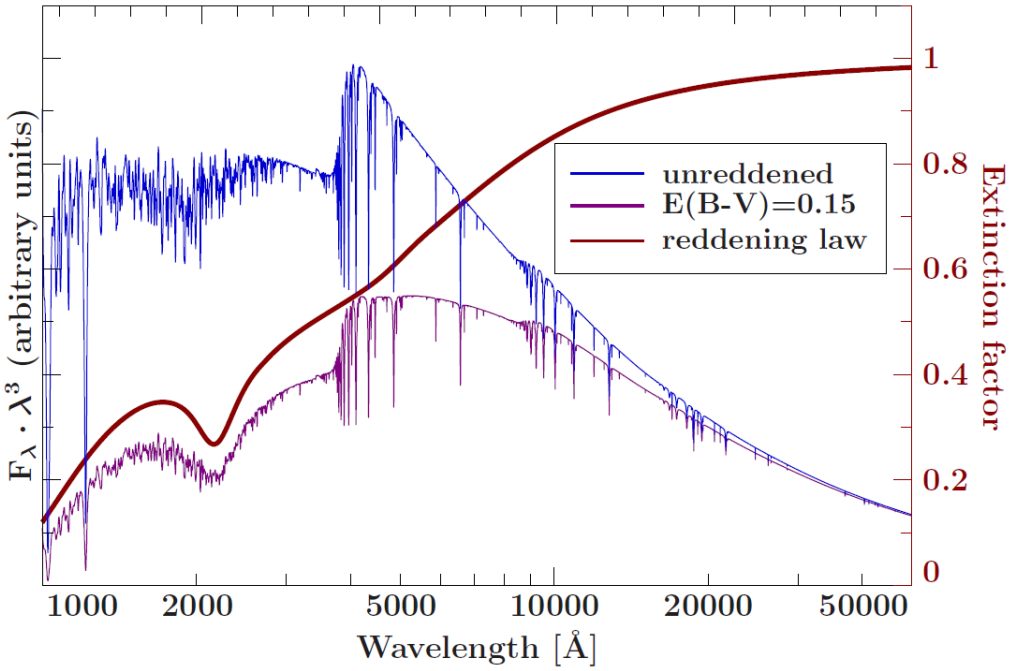


Figure 1.4: The effect of a certain reddening law (extinction curve) on a stellar SED. The extinction curve with $E(B-V) = 0.15$ mag is shown in red, the true SED in blue, and the SED affected by reddening in fuchsia. The flux distribution is flattened by the factor of λ^3 for better visibility. Otherwise, the slopes would be much steeper. The figure is taken from Heber et al. (2018).

1.3 Hot Subdwarf Stars

In the previous section the properties of the interstellar medium have been discussed as well as how it can be probed. Now it is time to have a closer look on the used background light bulbs to shine through the ISM. Those hot subdwarf stars are unique and intriguing objects themselves. Their formation and evolution is not yet well understood.

The formation, structure, and evolution of low-mass single stars, from the start of hydrogen fusion to their final fate as white dwarfs, is well-studied and mostly understood. However, the evolution of stars in close binary systems through phases of interaction and mass transfer is less clear. Hot subdwarfs represent a population of stars that is strongly associated with being the result of close binary interaction. Some studies even suggest that binary interaction is always required to form these stars (Pelisoli et al., 2020). In the next section the canonical stellar evolution of single low-mass stars is briefly described in order to understand why it can not account for the existence of hot subdwarf stars. Subsequently, the actual formation scenarios of these stars are summarized followed by an outline of their observational characteristics. In general, if not stated otherwise these sections follow the comprehensive reviews of Heber (2009, 2016) and the study of Dorsch (2023).

1.3.1 Canonical Single Low-Mass Star Evolution

The canonical stellar evolution of low-mass stars is well-studied and the results are recorded in many publications and textbooks. For example, a standard general astrophysical textbook including stellar evolution is Carroll & Ostlie (2017), while Clayton (1983) specifically focuses on the principles of stellar evolution and nucleosynthesis.

During the gravitational collapse of large molecular clouds many new young stars are formed. As soon as it becomes dense and hot enough in their cores, they begin to fuse hydrogen to helium. This is considered as start of their lives on the so-called main sequence, where they reside the longest time of their live, stably burning hydrogen in the core. Overall, their evolution comprises five stages from their birth to their end, which are all depicted in the Hertzsprung-Russell diagram (HRD) of Figure 1.5. It plots the stellar luminosity over the effective temperature. Each stage is briefly outlined in the following:

- **Main sequence:** All single stars spend most of their life on the main sequence (MS), where they burn hydrogen to helium in their core. Their masses can range from $0.07\,M_\odot$ up to $150\,M_\odot$, although the limit of the upper end is not clear. The lifetimes on the main sequence strongly depend on the mass: While the Sun is predicted to spend about 8 Gyr on the MS, a lower mass star with $0.8\,M_\odot$ would spend about 13 Gyr, and a heavy B-type star with about $10\,M_\odot$ only 30 Myr. Stars on the MS move little in the HRD (Figure 1.5).

- **Subgiant and red giant branch:** As soon as the hydrogen fuel is exhausted in the core, nuclear fusion continues in a shell around the helium core. The star leaves the MS as the core contracts and the envelope expands (subgiant branch). The subgiant evolution ends when the star begins to drastically expand and to ascend the red giant branch (RGB). For low-mass stars, the helium cores become electron-degenerate. These degenerate cores actually contract as their masses increase by dumped ashes from the hydrogen burning shell around it. Thus, the inner pressure, density, and temperature increases, which leads to higher nuclear luminosities of the burning shell. The stars answer with an expanding envelope on that higher luminosity. At some point the helium cores reach temperatures of about 10^8 K, which violently ignites helium fusion to heavier elements. Because the pressure of electron-degenerate matter is independent of temperature, the cores do not immediately expand. Thermonuclear run-away fusion heats up the core dramatically, which is called the helium-flash. This vicious cycle is interrupted only as soon as the thermal energy is sufficient to lift the degeneracy and the tip of the RGB is reached.

- **Horizontal branch:** Followed by the helium-flash, the stars contract again and reach a

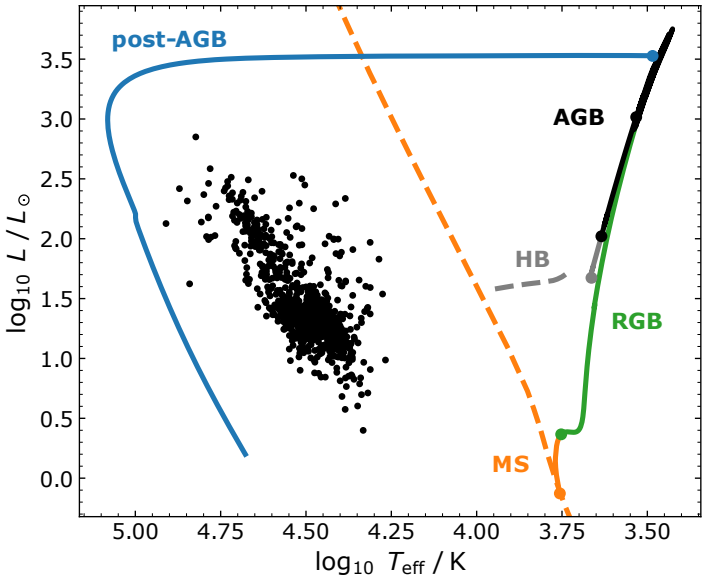


Figure 1.5: Stages of evolution for a Sun-like star in the Hertzsprung-Russell diagram (HRD) modeled with MIST (Choi et al., 2016). Depicted are the main sequence (MS) in orange, the red giant branch (RGB) in green, the horizontal branch (HB) in gray, the asymptotic giant branch (AGB) in black, and the post-AGB in blue. A sample of hot subdwarf stars is shown with black dots for comparison. The figure is taken from Dorsch (2023).

new equilibrium configuration. They stably burn helium in a non-degenerate core and hydrogen in a shell around it. This branch of core helium burning stars extends towards high temperatures at almost constant luminosity and is therefore called the horizontal branch (HB).

- **Asymptotic giant branch:** After even the helium is exhausted in the core, helium is burned in a shell around the carbon/oxygen (C/O) core and hydrogen in a shell around the helium burning one. They begin to ascend the asymptotic giant branch (AGB) similarly to the RGB before. It is called ”asymptotic“ because it asymptotically adapts to the RGB.

- **Post-asymptotic giant branch:** Unlike for the RGB, low-mass AGB stars never reach central temperatures to ignite the core with nuclear fusion (in this case carbon burning). It expands until the envelope is so weakly bound that radiation pressure, pulsations, and/or strong stellar winds lead to an ejection of the envelope, which is then observed as planetary nebula. Left behind is a hot C/O core, possibly with a thin layer of hydrogen or helium around it, without any energy source (no nuclear fusion). This remnant contracts and cools down entering the white dwarf cooling sequence. The way from the end of the AGB down to the white dwarf sequence is called the post-AGB.

This canonical stellar evolution picture can not explain the existence of hot subdwarf stars, which reside at the extreme blue end of the horizontal branch (EHB) or even beyond that on the so-called helium main sequence (He-MS), where stars purely consisting of helium are located (Paczyński, 1971). It is clearly visible from Figure 1.5 that none of the stellar evolution branches intersects the region of the hot subdwarf stars. Therefore, alternative formation scenarios including binary interaction need to be invoked in order to explain the existence of this unique population of stars.

1.3.2 Formation and Evolution of Hot Subdwarfs

In general, hot subdwarf stars are thought to be the stripped helium cores of evolved stars. Hence, some kind of mass loss is needed the remove most of the hydrogen envelope. To this end, binary interaction needs to be introduced with phases of mass transfer and loss. However, many hot subdwarfs appear to be single. Hence, the first formation scenario that is presented is a hypothetical one, which does not necessarily need a binary companion. All other ones include binary interaction, like common envelope evolution or merger events.

The Hot Flasher Scenario

Low-mass giant stars leave the tip of the RGB when their degenerate cores ignite helium fusion, while they are still burning hydrogen in a shell around the core. But hypothetically, the hydrogen shell might be exhausted before the core ignites nuclear fusion. Then the envelope would contract and the star leave the RGB early. While contracting, if the central temperature becomes high enough, a delayed helium-flash might occur that strips off parts of the envelope. This "late hot flash" is considered as possible explanation for single EHB stars (D'Cruz et al., 1996). However, calculations predict too much retained hydrogen, inconsistent with the formation of hot subdwarf stars. A common approach is to artificially increase the mass loss, which leads to increasingly hot and helium-rich HB stars. It is not yet clear whether this enhanced mass loss is possible.

Common Envelope Evolution

About 30-50% of all B-type hot subdwarfs (sdBs) are in close binary system with either white dwarf (WD), M-type MS, or brown dwarf companions. They show extremely short orbital periods from about one hour up to few days (Kupfer et al., 2015; Kawka et al., 2015; Schaffenroth et al., 2022) and strong radial velocity variability with amplitudes of up to several hundred $\mathrm{km\,s^{-1}}$. These tight orbits are not natural and can only be reached in binary interaction. Some interaction channels leading to the formation of hot subdwarfs are shown in Figure 1.6, where the first two result into close binaries. As example, the second scenario is outlined in more detail: Assuming an initial wide binary with two MS stars of unequal mass, the more massive star will evolve first to giant dimensions. If the system is close enough, it will fill its Roche lobe[11] and begin to transfer mass to the companion. If furthermore the mass ratio q is high ($q \gtrsim 1.2-1.5$), the mass transfer will be dynamically unstable and a common envelope (CE) is formed engulfing both components (Paczynski, 1976). Then the remaining helium core of the stripped star and its companion orbit each other inside of the CE, which leads to friction and drag forces. Therefore, orbital angular momentum is transferred to the surrounding material, which lets the orbit shrink and heats up the CE. As consequence the CE expands, becomes eventually gravitationally unbound to the system, and is ejected. If the central temperature of the helium core was already high enough to ignite nuclear burning, a hot subdwarf star is formed in a close binary with a low-mass MS companion (late K- or M-type). This is one way of forming binary systems with such short orbital periods.

In the first scenario of Figure 1.6, when the mass ratio is lower, a phase of stable Roche lobe overflow (RLOF) is required in advance of the common envelope phase. The final outcome is a hot subdwarf in a close binary system with a WD companion. Overall, given the large number of hot subdwarfs in close binary system, common envelope evolution must contribute significantly to their formation history.

Moreover, it is expected that some of the ejected common envelope material might persist gravitationally bound to the system. Still, direct detections of such circumstellar material (CSM) around hot subdwarf stars have been elusive yet. However, as this material expands after the ejection, it becomes thinner and thinner and cools down, perhaps mimicking interstellar conditions. Therefore, as the material is pervaded by the starlight of the host system, the same spectral lines as for the interstellar medium might emerge in the spectrum, opening up a tool to search for it. Hence, besides the investigation of the interstellar medium, the search for CSM around hot subdwarf stars is the second major aim of this work. More details on the physics of CE evolution and ejection can be found at the beginning of Chapter 6, in which the results regarding this search are presented.

[11] The Roche lobe is a teardrop-shaped region around a star in a binary system within which orbiting material is gravitationally bound to that star. If a star expands beyond its Roche lobe, matter can flow through the inner Lagrange point to its companion, leading to mass transfer. The Roche potential is given by the common gravitational potential of both components plus a rotation term.

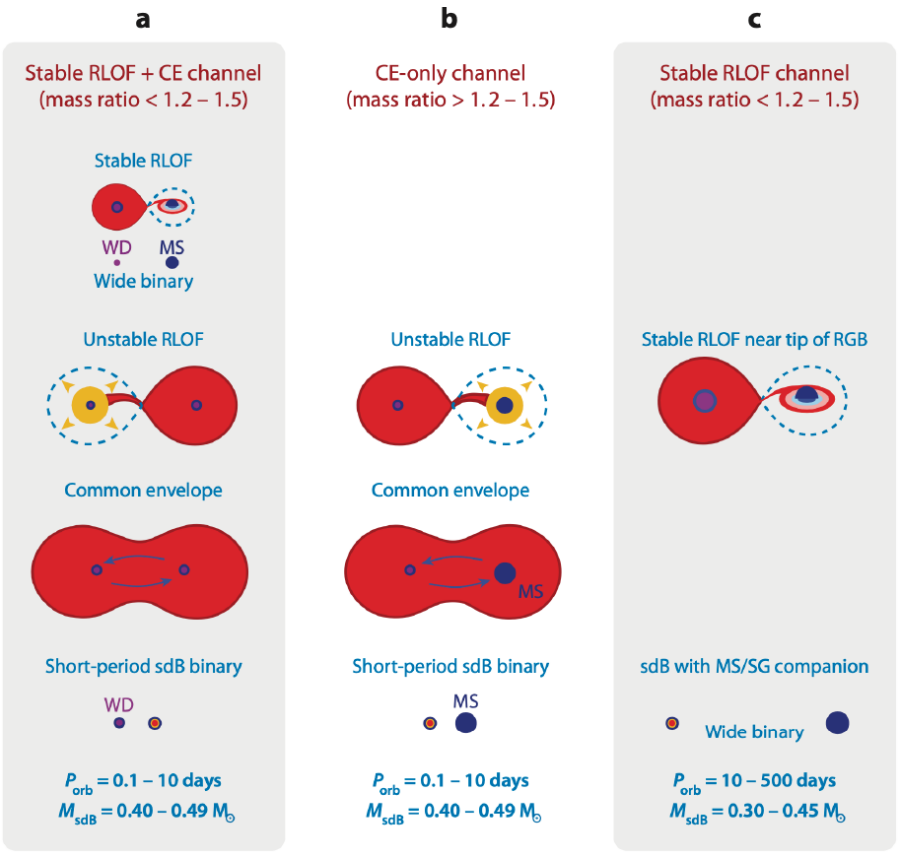


Figure 1.6: Three possible binary interaction channels resulting into the formation of hot subdwarf stars. They include phases of stable and unstable mass transfer as well as common envelope evolution. The figure is taken from Heber (2016), originally by Podsiadlowski et al. (2008).

Stable Roche Lobe Overflow

While the companions in close binary systems are too faint to be directly observed[12], brighter companions can be identified by a double-peaked SED with the hot subdwarf dominating the blue part and the cooler companion the red part. Such wide binary systems with F/G/K-type main sequence companions make up another 20-30% of the whole hot subdwarf population. The formation of such systems is the third scenario of Figure 1.6. If the mass ratio is low ($q \lesssim 1.2$), i.e. both components are roughly of equal mass, and the initial separation is not too large, the star that evolves first fills its Roche lobe and begins to transfer mass to its companion. Then the evolved star is stripped by stable RLOF and if the central temperature was high enough to ignite helium burning, a hot subdwarf plus MS star system in a wide orbit is formed. Even though Figure 1.6 states orbital periods of up to 500 days, there is evidence that the usual period might be slightly higher between 500 and 1,500 days (Barlow et al., 2012; Vos et al., 2019; Molina et al., 2022).

After the formation of such a system, the stars are evolving further. If the hot subdwarf has already exhausted its fuel becoming a C/O white dwarf and the companion expands while ascending the RGB, unstable RLOF starts. This exactly corresponds to the left panel of Figure 1.6 and the first scenario takes over, resulting into CE evolution and a close WD plus hot subdwarf binary system.

White Dwarf Mergers

Another mechanism to produce single hot subdwarfs is the merging of two low-mass objects, where at least one of them is a helium-core white dwarf (He-WD). Even without common envelopes, the orbits of close systems gradually shrink due to the emission of gravitational waves. As

[12] They are only evident from radial velocity variability or peculiar light curves of the system.

consequence both components come closer and closer until they eventually merge. Different progenitor systems are associated with different outcomes: A prominent example is the double He-WD merger scenario. It is expected to explain the population of extremely helium-rich hot subdwarfs of spectral type O (eHe-sdO). The peculiar surface compositions of eHe-sdOs are explained by different merger channels of two He-WDs, e.g. fast or slow merging. Another possibility is the merging of a He-WD with a C/O-WD. This might form peculiar He-sdOs too as well as intermediate helium-rich OB-type hot subdwarfs (iHe-sdOB). The last scenario is the merging of a He-WD with a low-mass main sequence star. It is predicted that this channel could form single helium-poor sdB stars.

In general, in such merger events both components do not just swallow each other and immediately mix their material. It is rather the case that one component gets disrupted, an accretion disk or a corona of material is formed, and the debris is accreted by the other component. It seems also possible (e.g, in the He-WD+MS case) that one star begins to unstably transfer mass, a common envelope is formed, and both components merge inside of the CE. Overall, in all scenarios a certain circumstellar environment might be present after the merger event. Again the same spectral lines as for the ISM could indicate the presence of such circumstellar material.

1.3.3 Spectral Classification and Observational Characteristics

Since the analysis of this work comprises the investigation of many hot subdwarf spectra, it is important to understand their observational characteristics. Historically, these stars were not distinguished from normal early O- and B-type stars until their subluminous nature was realized (Humason & Zwicky, 1947). In the following decades a lot of effort has been put into finding new candidates and collecting large samples of these stars. The most recent catalog of hot subdwarf candidates based on photometry and *Gaia* astrometry (see Section 2.1) was compiled by Culpan et al. (2022) and includes about 60,000 objects. Also the sample of this work is based on this catalog (Section 2.3).

Nowadays, it is clear that individual hot subdwarfs show very distinct spectral features in their spectra. Therefore, the whole population is divided mainly into five sub-populations depending on the presence of hydrogen and helium lines, which are determined by the surface temperature as well as the atmospheric hydrogen and helium abundances. Example spectra illustrating the distinction are shown in Figure 1.7. Most abundant among the whole population are the helium-poor B-type hot subdwarf (sdB) stars. They show mostly strong hydrogen Balmer lines together with weak neutral helium lines, while ionized helium lines are absent. Hotter helium-poor OB-type subdwarfs (sdOB) show Balmer lines together with both, weak neutral and ionized helium lines. Hot helium-poor O-type subdwarfs (sdO) show strong ionized helium, weaker Balmer lines, and no neutral helium lines. These are the three sub-populations of helium-poor hot subdwarfs. Furthermore, there is a class of intermediate helium-rich hot subdwarfs of OB-type (iHe-sdOB), which show strong neutral and ionized helium lines together with Balmer lines. Lastly, some O-type hot subdwarfs are extremely helium enriched (eHe-sdO) and show strong neutral as well as ionized helium lines, while Balmer lines are at most weak or even absent.

From the spectra of stars, e.g. the appearance and shape of spectral lines, characteristic physical properties about their atmospheres can be derived. This is particularly true for hot subdwarf stars. The most essential parameters are the following:

- **Effective temperature:** The effective surface temperature $T_{\rm eff}$ of a star is given by the Stefan-Boltzmann law $F = \sigma_{\rm SB} T_{\rm eff}^4$, where F is the stellar flux and $\sigma_{\rm SB}$ the Stefan-Boltzmann constant. It is arguably one of the most important parameters of a star, especially for the shape of its spectrum, because it determines the excitation and ionization state of atoms and ions in the atmosphere (see Subsection 1.4.2). For example, Balmer lines only emerge if enough hydrogen atoms are excited to the first energy level above the ground state. This is reached if the thermal velocity of the particles is high enough to excite this state, which predominantly occurs via collisions. It requires a certain temperature, but on the other hand, if $T_{\rm eff}$ is too high, hydrogen becomes more and more ionized and the Balmer line strengths

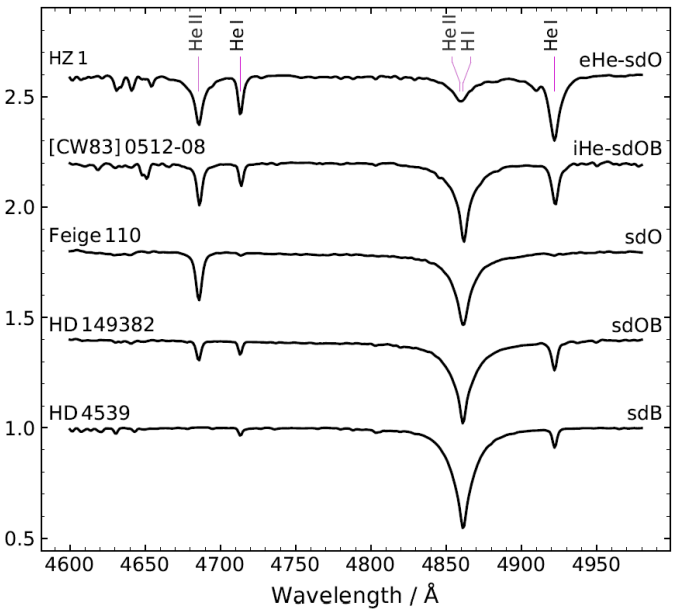


Figure 1.7: Rough classification of hot subdwarf stars into five sub-populations based on the appearance of hydrogen (H I) and helium lines (neutral He I and ionized He II) in their spectra. Shown are example spectra, where the name of the respective star is given on the left and its classification on the right. Responsible for the distinction are the different temperatures and significantly varying helium abundances. The figure is taken from Dorsch (2023).

drop again. Therefore, $T_{\rm eff}$ is of particular importance for hot subdwarf stars as it mainly determines the appearance of Balmer lines as well as neutral and ionized helium lines. Moreover, higher temperatures lead to broadening of spectral lines via thermal Doppler broadening (see Subsection 1.4.3). Overall, from the appearance of spectral lines and their shapes, $T_{\rm eff}$ can be spectroscopically determined. As their name says, hot subdwarfs show high values of $T_{\rm eff} > 20,000\,\mathrm{K}$.

- **Surface gravity:** Hot subdwarf stars are quite compact objects compared to main sequence stars. Hence, they have a high surface acceleration $g = GM/R^2$, where G is the gravitational constant, M the stellar mass, and R the stellar radius. This leads to an enhanced pressure in the atmosphere, which broadens spectral lines via pressure broadening (see Subsection 1.4.3) and influences the ionization balance. Therefore, g can be measured from the shape of spectral lines and since it varies over many orders of magnitude for different kinds of stars, it is usually given as the surface gravity $\log g$. Typical values for hot subdwarf stars are $4.5 < \log g < 6.5$, while O/B main sequence stars have $\log g < 4$ and white dwarfs $\log g > 7$.

- **Helium abundance:** Hot subdwarfs are the stripped helium cores remaining after binary interaction and have at most a thin outer hydrogen layer. They have both, high effective temperatures as well as a high surface gravities. The combination of these parameters leads to atmospheric diffusion processes, which is in principle the balance between gravitational settling of heavier elements and radiative levitation resulting from the absorption of photons[13]. This diffusion can alter the surface composition of hot subdwarfs significantly, in particular the surface helium abundance, leading to the different observed sub-populations. Besides the effective temperature, the surface helium abundance is mainly responsible for the appearance of helium lines in the spectra, from which it can be derived. It is typically given as the hydrogen-to-helium number density ratio $\log n(\mathrm{He})/n(\mathrm{H})$. A value close to -1 indicates solar helium abundance and most of the hot subdwarfs have sub-solar abundances. Only the iHe-sdOBs and eHe-sdOs show more helium in their atmospheres than the Sun.

- **Metal abundances:** The abundances of metals in the atmospheres are far less well-determined, because they require comprehensive spectral analyses based on high-quality spectra. Metals are mostly not even included in synthetic spectral models of hot subdwarf stars (see Section 3.1). However, there is evidence that the metal abundances can vary strongly between individual stars. Furthermore, metals are of course also subject to atmospheric diffusion. A general trend observed for many hot subdwarfs is that they are often

[13] For more details see, for example, Section 1.7 of Dorsch (2023).

enriched of heavier elements and deficient of lighter ones. The origin of this pattern is object of current research.

Together with a few other parameters, like the microturbulence, which is negligible for hot subdwarf stars, these four properties represent the full set of atmospheric parameters quantifying the stellar atmospheres. Even more fundamental parameters, describing a star as a whole, are the stellar mass M, radius R, and luminosity L. The mass of hot subdwarf stars is usually about $0.5\,M_{\odot}$, the radius is typically in the range of $0.1-0.3\,R_{\odot}$, and the the luminosity in the range of $10-1,000\,L_{\odot}$. The determination of those stellar parameters requires further information about the stars than only their spectra can provide, most importantly their distances (see Subsection 3.4.2). Both, atmospheric and stellar parameters, are determined for all stars of the investigated sample as part of this work. On the one hand, the atmospheric parameters are determined in order to use the stellar synthetic spectra as continuum for the ISM absorption lines and on the other hand, they are required to model the stellar SEDs to derive information about interstellar dust along the line of sight. Last but not least, they give conclusion about the hot subdwarfs themselves and for example with the help of diagnostic diagrams (see Section 4.2) their properties, formation, evolution, etc. can be constrained.

1.4 Quantitative Spectroscopy

A major aspect of this work is the analysis of stellar spectra. Typically, they are depicted as diagrams with two axes: The wavelength on the x-axis and the flux on the y-axis. However, it is a highly non-trivial process to arrive at the point that a file with the wavelength in one column and the flux in another one is on hand. This process begins with the observation of a star in a clear night using instruments like a telescope, a spectrograph, and a suited camera. The telescope collects and collimates the starlight, the spectrograph disperses it into its constituent wavelengths, and the camera digitally records what is seen. But this is still just a 2D image of pixels with different gray scale values and not yet ready to evaluate. A complicated process of data reduction follows, including multiple steps of calibration in order to remove noise and systematic effects, assigning wavelengths to the individual pixels (wavelength calibration), the removal of background light and high-energetic cosmic particles (”cosmics“), and the binning of the two dimensional image to a one dimensional wavelength-flux spectrum.

As soon as the data reduction is done, one can proceed with analyzing the resulting spectrum. The spectral features in stellar spectra are usually the shape of the continuum, ionization edges, and most importantly spectral lines. If the spectra get normalized to unity (see, for example, Subsection 3.3.3), spectral lines are actually in most cases the only features left. They appear in various kinds of shapes, some are stronger, some are weaker, sometimes they are broad, sometimes they are narrow. Many physical parameters can be derived from the analysis of these shapes, which is the topic of this section.

It will first briefly introduce general types of spectrographs, with which spectra are produced, and what influence their characteristics have on the final spectra, in particular concerning the spectral resolution. Since the spectra of this work are taken from large spectroscopic surveys and databases (see Section 2.2), they are all available in science-ready formats. Therefore, the complicated process of data reduction is already done and not further considered here. Instead it follows a discussion about the theory of excitation and ionization, which is important for both, stellar atmospheres as well as the interstellar medium. Finally, the mechanisms are outlined that form the shape of spectral lines and how in particular interstellar absorption lines can be modeled.

1.4.1 Spectrographs and the Role of Spectral Resolution

The physics as well as the structure and functionality of spectrographs are complicated and can perhaps fill whole textbooks. However, the basic principles are quite simple and can be mostly explained in the framework of ray optics. The aim of a spectrograph is to split infalling light into its constituent wavelengths. The resulting distribution can then be analyzed, which represents a

powerful tool in astronomy. The type of spectrograph and its components determine the quality of the recorded spectrum, especially the spectral resolution. In the following, the two most fundamental categories of spectrographs are briefly discussed as well as their spectral resolution behavior. It mainly follows Irrgang (2014).

Long-slit Spectrographs

Figure 1.8 shows the typical components of a (long-slit) spectrograph together with schematic ray traces. The most important components are:

- **Slit:** The slit ensures a sharp projection of the observed object by the telescope and prevents stray light to enter the light beam. Its width b influences the spectral resolution of the final spectrum.
- **Collimator:** The collimator parallelizes the incoming light rays before they hit the dispersion element. This is important in order to obtain a proper interference pattern.
- **Dispersion element:** The dispersion element is the main component of a spectrograph. It is responsible for splitting the light into its constituent wavelengths. It can be either a grating, a prism, or a mixture of both, a so-called "grism". Often blaze gratings are used, which are tilted reflection gratings minimizing the loss of intensity during dispersion (Figure 1.9).
- **Camera:** The camera focuses the resulting image of the spectrum on the detector.
- **Detector:** The detector records the image of the spectrum and is often part of the camera. While in former times photographic plates have been used, nowadays the most common instruments are digital CCD detectors.

The schematic of a blaze grating consisting of many periodic reflecting grooves is shown in Figure 1.9. The grooves are tilted by the blaze angle Θ with respect to the surface normal. If a light ray falls on the grating with an incident angle α relative to the surface normal, it leaves the grating under an angle β, which is determined by the blaze angle (given the signed definitions of α and β in Figure 1.9): $\alpha + \beta = 2\Theta$. The criterion for constructive interference is

$$d(\sin\alpha + \sin\beta) = \Delta s \stackrel{!}{=} n\lambda\,, \tag{1.6}$$

where d is the width of the grooves, Δs the path difference between two light beams, λ the wavelength, and n an integer multiple (called "order") of the wavelength. This equation allows to determine the reflection angle β in dependency on the wavelength. However, it also shows that in the optical ($3,800\,\text{Å} \lesssim \lambda \lesssim 7,800\,\text{Å}$) different diffraction orders overlap for $n > 1$. This means different wavelengths within the visible range give the same angle β. Plugging in the blaze angle Θ for β gives

$$d(\sin\alpha + \sin(2\Theta - \alpha)) = n\lambda_{\text{blaze},n}\,, \tag{1.7}$$

where $\lambda_{\text{blaze},n}$ is the blaze wavelength at which constructive interference occurs. For every order n the intensity is at its maximum at the blaze wavelength.

A central quantity of each spectrograph is its resolving power R, because it mainly determines how much spectral information is contained in the resulting spectrum. It is in general wavelength dependent and defined as the ratio of each wavelength λ to its corresponding minimal resolvable wavelength difference:

$$R := \frac{\lambda}{\Delta\lambda} \tag{1.8}$$

Two major processes contribute to a smearing of wavelength and a limited spectral resolution. The first is the intrinsic diffraction limited resolution of the grating, which is given by

$$R_{\text{grating}} = \left(\frac{\lambda}{\Delta\lambda}\right)_{\text{grating}} = nN\,, \tag{1.9}$$

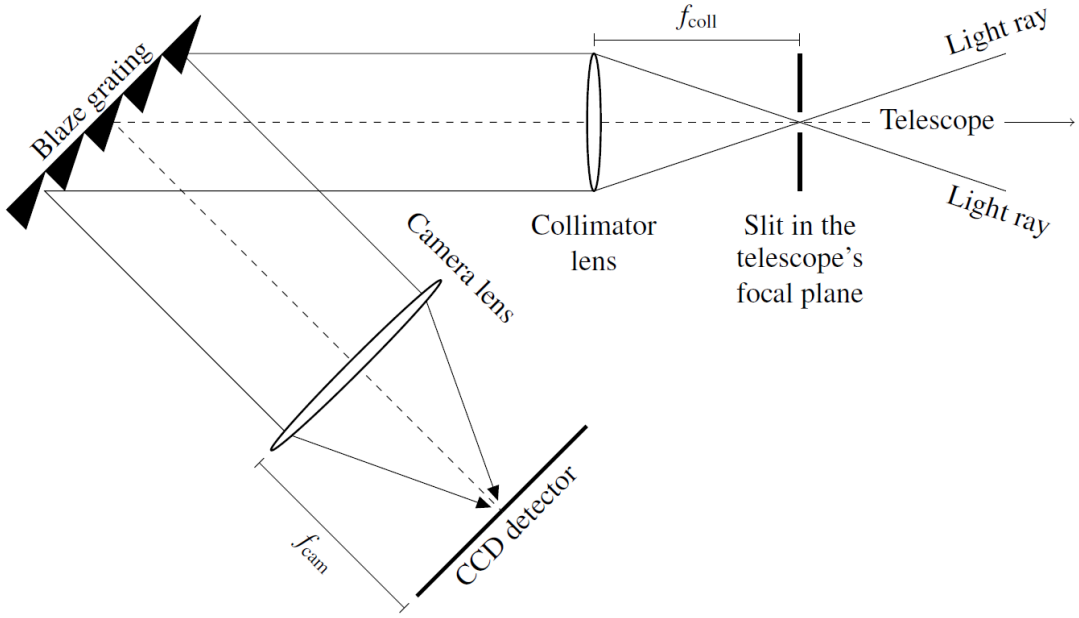


Figure 1.8: Schematic of a long-slit grating spectrograph with its fundamental components, which are the slit with width b, the collimator (with focal length f_{coll}), the dispersion element (here a blaze grating), the camera (with focal length f_{cam}), and the detector (here a digital CCD detector). The figure is taken from Irrgang (2014).

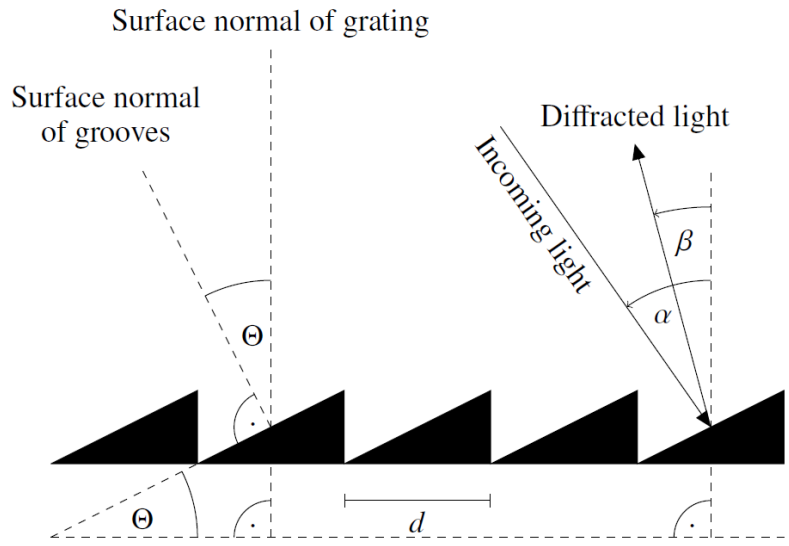


Figure 1.9: Schematic of a blaze grating, where d is the groove width, Θ the blaze angle, α the incident angle, and β the reflection angle of light rays with respect to the surface normal. The angles α and β are both signed. The figure is taken from Irrgang (2014).

where N is the number of illuminated grooves. The second one is the spatial extent of the projection of the star in the focal plane of the telescope, where the slit is located. Due to the finite slit width b, the infalling light is not entirely parallel, leading to a small variation $\Delta\alpha = b/f_{\text{coll}}$ of the incident angle α. From that it can be derived that the spectral resolving power of a long-slit spectrograph is given by

$$R_{\text{long-slit}}(\lambda) = \left(\frac{\lambda}{\Delta\lambda}\right)_{\text{long-slit}} = \frac{n\, f_{\text{coll}}\, \lambda}{d\, b\, \cos\alpha}\,, \tag{1.10}$$

which is in general wavelength dependent. Usually, R_{grating} is much higher than $R_{\text{long-slit}}$, such that in practice $R_{\text{long-slit}}$ determines the resolution of a produced spectrum. Since f_{coll}, d, and α are fixed device-specific quantities, the only ways to increase the resolution is to observe in higher orders n and/or to use smaller slit widths b. However, for higher n different diffraction orders overlap and the smaller b, the lower the amount of infalling light, if the source is not point-like, which is generally the case even for stars due to smearing from Earth' atmosphere ("seeing"). Of course, when building a new spectrograph, it is also possible to optimize the parameters f_{coll}, d, and α in order to obtain a higher resolution. Furthermore, if a fiber-fed spectrograph is used, the slit width b needs to be replaced by the fiber diameter.

Echelle Spectrographs

In order to observe at higher orders n to reach drastically higher resolutions, the issue of overlapping diffraction order can be conveniently overcome by placing a second dispersion element between the first one and the camera/detector. Its task is to disperse the light perpendicular to the initial dispersion direction and hence, it is called a cross disperser. This way, the overlapping diffraction orders are separated as shown in Figure 1.10 and it is possible to make use of diffraction orders up to $n \approx 100$. Due to the vertical alignment of each diffraction order as stripes, such spectrographs with a cross dispersion element are called "echelle" spectrographs according to the French word for ladder or stairs.

The spectral resolving power of an echelle spectrograph can be approximated by inserting the blaze wavelength (Equation 1.7) into Equation 1.10 for the wavelength:

$$R_{\text{echelle}} = \left(\frac{\lambda}{\Delta\lambda}\right)_{\text{echelle}} \approx \frac{f_{\text{coll}}\,(\sin\alpha + \sin(2\Theta - \alpha))}{b\,\cos\alpha} \tag{1.11}$$

Unlike in the long-slit case, the resolution of an echelle spectrograph solely depends on device-specific fixed parameters and no longer on the wavelength. This means the resolving power is constant over the obtained spectrum.

Another limiting factor can be the spatial resolution of the (CCD) detector, given by the pixel distribution and width. To resolve all spectral features the resolution of the detector in general has to be larger than that of the spectrograph. According to the Nyquist criterion, each resolvable wavelength element needs to be covered by at least two pixels of the detector.

Furthermore, for echelle spectrographs the data reduction becomes even more difficult. It includes another complicated step of order merging: The individual separated orders, with partly overlapping spectral regions, need to be merged into a single spectrum covering the whole spectral range.

A Comment on the Spectral Resolution

The most central parameter of a spectrograph quantifying the quality of the resulting spectrum, i.e. the amount of spectral information contained, is the spectral resolving power R. It is a unitless number and defined according to Equation 1.8. Pure long-slit spectrographs provide rather low resolution spectra with $R(\lambda) \approx 1,000 - 10,000$, while with the use of a cross dispersion element, echelle spectrographs are able to reach $R \approx 100,000$ or even higher. In the first case R is wavelength dependent, while in the second case it is constant across the spectrum. In most of the cases a

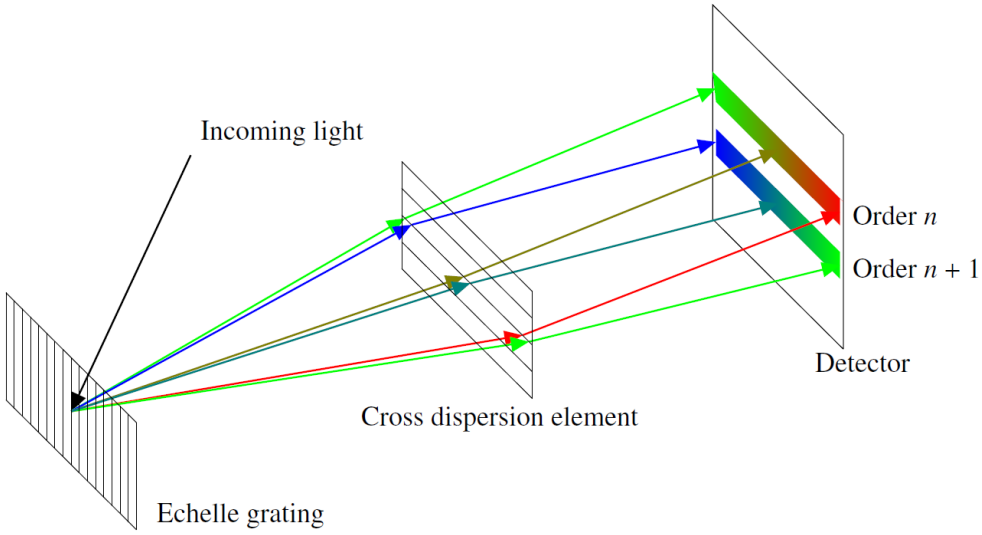


Figure 1.10: Additional component of an echelle spectrograph: The cross dispersion element. It is placed between the first dispersion element and the camera/detector in order to separate overlapping diffraction orders perpendicular to the initial dispersion. This way, it can be observed in high diffraction orders n. The figure is taken from Irrgang (2014).

higher R is desired in order to resolve weaker spectral features, which would be otherwise hidden or blended by neighboring stronger features.

However, the spectral resolving power is just an artificial number quantifying a spectrograph. In practice, the minimal resolvable wavelength difference $\Delta\lambda$ is sometimes more important, because it directly gives the minimal possible width of a narrow spectral line at a certain wavelength due to the used instrument. As such, it defines the so-called "instrumental broadening" (see Subsection 1.4.3) and is called the spectral resolution. Therefore, the spectral resolving power R and the spectral resolution $\Delta\lambda$ (measured in Å) are quite different parameters, even though they are often confused in the literature by also calling R the spectral resolution. It is important to note that for pure long-slit spectrographs R is wavelength dependent, but $\Delta\lambda$ is not. Vice versa, for echelle spectrographs R is constant, but $\Delta\lambda$ is wavelength dependent and increases with increasing wavelength. This means that in echelle spectra the spectral resolution drops[14] and the minimal line width increases at higher wavelength. Furthermore, if for example the resolving power of a long-slit spectrograph is stated to be $R = 3,000$, this just represents a rough average value for the resulting spectra, since it varies with wavelength.

A common way to parametrize the resolving power of both, long-slit and echelle spectrographs, is

$$R(\lambda) = \frac{1}{\Delta\lambda} \cdot \lambda + \text{constant} = R_{\text{slope}} \cdot \lambda + R_{\text{offset}} \,. \tag{1.12}$$

This is a linear approximation with a slope $R_{\text{slope}} := 1/\Delta\lambda$ and an offset R_{offset}. For a pure long-slit spectrograph the values are usually $R_{\text{slope}} > 0$ and $R_{\text{offset}} = 0$, while for an echelle spectrograph vice versa. Therefore, the resolving power increases linearly with wavelength in the long-slit case and stays constant in the echelle case. This approach is, for example, extensively used for the spectral analyses of this work and the developed analysis scripts (see Section 3.3), where a careful treatment of spectral resolution is crucial.

1.4.2 The Theory of Excitation and Ionization

The appearance and shape of spectral absorption lines from many astrophysical environments, such as stellar atmospheres or the interstellar medium, is mainly determined by the number of available

[14] Still, lower spectral resolution usually means higher $\Delta\lambda$.

absorbers. The number of available absorbers of course depends on the abundance of the respective element, but also strongly on their excitation and ionization states. For example, hydrogen atoms that are all in the ground state can not produce Balmer lines, which correspond to a transition from the first excited state to a higher one, even if a lot of hydrogen is available. On the other hand, ionized hydrogen that lost its only electron can not contribute with any electronic transition anymore. Analogously, the same is true for heavier elements with more electrons. Therefore, this section will briefly deal with the theory of excitation and ionization and follows Gray (2021).

The main way of exciting and ionizing particles in a gas is by collisions. The strength of the collisions is determined by their thermal velocities, i.e. predominantly by the gas temperature. In thermodynamic equilibrium the particles follow a Maxwellian velocity distribution given by

$$N(v)\mathrm{d}v = \left(\frac{m}{2\pi k_\mathrm{B} T}\right)^{3/2} \exp\left(-\frac{mv^2}{2k_\mathrm{B} T}\right) \mathrm{d}v\,, \tag{1.13}$$

where v is the velocity, m the particle mass, k_B the Boltzmann constant, and T the temperature. The ratio of populations in two different energy levels m and n of a species is given by

$$\frac{N_n}{N_m} = \frac{g_n}{g_m} \exp\left(-\frac{E_n - E_m}{k_\mathrm{B} T}\right)\,, \tag{1.14}$$

where the g's are the statistical weights[15] and the E's the energies above the ground state of the respective states. Equation 1.14 is called the Boltzmann or excitation equation. Furthermore, the ratio of populations in two consecutive ionization stages I of an element is given by

$$\frac{N_{I+1}}{N_I} = \frac{2u_{I+1}(T)}{u_I(T)} \frac{1}{n_e} \frac{(2\pi m_e k_\mathrm{B} T)^{3/2)}}{h^3} \exp\left(-\frac{\chi_I}{k_\mathrm{B} T}\right)\,, \tag{1.15}$$

which is called the Saha or ionization equation. Here, the $u(T)$ are the partition functions[16] of the respective ionization stages, n_e is the electron density of the gas, m_e the electron mass, h Planck's constant, and χ_I the ionization potential of the lower ionization stage. Notably, Equation 1.15 depends on the electron density of the gas. The higher it is, the lower the fraction of particles in the upper ionization stage, mostly because more electrons are available for recombination.

Overall, Equation 1.13, Equation 1.14, and Equation 1.15 determine the excitation and ionization states of gas particles in thermodynamic equilibrium. For virtually all conditions in the interstellar medium, at least for those considered in this thesis, this is the case. For the atmospheres of cooler stars thermal equilibrium can be often assumed as well. However, in that of hotter stars, like O- and B-type hot subdwarfs, it usually breaks down. Then alternative equations need to be invoked (see Section 3.1).

1.4.3 Line Broadening Mechanisms and Spectral Line Profiles

Many physical parameters can be derived from the appearance and, in particular, the shape of spectral lines. Intrinsic conditions like the temperature and density of a gas, but also extrinsic ones like rotation or the resolution of the used spectrograph, broaden the profiles of the observed lines in different ways. In general, this accounts for both, absorption and emission lines. Therefore, this section summarizes all the different mechanisms for line broadening and the corresponding line profiles. Again, it follows Gray (2021). The intrinsic mechanisms are the following:

- **Natural line broadening:** All energy levels except the ground states have a certain finite lifetime. However, according to Heisenberg's uncertainty principle, energy and time are two complementary quantities and can not be measured infinitely precise simultaneously. Since the energy is directly related to wavelengths or frequencies, spectral lines have naturally a

[15] The statistical weights are $2J+1$ of the state, where J is the inner quantum number. For example, for hydrogen $g_n = 2n^2$ for the n-th excited state.

[16] $u(T) = \sum_{i=0}^{\infty} g_i \exp(-E_i/k_\mathrm{B} T)$

certain width. This effect is called natural broadening and can be calculated in the formalism of a damped harmonic oscillator. The resulting line profile is given by

$$\varphi_{\mathrm{nat}}(\nu) = \frac{\gamma_{\mathrm{nat}}/4\pi}{(\nu_0 - \nu)^2 + (\gamma_{\mathrm{nat}}/4\pi)^2}\,, \tag{1.16}$$

where ν is the frequency and ν_0 the line center. It represents a Lorentzian function with the so-called damping constant γ_{nat} that determines the width. The natural line width of a spectral line is typically on the order of $\Delta\lambda \sim 10^{-4}$ Å and hence, often negligibly small.

- **Collisional (pressure) broadening:** Radiating or absorbing atoms/ions are affected by collisions with neutral and charged particles via Coulomb interactions (Stark effect). This leads to a modification of the energy levels (energy shift ΔE) and line broadening. The perturbations mainly depend on the distance r between the absorbers and perturbing particles: $\Delta E \sim r^{-n}$. Different orders n of collision processes are introduced, which account for different types of interactions:

  - $n = 2$ corresponds to the linear Stark effect, which is important for atoms/ions with intrinsic dipole moment like H I or He II ("hydrogenic"), perturbed by charged particles like electrons or ions.
  - The $n = 3$ case is called self-pressure broadening and mainly accounts for interactions between neutral atoms of the same kind. It particularly affects hydrogen as the lightest and most abundant element.
  - $n = 4$ corresponds to the quadratic Stark effect, which is important for collisions between metal ions and free electrons.
  - The $n = 6$ case is called Van der Waals broadening and accounts for interactions between neutral atoms of different kind.

  It is quite difficult to compute the corresponding line profiles analytically and approximations are needed. If the perturbation time is much smaller than the energy level lifetimes, e.g. for light and fast moving particles, often the "impact approximation" is used. It describes an interruption of the emitted or absorbed electromagnetic wave, creating a phase shift, which leads to a reduction of the lifetime. But as the lifetime decreases, the energy uncertainty increases according to Heisenberg and the line broadens. The corresponding line profile

$$\varphi_{\mathrm{n}}(\nu) = \frac{\gamma_{\mathrm{n}}/4\pi}{(\nu_0 - \nu)^2 + (\gamma_{\mathrm{n}}/4\pi)^2} \tag{1.17}$$

  is again Lorentzian with a damping constant γ_n depending on the considered order n. Unlike in the natural damping case, the damping "constant" is actually not constant here, but depends on the pressure and temperature. If the particles are heavier and move more slowly, the perturbation time is longer than the lifetimes and usually the "quasistatic approximation" is used. Here, the absorbers feel a static electric field by the disturbers during the transition, which leads to a distortion of the energy levels. But even if only the nearest neighbor ("nearest neighbor approximation") is considered, the resulting line profile is more complicated and non-Lorentzian.

- **Thermal Doppler broadening:** The absorbing or emitting particles of a thermal gas follow a Maxwellian velocity distribution (Equation 1.13). Their isotropic thermal motion leads to Doppler shifts for each transition. In total, the observed line broadens and the corresponding Gaussian line profile is

$$\varphi_{\mathrm{Doppler}}(\nu) = \frac{1}{\sqrt{\pi}\Delta\nu_{\mathrm{th}}} \exp\left(-\frac{\Delta\nu^2}{\Delta\nu_{\mathrm{th}}^2}\right) \tag{1.18}$$

  where $\Delta\nu_{\mathrm{th}} = v_{\mathrm{th}}\nu_0/c$ with the thermal velocity v_{th}, ν_0 is the line center, $\Delta\nu = \nu - \nu_0$, and c is the speed of light.

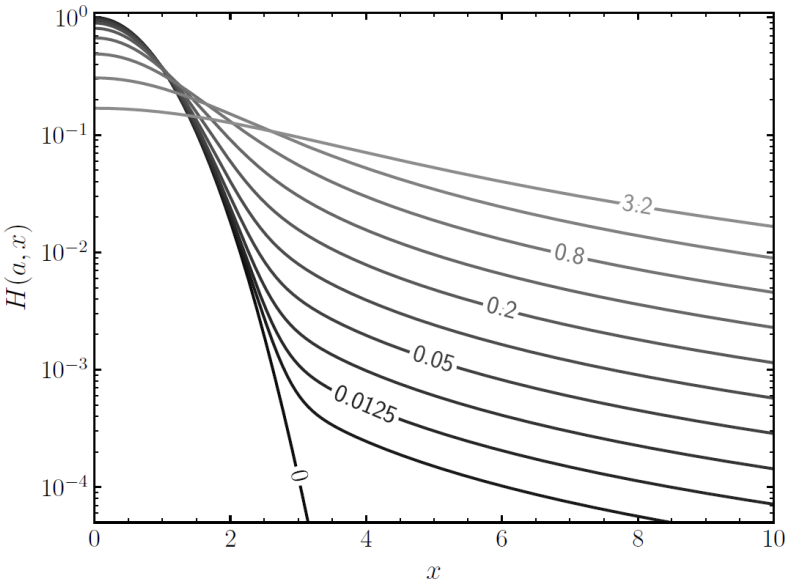


Figure 1.11: Hjerting function in dependency on the damping parameter a. For low values (black) the Gaussian contribution dominates, while for higher values (gray) the Lorentzian one does. The figure is taken from Dorsch (2023).

- **Microturbulence:** Microscopic small-scale motions of gas lead to additional Doppler broadening of spectral lines. This microturbulence can be treated similar to the Doppler broadening and the Gaussian line profile is exactly the same. Only the thermal velocity needs to be replaced by the microturbulent velocity ξ.

The contribution of all these broadening mechanisms can be combined to an overall line profile. Mathematically, all the individual line profiles need to be convolved with each other. The convolution of two functions f and g is defined as

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)\mathrm{d}y\,. \tag{1.19}$$

Then, the final line profile reads (neglecting non-Lorentzian/Gaussian contributions)

$$\varphi_{\text{total}} = \varphi_{\text{nat}} * \varphi_{\text{pressure}} * \varphi_{\text{Doppler}} * \varphi_{\text{micro}} \approx \varphi_{\text{Lorentz}} * \varphi_{\text{Gauss}}\,, \tag{1.20}$$

because the natural and pressure broadening line profiles are (mostly) Lorentzian, the thermal Doppler and microturbulent line profiles are purely Gaussian, and the convolution of two Lorentzians or Gaussians yield another Lorentzian or Gaussian, respectively. For the execution of the convolution the damping constants and Doppler parameters can be added up to $\gamma_{\text{total}} = \gamma_{\text{nat}} + \gamma_{\text{pressure}}$ and $\Delta\nu_{\text{Doppler}} = \frac{\nu}{c}\sqrt{v_{\text{th}}^2 + \xi^2}$, respectively. These in turn can be combined to a new damping parameter $a = \gamma_{\text{total}}/4\pi\Delta\nu_{\text{Doppler}}$. Furthermore, by defining a new running variable $x = \Delta\nu/\Delta\nu_{\text{Doppler}}$ the final line profile can be derived to

$$\varphi_{\text{Voigt}}(x) = \frac{1}{\sqrt{\pi}}\frac{a}{\pi}\int_{-\infty}^{\infty}\frac{e^{-y^2}}{(x-y)^2+a^2}\mathrm{d}y = \frac{1}{\sqrt{\pi}}H(a,x)\,. \tag{1.21}$$

The integral is called the Hjerting function $H(a,x)$ an can not be solved analytically. It needs to be approximated numerically or by a series expansion. Figure 1.11 shows the Hjerting function in dependency on the damping parameter a. For $a \to 0$ the line profile approaches a purely Gaussian shape and for higher values of a, Lorentzian wings become dominant. This exactly represents the general shape of spectral line profiles: A Doppler broadened line-core and pressure-broadened wings on both sides of the line.

There are two further extrinsic broadening mechanisms that should be taken into account in many observations:

- **Stellar rotation:** If stars are observed, they might rotate with some equatorial velocity v_{rot}. This leads to changing Doppler shifts across the stellar disk, because one side of the

star is continuously approaching, while the other one is receding. This causes additional broadening, but after the formation of the line in the stellar atmosphere. Hence, it conserves the equivalent width of the line. In general, one observes only the projected rotational velocity $v_{\rm rot}\sin i$, where i is the inclination angle between the rotation axis and the line of sight. The typical observed range is between zero and a few hundred $\mathrm{km\,s^{-1}}$. The rotational line profile can become quite complicated, because of other effects such as limb darkening that need to be included. For more details see Gray (2021).

- **Instrumental broadening:** The finite spectral resolution $\Delta\lambda$ of spectrographs also leads to a broadening of spectral lines, which is called instrumental broadening. The corresponding line profile can be complex and is usually measured by recording almost infinitesimally narrow emission lines from calibration lamps. However, it is often assumed to be Gaussian:

$$\varphi_{\rm instr}(\lambda) \approx \frac{2\ln 2}{\sqrt{\pi}\Delta\lambda}\exp\left(-4\ln 2\frac{(\lambda-\lambda_0)^2}{\Delta\lambda^2}\right) \tag{1.22}$$

  Just like for the rotational broadening, instrumental broadening of course arises after the formation of spectral lines in the corresponding astrophysical environment and therefore conserves the equivalent width.

1.4.4 The Theory of Interstellar Absorption

While all of the line broadening mechanisms from the previous section contribute to the formation of stellar spectral lines and need to be included in the modeling of synthetic spectra (Section 3.1), the conditions in the interstellar medium are quite different. Since the material is much thinner, pressure broadening can be usually neglected as well as (obviously) stellar rotation. Furthermore, the temperatures are either extremely low for cold interstellar clouds or at least lower than for early type O and B stars in warm clouds. Hence, thermal Doppler broadening is also weaker than for these stars. Overall, one is left with natural broadening, weak thermal broadening, and microturbulence as intrinsic broadening mechanisms. From these, the overall line profile of interstellar absorption lines can be derived, depending on physical parameters like the temperature and the density of absorbers.

The damping parameter of natural broadening $\gamma_{\rm nat}$ is related to the lifetime of the considered transition, which in turn is related to the inverse of the transition probability A_{lu} from the lower (l) to the upper (u) state. By converting the previous expressions from frequencies ν to wavelength λ, the overall damping parameter a can be written as

$$a = \frac{\lambda_0 A_{lu}}{4\pi b} \qquad \text{with} \qquad b = \sqrt{v_{\rm th}^2 + \xi^2}\,. \tag{1.23}$$

The quantity b, consisting of the thermal[17] and microturbulent velocity of the particles, is called the Doppler parameter. For interstellar conditions it is typically in the range of $b = 1-10\,\mathrm{km\,s^{-1}}$ (Welty et al., 1994; Redfield & Linsky, 2002; Welsh et al., 2010), at least for Ca II and Na I lines. Furthermore, the wavelength-dependent transmission of a spectral line is given by $I(\lambda) = \exp(-\tau(\lambda))$, where τ is the so-called optical depth. It is a dimensionless measure of how opaque a medium is to light, quantifying the cumulative absorption and scattering along the line of sight and is given by (Savage et al., 1977)

$$\tau(\lambda) = \frac{\sqrt{\pi}e^2}{m_e c}\cdot N\cdot b^{-1}\cdot f_{lu}\cdot\lambda_0\cdot H(a,x)\,. \tag{1.24}$$

The first factor just consists of general constants (elementary charge e, electron mass m_e, and speed of light c). N is the column density, which is defined as the number of absorbers along the line of sight and therefore has the unit $\mathrm{cm^{-2}}$. b is the Doppler parameter, f_{lu} the oscillator strength of

[17] The most frequent thermal velocity of a gas in thermal equilibrium depends on the temperature and is given by $v_{\rm th} = \sqrt{2k_{\rm B}T/m_{\rm u}A}$, where $m_{\rm u}$ is the atomic mass unit and A the atomic weight of the considered gas particles.

the transition, λ_0 the rest wavelength, and $H(a,x)$ the Hjerting function defining the line profile. The latter one includes the damping parameter a, which depends on the transition probability, the rest wavelength, and the Doppler parameter. Overall, apart from the atomic constants λ_0, A_{lu}, and f_{lu}, the appearance and shape of interstellar absorption lines $I(\lambda)$ solely depends on the column density N of absorbers and on the Doppler parameter b.

This is exactly the absorption model that is applied to the interstellar Ca II H&K as well as Na I D1&D2 lines (Subsection 3.3.4). From Voigt profile fits including the atomic constants from Table 1.1, the column density N and the Doppler parameter b can be derived. However, additionally the instrumental broadening due to finite spectral resolutions needs to be considered. Since ISM lines are usually very narrow, they are quickly dominated by $\Delta\lambda$, if the resolution is low. Then the intrinsic width of the lines determined by b is no longer resolved. It needs to be fixed to a reasonable value in order to derive at least reliable column densities (see Subsection 3.3.4).

1.5 Outline of the Thesis

The distribution and properties of interstellar material at high Galactic latitudes is poorly studied yet due to a lack of available Galactic background sources, which can be observed out to the halo. This thesis presents hot subdwarf stars as possible light bulbs to shine through the ISM predominantly at high Galactic latitudes. They represent a late stage of stellar evolution formed via binary interaction and are observable out to the Galactic halo. The first main aim of this work is to probe the ISM using hot subdwarf background sources. The second main aim is to search for CSM that possibly remained after their formation around hot subdwarfs using the same analysis methods as for the ISM.

In Chapter 2 the used surveys and instruments are summarized as well as the collected sample of hot subdwarf stars. Then in Chapter 3 the methods and tools are discussed, which are used to carry out the analyses of this work. On the one hand, they comprise pre-existing ones, like stellar model atmospheres and synthetic spectra or a spectral energy distribution (SED) fitting routine. On the other hand, a set of new Python scripts is developed in order to carry out comprehensive spectral analyses of both, stellar atmospheres as well as interstellar absorption lines. These scripts are not restricted to the needs of this study, but could be of general use for spectral analyses. Hence, Section 3.3 can be considered as kind of manual what the programs exactly do and how they are executed. The pipeline of the conducted analyses is outlined in Chapter 4. Furthermore, Chapter 4 includes a brief summary about the measured atmospheric and stellar parameters for the hot subdwarf stars, which are observed as a valuable side-effect of the analysis, as well as a simulation of measured interstellar parameters depending on the spectral resolution and signal to noise ratio of synthetic interstellar absorption lines. The simulation serves as preparation for the application of the scripts on real data and the quantification of possible systematic uncertainties. Detailed investigations of the interstellar medium are presented in Chapter 5:

- **Parameter distributions:** The raw distributions of the measured interstellar parameters, like column densities, equivalent widths, radial velocities, and reddening are evaluated on a statistical basis.

- **Parameter correlations:** The measured parameters are related in order to see how they correlate with each other. On the one hand, this serves as consistency check and on the other hand it provides new insights about the connection between calcium and sodium abundances in interstellar clouds.

- **Distribution of interstellar material on the sky:** Two dimensional maps are created in order to trace the distribution of interstellar gas and dust on the sky.

- **Variation of interstellar parameters:** It is investigated, how interstellar parameters vary with distance, direction, and height from the Galactic plane. In particular, the deviation of interstellar properties between the Galactic disk and the halo is probed.

- **Interstellar gas to dust ratios:** The connection between the abundances of interstellar gas and interstellar dust in the ISM is examined. It is investigated how the gas to dust ratios vary with distance, direction, and height from the Galactic plane.

As second part of this thesis it is searched for circumstellar material (CSM) around hot subdwarf stars in Chapter 6 using the same methods as before. Although CSM is expected to remain bound around these stars after their formation through common envelope evolution and ejection, detections of such material have been elusive yet. Chapter 6 comprises the following:

- **Overview:** At the beginning an overview about the physics of common envelope evolution and ejection is provided, from unstable Roche lobe overflow in binary interaction, to the formation of a common envelope, and its final ejection. Also the possible outcomes concerning the stellar remnants are discussed.
- **Recent detections:** Recent detections of ubiquitous CSM around hot subdwarf stars are claimed by Li et al. (2025). They did not only use similar methods, but also spectra that are part of the sample of this work. Therefore, their findings are evaluated and compared to the results of this study.
- **Search for new CSM candidates:** Finally, it is searched for new CSM candidates within a sample of higher resolution spectra in order to minimize uncertainties coming from low spectral resolutions. The main criterion for potential CSM is that the material from which Ca II H&K lines originate moves exactly with the velocity of the stars, while an interstellar origin can be most likely excluded.

Eventually, the findings and conclusions of this work are summarized in Chapter 7. Moreover, it is highlighted how the investigation of the interstellar medium using hot subdwarf stars as background light bulbs as well as the search for circumstellar material around these stars could be extended in the future.

Chapter 2

Surveys and Instruments

This study aims to collect and analyze a large number of stellar spectra. Subsequently, the spectral energy distributions (SEDs) of the corresponding objects are modeled by considering their photometry in different wavelength bands. Last but not least, parallax measurements are required to constrain the distances to the sources. Obviously, it is not possible to record this wealth of data by oneself, especially for several thousands of stars. Plenty of sophisticated instruments have been developed by science organizations and space agencies all around the globe in order to conduct large surveys that record these various kinds of data for the astronomical community.

In the following sections, the central surveys and instruments for this thesis are presented, beginning with the unprecedented *Gaia* space mission in Section 2.1. In Section 2.2 all the different spectroscopic surveys and spectrographs are briefly outlined, with the help of which the investigated sample of this work has been compiled. Lastly, in Section 2.3 an overview about this sample as a whole is provided. What is not considered in this chapter are the photometric surveys from which the data is taken to model the stellar SEDs. The automatic query of these databases as part of the SED fitting routine is discussed separately in Section 3.4.

2.1 ESA's *Gaia* Space Mission

The *Gaia* space mission of the European Space Agency (ESA) is the direct successor mission of HIPPARCOS (Clausen & Perryman, 1990; van Leeuwen, 1997) launched in 1989 by ESA as well. HIPPARCOS laid the foundation for space astrometry by measuring the positions, parallaxes[1], and motions of over 118,000 Galactic objects. It achieved a mean parallax accuracy of about 1 milliarcsecond (mas), significantly better than previous ground-based measurements. Since the parallax is directly related to the distance of an object, the HIPPARCOS catalog provided the first reliable distance measurements for a large number of stars (van Leeuwen, 2009) and provided groundwork for subsequent astrometry missions such as *Gaia*.

The main aim of *Gaia* is to compile a complete 3D map of the Milky Way in order to enhance our understanding of the formation, structure, and evolution of the Galaxy. It was launched December 19, 2013 and arrived at its operating point, the second Lagrange point of the Sun-Earth-Moon system (~1.5 million km from Earth) a few weeks later (Gaia Collaboration et al., 2016a). From there it measures the positions, parallaxes, and proper motions for more than 1.8 billion stars with unprecedented precision. It is capable of observing objects down to a limiting magnitude of about 20.7 and reaches astrometric precisions in the μas regime. Furthermore, it provides photometry across the G (generic), BP (blue), and RP (red) passbands and is even able to do low resolution spectroscopy for many objects. It continuously scans the full sky to gain uniform data. The longer it observes the more precise the data for each object becomes and the more objects are in total observed. Therefore, each data release (DR) supersedes the former ones. Originally, *Gaia* was planned to stay operational for five years, until 2019, but then it was still working so immaculate that it entered the extended mode until 2025. Unfortunately, now it is no longer

[1] The parallax angle ϖ is the annual angle a nearby astronomical object subtends with respect to distant background sources due to Earth's motion around the Sun. It is directly related to the distance d to the object: "$d = 1/\varpi$ ".

recording data and with its last remaining fuel it left its site of operation to keep the space free for future missions.

The major milestones of *Gaia* are its consecutive data releases: DR1 was released in 2016 and provided first astrometry and photometry for about 1 billion sources (Gaia Collaboration et al., 2016b; Fabricius et al., 2016; Lindegren et al., 2016). In DR2 (2018) about 1.7 billion objects are included with improved parallaxes and proper motions compared to DR1 (Gaia Collaboration et al., 2018; Arenou et al., 2018). In 2020, astrometric data with unprecedented precision for more than 1.8 billion objects was released as part of an early data release (eDR3; Gaia Collaboration et al., 2021; Fabricius et al., 2021). The most recent data release that complements the astrometric data from *Gaia* eDR3 is full DR3 and was published in 2022 (Gaia Collaboration et al., 2023a; Babusiaux et al., 2023). However, the amount of recorded data by *Gaia* before its retirement will be sufficient to provide us with two more releases. After all the data reduction and processing, DR4 is expected to be published in the near future, while at some point in the next years DR5 will complete the mission with a full data set.

This work makes extensive use of positions, parallaxes, and photometry from *Gaia* DR3. Precise coordinates are crucial to identify stars and query data from various catalogs and data bases. Parallaxes are, of course, used to determine distances to the observed objects. These distances are of central importance, e.g. to see how interstellar properties vary with distance to the background sources (Section 5.4) or to calculate stellar parameters that depend on the distance (Subsection 3.4.2). The photometry is mainly used to model stellar spectral energy distributions (SEDs) as explained in Section 3.4.

2.2 Spectroscopic Surveys and Spectrographs

As central part of this analysis spectra of hot subdwarf stars are investigated in order to derive information about the interstellar medium from interstellar absorption lines present in those spectra. They are not just taken from one predefined source, but collected manually from various surveys and databases. While the majority are low resolution spectra from large spectroscopic all-sky surveys, also a smaller number of higher resolution spectra is publicly available. The different surveys and spectrographs from which the spectra originate are briefly presented in the following.

2.2.1 SDSS

The Sloan Digital Sky Survey (SDSS) was planned in the 1980s-1990s to move from photographic plates to large scale and homogeneous digital sky surveys and officially started operations in 2000 (Szalay, 1999; York et al., 2000). It aims to create a multi-color, three-dimensional map of the nearby universe in order to constrain the models of cosmology and of galaxy formation and evolution (Loveday, 2002). To this end, it photometrically and spectroscopically observed several million celestial objects like galaxies and quasars, but also stars, down to a limiting magnitude of about 23. With that, it became a paradigm for open, public survey data in astronomy.

SDSS makes use of the dedicated 2.5 m wide-field Sloan Telescope at the Apache Point Observatory in New Mexico, USA. With a field of view of about 7 square degrees per exposure and a fiber-fed spectrograph with more than 1000 fibers, it scans about one half of the whole Northern sky. The recorded spectra have a spectral resolving power of about $R \sim 2,000$ and cover a spectral range from 3,600 Å to 10,000 Å (Smee et al., 2013).

The SDSS survey is conducted in multiple successive phases. From SDSS-I to SDSS-IV (2000-2020) the scientific goals as well as the instrumentation got permanently updated and extended. The SDSS-V phase (since 2020) also adopts Southern observatories to broaden the sky coverage (Kollmeier et al., 2017), but this extension is still ongoing. The SDSS survey regularly publishes public data releases. This work makes use of spectra from the first data release of SDSS-V. Including all releases from the previous phases, it represents in total DR18 of the overall SDSS survey (Almeida et al., 2023).

2.2.2 LAMOST

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), also called the Guo Shoujing Telescope, is used to carry out a spectroscopic survey with the same name (Cui et al., 2012; Zhao et al., 2012). It was developed by the Chinese Academy of Sciences and is operational since 2012. Its motivation is to complement other astrometric and photometric surveys like *Gaia* or SDSS with massive spectroscopic follow-up. To this end, it recorded millions of spectra yet for various kinds of objects, but mainly stars.

The telescope is located at the Xinglong Station in the Hebei Province, China, has an effective aperture of about 4 m, and a field of view of 5° in diameter. With a limiting magnitude of about 19 it is not capable of observing as faint objects as SDSS, but with 4,000 fibers feeding in total 16 spectrographs it can record spectra for a lot of objects at once. With more than 10 million recorded spectra it was the world's largest spectroscopic database. LAMOST offers two distinct modes: The low resolution (LRS) mode with $R \sim 1,800$ and a spectral coverage of 3,700-9,000 Å, which is used for the vast majority of recorded objects, and a medium/high resolution (MRS) mode with $R \sim 7,500$ or higher and a shorter coverage (two arms with 4,950-5,350 Å and 6,300-6,800 Å, respectively) for more dedicated programs. Due to its operation site it is restricted to the Northern hemisphere.

Just like SDSS, LAMOST follows an open data policy with staged public releases. This work makes use of spectra from the most recent tenth data release (DR10) of LAMOST LRS. The MRS spectra do not cover the spectral range of the considered interstellar absorption lines.

2.2.3 DESI

The Dark Energy Spectroscopic Instrument (DESI) is a next-generation spectroscopic survey carried out with the 4 m Mayall Telescope at Kitt Peak National Observatory, Arizona, USA (DESI Collaboration et al., 2016a). Its main aims are to measure the expansion of the universe, probe dark energy, and to map the 3D distribution of galaxies and quasars with high precision. The survey is operational since 2021 and is planned for a total duration of five years. Due to its operation site it observes the Northern hemisphere and the target classes are mainly galaxies and quasars, but additionally also Milky Way stars are observed. The expected yield in the end are about 40 million spectra, which would make it the largest spectroscopic survey ever undertaken, even outnumbering LAMOST.

The telescope has a 3.2° field of view in diameter and 5,000 fibers feed the in total 10 spectrographs. The spectra have about $R \sim 3,000$ in the optical and a wavelength coverage of 3,600-9,800 Å (DESI Collaboration et al., 2016b).

Apart from an early data release in 2023 (DESI Collaboration et al., 2024), the first full data set (DR1) was published recently this year (DESI Collaboration et al., 2025). It comprises about 18.7 million objects, of which about 4 million are stars. Serendipitously, the release took place before the finishing of this work. Therefore, DESI spectra from DR1 could be included in the investigated sample.

2.2.4 FEROS

The Fiber-fed Extended Range Optical Spectrograph (FEROS) was developed in the mid-1990s. It had first light in 1998 on the European Southern Observatory (ESO) 1.52 m telescope at the La Silla Observatory in Chile and was moved to the ESO 2.2 m telescope at La Silla in 2002, where it still operates. The motivation for FEROS' development was high resolution optical spectroscopy, in order to enable precise stellar radial velocity, atmospheric parameter, and chemical abundance measurements as well as the search for exoplanets in the Southern hemisphere.

FEROS is a fiber-fed echelle spectrograph with a spectral resolving power of $R \sim 48,000$ and a spectral range of 3,700-8,600 Å (Kaufer & Pasquini, 1998). It contains only two fibers, one for the science target and one for calibration purposes, e.g. wavelength calibration or sky background subtraction. Hence, it is not capable of recording spectra for multiple objects at the same time.

Unlike SDSS, LAMOST, or DESI, it is not part of a large all-sky survey, but the observing mode is given by principal investigator (PI) driven proposals. However, after a standard proprietary period the science ready spectra become publicly available in the ESO archive (Retzlaff et al., 2016; Mascetti et al., 2024), which also contains lots of spectroscopic data for hot subdwarf stars.

2.2.5 SPY/UVES

The Supernovae type Ia Progenitor surveY (SPY) was a spectroscopic survey dedicated to search for possible supernova type Ia progenitor systems, i.e. white dwarfs in close binary systems, by utilizing the Ultraviolet and Visual Echelle Spectrograph (UVES).

The UVES Spectrograph

UVES is installed on ESO's 8.2 m Unit Telescope 2 (Kueyen) as part of the Very Large Telescope (VLT) at the Paranal Observatory in Chile. It had first light in 1999 and is still one of the VLT's most heavily used high resolution instruments. It provides high resolution spectroscopy for a broad range of astrophysical conditions like the observations of stars or extragalactic objects in the Southern hemisphere.

It is an echelle spectrograph with configurable resolution of $R \sim 40,000 - 115,000$ depending on the chosen setup (Dekker et al., 2000). The wavelength range is typically 3,200-10,500 Å, but is also adjustable. Similar to FEROS, the observing mode is given by PI driven proposals and after some time, the recorded and science ready spectra are publicly available in the ESO archive.

The SPY Survey

The SPY survey was conducted with the UVES spectrograph in the early 2000s (Napiwotzki et al., 2001, 2020). The targets were about 1,000 white dwarfs and each object was observed at least twice in different nights in order to detect radial velocity shifts indicative of a close binary system (Napiwotzki et al., 2005). The optimal settings of UVES for this survey lead to a spectral resolution of $R \sim 18,500 - 20,000$, which is below the usual values for the spectrograph. Main reason for this setup was that white dwarfs are intrinsically faint objects and need a long exposure time, when observing them with high spectral resolution. Therefore, the settings were optimized for survey speed rather than resolution (Napiwotzki et al., 2003).

However, the SPY sample is contaminated with other objects than white dwarfs, like hot subdwarf stars, due to their photometric similarities. Lisker et al. (2005) identified the subdwarf B stars among the SPY sample, while Stroeer et al. (2007) identified the subdwarf O ones. All their spectra are available in the ESO archive and used in this work.

2.2.6 X-Shooter

X-Shooter is a medium resolution echelle spectrograph (D'Odorico et al., 2006). Like UVES, it is installed on ESO's 8.2 m Unit Telescope 2 (Kueyen) as part of the Very Large Telescope (VLT) at the Paranal Observatory in Chile. It had first light in 2009 and can be considered as second-generation VLT instrument. Its motivation was to provide simultaneous spectroscopy from the ultraviolet (UV) to the near-infrared (NIR) in one shot (Kaper et al., 2009). Similar to FEROS and UVES, it supports a broad range of astronomical science cases.

X-Shooter consists of three arms: An UVB arm (UV/blue, 3,000-5.600 Å), a VIS arm (optical, 5,600–10,200 Å), and a NIR arm (near-infrared, 10,200–24,800 Å). The spectral resolution varies depending on the arm and the chosen settings. In general, it is $R \sim 4,000 - 17,000$ (Vernet et al., 2011) with about $R \sim 9,800$ in the optical. The spectrograph is primarily designed for a high light throughput and efficiency to maximize the photon collection. Thus, it works more critically for faint targets, but often reaches high signal to noise ratios for brighter ones.

Just like FEROS and UVES, X-Shooter is operated on a PI proposal basis and the science ready spectra are publicly available in the ESO archive after a certain time. This includes a little number of spectra for hot subdwarf stars, which are included in the collected sample.

2.3 Overview of the Collected Sample

In order to query data from the databases of the large spectroscopic surveys and the ESO archive, an input list of targets is needed. For this purpose, the comprehensive catalog of hot subdwarf candidates by Culpan et al. (2022) is used. It contains 61,585 stars that were selected mainly based on color and magnitude criteria from *Gaia* eDR3.

This catalog was crossmatched with all databases, resulting into in total 17,467 spectra. From Table 2.1 it can be seen that the majority comes from the large all-sky surveys, especially SDSS and LAMOST. This means that most of the data are low resolution spectra with only a small contribution from higher resolved ones. It also has to be mentioned that Table 2.1 just represent the raw number of spectra, but many objects were observed multiple times. The number of unique stars is actually smaller and later on the individual results for each source are averaged (see Section 4.1).

In Figure 2.1 the quality of the spectra in terms of signal to noise ratio (SNR) is shown. In the overall distribution (left panel) it can be seen that the SNR drops off exponentially towards higher values. Most of the spectra are rather noisy with SNRs below 25. The peak at 200 comes from an overflow bin, which clips the distribution. Otherwise the tail would be even longer. In the right panel the distribution is shown per survey/spectrograph in different colors. As expected, X-Shooter spectra have on average the highest SNR with a mean value close to 100. All others peak well below 50. While SDSS, LAMOST, and FEROS all at least show a tail towards better SNR, DESI and SPY/UVES have on average the lowest quality. This is, however, no performance indicator for the used instruments, but rather a consequence of the properties of the observed objects. Both surveys are dedicated to record spectra of faint objects. Therefore, often long exposure times are needed, which naturally increases the noise in many cases. The SNR will later on play a crucial role whether interstellar lines are measurable in the spectra, in particular for low resolution spectra (see Section 4.3). Strict quality criteria are applied, which further reduces the number of objects of the investigated sample. Moreover, the sample is slightly contaminated with sources, which are apparently no hot subdwarf stars and need to be removed. The processes of cleaning the sample and averaging the results for multiply observed objects are discussed in Section 4.1 and Chapter 5.

One major goal of this thesis is to investigate the properties of the interstellar medium across the nearby Galaxy, especially at high Galactic latitudes, and at best to create interstellar gas and dust maps. To this end a good coverage of the sky by the background light bulbs is desired. In Figure 2.2 the distribution of the collected sample of hot subdwarf stars is shown in equatorial coordinates. The footprints of the large all sky surveys[2], which mostly cover only the Northern hemisphere, are clearly visible. Furthermore, they leave out a band ranging from the equatorial plane at a right ascension of about 75° towards the North pole and back to the equatorial plane at about 285°. This band corresponds to the Galactic disk, which is crowded with an extremely high number of young and bright stars. Due to these crowded fields with lots of blended objects and because the surveys are in general dedicated to observe fainter objects, this band is mostly excluded. Moreover, only a few stars are observed near the North celestial pole. This is most likely an observational limitation, because the pole region lies at high declinations for the considered telescopes and to point exactly overhead towards the zenith is mechanically difficult for all the heavy instrumentation. On the other hand, the Southern hemisphere is only covered very sparsely by FEROS, SPY/UVES, and X-Shooter observations. However, this underlines the importance of this rather small number of higher resolution spectra. They do not only provide much more spectral information than the low resolution ones, but are also solely responsible for covering at least some parts of the Southern hemisphere.

While equatorial coordinates are best suited for the orientation on Earth, Galactic coordinates are commonly used to represent the orientation within the Milky Way. Therefore, Figure 2.3 shows how Figure 2.2 is translated into the distribution of the sample in Galactic longitude l and Galactic latitude b. Here, the Galactic disk corresponds to the plane of zero latitude, which is mostly left empty as expected. Another two regions are covered very poorly: One is centered at $(l,b) \approx (120°, 30°)$ and corresponds to the celestial North pole region, while large parts of the

[2] Even though they cover at most half the sky, they are still called ”all sky“ surveys.

Table 2.1: Raw number of queried spectra N_{spectra} for each survey/spectrograph and in total. Many stars were observed multiple times, hence the number of unique objects is actually smaller.

	SDSS	LAMOST	DESI	FEROS	SPY/UVES	X-Shooter	In total
N_{spectra}	4,972	10,131	981	961	278	144	17,467

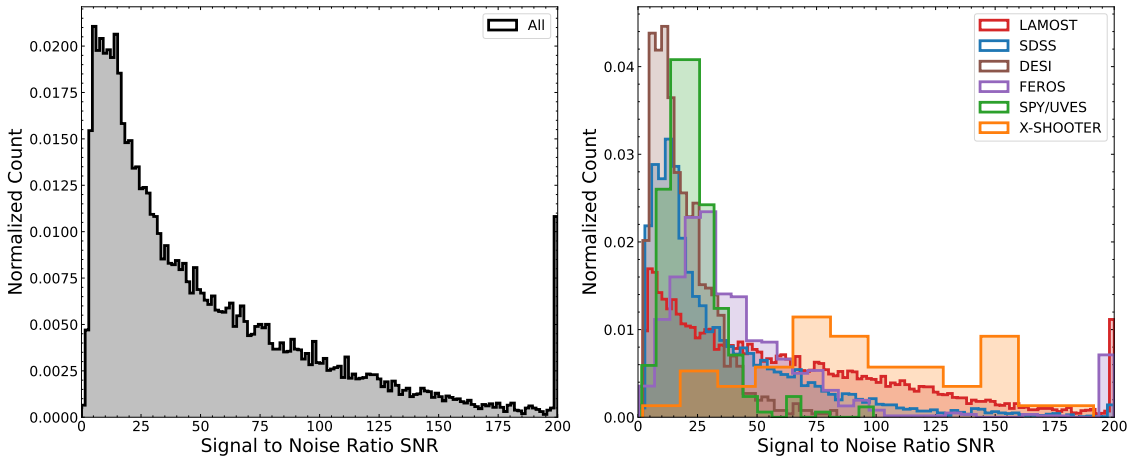


Figure 2.1: Signal to noise ratio (SNR) distribution for the collected sample of spectra. The left panel shows the overall distribution, while the right panel distinguishes between the different surveys/spectrographs in different colors. Each histogram contains $\sqrt{N}$ equally wide bins in the considered range, where N is the number of data points. This is adopted for all other distributions of this kind in this work.

Eastern side are empty, corresponding to Earth' Southern hemisphere. The rest of the sky is fairly well covered by background sources.

Figure 2.4 shows the Galactic latitude distribution of the sample. The dip at $b = 0°$ corresponding to the Galactic disk is nicely visible and confirms that this study mainly examines the properties of the interstellar medium towards higher Galactic latitudes. However, Figure 2.4 also shows an imbalance between the number of background sources on the Northern and Southern Galactic hemisphere. Unfortunately, this will introduce a little observational bias later on. If this is the case, it is mentioned at the respective point of this work.

Overall, it can be said that despite the rather exotic nature of hot subdwarf stars, a large sample of them has been collected to act as background light bulbs to shine through the interstellar medium. But even though a complete coverage of the sky is desired, this is essentially only the case for Earth' Northern hemisphere. The contribution from future large spectroscopic surveys, like SDSS-V (Kollmeier et al., 2017) or 4MOST (de Jong et al., 2014, 2022), is needed to reduce this issue.

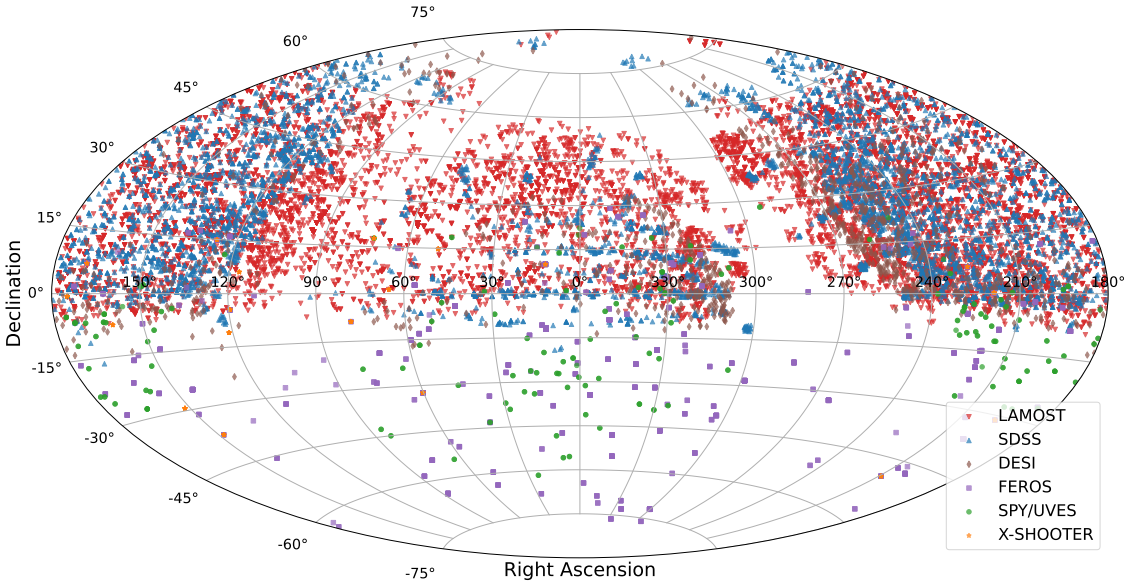


Figure 2.2: Aitoff projection of the distribution of the collected sample of hot subdwarf stars on the sky in equatorial coordinates (right ascension and declination). The corresponding survey/spectrograph in the course of which the objects were observed are indicated by different symbols and colors.

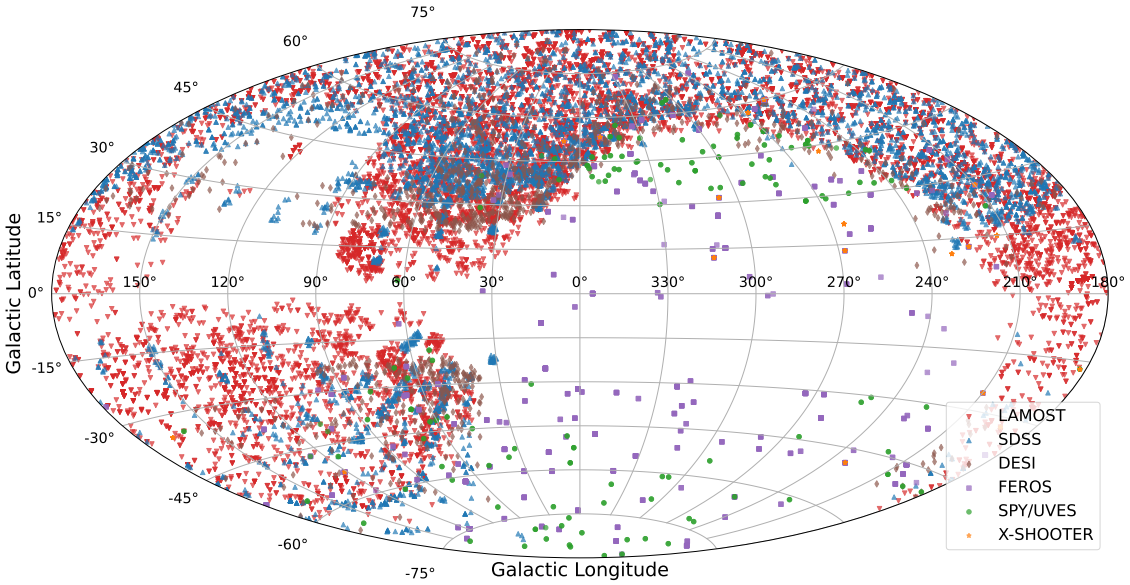


Figure 2.3: Same as Figure 2.2, but in Galactic coordinates (Galactic longitude and latitude).

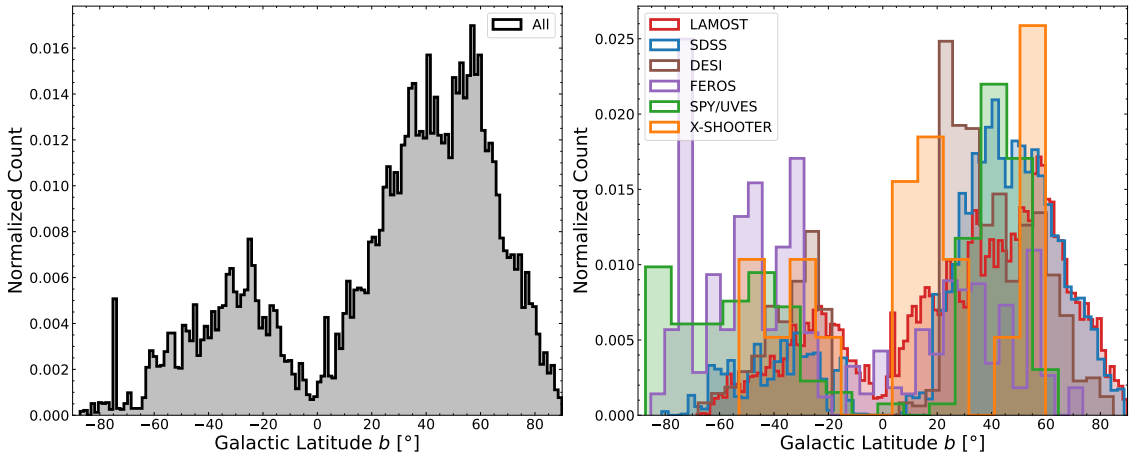


Figure 2.4: Galactic latitude distribution of the collected sample of background sources. The left panel shows the overall distribution, while the right panel distinguishes between the different surveys/spectrographs in different colors. A clear dip can be seen at $b = 0°$ corresponding to the Galactic disk as well as an imbalance between both Galactic hemispheres.

Chapter 3

Methods and Tools

For every investigation or analysis a kit of methods and tools is required. Those either already exist and can be readily used, at most with some adjustments for the specific needs, or have to be newly developed.

This chapter is devoted to all pre-existing as well as developed methods and tools used in the course of this work. Those are grids of model atmospheres and synthetic spectra discussed in Section 3.1, the theory and application of χ^2-minimization explained in Section 3.2, an extensive description of self-developed Python programs in Section 3.3, and a spectral energy distribution fitting routine outlined in Section 3.4. Since the core of the whole analysis was conducted using the self-developed Python scripts, they are discussed in great detail. Furthermore, they are not only limited to the purpose of this thesis, but could be of general interest, e.g. for the spectral analysis of stellar spectra or interstellar absorption lines. Therefore, Section 3.3 can be considered as kind of manual what the scripts exactly do and how they can be executed and applied to scientific data.

3.1 Model Atmospheres and Synthetic Spectra

Stars produce energy in their interiors by nuclear fusion. This energy is transported from the core to the surface mainly by radiation and/or convection[1]. On their way, photons are continuously absorbed and re-emitted until they reach a layer that is optically thin enough such they can escape from the star. This layer is by definition the surface of a star and called its photosphere. It represents the deepest part of the stellar atmosphere, which also comprises the chromosphere, the corona, and stellar winds. Photons emitted from the bottom of the photosphere can interact with particles of the outer atmosphere, which for example causes spectral lines. Most of what we can directly observe from stars are exactly those photons. The result of collecting this light and splitting it into its constituent wavelengths by a dispersion element is called a spectrum.

To deduce information about the atmosphere of a star, like its temperature, density, or chemical composition, recorded spectra need to be compared with theory. To this end, model atmospheres and synthetic spectra are calculated. A model atmosphere reflects the temperature and density stratification in the outer layers of a star, while a synthetic spectrum represents the emergent flux from these layers. The construction of those models is briefly outlined in this section. Since the whole procedure is quite complex, it is focused on the basic principles, which are explained in a qualitative manner. More quantitative discussions of the physics involved can be found for instance in Hubeny & Mihalas (2014), Irrgang (2014), Gray (2021), or Dorsch (2023).

3.1.1 Standard Assumptions for Model Atmosphere Calculations

In order to simplify the complex task of model atmosphere calculations, a number of approximations and assumptions are often made, which are mostly based on the conservation of energy and momentum:

[1] Hot O and B stars are mostly radiative.

- **Plane-parallel geometry:** The thickness of the stellar atmosphere is usually small compared to the radius of a star. Hence, the curvature of the star is negligible for the length scales on which the photon interactions occur.
- **Homogeneity:** The structure and composition of the stellar atmosphere is supposed to be homogeneous in each layer, which means there are no variations in radial direction. Therefore, one single coordinate is sufficient, usually either the dimensionless optical depth τ_ν or the spatial z-coordinate, which increases from the interior of a star to its outside and equals the stellar radius at the surface. However, due to the frequency dependence of τ_ν, the same value of τ_ν does not correspond to the same spatial z-coordinate for different wavelengths.
- **Stationarity:** All time dependent effects, like pulsations, are neglected. The atmosphere is said to be in a "steady state".
- **Hydrostatic equilibrium:** As long as the assumption of hydrostatic equilibrium is valid, i.e. the outwards directed radiation and gas pressure is in balance with the inwards directed gravitational pressure, the whole star is in a stationary state. This means the gravitational acceleration or surface gravity g, which is one of the fundamental parameters of stellar atmospheres since it is a proxy for the (electron) density, is constant along the stellar surface. For $g_{\text{radiation}} > g$ hydrostatic equilibrium breaks down, leading to a dynamic outflow of stellar material into the interstellar medium. Therefore, such stellar winds can not be modeled under this assumption.
- **Radiative equilibrium:** The absorption and emission of photons in each volume element of the stellar atmosphere is in balance, which means the energy is conserved for each point. This directly leads to flux conservation in a plane-parallel geometry and to the Stefan-Boltzmann law $F = \sigma_{\text{SB}} T_{\text{eff}}^4$, where F is the total flux, T_{eff} the effective temperature, another fundamental parameter of the stellar atmosphere, and σ_{SB} the Stefan-Boltzmann constant.
- **Local thermodynamic equilibrium (LTE) vs. non-LTE:** If the mean free path of photons is much smaller than the distances over which the temperature changes, thermodynamic equilibrium can be locally assumed for each layer with roughly constant temperature. This considerably simplifies model atmosphere calculations. The particle velocities follow a Maxwell-Boltzmann distribution (Equation 1.13), the ratios of energy level populations of the same atoms or ions are given by the Boltzmann excitation equation (Equation 1.14), and the number ratio of atoms or ions of an element in subsequent ionization stages is given by the Saha ionization equation (Equation 1.15). However, if LTE can not be assumed, which is often the case for hot and thin atmospheres of early type stars, the Boltzmann and Saha equations need to be replaced by more general ones, namely the so-called kinetic (statistical) equilibrium or rate equations (see, for example, Irrgang (2014) or Dorsch (2023)).

3.1.2 Radiative Transfer

The specific photon intensity $B_\nu(T)$ emitted from the bottom of the photosphere follows a Planck distribution, where ν is the photon frequency and T the local temperature. The radiative transfer of a photon beam through the stellar atmosphere is governed by the fundamental radiative transfer equation, which reads

$$\mathrm{d}I_\nu = -\kappa_\nu I_\nu \mathrm{d}s + \eta_\nu \mathrm{d}s\,. \tag{3.1}$$

I_ν is the specific intensity, κ_ν the total (macroscopic) absorption coefficient, sometimes also called opacity, η_ν the total (macroscopic) emission coefficient, and $\mathrm{d}s$ an infinitesimal distance in direction of the solid angle $\mathrm{d}\Omega$. The ratio of absorption to emission $S_\nu = \kappa_\nu/\eta_\nu$ is called the source function. Now Equation 3.1 can be rewritten to

$$\cos\theta \frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = I_\nu - S_\nu\,, \tag{3.2}$$

where θ is the normal angle to the surface and τ_ν the optical depth, which is a measure of how opaque a medium is, describing how much light is absorbed or scattered as it passes through. It decreases from $\tau_\nu = \infty$ in the interior to $\tau_\nu = 0$ at the outer atmosphere of a star. The photosphere (surface) has $\tau_\nu \approx 1$. Introducing the variable $\mu = \cos\theta$ the emergent specific intensity is given by

$$I_\nu(\tau_\nu = 0) = \int_{\tau_1=0}^{\tau_2=\infty} S_\nu(\tau'_\nu) \exp\left(-\frac{\tau'_\nu}{\mu}\right) \frac{\mathrm{d}\tau'_\nu}{\mu}, \tag{3.3}$$

i.e. essentially by integrating the source function from the interior of a star out to the interstellar medium. If LTE is valid, the source function is identical to the Planck function: $S_\nu = B_\nu$. The exponential factor in Equation 3.3 is called the extinction. Finally, the flux $F_\nu(\tau_\nu = 0)$ of a model spectrum is given by integrating $I_\nu(\tau_\nu = 0)$ over all angles and is typically expressed in the units of erg $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\text{Å}^{-1}$. However, this flux can only be determined with detailed knowledge about the structure of the stellar atmosphere as well as its physical conditions. Furthermore, atomic data and cross sections for all possible interactions of the photons with atoms, ions, and molecules of the atmosphere are required.

3.1.3 Hot Subdwarf Model Grids

Usually, the goal is not to calculate only one specific synthetic spectrum, but a whole grid of them, where every grid point consists of a spectrum for a certain combination of atmospheric parameters. Then in spectral fits, for example, it can be interpolated between those grid points in order to obtain a continuous parameter range. Typically, linear interpolation is sufficient.

The two model grids for hot subdwarf stars used in this work were created by many contributors of the stellar research group at the Karl Remeis-Observatory in Bamberg, mainly by Andreas Irrgang and Uli Heber, and are regularly updated and extended. Since calculating synthetic spectra in particular in non-LTE is an expensive numerical task, a hybrid LTE/non-LTE approach was used. A detailed discussion can be found in Nieva & Przybilla (2007) and Przybilla et al. (2011), who also show that this method is consistent with pure non-LTE calculations for B-type stars. Three different codes were applied to compute the model atmospheres and to obtain synthetic spectra:

1. ATLAS12 (Kurucz, 1996): Used to compute the atmospheric structure in LTE.

2. DETAIL (Butler & Giddings, 1985): An updated and extended version of this code is used to solve the coupled radiative transfer and rate equations in non-LTE. It introduces departure coefficients from LTE for the emergent flux. In former times, this step was often omitted by sticking to pure LTE.

3. SURFACE (Butler & Giddings, 1985): Builds upon the results of `DETAIL` to calculate more detailed line profiles and returns the final synthetic spectrum in both, absolute and normalized form[2].

The atmospheres of hot subdwarf stars predominantly consist of hydrogen and helium. Hence, the synthetic spectra include only absorption lines from these two elements, also because usually almost no metal lines are found in the optical and the H and He lines are sufficient to determine the atmospheric parameters. However, even if metal lines are not used for the spectral analysis, metal line opacity has a strong impact on the atmospheric structure through "line blanketing" (Hubeny & Mihalas, 2014). Metal lines act as "blanket" heating up the continuum-forming layers ("backwarming") and cooling the layer close to the surface. In general, it is quite difficult to implement line blanketing in non-LTE calculations, but it is fairly easy to treat it in LTE. Therefore, the ADS (ATLAS/DETAIL/SURFACE) approach combines the advantages of LTE (fast computation, treatment of metal line blanketing) with the approximated effects of non-LTE.

[2] "Absolute" means the shape of the continuum is included and the flux given in units of erg $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\text{Å}^{-1}$, while "normalized" means the spectrum is divided by the continuum such that it is normalized to unity and only spectral features like absorption lines are visible.

Table 3.1: Effective temperature $T_{\rm eff}$, surface gravity $\log g$, and helium abundance $\log n(\mathrm{He})$ ranges of the sdB and He-sdO model grids.

grid	$T_{\rm eff}$ [K]	$\log g$ [cgs]	$\log n(\mathrm{He})$
sdB	15,000 to 55,000 in steps of 1,000	4.6 to 7.0 in steps of 0.2	-5.05 to -0.041 in non-uniform steps
He-sdO	25,000 to 55,000 in steps of 1,000	5.0 to 6.6 in steps of 0.2	-1.05 to -0.001 in non-uniform steps

The two model grids used in this work are the so-called sdB and He-sdO grids. The former one comprises a wide range of atmospheric parameters for typical sdB stars, while the latter one is an extension of the sdB grid towards higher helium abundances. The parameter ranges of both grids are listed in Table 3.1, where $\log n(\mathrm{He}) \coloneqq \log n(\mathrm{He})/n(\mathrm{H})$ is the helium abundance given as number density ratio of helium to hydrogen. It can be converted to the sometimes more commonly used helium abundance $\log y \coloneqq \log n(\mathrm{He})/n(\mathrm{All}) \approx \log n(\mathrm{He})/(n(\mathrm{H})+n(\mathrm{He}))$ according to

$$\log y = -\frac{\ln\left(10^{-\ln\log n(\mathrm{He})}-1\right)}{\ln 10}\,, \tag{3.4}$$

which is the number density ratio of helium to all other elements. These two model grids are used for the entire spectral analysis of this work.

3.2 χ^2-Minimization

Comparing theoretical models to observations in order to derive physical parameters is a key principle across all fields of science. A common technique to automatically carry it out using computer programs is the χ^2-minimization. Also in this work it is made extensive use of the method. Therefore, it is briefly discussed in this section how it works and how it is implemented in the developed Python scripts, which are explained in Section 3.3.

3.2.1 Overview

The aim of χ^2-minimization is to extract meaningful parameters from observations by finding the best matching model. Also it allows to assess how well a model fits the data involving the uncertainties on the data points. In general, it tries to minimize the χ^2-statistic, which is defined as

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{y_i - f(x_i;\theta)}{\sigma_i}\right)^2, \tag{3.5}$$

where N is the number of data points, y_i the observed data, $f(x_i;\theta)$ the model function with a set of fit parameters θ, and σ_i the uncertainties on the recorded data. Minimizing the χ^2 yields the parameters θ that describe the observation best under the given model function. The lower the χ^2-value, the better the agreement between model and observed data, but always assuming that the uncertainties are correctly estimated. The resulting value can then be used further to evaluate the quality of the fit. If the reduced χ^2,

$$\chi^2_{\rm red} = \frac{\chi^2}{N - N_\theta}\,, \tag{3.6}$$

where the difference between the number of data points N and the number of fit parameters N_θ is called the degrees of freedom, roughly equals unity, the fit is considered as good and consistent with the assumed uncertainties. However, if $\chi^2_{\rm red} \gg 1$, the fit is considered as poor or the uncertainties are underestimated and if $\chi^2_{\rm red} \ll 1$, the data is possibly overfitted (the number of data points is

not sufficient to constrain the model) or the uncertainties are overestimated. From this it can be seen that it is important to properly estimate the uncertainties on the observed data. Furthermore, they should be normally distributed. Otherwise, they can bias the fit or lead to invalid conclusions about the fit results. Common practice is also to fit the data once with the estimated uncertainties, rescale them with the "noise scale factor" nsf $= \sqrt{\chi^2_{\mathrm{red}}}$, and to fit everything again with the updated uncertainties. It can be shown that then the final χ^2_{red} always equals unity, meaning the resulting uncertainties are properly estimated, if the model is suited to reproduce the observation. What can also give conclusion about the goodness of a fit are the χ values

$$\chi_i = \frac{y_i - f(x_i;\theta)}{\sigma_i}\,, \tag{3.7}$$

which are often shabbily called "residuals", while residuals are in principle just the differences between the observed and model data points without weighting them with the uncertainties. However, during the course of this work when saying "residuals", always the χ-values are meant. Now, if the residuals show unusual high values somewhere, it can either mean that the model is not sufficient to reproduce the observation there, or hint towards irregularities in the recorded data in this region. If this is the case, it is possible to increase the uncertainties on these data points to improve the fit, because they are then weighted less. Of course, it should be taken care that the uncertainties are always manipulated in a reasonable way such that not just every model can arbitrarily match the data and result into a good fit. Plotting the data together with the fit as well as the residuals is always recommended to check the outcome visually.

Finally, to carry the whole minimization procedure out, a fitting framework as well as efficient minimization algorithms are needed, which are both discussed in the next two sections.

3.2.2 The LMFIT Framework

One flexible and user-friendly open-source Python library for non-linear least-squares minimization and curve-fitting problems is the LMFIT package[3] (Newville et al., 2014). Essentially, it is built on top of the widely used Scipy library (Virtanen et al., 2020), especially `scipy.optimize`, but allows for an easier handling of complex fitting tasks and offers several key features. The first one is the creation of a `Parameters` object to define the fit parameters with initial values, bounds (minimum/maximum value), and further constraints. It is even possible to link different fit parameters via mathematical expressions. Fit parameters can be fixed during the fit or allowed to vary. Furthermore, it provides a `Model` class that can wrap user-defined functions into fit-ready models and supports automatic handling of residuals and χ^2 computation. Also a lot of pre-defined models for common fits (linear, Gaussian, Lorentzian, etc.) are accessible. However, for a better control and to avoid too many "black boxes", it was decided to not use the `Model` class for the development of the scripts. Instead, an objective function ("fit function") is always manually set up, which calculates or defines the model and computes and returns the residuals according to Equation 3.7. Both, the `Parameters` object as well as the fit function are then combined to a `Minimizer` object. This `Minimizer` object is finally passed to the minimization routine with a selected algorithm (see Subsection 3.2.3) and a `FitResult` object is returned, which contains all relevant fit results. Those are, for example, the best-fit parameter values, properly estimated uncertainties and correlations, and the fit statistics. It also includes a complete fit report that can be printed out in the terminal to directly check the outcome.

A recommended way of setting up the whole LMFIT framework, which is also preferred in the developed scripts, is to split all the different steps into distinct functions. One function is of course the fit function, which needs to be created anyway. Another one is the function that defines the fit parameters, creates the `Parameters` object, and combines it together with the fit function to a `Minimizer` object. The third one is the "main" function that runs the fitting procedure on the `Minimizer` object and, for example, structures as well as executes the initial fit, the rescaling of

[3] https://lmfit.github.io/lmfit-py/

the uncertainties, the refitting, and all other steps that might be necessary. However, this is just one way of structuring it and the LMFIT package offers a great flexibility with lots of possibilities.

Lastly, for every fitting run a minimization algorithm needs to be chosen. It is another advantage of LMFIT that many commonly used algorithms are already implemented and just need to be selected in the fitting procedure without the need to change the objective function. Therefore, some of the most common ones are discussed in the next section.

3.2.3 Minimization Algorithms

A lot of numerical methods have been invented to solve minimization problems, to which also the χ^2-minimization belongs. In general, one can imagine the χ^2 (Equation 3.5) in dependency on the set of fit parameters θ as multidimensional landscape. For instance, if θ consists of only two parameters, the landscape is three dimensional with the parameter values as "sea level ground" axes and the χ_i^2 values as "elevation" axis. A lot of hills and valleys, heights and dips arise, which are technically minima and maxima of the χ^2-distribution. The task of the minimization algorithm is to begin from a certain starting point (the initial values) and to search for minima in the landscape, in the best case the global minimum. Most of them work quite differently, all with advantages and disadvantages. Some are very efficient, but strongly depend on the choice of initial values, because they are just capable of finding the next local minimum, while others are very computation heavy, because they always try to find the global minimum. Some are well suited to properly estimate the uncertainties on the determined fit parameters, some are not. Hence, depending on the circumstances of the fitting problem, a proper choice of the minimization algorithm is essential. The in this work commonly used ones are briefly outlined in the following. However, a lot more algorithms are implemented in LMFIT and a list can be found on the website.[4]

- **Leastsq (Levenberg-Marquardt):** The Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) is most likely the most commonly used one for fitting problems. It is implemented with the key word "leastsq" in LMFIT and set as default, if the fitting method is not specified. It is very fast and efficient, but can easily get stuck in a local minimum, if the starting point is far from the global best solution. Hence, a decent initial guess is needed and it is often used when the model is smooth and well-behaved.

- **Basinhopping:** The "basinhopping" method (Wales & Doye, 1997) is designed to explore the whole landscape instead of just the area around the starting point. It combines many local minimization steps with random jumps to escape from local minima. It is useful when it is likely that the problem has several minima and the goal is to find the best one, the global minimum. Typically, it is very computation heavy and using it for complex fits with lots of parameters can become extremely time-consuming.

- **Emcee (Markov Chain Monte Carlo - MCMC):** The stochastic "emcee" method (Metropolis et al., 1953; Hastings, 1970) is strictly speaking no fitting algorithm and actually, does not perform a fit at all. Instead, it just explores the parameter space around a given solution. Therefore, it is not suited to iteratively find a solution in the first place, but to properly estimate the parameter space and uncertainties in a second run, after a solution has already been found with another fitting method. Moreover, some finetuning is needed to use this method. In general, MCMC methods use "walkers", which explore the parameter space in

[4] Two more algorithms that were tested during the development of the scripts, but are not used in the end, are Powell and Nelder-Mead:

- **Powell:** The Powell method (Powell, 1964) implemented with the key word "powell" in LMFIT is a bit more robust and can handle more complex shapes of the χ^2-landscape compared to "leastsq". However, it is slower and less efficient, but sometimes more reliable for complex fitting problems.
- **Nelder-Mead:** Nelder-Mead ("nelder" in LMFIT) is another method that can handle complex fitting problems (Nelder & Mead, 1965). It usually works well even if the χ^2-landscape is irregular or noisy. However, it is often slower and less precise, but a good fallback if other methods do not work.

a certain number of "steps" to build up a full picture of the distribution (see, e.g., Speagle, 2020). Each walker is an independent chain exploring the parameter space and a proper number of walkers needs to be chosen. A rule of thumb is to take $2-4$ times the number of free parameters, where more walkers lead to a better sampling, but also a longer computation time. Also the number of steps per walker needs to be specified. Typical values are between 1,000 and 10,000. Additionally, a "burn" period needs to be chosen. The initial part of the chains depends on the starting positions and should, thus, be discarded. Usually, $10-30\%$ of the total steps are burned. All these settings can be adjusted in LMFIT and when this method is used later on in the developed scripts, always four times the number of free parameters is chosen for the number of walkers, a number of 10,000 for the total steps, and a burn of 2,500 steps (25%).

What the fitting routine technically does, is to iteratively call the fit function once and once again with a different choice of parameters. The chosen algorithm is responsible to decide how this choice is varied from one call to another in order to minimize the residuals/the χ^2-statistic. It also determines, when this goal is reached and the best matching parameters are returned.

3.3 Stellar Python Scripts

This work mainly consists of comprehensive spectral and photometric analyses for a large sample of stars (Section 2.3). Hence, tools are needed to carry out those analyses. A collection of such tools has been developed by Andreas Irrgang and is now maintained by Matti Dorsch. All the programs are written for the data analysis software ISIS[5] (Interactive Spectral Interpretation System) in the scripting language SLang[6] (Houck & Denicola, 2000). One of these programs is a spectral fitting tool that fits synthetic spectra based on model atmospheres to observed stellar spectra in order to determine atmospheric parameters (Irrgang, 2014), a procedure that is also central for this work. However, modifications would have been needed for the purposes here. Therefore, the idea came up to write own scripts in the more commonly used programming language Python. This idea has been realized during the course of this work and four main stellar Python scripts have been developed, which built upon each other. One of them is the script `spectral_fit.py` (Subsection 3.3.3) that is comparable to the ISIS spectral fitting tool. A simple script to read in, process, and display a recorded spectrum is `read_observation.py` (Subsection 3.3.1). The script `interpolate_in_grid.py` (Subsection 3.3.2) reads a selected grid of model spectra and carries out a multidimensional linear interpolation between grid points, which is needed especially for the spectral fitting. The last of the four main scripts is `ism_analysis.py` (Subsection 3.3.4), which runs a comprehensive quantitative analysis of the interstellar absorption lines present in stellar spectra. This functionality is not available in the set of ISIS scripts. The script is crucial for the main aim of this project and actually the most complex and complicated one, because the outcome of the analysis is highly sensitive to the quality (mainly spectral resolution and signal to noise ratio) of the given spectrum. Thus, a lot of effort was needed to make this program as robust as possible by implementing a careful treatment of spectral resolution and noise. The cores of the four scripts constitute the classes `Observation`, `GridInterpolator`, `SepctralFit`, and `ISM`. Before their discussion in the following sections, an overview about their functionality and interplay is illustrated in form of a flow chart in Figure 3.1.

In general, like the ISIS programs, the stellar Python scripts are developed for the use on a Linux operation system as installed on the computation cluster at the observatory. All scripts can be run in the command line and the functionality as well as the usage is explained in the following sections. Since in particular the fitting of observed spectra with models is very common practice in stellar astrophysics, all scripts can be found online on the internal observatory gitlab[7]. For non-observatory members, they can also be provided upon request.

[5] https://space.mit.edu/cxc/isis/

[6] https://www.jedsoft.org/slang/

[7] https://www.sternwarte.uni-erlangen.de/gitlab/weich/stellar-Python-scripts

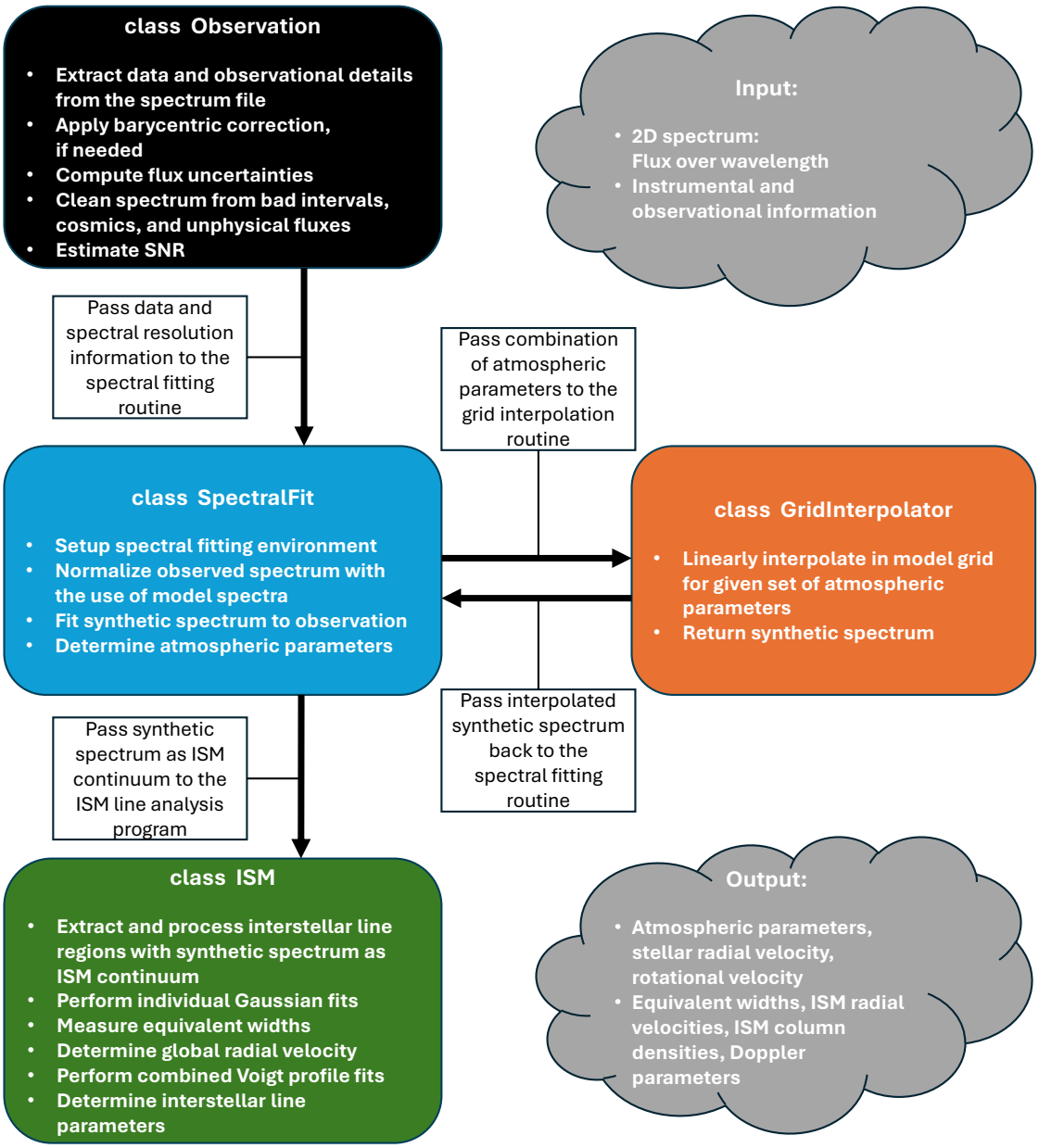


Figure 3.1: Overview flow chart for the functionality and interplay between the four main stellar Python scripts. The input spectrum file containing the wavelength and flux values as well as instrumental and observational information is read and processed by the `Observation` class. The data plus spectral resolution information is passed to the spectral fitting routine `SpectralFit`. Synthetic model spectra are fitted to the observation by successively passing different combinations of atmospheric parameters to the grid interpolation routine `GridInterpolator` until the best matching combination is found. As part of the fitting process, the observed spectrum gets normalized. The results are the atmospheric parameters of the observed star, like temperature, surface gravity, and helium abundance, as well as the stellar radial velocity and the projected rotational velocity. Finally, the synthetic spectrum is passed to the ISM line analysis tool `ISM`, which first extracts and processes the observed spectrum by taking the synthetic spectrum as ISM continuum. Subsequently, the interstellar line analyses are carried out for the four interstellar lines, like individual Gaussian fits, equivalent width measurements, the global ISM radial velocity determination, and combined Voigt profile fits. From the latter one, column densities and Doppler parameters are derived.

3.3.1 `read_observation.py`

The `read_observation.py` script simply reads the observed two dimensional spectrum (flux over wavelength) as well as instrumental properties and observational information (object name, coordinates, time and location of observation, etc.) from the given file, processes the spectrum, and calculates some important parameters like the signal to noise ratio (SNR) and the uncertainty on each flux pixel. It contains the class `Observation`, which provides the functionality and can be imported and used in other scripts. However, it is also possible to use the script independently, e.g., to have a first look at an recorded spectrum. Then several positional and optional arguments have to be specified in the command line:

- `file_path`: The path to the file containing the spectrum (positional).
- `survey`: The survey the spectrum originates from (positional).
- `--wmin`: Lower wavelength limit in Å (optional, default: 3,000).
- `--wmax`: Upper wavelength limit in Å (optional, default: 9,000).
- `--noclean`: Disable cosmic ray removal (optional, default: enabled).
- `--show`: Show the plot of the spectrum (optional, default: disabled).
- `--save`: Save the plot of the spectrum (optional, default: disabled).

The `survey` argument is of particular importance, because it specifies how the data in the spectrum file is read. Usually, the spectra coming from spectroscopic surveys are stored in FITS files and every survey provides them with a uniform structure and observational details in the header. Some of these information are essential, for example, in order to calculate barycentric corrections. Hence, the `Observation` class contains a set of reader functions for many popular spectroscopic surveys and the `survey` argument specifies, which reader function is called for the respective spectrum file. A list of currently implemented surveys can be printed out with the help command (`-h` or `--help`). As example, the command for running the script could look like this:

```
python3 read_observation.py my_sdss_spectrum.fits SDSS --wmax=7000 --show
```

With that, the SDSS spectrum `my_sdss_spectrum.fits` would be read in, cropped to 3000 – 7000 Å, and a plot of the spectrum is shown, but not saved.

If a spectrum comes from a survey that is not implemented in the script or the spectrum file is given in a different format, it is also possible to provide it as ASCII file with two columns (wavelength and flux). In this case, the `survey` argument must be set to `ASCII` and auxiliary data can be entered in the `get_ascii_info` function at the top of the script (object name, coordinates, real survey, resolution parameters, and barycentric correction). Some of these information, like the spectral resolution, might be crucial for further scripts, for example the spectral fitting routine (Subsection 3.3.3), and are therefore important to provide. Furthermore, if certain parts of the spectrum should be excluded for further analysis steps like spectral fits, for instance because of calibration issues in that region or if certain spectral lines should be clipped, a list of wavelength tuples can be defined in the `get_bad_intervals` function, which are then cut out from the spectrum. This function is also found at the top of the script.

In the following, the different steps of the script are described in more detail. First, the respective reader function is called. It reads the header information, calculates Galactic coordinates from the given equatorial coordinates and extracts the observed wavelength and flux values. If necessary, the wavelength grid given in vacuum is transformed to the corresponding wavelength grid in air. For this transformation the equation from Morton (2000) is used. Moreover, the spectral resolution parameters are computed or defined for the respective spectrograph of the survey and a barycentric correction is applied if needed. In a next step, the uncertainty on each flux pixel is determined. For this purpose a function is used that was provided by Matti Dorsch and is based on the technique

from Stoehr et al. (2008). Then, the whole spectrum is cropped to the range defined by `wmin` and `wmax` and all unwanted intervals given in the `get_bad_intervals` function are clipped. Next, the spectrum is checked for unphysical negative flux values and substitutes those with the median of the 10 Å neighborhood (in each direction). Also, cosmics[8] or unusual high noise peaks are identified and set to the median of the neighborhood of that pixel as well, if not disabled by the user with the `noclean` argument. The criterion for the detection of a cosmic is if a flux value exceeds the 3σ range of the direct neighborhood. Note that the script was mainly developed for stellar absorption line spectra. Hence, if dealing with an emission line spectrum, the cosmic ray removal must be disabled. Otherwise, emission lines might be identified as cosmics and removed from the spectrum. Finally, after the spectrum is fully processed and cleaned, the SNR is estimated. For this, the flux over flux uncertainty ratio is calculated for each pixel of the spectrum and the median of all these ratios is taken as average value for the SNR. Here, it should be kept in mind that this really only gives a rough estimate of the overall SNR of the spectrum. Typically it is not constant, but varies with wavelength across a spectrum. In the optical, it is for example often much higher at the red end than at the blue end.

3.3.2 `interpolate_in_grid.py`

The script `interpolate_in_grid.py` contains the class called `GridInterpolator`. It reads the selected grid of synthetic spectra (e.g., one of those described in Section 3.1) and linearly interpolates between the grid points to generate a model spectrum for a given set of atmospheric parameters. Usually, the chosen type of interpolation should depend on the spacing between the grid points. While polynomial or spline interpolation is more precise, it is also much slower than linear interpolation. For close meshed grids as those fore the hot subdwarf stars, however, linear interpolation is sufficient. Again, the functionality of the `GridInterpolator` class can be imported and implemented in other scripts as well as used independently. For example, it is one of the core ingredients in the spectral fitting routine described in Subsection 3.3.3. If it is used independently, the following positional and optional arguments can be specified:

- `grid`: The model grid to be interpolated in (positional).
- `--norm`: Generate a normalized spectrum (optional, default: disabled).
- `--show`: Show the plot of the spectrum (optional, default: disabled).
- `--save`: Save the plot of the spectrum and a table containing the wavelength and flux values (optional, default: disabled).

Here, the `grid` argument defines the used model grid and again, a list of currently implemented grids is available when calling the help command (`-h` or `-help`). In principle, all grids can be used that have the same structure as those, which are already implemented and just the path to the grid needs to be added in the script. If the `norm` argument is given, the generated model spectrum will be normalized. Additional to the arguments in the command line, the atmospheric parameters (temperature, surface gravity, metallicity, helium abundance, and microturbulence) need to be adjusted in the `get_input_values` function at the top of the script. Then, the interpolated spectrum will be the model spectrum for exactly those atmospheric parameters. More details about the interpolation routine can be found in Section B.1.

3.3.3 `spectral_fit.py`

The script `spectral_fit.py` with the class `SpectralFit` fits model spectra to recorded ones in order to determine the atmospheric parameters of the observed star. To this end, it needs both classes of the previous scripts, the `Observation` class to read in the observed spectrum and

[8] If cosmic high energetic charged particles ("cosmics") hit the detector during exposure, they cause a feature similar no an emission line in the spectrum.

to derive instrumental and observational information like the spectral resolution as well as the `GridInterpolator` class to interpolate between the grid points. Furthermore, it uses functions of the two side scripts `plot_results.py` and `generate_results.py` to generate proper output that is suited for further processing. The positional and optional arguments are just a mix of those for the previous scripts and thus, do not need further explanation:

- `file_path`: The path to the file containing the spectrum (positional).
- `survey`: The survey the spectrum comes from (positional).
- `grid`: The model grid to be fitted to the observation (positional).
- `--wmin`: Lower wavelength limit in Å (optional, default: 3600).
- `--wmax`: Upper wavelength limit in Å (optional, default: 7200).
- `--noclean`: Disable cosmic ray removal (optional, default: enabled).
- `--show`: Show the plot of the spectral fit (optional, default: disabled).
- `--save`: Save the outcomes of the spectral fit (optional, default: disabled).

Note that the default wavelength range changed to properly cover the Balmer jump in the blue regime of the spectrum without taking too much of the often noisy parts at shorter wavelengths into account and to cut the spectrum at the red end of the optical band. These two wavelength cuts are actually very important. On the one hand, recent investigations in our working group have shown that spectral fits, which do not cover the whole Balmer series of the photospheric absorption line spectra, might lead to systematic uncertainties in the atmospheric parameters (especially the surface gravity). On the other hand, hot subdwarfs do not contribute with a lot of flux in the near infrared such that the Paschen series (additionally weakened by pressure ionization of hydrogen) is often swallowed by noise and blended with telluric lines[9]. Of course, if the spectral coverage of a certain spectrum does not range down to the Balmer jump or up to the red end of the optical, the upper and lower wavelength limits are given by the edges of the observed spectrum. Additional to the `get_bad_intervals` and `get_ascii_info` functions (see Subsection 3.3.1), there are two more functions at the top of the script that can be adjusted by the user to change the fit settings: `setup_fitting_parameters` and `setup_cspline_anchorpoints`. In the first one, the initial values for the fitting parameters can be set up as well as the setting, which parameters should be varied and which one stay frozen during the fitting procedure. In the latter one, the anchor points for the spline (continuum) fitting can be adjusted, which are important for the normalization of the spectrum. The normalization procedure is explained in more detail in Section B.2 and below. Overall, a command for running the fit of a sdB star model to an observed SDSS spectrum could look like this:

```
python3 spectral_fit.py my_sdss_spectrum.fits SDSS sdB --save
```

As output (if requested), the program generates a multi-page PDF showing a plot of the spectral fit (Figure 3.2), a CSV table containing the fit results, and a PDF summarizing the fit results (Figure 3.3). The plot consists of three panels, the upper panel showing the original spectrum together with the continuum fit, the mid panel showing the normalized spectrum together with the spectral fit, and the lower panel showing the residuals of the fit.

Since the `SpectralFit` class is one of the core methods used in this work, it is explained step by step and in more detail than the previous scripts in the following. After the observed spectrum is read in and processed with the `Observation` class, it is passed to `SpectralFit` together with the positional and optional arguments as well as the parameter settings and the spline anchor points. Then everything is implemented in a LMFIT framework as described in

[9] Telluric lines are caused by atoms and molecules in the Earth' atmosphere and are predominantly located at longer wavelengths.

Section 3.2, consisting of the main function `fit_spectrum` of the class, the `get_minimizer` function setting up a `Minimizer` object for the fitting routine, and the actual fit function `fit_func`. While the main and the fit function are discussed below, the setting up of the `Minimizer` object by `get_minimizer` is outlined in Section B.2.

The Fit Function `fit_func`

The fit function `fit_func` includes all the relevant steps for the fitting process and is later on called once and once again, until the residuals of the fit are minimized and the best matching model has been found. First, the `Parameters` object included in the `Minimizer` is read in and the x- and y-values of the anchor points for the normalization are passed to a spline interpolation routine. In principle, multiple such routines are implemented in the script, but as default the `InterpolatedUnivariateSpline`[10] object from the scipy library `scipy.interpolate` is used to calculate a second order spline as continuum estimate. Then, after checking that the spline interpolation did not return unphysical negative values (which might happen, if measured flux values are close to zero), the normalization is carried out by dividing the observed spectrum by the continuum estimate. As soon as this is done, the `GridInterpolator` class is initialized with the selected model grid and the atmospheric parameters. The wavelength grid of the synthetic spectra is extracted and a radial velocity shift is applied using the Doppler formula. Next, the actual normalized[11] model spectrum is calculated by carrying out the multidimensional interpolation within the grid (see Subsection 3.3.2). Of course, the resolution of the synthetic spectrum is in general higher than the resolution of the observed spectrum and effects like instrumental broadening and stellar rotation are not considered yet (see Subsection 1.4.3). Therefore, on the one hand the model needs to be resampled to the observation conserving the strength and shape of spectral lines, and on the other hand a convolution of the model with the instrumental profile as well as the rotational broadening needs to be executed. To this end, the function `convolve_single_simple` is used, which was provided by Matti Dorsch and is a wrap-around of a function originally written by Natasa Dragovic, Andreas Irrgang, and Simon Kreuzer in C. It is implemented this way, since it is computationally demanding and can be ran much faster in C than in Python. As arguments, it takes the wavelength grid of the model, the synthetic flux values, the resolution parameters (resolution slope and offset, see Subsection 1.4.1), the projected rotational velocity, as well as the wavelength grid of the observation and returns the modified model spectrum. If dealing with very high resolution spectra, it can still happen that the spectral resolution is higher than that of the models. In this case, the resulting synthetic flux values from `convolve_single_simple` are interpolated to the observed wavelength grid. Finally, the residuals are calculated by taking the differences between the normalized flux values of the observation and that of the model, weighted by the normalized flux uncertainties of the recorded spectrum (Equation 3.7). These residuals are then checked for finiteness and returned by the function.

The main function `fit_spectrum`

The main function `fit_spectrum` of the `SpectralFit` class structures and executes the whole fitting process in three stages. Before running the first fit, the `Minimizer` object is created by calling `get_minimizer` (see Section B.2). Then, the first whole fitting run is executed employing the commonly used Levenberg-Marquardt least-squares (”leastsq“ in LMFIT, see Section 3.2) minimization algorithm. This means, the fit function `fit_func` is successively called again with varying fit parameters until the best matching model has been found, where the residuals are minimal. The fit results and statistics are printed out in the terminal and collected in compact form by the function `get_fit_params`. In the second stage, the flux uncertainties, which play a

[10] https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.InterpolatedUnivariateSpline.html

[11] It is important that the normalized version of the synthetic spectra is taken, because otherwise the shape of the continuum of the observed spectrum would contribute to the atmospheric parameter determination, mainly to that of the effective temperature. However, observed spectra are often not absolute flux calibrated, which means the shape of their continuum is not properly reflecting the energy output of the star.

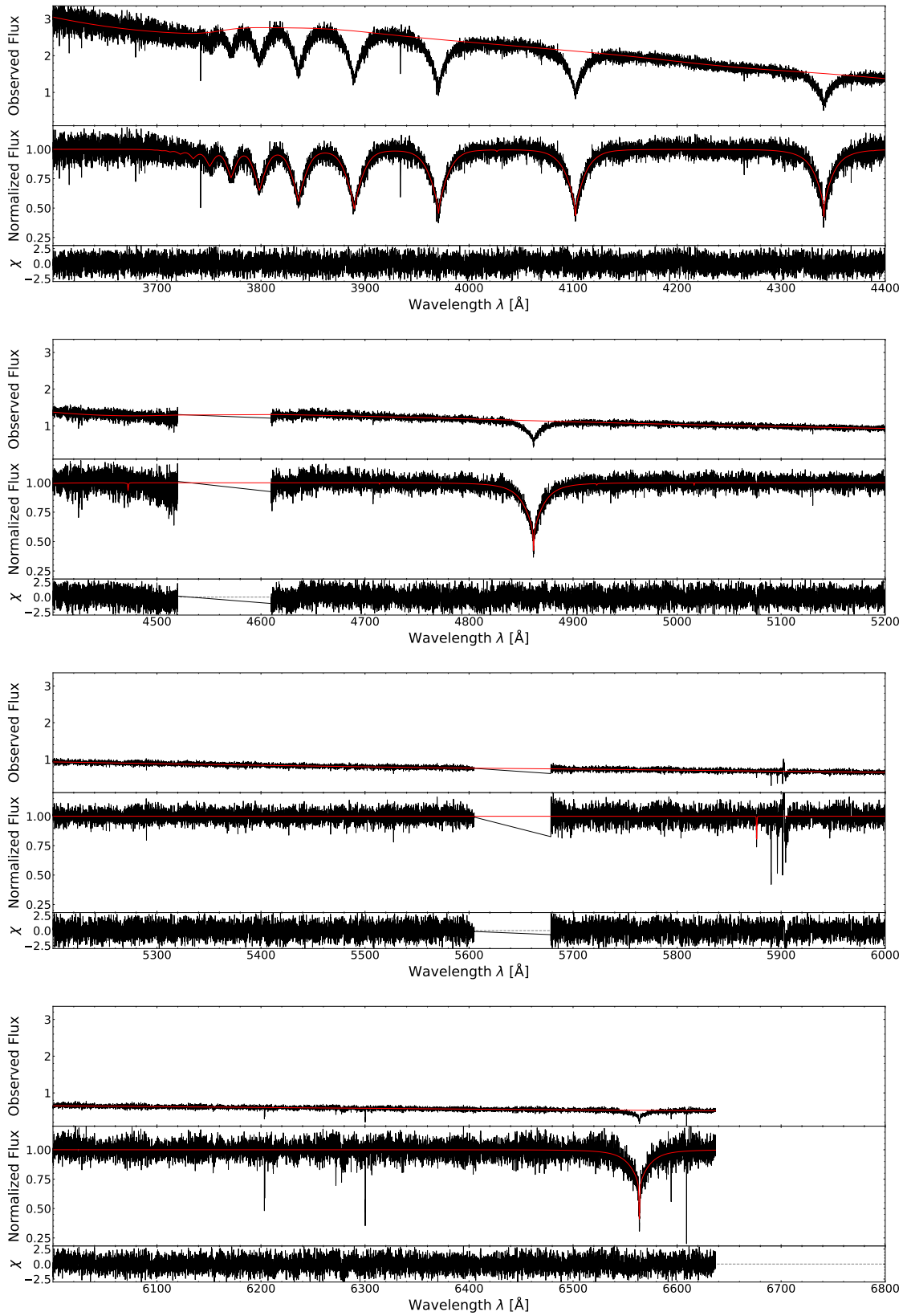


Figure 3.2: Example multi-page plot of a spectral fit showing the original spectrum plus continuum estimate in the upper panel, the normalized spectrum with spectral fit in the mid panel, and the residuals of the fit in the lower panel. There are two gaps present in spectrum, because the spectrum comes from the SPY survey recorded with UVES (Subsection 2.2.5) and those spectra were recorded on three CCD chips. The synthetic spectra hold hydrogen and helium lines only. The sharp absorption lines, which are both not modeled, are from metals in the atmosphere or from interstellar material.

Parameter	1σ confidence interval
Information	
Object name	he0539-4246
Right ascension	85.27796°
Declination	-42.75886°
Galactic longitude l	248.64021°
Galactic latitude b	-30.46122°
Survey/spectrograph	SPY
Estimated SNR	22.2
Effective SNR	21.9
Spectral fit	
Model grid	sdB
Resolution offset $R = \lambda/\Delta\lambda$ (fixed)	20000
Resolution slope $1/\Delta\lambda$ (fixed)	0.000 Å^{-1}
Projected rotational velocity $v\sin(i)$	3 ± 1 $\text{km}\,\text{s}^{-1}$
Effective temperature T_{eff}	23381 ± 79 K
Surface gravity $\log(g\,(\text{cm}\,\text{s}^{-2}))$	5.615 ± 0.009
Microturbulence ξ (fixed)	0 $\text{km}\,\text{s}^{-1}$
Metallicity z (fixed)	0.00 dex
He abundance $\log(n(\text{He}))$	-3.784 ± 0.063
Radial velocity v_{rad}	41.8 ± 1.1 $\text{km}\,\text{s}^{-1}$
$\chi^2_{\text{red},0}$	1.132
$\chi^2_{\text{red,initial}}$	1.021
$\chi^2_{\text{red,final}}$	1.000
Noise scale factor	1.011

Figure 3.3: Summary PDF of the spectral fit corresponding to Figure 3.2. Information about the recorded object and spectrum are given as well as the determined atmospheric parameters and fit statistics (the reduced chi squares of the three fitting stages and the noise scale factor as explained below). It is also noted, which fit parameters were varied during the fit and which ones were frozen (fixed).

crucial role in the fitting process as weighting factors for the residuals (Equation 3.7), are locally increased in regions of uncertain fluxes. It is checked, where the residuals of the first fit are larger than three (3σ clipping) and the corresponding flux uncertainties are increased by a factor of five. In general, this can be the case for artifacts or other irregularities in the observed spectrum, but also for real features that are not included in the model spectrum like interstellar spectral lines or photospheric metal lines. By rescaling the uncertainties of such flux pixels, they have a lower weight such that they become less relevant than the well behaving ones for the fitting process and the atmospheric parameter determination. After rescaling, another full fitting run is carried out using the fit parameters from the first run as initial parameters as well as the updated flux uncertainties. In the beginning of the third stage, the flux uncertainties are rescaled globally with the so-called noise scale factor, which equals the square root of the reduced chi square (see Section 3.2) from the previous fitting run. All flux uncertainty values are multiplied with this factor such that the next run results in a proper fit with a reduced chi square of unity. Technically, this step compensates the effect of initially over- or underestimated uncertainties on the recorded data as described in Section 3.2. Afterwards, these updated flux uncertainties are used to refit the observed spectrum three times, always with the fit parameters from the previous run as initial values, respectively. As last step, the effective SNR is calculated analogously to Subsection 3.3.1, but using the rescaled flux uncertainties.

Overall, it can be said that this fitting routine works robustly and efficiently to derive decent atmospheric parameters for various stellar spectra. However, since the Levenberg-Marquardt algorithm is used for all fitting runs, one could think about varying and switching the algorithm during the fitting process. Several other methods, for instance the Powell or Nelder-Mead algorithms (see Section 3.2) would be possible. In fact, this was tested when developing the script, but it was realized that other algorithms on average take much longer than the very efficient

Levenberg-Marquardt one. Also, the resulting mean values for the atmospheric parameters did in most of the cases not change anymore after multiple fitting runs, even when switching the fit method. Moreover, the derived statistical uncertainties on the fit parameters were sometimes unrealistically low with other algorithms. Therefore, primarily for the sake of efficiency, it was entirely stick to the Levenberg-Marquardt method in this script.

3.3.4 `ism_analysis.py`

A central part of this work is to analyze interstellar absorption lines, which are often present in the spectra of stars. The script `ism_analysis.py` with the class `ISM`, is an extension of the spectral fitting routine (Subsection 3.3.3) that also investigates and fits these interstellar absorption lines, which are, of course, not part of the model atmospheres. Thus, it takes exactly the same positional and optional arguments as `spectral_fit.py` and the same functions can be found at the top of the script to adjust the settings. In addition, the optional argument `csm` can be given. Then the radial velocity of the ISM lines is fixed to that of the star with a small buffer to actively search for circumstellar material (CSM, see Subsection 1.3.2 and Chapter 6). The fit parameters for the interstellar lines (column densities, Doppler parameters, and radial velocity, see Subsection 1.4.4) can be defined in the `setup_ism_fitting_parameters` function also at the top of the script. Unlike for the spectral fitting script, if here the `show` argument is given, not only the final result plots are shown, but also intermediate results. Since the whole analysis is highly sensitive to the quality of the recorded spectrum, this can be helpful to find the error sources, if something does not behave as expected. The command to run the script could, for example, look like this:

```
python3 ism_analysis.py my_sdss_spectrum.fits SDSS sdB --save --show
```

Then, first the spectral fit of the sdB model grid to a recorded SDSS spectrum is carried out according to Subsection 3.3.3, the ISM flux is calculated by taking the synthetic spectrum as continuum, the interstellar line regions are extracted and re-normalized, all ISM lines are analyzed individually, a global radial velocity for these lines is determined, and finally, a combined Voigt profile fit to the lines is run to find column densities and Doppler parameters from their line shapes. Since almost all results characterizing the interstellar medium in this work come from this program, all the individual steps are described in detail below. As output (if requested), several results plots are generated (Figure 3.6, Figure 3.8, and Figure 3.9) and the summary PDF of Figure 3.3 as well as the CSV table are extended with the results from the ISM analysis (Figure 3.10).

ISM Flux Calculation and Re-Normalization

The first step is to calculate the ISM flux and extracting as well as re-normalizing the regions around the interstellar absorption lines, which is all carried out by the function `get_ism_flux`. The four ISM lines that can be typically found in the optical are the Ca II H&K lines and the Na I D1&D2 lines as described in Subsection 1.2.2. To properly analyze them, they need to be clearly distinguished from photospheric lines. One advantage of using the spectra of hot subdwarf stars is that they are in most cases quite simple, only showing hydrogen and helium lines. Hence, this distinction is often not a big issue. However, one of the lines, the Ca II H line at ~3968 Å, is usually heavily blended by the prominent Hε Balmer line at ~3970 Å. To solve this issue, the synthetic spectrum from the spectral fit is taken as ISM continuum and removed from the spectrum by dividing the normalized observation by it. This procedure is visualized in Figure 3.4. The same is done for the flux uncertainties and in both cases all resulting pixels are checked for finiteness. As next step, the 10 Å regions around the rest wavelengths of each ISM line are extracted. It is also checked if the respective regions are covered by the given spectrum. If not, these regions are no longer considered for the further course of the analysis. Now, even though the flux already got normalized during the spectral fit, the normalization procedure there might sometimes not account for local offsets due to the spacing of the anchor points. Therefore, simple Gaussian fits with linear background are carried out to each line region in order to re-normalize them using the modeled background. The only constraints on the Gaussians here are that they should represent

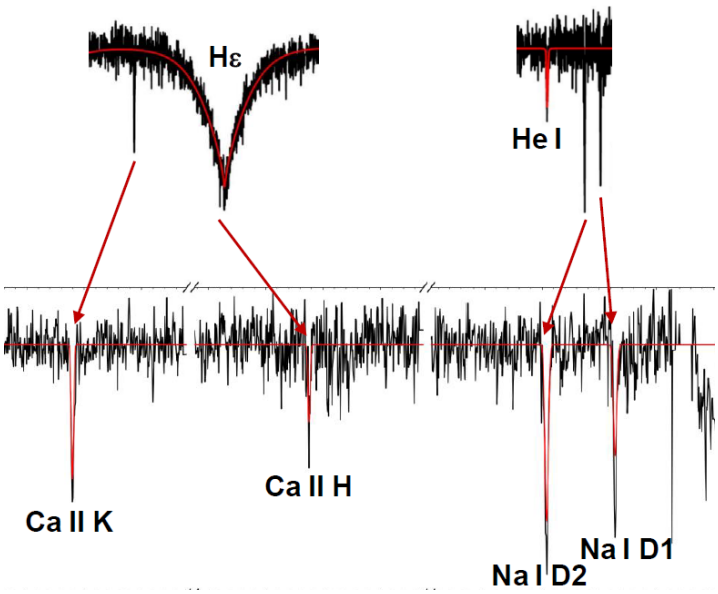


Figure 3.4: The principle of the ISM flux calculation. The synthetic spectrum shown in red in the upper half of the figure is taken as ISM continuum to divide the observation (black) by it. This results into nicely isolated interstellar absorption lines as shown in the lower half. In particular, it solves the heavy blend of the Ca II H line (the tiny line at the tip of Hε) with the Hε Balmer line. Also shown in red is a fit to the four ISM lines, which will be discussed later on as well as why the line cores are sometimes not properly modeled.

absorption lines rather than emission lines and their width σ is fixed to a reasonable range between the spacing of the wavelength pixels as lower limit and 2.5 Å as upper limit. While in some cases, especially if the interstellar lines are weak, it can be difficult for the program to automatically identify the lines as explained in the next paragraph, here this problem is actually not important. The only relevant issue is the linear background, which is properly modeled even if the Gaussian fit was applied to some prominent noise peak instead of a real line. Hence, these constraints on the fit are sufficient and for simplicity the `scipy.optimize.curve_fit`[12] function from the Scipy library is used to carry out the fits instead of setting up a whole LMFIT framework. For each Ca II line a single Gaussian is fitted, while for the two Na I lines a double Gaussian is used because of their proximity. Then finally, each line region is divided by its linear background in order to re-normalize it. Next, "cosmics" (or rather unusual high noise peaks or artifacts) are removed similarly to Subsection 3.3.1. Here, the SNR of each region is estimated and all flux values above the 3σ noise level are set to unity. This can be important, because strong noise peaks or artifacts next to a line might disturb its equivalent width measurement as described in the next section. In the end, if the `show` argument is given, two plots of the outcome of this function are shown as intermediate results (Figure 3.5).

Individual ISM Line Analysis - Constrained Gaussian Fits

Now, after the interstellar line regions are extracted and processed, the aim is to measure as many parameters of these lines and properties of the material they originate from as possible. However, to carry out all measurements in one step, a lot of information needs to be contained in each spectral line. Unfortunately, there are several factors that limit the amount of spectral information, mainly the spectral resolution and noise. The higher the resolution and SNR, the easier it is for the program to derive the parameters describing the line. But the lower both quantities are, the less reliable the measurements become. Since it is dealt with various kinds of spectra of different quality in this work, which are mostly of low resolution (see Section 2.3), the needed amount of information is often not given. Therefore, the measurements were divided into several steps to make them more robust, beginning with an individual analysis of each line carried out by the function `run_single_line_analysis`. Its aim is to reliably identify each line or rather check whether a line is present at all, fit its Gaussian properties (amplitude, FWHM or $\Delta\lambda$, and radial velocity), and to compute its equivalent width. To this end, Gaussian fits are set up using the LMFIT framework with a proper fit function and a function to create the `Minimizer` object. But as described further below, due to the limited amount of spectral information even in higher resolution

[12] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html

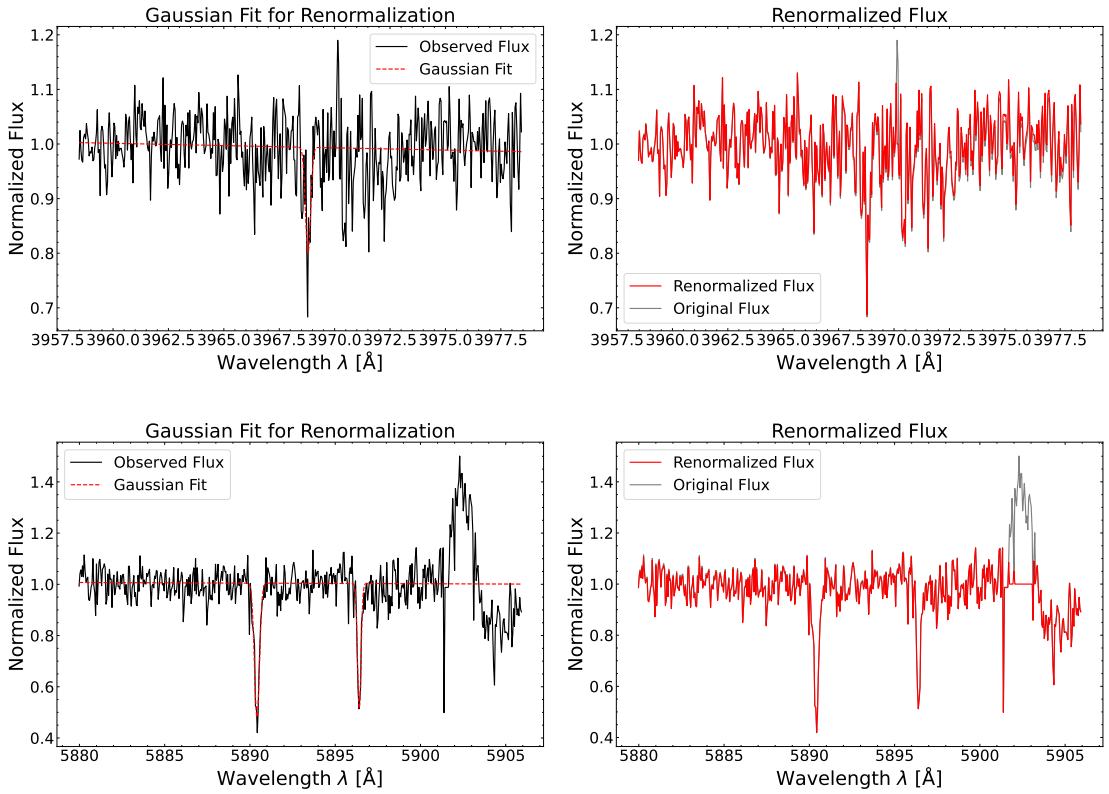


Figure 3.5: *Upper left panel:* Single Gaussian fit with linear background to the Ca II H line. A small slope is visible in the background. *Upper right panel:* Comparison between the re-normalized flux and the original flux of the Ca II H region. A small shift is visible on the right. Also a noise peak in the middle got removed. *Lower left panel:* Double Gaussian fit with linear background to the Na I D1&D2 lines. *Lower right panel:* Comparison between the re-normalized flux and the original flux of the Na I D1&D2 region. A strong artifact often visible at this position in SPY spectra was excluded at the right.

spectra, it is here not possible to include the convolution with the instrumental profile in the fit. Hence, prior knowledge is needed to find proper initial values and constraints for the fit parameters in dependency on the spectral resolution. In general, the simple Gaussian absorption line profile

$$f(\lambda) = 1 - A \cdot \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma^2}\right) \tag{3.8}$$

is fitted to the normalized lines, where A is the amplitude (depth), λ_0 the position of the line center, and σ the width of the lines. However, both the amplitude and the width are just the apparent properties of the lines after getting smeared out by the instrumental broadening. Thus, initial values as well as fit limits are calculated based on the observed spectrum.

To estimate the amplitude, it is started with one minus the lowest flux value ($1 - f_{\min}$) within a region of 3 Å around the line's rest wavelengths. However, this value is never allowed to be smaller than half the noise level, i.e. the amplitude should exceed average noise peaks. For instance, in the upper left panel of Figure 3.5 the minimum flux is about 0.7, which gives an initial amplitude of 0.3 (a 30% decrease in flux). The allowed range for the amplitude is set to $\pm$20% of this initial value, but it is always kept between half the noise level (as a lower limit) and 1.0 (a full 100% flux decrease) as an upper limit.

More complicated is the estimation of the width of the lines. In principle, every line has an intrinsic width mainly depending on the temperature and the density of the material (see Subsection 1.4.3). Measurements of the Doppler parameter b of interstellar Ca II H&K and Na I D1&D2 lines (see Subsection 1.4.4) based on very high resolution spectroscopy have shown that

their values are usually in the range of $0.5 - 3.5\,\mathrm{km\,s^{-1}}$ (Welty et al., 1994; Redfield & Linsky, 2002; Welsh et al., 2010). However, these results are based on observations of very close-by objects (a few hundred pc) and even there a tail towards $b \sim 7\,\mathrm{km\,s^{-1}}$ is found that is explained by multiple unresolved ISM components. In this work, most of the stars are much more distant up to several kpc. Therefore, high b values can be expected. Converting these widths given as Doppler parameters b to $\Delta\lambda$ results in an intrinsic width of about $0.1 - 0.2\,\text{Å}$ for the interstellar lines and a value of 0.15 Å is chosen here. Now, the lines are additionally broadened by instrumental broadening, which depends on the spectral resolution. The spectral resolving power is given by $R(\lambda) = \lambda/\Delta\lambda$ and hence, the minimal width of a spectral line due to instrumental broadening is $\Delta\lambda = \lambda/R(\lambda)$. Both widths add up quadratically to the observed total width of a line:

$$\Delta\lambda_{\mathrm{total}} = \sqrt{\Delta\lambda^2_{\mathrm{instrumental}} + \Delta\lambda^2_{\mathrm{intrinsic}}} \tag{3.9}$$

So $\Delta\lambda_{\mathrm{instrumental}}$ is calculated for each line from its rest wavelength and the spectral resolution of the spectrum before the fit is carried out. Then $\Delta\lambda_{\mathrm{total}}$, which corresponds to the full width at half maximum (FWHM) of a line, is taken, converted to a Gaussian width according to $\Delta\lambda_{\mathrm{total}} = 2\sqrt{2\ln 2}\cdot\sigma$, and used as initial value for the fit. As lower limit the pure $\Delta\lambda_{\mathrm{instrumental}}$ is used and as upper limit the $\Delta\lambda_{\mathrm{total}}$ corresponding to $3\cdot\Delta\lambda_{\mathrm{intrinsic}}$. Here it can be seen that for low resolution spectra ($R \sim 2{,}000$), the $\Delta\lambda_{\mathrm{instrumental}}$ is on the order of ~2.5 Å, which means the line shape is completely dominated by instrumental broadening (see, for example, Figure 3.11).

Fitting the radial velocity (shifted position λ_0) of the lines is actually also not trivial. While the velocities of the ISM are usually small ($|v_{\mathrm{rad,ISM}}| \lesssim 30\,\mathrm{km\,s^{-1}}$), they can in extreme cases range up to $|v_{\mathrm{rad,ISM}}| \gtrsim 100\,\mathrm{km\,s^{-1}}$, for example for so-called high-velocity clouds (Wakker & van Woerden, 1997). Thus, the fit limits should not be chosen smaller than $\pm 200\,\mathrm{km\,s^{-1}}$. But when testing the fitting procedure, some difficulties appeared, in particular for noisy spectra with weak ISM lines. If the usual Levenberg-Marquardt algorithm is used, it can happen that it gets stuck in local minima before finding the right line. Therefore, other fitting algorithms were tested that are more suited to find the global minimum. In the end, the "basinhopping" method (see Subsection 3.2.3) has proven to be the best choice even though it is more computationally demanding. On the other hand, when using this algorithm, the real lines can get confused with other features at high radial velocities, since right now higher radial velocities are considered as likely in the fit than lower ones within the allowed range. So the idea came up to somehow manipulate the fit utilizing the prior knowledge on the radial velocity distribution of the ISM. This is realized by adding the artificial penalty term

$$\left(\frac{v_{\mathrm{rad}} - v_{\mathrm{rad,prior}}}{\Delta v_{\mathrm{rad,prior}}}\right)^2 \tag{3.10}$$

to the residuals of the fit function, where v_{rad} is the fitted radial velocity, $v_{\mathrm{rad,prior}}$ the initial value of the radial velocity, i.e. the value given by the user in the `setup_ism_fitting_parameters` function (usually $v_{\mathrm{rad,prior}} = 0\,\mathrm{km\,s^{-1}}$), and $\Delta v_{\mathrm{rad,prior}}$ the 1σ dispersion of $v_{\mathrm{rad,prior}}$. For $\Delta v_{\mathrm{rad,prior}}$ a value of $60\,\mathrm{km\,s^{-1}}$ was chosen in consistency with the allowed radial velocity range and with tests of the script on spectra with high radial velocity lines. However, it is just an arbitrary value that seemed to properly work and does not reflect any physical velocity dispersion of the ISM. Now, if the fit wants a high radial velocity, the penalty term artificially increases the reduced χ^2 such that it tends to find a lower radial velocity, i.e. another line position and hence a different line, in the minimization process, unless the high radial velocity is indeed by far the best match. Due to the high dispersion value, the fit should also not get biased too much. It just ensures that the program prefers lines at low radial velocities, which are more likely real, instead of noise peaks at high radial velocities in the case of weak ISM lines or very noisy spectra.

As last refinement of the fit, the model (Equation 3.8) is defined on a wavelength grid corresponding to a spectral resolving power of $R = 50{,}000$ and resampled to the wavelength grid of the observation. Since the convolution of the model with the instrumental profile is not possible, this step is executed to conserve at least some information that gets lost due to instrumental broadening, especially when working with low resolution spectra. For the calculation of the model wavelength

grid a function provided by Matti Dorsch is used that gives a properly sampled grid for a specific resolution. It uses a slightly higher sampling of 2.7 instead of 2 suggested by the Nyquist criterion, which is in general not sufficient (Robertson, 2017). The resampling is carried out using a Fortran function implemented in Python and provided by Matti Dorsch as well, which is based on Carnall (2017).

Finally, taking all these adjustments into account, the Gaussian fits are carried out on each ISM line. One disadvantage of the ”basinhopping“ algorithm is that it often can not properly estimate the uncertainties on the fit parameters. Therefore, the results are used to refit the lines with the Markov Chain Monte Carlo (MCMC) method ”emcee“ (see Subsection 3.2.3, also for the needed settings of this method), which is capable of properly estimating the parameter space, if the mean values of the fit parameters are already decently determined. This whole procedure shows how complex even simple Gaussian fits can become. In the end, reliable amplitudes, FWHMs, and radial velocities are obtained for each of the ISM lines, which are also needed for the subsequent equivalent width measurements.

Individual ISM Line Analysis - Equivalent Width Measurements

The next step of the `run_single_line_analysis` function is to measure the equivalent width of each line and its uncertainty, which is another valuable parameter describing the strength of a spectral line. In general, it represents the width of a rectangle with a height equal to the continuum that has the same area as the spectral line itself and is conserved under instrumental broadening. When dealing with normalized spectra as here, the equivalent width is in principle directly given by the intensity (covered area) of the Gaussian fit. However, the line shape needs to be well reproduced by that Gaussian to estimate the equivalent width properly, which is often hindered by partly resolved multiple ISM components (see the right panel of Figure 3.6 or Figure 3.7), in particular for high resolution spectra. A second possibility to derive equivalent widths is to integrate over a spectral line. But then, the question is the chosen integration range, especially since the apparent width of the lines varies with spectral resolution as depicted in Figure 3.11. A good choice seems to be the 5σ Gaussian width as measured before. Then, the integration range is broad enough to account for all close-by ISM components and the influence of noise at the edges of the lines or the surrounding continuum should be averaged out. Only for very low resolution spectra the proximity of the Na I D lines can become a problem, because the integration ranges might extend from one line to the other, which would lead to a heavy overestimation of the equivalent widths. To avoid this issue, the integration range for the Na I D1&D2 lines is capped to 3 Å around the line center, respectively. The actual integration is executed using the function `simpson`[13] from the scipy.integrate library, which applies a numerical integration according to the Simpson's rule to one minus the observed flux. To measure the equivalent width as precisely as possible a Monte Carlo run with 10,000 samples is carried out considering the uncertainties on the flux values. As mean value of the equivalent width the median of these samples is taken and as uncertainty the standard deviation[14]. Of course, if the integration is applied on multiple ISM components, the result does not represent the strength of a single spectral line as usual. The value will be inflated, which can also be seen in Figure 3.6. Still, it is a good proxy for the total amount of intervening interstellar material. Important is that there are no pronounced noise peaks or artifacts next to a line, which is why at least those in upward direction are removed by the `get_ism_flux` function. If, for example, an artifact like in the lower panels of Figure 3.5 is within the integration range of the equivalent width, the resulting value can easily become negative.

Finally, the quality of the analysis has to be evaluated. Also, not all stellar spectra will show interstellar absorption lines. If a star is close-by or located in a direction with little intervening interstellar material in the line of sight, ISM lines can be very weak or not present at all. Or Ca II lines are visible and Na I ones are not (or vice versa). Therefore, criteria have to be found to

[13] https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.simpson.html

[14] In the future, the confidence interval might be replaced by more sophisticated quantiles (see Appendix A) instead of the standard deviation.

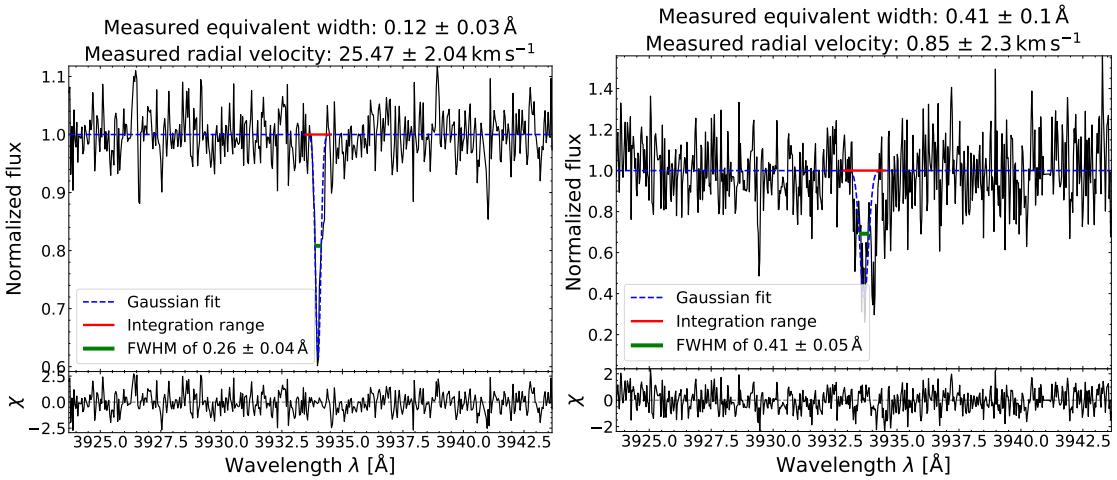


Figure 3.6: *Left panel:* Example results plot of the `run_single_line_analysis` function. Depicted is the observed flux in black, a Gaussian fit to the ISM line in blue, the integration range for the equivalent width in red at the height of the continuum, the apparent FWHM in green at half the height of the line, and the residuals of the fit. Also given are the measured values of the equivalent width, radial velocity, and FWHM. *Right panel:* Same plot as in the left panel, but for a spectrum, where at least one additional ISM component is clearly visible. The full set of lines is shown in Figure 3.7. Note the different scale of the *y*-axis and the considerably higher equivalent width compared to the left plot.

check if the detected and analyzed lines are real. When inspecting spectra visually (”by eye“), we recognize spectral absorption lines if the flux drops below the surrounding continuum and/or if a feature is broader than average noise peaks. Hence, suitable criteria seem to be that the amplitude has to clearly exceed the noise level or the equivalent width has to exceed a certain threshold. When testing the function, a threshold of 0.04 Å was established, in consistency with a minimum measurable equivalent width of ~0.01 Å in decent quality high resolution spectra. Only if at least one of the both criteria is fulfilled, the detected ISM line is considered as real and all measured quantities are returned by the function. Lastly, depending on the choice of the `show` and `save` arguments, results plots like in Figure 3.6 are shown and/or saved for each ISM line region.

Global Radial Velocity Determination

When there are multiple interstellar lines present in a spectrum, they are expected to originate from the same interstellar cloud and hence, should show the same radial velocity. While previously the radial velocities have been determined individually for each ISM line, now the aim is to find one uniform global ISM radial velocity, just like for the stellar radial velocity determined from shifting the whole spectrum of photospheric lines. To this end, a LMFIT framework is set up, which tries to derive this global radial velocity from the results of the single line analyses. A Gaussian model is defined, which takes the parameters (amplitudes and FWHMs) from before as fixed parameters and the radial velocity as only free fit parameter. As initial value the mean of the individually determined radial velocities is taken and as limits the minimum/maximum value of the set of radial velocities from before $\pm 20\,\mathrm{km\,s^{-1}}$, respectively. This choice of fit limits has been established while testing the script, because on the one hand it reasonably constrains the fit and on the other hand it permits enough freedom to account for the possibility that the individual line analyses detected lines from different fully resolved ISM components with different radial velocities. Unfortunately, this can happen in the rare cases of high resolution spectra with two or more fully resolved and almost equally strong components arising from multiple intervening interstellar clouds. If one component is much stronger than the other one, the fitting routine from before will prefer this one and that issue does not appear. By using these settings here, at least the

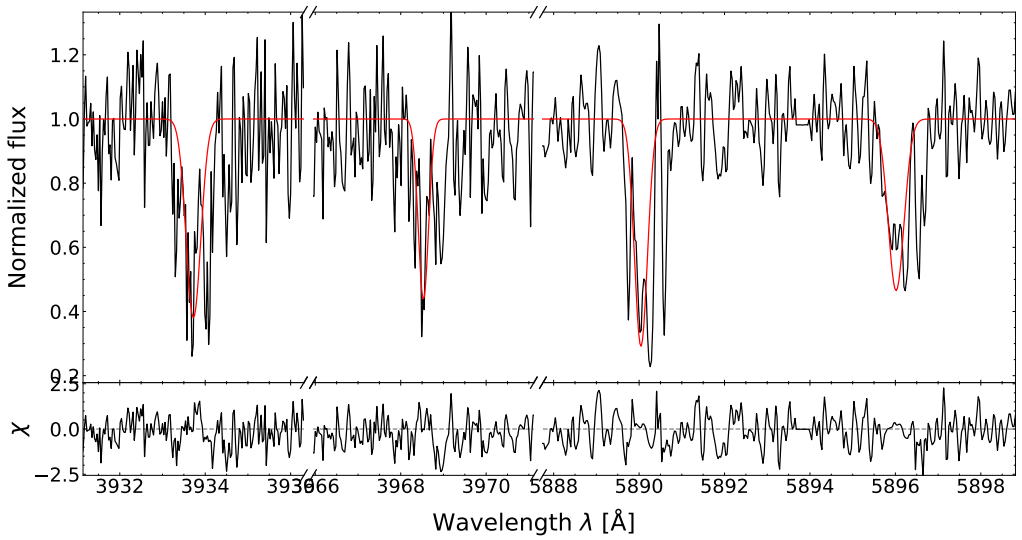


Figure 3.7: Example full set of ISM lines with multiple partly resolved components.

radial velocity of one component, most likely the slightly stronger one, will be reliably determined. In turn, this means that in these rare cases the individual measurements do not necessarily belong to this component. But if this problem occurs, it is clearly visible from the different individual radial velocities of the ISM lines.

However, in most of the cases the global radial velocity determination procedure will allow for an unproblematic and precise measurement of the ISM velocity. In general, the precision is increased by a factor of $2-3$ compared to the individual radial velocities, depending on the spectral resolution, and also the reliability is improved. For instance, in the very low resolution case the radial velocities from just one spectral lines are often accurate to only $30\,\mathrm{km\,s^{-1}}$. By taking all four interstellar lines into account, the uncertainty in most cases decreases to about $10\,\mathrm{km\,s^{-1}}$. In the high resolution case, the uncertainty is on average reduced from about $1\,\mathrm{km\,s^{-1}}$ to about $0.5\,\mathrm{km\,s^{-1}}$. The fits are again carried out using the ”basinhopping“ algorithm for the same reasons as before, followed by another run using the ”emcee“ algorithm to properly estimate the parameter space. Unlike before, no prior on the ISM radial velocity is implemented as it is already well constrained. Depending on the choice of the `show` and `save` arguments, an output as in Figure 3.8 is generated. All four interstellar line regions are shown with exactly the same scale, such that the global radial velocity measurement can be checked. The plot can also serve as easy comparison between the shapes of the different lines.

Combined Voigt Profile Fits

In the previous steps, the interstellar line regions have been extracted and processed, it was checked if ISM lines are present and at which radial velocity, and Gaussian fits have been executed as well as an equivalent measurement. While the first parts are a requirement for further investigation of the lines, the latter ones give first results describing their properties. However, no theory has been put into the analysis yet. The amplitudes and FWHMs are just the apparent ones that highly depend on the spectral resolution of the observed spectrum and equivalent widths are a good proxy for the overall strength of a spectral line, but do not directly give a conclusion about the properties like temperature or density of the interstellar material they originate from. To obtain insights into the nature of the interstellar clouds from the shapes of the ISM lines, now Voigt profile (see Subsection 1.4.3) fits are carried out with column densities N (or $\log N$) and Doppler parameters b as free parameters. As explained in Subsection 1.4.4, the column densities are a measure for the amount of interstellar material along the line of sight, which is mainly responsible for the depth of a line, while the Doppler parameters are a combination of the thermal motion of the gas and its microturbulent velocity, which both broaden a line and are hence mainly responsible for the width.

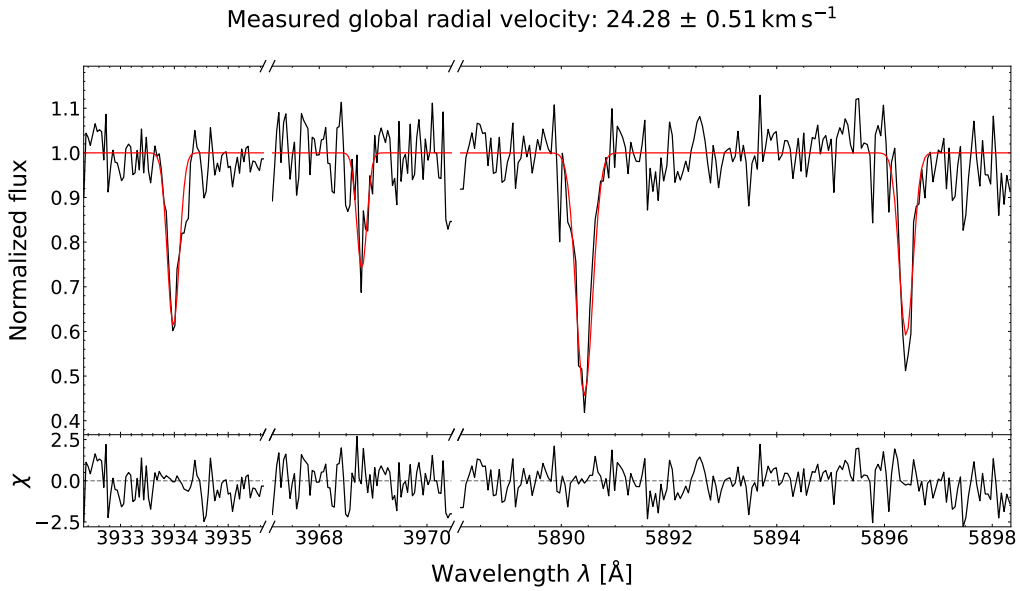


Figure 3.8: Broken-axis plot showing all four observed interstellar lines (black), together with the Gaussian fits from the individual line analysis (red), and the corresponding residuals. The measured global ISM radial velocity is noted as well.

Additional to these two parameters, atomic data like transition probabilities and oscillator strengths (both constants) for the respective lines are needed as input to properly describe them (Table 1.1).

Altogether, another LMFIT framework is set up. First, a `Minimizer` object is created defining the fit parameters, which are the column densities and Doppler parameters for each element as free parameters, as well as the global ISM radial velocity and the resolution parameters (resolution slope and offset, see Subsection 1.4.1 as fixed parameters. The free parameters are constrained to $N = 10^{10} - 10^{15}\,\mathrm{cm}^{-2}$ and $b = 0 - 10\,\mathrm{km\,s}^{-1}$ in consistency with the physically reasonable range for interstellar clouds (Welty et al., 1994; Redfield & Linsky, 2002; Welsh et al., 2010). Then, as first step of the fit function, a wavelength grid is calculated on which the model will be defined. For this, a spectral resolving power of $R = 50,000$ is chosen and the same function as for the Gaussian fits is used. In principle, the higher the chosen resolution the better, but on the other hand with increasing resolution also the number of generated data points increases, which possibly makes the resampling to the observed wavelength grid more time consuming. For the sake of efficiency, this moderately high value is taken, which is still larger than the resolution of every spectrum in the sample of this work. This wavelength grid together with all relevant parameters (rest wavelengths, oscillator strengths, transition amplitudes, column densities, Doppler parameters, and radial velocity) are then used to generate a model including both line doublets. A Voigt profile optical depth τ is calculated for each line according to Equation 1.24, such that the flux transmission visible in a normalized spectrum is given by $I(\lambda) = \mathrm{e}^{-\tau(\lambda)}$. All fluxes (one $I(\lambda)$ for each line) are then multiplied to reproduce the whole observed ISM spectrum with the respective lines. The optical depth calculation follows the absorption theory as described in Subsection 1.4.4 and the Voigt profiles are approximated using the `voigt_profile`[15] function from the scipy.special library. Afterwards, the model is convolved with the instrumental profile of the recorded spectrum and resampled to the observed wavelength grid. Finally, the residuals are computed, checked for finiteness, and returned by the fit function. Since the positions of the lines are already known and the only free fitting parameters are the column densities and Doppler parameters, the actual fit is carried out using the common "leastsq" algorithm. After a first run, the flux uncertainties are rescaled with the noise scale factor and everything is refitted three more times, completely analog to the spectral fitting routine of Subsection 3.3.3. As result, depending on the `show` and `save` arguments, a plot like Figure 3.9 is generated. It is very similar to Figure 3.8,

[15] https://docs.scipy.org/doc/scipy/reference/generated/scipy.special.voigt_profile.html

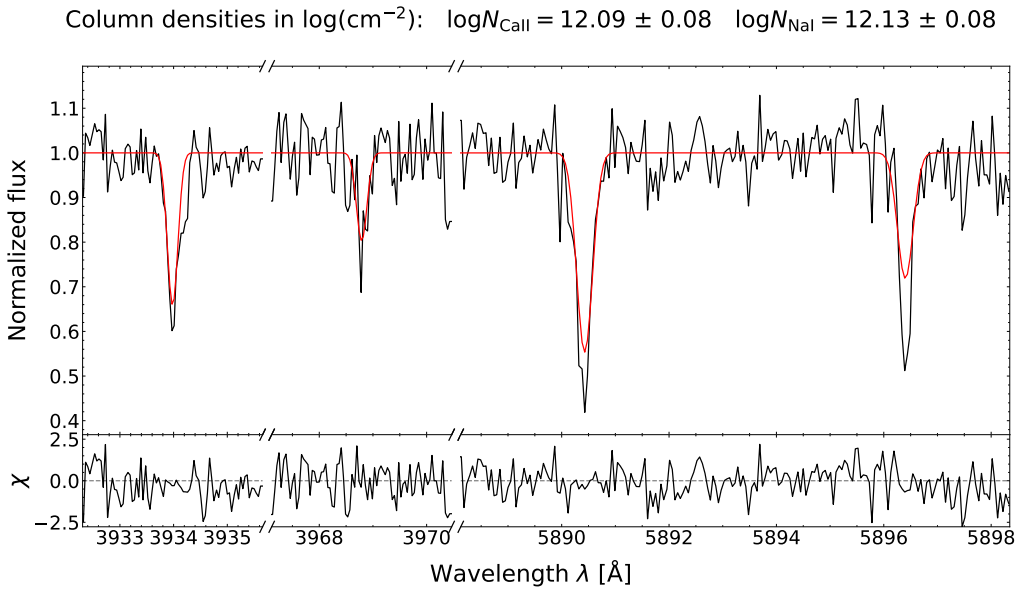


Figure 3.9: Similar plot as Figure 3.8, but showing the Voigt profile fits instead of the Gaussian fits. The measured column densities for each element are noted at the top in units of log(cm^{-2}). The line cores are often not properly reproduced either due to the fixing of the Doppler parameter or because of multiple unresolved ISM components. In the latter case, a single Voigt profile can not fully account for these multiple components.

but this time showing the Voigt profile fits instead of the Gaussian fits.

One big caveat of this fitting procedure is that a simultaneous determination of column densities and Doppler parameters reliably works for high spectral resolution spectra only ($R \gtrsim 100,000$). This problem is discussed further and in more detail below. When working with low resolution spectra, it is recommended to freeze one of the two parameters to a physically reasonable value. Then at least the other one can be decently determined. Another problem can be multiple ISM components again. While the model considers only one single line, the observed lines might consist of multiple unresolved components. If this is the case, sometimes the fit fails to account for the full depth of a line, as for example visible for the Na I D1 line in Figure 3.9. However, in most of the cases the determined column densities are still a valid measure for the total amount of interstellar material along the line of sight. In principle, it is also possible to carry out multiple component fits, as for instance done in Welsh et al. (2010). But for this, again super high spectral resolution is required, which is unfortunately not available in this work. Hence, it did not appear worthwhile to implement it.

The Main Function `run_ism_analysis`

The main function of the `ISM` class is called `run_ism_analysis` and structures as well as operates all the individual steps described above. First, the `get_ism_flux` function is called to isolate and process the interstellar line regions. Then, the individual line analyses including the equivalent width measurements are run by the `run_single_line_analysis` function, which also gives conclusion about each line presence. The results as well as information about the line presences are stored and before continuing with the subsequent steps another consistency check is carried out. In general, the thermal motion and the density of each species of particles is identical within the same interstellar cloud. Therefore, Doppler and pressure broadening (Subsection 1.4.3) are also identical for each line doublet as well as the instrumental broadening. The only parameters that make the line profiles different are the atomic constants. That is why, for instance, the Ca II K line always appears stronger than the Ca II H line (see Figure 3.8 or Figure 3.9). However, since the broadening mechanisms given by the environment and the spectrograph are the same for each element, the difference in width is only determined by the natural line broadening. This in turn is determined by

the life time of the respective atomic state, which is given by the inverse of the transition probability (Subsection 1.2.2 and Subsection 1.4.4). But the transition probabilities within the Ca II and Na I line doublets are actually very similar (Table 1.1), which means the apparent widths (FWHMs) should be comparable. Therefore, the (rather soft) condition is implemented that the FWHM of one line of each doublet is not allowed to be larger than five times the FWHM of the other line. Otherwise, the whole line doublet is considered not to be present. This ensures that always either both lines of a doublet are considered as present or none, which is the only physically reasonable actuality. The presence of only one out of two lines of a doublet is not realistic and if so, it is more likely that a noise peak was identified as line. The rather large factor of five is chosen, because the similar width is just a rough estimate and the measured FWHMs are also subject to uncertainties. Whole line doublets are only removed, if something is really going wrong. But in fact, this is just an edge case. In the vast majority of all spectra of decent quality, it is clearly visible whether the ISM line doublets are present or not.

After this additional check, the global radial velocity determination and the Voigt profile fitting procedure is run on the detected line doublets. Lastly, if the `save` argument is set, the summary PDF as well as the results CSV table of the spectral fit is extended with the results from the ISM analysis (Figure 3.10).

As future prospect, it is planned to additionally incorporate the functionality of searching for a second ISM component, at least in higher resolved spectra, by removing the first component and re-running the analysis. While a first approach is already implemented, it does not seem to reliably work yet and is not yet well tested. Therefore, it is just left as an outlook.

Problems and Caveats

In this section, the problems and caveats of the ISM analysis script are summarized and further discussed. While the previous descriptions can be considered as general manual for the scripts and not necessarily only in the context of hot subdwarf stars and the main aims of this work, the following will specifically deal with the issues that appeared in the course of this project.

In general, most issues appear only in rare cases. All the individual steps and consistency checks ensure that the procedure robustly works and produces decent results. However, the major limitation is the amount of information contained in an observed spectrum, which strongly scales with the spectral resolution and signal to noise ratio. The higher both quantities are, the more information describing the ISM can be deduced and the more precise are the results. The effects of instrumental broadening and SNR are depicted in Figure 3.11, where the Ca II K line is simulated for different spectral resolutions and signal to noise ratios. As it can be seen in the panel without noise on top of the model line, the instrumental profile plays only a minor role for resolving powers $R \geq 50,000$. For lower resolutions, however, instrumental broadening begins to deform the line. It becomes less deep and much broader until even moderately strong spectral lines are smeared out and are not detectable any more ($R < 2,000$). Weaker lines will disappear much earlier. This situation gets even more complicated, if the quality of a spectrum additionally suffers from a low SNR. For SNR ≥ 50 also in the low resolution case, the line is still clearly visible. For SNR ≤ 25, however, it gets completely distorted by the Gaussian noise, while higher resolved lines stay detectable. Obviously, it will be difficult to deduce information from low resolution and low SNR spectra, which is most likely the biggest caveat of this work, since many observations in the collected sample are $R \lesssim 2,000$ SDSS and LAMOST spectra. On the other hand, not all low resolution spectra are super noisy, but can be of decent quality and are therefore still useful. Figure 3.11 also underlines the importance of the lower number of high resolution spectra in the sample. In the following, a few implications of the resolution and SNR problem are briefly summarized and discussed as well as some further complications for the ISM analysis:

- **Uncertainty on the radial velocity:** One simple implication of low resolution spectra is that the radial velocity of single spectral lines cannot be determined with high precision. Fitting individual lines always gives uncertainties of several tens of $\mathrm{km\,s^{-1}}$ ($\Delta v_{\mathrm{rad,ISM}} \sim 30\,\mathrm{km\,s^{-1}}$). By combining the radial velocity measurement of all four ISM lines, the uncertainty reduces

Parameter	1σ confidence interval
Information	
Object name	he0539-4246
Right ascension	85.27796°
Declination	-42.75886°
Galactic longitude l	248.64021°
Galactic latitude b	-30.46122°
Survey/spectrograph	SPY
Estimated SNR	22.2
Effective SNR	21.9
Spectral fit	
Model grid	sdB
Resolution offset $R = \lambda/\Delta\lambda$ (fixed)	20000
Resolution slope $1/\Delta\lambda$ (fixed)	0.000 Å^{-1}
Projected rotational velocity $v\sin(i)$	3 ± 1 km s^{-1}
Effective temperature $T_{\rm eff}$	23381 ± 79 K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$	5.615 ± 0.009
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0.00 dex
He abundance $\log(n(\text{He}))$	-3.784 ± 0.063
Radial velocity $v_{\rm rad}$	41.8 ± 1.1 km s^{-1}
$\chi^2_{\rm red,0}$	1.132
$\chi^2_{\rm red,initial}$	1.021
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	1.011
ISM line analysis	
Equivalent width $W_{\rm CaII\,K}$	0.12 ± 0.03 Å
Equivalent width $W_{\rm CaII\,H}$	0.06 ± 0.02 Å
Equivalent width $W_{\rm NaI\,D2}$	0.19 ± 0.04 Å
Equivalent width $W_{\rm NaI\,D1}$	0.14 ± 0.04 Å
Column density $\log(N_{\rm CaII}\,(\text{cm}^{-2}))$	12.09 ± 0.08
Column density $\log(N_{\rm NaI}\,(\text{cm}^{-2}))$	12.13 ± 0.08
Doppler parameter $b_{\rm CaII}$ (fixed)	5.00 km s^{-1}
Doppler parameter $b_{\rm NaI}$ (fixed)	5.00 km s^{-1}
Radial velocity $v_{\rm rad}$	24.25 ± 0.50 km s^{-1}
$\chi^2_{\rm red,initial}$	1.017
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	1.009

Figure 3.10: Extension of the summary PDF of Figure 3.3 with the results of the ISM analysis. Additionally given are the measured equivalent widths of each line, the fitted column densities of each element, the Doppler parameters frozen to a reasonable value, the global ISM radial velocity, and the fit statistics.

to $\Delta v_{\rm rad,ISM} \sim 10\,\text{km s}^{-1}$. However, if high precision is needed one still has to rely on higher resolution spectra, where the statistical uncertainties are mostly less than $\Delta v_{\rm rad,ISM} \sim 2\,\text{km s}^{-1}$.

- **Voigt profile fits:** For the combined Voigt profile fits, it is crucial to convolve the model that is fitted to the observation with the instrumental profile. However, if the instrumental width completely dominates the true width of a line and the model is resampled to a widely spaced wavelength grid as it is the case for low resolution spectra, the fit is no longer sensitive to the physical fit parameters. The column densities and Doppler parameters become highly correlated, because many different combinations of them can reproduce the observation.

The simultaneous fit of both quantities does no longer work and sometimes even crashes. The only way to derive at least one of them more or less reliably is to freeze the other one to a reasonable value. Then, however, a systematic uncertainty is raised depending on the choice of the frozen parameter. If, for example, the Doppler parameter is frozen to $5\,\mathrm{km\,s^{-1}}$, while the true value would be $3\,\mathrm{km\,s^{-1}}$, the fitted column density will be underestimated to reproduce the full strength of the line. On the other hand, if the true Doppler parameter is $7\,\mathrm{km\,s^{-1}}$, the column density will be overestimated. This issue is further investigated and quantified by applying the script on simulated data in Section 4.3. Even for higher resolved spectra ($R = 10,000 - 50,000$) the issue of the strong correlation does not fully disappear and noise complicates the situation further. In the literature, the simultaneous determination of both parameters is carried out for spectra with $R \gtrsim 100,000$.

- **Equivalent width vs. column density:** During the ISM analysis, equivalent widths and column densities, are determined, which both provide a measure for the strength of the spectral lines. But while the equivalent width represents the overall strength of a line (depth plus width), the column density mostly represents the depth of a line only. However, from equivalent widths alone no conclusion about the physical properties of the interstellar material from which the lines originate can be drawn, while the column density directly gives the number of gas particles along the line of sight. So the question is, which one should be preferred? The answer again depends on the quality of the investigated spectrum, mainly on the spectral resolution. Since instrumental broadening conserves the equivalent widths of a line, no convolution with the instrumental profile is needed. Hence, if the spectra are not too noisy, its determination works robustly for both, high and low resolution ones. On the other hand, the column density determination suffers from ambiguities and (systematic) uncertainties in the low resolution case, especially if spectra are noisy. But as the resolution increases, column densities become more and more accurate. Therefore, it can be said equivalent widths might be preferred for low resolution spectra, because they are determined more robustly than column densities, but for high resolution spectra column densities might be the preferable choice, because they contain more direct information about the interstellar material.

- **Multiple ISM components:** As discussed earlier, for distant stars it is highly likely that their emitted light crossed more than one interstellar cloud before reaching us as observers. This leads to multiple ISM line components (resolved or unresolved) in their spectra, which in turn leads to unusual broad ISM lines if the components are unresolved, distorted line profiles if they are partly resolved, or multiple lines if they are fully resolved. For example, partly resolved multiple component lines can be seen in the right panel of Figure 3.6 or in Figure 3.7. Obviously, this issue can also influence the outcome of the ISM analysis. In the first place, it complicates the line detection and position (radial velocity) determination. If two components are almost equally strong, the program might also confuse lines from both components during the individual line analysis. However, this problem is fixed due to the global radial velocity determination for the combined Voigt profile fits, because there only one of the components is selected and fitted with a fixed radial velocity afterwards. The only issue left concerning the column densities in the case of unresolved components is that the Voigt profile fit expects only one distinct line. Thus, it might have problems to account for the full strength of a line if multiple components are present. This is, however, not an issue for the equivalent width measurement due to the integration over the 5σ Gaussian range.

- **The Hε blend:** The issue that mainly motivated to take the synthetic spectrum from the spectral fit as ISM continuum is the the blend of the Ca II H line with the Hε Balmer line. While the other three ISM lines are usually nicely isolated in the spectra of hot subdwarf stars, this line sits exactly at the tip of the mostly very strong hydrogen line. To overcome this problem, the stellar spectral fit is taken as ISM continuum and removed from the spectrum (Figure 3.4). Then however, it must be ensured that the model atmosphere reproduces the

observation very well, especially the Balmer lines. If, for example, the spectrum of a cooler horizontal branch star is fitted with the sdB model grid, the lower surface gravity of the star compared to hot subdwarfs does not allow to account for the full depth of the Balmer lines. This leads to residual hydrogen lines, when removing the spectral fit from the spectrum, which can then be confused with the Ca II H line. Hence, it should be ensured beforehand that the model really fits the recorded spectrum.

- **Binary companions:** Hot subdwarf stars often reside in binary systems with, for example, F/G/K-type main sequence companions. When recording the spectrum, a composite spectrum is observed including the light from both stars. But now, cooler main sequence stars typically show strong photospheric Ca II and Na I lines. So for example, the strongest spectral lines of the Sun (a G2 star) are not the Balmer lines, but Ca II lines. Hence, those lines from the companion are often clearly visible in the composite spectrum and are not removed from the spectrum, except if the secondary star is included in the spectral model. Hence, it is not reasonable to run the ISM analysis script on such composite spectra. It should be checked beforehand, if the observed star is of single or binary nature. Composite sdB binaries can, for instance, be identified from their SEDs (see Section 3.4) and removed from the sample in the first place. Furthermore, due to the high temperatures and densities in stellar atmospheres compared to the ISM, those intrinsic lines are much broader than the narrow interstellar ones, which should in doubt allow for an easy distinction. Also, if multiple spectra are recorded of the same object, photospheric lines will move with the star and show the same radial velocity shift if it is a binary system, while interstellar lines will always appear at the same position. Therefore, the radial velocity is another indicator for the interstellar or photospheric origin of the spectral lines.

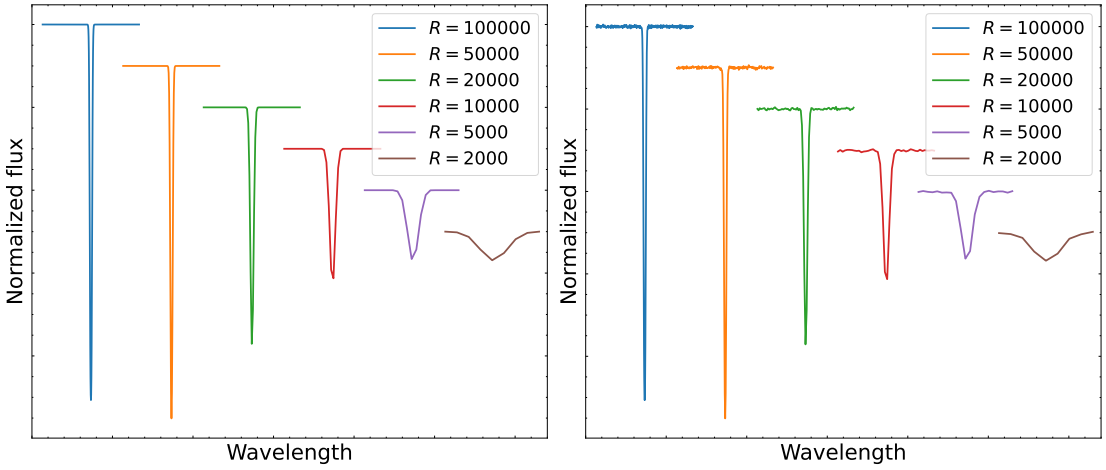
3.4 Spectral Energy Distributions

Apart from the spectral analyses, the fitting of stellar spectra and the investigation of interstellar absorption lines, subsequent photometric analyses are carried out in this work. Spectral energy distributions (SEDs) are constructed from observed magnitudes of stars taken from various photometric surveys. To this end, a tool is needed and one of the above-mentioned ISIS scripts (Section 3.3) exactly fulfills the needs of this work. It was originally implemented by Andreas Irrgang, is now maintained by Matti Dorsch, and constitutes a direct update to the method of Heber et al. (2018).

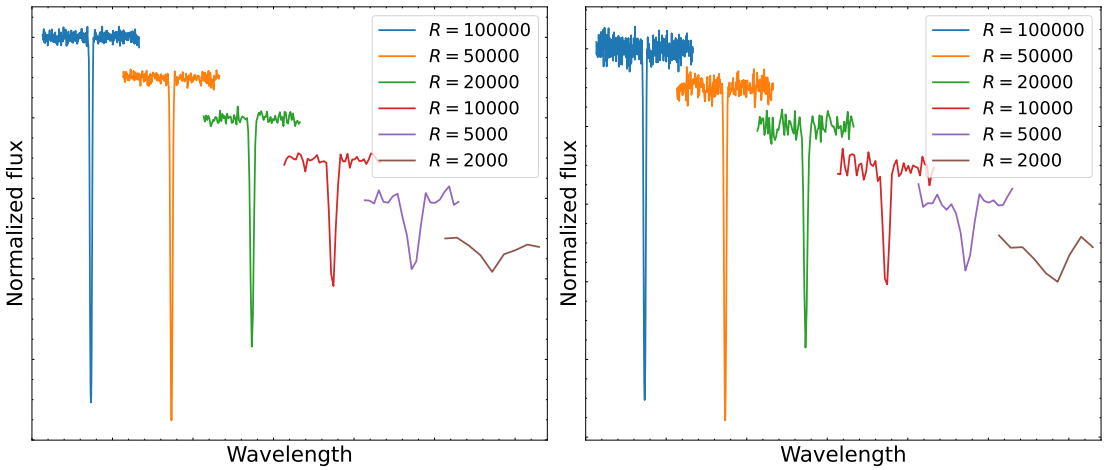
In general, a SED is the total energy output of a star over all wavelengths. Here, as usual in stellar astrophysics, the flux density F_λ in units of erg cm^{-2} s^{-1} Å^{-1} is used as proxy for the energy. Many large and deep photometric surveys scan the sky using numerous filters to obtain apparent magnitudes for various objects in different wavelength bands. Spectral regions from the ultraviolet (UV) to the infrared (IR) are covered. Those pure photometric observations can be conducted for many more stars than spectroscopic ones, in principle for all except the faintest or blended objects. Therefore, from these measurements SEDs can be constructed for the vast majority of observable hot subdwarf stars as well as all other stellar populations. Some of the most important photometric surveys are the following:

- **UV:** The UV is actually the least well covered spectral region. The only comprehensive surveys were performed by the GALEX (Bianchi et al., 2017) and IUE (González-Riestra et al., 2001) satellites. The latter one does not even directly provide photometric measurements, but low resolution spectra, from which three apparent magnitudes are constructed (Heber et al., 2018). Unlike optical light, the UV is absorbed by Earth' atmosphere, which prohibits ground-based observations. Hence, one has to rely on dedicated spacecrafts, but unfortunately, not many of such have been developed so far. This is the reason why the coverage in the UV is the worst.

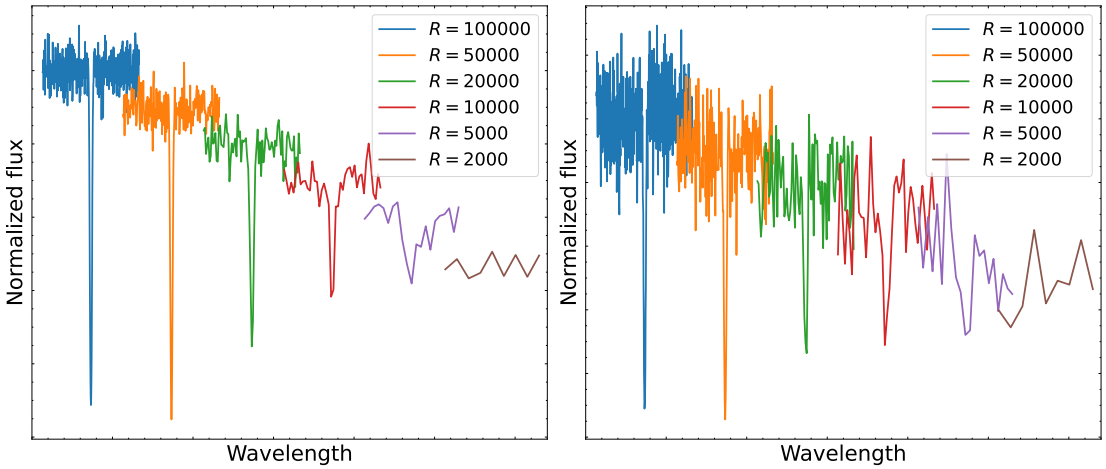
- **Optical:** Since the optical regime is accessible to ground- and space-based observations,



(a) SNR $= \infty$ (b) SNR $= 500$



(c) SNR $= 100$ (d) SNR $= 50$



(e) SNR $= 25$ (f) SNR $= 10$

Figure 3.11: Simulation of the Ca II K line ($\log N = 12.5$ and $b = 5\,\mathrm{km\,s^{-1}}$) for different spectral resolving powers R and signal to noise ratios (SNRs). The Voigt profiles with instrumental broadening are calculated the same way as in the `ism_analysis.py` script on a proper wavelength grid for each resolution and following the absorption theory of Subsection 1.4.4. Gaussian noise is added on top of the model flux to simulate observed absorption lines. For better visibility, the lines are shifted with respect to each other in both, wavelength and flux direction. ISM lines are in the low R plus low SNR case clearly no longer detectable. This issue is further investigated in Section 4.3.

it is very well covered by numerous surveys. Some of the most important ones are SDSS (Alam et al., 2015), APASS (Henden et al., 2015), Pan-STARRS (Magnier et al., 2020), SkyMapper (Onken et al., 2019), DES (Abbott et al., 2021), and *Gaia* (Riello et al., 2021, see Section 2.1). While *Gaia* already provides three photometric measurements, a set of 14 additional apparent magnitudes are constructed from its low resolution XP spectra (Gaia Collaboration et al., 2023b).

- **IR:** The IR regime can be separated into the near-infrared, the mid-infrared and the far-infrared. The near-infrared is closest to the optical and Earth' atmosphere is mostly transparent to it as well, enabling ground-based observations. It is covered by 2MASS (Cutri et al., 2003), UKIDSS (Lawrence et al., 2007), and for example VHS (McMahon et al., 2013) as well as VIKING (Edge et al., 2013), which are two of several surveys carried out at the VISTA telescope. With increasing wavelength, Earth' atmosphere becomes more opaque. Thus, further photometry in the near- to mid-infrared is provided by the WISE satellite (Schlafly et al., 2019). The far-infrared does not play a role for stellar SEDs, because even cooler stars do not significantly contribute with flux in this spectral region.

The complete list of in total 66 available photometric surveys for the SED construction is, for example, provided by Table B.1 of Culpan et al. (2024). In the next section it is described, how meaningful parameters can be derived from all these photometric measurements.

3.4.1 The SED Fitting Method

To access the photometric data, the script first automatically queries from all available databases and catalogs based on the coordinates of the star. This functionality was mainly implemented by Kreuzer (2021). Then, the program tries to find the best-matching model SED to the observed apparent magnitudes using χ^2-minimization. These models are the same as those described in Section 3.1 in non-normalized form. To actually compare the measurements with the models, synthetic magnitudes need to be calculated. According to Heber et al. (2018), these are given by

$$\mathrm{mag}_x = -2.5 \log \left(\frac{\Theta^2 \int_0^\infty r_x(\lambda) 10^{-0.4A(\lambda)} F(\lambda) \lambda \mathrm{d}\lambda}{4 \int_0^\infty r_x(\lambda) f^{\mathrm{ref}}(\lambda) \lambda \mathrm{d}\lambda} \right) + \mathrm{mag}_x^{\mathrm{ref}}, \tag{3.11}$$

where mag_x is the magnitude of an arbitrary pass-band x, Θ is the angular diameter of the star, $r_x(\lambda)$ is the response function of the filter, $A(\lambda)$ is the interstellar extinction as function of the color excess $E(B-V)$ and the extinction parameter R_V (see Subsection 1.2.3), $F(\lambda)$ is the model flux at the stellar surface, and the flux of a reference star $f^{\mathrm{ref}}(\lambda)$ (usually Vega) is needed to set the zero point of the filter to a predefined magnitude $\mathrm{mag}_x^{\mathrm{ref}}$. More details on the synthetic magnitude calculation can be found in Dorsch (2023). Interstellar extinction is modeled using the empirical extinction curves of Fitzpatrick et al. (2019). For more details on interstellar extinction and reddening in general also see Subsection 1.2.3, while more details on the implementation in the SED fitting script can be found in Dorsch (2023) as well.

According to Planck's law, the emergent flux distribution of a star has one distinct peak at a certain wavelength depending on the effective temperature T_{eff}. Hence, the SED fits are sensitive to the temperature of a star, which constrains the selected model. However, if prior knowledge on the atmospheric parameters is given, for example from spectroscopy as in this work, these parameters can be specified as input to choose the right model beforehand. Since in particular the surface gravity, which is also an essential parameter of the models, is completely unconstrained from SED fits alone, this makes the whole routine more robust. Moreover, hot stars emit most of their flux in the far-UV (or at even smaller wavelength), which is usually not covered by photometric measurements. This makes even the effective temperature determination less accurate. Therefore, providing atmospheric parameters is the best way to derive the most reliable results.

Furthermore, the fitting routine does not only consider single SEDs. If, for example, a F/G/K-type main sequence (MS) companion is present, the composite SED of both components will be double peaked. In the case of a primary hot subdwarf, there is one peak in the UV (hotter star) and

another one at longer wavelengths (cooler star). Thus, the SEDs can give conclusion about the single or binary nature of an object. Since the spectral regions from the near-UV to mid-IR are often well covered by photometric measurements, even the temperature of the cooler star can be accurately determined. To model the SED of the cool companion the PHOENIX grid from Husser et al. (2013) is used.

3.4.2 Transition from Atmospheric to Stellar Parameters

One fit parameter of the SED fitting routine is the angular diameter Θ of a star (Equation 3.11). Combined with parallax measurements ϖ from *Gaia* (Section 2.1), which directly gives the distance d to a star, stellar radii can be determined with the geometric relation

$$R = d\frac{\Theta}{2} = \frac{\Theta}{2\varpi}\,. \tag{3.12}$$

Next, from the radius R and the effective temperature $T_{\rm eff}$ the stellar luminosity can be calculated via the Stefan-Boltzmann law

$$L = 4\pi R^2 \sigma_{\rm SB} T_{\rm eff}^4\,, \tag{3.13}$$

where $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant. Furthermore, if also the surface gravity g is known, e.g. from spectroscopy, the stellar mass M can be derived from Newton's law

$$M = g\frac{R^2}{G}\,, \tag{3.14}$$

where G is the gravitational constant. Especially this last step is of particular importance, because even though the mass M is one of the most crucial parameters that describe the evolution and the fate of a star, it is often hardly determinable. But by combining spectroscopy, photometry, and *Gaia* astrometry, the full set of atmospheric and stellar parameters can be obtained. This impressively shows the significance of the interplay between different analysis techniques and the power of the unprecedented *Gaia* mission for stellar astrophysics.

3.4.3 The Power of SEDs

Constructing and fitting model SEDs to the observations of apparent magnitudes in different wavelength bands allows for the determination of various parameters. If not already available from spectroscopy, the effective temperature of star can be derived as well as the angular diameter and the color excess from interstellar extinction and reddening. Moreover, in the case of a F/G/K-type MS companion, the binary nature of a hot star can be confirmed from its double peaked SED.[16] Even the temperature of the cool companion can be accurately determined as well as the surface ratio $A_{\rm cool}/A_{\rm hot}$ of both components. Examples of an SED fit for the single and binary case, respectively, are shown in Figure 3.12.

In this work, it is made extensive use of the SED fitting method, which complements the spectral analysis. The atmospheric parameters derived from spectroscopy for each object in the sample are passed as input to obtain the most reliable results. In particular, the color excess is an essential parameter for the aim of this work as it quantifies the amount of intervening interstellar dust between us and the observed object (see Subsection 1.2.3). Moreover, information about the potential binary nature of a star is crucial, since those systems necessarily need to be excluded from the sample as the binary companions possibly contribute with photospheric Ca II H&K and Na I D lines, which would be confused with the interstellar lines during the spectral analysis. Lastly, even though it is not the main aim of this work, the SEDs combined with precise parallaxes provide stellar parameters for a large sample of hot subdwarf stars. This is only possible owing to the unprecedented *Gaia* mission and was never feasible in the pre-*Gaia* era.

[16] Even fainter companions such as M-dwarfs, white dwarfs, or sub-stellar objects like brown dwarfs or planets can not be detected with the SED method, because they are outshone by the hotter component even at longer wavelengths. Vice versa, O/B/A-type MS companions usually outshine subdwarfs making it hard to detect such systems.

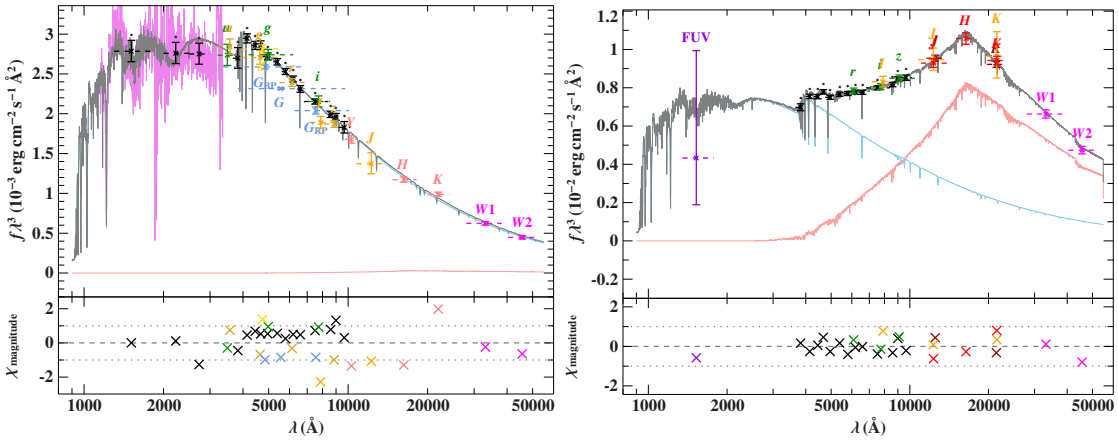


Figure 3.12: Example SED fits. Shown in the upper panel of each plot are the data points (apparent magnitudes with corresponding filter labels), the composite model SED in gray, the hot subdwarf model SED in blue, and the MS companion model SED in red. The flux density distribution is flattened for better visualization by a factor of λ^3. The lower panels show the residuals of the fit. *Left figure:* Single peaked SED, where the flux of the MS companion is zero. Additionally depicted in magenta is the IUE spectrum from which the data points in the UV colored in black were constructed. *Right figure:* Clearly double-peaked SED proving the existence of a cool companion to the primary hot subdwarf star.

Chapter 4

Application of the Scripts and First Results

When carrying out analyses in stellar astrophysics, there are typically two choices. Either investigating a small number of objects in great detail, for example to understand their formation and evolution, or studying one or more whole populations of stars to obtain their parameters ranges or relative frequencies of their formation channels. In this work, the latter one is the case, for which a large sample of spectra of hot subdwarf stars has been collected (Section 2.3). However, the main aim here is actually not to study the stars themselves, but to extract information about the interstellar medium from their spectra. Still, an important aspect is to obtain model spectra for the stars in order to use them as continuum for the ISM and to resolve blends between photospheric and interstellar lines (see Section 3.3) as well as to feed the derived atmospheric parameters into the SEDs in order to obtain information about interstellar dust. As side-effect, fundamental parameters of the hot subdwarf stars are derived. Overall, a set of dedicated scripts is used for the analyses, which are all described in Chapter 3. To conduct the whole analysis, all the individual steps that built upon each other need to be structured and connected. Therefore, an analysis pipeline has been developed, which is outlined in Section 4.1. In Section 4.2 the results regarding the parameters of the hot subdwarf stars are presented. Even though those are not central for this thesis, it is important to check them for consistency, since the subsequent steps are based on them. After this is done, in Section 4.3 the `ism_analysis.py` script is applied to simulated synthetic spectra in order to test the performance and to quantify potential caveats and systematic uncertainties. Additionally, supplementary material can be found in Appendix C. It contains further parameter distributions as well as additional plots regarding this chapter.

4.1 Analysis Pipeline

The analysis pipeline of this work comprises multiple distinct steps such as spectral fits, ISM line analyses, and SED fits, for which extensive use of the slurm computation cluster at the Remeis observatory in Bamberg was made[1]. Furthermore, the results of each step need to be processed and combined. An outline of the pipeline is depicted in Figure 4.1.

The starting point is the raw collection of spectra from each survey/spectrograph (Section 2.2). Then initial spectral fits using `spectral_fit.py` (see Subsection 3.3.3) are applied on the sample, for which the sdB model grid (see Section 3.1) is entirely used. Proper initial values are chosen[2], but only the effective temperature $T_{\rm eff}$, the surface gravity $\log g$, the helium abundance $\log n(\mathrm{He}) \coloneqq \log n(\mathrm{He})/\log n(\mathrm{H})$, and the radial velocity $v_{\rm rad}$ are varied during the fit for all spectra. The metallicity z and the microturbulence ξ are always kept frozen, while the projected rotational velocity $v\mathrm{rot}\sin\mathrm{i}$ is fixed to zero for the low resolution spectra (SDSS, LAMOST, and DESI), but varied for the higher resolution ones (X-Shooter, SPY/UVES, and FEROS). The reason for the

[1] To automatically set up the slurm jobs, another set of python scripts was developed. These can also be found on the internal observatory gitlab or provided upon request.

[2] $T_{\rm eff} = 30{,}000\,\mathrm{K}$, $\log g = 5.5$, $z = 0.0$, $\log n(\mathrm{He}) = -1.0$, $\xi = 0\,\mathrm{km\,s^{-1}}$, $v_{\rm rad} = 0\,\mathrm{km\,s^{-1}}$, and $v_{\rm rot}\sin\mathrm{i} = 0\,\mathrm{km\,s^{-1}}$

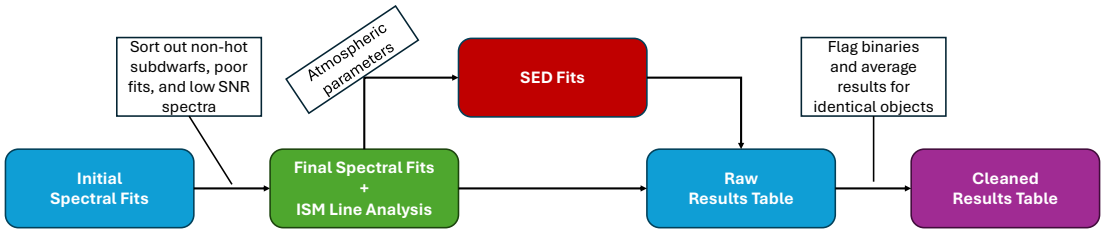


Figure 4.1: Flow chart of the analysis pipeline. First, initial spectral fits are executed on the whole sample of objects. All spectra are sorted out, which are considered not to belong to hot subdwarf stars, show poor spectral fits, or have low signal to noise ratios. Subsequently, final spectral fits are carried out as part of the ISM line analysis. The resulting atmospheric parameters are passed to the SED fits. The results from both, the ISM line analysis including the spectral fits as well as the SED fits, are combined to a common table. Systems with F/G/K-type MS companions are flagged and the results for identical objects are averaged to end up with a cleaned final results table. More details about the individual steps can be found in the text.

latter distinction is that to reliably fit the rotational velocity it is favorable to have better resolved line shapes and the spectral resolution parameters need to be well known. Both is only the case for the higher resolution spectra. These initial spectral fits are carried out in order to sort out non-hot subdwarf stars (e.g., horizontal branch stars, white dwarfs, and B-type main sequence stars), as which all objects are considered that approach the grid borders in $T_{\rm eff}$ or $\log g$, poor fits with high uncertainties ($\Delta T_{\rm eff} > 5,000$ K, $\Delta \log g > 2$, or $\Delta v_{\rm rad} > 50\,{\rm km\,s^{-1}}$), and low signal to noise spectra (SNR < 10)[3]. Additionally, a lower effective temperature cut is set at $T_{\rm eff} = 20,000$ K, because for cooler stars it gets more and more likely that their spectra show photospheric Ca II H&K as well as Na I D lines, which would blend the interstellar ones. Moreover, it is distinguished between helium-poor ($\log n$(He) smaller than the upper grid border) and helium-rich objects ($\log n$(He) approaching the upper grid border).

Subsequently, the final spectral fitting run is executed with the same initial values as before as part of the ISM line analysis[4] using `ism_analysis.py` (Subsection 3.3.4), where the model grid for the helium-rich objects is replaced by the He-sdO model grid (see Section 3.1). For the ISM analysis the initial values $N_{\rm Ca\,II,Na\,I} = 10^{12}\,{\rm cm^{-2}}$, $b_{\rm Ca\,II,Na\,I} = 5\,{\rm km\,s^{-1}}$, and $v_{\rm rad,ISM} = 0\,{\rm km\,s^{-1}}$ are used, but the Doppler parameter b is kept frozen for both elements due to the issues outlined in Subsection 3.3.4.

The derived atmospheric parameters ($T_{\rm eff}$, $\log g$, and $\log n$(He)) are then passed as input to the spectral energy distribution (SED) fitting method (see Section 3.4) to constrain the model SED before running the script and to calculate stellar parameters.

The results from both, the ISM line analysis (plus spectral fits) as well as the SED fitting, are then combined to a common table. During the process of creating the table, also the resulting stellar parameters from the SED fits are checked for consistency. If the parallax measurement of *Gaia* has an uncertainty of more than 20%, a bad parallax flag is raised and if the resulting stellar parameters are completely off from the expected values for hot subdwarf stars (radius $R > 1\,R_\odot$, mass $M > 2\,M_\odot$, or luminosity $L > 3,000\,L_\odot$) or the number of photometric data points is below 10, a poor SED flag is raised.

Now, this table still contains the systems with F/G/K-type main sequence companions, which are detected in the SEDs. Such binary systems need to be removed from the sample as the companions might contribute with photospheric Ca II and Na I lines. In the following, with the term "binary" always such systems with composite SEDs are meant. Apparently single stars are called "single", but still might be in close binary systems, where the companion is too faint to

[3] For the results of the ISM line analysis even stricter SNR cuts are applied later on to the low resolution spectra (see Section 5.1).

[4] A spectral fit is always the first step of the ISM line analysis script to take the resulting synthetic spectrum as continuum for the ISM flux (see, for example, Figure 3.4).

be observed in the SED. Furthermore, many stars have multiple spectroscopic observations and the pipeline was ran on all of those individually, possibly resulting into multiple rows per unique object. The first issue is solved by flagging all objects with a binary flag, where the minimum surface ratio $A_{\rm cool}/A_{\rm hot}$ of both components determined during the SED fit is larger than three[5], which is a conventional choice. Overall, for about one third of all objects a binary companion has been detected (Figure 4.2 and Table 4.1). The latter issue is actually more complicated, because the different spectra for the same object can be of various quality. Higher or lower SNRs often only lead to smaller or larger statistical uncertainties on the derived parameters, respectively, and one could think about simply averaging the results. However, varying spectral resolutions or wavelength coverages (especially in the blue/near UV regime) can lead to systematics on the derived parameters. Hence, it did not appear reasonable to average the results from spectra that come from different surveys/spectrographs and only multiple rows for identical objects are averaged, if the survey/spectrograph is the same. To this end, the commonly used formulas for the weighted average are used:

$$\bar{x}_{\rm w} = \frac{\sum\limits_{i=1}^{N} w_i x_i}{\sum\limits_{i=1}^{N} w_i} \qquad \text{and} \qquad \sigma_{\bar{x}_{\rm w}} \sqrt{\frac{1}{\sum\limits_{i=1}^{N} w_i}} \qquad \text{with} \qquad w_i = \frac{1}{\sigma_i^2} \tag{4.1}$$

Here, $\bar{x}_{\rm w}$ is the weighted average with uncertainty $\sigma_{\bar{x}_{\rm w}}$ and w_i, x_i, and σ_i are the statistical weight, the mean value, and the uncertainty of each measurement, respectively. However, the weighted average does not intrinsically account for systematic deviations, for example if the mean values scatter more than expected. This can for instance happen, if multiple spectra of the same object give fundamentally different results because of their differing quality. Also, if an observed star resides in a close binary system with an unseen companion, the radial velocity can change drastically between different epochs. To account for this possibilities, a χ^2-rescaling of the resulting uncertainty is implemented analog to Section 3.2. The reduced χ^2 is calculated according to

$$\chi^2_{\rm red} = \frac{\chi^2}{N-1} \qquad \text{with} \qquad \chi^2 = \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}_{\rm w}}{\sigma_i} \right)^2 , \tag{4.2}$$

where N is the number of measurements. Then, $\sigma_{\bar{x}_{\rm w}}$ is rescaled with the factor $\sqrt{\chi^2_{\rm red}}$, but only if $\chi^2_{\rm red} > 1.5$, which means the mean values of the parameter scatter much more than expected. Inflated uncertainties then indicate potential systematic deviations between the measurements. If $\chi^2_{\rm red} \approx 1$ everything behaves as expected and the uncertainty is kept unchanged. On the other hand, if $\chi^2_{\rm red} < 1$ the uncertainty is theoretically overestimated. However, this condition can occur very quickly, when only a small number of measurements is considered and the results perfectly match. Rescaling would then lead to a unreasonably small uncertainty and is therefore not done. This averaging procedure is applied to all parameters in the table and a new final results table is generated. Also a new column is added specifying the number of spectra the results were averaged from. This final table is used for all the following findings and investigations of this work.

A summary table about the analyzed sample after sorting out unwanted spectra as first step of the analysis pipeline is shown in Table 4.1.

4.2 Atmospheric and Stellar Parameters

While one of the aims of this work is to investigate the interstellar medium leveraging the spectra of hot subdwarf stars, a significant side-effect is the determination of various new atmospheric and stellar parameters for a large sample of these stars through extensive spectroscopic and photometric analyses. Those parameters are briefly discussed in this section, even though a detailed study is

[5] If the object is a single star or the companion is too faint to be detected by the SED method (e.g., a white or M dwarf), the surface ratio is (close to) zero.

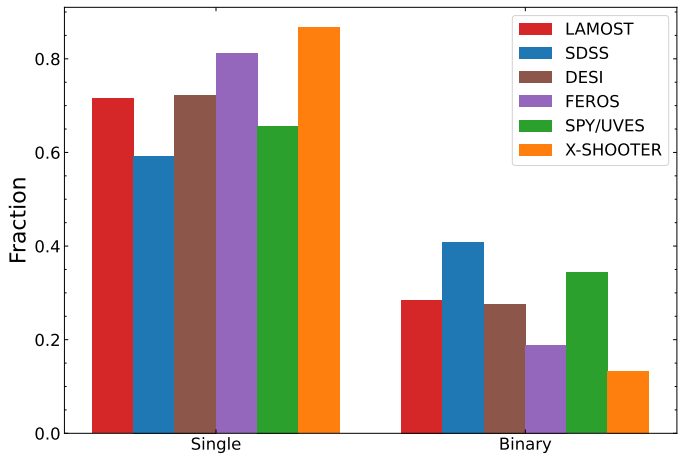


Figure 4.2: Fraction of single and binary stars within each survey/spectrograph. Only systems with F/G/K-type MS companions are detected via their SEDs. Apparently single stars might still reside in close binary systems with faint companions that con not be detected in the SED.

Table 4.1: Number statistics of the analyzed sample per survey/spectrograph and in total after sorting out non-hot subdwarf stars, poor spectral fits, and low SNR spectra. Shown is the number of spectra N_{spectra}, the number of unique objects N_{objects}, the number of unique single stars N_{single}, and the number of unique binary stars N_{binary}. The sum of the objects of all surveys is higher than the total number of objects given in the last column, because identical stars are sometimes observed across different surveys. Those are not counted multiple times for the total number of unique objects.

	SDSS	LAMOST	DESI	FEROS	SPY/UVES	X-Shooter	In total
N_{spectra}	2,235	2,581	666	548	235	135	6,400
N_{objects}	1,587	1,705	666	114	121	18	3,431
N_{single}	934	1,231	485	88	80	16	2,349
N_{binary}	653	474	181	26	41	2	1,092

not in the scope of this work. However, the potential for comprehensive follow-up using this data is outlined at the end of this section.

The atmospheric parameters (T_{eff}, $\log g$, and both equivalent helium abundances $\log n(\mathrm{He})$ and $\log y$) were determined via spectral fits carried out as part of the ISM analysis as described in Section 3.3 and Section 4.1. To run the corresponding script on in total 6,400 spectra (Table 4.1) it was made use of the computation cluster at the Remeis observatory in Bamberg. If enough resources are available, it allows for the execution of up to 155 jobs in parallel. Therefore, it is easily possible to complete the whole spectral analysis within one day. Subsequently, these atmospheric parameters are passed to the SED fitting routine. Systematic uncertainties of 3% in T_{eff} and 0.01 in $\log g$ are added quadratically to the pure statistical uncertainties from the spectral fits, which is conventionally done. The SED fits were run in parallel analog to the spectral fits using the Remeis cluster as well. The photometric data is automatically collected from numerous catalogs and data bases using variable search radii depending on the astrometric accuracy of each queried survey. When dealing with individual SED fits, the resulting data points can be checked manually to reject spurious ones from the fit. For a large sample of stars, however, this is no longer possible. Instead, observations with various bad quality flags are excluded by the program beforehand and 5σ outlier magnitudes are automatically removed in consecutive fitting runs. Then, the SED fits are carried out using χ^2-minimization. Usually, the program tries to find the best matching model grid to the observation. But if the atmospheric parameters are provided as input, this step is redundant, since they already constrain the model. Direct fit parameters are the angular diameter Θ (see Equation 3.11) and $E(44-55)$, which is the monochromatic analog to the color excess $E(B-V)$. Interstellar reddening is treated following Fitzpatrick et al. (2019) and assuming an extinction parameter of $R(55) = 3.02$, typical for the Milky Way (see Subsection 1.2.3). The resulting reddening from the SEDs is of particular importance for this

work, but discussed separately in Section 5.1. For systems with available *Gaia* DR3 parallax measurements, from the atmospheric parameters together with the determined angular diameter the transition to stellar parameters (radius R, luminosity L, and mass M) can be made according to Subsection 3.4.2. In the SED script, this is automatically done and the uncertainties on the input parameters are propagated using the Monte Carlo method. From the resulting parameter distributions the median and 1σ quantiles (for an explanation see Appendix A) are used for the final values and uncertainties, respectively. Moreover, potential cool binary companions are detected in the SEDs, where the temperature of the companion is capped to 7,000 K, because for higher ones the SED would get more and more similar to that of the hot subdwarf such that both components are no longer distinguishable. The minimum temperature of 2,300 K is given by the PHEONIX grid border (Section 3.1). The surface ratio $A_{\rm cool}/A_{\rm hot}$ is capped to 900.

In Figure 4.3 to Figure 4.8 plots with different combinations of atmospheric and stellar parameters are shown. The different surveys/spectrographs from which the spectra come are indicated by colors and symbols. If only atmospheric parameters are considered, the measurements of all objects are shown excluding only those with detected binary companions. If stellar parameters are considered, those are additionally rejected, which are flagged indicating a poor SED fit or uncertain parallax measurement ($\Delta\varpi > 20\%$).

Figure 4.3 shows the Kiel diagram of the analyzed sample, where the surface gravity $\log g$ is plotted over the effective temperature $T_{\rm eff}$. Additionally depicted is the position of the extreme horizontal branch (EHB) from Dorman et al. (1993) (upper right corner down to the middle of the plot) as well as the position of the helium main sequence (He-MS) from Paczyński (1971) (middle to the left side of the plot). The first one represents the blue end of the horizontal branch sdB and sdO stars are associated with, while the latter one is a theoretical sequence for stars purely consisting of helium (see, e.g., Heber (2016)). It can be seen that the vast majority of objects is located exactly on or close to these sequences. Furthermore, three distinct clusters of stars are visible: The most prominent one is elongated on the EHB and centered at $(T_{\rm eff},\ \log g) \approx (27,000\,{\rm K},\ 5.50)$. It is associated with He-poor sdB stars. Another big one is located close to the junction of the EHB and He-MS at $(T_{\rm eff},\ \log g) \approx (35,000\,{\rm K},\ 5.85)$ and belongs to the population of (iHe-)sdOB[6] stars. The origin of these two clusters is not yet fully understood, but matches the groups identified by Németh et al. (2012), Luo et al. (2021), and Geier et al. (2022). The last clustering located on the He-MS and ranging towards higher temperatures corresponds to all kinds of sdO stars. Actually, an additional distinct population of objects below the EHB was recognized by Geier et al. (2022) and He et al. (2025) and called bEHB stars, however, their origin is unclear. Many stars in Figure 4.3 seem to match that position as well.

Another diagnostic tool with which different types of hot subdwarf stars can be clearly distinguished, is the helium abundance $\log y$ over effective temperature $T_{\rm eff}$ plot as shown in Figure 4.4. The dashed black line represents the solar helium abundance and the dotted black line the transition between intermediate and extreme helium enrichment (Dorsch, 2023). The clusters of sdB, sdOB, iHe-sdOB, and eHe-sdO stars are marked with black ellipses. All objects in the lower left corner at high temperatures and low helium enrichment are considered to be helium-poor sdO's. Figure 4.3 and Figure 4.4 can be combined by color-coding the helium abundance in the Kiel diagram according to poor-, intermediate-, and extreme-helium enrichment (Figure 4.5). The three clusters of sd(O)B, iHe-sdOB, and eHe-sdO stars are clearly visible.

Figure 4.6 shows the Hertzsprung-Russell diagram (HRD, luminosity L over effective temperature $T_{\rm eff}$), Figure 4.7 the radius R over $T_{\rm eff}$ diagram, and Figure 4.8 the L over R diagram. The same sequences as before are depicted in gray. Even though less objects are included, because those with a poor SED or bad parallax flag were excluded, the same pattern as in Figure 4.3 is visible. The position of the vast majority of stars agrees well with the EHB or He-MS and there are the same three apparent clusters of stars in each plot. However, in this three plots as well as in Figure 4.3 some positions of objects are also quite off from the shown theoretical tracks. The main reason for that is, that in fact, hot subdwarf stars can of course also be observed in evolutionary stages, when they evolve towards or away from these sequences (Dorman et al., 1993; Han et al.,

[6] An explanation of the abbreviations used for the different hot subdwarf populations can be found in Figure 1.7.

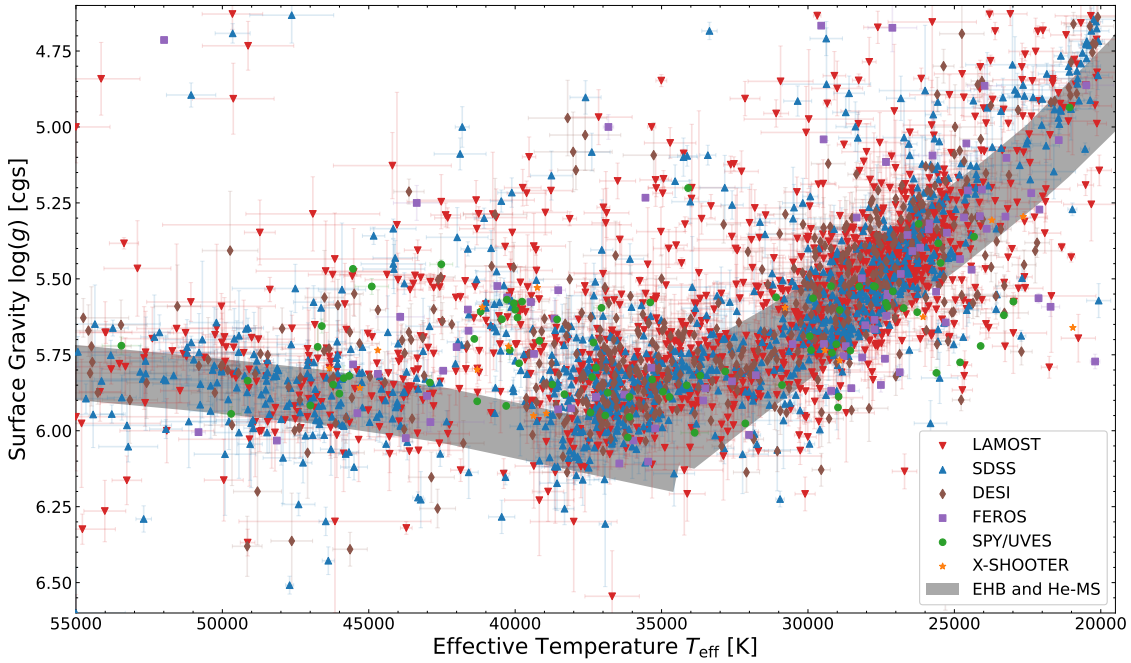


Figure 4.3: Kiel diagram of the analyzed sample of hot subdwarf stars. The observations from different surveys/spectrographs are indicated by different colors and symbols. Additionally shown are the extreme horizontal branch (EHB) from Dorman et al. (1993) and the helium main sequence (He-MS) from Paczyński (1971). Their shapes were slightly modified by Matti Dorsch for a better visualization.

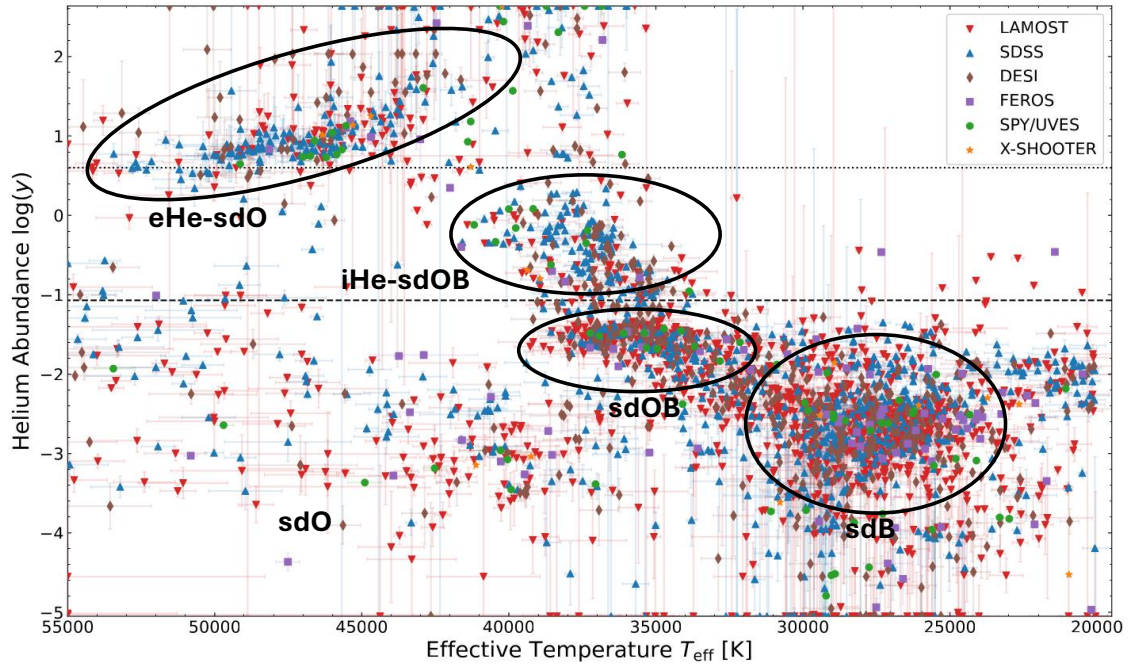


Figure 4.4: Helium abundance $\log y \coloneqq \log n(\mathrm{He})/n(\mathrm{All}) \approx \log n(\mathrm{He})/(n(\mathrm{H})+n(\mathrm{He}))$ over effective temperature $T_{\rm eff}$ diagram of the analyzed sample of hot subdwarf stars. The observations from different surveys/spectrographs are indicated by different colors and symbols. The dashed black line represents the solar helium abundance, while the dotted black line distinguishes intermediate helium-enrichment from extreme enrichment (Dorsch, 2023). Abundances that approach the model grid borders are shown at the plot limits.

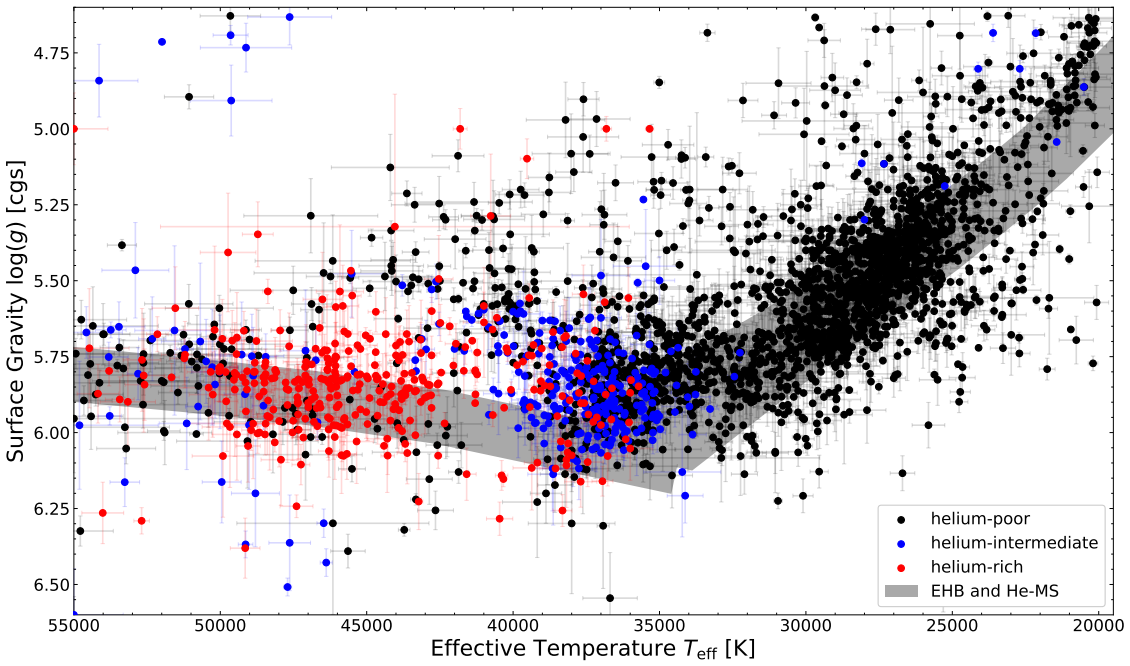


Figure 4.5: Same as Figure 4.3, but with a color-coded helium abundance $\log y$. Different clusters with a low helium abundance ($\log y < -1.07$) in black, an intermediate helium abundance ($-1.07 \leq \log y < 0.06$) in blue, and an extreme helium enrichment ($\log y \geq 0.06$) in red are clearly visible. They correspond to sd(O)B, iHe-sdOB, and eHe-sdO stars, respectively.

2003; Yu et al., 2021). Then they are located above or below them. Furthermore, the evolutionary tracks of other kinds of stars can also intersect the parameter space of hot subdwarf stars such as post-AGB stars (Miller Bertolami, 2016) or pre-extremely low mass white dwarfs (pre-ELMs, Driebe et al., 1998; Althaus et al., 2013; Istrate et al., 2016). A little contamination of the sample with such objects cannot be excluded.

The last stellar parameter not considered yet is the stellar mass M. A histogram showing the distribution of measured masses for the analyzed sample is shown in Figure 4.9. It peaks very close to the canonical mass of $0.47\,M_\odot$ for the sdB component (Heber, 2016). However, the mass ranges for different hot subdwarf populations can be quite different. So for example, while typical masses are subsolar, the masses of eHe-sdO's often exceed $1\,M_\odot$ (Dorsch, 2023) explaining the tail towards higher masses in Figure 4.9. Therefore, it is actually not reasonable to lump all types of hot subdwarfs together, but to consider their mass distributions individually. This is for instance done in Dorsch (2023), Dawson et al. (in prep.), Latour et al. (in prep.), or Heber et al. (in prep.), but is not in the scope of this work.

All these plots do not only show the agreement between observations and theory, but also prove that the developed methods for the spectral analysis provide decent results. None of the parameters seem to be systematically over- or underestimated compared to the predictions. Furthermore, no systematics between the different surveys/spectrographs are visible. This is actually an important conclusion and is also supported by the supplementary material found in Section C.1.

Lastly, in Figure 4.10 the distribution of measured barycentric radial velocities on the sky is shown in Galactic coordinates. Apparently, the Western side is dominated by negative velocities, meaning the stars are approaching us, while the Eastern side is dominated by positive velocities, meaning the stars are receding from us. Unfortunately, the Eastern side is much more sparsely covered, but this trend is still clearly visible. It is most likely the consequence of differential Galactic rotation.

The determination of atmospheric and stellar parameters for a large sample of stars allows for a comprehensive analysis of their intrinsic properties as well as their formation and evolution channels. This is for example done in Dorsch (2023), where much more detailed information on this topic can be found. In principle, the collected data set here allows to do some of the

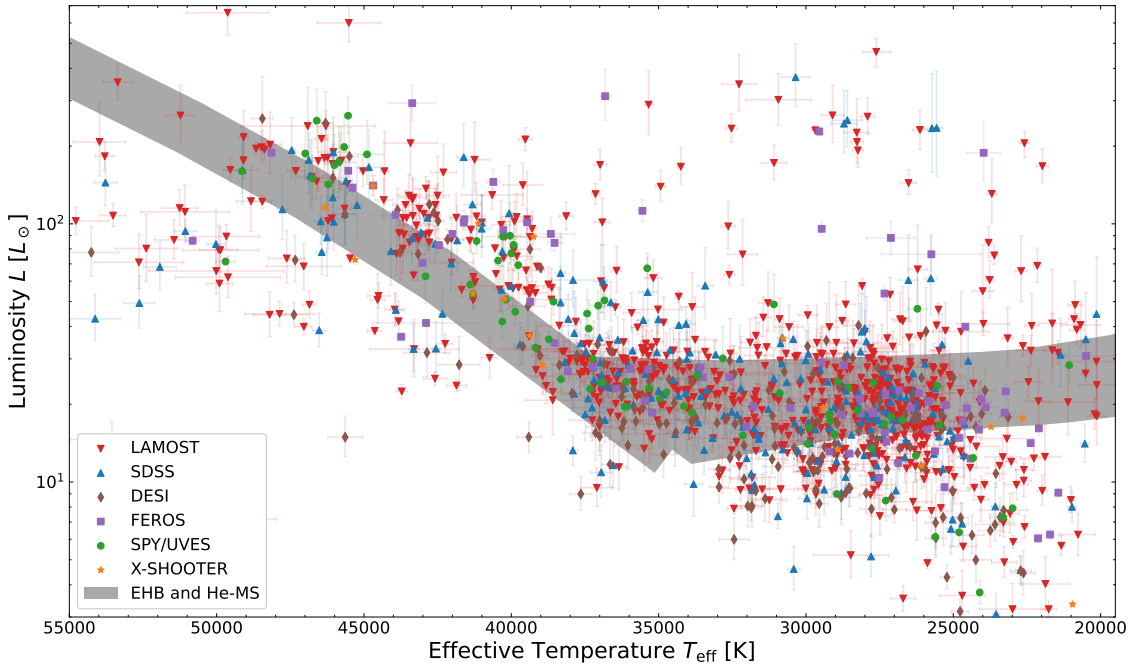


Figure 4.6: Hertzsprung-Russell diagram (HRD) analog to Figure 4.3. Less stars are visible as all SED fits flagged with a poor SED or bad parallax flag are excluded.

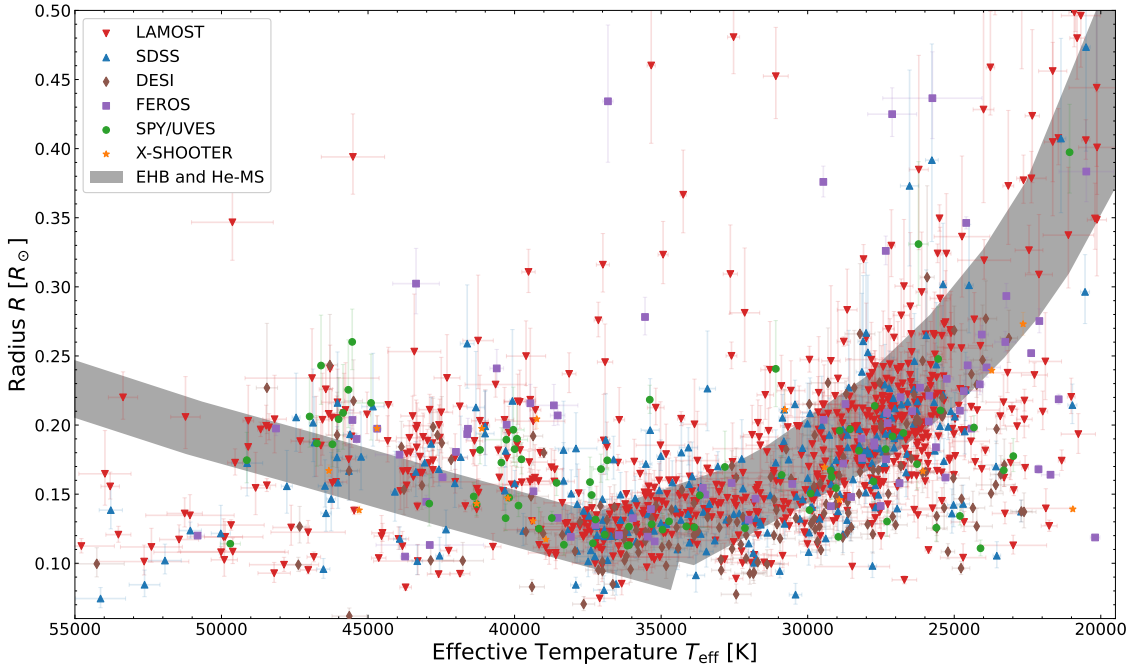


Figure 4.7: Radius R over effective temperature $T_{\rm eff}$ diagram analog to Figure 4.6.

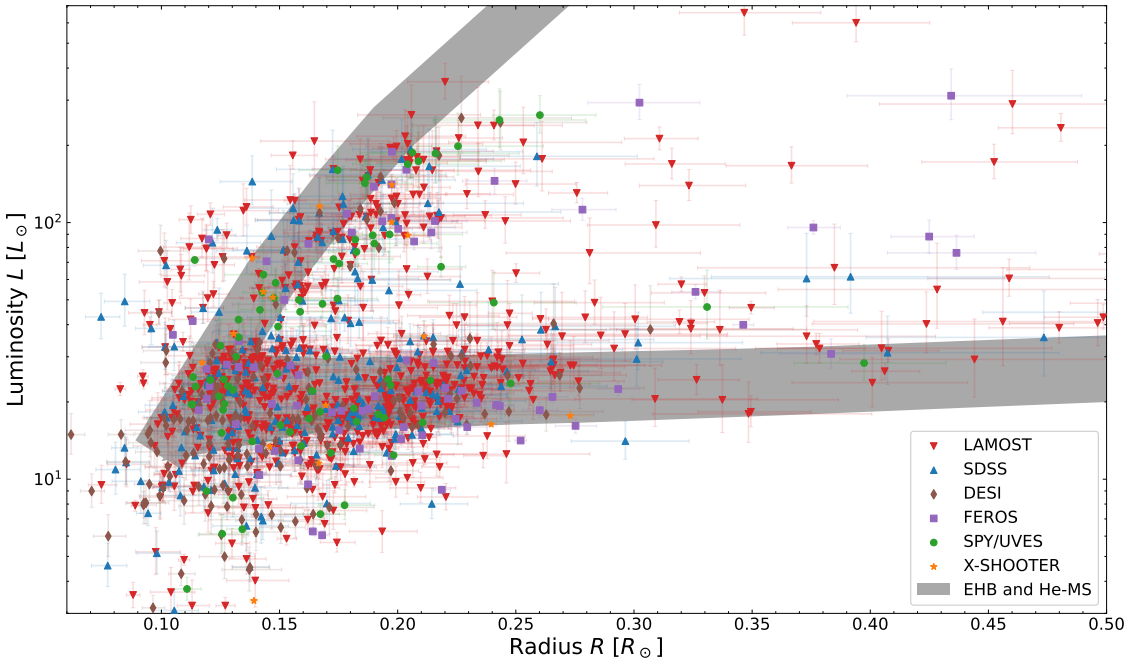


Figure 4.8: Luminosity L over radius R diagram analog to Figure 4.6.

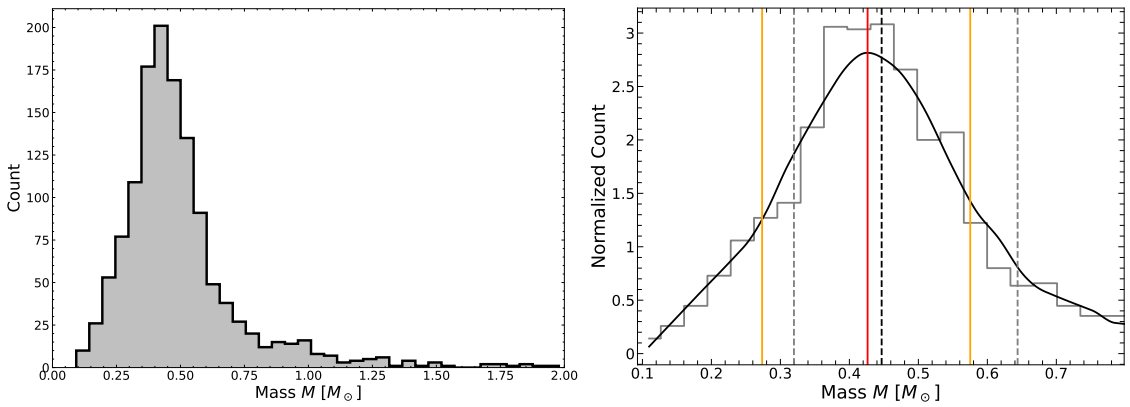


Figure 4.9: Mass distribution of the analyzed sample of hot subdwarf stars. Measurements with a poor SED or bad parallax flag are excluded. A clear peak close to the canonical mass of $0.47\,M_\odot$ for sdB stars (Heber, 2016) is visible in the left panel. The tail towards higher masses can be explained with massive eHe-sdO's (Dorsch, 2023) or non-hot subdwarf contaminants. The mode and HDI as well as median and quantiles determination (for the definitions see Appendix A) is shown in the right panel. The mode and HDI of $0.43^{+0.15}_{-0.16}\,M_\odot$ is depicted with red and yellow vertical lines, respectively, while the median and quantiles of $0.45^{+0.20}_{-0.13}\,M_\odot$ is depicted with dashed vertical lines.

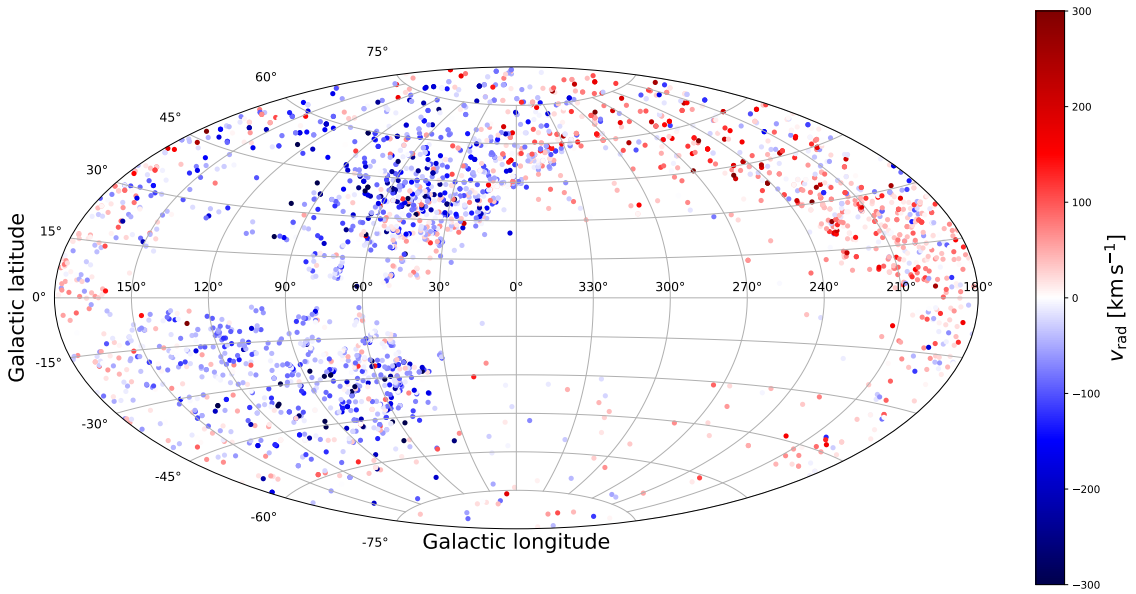


Figure 4.10: Distribution of the analyzed sample of hot subdwarf stars on the sky with their color-coded radial velocity v_{vrad} measured via spectral fits. Signs for differential Galactic rotation are clearly visible.

work in exactly the same way. With Figure 4.4 different populations of hot subdwarf stars can be distinguished and their differences could be studied in great detail leading to conclusions about their formation and further evolution. Therefore, it would be interesting whether the findings of Dorsch (2023) or other works can be reproduced. Additionally, combining the obtained data with *Gaia* proper motions would allow for a comprehensive kinematic analysis of the hot subdwarf population. However, these considerations are left as an outlook, since it is not in the scope of this thesis. All the following sections will mainly deal with the results regarding the interstellar medium and the search for circumstellar material.

4.3 Simulation of Interstellar Line Parameters

Before having a look on the actually observed interstellar line parameters, it is desirable to get a more quantitative impression of the impact of low quality spectra and the fixing of the Doppler parameter on the outcomes of the spectral ISM analysis. During the development of the script (Subsection 3.3.4) it was already realized that low spectral resolutions and low signal to noise ratios (SNRs) are a major issue for the investigation of the interstellar Ca II H&K as well as Na I D1&D2 lines. As consequence it does not appear feasible to fit the two physical parameters of the lines, the column density N and the Doppler parameter b, simultaneously. This seems to be possible only for spectra with much higher spectral resolution ($R \geq 100,000$) than available in this work ($R_{\mathrm{max}} \sim 48,000$), where the true line profile begins to dominate over the instrumental profile. Therefore, the Doppler parameter is fixed to a reasonable average value of $b = 5\,\mathrm{km\,s^{-1}}$ (Welty et al., 1994; Redfield & Linsky, 2002; Welsh et al., 2010) for the entire analysis of this work, where the possibility of line broadening through unresolved multiple ISM components is included.

To quantify the capabilities of the ISM analysis and the systematic uncertainties raised by freezing the Doppler parameter to a fixed value, synthetic spectra including both components, the photospheric spectrum and the interstellar lines, were generated for different spectral resolutions and Doppler parameters. Furthermore, Gaussian noise was added to simulate observations with varying SNR (analog to Figure 3.11). The atmospheric parameters of the stellar model spectra are $T_{\mathrm{eff}} = 30,000\,\mathrm{K}$, $\log g = 5.5$, and $\log n(\mathrm{He}) = -1.0$ in all cases. The column density of the interstellar lines is fixed to $\log N = 12.5$ for both elements, while the Doppler parameters vary

between $b = 1\,\mathrm{km\,s^{-1}}$ and $b = 10\,\mathrm{km\,s^{-1}}$ in steps of $1\,\mathrm{km\,s^{-1}}$. The choice of different SNRs and spectral resolutions can be found in the figures below, where the latter ones exactly correspond to the observed spectra of the analyzed sample. Only the $R = 100,000$ case is hypothetical. Finally, the `ism_analysis.py` script was run on all of the synthetic spectra. The derived atmospheric parameters are in perfect agreement with the true values, even in the low resolution plus low SNR case. This result underlines the robustness of the spectral fits as well as the performance of the used method (Subsection 3.3.3).

4.3.1 Simultaneous Column Density and Doppler Parameter Fits

The first concern is to prove that simultaneous column density and Doppler parameter fits just do not reliably work, especially for low resolution spectra. Hence, in a first analysis run they were both fitted freely.

The resulting column densities for Na I are shown in Figure 4.11, where the true value of $\log N = 12.5$ is indicated by a dashed horizontal line. Obviously, in the low resolution case ($R = 2,000$ and $R = 3,000$) the measurements scatter randomly around and have overwhelmingly large uncertainties. For $R = 9,800$ the mean values come on average close to the true value, but the uncertainties are still enormous, in particular for low SNR spectra. For $R = 20,000$ the mean values more or less approach the true value and the uncertainties decrease with increasing Doppler parameter. This makes sense, because the higher b, the stronger the line and the less dominant the effect of instrumental broadening becomes compared to the intrinsic width of the line. For higher resolutions ($R = 48,000$ and $R = 100,000$) the measurements seem to converge with the true value, especially for $b_{\mathrm{true}} \gtrsim 3\,\mathrm{km\,s^{-1}}$. Only the uncertainties slightly decrease for higher resolutions.

In Figure 4.12 the resulting Doppler parameters $b_{\mathrm{Na\,I}}$ of the same measurements are shown, where the true values are indicated by a dashed identity line. The same random scattering of the mean values and overwhelmingly large uncertainties as before are visible for resolutions up to including $R = 9,800$. For $R = 20,000$ they seem to be slightly more confined to the true values, but still a lot of outliers are observable and the uncertainties are large. Again, only for $R \geq 48,000$ the measurements converge with the true values and the uncertainties slightly decrease with increasing spectral resolution.

Analog plots for Ca II can be found in Section C.2 and show the same behavior. In conclusion it can be said that fitting both parameters freely is indeed not reasonably possible for the available observed spectra, in particular the low resolution ones. Even for high resolution FEROS spectra ($R = 48,000$), for which the simulated results indicate that it might work, multiple ISM components further complicate the analysis in practice. It was actually tried to fit both parameters simultaneously for these spectra in order to derive a reasonable guess for the Doppler parameter apart from the literature. However, many measurements approach the physical reasonable upper fit limit of $b = 10\,\mathrm{km\,s^{-1}}$ (Figure 4.13), especially for Ca II. Visual inspection of the fits showed that this is exactly the consequence of multiple only partly or even unresolved ISM components. Multiple component fits are in principle possible (Welsh et al., 2010), but again only for higher resolution spectra, where the different lines are fully resolved. At least the small peaks in Figure 4.13 in the range of $b = 4 - 7\,\mathrm{km\,s^{-1}}$ confirm that fixing the Doppler parameter to $b = 5\,\mathrm{km\,s^{-1}}$ seems to be the best choice that can be made. However, this raises a systematic uncertainty on the measurements, which is investigated next.

4.3.2 Fixed Doppler Parameter Fits

The question is, how fixing the Doppler parameter to a certain value influences the resulting column densities. Figure 4.14 shows an analog plot as Figure 4.12 for the derived column densities of Na I, but this time with a fixed Doppler parameter of $b_{\mathrm{fixed}} = 5\,\mathrm{km\,s^{-1}}$. The true column density of $\log N_{\mathrm{Na\,I}}$ is indicated by a dashed horizontal line, while the observations, where the true and fixed Doppler parameters match, are marked with a dotted vertical line.

First of all, it can be seen that those measurements lying on the dotted vertical line perfectly agree with the true value in all cases, except of one outlier at $R = 3,000$ and low SNR. However, the

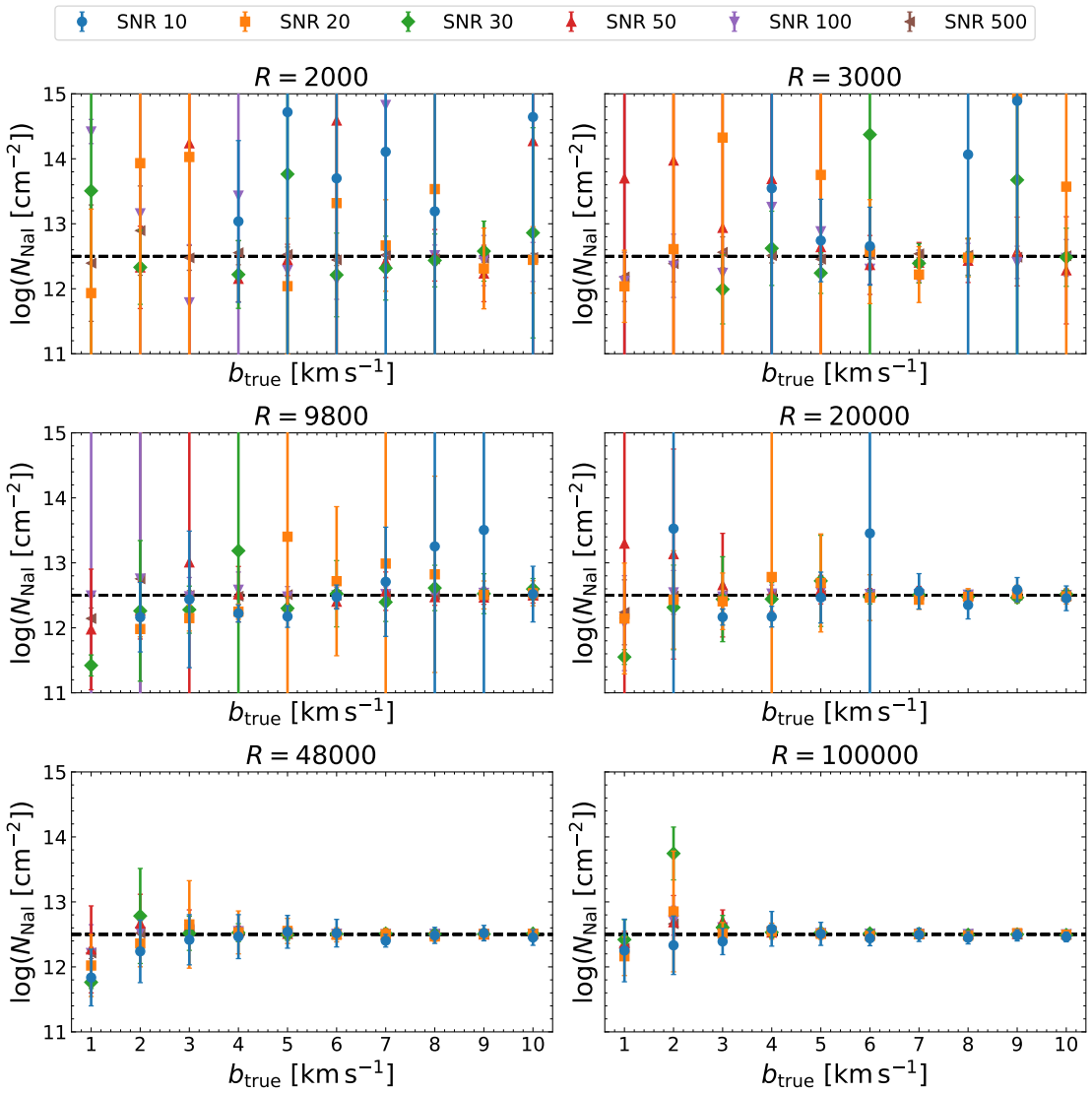


Figure 4.11: Resulting column densities for Na I after simultaneously freely fitting it and the Doppler parameter b. The true value of $\log N = 12.5$ is indicated by a dashed horizontal line, while the true Doppler parameters $b_{\rm true}$ are given on the x-axis. Each subplot corresponds to a different spectral resolution R as stated in each title and different signal to noise ratios (SNRs) are coded with different colors and symbols. Missing data points (e.g., in the low resolution and low SNR cases) mean that no lines have been detected by the program. This happens, if the line is swallowed by noise. The measurement for $R = 100,000$, $b_{\rm true} = 2\,{\rm km\,s^{-1}}$, and a SNR of 30 seems to be an outlier. In conclusion from the scattering and the large uncertainties, simultaneous free fits of $\log N$ and b do not seem to be reasonable for the real data, except perhaps for the $R \sim 48,000$ FEROS spectra.

trend is clearly visible that for overestimated Doppler parameters the measured column densities are strongly underestimated, while for underestimated Doppler parameters the measured column density is only slightly overestimated. This trend can be explained with the correlation between column densities and Doppler parameters. In principle, both parameters increase the equivalent width of a spectral line, the column density mainly by increasing the depth and the Doppler parameter by mainly increasing the width. The Voigt profile fits aim to reproduce the equivalent width of each line. But if, for example, the width cannot be reproduced, because it is fixed to a wrong value, the fitting routine tries to counteract this issue by increasing or decreasing the column density. This correlation is also shown in Figure 4.15, where the ISM line fits are shown for three different true Doppler parameters, but identical fixed ones. In the upper panel the width (b) is

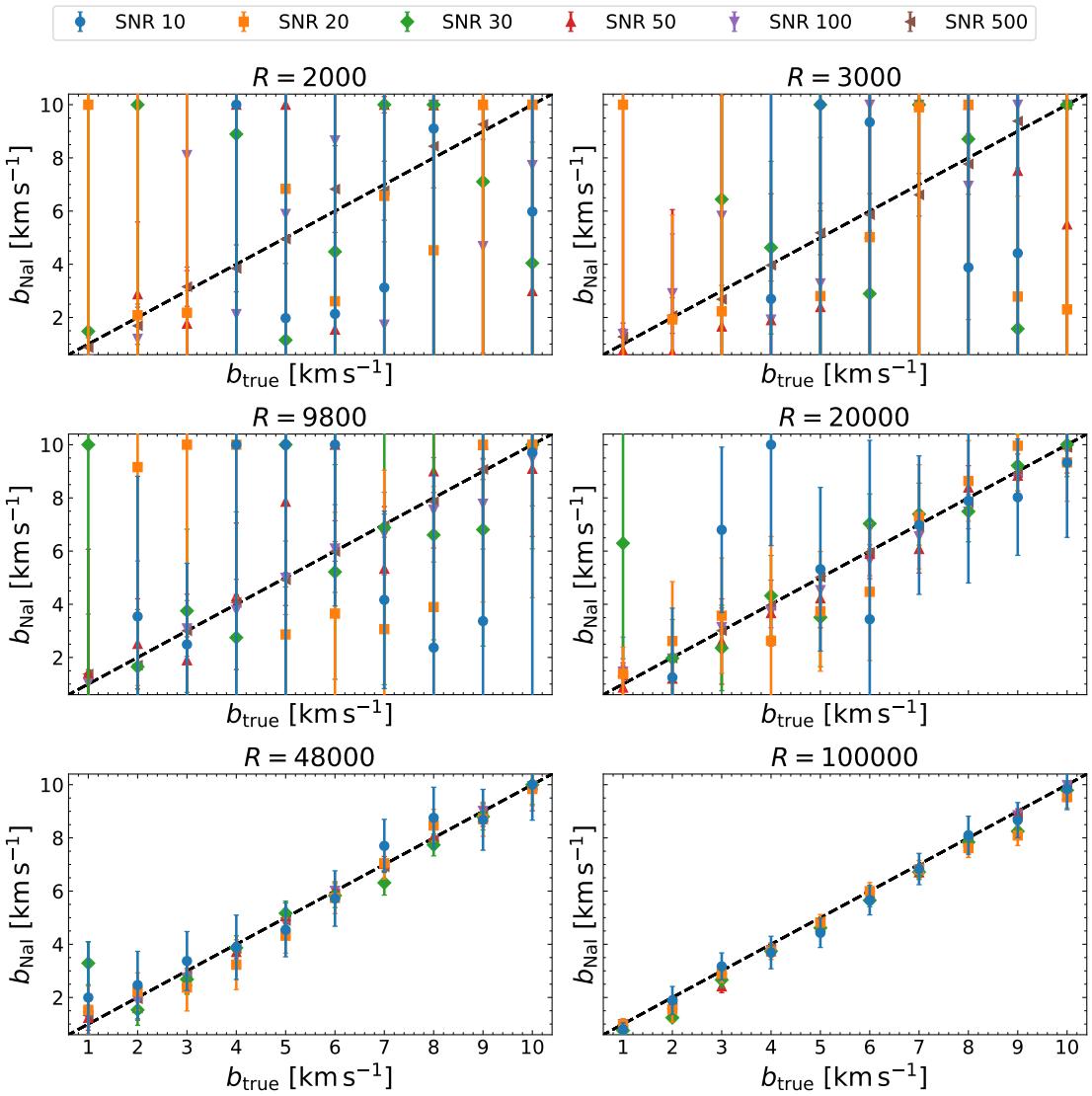


Figure 4.12: Resulting Doppler parameters for Na I after simultaneously freely fitting it and the column density. The true values are indicated by a dashed identity line. Each subplot corresponds to a different spectral resolution R as stated in each title and different signal to noise ratios (SNRs) are coded with different colors and symbols. Missing data points (e.g., in the low resolution and low SNR cases) mean that no lines have been detected by the program. This happens, if the line is swallowed by noise. The scattering and uncertainties are even higher as for Figure 4.11, which leads to the same conclusion as there.

overestimated and hence, the fit does not account for the full depth ($\log N$) of the line. On the other hand, in the lower panel the width is underestimated and hence, the depth is overestimated. In the mid panel the fixed and true value matches and so does the fit. Overall it can be said that fixing the Doppler parameter raises a systematic uncertainty on the derived column densities. When dealing with a small number of observation, in the best case higher resolution ones, one could try to manually adjust the Doppler parameter until the line profile matches. However, this can also be quite uncertain and is not possible at all, when automatically fitting a large sample of observations as in this work. Therefore, it is reasonable to agree on a generic value for the systematic uncertainty that is added on top of the statistical one. According to Figure 4.14, $\Delta \log N_{\mathrm{systematic}} = 0.5$ seems to be acceptable and is used for all observed column densities later on in this work. Then it is roughly accounted for the deviation at larger true Doppler parameters and also for smaller ones, which are not too far from the fixed value. Really small values of $b \sim 1\,\mathrm{km\,s^{-1}}$, where the deviation is larger, are actually not often expected for the observed sample, because most of the objects are

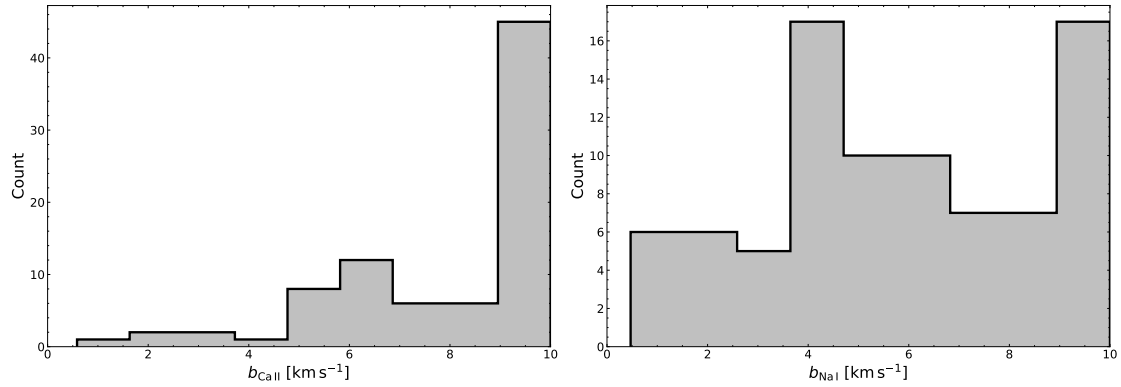


Figure 4.13: Doppler parameter distribution for both elements after freely fitting it using observed high resolution FEROS spectra. Many measurements approach the upper fit limit of $b = 10\,\mathrm{km\,s^{-1}}$ because of multiple ISM component blends.

fairly distant. Thus, multiple unresolved ISM components broaden the observed interstellar lines anyway (Figure 4.13). This systematic uncertainty and the pure statistical one add up in quadrature: $\Delta \log N_{\mathrm{total}} = \sqrt{\Delta \log N_{\mathrm{stat}}^2 + \Delta \log N_{\mathrm{sys}}{}^2}$.

Another concern is the signal to noise ratio, especially for low resolution spectra. From Figure 4.14 it can be clearly seen that the trend observed for high resolution spectra is more and more smeared out in the low resolution ones. This is not only the consequence of the lower resolution, but also of low SNRs. Even for the higher resolved spectra the low SNR measurements deviate from the trend at high Doppler parameters. This is even more critical in the low resolution cases. The derived column densities in the low SNR cases scatter much more heavily and have much larger uncertainties. Some even approach the upper fit limit of $\log N = 15$. Therefore, it appears reasonable to set even stricter SNR cuts for low resolution spectra than for higher resolved ones. The choices adopted later on for the real observations are SNR ≥ 40 for LAMOST spectra ($R \sim 1,800$), SNR ≥ 35 for SDSS spectra ($R \sim 2,000$), SNR ≥ 25 for DESI spectra ($R \sim 3,000$), and SNR ≥ 15 for X-Shooter spectra ($R \sim 9,800$). For all other ones the cut is set to SNR ≥ 10.

4.3.3 Simulated Equivalent Widths

Besides column densities also equivalent widths are measured for each line as proxy for the overall line strength. These are investigated for the simulated synthetic spectra as well. In Figure 4.16 for example, the derived equivalent widths for the Na I D2 line are shown, which is on average the strongest of the four interstellar lines. Analog plots for the other three lines can be found in Section C.2.

Obviously, the measured equivalent widths increase with increasing Doppler parameter as expected. But this increase does not seem to be linear. It is stronger at lower b values and drops off at higher ones. From a theoretical point of view, the dependence of the equivalent width of a spectral absorption line on the number of absorbers (column density) and other physical parameters like temperature and microturbulence (Doppler parameter) is described by the "curve of growth" (e.g., Draine, 2011; Gray, 2021; Bartelmann, 2021). This curve of growth in general exhibits three distinct regimes. For low column densities N, the equivalent width increases linearly with N ($W_\lambda \propto N$) and is independent of the Doppler parameter b. At moderate N, the Gaussian Doppler broadened line core begins to saturate and adding more absorbers just broadens the line rather than deepening it. In this regime b matters most, because the dependency is approximately

$$W_\lambda \propto b\sqrt{\ln N/b}\,. \tag{4.3}$$

Eventually for high column densities N, the growth is dominated by Lorentzian tails of the line profile and the equivalent width scales with $W_\lambda \propto \sqrt{N}$, again independent of b. Interstellar

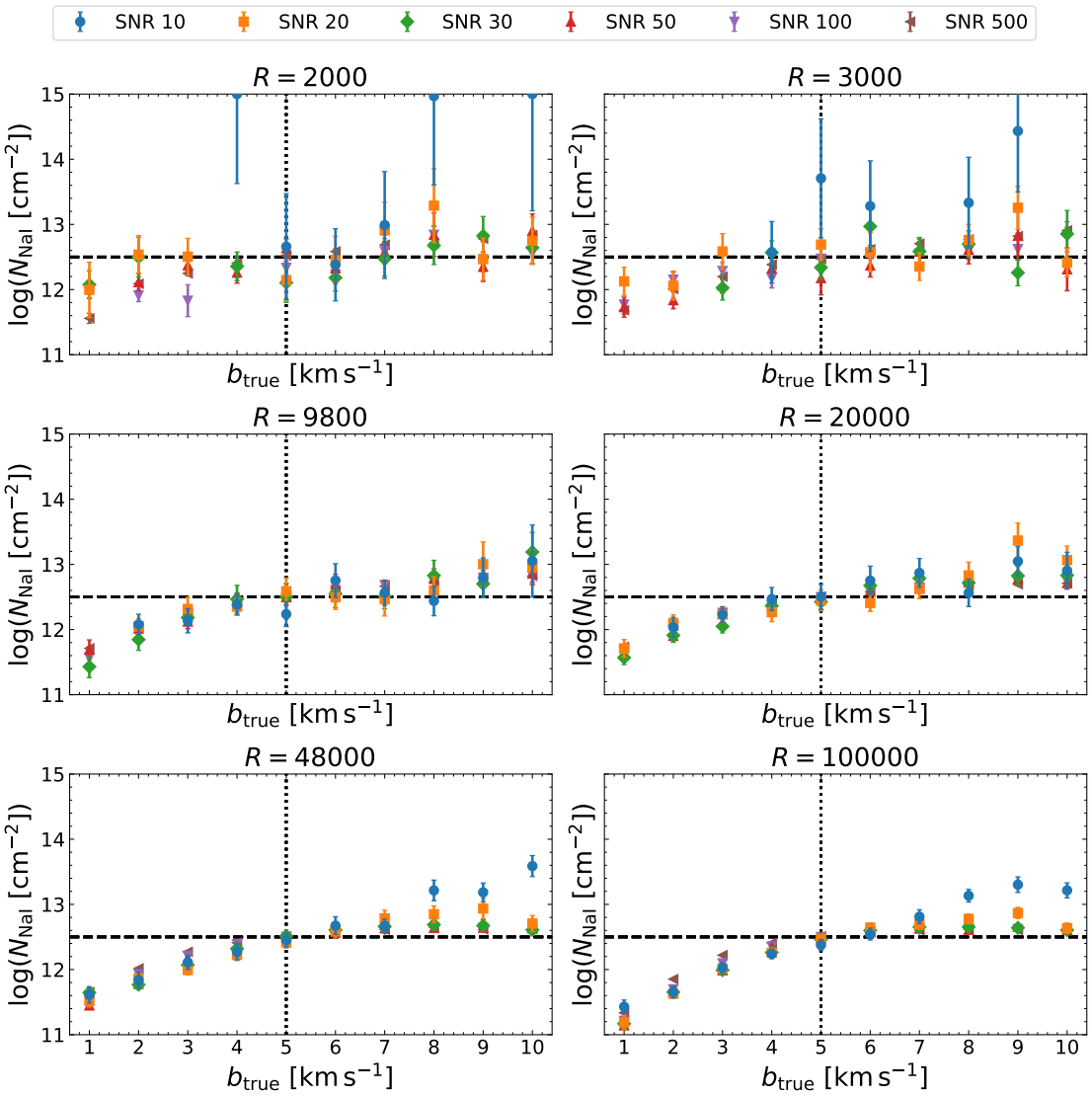


Figure 4.14: Resulting column densities for Na I after fitting it with a fixed Doppler parameter of $b_{\mathrm{fixed}} = 5\,\mathrm{km\,s^{-1}}$. The true value of $\log N = 12.5$ is indicated by a dashed horizontal line, while the observations, where the fixed and true Doppler parameters match, are marked by a dotted vertical line. Each subplot corresponds to a different spectral resolution R as stated in each title and different signal to noise ratios (SNRs) are coded with different colors and symbols. Missing data points (e.g., in the low resolution and low SNR cases) mean that no lines have been detected by the program. This happens, if the line is swallowed by noise. The uncertainties on the column densities decrease with increasing R and SNR, but are still on average about $\Delta \log N \sim 0.5$ or even higher for the lowest R values. The systematic offset between measured and true column densities is larger for lower Doppler parameters than for higher ones. Overall, a systematic uncertainty due to the fixing of the Doppler parameter of about $\Delta \log N_{\mathrm{sys}} \sim 0.5$ appears reasonable, as very low b values are barely observed. A analog plot for the column densities of Ca II can be found in Section C.2. It shows exactly the same behavior.

conditions and column densities of roughly $N = 10^{11} - 10^{15}\,\mathrm{cm^{-2}}$ for Ca II and Na I are typical for the second regime, where Equation 4.3 is valid. This dependency is perfectly reflected by Figure 4.16: A linear rise of $W_{\mathrm{Na\,I\,D2}}$ at small b and progressive flattening as b grows. This behavior is actually also transferred to the column density measurement of Figure 4.14. Since the Voigt profile fits try to reproduce the equivalent width of a line, which depends on both, the column density N and the Doppler parameter b. Fixing b to a certain value leads to the systematic deviation of $\log N$ from the true value according to Equation 4.3 and observed in Figure 4.14.

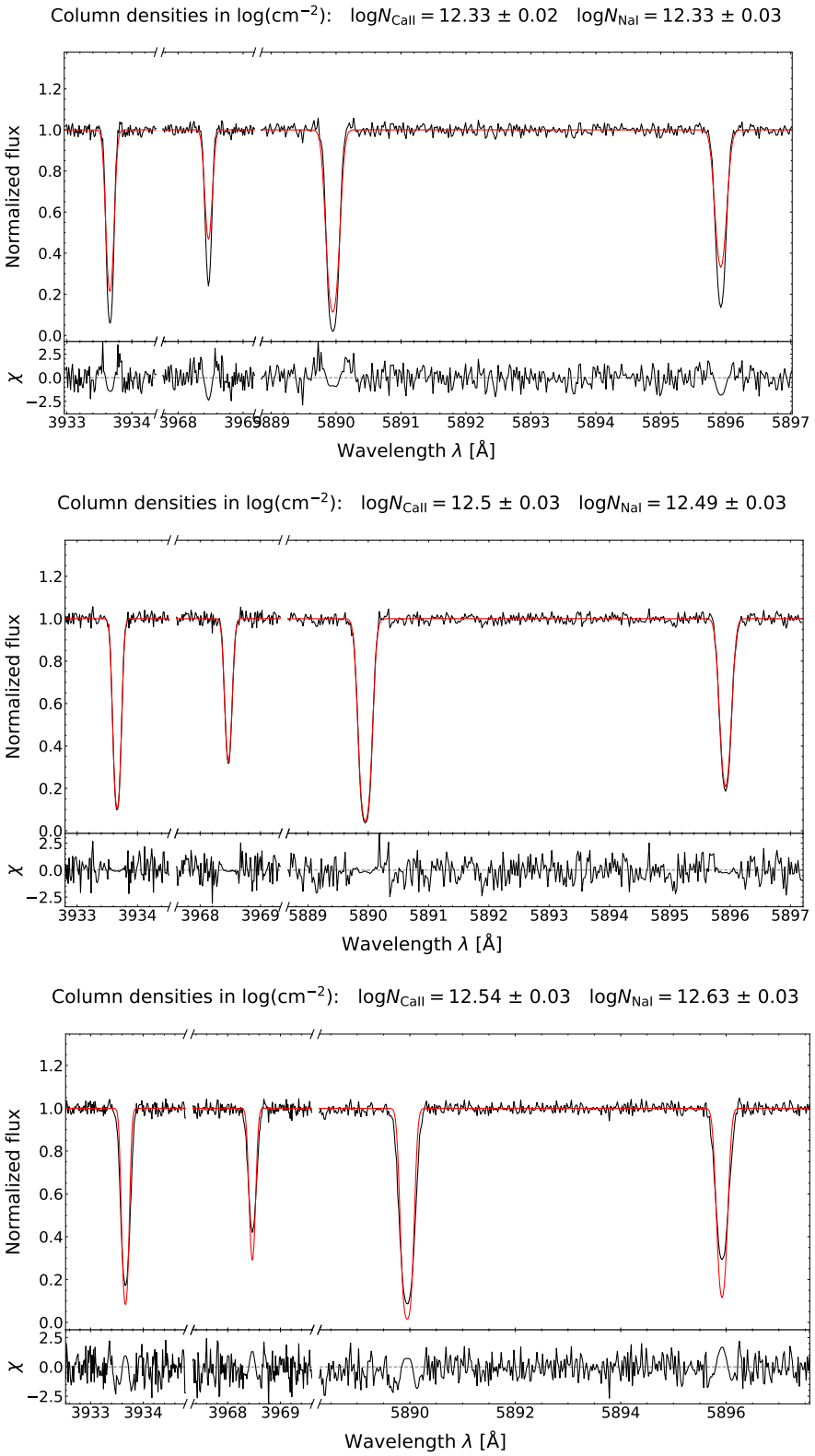


Figure 4.15: Three Voigt profile fits to synthetic ISM lines ($R = 100,000$, SNR $= 50$, $\log N_{\mathrm{true}} = 12.5$, $b_{\mathrm{fixed}} = 5\,\mathrm{km\,s^{-1}}$) showcasing the bias from fixing the Doppler parameter to a certain value. In the upper panel the Doppler parameter is overestimated ($b_{\mathrm{true}} = 4\,\mathrm{km\,s^{-1}}$) leading to underestimated column densities and fits that cannot account for the full depth of the lines. In the mid panel the fitted and true values perfectly match and the fit reproduces the lines well. In the lower panel the Doppler parameter is overestimated ($b_{\mathrm{true}} = 7\,\mathrm{km\,s^{-1}}$) leading to slightly overestimated column densities and fits that become deeper than the lines.

Finally, the equivalent widths measurements are also affected by the SNR, perhaps even more than the column density measurements (Figure 4.16). While the trend of Equation 4.3 is clearly visible for higher resolved spectra and low SNR is only affecting the uncertainties, for low resolution spectra it is again smeared out. The low SNR observations scatter a lot and have huge uncertainties of up to more than $\Delta W \sim 0.1\,\text{Å}$. Therefore, in order to measure accurate equivalent widths, this is another reason to apply strict SNR cuts for low resolution spectra.

4.3.4 Interstellar Radial Velocities

Another important parameter that is measured during the ISM analysis is the radial velocity of the interstellar lines. In Figure 4.17 the resulting $v_{\text{rad,ISM}}$ distributions are shown for different spectral resolutions and SNRs as measured from the synthetic spectra. The true radial velocity of $v_{\text{rad,ISM}} = 0\,\text{km}\,\text{s}^{-1}$ is indicated with a dashed vertical line. Obviously, the accuracy of the measurements increases drastically from low to high resolution. While it scatters up to more than $\pm 40\,\text{km}\,\text{s}^{-1}$ around the true value for a resolution of $R \leq 3,000$, it always hits the target for higher resolutions independent of SNR. To get a decent accuracy in the low resolution case, extremely high SNRs are needed (SNR ~ 500), which are in practice often unrealistic.

Also, when looking at the uncertainties on the radial velocity (Figure 4.18), it can be seen that it is always well constrained for higher resolutions ($\Delta v_{\text{rad,ISM}} \lesssim 0.5\,\text{km}\,\text{s}^{-1}$), even at low SNR. However, even though the fitting method combines the four individual lines to one radial velocity fit, which increases the goodness of the measurement, it is not constrained at all for low resolution spectra ($R < 3,000$). The uncertainties range up to $\Delta v_{\text{rad,ISM}} = 40\,\text{km}\,\text{s}^{-1}$ and even for super high SNR it only drops to about $\Delta v_{\text{rad,ISM}} \sim 10\,\text{km}\,\text{s}^{-1}$. As comparison, for the highest resolutions $\Delta v_{\text{rad,ISM}}$ can reach values of less than $0.1\,\text{km}\,\text{s}^{-1}$. For all intermediate resolutions $\Delta v_{\text{rad,ISM}}$ is somewhere in between, but is already quite low for $R = 9,800$ ($\Delta v_{\text{rad,ISM}} \lesssim 5\,\text{km}\,\text{s}^{-1}$). Compared to the typical uncertainties of the stellar radial velocities of about $\Delta v_{\text{rad}} \sim 1\,\text{km}\,\text{s}^{-1}$, this is still rather high. Later on, the difference between stellar and ISM radial velocity will play an important role in the search for circumstellar material and high precision is needed. The results here already indicate, that for this purpose only those spectra with $R \geq 20,000$ should be used.

In general, it again appears as a good choice to reject spectra that have both, low resolution and low SNR for the whole analysis of this work. Furthermore, a lot of caution is needed when drawing conclusions from radial velocities derived from low resolution spectra. It is generally not possible to obtain accurate and well constrained values. If this is required though, one necessarily has to rely on high resolution spectra.

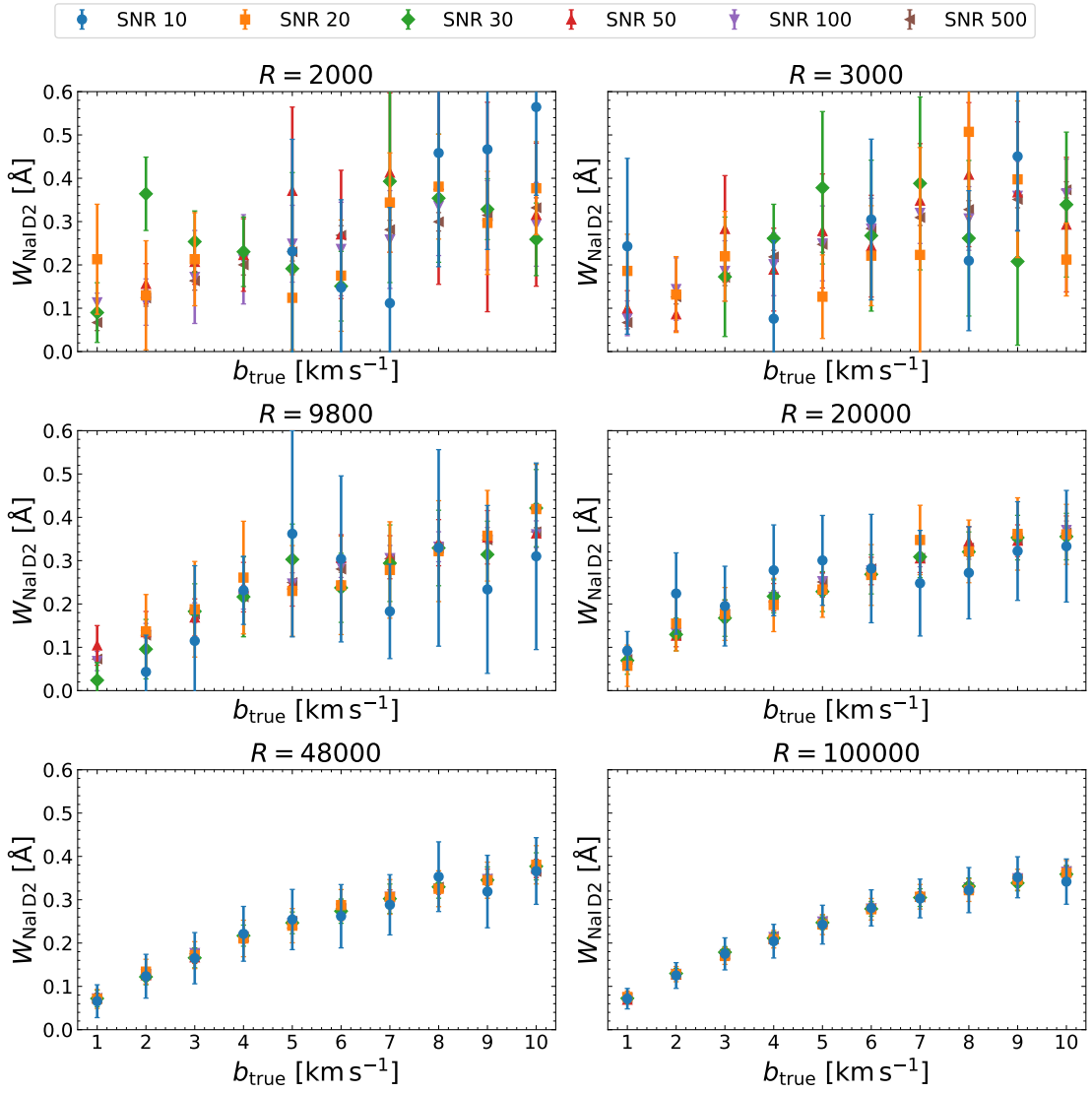


Figure 4.16: Resulting equivalent widths W for the Na I D2 line. Each subplot corresponds to a different spectral resolution R as stated in each title and different signal to noise ratios (SNRs) are coded with different colors and symbols. Missing data points (e.g., in the low resolution and low SNR cases) mean that no lines have been detected by the program. This happens, if the line is swallowed by noise. The equivalent widths increase with increasing Doppler parameter as expected, but this increase is stronger at lower b values. For higher ones the additional increase seems to drop off according to Equation 4.3. For lower resolutions the uncertainties become very large (up to more than $\Delta W \sim 0.1\,\text{Å}$) and even in the high resolution case they can be quite high (about $\Delta W \sim 0.05\,\text{Å}$) for low SNR. Analog plots for the equivalent widths of the other interstellar lines can be found in Section C.2.

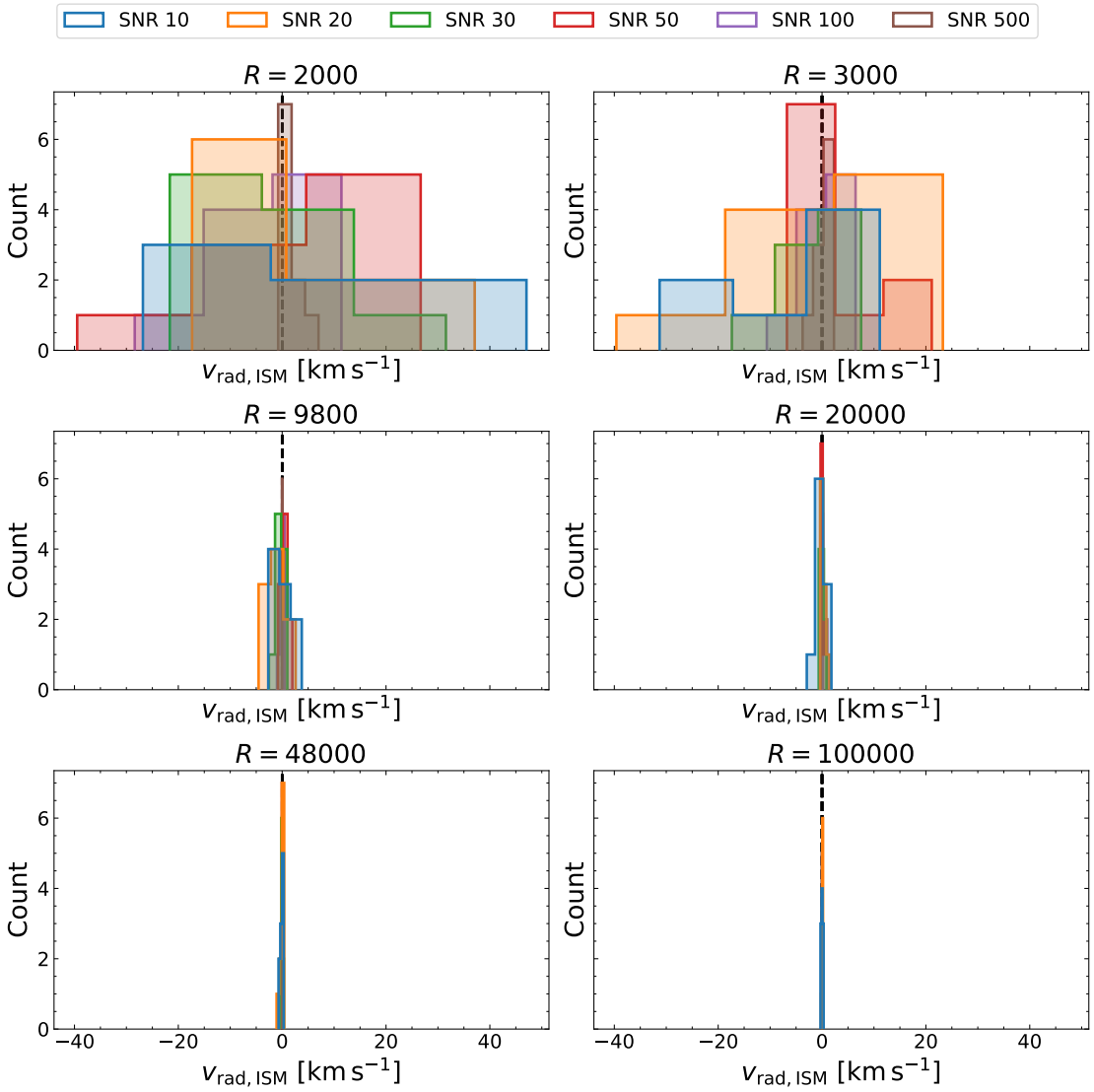


Figure 4.17: Resulting radial velocity distributions from the four interstellar lines. Each subplot corresponds to a different spectral resolution R as stated in each title and different signal to noise ratios (SNRs) are coded with different colors. The true radial velocity of $v_{\mathrm{rad,ISM}} = 0\,\mathrm{km\,s^{-1}}$ is indicated by a dashed vertical line. Each histogram contains $\sqrt{N_{\mathrm{data}}}$ bins within the observed parameter range, where N_{data} is the number of measurements. For $R \leq 3{,}000$ the values scatter drastically around the true value even for higher SNR ($\pm 40\,\mathrm{km\,s^{-1}}$). For $R = 9{,}800$ this scattering is already quite low (about $\pm 5\,\mathrm{km\,s^{-1}}$) and for higher R it is no longer visible.

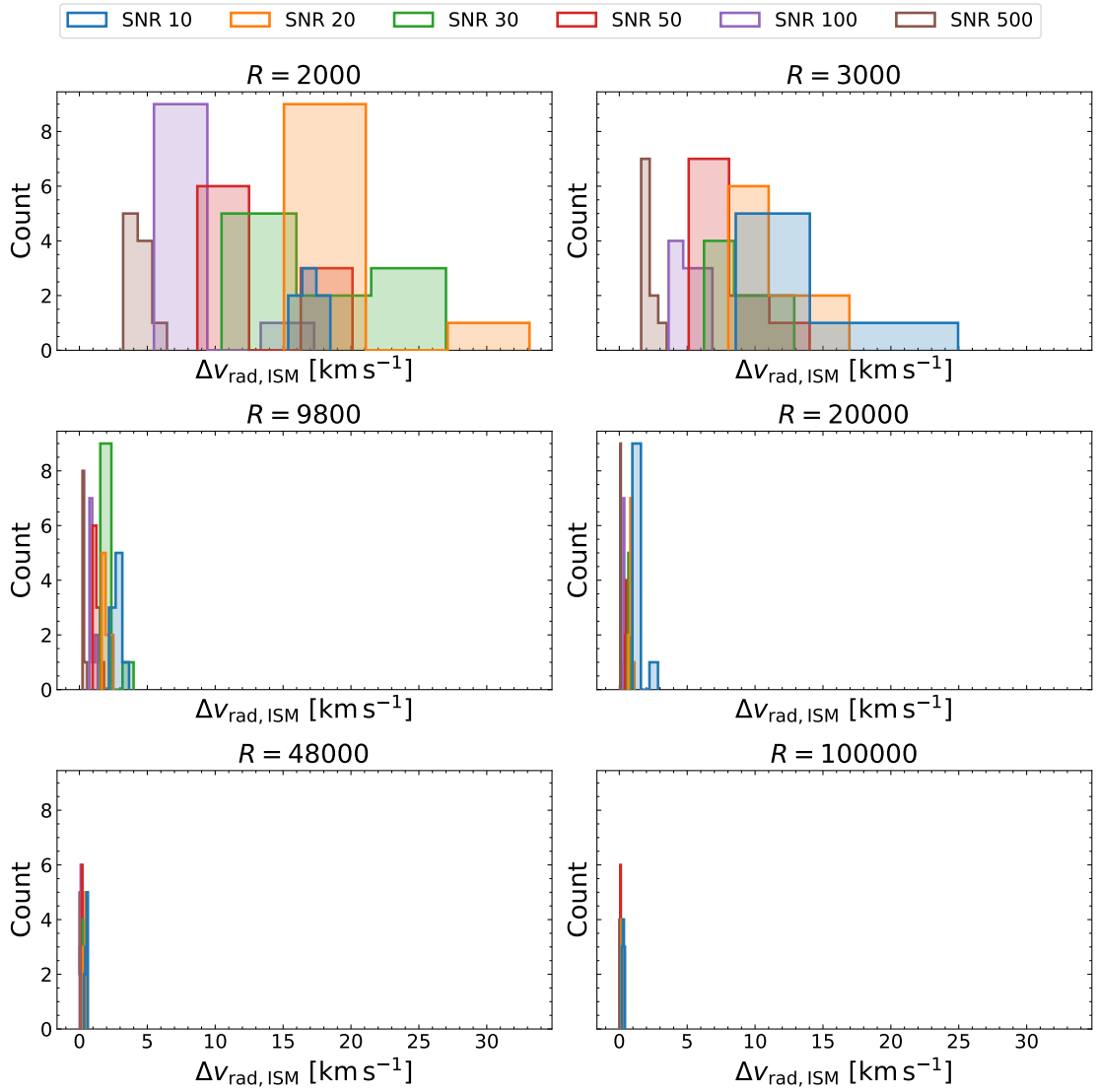


Figure 4.18: Resulting radial velocity uncertainty distributions from the four interstellar lines. Each subplot corresponds to a different spectral resolution R as stated in each title and different signal to noise ratios (SNRs) are coded with different colors. Each histogram contains $\sqrt{N_{\text{data}}}$ bins within the observed parameter range, where N_{data} is the number of measurements. In the low resolution case $R \leq 3{,}000$ the radial velocities are highly unconstrained with up to $\Delta v_{\text{rad,ISM}} \sim 30\,\text{km}\,\text{s}^{-1}$ or even higher, except for extremely high SNR. From $R = 9{,}800$ on ($\Delta v_{\text{rad,ISM}} \lesssim 5\,\text{km}\,\text{s}^{-1}$) the uncertainty decreases drastically even for low SNR. In conclusion, if high precision on the radial velocity is required, one should stick to $R \geq 20{,}000$ spectra.

Chapter 5

Investigation of the Interstellar Medium

By running the `ism_analysis.py` script (Subsection 3.3.4) on numerous stellar spectra of hot subdwarf stars, various parameters of the interstellar Ca II H&K and Na I D1&D2 lines present in those spectra have been obtained, describing the properties of the interstellar gas they originate from. Subsequently, the atmospheric parameters of the hot subdwarf stars, derived as part of the ISM line analysis, are passed to the SED fitting routine (Section 3.4) in order to obtain the reddening parameter, which represents a proxy for the amount of interstellar dust along the line of sight.

But before looking at the outcomes, quality cuts have to be introduced to reject unreliable or even spurious results. First, as extensively discussed in Section 4.3, spectra with a low signal to noise ratio (SNR) have to be rejected. Since in high resolution spectra much more spectral information is contained than in low resolution ones, even if the SNR is low, individual SNR cuts are applied for the different surveys/spectrographs depending on the spectral resolution. The concrete choices are SNR $\geq$40 for LAMOST, SNR $\geq$35 for SDSS, SNR $\geq$25 for DESI, SNR $\geq$15 for X-Shooter, and SNR $\geq$10 for SPY/UVES as well as FEROS. These cuts are applied on all outcomes of the ISM line analysis (equivalent widths, column densities, radial velocities, etc.). However, if those results are correlated with parameters involving the distance to the stars, another quality cut based on the *Gaia* parallax has to be introduced. In particular for distant stars, the measured parallax gets closer and closer to zero[1] and highly uncertain. Therefore, if parallaxes are involved, additionally objects for which the measured parallax uncertainty is larger than 20% are rejected. Table 5.1 gives an overview about the number of objects with and without quality cuts. Earlier in the analysis pipeline (Section 4.1) binary systems detected in the SED as well as spectra with SNR $\leq$ 10 were already rejected. Hence, N_0 is the number of objects before applying the new quality cuts and is identical to N_{single} in Table 4.1. N_{SNR} is the number of objects after the new SNR cuts, from which only the low resolution surveys with stricter cuts than SNR $\geq$ 10 are affected. $N_{\mathrm{SNR}+\varpi}$ is the number of objects left after applying both, the SNR and bad parallax cut. Again, predominantly the low resolution spectra are affected by the additional cut. The reason for this is that high resolution spectrographs in general need a higher light throughput than low resolution ones. Therefore, the stars observed by those are brighter and closer and thus, have a better constrained parallax. Low resolution surveys can observe much fainter and more distant stars (also see Figure C.4), which is why they are affected more. Obviously, the number of SDSS objects is reduced the most, because it observes the faintest objects compared to the other surveys (see Section 2.2). Furthermore, Table 5.1 lists the number of detected interstellar line doublets $N_{\mathrm{Ca\,II,Na\,I}}$ out of N_{SNR} for each species. Due to the strict SNR cuts, most of the spectra show interstellar lines. Only for a small fraction of objects no interstellar lines were visible. In total, the spectra of 87% of all objects showed Ca II H&K lines, while for Na I D1&D2 lines the fraction even increases to 92%. This result alone confirms the ubiquity of interstellar material along all directions, even at higher Galactic latitudes.

In Section 5.1 a first look at the different obtained parameters is taken and their distributions

[1] In fact, the measured *Gaia* parallaxes can even become negative (see Figure C.4), which is unphysical, but possible due to the data processing procedure (Lindegren et al., 2021).

Table 5.1: Number of objects used for the investigation of the interstellar medium per survey/spectrograph and in total. N_0 is the number of objects without any quality cuts, N_{SNR} the number of objects after rejecting spectra with low signal to noise ratio, and $N_{\mathrm{SNR}+\varpi}$ the number of objects after applying the SNR cut combined with a bad parallax cut. Furthermore, the number of detected interstellar line doublets $N_{\mathrm{CaII,NaI}}$ out of N_{SNR} is listed for each element. The sum of objects per survey is higher than the total number of objects in each row, since some stars have been observed across different surveys.

	SDSS	LAMOST	DESI	FEROS	SPY/UVES	X-Shooter	In total
N_0	934	1,231	485	88	80	16	2,349
N_{SNR}	466	586	181	88	80	16	1,237
$N_{\mathrm{SNR}+\varpi}$	263	567	147	88	77	16	996
N_{CaII}	418	478	131	84	79	16	1,075
N_{NaI}	416	539	158	85	69	16	1,136

are statistically evaluated. Then in Section 5.2 the dependencies and correlations between those parameters are investigated, before it is looked at the distribution of interstellar gas and dust on the sky in form of Galactic sky maps in Section 5.3. In Section 5.4 the variation of the observed parameters with distance, direction, and height above and below the Galactic plane is studied in more detail. Finally, in Section 5.5 the connection between interstellar gas and dust is made by calculating and analyzing gas to dust ratios and Section 5.6 provides a brief comparison between measured and predicted reddening values from dust maps.

5.1 Observed Interstellar Line Parameters and Reddening

In this subsection the first results regarding the observed parameters of the interstellar medium are presented. First, in Figure 5.1 and Figure 5.2 the distribution of the measured column densities is shown before and after applying the new SNR cuts, respectively. Obviously, the SNR cuts are justified, since they considerably reduce the contribution from unreasonably high and low column densities, which approach the fit limits (the borders of the histograms). Additionally, the distributions get more confined (narrower), which happens if measurements with high uncertainties and mean values off from the true value are removed. It is also clearly visible from the right panels of Figure 5.1 that all the unrealistically high column densities exclusively come from low resolution spectra as expected from the discussions in Section 4.3. In Figure 5.2 leftover peaks at $\log N = 15$ are still visible. They either come from remaining spurious measurements or they are actually true. Most of those high column densities come from LAMOST spectra, but LAMOST comprises the LAMOST Spectroscopic Survey of the Galactic Anticentre (LSS-GAC, Yuan et al., 2015), which is a survey dedicated to investigate the structure and the history of the Galactic thin and thick disk. To this end, a large number of stellar spectra were recorded at low Galactic latitudes (the Galactic disk) in the direction of the Galactic anticentre. This footprint can be clearly seen in Figure 2.3 for $|b| \leq 30^\circ$ and $150^\circ < l < 210^\circ$. Because of the dense interstellar material within the Galactic disk, for those stars extremely high column densities are indeed expected.

Moreover, a deviation between column densities from high and low resolution spectra is observed: Those from higher resolution spectra (FEROS, SPY/UVES, X-Shooter) are systematically lower than those from low resolution spectra (SDSS, LAMOST, DESI). This is actually not expected to be a discrepancy in the analysis routine, but reflects the greater distances towards fainter stars observed with low resolution spectrographs (Figure C.4), which do not require a light throughput as high as often needed for high resolution ones. The further a star is away from us, the more intervening interstellar material is possible, resulting into higher column densities. In addition, the test of the analysis methods on simulated spectra (Section 4.3) did not indicate any clear systematics between the measured column densities from high and low resolution spectra.

Furthermore, Figure 5.2 shows the different detection limits of weak ISM lines for each survey. In general, the higher the spectral resolution the weaker lines can be detected, but also SNR plays a role. This is also a reason why overall lower column densities for FEROS, SPY/UVES, and X-Shooter spectra are measured, whereas the distributions for SDSS, LAMOST, and DESI are already cut off at this point. However, this also means that the column density distributions might be affected by an observational bias. While it looks like they are closely confined and show a distinct peak, perhaps measurements of lower column densities are just missing, because they fall below the detection limit. Anyway, in the following the distributions are statistically examined further. For both elements the mode and highest density interval (HDI) as well as the median and quantiles (see Appendix A) were determined as shown in Figure 5.3. For Ca II the mode and the median as well as the confidence intervals are identical and the resulting parameter space is

$$\log \bar{N}_{\mathrm{Ca\,II,mode/median}} = 12.7^{+0.6}_{-0.5}\,. \tag{5.1}$$

For Na I the results are slightly different with

$$\log \bar{N}_{\mathrm{Na\,I,mode}} = 12.3^{+0.7}_{-0.6} \qquad \text{and} \qquad \log \bar{N}_{\mathrm{Na\,I,median}} = 12.4^{+0.9}_{-0.5}\,. \tag{5.2}$$

Since the distributions are clearly unimodal, the mode and HDI might be the more appropriate choice. Assuming that the observational bias is negligible, these results can be interpreted as the average column densities of Ca II and Na I with corresponding 1σ dispersions in a volume of about 4 kpc (Figure C.4) around us, (mostly) excluding sight lines within the Galactic disk. In previous analyses with high resolution spectra ($R \gtrsim 100,000$) the column densities are mostly lower than these average values (Welty et al., 1994; Redfield & Linsky, 2002; Welsh et al., 2010). However, they observed much closer objects and hence, the higher column densities are not surprising, assuming that more interstellar material is accumulated along the line of sight.

The distributions of the equivalent widths of the four interstellar lines are shown in Figure 5.4. For all of them the possible range seems to be within 0.6 Å, which is consistent with, e.g., Megier et al. (2005) and Welsh et al. (2010). Those distributions can also be statistically evaluated (Figure D.1), even though they show multiple smaller peaks and have zero as physical lower border, which makes them skewed. The results are as follows:

$$\bar{W}_{\mathrm{CaII\,K,mode}} = 0.21^{+0.09}_{-0.14}\,\text{Å} \qquad \text{and} \qquad \bar{W}_{\mathrm{CaII\,K,median}} = 0.22^{+0.14}_{-0.11}\,\text{Å} \tag{5.3}$$

$$\bar{W}_{\mathrm{CaII\,H,mode}} = 0.11^{+0.10}_{-0.07}\,\text{Å} \qquad \text{and} \qquad \bar{W}_{\mathrm{CaII\,H,median}} = 0.14^{+0.12}_{-0.08}\,\text{Å} \tag{5.4}$$

$$\bar{W}_{\mathrm{NaI\,D2,mode}} = 0.17^{+0.20}_{-0.11}\,\text{Å} \qquad \text{and} \qquad \bar{W}_{\mathrm{NaI\,D2,median}} = 0.25^{+0.20}_{-0.15}\,\text{Å} \tag{5.5}$$

$$\bar{W}_{\mathrm{NaI\,D1,mode}} = 0.14^{+0.14}_{-0.11}\,\text{Å} \qquad \text{and} \qquad \bar{W}_{\mathrm{NaI\,D1,median}} = 0.19^{+0.17}_{-0.11}\,\text{Å} \tag{5.6}$$

These values can again be interpreted as the average values and corresponding dispersions of the possible parameter space for the interstellar environment of the Milky Way. However, while column densities directly give physical conditions, equivalent widths only represent the overall line strengths and no direct conclusions about the properties of the interstellar material can be deduced from them alone. Therefore, for the further analysis of this work column densities are preferred. However, as expected from visual inspections of the spectra and the corresponding atomic data of the lines, the Ca II H line is on average the weakest one, followed by the Na I D1 line. Whether the Ca II K or the Na I D2 line is the strongest one depends on the choice of the mode or median. Since the distributions are a bit oddly shaped, the median might be the more appropriate choice. Furthermore, the distributions for Na I are slightly broader than those for Ca II, which is also reflected in the HDI and the quantiles. Since the column density distributions seem to be similarly broad for both elements, the physical reason could be that sodium is lighter (atomic number of $Z = 11$) compared to calcium ($Z = 20$). This allows for higher thermal velocities, leading to increased Doppler broadening and equivalent widths.

One more parameter measured in the ISM line analysis is the radial velocity. Its distribution is shown in Figure 5.5. It clearly peaks around zero and is much narrower than the distribution of stellar radial velocities (Figure C.2). This is expected, because the ISM is neither thought to prefer

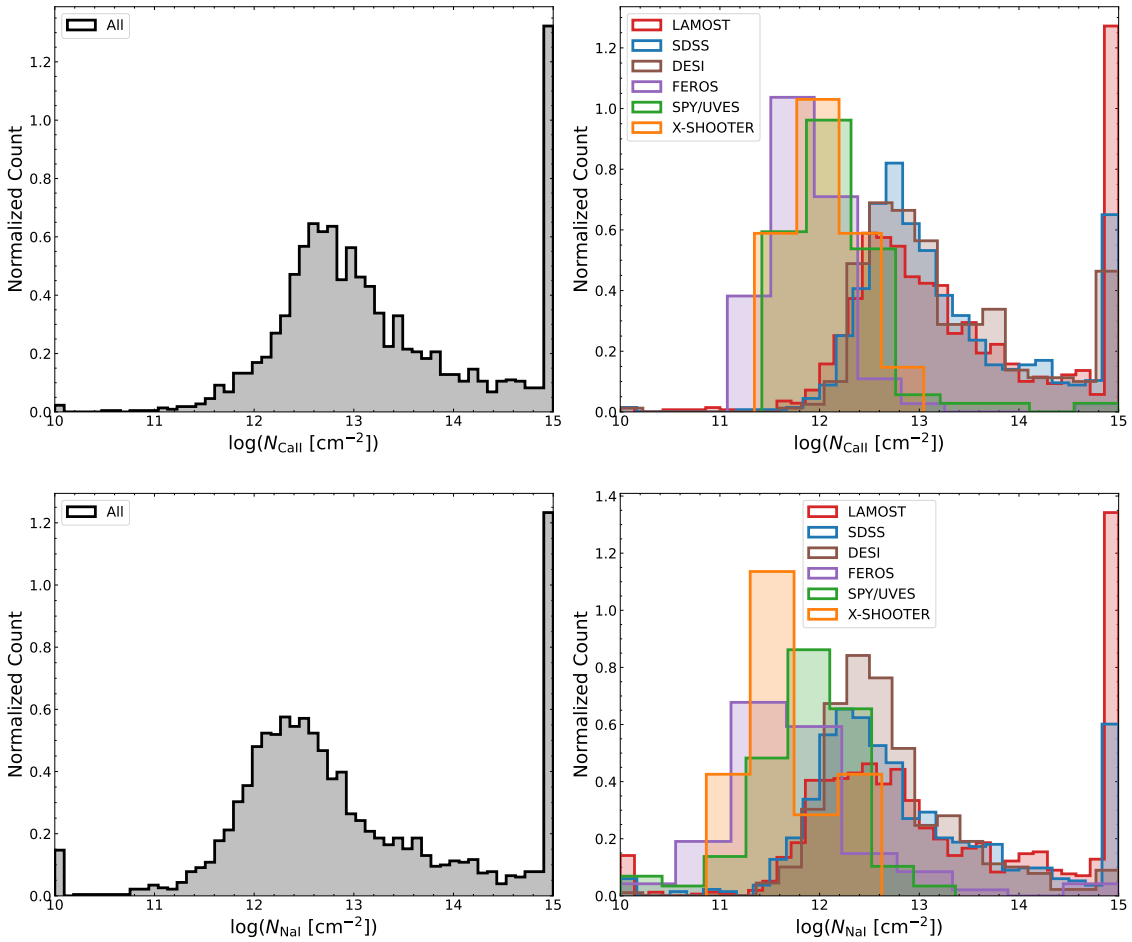


Figure 5.1: Column density distributions of Ca II (upper left panel) and Na I (lower left panel) before applying the new SNR cuts. The number of objects is identical to N_0 in Table 5.1. The number of bins is $\sqrt{N_{\rm data}}$ within the observed parameter range, where $N_{\rm data}$ is the number of data points. The right panels show the same, but for each survey/spectrograph individually. Clearly visible is the dominant peak at unrealistically high column densities.

one direction, nor to move very fast. The only exceptions are high and intermediate velocity clouds (HVCs and IVCs, respectively), which show radial velocities of more than $|v_{\rm rad,ISM}| > 100\,{\rm km\,s^{-1}}$ (Wakker & van Woerden, 1997). However, those clouds are located in the Galactic halo at distances $d \gtrsim 5\,{\rm kpc}$ or are even of extragalactic origin (Richter et al., 2017; Werk et al., 2019; Marasco et al., 2022). Therefore, only little overlap with the analyzed sample is expected and no clear evidence for interstellar lines originating from those clouds has been found. Most of the high radial velocity measurements at the outer wings of the distribution of Figure 5.5 seem to be spurious or very uncertain ones from low resolution spectra. Statistical evaluation of the distribution gives

$$\bar{v}_{\rm rad,ISM,mode} = -7^{+26}_{-31}\,{\rm km\,s^{-1}} \qquad \text{and} \qquad \bar{v}_{\rm rad,ISM,median} = -8^{+29}_{-28}\,{\rm km\,s^{-1}}\,. \tag{5.7}$$

This quantitatively shows that the dispersion of interstellar radial velocities is about $\pm 30\,{\rm km\,s^{-1}}$, while for stellar ones it is up to $\pm 80\,{\rm km\,s^{-1}}$ (Figure C.2). The very little skewness towards negative radial velocities is most likely a consequence of the more sparsely coverage of the Eastern part of the sky compared to the Western part (Figure 4.10). Due to differential Galactic rotation the Western side predominantly shows negative radial velocities, while the Eastern one shows predominantly positive ones (also Figure 4.10). Hence, there are just more data points at negative radial velocities explaining the little skewness. This is also observed for the stellar radial velocities (Figure C.2) and would most likely disappear in a fully homogeneous sample.

Lastly, lots of monochromatic reddening parameters $E(44-55)$ have been measured using the SED method, which quantify the amount of intervening interstellar dust. In principle, it is

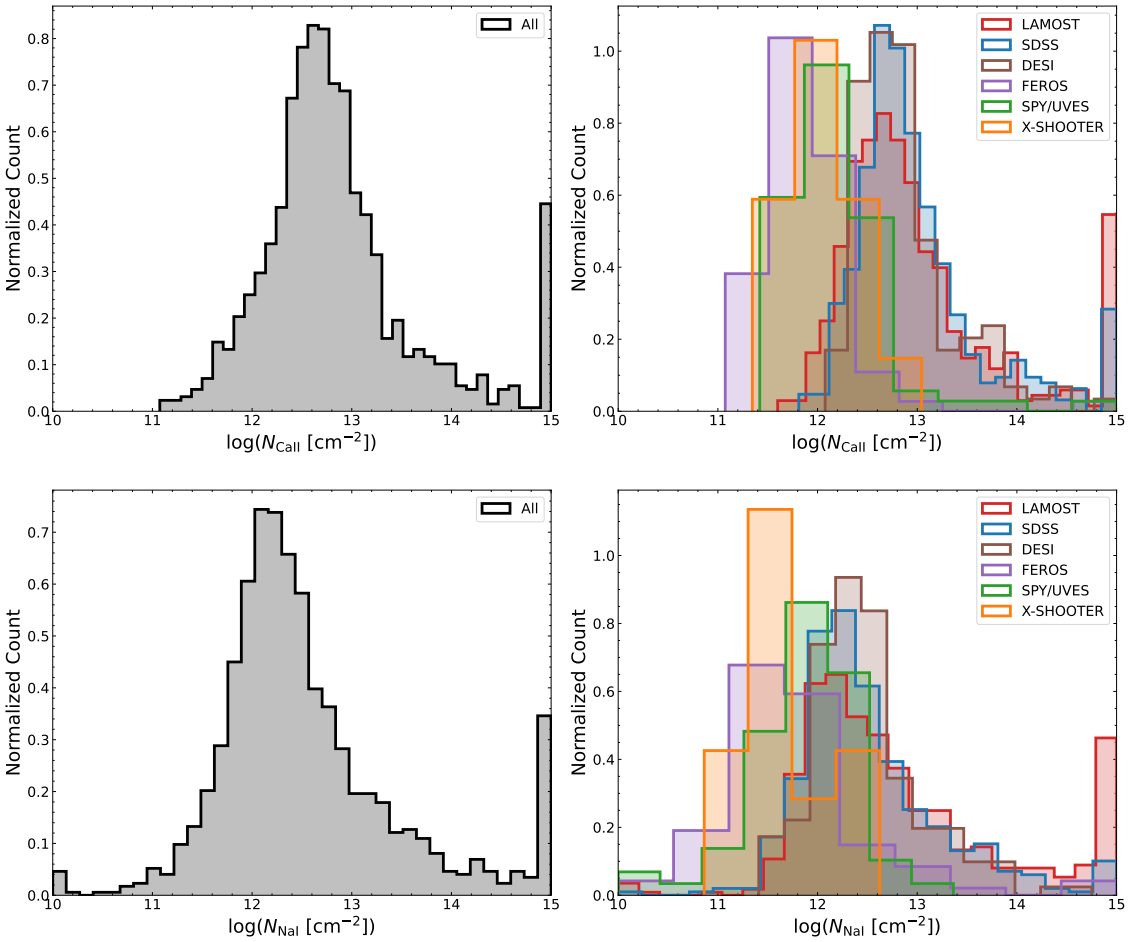


Figure 5.2: Same as Figure 5.1, but after applying the new SNR cuts. The large peak at unrealistically high column densities is considerably reduced compared to Figure 5.1.

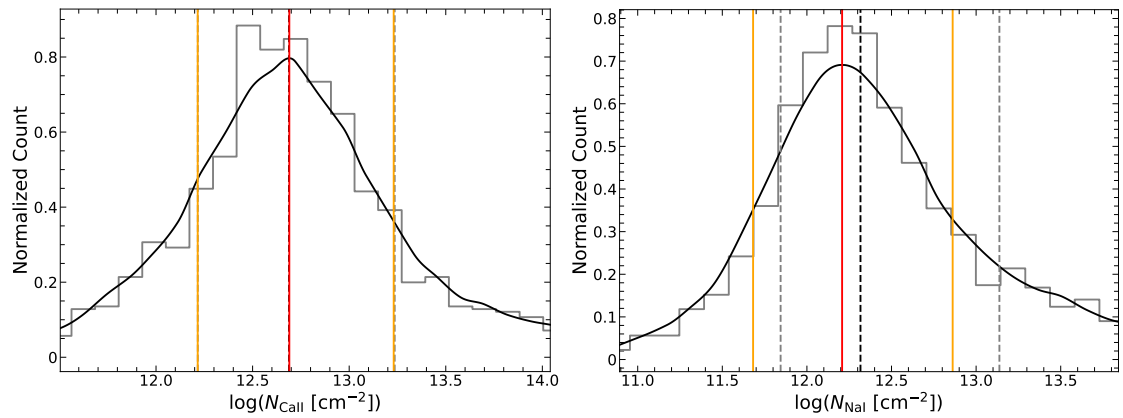


Figure 5.3: Mode (red line) and highest density interval (HDI, yellow lines) as well as median and quantiles (dashed lines) determination (see Appendix A) for the distributions in the left panels of Figure 5.2. The results are identical for the column densities of Ca II with $\log \bar{N}_{\rm CaII,mode/median} = 12.7^{+0.6}_{-0.5}$, while they are slightly different for the column densities of Na I with $\log \bar{N}_{\rm NaI,mode} = 12.3^{+0.7}_{-0.6}$ and $\log \bar{N}_{\rm NaI,median} = 12.4^{+0.9}_{-0.5}$.

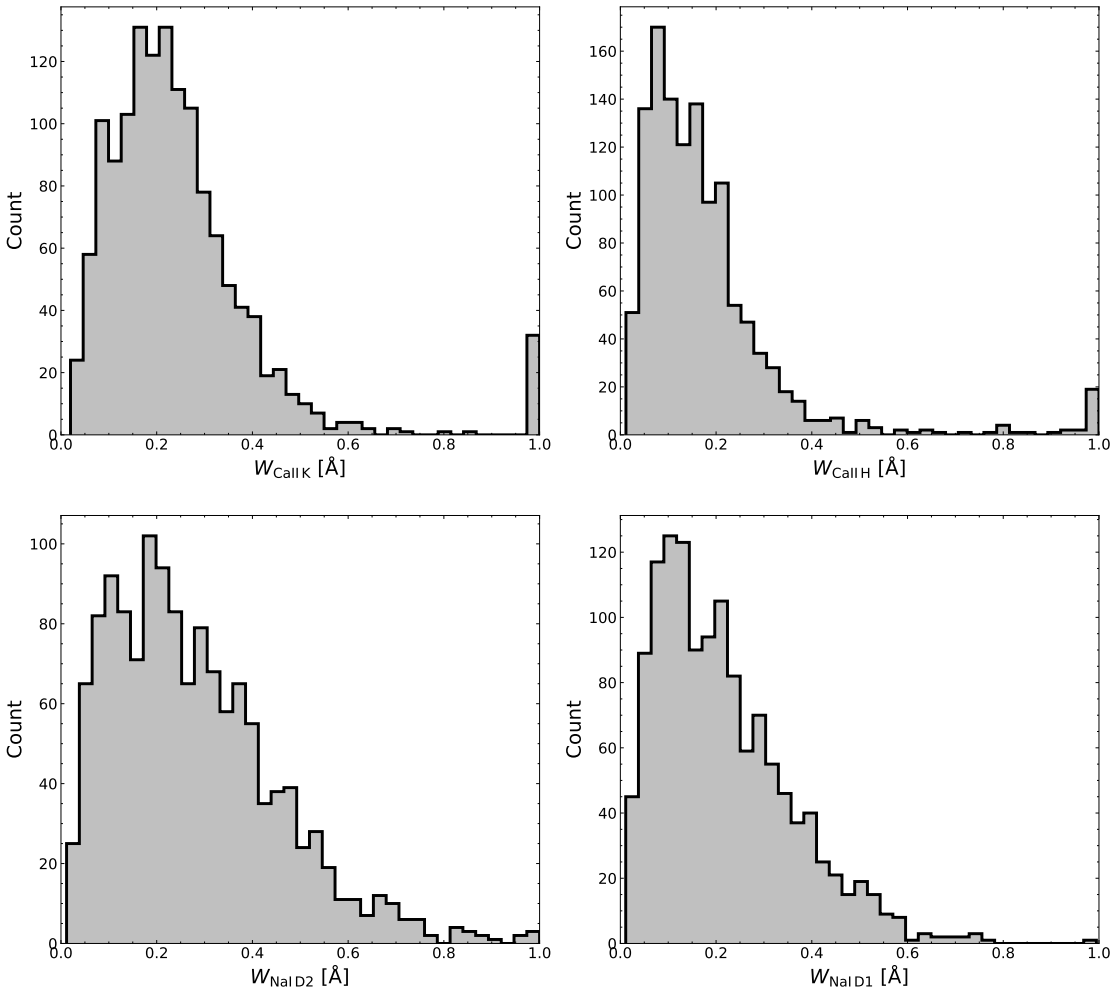


Figure 5.4: Measured equivalent distributions for the four interstellar lines. The small peaks especially for Ca II at unrealistically high values of $W > 1$ Å most likely again come from spurious measurements using low resolution spectra. The statistical evaluations can be found in Figure D.1.

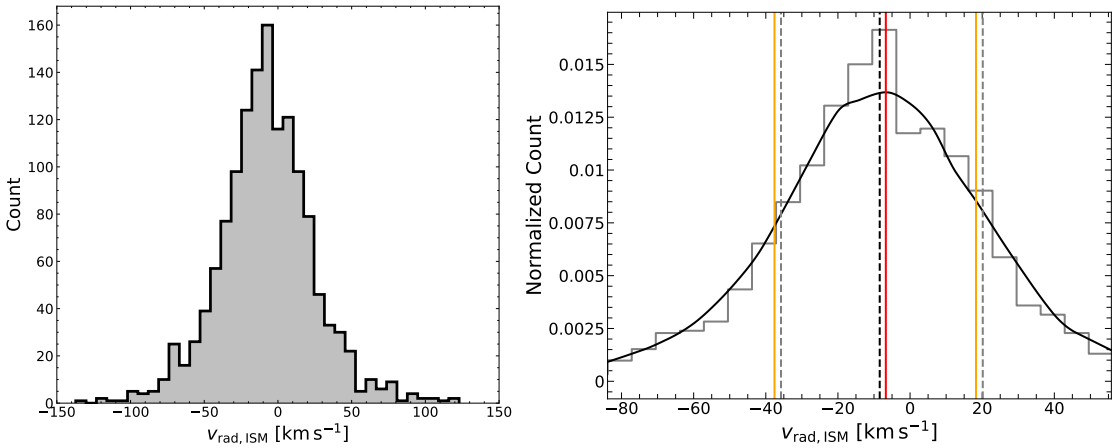


Figure 5.5: Overall distribution of measured interstellar radial velocities (left panel) and its statistical evaluation (right panel). The red and yellow vertical lines depict the determined mode and HDI of $\bar{v}_{\mathrm{rad,ISM,mode}} = -7^{+26}_{-31}\,\mathrm{km\,s^{-1}}$, while the dashed vertical lines depict the median and quantiles of $\bar{v}_{\mathrm{rad,ISM,median}} = -8^{+29}_{-28}\,\mathrm{km\,s^{-1}}$.

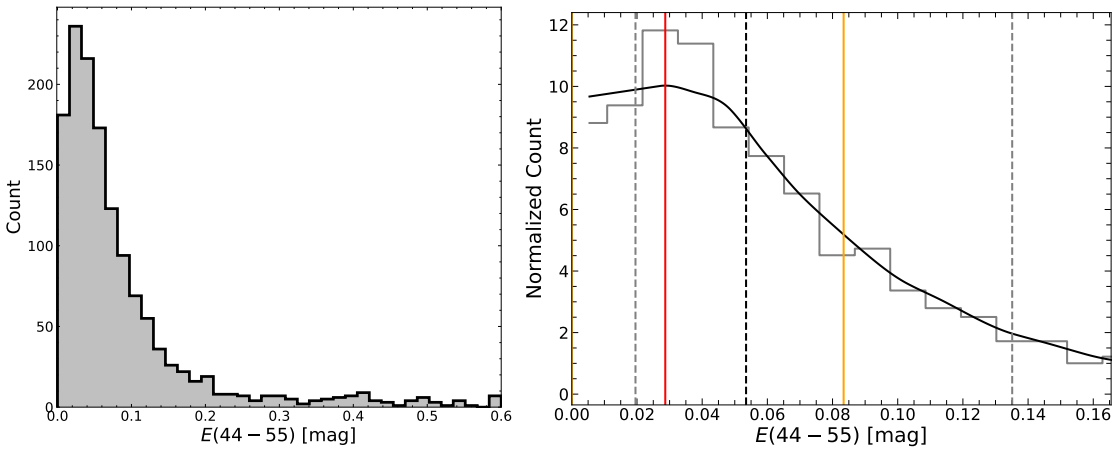


Figure 5.6: Overall distribution of measured interstellar reddening parameters (left panel) and its statistical evaluation (right panel). The red and yellow vertical lines depict the determined mode and HDI of $\bar{E}(44-55)_{\mathrm{mode}} = 0.03^{+0.06}_{-0.03}$ mag, while the dashed vertical lines depict the median and quantiles of $\bar{E}(44-55)_{\mathrm{median}} = 0.05^{+0.09}_{-0.04}$ mag.

possible to convert it to the more commonly used $E(B-V)$ reddening parameter (Fitzpatrick et al., 2019), but this conversion is actually not trivial. Therefore, and because the conversion factor is usually close to one, especially in the UV/blue regime where hot subdwarfs emit most of their light, it did not appear worthwhile to carry out the conversion and in the further course of this work $E(44-55)$ is used. Its distribution is shown in Figure 5.6. It comprises the physically expected parameter space up to 1 mag (Green et al., 2015), but is clipped at 0.6 mag for a better visualization. However, the comprehensive three dimensional dust maps of Green et al. (2015) show that high reddening values are only reached for sight lines within the Galactic plane, where most of the interstellar dust is concentrated. The stellar as well as interstellar density falls off exponentially with increasing distance from the disk (exponential disk). This explains, why the reddening distribution of Figure 5.6 mostly shows small values approaching zero: The Galactic plane is almost entirely excluded (Figure 2.3) and most of the hot subdwarf stars are located at higher Galactic latitudes. The determined mode with HDI and median with quantiles are

$$\bar{E}(44-55)_{\mathrm{mode}} = 0.03^{+0.06}_{-0.03}\,\mathrm{mag} \qquad \text{and} \qquad \bar{E}(44-55)_{\mathrm{median}} = 0.05^{+0.09}_{-0.04}\,\mathrm{mag}\,. \tag{5.8}$$

Especially the right panel of Figure 5.6 shows that the skewness caused by the proximity to zero of most of the values makes the distribution not well behaved. This complicates its evaluation and the median is definitely the better choice.

To ensure a consistent comparability between all the measured parameters, in the following if not stated otherwise, always the median and the 3σ confidence range is chosen for the parameter spaces. This mainly concerns the chosen axis limits of plots and the color codes of sky maps.

5.2 Correlations Between the Interstellar Line Parameters - A Consistency Check

The measurements of equivalent widths and column densities for each line and element are, of course, not independent. If, for example, the equivalent width of one line of a doublet increases, also that of the other line of the doublet should increase, since both originate from the same interstellar environment. Furthermore, the amounts of interstellar Ca II and Na I are also not independent of each other and the column density ratio $N_{\mathrm{Na\,I}}/N_{\mathrm{Ca\,II}}$ mainly scales with the temperature of the interstellar environment. In cold neutral gas clouds ($T \sim 30$ K) most of the gaseous calcium is depleted onto dust grains and the ratio is $N_{\mathrm{Na\,I}}/N_{\mathrm{Ca\,II}} > 100$, while in warm partially ionized gas clouds Ca II remains gaseous, Na I becomes ionized, and the ratio drops to $N_{\mathrm{Na\,I}}/N_{\mathrm{Ca\,II}} < 0.5$

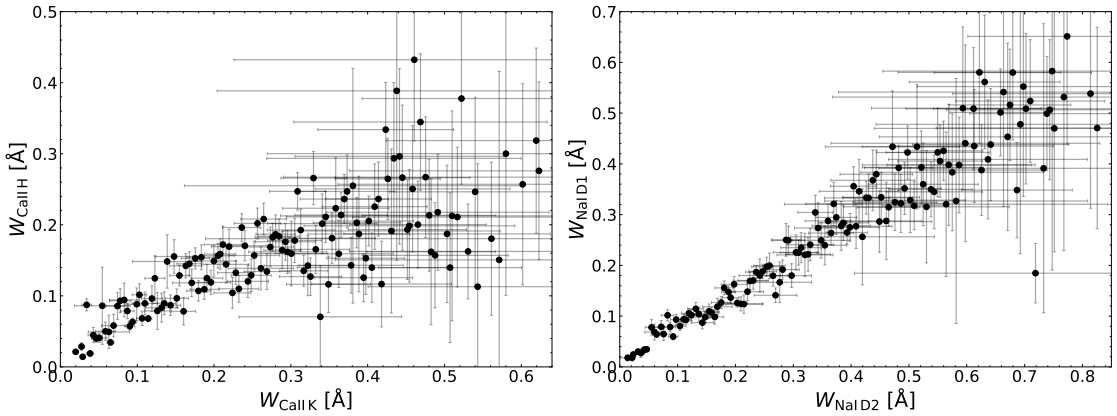


Figure 5.7: Correlation of the equivalent widths of the Ca II line doublet (left panel) and the Na I doublet (right panel). The measured data is binned into 150 equally wide bins in x-direction. The linear increase is clearly visible in both cases.

(Welsh et al., 1991, 2010). Taking the average column densities of each element from Equation 5.1 and Equation 5.2, the average ratio here is about

$$\frac{N_{\mathrm{Na\,I}}}{N_{\mathrm{Ca\,II}}} = \frac{10^{12.3}\,\mathrm{cm}^{-2}}{10^{12.7}\,\mathrm{cm}^{-2}} \approx 0.4\,. \tag{5.9}$$

This result is in perfect agreement with the findings of Linsky et al. (2022) that the space within 4 kpc around the Sun is filled with warm and partially ionized interstellar clouds having temperatures in the range of 2,450 K up to 12,900 K. However, the Na I to Ca II ratio should stay approximately constant across an interstellar cloud. Therefore, an increase of the column density of one element should also come with an increase of the column density of the other element (or equivalent width, respectively). These patterns are investigated further in this section.

But first of all, just plotting the measured parameters over each other with more than 1,000 data points does not appear reasonable. The figures would be heavily crowded and potential trends smeared out. Thus, it was decided to bin the data for a better view and to isolate those potential trends. So each of the following plots contains 150 equally wide bins in x-direction, where the median and the standard error is taken for the x- and y-values.

Figure 5.7 shows the correlation of the equivalent widths among the two line doublets. Even though the uncertainties and the scattering increases at large equivalent widths in both cases[2], the linear trend is clearly visible. In particular for the Na I D lines, it is nicely confined and well pronounced. The reason, why for the Ca II lines the scattering is a bit higher, is most likely that the SNR in the blue regime of optical spectra is usually lower than in the red one. But the weakness of the Ca II H line combined with the Hε blend could also play a role. Still, the linear trend is also visible there. It exactly confirms the expectation that the equivalent widths of both lines of a doublet increase equally strong with changing interstellar environments like density, temperature, and microturbulent velocity.

Next, the correlation of the line strengths across the two elements is investigated. In Figure 5.8 the equivalent widths of both Ca II lines are plotted against that of the two Na I lines. For the strength of the Ca II K line again a linear correlation with the strengths of the Na I D lines is clearly visible (upper panels). For the Ca II H line (lower panels), however, the trend is smeared out and does not necessarily appear to be linear. This can be again explained with the overall weakness of the line combined with the Hε blend, which is certainly a source of uncertainty. But it is still obvious that the equivalent width of the Ca II H line increases simultaneously with that

[2] This is not only a result of indeed increasing uncertainties for increasing equivalent widths, but also a result of the binning. The number of data points in the high equivalent width bins is much smaller than in the lower ones. If additionally the values for these data points scatter a lot, the uncertainties of the binned data is quickly inflated. This is also seen in further plots in this chapter.

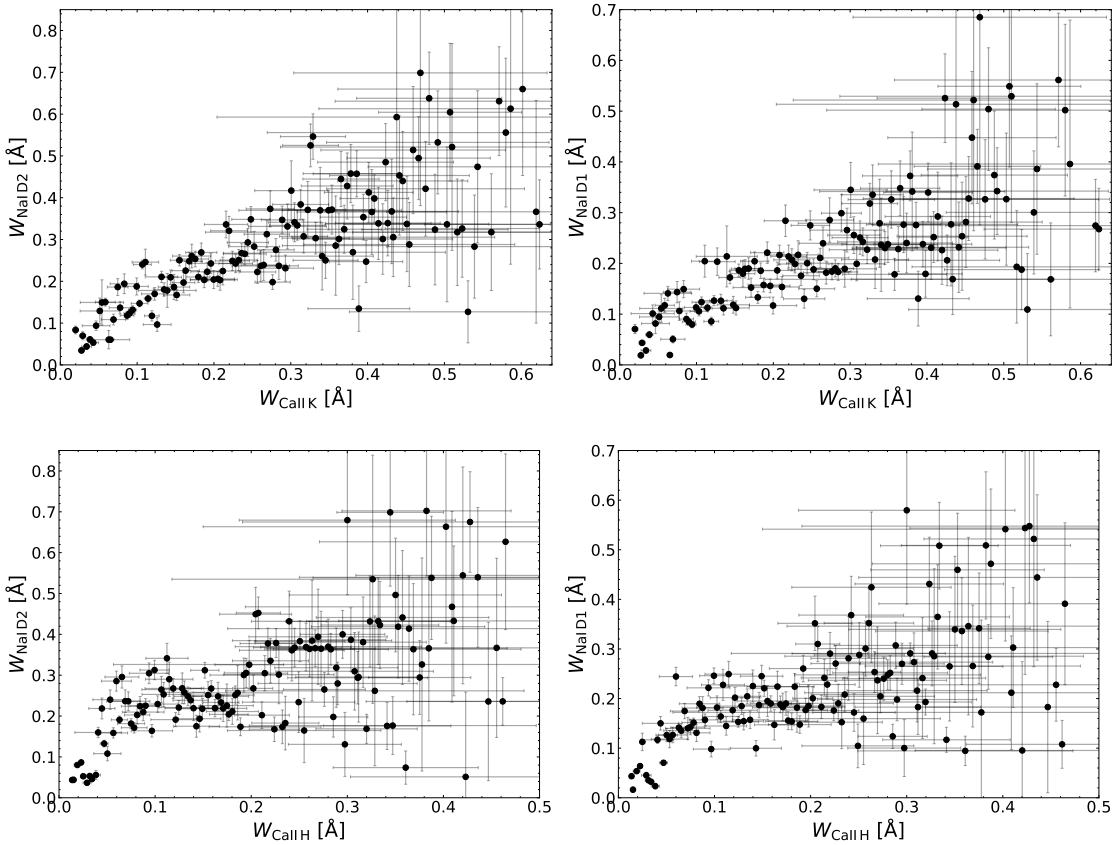


Figure 5.8: Correlation of the equivalent widths of the Ca II H&K lines with that of the Na I D1&D2 lines. Apparently, the amounts of interstellar Ca II and Na I causing the spectral lines are not independent of each other, since the equivalent widths simultaneously increase.

of the Na I D lines. In Figure 5.9 the same is shown for the measured column densities of both elements. Again, even though it can not be concluded that the dependence is necessarily linear, it is clearly visible that both column densities correlate with each other. If one of them increases, on average also the other one increases. Together with Figure 5.7 this confirms the expectation that the abundances of Ca II and Na I in interstellar clouds are not independent, even though they vary with temperature. Except of in the extreme cases of cold neutral clouds ($T \sim 30\,\mathrm{K}$) and warm ionized clouds ($T \sim 15,000\,\mathrm{K}$), there are barely clouds either containing large amounts of calcium and totally lacking of sodium or vice versa. Hence, on average the densities of both scale with each other.

The last relation to check is the increase of the equivalent width with column density. As already stated in Section 4.3, it is theoretically described by the curve of growth, which has three distinct regimes depending on the column density N (Draine, 2011; Bartelmann, 2021; Gray, 2021):

$$W(N) \propto \begin{cases} N & \text{small } N\text{, linear regime} \\ \sqrt{\ln N} & \text{moderate } N\text{, saturated line core regime} \\ \sqrt{N} & \text{high } N\text{, Lorentzian wings regime} \end{cases} \tag{5.10}$$

In Figure 5.10 the equivalent widths of all four interstellar lines are plotted against the respective column density. In order to stick with the convention, the equivalent widths are also plotted in a logarithmic scale. In all of the cases the $\sqrt{\ln N}$ dependence is strongly pronounced and confined, especially when comparing the curves to textbook figures or simply to that on Wikipedia[3]. Apparently, the measured column densities lie at the lower end of the saturated core regime. With spectra that are more sensitive to weaker lines, i.e. having higher resolution or signal to noise

[3] https://en.wikipedia.org/wiki/Curve_of_growth

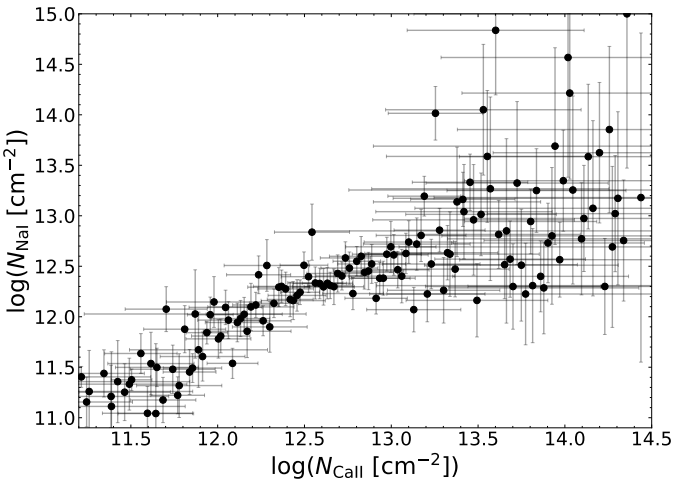


Figure 5.9: The correlation of measured Ca II and Na I column densities. Obviously, both scale with each other, even though the dependence does not appear necessarily linear. Physically this means that the ratio of both on average comprises only a small range, at least for the analyzed sample. If the density of calcium increases in an interstellar cloud, also the density of sodium increases and vice versa.

ratio, a lower detection limit could be reached, perhaps also revealing the transition towards the linear regime. While the linear and the saturated line core regime are indeed those expected for the densities of the interstellar medium, the Lorentzian wings regime is most likely never reached for interstellar conditions, even for sight lines within the Galactic disk. Such high densities regions might only be found in stellar environments or other very high column density sites, but not in the average interstellar medium.

Overall it can be said from the findings of this section, that all measurements perfectly fulfill the expectations. This does not only demonstrate the quality of the derived data, but also underlines the robustness of the developed analysis methods through all of the complications and caveats discussed in Subsection 3.3.4. Furthermore, it showcases the effectiveness of the applied quality cuts. Without them, the results would by far not look as conclusive.

5.3 Distribution of Interstellar Gas and Dust on the Sky

In this section the two dimensional distribution of gas and dust on the sky is investigated. In principle, it would be desirable to compile comprehensive gas and dust maps, as for example done by Murga et al. (2015), Green et al. (2019), or Anders et al. (2022). With distance information even the third dimension is accessible, instead of just the two dimensional distribution on the sky. But unfortunately, with only about 1,000 sight lines as in this work, even that 2D distribution on the sky is only poorly covered. Therefore, compiling three dimensional gas and dust maps, for which hundred thousands or millions of background sources are needed, is far beyond the capabilities of the obtained data set. Still some major structures and differences between the distributions of gas and dust can be resolved.

In Figure 5.11 the distribution of Ca II column densities, in Figure 5.12 the distribution of Na I column densities, and in Figure 5.13 the distribution of the measured reddening parameters is shown. The column densities represent gaseous Ca II and Na I, which can be both considered as proxy for the overall amount of intervening gas, while the reddening parameter stands as proxy for the overall amount of intervening dust. For all sky maps the Aitoff projection is used and all measurements exceeding the borders of the color-coded range are set to the border values. The same maps for the determined equivalent widths are shown in Section D.2, but do not provide new insights.

What all sky maps have in common, is a gradient from the Galactic plane towards higher Galactic latitudes. Much higher values are measured close to the disk than towards the poles, which is expected, since most of the interstellar gas and dust is confined to the Galactic disk. Furthermore, the highest values are measured in all maps in a region towards the Galactic anticenter. With a range in Galactic longitude of about $130^\circ < l < 200^\circ$ and in Galactic latitude of about $-30^\circ < b < 10^\circ$ this major structure can most likely be identified with the Taurus–Perseus–Auriga–California complex as part of the Gould belt (Ungerechts & Thaddeus, 1987; Alves et al., 2020). This

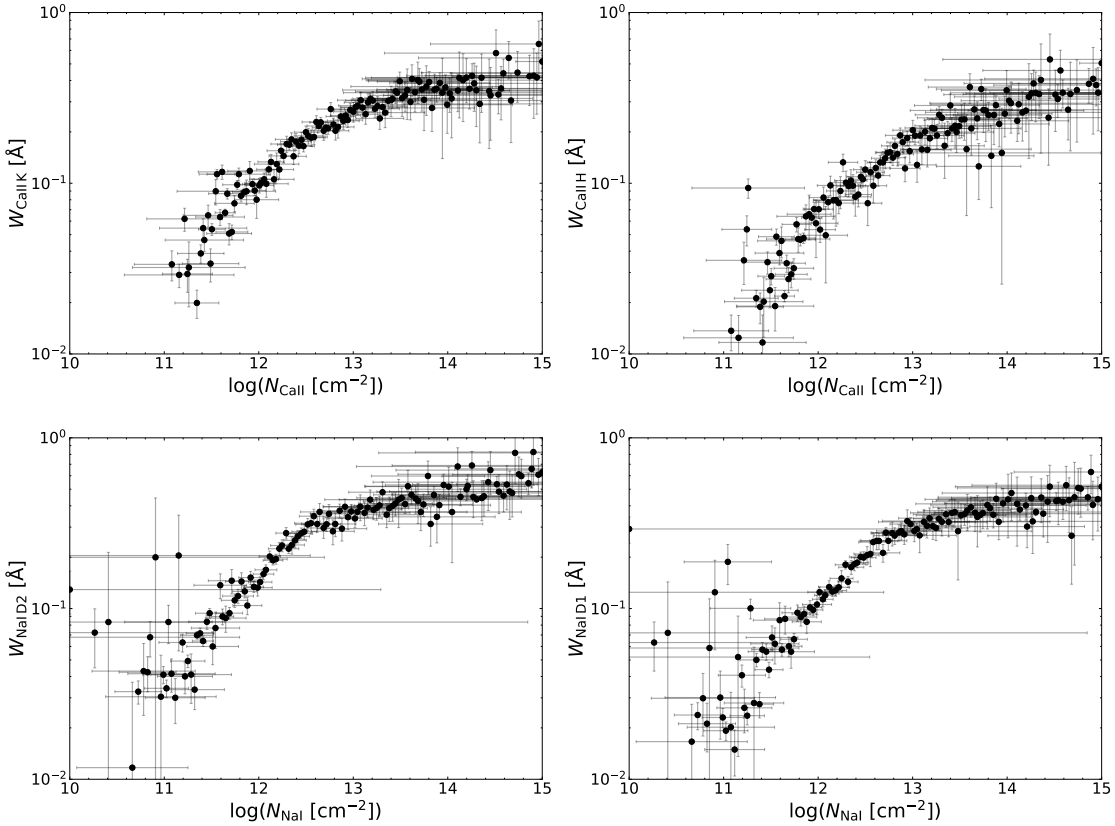


Figure 5.10: Correlation of the equivalent widths of the four interstellar lines with the respective column densities, which is theoretically described by the curve of growth (Equation 5.10). The $W(N) \propto \sqrt{\ln N}$ behavior is clearly pronounced and the measured column densities all lie within the saturated line core region. At best the transition towards the linear regime is touched at the lower column density end, but the high density Lorentzian wing regime is never reached, as expected for interstellar environments. Uniform axis limits instead of the 3σ parameter space from the distributions are chosen for the best comparability of all four plots.

structure is also, for example, seen in Figure 1.1. With distances of only $120-400$ pc this giant accumulation of dark clouds, molecular clouds, and star forming regions, is very close by and covers a large area of the sky.

But the maps also show some differences among each other, even the Ca II and Na I gas maps. According to Figure 5.11, moderately large Ca II column densities scatter up to higher Galactic latitudes, especially in the North. The South is much more poorly covered at high Galactic latitudes, but also shows some isolated larger values down to $b \sim -60°$. Since on average no asymmetry between the Northern and Southern side is expected, the pattern should be the same for a homogeneously distributed sample. That more higher values are measured towards the North is most likely an observational bias. The Na I column density map (Figure 5.12) shows the same patterns and structures up to $|b| \sim 40°$, but at higher Galactic latitudes the measured values drop off rapidly. Unlike for Ca II, almost no higher column densities are found for $|b| \gtrsim 60°$. This means that gaseous Na I, mostly found in cold neutral clouds, is more closely confined to the Galactic plane than gaseous Ca II, which is mostly found in warm partially ionized clouds and extends more uniformly to higher Galactic latitudes. This result is in great consistency with, for instance, Welsh et al. (2010) or Murga et al. (2015). While the background sources of Welsh et al. (2010) have distances of up to only a few hundred parsecs, making it hard to directly compare the results, Murga et al. (2015) used several hundred thousand extragalactic sources (quasars and galaxies) to shine through the interstellar and circumgalactic media of the Milky Way, exclusively at high Galactic latitudes. Their Ca II and Na I gas maps show exactly the same patterns and structures as

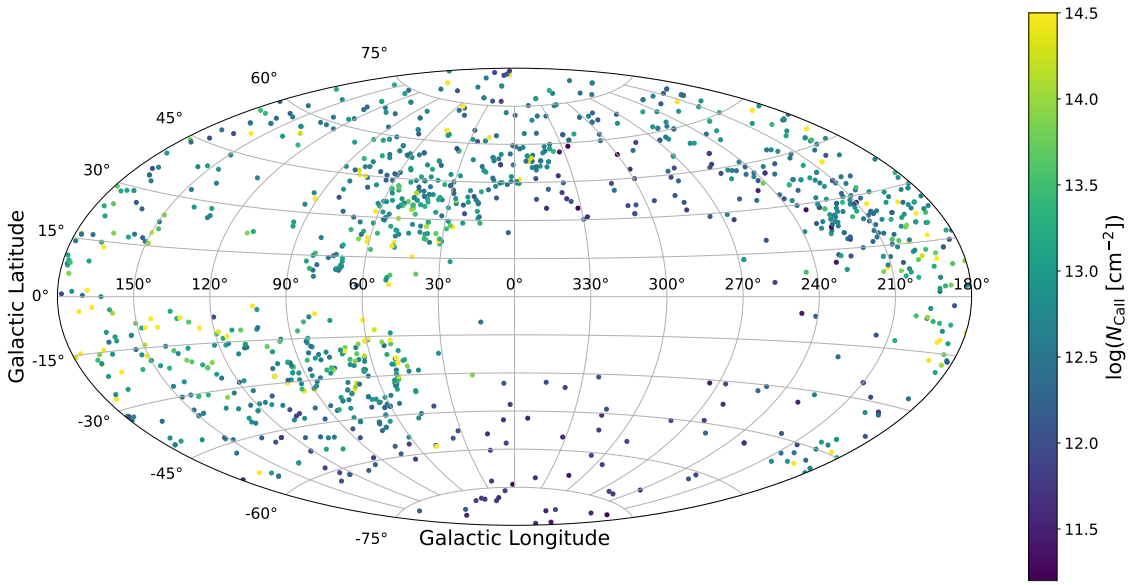


Figure 5.11: Aitoff projection of the measured Ca II column densities on the sky in Galactic coordinates. Values exceeding the color-coded parameter range are saturated. The most prominent feature is observed towards the Galactic anticenter, most likely matching with the Taurus–Perseus–Auriga–California complex as part of the Gould belt (Alves et al., 2020). High column density values extend up to high Galactic latitudes.

Figure 5.11 and Figure 5.12.

Now when looking at the distribution of interstellar dust in Figure 5.13, the structure towards the Galactic anticenter is even more pronounced than in the gas maps. Except of a few measurements close to the Galactic plane, it is essentially the only region, where the reddening parameters exceed values of about 0.2 mag. This is actually also the biggest contrast to the previous maps. For Galactic latitudes of $|b| \gtrsim 30°$, the vast majority of the data points are black, indicating a reddening close to zero. Apparently, interstellar dust seems to be heavily concentrated in the Galactic disk and not extending at all towards higher Galactic latitudes. This finding agrees well with, for example, Sodroski et al. (1997), Demyk (2011), or Anders et al. (2022). Interstellar dust is predominantly found in the Galactic disk, where it closely correlates with the density of gas. It is primarily concentrated in regions with dense gas such as molecular clouds and areas of star formation like the Taurus–Perseus–Auriga–California complex. While interstellar gas can extend to higher Galactic latitudes, especially in the form of diffuse clouds, dust tends to be more localized within the disk. This is due to the processes of dust formation and destruction, which are heavily influenced by the local environment, particularly in regions of star formation where dust is created and subsequently mixed with gas (Demyk, 2011).

Lastly, in Figure 5.14 the distribution of measured ISM line radial velocities is shown. Again like for the stellar radial velocities (Figure 4.10), the Western side predominantly shows negative radial velocities (blue dots), while the Eastern side predominantly shows positive ones (red dots). This is again a clear sign for Galactic differential rotation. Deviations from this trend are most likely explained by the uncertain radial velocity measurements using low resolution spectra.

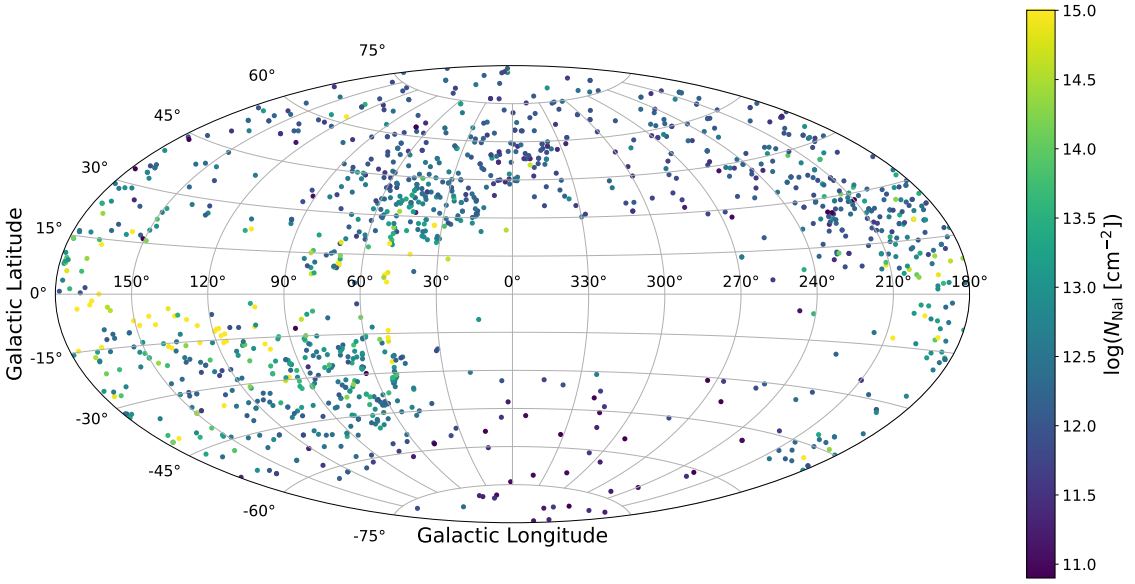


Figure 5.12: Aitoff projection of the measured Na I column densities on the sky in Galactic coordinates. Values exceeding the color coded parameter range are saturated. The most prominent feature is observed towards the Galactic anticenter, most likely matching with the Taurus–Perseus–Auriga–California complex as part of the Gould belt (Alves et al., 2020). Higher column density values fall off more rapidly towards higher latitudes as for Ca II in Figure D.2.

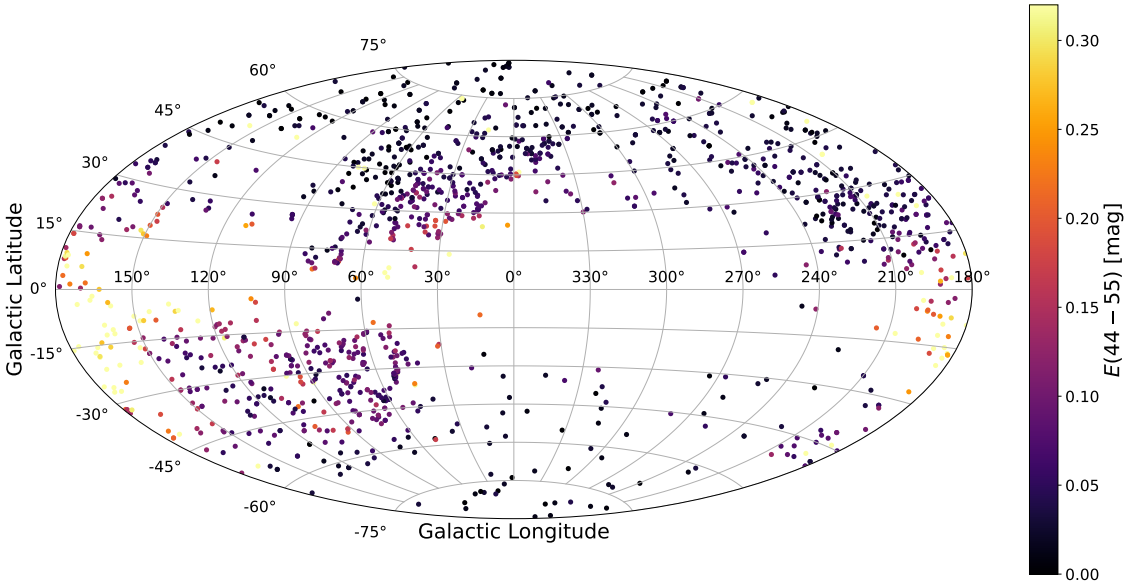


Figure 5.13: Aitoff projection of the measured reddening parameters on the sky in Galactic coordinates. Values exceeding the color coded parameter range are saturated. The most prominent feature is observed towards the Galactic anticenter, most likely matching with the Taurus–Perseus–Auriga–California complex as part of the Gould belt (Alves et al., 2020). Apart from that feature almost no larger reddening values are visible. In particular at high Galactic latitudes no significant reddening is measured at all.

5.4 Variation of Interstellar Properties with Distance, Direction, and Height from the Galactic Plane

In the previous section the two dimensional distributions of the measured interstellar parameters on the sky were inspected. Thanks to *Gaia*, also the third dimension is accessible in form of parallaxes and the corresponding distances. Hence, this section will dig a bit deeper and have a closer look on the correlations of the interstellar parameters with the distance itself, the directions of the sight lines combined with distances, and the heights above and below the Galactic disk.

In Figure 5.15 the interstellar parameters (equivalent widths of the four ISM lines, column densities of both elements, and reddening) are plotted against the stellar distances. Again, the data is binned into 150 equally wide bins in x-direction in order to make the figures not overcrowded and to better isolate potential trends. Additionally, the median of each parameter distribution as determined in Section 5.1 is depicted as red dashed line. For the equivalent widths and column densities a clear trend can be observed for distances smaller than about 1 kpc, where the line strengths increase with increasing distance. For larger distances, the line strengths reach average values (the dashed lines) and begin to saturate. Additionally, at about 2.5 kpc the measurements begin to scatter a lot with large uncertainties smearing out potential trends. As explained before, this is not only a consequence of indeed larger uncertainties, but also of the binning. Overall, the observed behavior can be explained as follows: For very low distances there might on average not be a lot of intervening gas in the line of sight, resulting into low column densities, which increase successively with distance as additional clouds are intersected. At the same time, for low column densities the equivalent widths increase linearly with column density and hence, distance, according to the curve of growth. At some distance, when enough material is added to the line of sight, the absorption line cores begin to saturate and the growth of the equivalent widths with column density begins to flatten (Figure 5.10). The reason why also the column densities begin to saturate with increasing distance is most likely a geometric effect. Since most of the stars are observed towards higher Galactic latitudes, large distances place them into the Galactic halo rather than the Galactic disk, where most of the interstellar gas is concentrated. Thus, increasing the distance further will not add significant amounts of interstellar material to the line of sight, assuming that the ISM density drops off in the halo, making the column densities saturate as well as the equivalent widths. In particular, the roughly linear increase of the equivalent widths at low distances was already observed formerly. Therefore, attempts were made to model this correlation as precise as possible in order to estimate distances to early type stars based on the interstellar absorption lines in their spectra (Megier et al., 2005, 2009). However, this approach is obviously only valid for small distances up to a few hundred parsecs before the equivalent widths begin to saturate. Furthermore, it is highly uncertain and it was already predicted by Megier et al. (2009) that the parallaxes from future astrometric surveys will supersede the distance estimates from interstellar line strengths. With the advent of the unprecedented *Gaia* mission this became clearly the case, making distances inferred from interstellar line strengths redundant. Hence, even though it might be an interesting exercise, it did not appear worthwhile to reproduce the relations of Megier et al. (2005) and Megier et al. (2009) here. Otherwise, they also investigate the distance dependence of interstellar line strengths from other metals as well as reddening. However, they found that those are no suitable candidates for distance estimates as their relations are far less tight than for Ca II (Na I is not discussed). This is also reflected in the lowest panel of Figure 5.15, where the reddening is plotted over the distance. It shows almost no pattern at all, which most likely means that most of the interstellar dust is located very close to us. Increasing distance, especially towards higher Galactic latitudes, does on average not add any significant amount of dust to the line of sight such that the distribution rapidly flattens.

In the sky maps of Section 5.3 a clear gradient of the interstellar parameters from low to high Galactic latitudes was observed. This gradient is further investigated in Figure 5.16, where the measured interstellar parameters are plotted over Galactic latitude. To include distance information, the data is binned into three different distance bins ($d < 1\,\mathrm{kpc}$, $1\,\mathrm{kpc} \leq d < 2\,\mathrm{kpc}$, and $d \geq 2\,\mathrm{kpc}$). The measurements of each distance bin are again binned into 50 equally wide bins in x-direction.

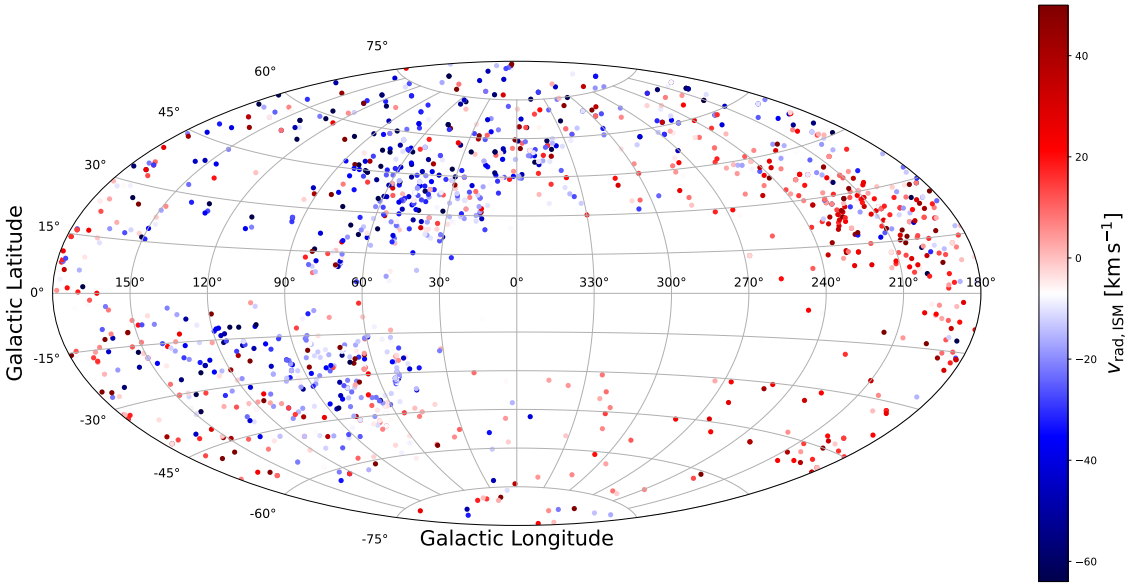


Figure 5.14: Aitoff projection of the measured ISM line radial velocities on the sky in Galactic coordinates. Values exceeding the color coded parameter range are saturated. The trend of predominantly negative radial velocities (blue dots) on the Western side and predominantly positive radial velocities (red dots) on the Eastern side can be explained with differential Galactic rotation. Deviations most likely originate from uncertain radial velocity measurements using low resolution spectra.

Additionally, the median of the respective parameter distribution is depicted as dashed horizontal line. In the panels of Figure 5.16, showing the interstellar line parameters (equivalent widths and column densities), it is clearly visible that on average higher values are measured for larger distances, at least at lower Galactic latitudes. For higher Galactic latitudes the three curves more or less seem to merge, which means increasing the distance does not add significant amounts of interstellar material. This makes sense for objects located in directions close to the Galactic poles, where the sight lines traverse mostly the Galactic halo rather then the disk and little intervening interstellar material is added with increasing distance. For the reddening, however, almost no distance dependence is observed. This was already indicated by the last panel of Figure 5.15. Otherwise, Figure 5.16 strongly confirms the findings of Section 5.3:

- Most of the intervening gas and dust is found for sight lines within the Galactic disk at low latitudes as seen from the huge peak in the middle of each panel.

- A second smaller peak or rather plateau is visible for about $-50^\circ < b < -10^\circ$, which most likely matches with the Taurus–Perseus–Auriga–California complex as part of the Gould belt. Since it consists mainly of molecular and dark clouds, which are associated with cold neutral medium, this feature is more pronounced for Na I and reddening than for Ca II, predominantly found in warm partially ionized medium.

- The asymmetry between the North and the South is most likely an observational effect. The South is much more sparsely covered than the North and by chance predominantly sight lines are observed with little intervening interstellar material, while the North covers also sight lines with denser material. In a homogeneously distributed sample, this asymmetry would most likely disappear. Hence, in the following the North is considered as statistically more representative.

- Gaseous Ca II is more uniformly distributed towards higher Galactic latitudes than Na I, which is more closely confined to the Galactic plane. This can be seen from the less dominant peak for Ca II equivalent widths and column densities at low Galactic latitudes and the less

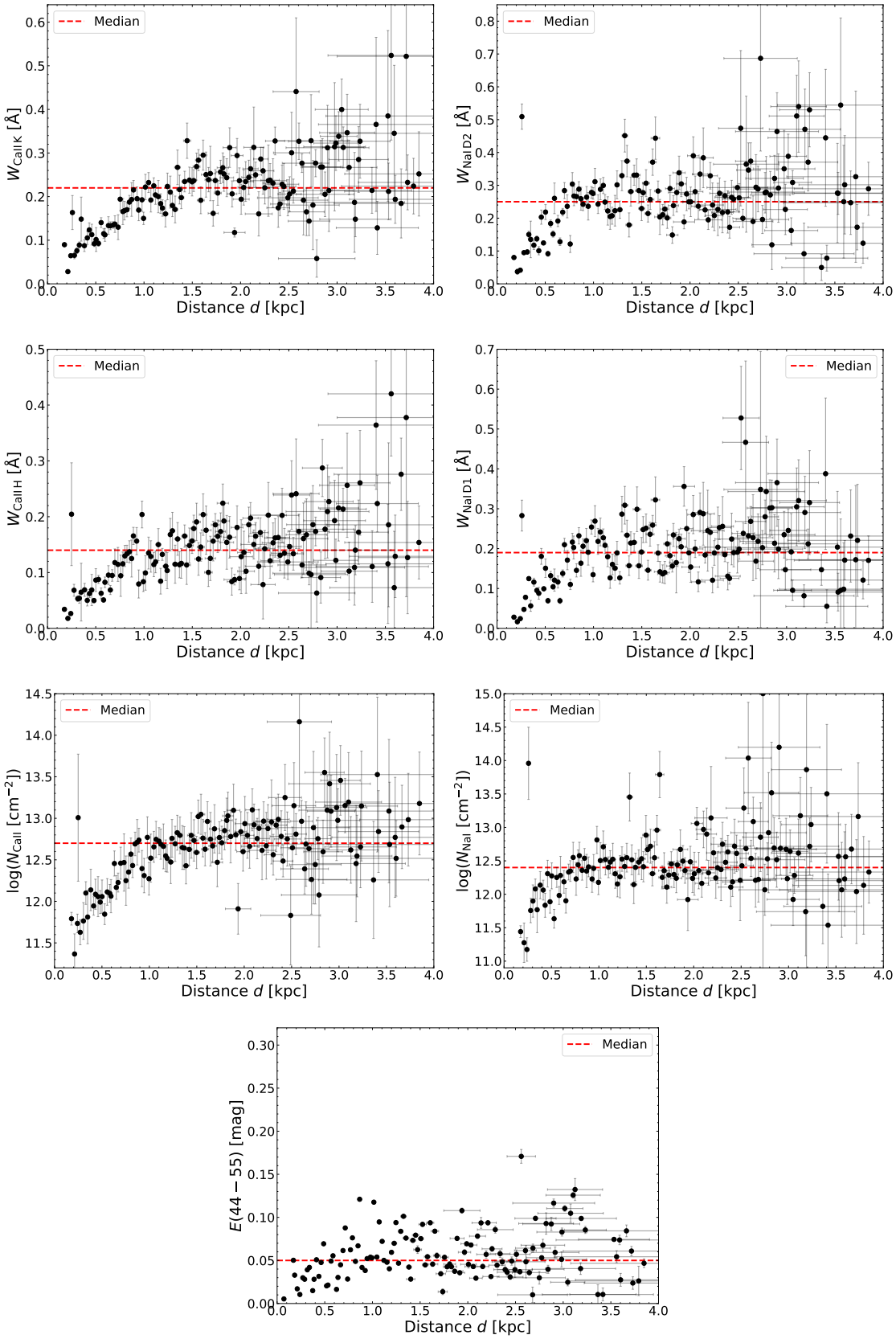


Figure 5.15: Dependence of the interstellar line parameters (equivalent widths and column densities) as well as reddening on the distance. The median of each parameter distribution is depicted as dashed red line. The data is binned to 150 equally wide bins in x-direction for better visibility. In the case of the ISM line parameters a clear correlation is observed for low distances. At higher ones the values begin to saturate. For the reddening parameter almost to pattern is visible.

rapid decrease towards higher Galactic latitudes. The measurements for Ca II stay at the average (dashed line) even for the highest Galactic latitudes, while those for Na I fall well below it.

- Interstellar dust is most strongly confined to the Galactic disk, even stronger than gaseous Na I. While the Na I line strengths fall below the average at high Galactic latitudes, but still show moderate values, the reddening almost drops to zero.

Analog plots as in Figure 5.16 for the Galactic longitude are shown in Section D.3. However, except of the Taurus–Perseus–Auriga–California complex at about $100^\circ < l < 210^\circ$ they do not show any features or new insights and are most likely also biased by observational effects.

Furthermore, analog plots are also shown for the height z above and below the Galactic plane in Figure 5.17, where the binning is the same as before. From simple geometry the height can be calculated via $z = d \sin b$ from the distance d and the Galactic latitude b. Of course, if the distance is limited to a certain value, also the height cannot exceed that value. This explains the different ranges for the three distance ranges in Figure 5.17. In this case, the distance dependence is clearly visible for all parameters, especially at low heights, even for the reddening. This behavior is explained by the long light paths through the Galactic disk for high distances at low heights. Here, increasing the distance adds a lot of intervening interstellar gas as well as dust to the line of sight. On the other hand, for low distances (the blue curves) the height does not seem to play a role and no clear trends are observed. Since the distances are too low for the sight lines to leave the disk towards the halo, the interstellar parameters are essentially independent of direction there. The only feature observed at least for Na I and reddening at about $-1\,\mathrm{kpc} < z < 0\,\mathrm{kpc}$ again corresponds to the Taurus–Perseus–Auriga–California complex and is weaker for Ca II, because it is mostly found in warm partially ionized medium rather than in cold neutral clouds. Otherwise, the same patterns as before can be seen. For large $|z|$ values at high distances, at which the background stars are located in the halo, the measured parameters for Ca II barely fall below the average (dashed line), indicating the uniform distribution of Ca II even outside of the disk. For Na I and reddening, the measured values rapidly fall below the average, again indicating that gaseous Na I and interstellar dust is much more closely confined to the Galactic disk than gaseous Ca II.

5.5 Interstellar Gas to Dust Ratios

Another valuable characteristic of interstellar clouds is the ratio of their gas and dust contents. Therefore, the correlation between interstellar gas and dust is investigated in this section as well as the concrete gas to dust ratios.

5.5.1 Correlation Between Interstellar Line Strengths and Reddening

In Figure 5.18 the measured interstellar line parameters for both elements are plotted against the determined reddening. The data points are binned the same way as before and the medians of the parameter distributions are depicted as red dashed lines.

Especially for low reddening, for which also the line parameters are below average, a tightly confined and roughly linear correlation can be seen. It is even more pronounced for Na I than for Ca II, which again reflects the interstellar environments in which these elements are found. Ca II is predominantly found in warm partially ionized clouds, while Na I is predominantly found in cold neutral clouds, where also a lot of dust is expected, since dust grains survive and grow better in cold, dense, and neutral environments (Zhukovska et al., 2018; West et al., 2023). This means the abundances of Na I and dust are much more coupled than that of Ca II and dust. For larger reddening values, the measured line parameters begin to scatter a lot with large uncertainties. However, for Ca II it seems that they scatter around the average, while for Na I they scatter well above the average. In both cases the line parameters seem to saturate at higher reddening values. This linear correlation with subsequent saturation was already formerly realized for the Na I D doublet and attempts were made to model it (Richmond et al., 1994; Munari & Zwitter, 1997;

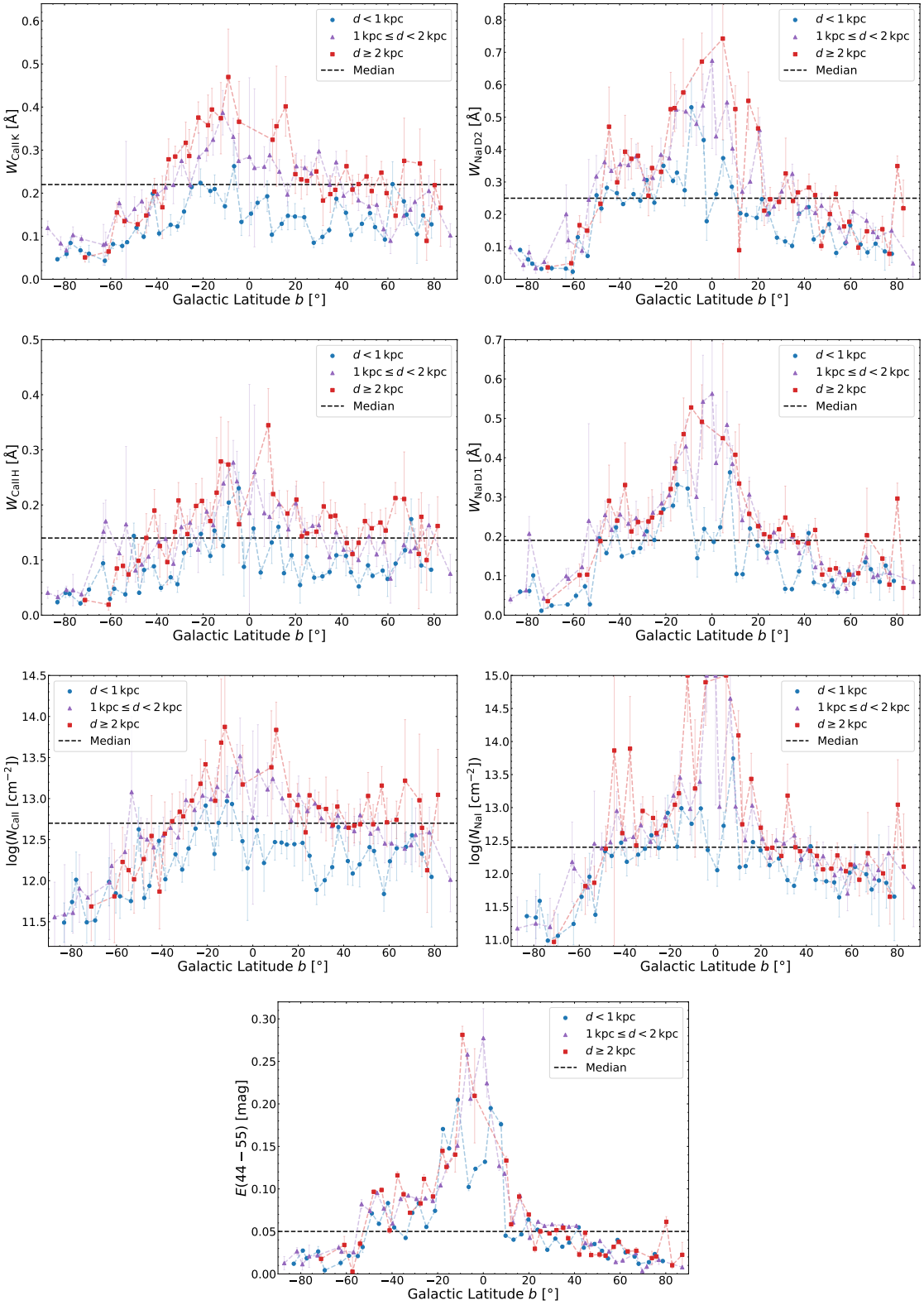


Figure 5.16: Dependence of the interstellar line parameters (equivalent widths and column densities) as well as reddening on the Galactic latitude and distance. The median of each parameter distribution is depicted as dashed line. The data is binned into three distance ranges with $d < 1\,\mathrm{kpc}$, $1\,\mathrm{kpc} \leq d < 2\,\mathrm{kpc}$, and $d \geq 2\,\mathrm{kpc}$. Each distance range contains 50 equally wide bins in x-direction. For the equivalent widths and column densities on average higher values are measured for larger distances, at least at lower Galactic latitudes. The reddening shows almost no distance dependence.

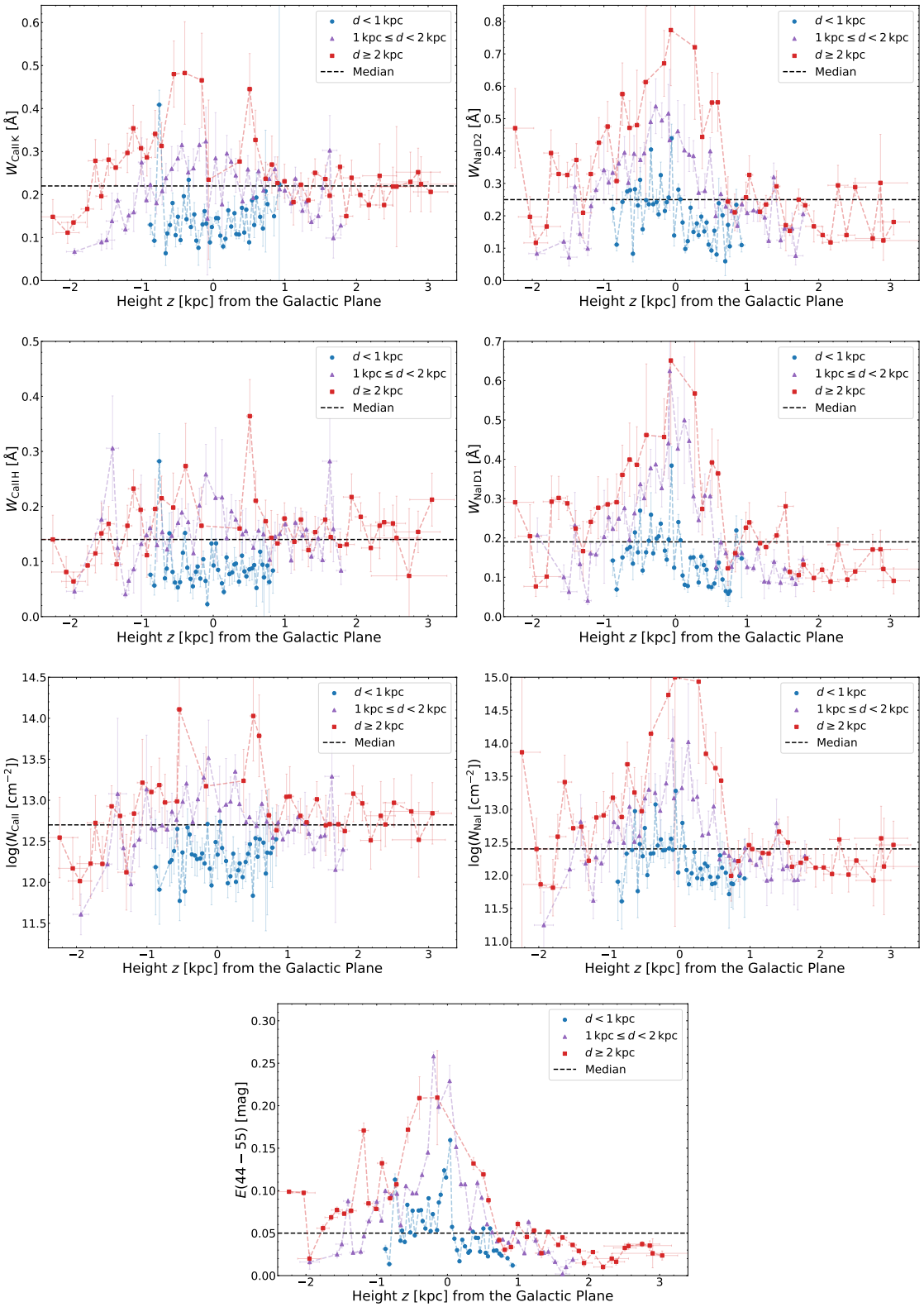


Figure 5.17: dependence of the interstellar line parameters (equivalent widths and column densities) as well as reddening on the height above and below the Galactic plane and distance. The median of each parameter distribution is depicted as dashed line. The data is binned into three distance ranges with $d < 1\,\mathrm{kpc}$, $1\,\mathrm{kpc} \leq d < 2\,\mathrm{kpc}$, and $d \geq 2\,\mathrm{kpc}$. Each distance range contains 50 equally wide bins in x-direction. On average higher values are measured for larger distances, especially at lower heights.

Poznanski et al., 2012). This would allow for a reddening estimate based on interstellar Na I D line strengths.

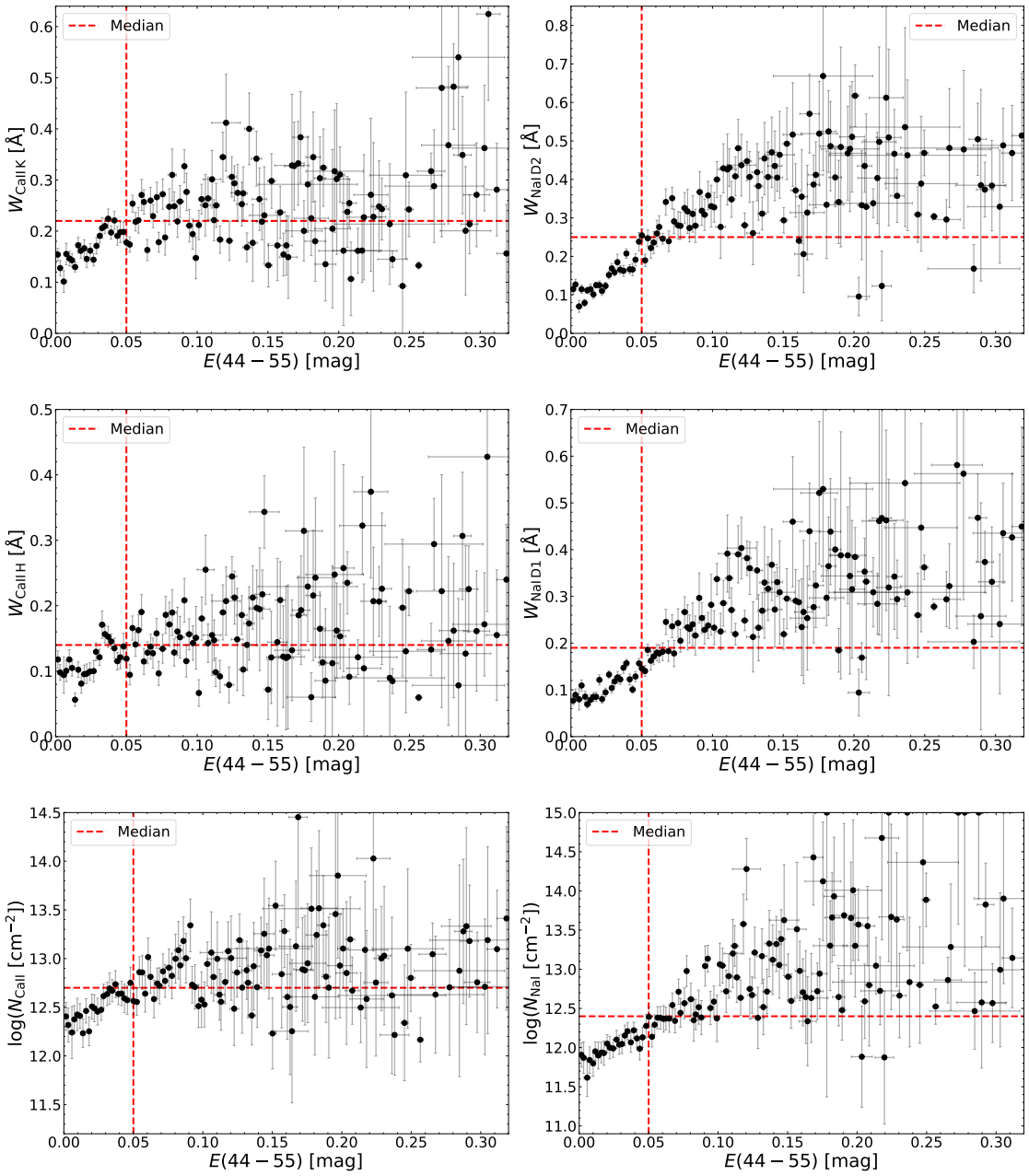


Figure 5.18: Correlation between interstellar line parameters (gaseous Ca II and Na I) and reddening (interstellar dust). The data is binned into 150 equally wide bins in x-direction and the medians of the respective parameter distributions are indicated with red dashed lines. For low reddening values a tight and roughly linear relation is observed, while for larger ones the scattering dominates and saturation sets in. The abundance of Na I seems to be more strongly coupled to interstellar dust than that of Ca II.

Unfortunately, their models show large discrepancies. Therefore, Poznanski et al. (2012) extensively discusses the complications for modeling the correlation from both, a theoretical and observational point of view. On the one hand, the theory is quite complex with a large number of proportionality factors that can all drastically vary in the Milky Way. On the other hand, the correlation highly depends on the interstellar physical conditions and the used background sources. As short summary, the sources of uncertainty are: The low ionization potential of sodium (5.1 eV), depletion of sodium on dust grains, multiple ISM components, as well as varying gas to dust ratios,

sodium abundances, and dust properties across the Galaxy. Overall they conclude that the simplest empirical, but still precise way to model the correlation is to take a linear relation between the logarithm of the color excess (reddening parameter) with respect to Na I D line equivalent widths. This approach is reproduced for the obtained data set of this work and shown in Figure 5.19. It comprises not only the individual equivalent widths of the Na I D doublet, but also the sum of both, which is stated as the most accurate relation in Poznanski et al. (2012). Furthermore, the same is tested for the Na I column densities as well as for Ca II equivalent widths and column densities, which are not considered in Poznanski et al. (2012). The resulting relations for the Na I D equivalent widths from the linear fits are

$$\log E(44-55) = (1.55 \pm 0.06) \times W_{\mathrm{NaI\,D2}} - (1.64 \pm 0.01) \tag{5.11}$$

$$\log E(44-55) = (1.95 \pm 0.08) \times W_{\mathrm{NaI\,D1}} - (1.61 \pm 0.01) \tag{5.12}$$

$$\log E(44-55) = (0.95 \pm 0.04) \times W_{\mathrm{NaI\,D1+D2}} - (1.67 \pm 0.01)\,, \tag{5.13}$$

while Poznanski et al. (2012) found[4]

$$\log E(B-V) = 2.16 \times W_{\mathrm{NaI\,D2}} - (1.91 \pm 0.15) \tag{5.14}$$

$$\log E(B-V) = 2.47 \times W_{\mathrm{NaI\,D1}} - (1.76 \pm 0.17) \tag{5.15}$$

$$\log E(B-V) = 1.17 \times W_{\mathrm{NaI\,D1+D2}} - (1.85 \pm 0.08)\,. \tag{5.16}$$

Although the trend seems in all cases to nicely match, the mean values of the slopes and offsets are off by many standard deviations. Also the pure statistical uncertainties on the offsets are much smaller here. Overall the mismatch can have several reasons: Firstly, the background sources used by Poznanski et al. (2012) are extragalactic quasars and galaxies, while in this work nearby stars in a volume of about 4 kpc around the Sun are used. So the average number of crossed interstellar clouds by the lines of sight might be higher in their sample than here. However, when more and more clouds are traversed, the measured reddening from increased dust column densities adds up linearly, while the line strengths begin to saturate at some point, if more gas is added to the line of sight. This observational bias is most likely one source of systematic deviation. Secondly, Poznanski et al. (2012) take the two dimensional line of sight reddening parameters from Schlegel et al. (1998) for their background sources, which are $E(B-V)$ values. Even though the deviation is small (Fitzpatrick et al., 2019), the monochromatic $E(44-55)$ values measured by the SED method (Section 3.4) are not identical. This is clearly another source of systematic bias. Still, except of the systematic offsets, the results match quite well. If no independent reddening measurements are possible for stars or other astronomical background sources, Equation 5.11 to Equation 5.13 represents a valid tool to estimate it from interstellar Na I D1&D2 absorption line equivalent widths. According to Poznanski et al. (2012), with this method the uncertainty on the derived reddening will be about 30%-40%.

For test purposes, the same procedure was applied for the Na I column densities as well as for Ca II equivalent widths and column densities. The resulting linear fits, slopes, offsets, and reduced χ^2 values are also shown in Figure 5.19. While the Ca II equivalent widths seem to decently work for this purpose as well, even though the scattering of the data points and the uncertainties on the fit parameters are a bit larger than for Na I, column densities are obviously not suited in both cases. The uncertainties, in particular on the obtained offsets, are just too large. Overall, using Na I equivalent widths instead of Ca II ones still seems to be the better choice. This can be seen from the reduced χ^2 values alone, which are much closer to unity for Na I, which indicates that the linear model fits better there.

5.5.2 Variation of Galactic Gas to Dust Ratios

What is left to investigate are the Galactic gas to dust ratios. In the literature they are typically expressed as $R_{\mathrm{GtD}} \coloneqq N_{\mathrm{H}}/E(B-V)$ in units of $10^{21}\,\mathrm{cm}^{-2}\,\mathrm{mag}^{-1}$, where N_{H} is the total hydrogen

[4] For some reason they did not give uncertainties on the fitted slopes.

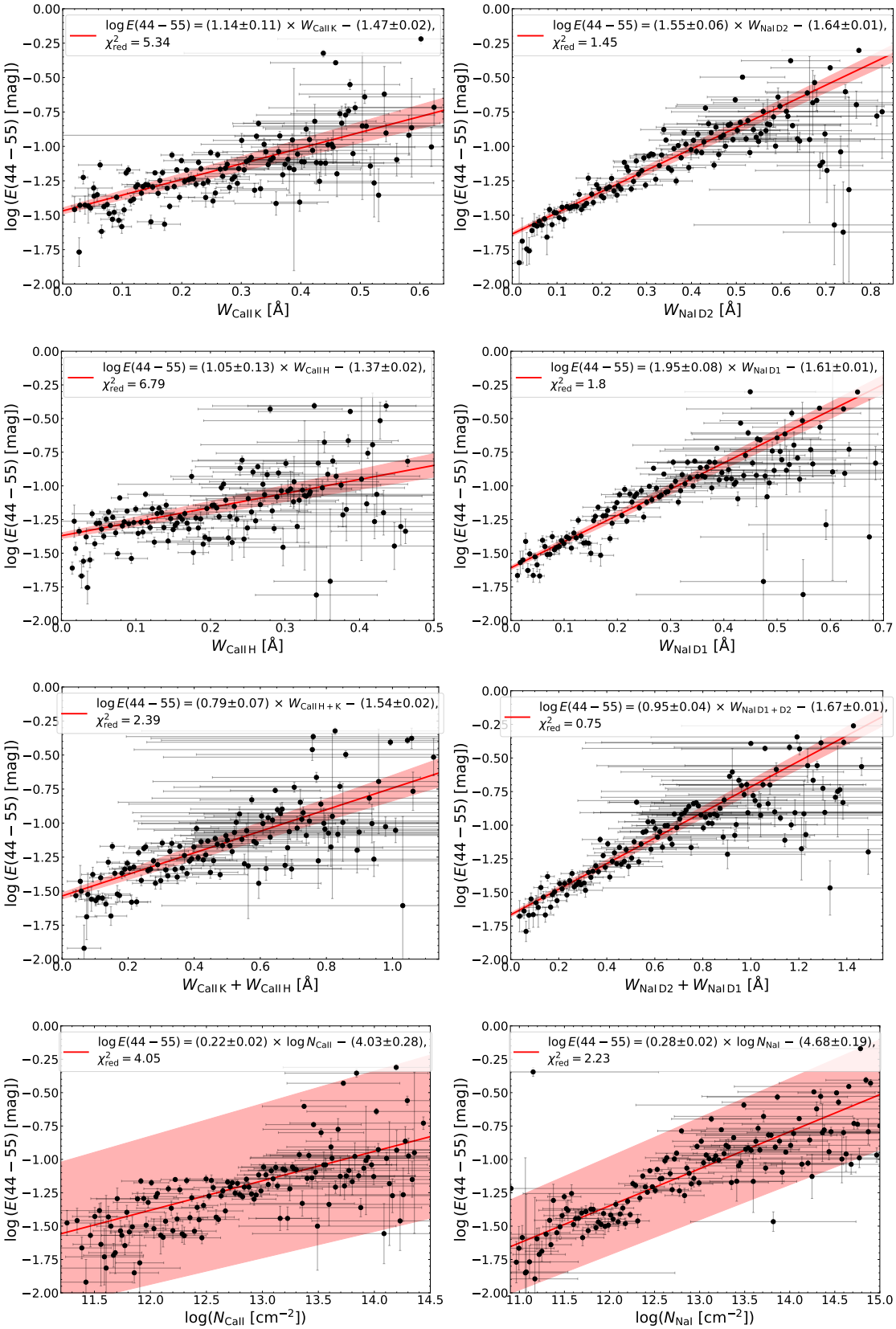


Figure 5.19: Logarithm of the reddening parameter versus interstellar line parameters with linear fits (red solid lines). The 1σ uncertainty interval is depicted as red contour. The derived linear relations as well as the corresponding reduced χ^2 values are given in the labels of each panel. The relations are constrained best for Na I equivalent widths. Ca II ones also give decent results, while those for the column densities are poor in both cases.

column density (neutral hydrogen H I and molecular one H_2). On average[5] they are not expected to vary a lot throughout the Milky Way disk, but with Galactic latitude as interstellar dust is closely confined to the disk. For the Galactic disk typical values of about $6 \times 10^{21}\,\mathrm{cm^{-2}\,mag^{-1}}$ have been found (Savage et al., 1977; Bohlin et al., 1978; Rachford et al., 2009; Gudennavar et al., 2012; Shull et al., 2021; Liszt & Gerin, 2023). For higher Galactic latitudes, however, somewhat larger values of about $8 - 10 \times 10^{21}\,\mathrm{cm^{-2}\,mag^{-1}}$ are observed (Liszt, 2014a,b; Lenz et al., 2017; Liszt & Gerin, 2023; Shull & Panopoulou, 2024). For the determination of these ratios, various kinds of background sources have been used, like early type stars and active galactic nuclei (AGNs), combined with radio observations of neutral hydrogen. Hydrogen column densities were derived either using the UV absorption in the spectra of the background sources or the 21 cm radio emission, while reddening was either obtained from stellar photometry or far infrared dust emission.

In this work, however, there is no direct access to hydrogen column densities, since only that of Ca II and Na I are measured. Nevertheless, the gas to dust ratios $R_{\mathrm{GtD,CaII,NaI}} := N_{\mathrm{CaII,NaI}}/E(44-55)$ for those elements are calculated in units of $10^{12}\,\mathrm{cm^{-2}\,mag^{-1}}$. The distributions of both ratios can be seen in Figure 5.20, where they are shown in logarithmic scale due to the large dispersions. Both distributions exhibit a constrained peak with tails towards higher values. These tails can most likely be explained with measurements at high Galactic latitudes, where the dust content rapidly falls off leading to high gas to dust ratios. Towards lower values only the ratios for Na II show a little tail. This is also reasonable, because it means the dust content dominates over the gaseous Na I there, which is only possible in cold neutral clouds, where both can coexist. Gaseous Ca II is predominantly found in warm partially ionized clouds, where only little amounts of dust are expected, while in colder clouds calcium is depleted onto dust grains. This explains the sharp cut towards smaller values for the Ca II ratios. Both distributions have been statistically evaluated (also see Figure 5.20) and the resulting average parameters are

$$\log \bar{R}_{\mathrm{GtD,CaII,mode}} = 2.1^{+0.4}_{-0.9} \qquad \text{and} \qquad \log \bar{R}_{\mathrm{GtD,CaII,median}} = 2.0^{+0.7}_{-0.6} \tag{5.17}$$

$$\log \bar{R}_{\mathrm{GtD,NaI,mode}} = 1.6^{+0.7}_{-0.6} \qquad \text{and} \qquad \log \bar{R}_{\mathrm{GtD,NaI,median}} = 1.7^{+0.8}_{-0.5}\,. \tag{5.18}$$

Now it would be desirable to compare these results with the above mentioned gas to dust ratios for hydrogen. Unfortunately, there is no certain way to reliably convert Ca II and Na I column densities to hydrogen ones. Murga et al. (2015) investigates how the Ca II and Na I abundances vary with hydrogen abundance in the interstellar medium and try to find a relation to convert Ca II and Na I equivalent widths to hydrogen column densities. However, this procedure is highly uncertain, especially for different interstellar environments, and only applicable for weak lines before saturation sets in. Another simple, but more straight forward way is to just scale the Ca II and Na I column densities with solar abundances to estimate the hydrogen column density. Even though interstellar abundances can in specific cases deviate considerably from solar ones, this approach might be on average not too bad and worth a try. According to Asplund et al. (2021) the solar abundances of calcium and sodium are

$$\frac{N_{\mathrm{Ca}}}{N_{\mathrm{H}}} \approx 2.00 \times 10^{-6} \qquad \text{and} \qquad \frac{N_{\mathrm{Na}}}{N_{\mathrm{H}}} \approx 1.66 \times 10^{-6}\,. \tag{5.19}$$

So when taking the median values of Equation 5.17 and Equation 5.18, scaling them with Equation 5.19, and converting them to linear scale, one obtains

$$\bar{R}_{\mathrm{GtD,H(CaII)}} = (5.00^{+0.25}_{-0.20}) \times 10^{19}\,\mathrm{cm^{-2}\,mag^{-1}} \tag{5.20}$$

$$\bar{R}_{\mathrm{GtD,H(NaI)}} = (3.02^{+0.38}_{-0.19}) \times 10^{19}\,\mathrm{cm^{-2}\,mag^{-1}} \tag{5.21}$$

for the hydrogen column density gas to dust ratios. Those are both more than two orders of magnitude off from the literature values, which means they clearly must be wrong. The assumption of solar abundances for the interstellar medium, combined with a very simplified approach not taking into account ionization fractions and dust depletion of the elements, seems to be very poor.

[5] Locally, they might change a lot with different interstellar environments.

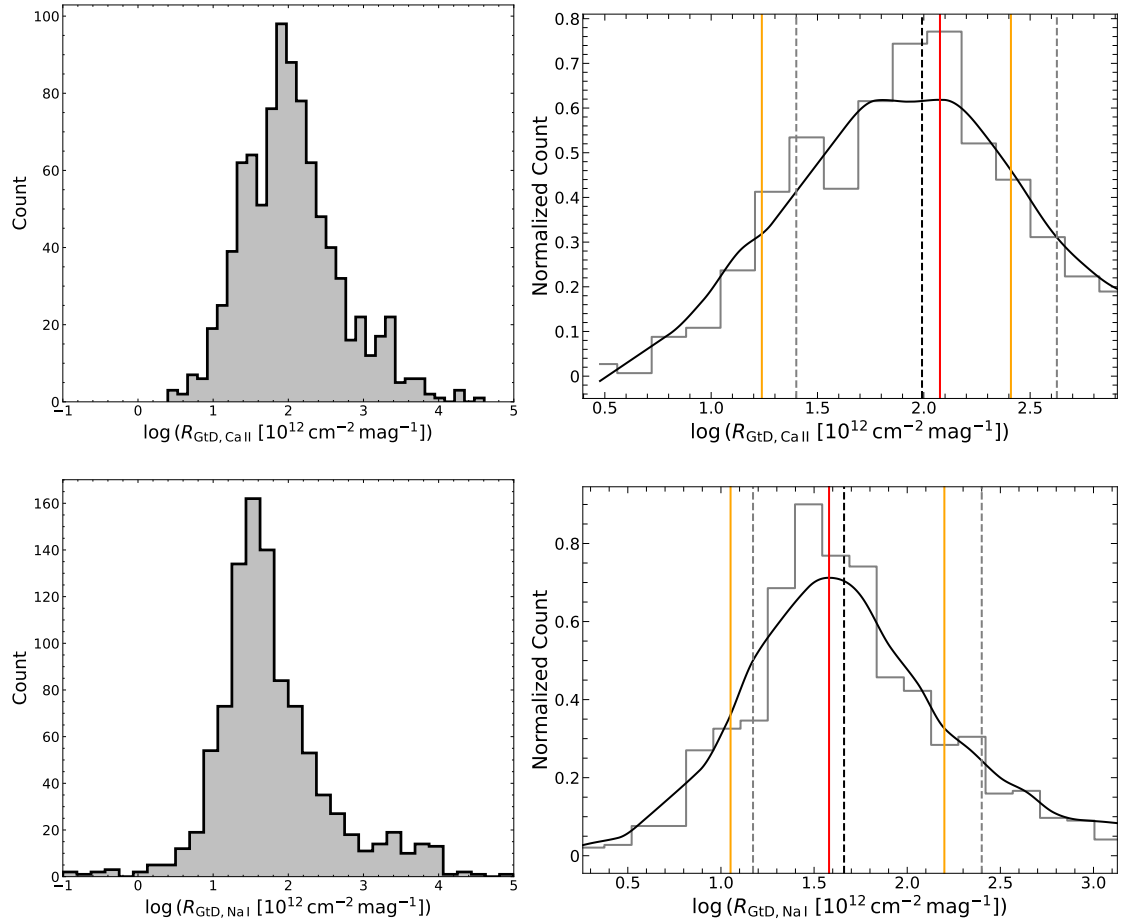


Figure 5.20: Distributions of the determined interstellar gas to dust ratios for Ca II and Na I (left panels) and their statistical evaluations (right panels). The red and yellow vertical lines depict the modes and HDIs of $\log \bar{R}_{\mathrm{GtD,CaII,mode}} = 2.1^{+0.4}_{-0.9}$ and $\log \bar{R}_{\mathrm{GtD,NaI,mode}} = 1.6^{+0.7}_{-0.6}$, respectively, while the dashed vertical lines depict the medians and quantiles of $\log \bar{R}_{\mathrm{GtD,CaII,median}} = 2.0^{+0.7}_{-0.6}$ and $\log \bar{R}_{\mathrm{GtD,NaI,median}} = 1.7^{+0.8}_{-0.5}$, respectively.

Therefore, the only alternative is to indeed use the relations of Murga et al. (2015) to convert Ca II H&K and Na I D1&D2 equivalent widths to hydrogen column densities:

$$N_{\mathrm{H}} = N_0 \cdot W_{\mathrm{x}}^{1/\alpha_0}\,, \tag{5.22}$$

where N_{H} is the hydrogen column density, W_{x} the equivalent width of the respective line and N_0 and α_0 are fit parameters determined by Murga et al. (2015) for each line[6]. But from uncertainties on the order of the mean values themselves[7], it can be seen that the results should be treated with caution. Additionally, the hydrogen column densities for the same observations calculated from the four different equivalent widths sometimes show a strong mismatch of up to two orders of magnitude, although they should be roughly the same. For this reason it did not appear reasonable to average the four hydrogen column densities derived from the interstellar lines of each spectrum and they are considered individually. In Figure 5.21 and Figure 5.22 the distributions of the four hydrogen gas to dust ratios are shown derived from the Ca II and Na I equivalent widths, respectively. They were calculated the same way as the calcium and sodium gas to dust ratios and

[6] The fit parameters N_0 and α_0 for each line are listed in Table 1 of Murga et al. (2015).

[7] The uncertainties were calculated with simple Gaussian error propagation, including the uncertainties on the equivalent width as well as on the fit parameters N_0 and α_0.

are in units of $10^{21}\,\mathrm{cm}^{-2}\,\mathrm{mag}^{-1}$. The statistical evaluations gave

$$\log \bar{R}_{\mathrm{GtD,H(CaII\,K),mode}} = 0.5^{+0.6}_{-0.8} \quad \text{and} \quad \log \bar{R}_{\mathrm{GtD,H(CaII\,K),median}} = 0.4^{+0.6}_{-0.8} \tag{5.23}$$

$$\log \bar{R}_{\mathrm{GtD,H(CaII\,H),mode}} = 0.5^{+0.7}_{-0.8} \quad \text{and} \quad \log \bar{R}_{\mathrm{GtD,H(CaII\,H),median}} = 0.5^{+0.7}_{-0.9} \tag{5.24}$$

$$\log \bar{R}_{\mathrm{GtD,H(NaI\,D2),mode}} = 0.8^{+0.3}_{-0.5} \quad \text{and} \quad \log \bar{R}_{\mathrm{GtD,H(NaI\,D2),median}} = 0.7^{+0.4}_{-0.4} \tag{5.25}$$

$$\log \bar{R}_{\mathrm{GtD,H(NaI\,D1),mode}} = 0.9^{+0.3}_{-0.4} \quad \text{and} \quad \log \bar{R}_{\mathrm{GtD,H(NaI\,D1),median}} = 0.8^{+0.3}_{-0.4}. \tag{5.26}$$

It can be seen that the mode and median values match within each line doublet, but disagree across the two elements. Additionally, the dispersion is much stronger for calcium, while the hydrogen column densities derived from the sodium lines are more closely confined. Physically, this would mean that the abundances of Na and H in the interstellar medium are more tightly bound than that of Ca and H. This pattern is also found and discussed in Murga et al. (2015). Within the 1σ dispersion the average values of the hydrogen column densities still match across the elements and converting them to linear scale roughly gives

$$\bar{R}_{\mathrm{GtD,H(CaII)}} = (3.2^{+12.7}_{-\ 2.8}) \times 10^{21}\,\mathrm{cm}^{-2}\,\mathrm{mag}^{-1} \tag{5.27}$$

$$\bar{R}_{\mathrm{GtD,H(NaI)}} = (6.3^{+9.6}_{-4.4}) \times 10^{21}\,\mathrm{cm}^{-2}\,\mathrm{mag}^{-1}. \tag{5.28}$$

Unfortunately, the uncertainties are too high to make clear conclusions or to distinguish between disk and halo gas to dust ratios. But at least the 1σ parameter range matches well with the literature values ($\bar{R}_{\mathrm{GtD,H}} = 6 - 10 \times 10^{21}\,\mathrm{cm}^{-2}\,\mathrm{mag}^{-1}$).

In order to not only find average values for interstellar gas to dust ratios, but to also investigate the variations with different stellar environments, it is continued in the following with the Ca II and Na I gas to dust ratios directly obtained from the analysis pipeline. Even though the uncertainties are still high, they are much better constrained than the hydrogen ones and are not determined under so many sources of uncertainties.

In Figure 5.23 the variation of the two gas to dust ratios is shown over distance, Galactic latitude, and height above and below the Galactic plane. Overall, the patterns are not very distinctive. For Ca II the dependence on the distance seems to be more pronounced than for Na I. It was already observed before that Ca II is more uniformly distributed towards larger distances, especially at high Galactic latitudes, than Na I and interstellar dust. This is apparently also seen in the top panels of Figure 5.23 for the ratios of the gas and dust abundances. In the mid panels the dependence on the Galactic latitude is shown, where the South is again affected by the observational bias and, hence not considered. Towards the North pole the gas to dust ratios seem to increase, in particular for Ca II, which again confirms that interstellar gas is less closely confined to the Galactic disk than interstellar dust. For latitudes close to zero a large scattering is visible. This most likely reflects the variety of different interstellar environments within the Galactic disk. Some clouds contain more gas than dust, some vice versa, and for some clouds the gas to dust ratio is almost constant. The lower panels show the dependence on the height above and below the Galactic disk. While for Na I the data only scatters around the average value, for Ca II a clear trend is visible: At low height dust dominates over gas (low gas to dust ratios) as visible from the dip around zero and at large heights the gas dominates over dust (increasing gas to dust ratios). This confirms the results from before as well.

5.6 Comparison With Line of Sight Dust Maps

The SED fitting routine (Section 3.4) does not only measure the reddening, but also a comparison with the dust maps from Schlegel et al. (1998) and Schlafly & Finkbeiner (2011) for the respective line of sight is implemented. Both are two dimensional maps compiled from infrared dust emission and dust extinction of starlight, respectively. Incorporating the predictions from more recent three dimensional dust maps like Green et al. (2019) is planned for the future. In Figure 5.24 the predictions from both maps for the sight lines of this sample are shown. The distributions look similar to Figure 5.6.

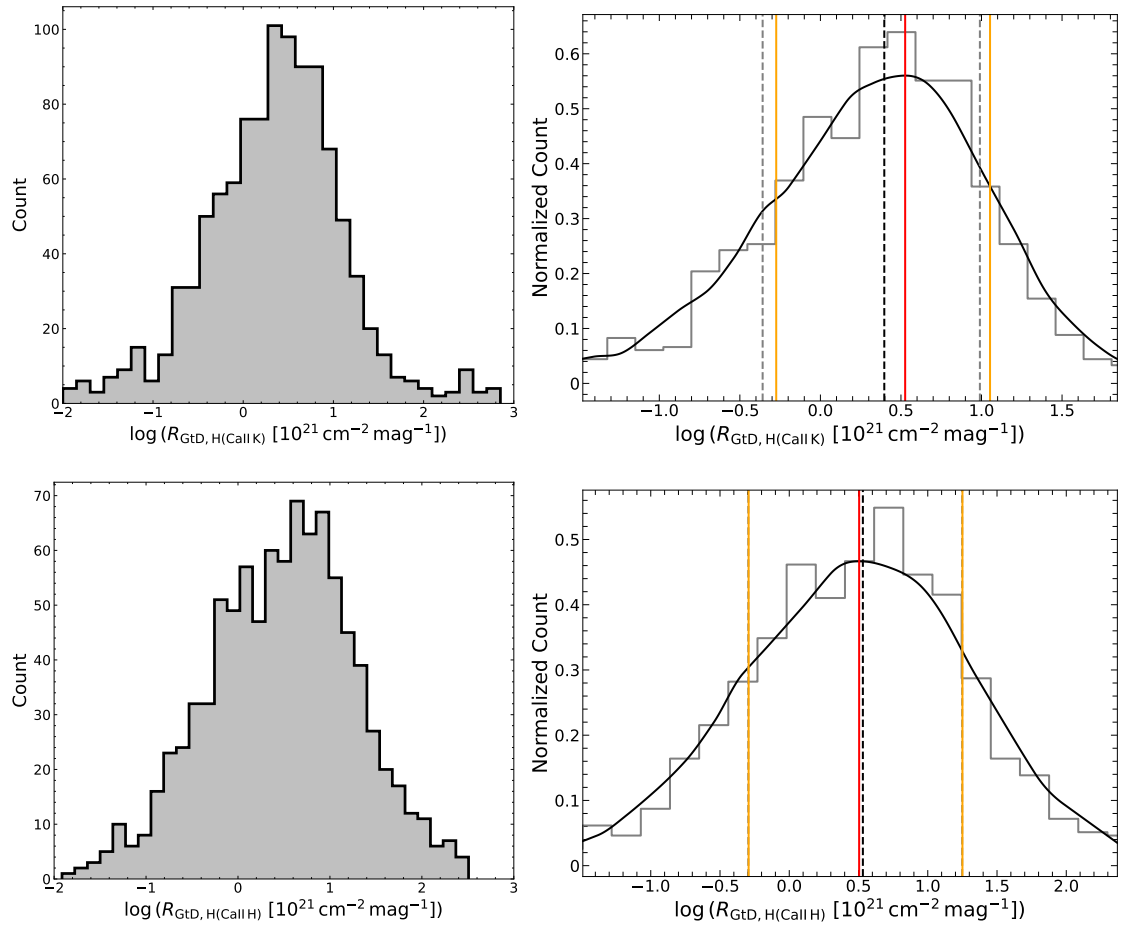


Figure 5.21: Distributions of the determined hydrogen gas to dust ratios derived from the Ca II H&K line equivalent widths using Equation 5.22 (left panels) and their statistical evaluations (right panels). The red and yellow vertical lines depict the modes and HDIs of $\log \bar{R}_{\mathrm{GtD,H(CaII\,K),mode}} = 0.5^{+0.6}_{-0.8}$ and $\log \bar{R}_{\mathrm{GtD,H(CaII\,H),mode}} = 0.5^{+0.7}_{-0.8}$, respectively, while the dashed vertical lines depict the medians and quantiles of $\log \bar{R}_{\mathrm{GtD,H(CaII\,K),median}} = 0.4^{+0.6}_{-0.8}$ and $\log \bar{R}_{\mathrm{GtD,H(CaII\,H),median}} = 0.5^{+0.7}_{-0.9}$, respectively.

The question is, if the measured and predicted reddening values align. Therefore, in Figure 5.25 the difference between the measured E(44-55) and the mean[8] E(B-V) from both maps is shown and statistically evaluated. The resulting values from this sharply peaking distribution are:

$$\bar{E}_{\mathrm{diff,mode}} = 0.000^{+0.023}_{-0.018}\,\mathrm{mag} \qquad \text{and} \qquad \bar{E}_{\mathrm{diff,mode}} = 0.002^{+0.023}_{-0.019}\,\mathrm{mag} \tag{5.29}$$

On average, predictions and measurements are in astonishing good agreement, which is seen from the average values very close to zero. This impressively showcases the accuracy of the reddening estimates via SED modeling. However, this result is perhaps even slightly more relevant for the search for circumstellar material (CSM) than for the investigation of the interstellar medium. Assuming that dust is produced within the ejected material after common envelope ejection, one could think about a possible reddening excess compared to ISM predictions (Li et al., 2025), meaning a higher measured than predicted value. Figure 5.25 at least demonstrates that there is no systematic offset towards higher SED reddening. Nevertheless, the distribution is slightly skewed towards positive values. This could possibly indicate CSM around some hot subdwarfs. However, other explanations seem to be more likely: On the one hand, the limited spatial resolution of dust maps might not account for localized dense structures in the ISM. On the other hand, the

[8] It is actually known that both maps slightly deviate from each other. However, as rough estimate considering the mean of both is sufficient.

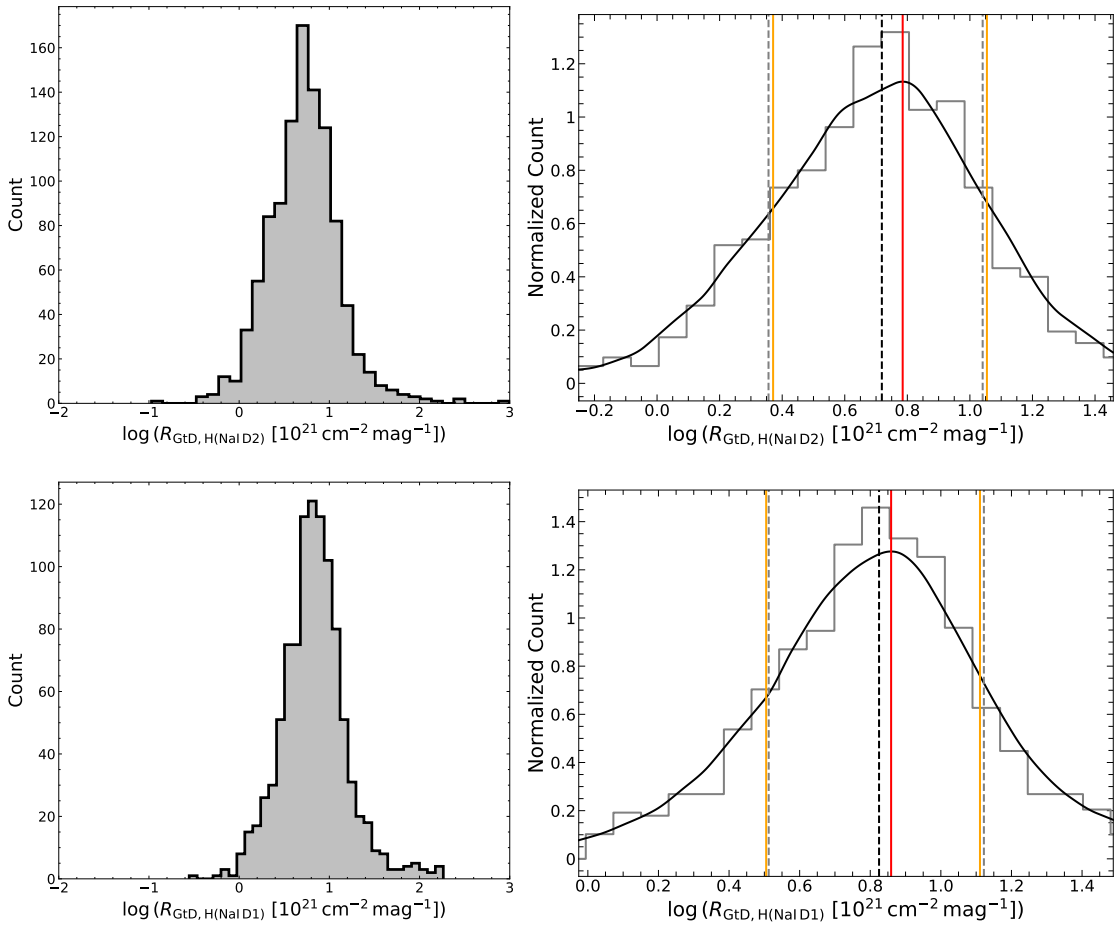


Figure 5.22: Distributions of the determined hydrogen gas to dust ratios derived from the Na I D1&D2 line equivalent widths using Equation 5.22 (left panels) and their statistical evaluations (right panels). The red and yellow vertical lines depict the modes and HDIs of $\log \bar{R}_{\mathrm{GtD,H(NaI\,D2),mode}} = 0.8^{+0.3}_{-0.5}$ and $\log \bar{R}_{\mathrm{GtD,H(NaI\,D1),mode}} = 0.9^{+0.3}_{-0.4}$, respectively, while the dashed vertical lines depict the medians and quantiles of $\log \bar{R}_{\mathrm{GtD,H(NaI\,D2),median}} = 0.7^{+0.4}_{-0.4}$ and $\log \bar{R}_{\mathrm{GtD,H(NaI\,D1),median}} = 0.8^{+0.3}_{-0.4}$, respectively.

monochromatic $E(44-55)$ reddening is slightly larger than the broad band counterpart E(B-V) (Fitzpatrick et al., 2019). This might be a natural explanation for these little reddening excesses. The idea of reddening excess as indicator for CSM is discussed more extensively in Chapter 6.

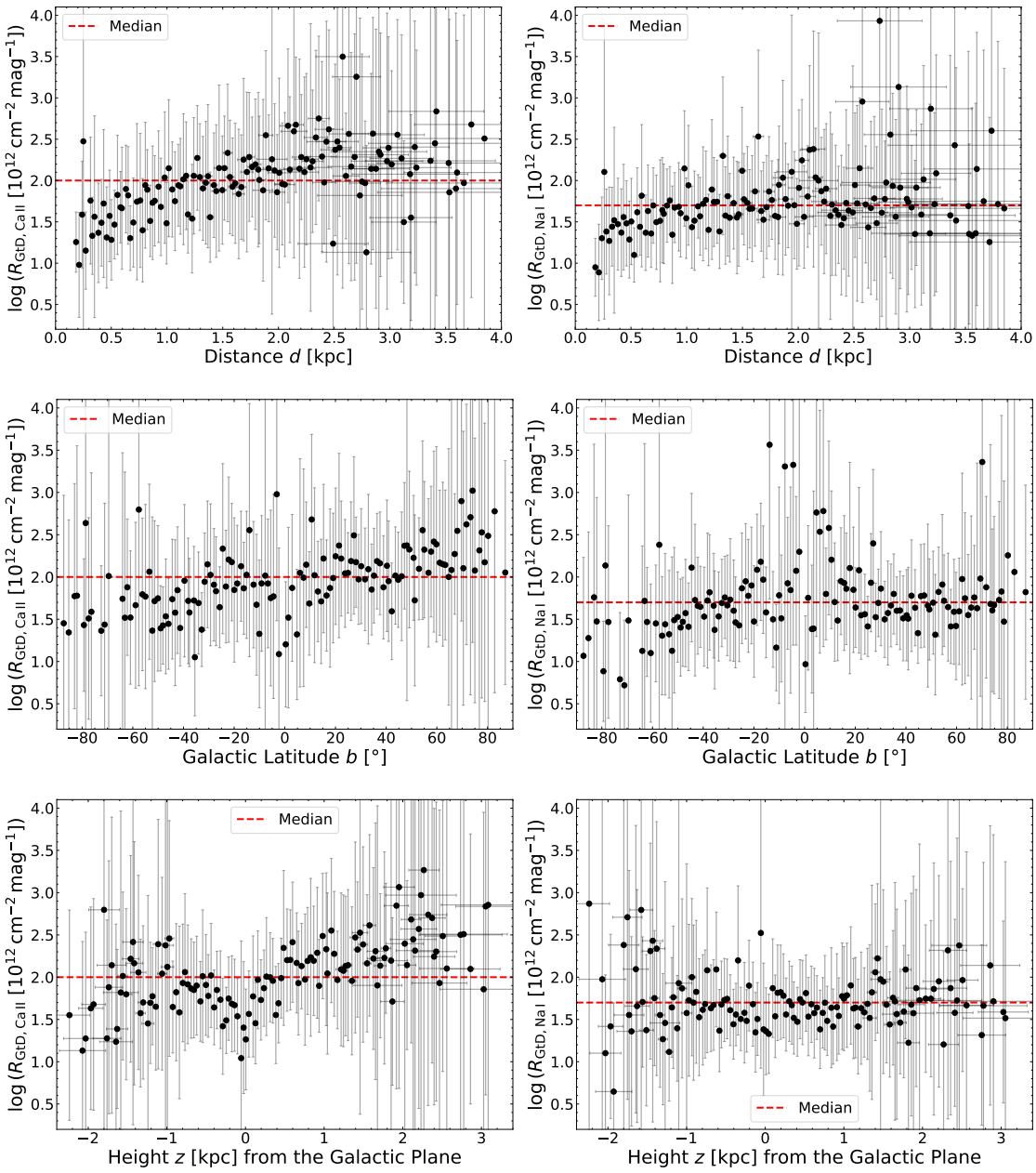


Figure 5.23: Variation of the interstellar gas to dust ratios of Ca II (left panels) and Na I (right panels) with distance, Galactic latitude, and height above and below the Galactic plane. The average value of each parameter distribution is depicted with a red dashed line, respectively.

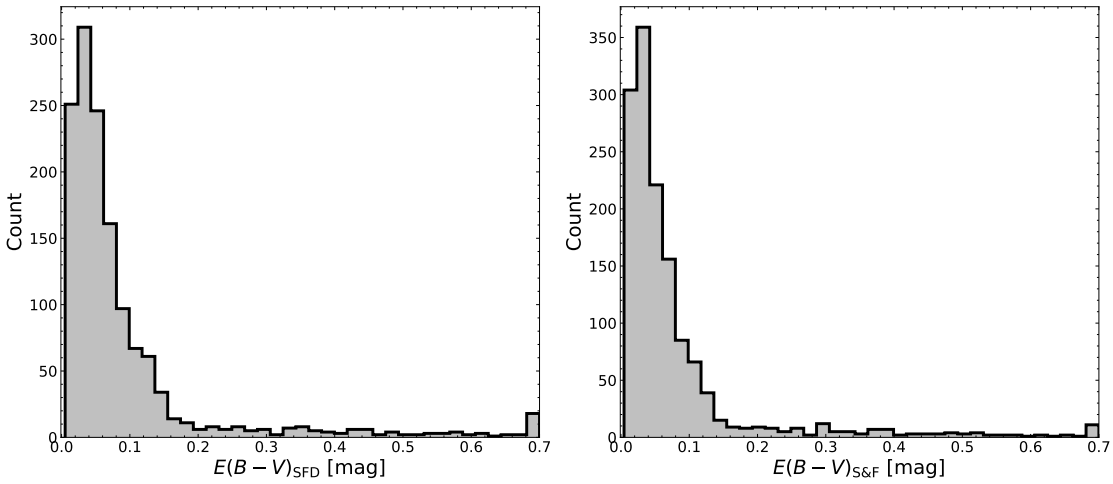


Figure 5.24: Reddening predictions from the 2D dust maps of Schlegel et al. (1998) (left panel) and Schlafly & Finkbeiner (2011) (right panel) for the sight lines of this sample. Both distributions look similar to the measured reddening values of Figure 5.6.

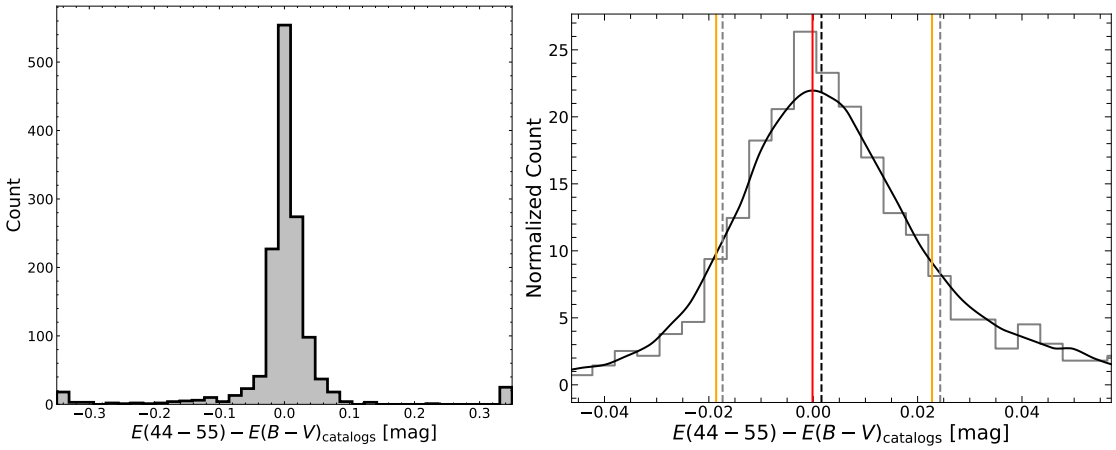


Figure 5.25: Difference between the measured SED reddening $E(44-55)$ and the mean predictions $E(B_V)$ from Schlegel et al. (1998) and Schlafly & Finkbeiner (2011). The distribution peaks sharply at zero, which means measurements and predictions agree well. The statistical evaluation confirms this result: $\bar{E}_{\mathrm{diff,mode}} = 0.000^{+0.023}_{-0.018}$ mag and $\bar{E}_{\mathrm{diff,mode}} = 0.002^{+0.023}_{-0.019}$ mag.

Chapter 6

Search for Circumstellar Material (CSM)

Many formation channels of hot subdwarf stars include common envelope (CE) evolutions (see Subsection 1.3.2). Inside the common envelope the orbit of the binary shrinks mainly because of friction and drag forces, which leads to a loss of orbital energy and momentum transferred to the surrounding envelope. As consequence, the envelope expands, becomes gravitationally unbound from the binary, and is finally ejected. However, due to the transient nature of this evolutionary stage, direct observations of binaries in the CE phase have been elusive yet. But it is expected that some of the ejected material might remain bound around the binary as circumstellar material (CSM). As it expands it becomes thinner and thinner and cools down, eventually mimicking the conditions of the diffuse interstellar medium analog to the ejecta of supernova explosions. As the light of the binary passes through this CSM, the same Ca II and Na I absorption lines might emerge in the spectrum as for the ISM. Therefore, it can be searched for this material using the same methods as before for the investigation of the interstellar medium. While there was no clear observational evidence for CSM around hot subdwarf stars coming from CE ejection up to now, Li et al. (2025) claims the detection of ubiquitous CSM around hot subdwarf stars based on Ca II K line strengths and reddening.

Therefore, this chapter makes a transition from the investigation of the interstellar medium to the search for this circumstellar material as second main aim of this work. First, an overview of the physics of CE evolution and ejection is given in Section 6.1. In Section 6.2 the detections of Li et al. (2025) are summarized and assessed by comparing them to the results of this work. Finally, in Section 6.3 it is searched among the sample of hot subdwarf stars of this thesis for the presence of CSM.

6.1 Overview

When stars have burned most of their fuel in the core, they expand. This is a general result of stellar evolution. During these late stages of stellar evolution like the red giant or asymptotic giant branch (RGB and AGB, respectively, see Subsection 1.3.1), stellar radii can reach sizes of several astronomical units. If such stars resides in binary systems and the orbital separation is wide enough, both components evolve without interacting with each other. However, if they are close enough, the giant component begins to transfer mass to its companion at a rate that might be too large for the companion to digest. Then, the accreting star swells until the envelopes of both stars meet and begin to engulf the whole system. The result is called a common envelope and was first introduced by Paczynski (1976).

The following evolutionary phase is one of the most poorly understood ones in binary stellar evolution (Ivanova et al., 2013a, 2020). Due to the transient nature of the common envelope phase, which is thought to be on the order of only a thousand or at most a few thousand years (Izzard et al., 2012), direct observations of a CE have been elusive yet. Even if such a system would be observed, it might mimic the properties of only one ”ordinary“ star, making it hard to distinguish it

from normal single stars. There is only indirect evidence for the CE phase due to the existence of close binary systems, whose orbits shrank inside of the CE (Iben & Livio, 1993), as well as that of objects like hot subdwarf stars, which can hardly be explained with single star evolution (Baschek & Norris, 1975; Mengel et al., 1976; Heber, 2016). One direct evidence for CEs could be the detection of CSM around such systems. In order to gain insights into the physical processes involved in the formation of such CSM, the current understanding of CE ejection is discussed in more detail in the following.

6.1.1 The Physics of Common Envelope Evolution and Ejection

In the single star evolution case for low mass objects the expanding envelope of a giant star is eventually lost by radiation-driven stellar winds, while in the more massive case the star explodes in a supernova. In binary systems this scenario is often complicated by the presence of the binary companion and a common envelope is formed. The question is, what happens to both components inside of the CE and which mechanisms ultimately lead to the CE ejection possibly leaving behind circumstellar material?

The details are actually far from being well understood. However, the common answer is that the binary orbit inside of the CE decays because of friction and a loss of co-rotation between the core and envelope (Podsiadlowski, 2001). This decay increases as the components descend into ever denser material, which is called the plunge phase (Izzard et al., 2012). Additional to the frictional energy, energy and angular momentum is lost from the orbit and transferred to the surrounding CE very efficiently. As consequence, the envelope heats up and expands until it eventually becomes unbound to the binary and ejected. This ejection leads to a sudden decrease of the total mass and density of the CE reducing the drag forces and thus, the rate of orbital energy loss, which slows down the inspiral of both components. In fact, other energy sources like magnetic fields (Tout & Regos, 2003), energy from the recombination of ions and electrons (Han et al., 1995; Dewi & Tauris, 2000), and nuclear burning (Ivanova & Podsiadlowski, 2002; Podsiadlowski et al., 2010) might also contribute to or even dominate the CE ejection.

A common way to parametrize the CE evolution is the so-called α-formalism first introduced by Webbink (1984). It treats the energy release during the CE phase as fraction α of the change in orbital energy $\Delta E_{\rm orb}$. This release of orbital energy $\alpha\Delta E_{\rm orb}$ is assumed to be responsible for the expansion and ejection of most of the CE. However, it needs to overcome the binding energy

$$E_{\rm bind} = \frac{GM_{\rm env}M_{\rm core}}{\lambda R} \tag{6.1}$$

of the envelope, where G is the gravitational constant, $M_{\rm env}$ the mass of the envelope, $M_{\rm core}$ the core mass of the primary star, R its radius, and λ a dimensionless constant further quantifying the binding energy. The latter one depends on the stellar density distribution (Dewi & Tauris, 2000) and is usually on the order of unity (Izzard et al., 2012). So besides masses, radii, and orbital properties, the main parameters describing the CE evolution and ejection are α and λ. The product $\alpha\lambda$ is used to calculate, for example, the final separation of the binary after the CE ejection. Furthermore, the final outcome of the system depends on the time, when the CE is ejected. If the orbit becomes too small inside of the CE, the binary might merge before the ejection. Hence, the lower the efficiency of the energy transfer (small α) and the higher $E_{\rm bind}$ (small λ), the longer the CE evolution lasts and the more likely it is that both components merge. This means a small $\alpha\lambda$ indicates the merging of the binary inside of the CE, while on the other hand a high $\alpha\lambda$ indicates that two distinct objects remain after the CE ejection.

In principle, it is possible to calculate λ from detailed stellar models (Dewi & Tauris, 2000; Loveridge et al., 2011; Xu & Li, 2010; Ivanova, 2011). However, this is rather uncertain and the values of λ can differ by an order of magnitude (Tauris & Dewi, 2001; Ivanova & Chaichenets, 2011). Most studies assume a constant value of either 0.5 or 1.0. Even more difficult and uncertain is the estimation of the α parameter. Due to the lack of direct common envelope observations, this is at the moment only possible via simulations.

6.1.2 Predictions from Simulations and Observations

A lot of effort is put into detailed numerical simulations of common envelope evolution[1]. In general, there are two approaches: Either one dimensional stellar evolution modeling (Podsiadlowski, 2001) or full three dimensional hydrodynamical modeling (Taam & Ricker, 2010). In the 1D case the common envelope phase can be followed from the initial mass transfer to the final ejection, but it lacks of multi-dimensional geometric details. On the other hand, the 3D approach is ideal for modeling the dynamical plunge phase, but can not follow the evolution beyond that. Such simulations suggest values of α between 0.1 and 1.0 for most of the configurations[2]. However, each stage in the models is subject to uncertainties, which often leads to an accumulated uncertainty on α on the order of one magnitude. Overall, it can be said that common envelope evolution is far from being well understood and the resulting parameters from simulations are anything but constrained. Moreover, the final outcome highly depends on the initial configuration, i.e. the (core) masses of both components, their radii, and orbital parameters. Also not only hot subdwarf stars are formed through common envelope channels, but it is thought to be a central phase in many cases of binary stellar evolution. If the CE evolution is fast enough, it can result into all kinds of close binary systems, while if it is not, both components might merge into various kinds of single objects. Concerning hot subdwarf stars, the expected outcomes in the former case are short period binary systems with low mass main sequence or white dwarf companions (see Figure 1.6). According to Politano et al. (2008) also the CE merger scenario seems possible, where for example a RGB star merges with a low mass M-dwarf or even substellar object (brown dwarf or giant planet), leading to single sdB stars. In both cases circumstellar material might persist after the CE ejection, ready to be detected[3].

Unfortunately, all these numerical simulations do not directly predict the properties like temperature, density, dust content, or distance from the host system of the possibly remaining CSM after CE ejection. Only Ivanova et al. (2013b) discuss the possibility that the phenomenon of luminous red novae (e.g., Pastorello et al., 2019) could be explained with CE ejection events and find temperatures for the ejecta of about 5,000 K. From the observational side, Steinmetz et al. (2024) found a circumstellar environment around the red nova V1309 Sco with a temperature of $5,000 - 15,000$ K for the gaseous component. Hence, also for the expected CSM around hot subdwarf stars a temperature in that range can be assumed. Furthermore, Li et al. (2022) claim the detection of CSM around J1920-2001, a short period sdO + WD binary system, and find a gas density of about $6 \times 10^{-10}\,\mathrm{g\,cm^{-3}}$. They base their detection on strong Ca II H&K lines, but at the same time rather low reddening (derived from the reddening map of Green et al. (2019)) and argue that this is in discrepancy to the relation between reddening and Ca II equivalent widths of Megier et al. (2005). However, on the one hand the system has a distance of about 7 kpc, which makes the reddening prediction from dust maps very uncertain due to the finite spatial resolution, and on the other hand, the relation of Megier et al. (2005) was derived for smaller distances. According to the findings of Chapter 5, gaseous Ca II is more uniformly distributed towards greater distances and the Galactic halo, while dust is closely confined to the Galactic disk. This means the strong Ca II H&K lines are not necessarily in disagreement with the (rather uncertain) low reddening. Also they find a radial velocity for the lines of about $30\,\mathrm{km\,s^{-1}}$, while the systemic velocity of the binary is about $200\,\mathrm{km\,s^{-1}}$ and explain it with the expansion of the ejected CE material with about $-170\,\mathrm{km\,s^{-1}}$. But since a radial velocity of $30\,\mathrm{km\,s^{-1}}$ is still roughly in agreement with ISM radial velocities (Figure 5.5), an interstellar origin of the lines seems much more likely, even though the CSM case can also not be entirely excluded.

[1] See, for example, Passy et al. (2012); Ivanova & Nandez (2016); Clayton et al. (2017); Iaconi et al. (2018); Kramer et al. (2020); Sand et al. (2020); Glanz & Perets (2021); González-Bolívar et al. (2022). A review on numerical simulations of the common envelope phase is given by Röpke & De Marco (2023)

[2] If only energy transferred to the envelope by the loss of angular momentum (decay of the orbit) is considered, the natural range for the efficiency is $0 < \alpha < 1$. Practically, due to recombination and nucleosynthesis, it is actually possible to exceed values of 1.

[3] Also in the merger scenario of, e.g., two He-WDs leading to the formation of He-sdOs (see Subsection 1.3.2) a certain circumstellar environment is expected, even without a common envelope phase.

6.2 Recent Detections of CSM Around Hot Subdwarf Stars

Apart from the one (highly uncertain) detection of CSM around a binary system containing a hot subdwarf component by Li et al. (2022), Li et al. (2025) claim the recent detection of ubiquitous CSM around hot subdwarf stars mainly based on LAMOST spectroscopy and SED modeling. Exactly the same spectra and objects have been investigated in the LAMOST sample of this work (see Subsection 2.2.2 and Chapter 5). Therefore, in this section the results of both studies are compared with each other in order to see, whether the findings of Li et al. (2025) are reproduced.

6.2.1 Summary

The spectra of Li et al. (2025) are taken from the seventh data release (DR7) of LAMOSTs low resolution survey, using the catalog of hot subdwarf (sdB) star candidates from Luo et al. (2021), which is based on the *Gaia* DR2 catalog of Geier et al. (2019). Luo et al. (2021) already excluded composite spectra by visual inspection, which is crucial, because F/G/K-type companions might contribute with photospheric calcium and sodium lines, making it hard to distinguish them from potential CSM lines, exactly like in the ISM case before. Furthermore, Luo et al. (2021) determined the atmospheric parameters of the stars. Li et al. (2025) applied a signal to noise ratio (SNR) cut[4] of better than 30 and finally ended up with a sample of 727 sdB stars. The methods and findings of their analysis are the following:

- They applied Gaussian fits to the Ca II K line in order to derive equivalent widths and radial velocities. The Ca II H line is not considered due to the blend with the Hε line (see Subsection 3.3.4, in particular Figure 3.4). Other absorption lines are also not considered for the central part of their analysis. After visual inspection of the spectra, they find 623 out of 727, which clearly exhibit the Ca II K line.

- Using *Gaia* eDR3 coordinates and parallax distances, they derive predicted $E(B-V)$ values from the 3D Bayestar19 dust maps of Green et al. (2019) as measure for the amount of interstellar dust along the line of sight to the stars. Furthermore, by taking the sample of 284 OB stars from Megier et al. (2009), they model a linear relation between the Ca II K line equivalent widths and reddening values for these stars[5]:

$$W_{\mathrm{CaII\,K}} = (0.52 \pm 0.06) \times E(B-V) + (0.07 \pm 0.01) \tag{6.2}$$

  This relationship is stated as the predicted behavior for absorption lines purely originating from the ISM. According to Li et al. (2025), Ca II K equivalent widths in the spectra of the sdB stars exceeding this trend indicate the presence of additional circumstellar material. It is the case for 188 out of 623 stars exhibiting Ca II K lines.

- Moreover, they argue that the reddening values from large-scale dust maps do not account for dense, localized structures around the stars, while the $E(B-V)$ from SED fits (see Section 3.4) does. Hence, they model the SEDs for all sdB stars in their sample and compare the resulting reddening to the dust map predictions. 145 out of the previous 188 CSM candidates exceed that predictions and are treated as the final CSM candidate sample. The difference $E(B-V)_{\mathrm{SED}} - E(B-V)_{\mathrm{map}}$ is taken as the observed reddening caused by dust contained in the CSM[6]. None of their SEDs showed infrared excess indicating the presence of a cool companion.

- The radial velocity differences $v_{\mathrm{rad,CaII\,K}} - v_{\mathrm{rad}}$ between the Ca II K line and the stars are investigated. Their distribution shows a peak around zero with long tails towards high

[4] In the analysis of this work a SNR cut of better than 40 is applied.

[5] The uncertainties are 1σ ones.

[6] The question is, if dust is actually produced during a common envelope ejection. But since a large fraction of the total interstellar dust content in the Galaxy is produced by ejections from AGB stars during their late evolutionary stages and by supernova explosions (e.g., Gupta & Sahijpal, 2020), it can probably also be assumed for CE ejections.

(positive as well as negative) differences. On the one hand, they conclude with long-lived CSM gravitationally bound to the system due to the agreement of the radial velocities for many spectra. On the other hand, they explain the long tails with still expanding or inwards drifting CSM.

- Gaussian fits are also applied to the Na I D lines. The equivalent widths are plotted over the reddening and the relations of Poznanski et al. (2012) and Murga et al. (2015) (for both, see Section 5.5) are fitted to the results. Even though there are some high equivalent widths outliers exceeding the relations, most CSM candidates do not show coincident high Ca II K and Na I D line equivalent widths. Li et al. (2025) conclud that neutral sodium is unlikely to be present near stars with effective temperatures of $T_{\rm eff} > 20,000\,\mathrm{K}$, because it becomes easily ionized. The outliers with high equivalent widths are explained with Na I in outer regions of the CSM shielded by dust (Lallement et al., 2003).

- Finally, even though they discuss that high-resolution spectroscopy is needed to distinguish between the ISM and the potential CSM component, they conclude that their study represents a clear detection of CSM remaining after CE ejection around sdB stars. They even try to constrain the mass, radius, density, and shape of the material.

6.2.2 Comparison with the Results of This Work

Exactly the same parameters as in Li et al. (2025) were determined during the analysis of this work, albeit with slightly different methods. Therefore, the results for the LAMOST objects can be readily compared to those from Li et al. (2025). The aim is, of course, to see if the parameters match or if there are any (systematic) deviations.

SED reddening plays an important role in the CSM criteria of Li et al. (2025). Since the derived atmospheric parameters are required in order to model the SEDs and to obtain information about reddening, the atmospheric parameters are checked first. In Figure 6.1 these and the stellar radial velocities from the spectral fits of Luo et al. (2021) are plotted against those of this work. It can be seen that the helium abundance and the radial velocity (lower panels) match quite well with little scattering, except of a few outliers. The large uncertainties and increasing scattering for low $\log y$ values is natural, since it is measured from helium lines in the spectra, which are no longer visible for low helium abundances in low resolution spectra. The surface gravity (upper right panel) scatters much more, because it is often the least well constrained parameter in spectral fits. Perhaps a slight offset towards higher values in Li et al. (2025) is visible, but it does not seem to be significant. The effective temperature (upper left panel), however, seems to deviate systematically, at least for lower values of $T_{\rm eff} \lesssim 32,000\,\mathrm{K}$. Their temperatures are on average $2,000 - 3,000\,\mathrm{K}$ higher than those of this work in that range. There seems to be a little discrepancy between the models of Luo et al. (2021) and here, but at the moment it can not be said, where this disagreement comes from or which models perform better. Unfortunately, the effective temperature is strongly correlated with the derived SED reddening. This will become clear in a moment.

Next, the measured Ca II K line parameters as well as the reddening values are compared. First of all, surprisingly, although Luo et al. (2021) excluded composite spectra by visual inspection and Li et al. (2025) did not find any infrared excess in their SEDs indicative for a cool main sequence companion in a wide orbit, 92 objects are flagged as binaries from the analysis of this work. Such companions might contribute with strong photospheric Ca II absorption. Also such systems are thought to be formed through stable Roche lobe overflow, not a CE channel (see Figure 1.6), which means no CSM is expected. Furthermore, it is not clear, which relation Li et al. (2025) exactly applied to find the CSM candidates. If equivalent widths larger than 0.2 Å (see Figure 4 in Li et al. (2025)) are taken that exceed the 3σ range of their ISM relation, where at the same time a reddening excess, $E(B-V)_{\rm SED} - E(B-V)_{\rm Map} > 0$, is observed, one is left with 147 objects. These are two more than their 145 candidates, but still considered as the candidate sample in the following. 30 out of the 147 candidates are flagged as binaries with F/G/K-type companions from the SEDs in this work.

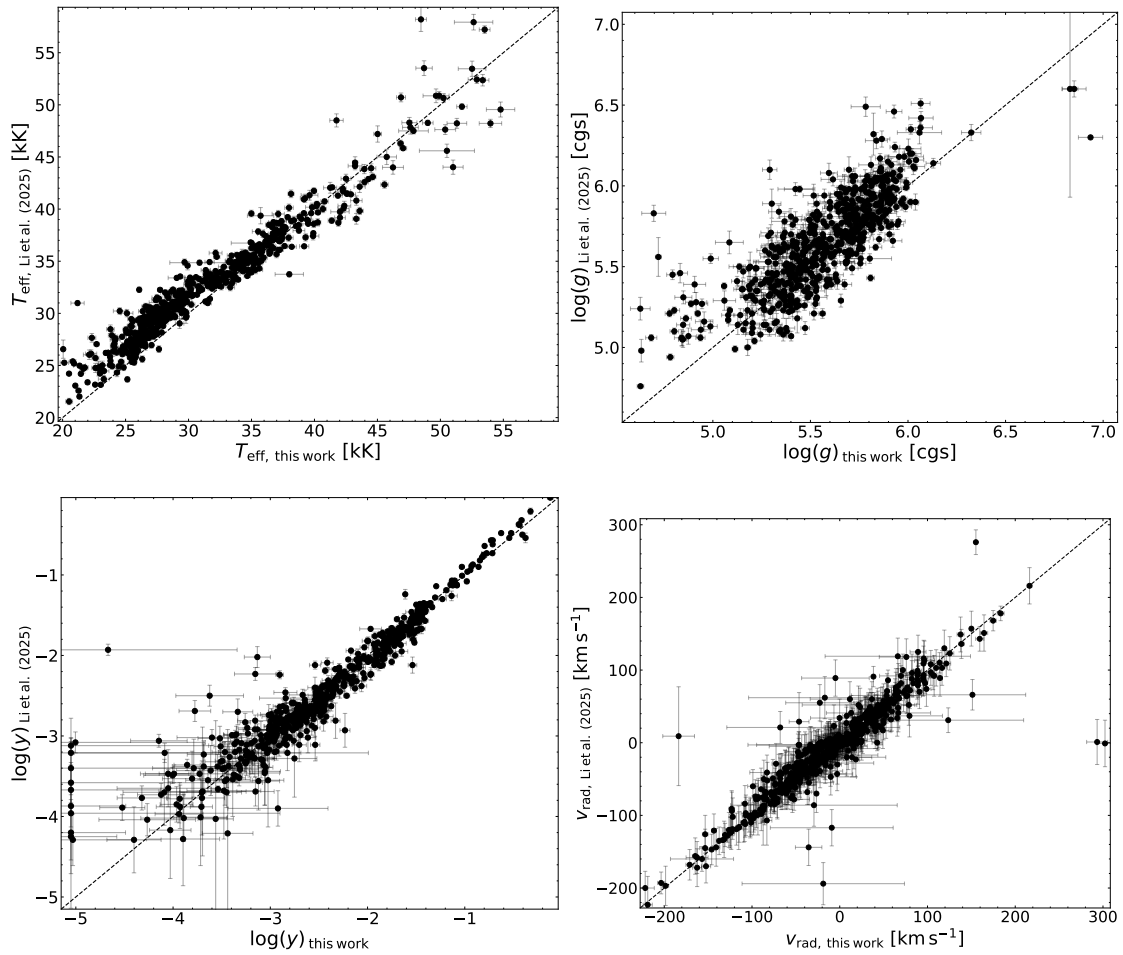


Figure 6.1: Comparison between the atmospheric parameters and radial velocities derived by Luo et al. (2021) and used by Li et al. (2025) with that of this work. The helium abundance $\log y$ and the radial velocity v_{rad} seems to match well. The surface gravity $\log g$ shows more scattering and perhaps a little offset, which is most likely negligible. The effective temperature T_{eff}, however, shows systematically slightly higher temperatures in their works, at least for cooler stars.

In Figure 6.2, the SED reddening, the reddening excess, the equivalent widths, and the radial velocities of the Ca II K lines are compared. It is distinguished between all stars (black dots), apparently single CSM candidates (blue dots) and the 30 composite CSM candidates (red dots). It can be seen that the measured reddening parameters from the SEDs of both studies (upper left panel) on average match[7], except of a some observations towards systematically higher values in Li et al. (2025). These can be explained with the deviation of the effective temperatures, which are passed to the SED fits. The higher the prescribed temperature, the more UV/blue flux the synthetic SEDs expect. But if the observed photometric data indicates a lower flux, the fits try to suppress the model flux by introducing enhanced reddening. This correlation between temperature and reddening is illustrated in Figure 6.3. However, even though it explains the systematically deviating measurements in general, only some CSM candidates are clearly affected.

The reddening excess values (upper right panel) scatter around zero for both studies. This is expected, since for a pure ISM component the SED and dust map reddening should roughly match. The individual measurements do not deviate too much, except of the same systematically higher values in Li et al. (2025) as for the SED reddening. However, a non-negligible number of CSM candidates have a negative reddening excess, $E(44-55)_{\mathrm{SED}} - E(B-V)_{\mathrm{Map}} < 0$, in this work. This contradicts the criteria for CSM candidates of Li et al. (2025) for these stars. Notably, the three

[7] Even though the SED reddening from Li et al. (2025) shows extremely low uncertainties, perhaps unrealistically small ones.

objects with the most extreme reddening excesses are identified as wide binaries in this work, for which no CSM is expected. Two of them show hydrogen emission lines in their spectra, indicating the presence of an accretion disk, and the other one shows an SED, which is completely outshone by the cooler companion, combined with only a small number of photometric measurements. Thus, these objects should certainly be excluded from the sample.

Except of some scattering, but at the same time also large uncertainties, the Ca II K line equivalent widths (lower left panel) seem to match. The extreme observations are for the same objects as before. Only a little trend towards higher values in this work is visible, which most likely comes from the different methods. Li et al. (2025) determines the equivalent widths from Gaussian fits, while in this analysis it is integrated over the absorption lines (see Subsection 3.3.4).

The Ca II K line radial velocity comparison (lower right panel) looks similar to the equivalent width one. Even though there is a large scattering, the overall trend matches, except of a few outliers. But with an average uncertainty of about $30\,\mathrm{km\,s^{-1}}$, which is quite common for measurements from only one spectral line in a low resolution spectrum (see Subsection 3.3.4 and Section 4.3), the radial velocities are not well constrained. Anyway, the apparent Ca II K lines, even if there is a CSM contribution, also contain the ISM component. In low resolution spectra, even if the radial velocity difference between both components comprises several tens of $\mathrm{km\,s^{-1}}$, they are blended and can not be distinguished. Therefore, it is not reasonable to identify the measured radial velocities with that of the CSM and they are no longer considered, e.g. to argue that they match with the stellar radial velocities, which could be another sign for CSM, as it is done in Li et al. (2025).

6.2.3 Updated Sample of CSM Candidates

Apart from leftover binaries with composite SEDs, also observations among the Li et al. (2025) sample are found that are flagged with either a bad parallax measurement or a poor SED fit in this work. This can become very critical for the identification of CSM candidates, since the prediction from the dust maps are based on the 3D positions of the stars, which depend on the parallax distances. On the other hand, poor SED fits might lead to spurious SED reddening values. Hence, it makes sense to remove those observations as well as the binaries from the sample. Furthermore, the CSM candidates are reduced to only those, where the Ca II K equivalent widths exceed 0.2 Å and the reddening excesses are positive in both analyses (additional to the ISM relation criterion). Overall, one is left with a total sample of 424 out of originally 623 stars, among which 49 out of originally 145 stars are CSM candidates. Hence, the number of potential CSM candidates decreases drastically.

In Figure 6.4 the same comparison plots as in Figure 6.2 are shown for the cleaned sample, where the CSM candidates are depicted in red. It can be clearly seen that many extreme or mismatching measurements got removed. Figure 6.5 represents analog plots to Figure 3 and Figure 4 of Li et al. (2025), but with the results of this work instead. The left panel shows the dependency of the Ca II K equivalent width on the dust map reddening together with the linear ISM relation of Equation 6.2, which Li et al. (2025) derived from the sample of Megier et al. (2005). In the right panel the reddening excess is plotted over the Ca II K equivalent width. The reduced sample of CSM candidates (depicted in red) are exactly those with $W_{\mathrm{CaII\,K}} > 0.2$ Å, exceeding the 3σ range of the ISM relation, and showing a positive reddening excess. All three criteria need to be fulfilled in both studies for a higher statistical significance.

6.2.4 Conclusion

So what are the final conclusions about the recent detections of circumstellar material around hot subdwarf stars by Li et al. (2025)? First, it was found that their sample is still contaminated with a large number of wide binaries, visible from composite SEDs, which are not expected to show CSM, because of their formation through stable Roche lobe overflow instead of CE evolution. Furthermore, many observations were flagged with bad parallax measurements or poor SED fits. Excluding all these cases results in a total sample of 424 stars, a decrease of more than 30%. As

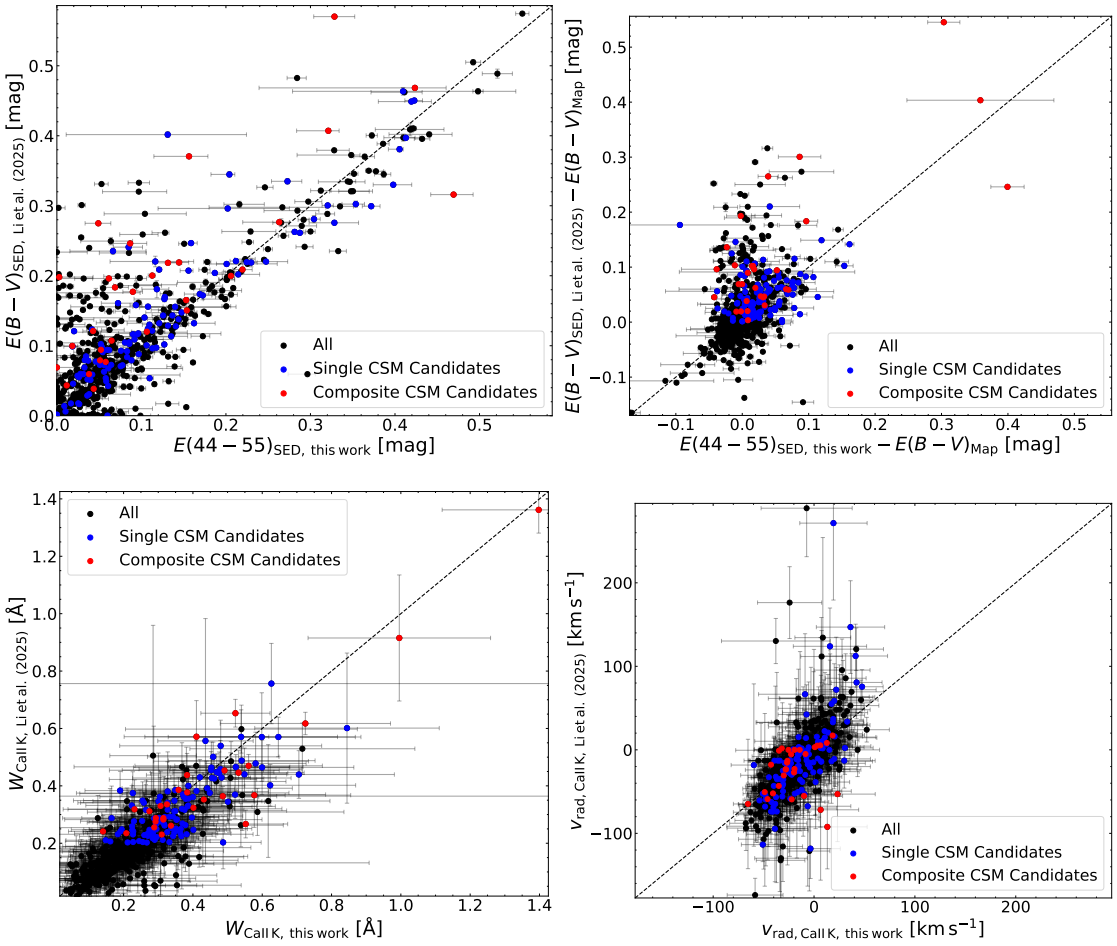


Figure 6.2: Comparison of the measured SED reddening (upper left panel), the reddening excess (upper right panel), the Ca II K equivalent width, and the Ca II K radial velocity (lower right panel) between Li et al. (2025) and this work. The whole sample is shown as black dots, the apparently single CSM candidates as blue dots, and the 30 composite SED CSM candidates as red dots. Despite some scattering, overall a matching trend is visible in all cases. Only the SED reddening as well as the reddening excess shows a bunch of systematically higher values for the Li et al. (2025) measurements. These can be explained with the temperature discrepancy of Figure 6.1, because higher temperatures passed to the SED fits lead to enhanced reddening values (Figure 6.3).

criteria for the identification of CSM candidates Li et al. (2025) uses high Ca II K equivalent widths, equivalent widths that are beyond the ISM relation of Equation 6.2, and positive reddening excesses. Applying these criteria to the reduced sample of stars and neglecting measurements, where the parameters are in disagreement between both analyses, one ends up with 49 CSM candidates. These are more than 65% less than in Li et al. (2025). The question is, does this final sample represent good candidates that are likely surrounded by circumstellar material left from common envelope ejection? To answer this question, the methods and criteria of Li et al. (2025) need to be assessed from the point of view of this thesis:

- **Strong Ca II K lines:** The first criterion for the identification of CSM candidates suggested by Li et al. (2025) is that the Ca II K lines present in the spectra of the sdB stars show high equivalent widths of $W_{\mathrm{CaII\,K}} > 0.2$ Å. However, Li et al. (2025) neglects the large uncertainties on the measurements (see Figure 6.2, Figure 6.4, or Figure 6.5), which are according to the publication on average 0.23 Å. Also the slightly lower SNR cut of 30 compared to that of 40 in this work might introduce further enhancing of the uncertainties (see Subsection 4.3.3).

- **The ISM relation:** Another central criterion by Li et al. (2025) is that Ca II K equiva-

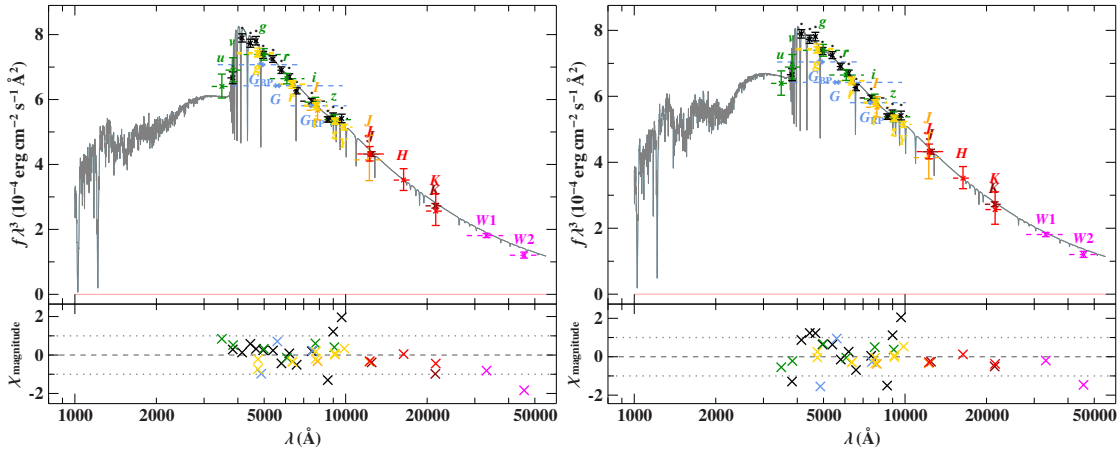


Figure 6.3: SED fits for an arbitrary sdB star with an effective temperature of about 23,400 K. If this temperature is passed to the SED fit, it leads to the result shown in the left panel and a reddening parameter of $E(44-55) \approx 0.023$ mag. However, if the temperature is artificially increased by 4,000 K, the SED fit expects a higher flux in the UV/blue regime as shown in the right panel. If this flux is not reproduced by the observed photometric data, the program tries to suppress it by introducing an enhanced reddening of $E(44-55) \approx 0.058$ mag. This is more than twice the value from the true temperature and highlights the correlation between temperature and reddening.

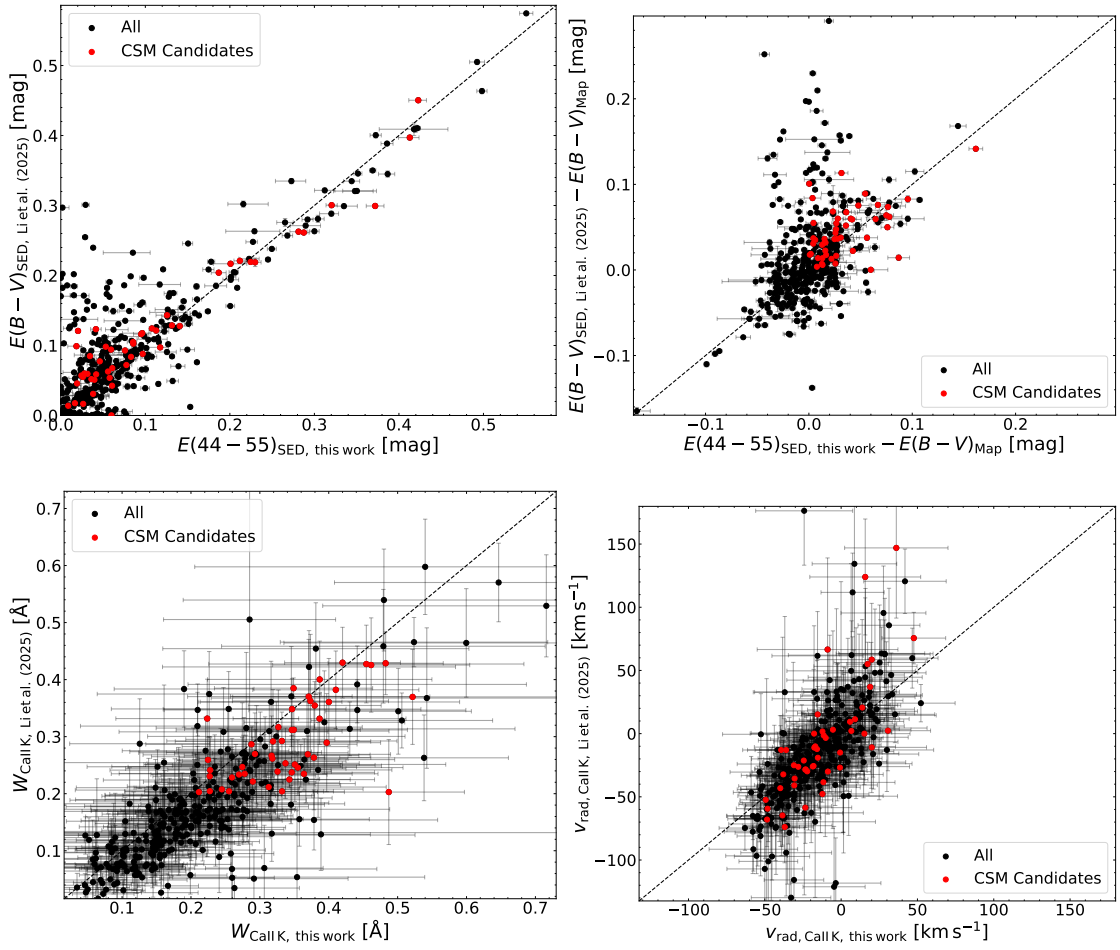


Figure 6.4: Same as Figure 6.2, but for the cleaned sample, where binaries and observations with bad parallax measurements or poor SED fits are removed. Also the sample of CSM candidates depicted in red is reduced to those stars, where the parameters used for their identification match in both studies.

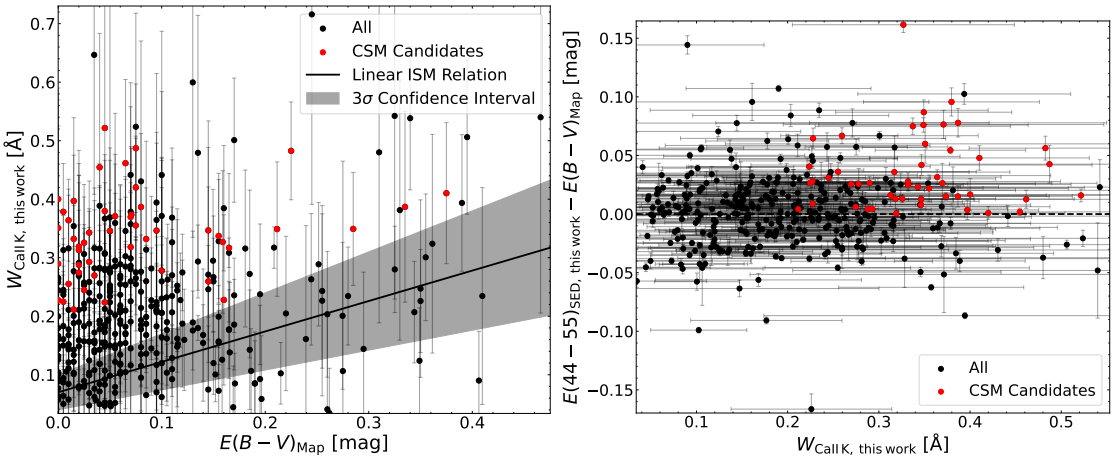


Figure 6.5: Analog plots to Figure 3 and Figure 4 of Li et al. (2025), but with the results of this work. The CSM candidates are depicted in red. *Left panel:* Dependency of the Ca II K equivalent width on the dust map reddening together with the linear ISM relation of Equation 6.2 and its 3σ confidence interval, which Li et al. (2025) derived from the sample of Megier et al. (2005). *Right panel:* Reddening excess over Ca II K equivalent width.

lent widths clearly exceed the linear relation between equivalent widths and reddening of Equation 6.2, which they derived from the sample of Megier et al. (2005). However, the connection between Ca II and dust content in the ISM is highly non-trivial, as extensively discussed in Section 5.5. It strongly depends on many proportionality factors that can all vary drastically between different interstellar environments in the Milky Way. Therefore, there is no consistent way yet to model it from a theoretical point of view. Some effort is made to model the dependency empirically, similarly to what Li et al. (2025) does. But these relations are usually derived from close-by stars. For higher distances of a few kpc, especially towards higher Galactic latitudes, as for the stars in the LAMOST sample, interstellar Ca II is even less closely linked to interstellar dust (Section 5.5). The left panel of Figure 6.5 impressively shows, that for the non-CSM candidates, which should follow the ISM relation, the mean values of not even half of the sample is located within the 3σ range. Hence, it does not appear reasonable to use Equation 6.2 as foundation for the identification of CSM candidates.

- **Reddening excess:** The third criterion by Li et al. (2025) is that the SED reddening exceeds the prediction from 3D dust maps, what they call reddening excess. They argue that the dust maps can not account for dense localized structures around the stars, which the SEDs on the other hand are able to capture. So if the predictions from dust maps were highly reliable, this criterion might actually be the most promising one. Therefore, it is crucial to remove observations with uncertain parallax distances, as the distances are central for the determination of the positions of the stars needed for the comparison with 3D dust maps. However, a further source of uncertainty is the finite spatial resolution of the maps. Exactly, like for dense localized structures around the stars, they might be also not capable of resolving such dense local structures in the interstellar medium, which are certainly present. This means the reddening predicted from dust maps should rather be considered as rough estimate than as precise measurement. For example, the spectrum and the SED of the CSM candidate with the highest reddening excess in the right panel of Figure 6.5 were visually inspected. From its position very close to the Galactic disk, where most of the dense interstellar structures are located, and strong Na I D lines in the spectrum, it is clear that the high SED reddening is most likely explained by strong interstellar extinction rather than circumstellar one. This is not predicted by the 3D dust map. Thus, even though the reddening excess seems to be a reasonable argument for CSM, it also suffers from several

sources of uncertainties.

- **Radial velocities:** Probably the most convincing argument for CSM gravitationally bound to the star system and co-moving with it is, if the radial velocity exactly matches with the stellar one. This is also noticed by Li et al. (2025), hence they investigate the distribution of the radial velocity difference between the Ca II K line and their candidate stars. It peaks close to zero, what they take as confirmation for the circumstellar origin of the spectral lines. However, there are a few crucial things to note:

  1. As already mentioned earlier, in low resolution spectra the ISM and potential CSM component can not be distinguished. They will in the vast majority of all cases appear as only one spectral line. Since in general an ISM component is certainly expected for such distant stars in the sample, it is just wrong to consider the determined radial velocities as that of CSM. Therefore, Li et al. (2025) simulates the effect of blended components, but under the assumptions of a dominant CSM and weak ISM component ($W_{\mathrm{CaII\,K,CSM}} > 0.2$ Å and $W_{\mathrm{CaII\,K,ISM}} \sim 0.1$ Å) as well as an ISM radial velocity of always zero. These assumptions are in disagreement with the results of Chapter 5, in particular Figure 5.4 and Figure 5.5, where the average ISM equivalent width is also about 0.2 Å and the ISM radial velocity can scatter up to several tens of $\mathrm{km\,s^{-1}}$. Hence, their conclusion that the blend does not matter seems not to be reasonable[8].
  2. If the ISM component can not be neglected, it is not surprising that the distribution of the radial velocity differences between Ca II K lines and the stars peaks at zero, since the distributions of both, the ISM radial velocity and the stellar radial velocity is distributed around zero. Then, it does not necessarily indicate CSM.
  3. Even if the Ca II K lines in the spectra would purely originate from CSM, the large uncertainties on the radial velocity of on average about $30\,\mathrm{km\,s^{-1}}$ does not really allow for the conclusion that the line and stellar radial velocities align. Within one or two standard deviations they will in most of the cases do.

  Overall, only high resolution spectroscopy can help to distinguish between the ISM and CSM component and to derive valuable information from the determined radial velocities.

As final conclusion it can be said that even for the reduced sample of 49 CSM candidates, the results based on LAMOST spectroscopy are rather uncertain. It is more likely that the absorption features as well as the reddening originates from the ISM instead of potential CSM. Nevertheless, it would be highly interesting to collect high resolution spectra for these stars, in order to see if distinct ISM and CSM components are visible. Analyzing those spectra would also yield much better constrained equivalent widths and radial velocities for the Ca II K lines, with which a possible circumstellar origin could be investigated. Unfortunately, there is no overlap of the CSM candidates with the sample of higher resolution spectra of this work, even for the larger original sample of 145 candidates from Li et al. (2025). Therefore, this is just left as an outlook.

6.3 Search for New CSM Candidates

The previous section demonstrated that LAMOST spectroscopy is not sufficient to search for circumstellar material. The spectral resolution is just too low, even at high SNR it is not possible to distinguish between the ISM and a potential CSM component. High resolution spectroscopy is needed and hence, this section summarizes the search for CSM among the higher resolution X-Shooter, SPY/UVES, and FEROS spectra from the sample of this work (see Section 2.2), which comprises in total 918 spectra for 170 unique stars.

Due to the proximity of the CSM to the hot O- and B-type host stars, sodium is most likely ionized to higher ionization stages and no Na I D lines of circumstellar origin are expected.

[8] Except if it is assumed that almost all hot subdwarf stars show a dominant CSM component. But since the detection of CSM has been elusive for a long time, this is not very likely.

Therefore, similar to Li et al. (2025) it is essentially focused on the Ca II lines, in particular the K line, since it is intrinsically much stronger than the H one. The most convincing argument for CSM then is, if the radial velocity of the Ca II K line closely aligns with the stellar one, while an interstellar or photospheric (i.e. from a potential cooler companion) origin can be excluded.

For the search only apparently single objects are considered, since those with F/G/K-type main sequence companions detected in the SED are not thought to be formed via common envelope evolution. However, those apparently single hot subdwarfs can be either truly of single nature or in close binary systems with faint white or M dwarf companions. In the latter case, the SEDs are not sensitive to the companion, because it is outshone by the hot subdwarf star. Then the binary nature can be confirmed spectroscopically only by radial velocity variability of the system due to its orbital motion. Close binary systems containing a hot subdwarf star are thought to be formed via CE evolution channels, where both components dit not yet merge, and CSM might persist after the CE ejection. On the other hand, single hot subdwarfs are thought to be formed via merger events, if the hot flasher scenario (see Figure 1.6) or other potential formation channels without binary interaction are neglected. Either at least one He-WD merges with another low-mass component, resulting for example into a He-sdO, or the two former components merge inside of a common envelope before the ejection, resulting into single sdBs (Politano et al., 2008). In both cases a certain circumstellar environment can be expected.

Serendipitously, for many objects more than one higher resolution spectrum is available. Thus, the search is restricted to stars with multiple observed epochs (at least three), such that from the stellar radial velocities the single or close binary nature of the systems can be identified. Another advantage of the high resolution sample is that at least for X-Shooter and FEROS most of the objects are close-by (see Figure C.4). This means, the ISM contribution in the spectra is on average much smaller than for more distant stars like in the LAMOST sample of Li et al. (2025), making the distinction between ISM and potential CSM easier. The SPY/UVES objects are mostly much more distant. This on average increases the ISM contribution and makes the data noisier. Hence, it is more difficult to find CSM candidates among the SPY/UVES stars. Even though X-Shooter spectra have only an intermediate spectral resolution of $R \sim 9,800$, they have in general very high SNR (Figure 2.1). This makes them still suited for the search of CSM candidates.

To actively search for CSM, the whole analysis pipeline (Section 4.1) has been repeated for the higher resolution sample with setting the optional `csm` argument for the ISM line analysis script (see Subsection 3.3.4). Then the radial velocity of the calcium and sodium absorption lines is fixed to an interval of $\pm 5\,\mathrm{km\,s^{-1}}$ around the stellar radial velocity. If such lines are found exactly at the radial velocity of the observed system, it might indicate CSM gravitationally bound to the host system and co-moving with it.

6.3.1 Single CSM Candidates

The search for single CSM candidates formed via merger events has been limited to objects with non-variable stellar radial velocities of $|v_{\mathrm{rad}}| > 30\,\mathrm{km\,s^{-1}}$ and present Ca II lines in their spectra. The radial velocity restriction is useful to reduce the probability of confusing the potential CSM component with an ISM one, where the radial velocities are usually much closer to zero (see Figure C.2 and Figure 5.5). Overall, three CSM candidates have been identified and their arguments for the presence of CSM together with their other properties are summarized in the following individually. The uncertainties on all following parameters are only the pure statistical ones and hence, often unrealistically low. Still it was abstained from adding generic systematic uncertainties yet. Furthermore, the stated values are averaged ones over all available observations.

HD 205805

The first candidate with the most extreme radial velocity of $v_{\mathrm{rad}} = -69.31 \pm 0.07\,\mathrm{km\,s^{-1}}$ is HD 205805 (or Gaia DR3 6563704506492363392) and has five available FEROS spectra. The deviation between the measured stellar radial velocities is less than $1\,\mathrm{km\,s^{-1}}$ among the spectra, which strongly suggests a single nature. The SED (Figure 6.6) is well covered by photometric data

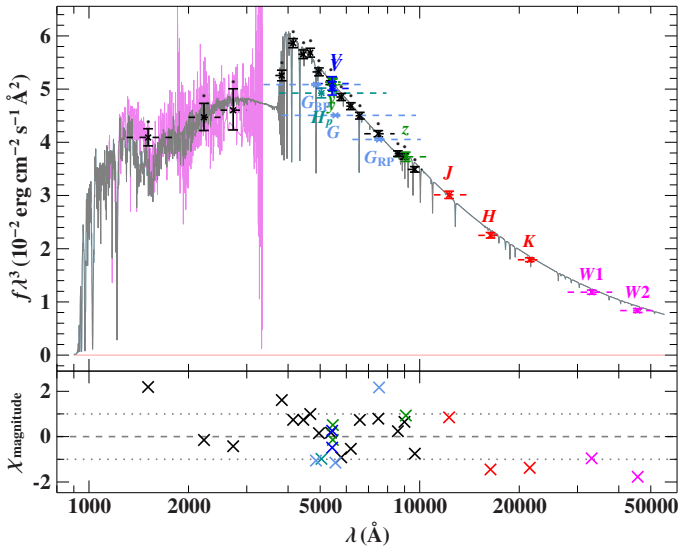


Figure 6.6: Observed SED for HD 205805. It is well covered by photometric data and does not show any peculiarities, especially no indication for a cool companion. For an explanation of the plot see Section 3.4. The averaged reddening from all five observations is $E(44-55) = 0.009 \pm 0.003$ mag.

and also does not show any peculiarities indicating the presence of a companion. The derived reddening of $E(44-55) = 0.009 \pm 0.003$ mag is quite low and an order of magnitude lower than the predictions from the 2D line of sight reddening maps[9] of Schlegel et al. (1998) and Schlafly & Finkbeiner (2011). Despite the low reddening a weak but clearly visible Ca II K line is observed in all spectra (Figure 6.7), at a radial velocity of $v_{\mathrm{rad,CaII\,K}} = -69.18 \pm 0.12\,\mathrm{km\,s^{-1}}$, which exactly matches with that of HD 205805 . It has a rather low equivalent width of $W_{\mathrm{CaII\,K}} = 0.065$ Å, but with an uncertainty of only $\Delta W_{\mathrm{CaII\,K}} = 0.003$ Å it is very well constrained. In Figure 6.8 the Voigt profile fit to the Ca II lines is shown, which gives a also well constrained Ca II column density of $\log N_{\mathrm{CaII}} = 11.93 \pm 0.04$[10]. Furthermore, it can be seen that the Ca II H line is present as well, but weaker than expected from the strength of the K line, as indicated by the combined fit (Subsection 3.3.4), which does not match both lines simultaneously. Na I D lines are not present at all, which further supports the assumption of CSM, since sodium is most likely ionized in the vicinity of hot stars. Figure 6.9 shows the blue part of HD 205805's spectrum, which also contains the Ca II H&K lines. Apparently, it shows a lot of metal lines besides the usual hydrogen and helium lines. The effective temperature of the star is $T_{\mathrm{eff}} = 24,590 \pm 60$ K and the surface gravity is $\log g = 5.054 \pm 0.009$. Moreover, with a distance of only $d = 243 \pm 3$ pc HD 205805 is very close to us and not a lot of intervening interstellar material is expected. This explains why no ISM component, in particular no sodium lines and also a low reddening, is observed and represents another argument for the circumstellar origin of the calcium lines. More information and parameters of the star, like the stellar parameters, are provided in Table E.1 and one whole spectrum is shown in Figure E.1.

EC 03591-3232

For the second CSM candidate, EC 03591-3232 (or CD-32 1567, Gaia DR3 4883474679951350656), even more spectra are available: Seven medium resolution X-Shooter spectra and eight high resolution FEROS spectra. It shows an stellar radial velocity of $v_{\mathrm{rad}} = 52.0 \pm 1.1\,\mathrm{km\,s^{-1}}$. Among the X-Shooter spectra the radial velocities scatter only with $\pm 1\,\mathrm{km\,s^{-1}}$, while among the FEROS spectra the scattering is about $\pm 3\,\mathrm{km\,s^{-1}}$. Still this scattering is far too low to claim the star as radial velocity variable. However, there is a small discrepancy between the X-Shooter and FEROS measurements. For X-Shooter the mean radial velocity is about $48\,\mathrm{km\,s^{-1}}$, while for FEROS it is about $55\,\mathrm{km\,s^{-1}}$. This is either a (wavelength) calibration issue or, since both sets

[9] Line of sight reddening maps give a rough estimate of the maximum reddening that is expected for a certain direction. These two maps are incorporated in the SED fitting program and the predicted values are automatically determined. A 3D dust map like that of Green et al. (2015) is unfortunately not yet implemented.

[10] The systematic uncertainty of 0.5 due to the fixing of the Doppler parameter (see Section 4.3) is not yet applied. It is also not done in the further course of this section.

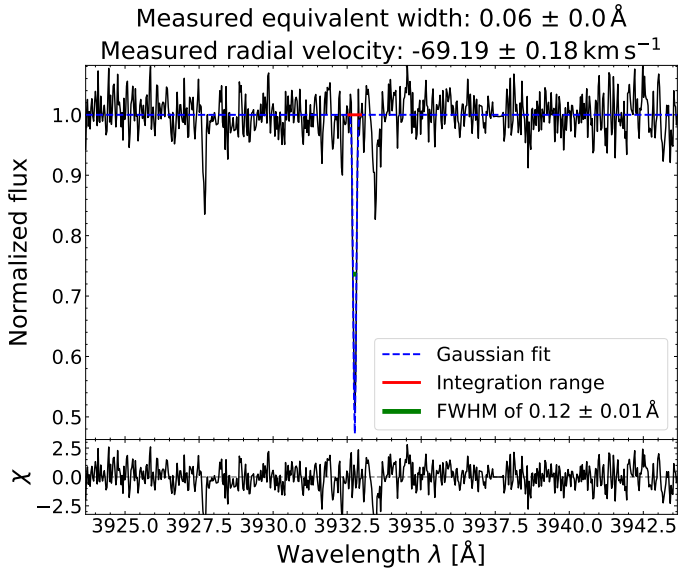


Figure 6.7: Example Gaussian fit to the Ca II K line present in one of the spectra of HD 205805. The averaged radial velocity of $v_{\mathrm{rad,CaII\,K}} = -69.18 \pm 0.12\,\mathrm{km\,s^{-1}}$ from the five spectra perfectly matches with that of the star ($v_{\mathrm{rad}} = -69.31 \pm 0.07\,\mathrm{km\,s^{-1}}$). The integration over the 5σ range gives an averaged equivalent width of $W_{\mathrm{CaII\,K}} = 0.065 \pm 0.003$ Å.

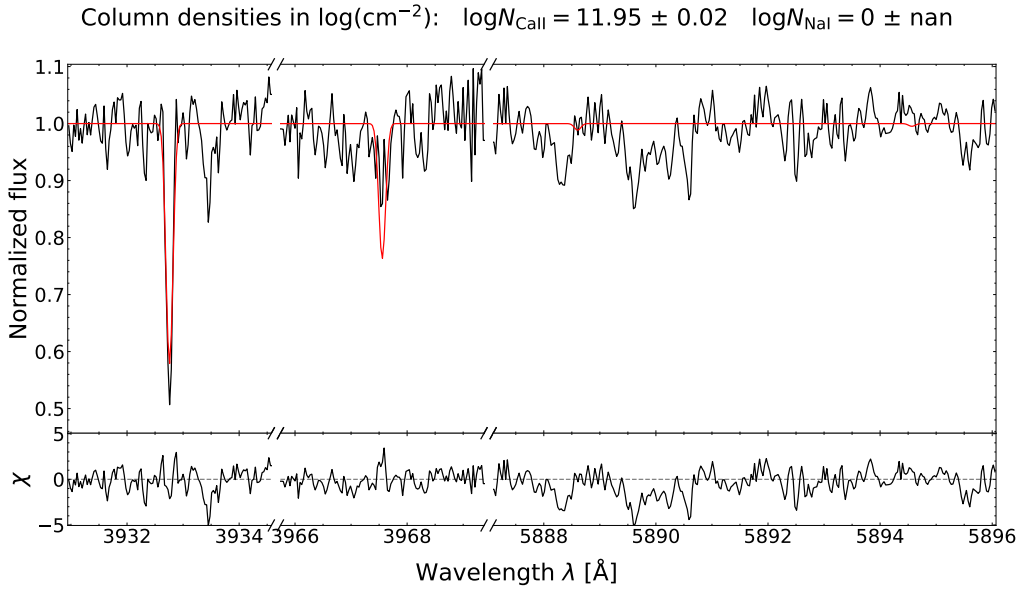


Figure 6.8: Example Voigt profile fit to the Ca II H&K lines present in one of the spectra of HD 205805, resulting into an averaged ionized calcium column density of $\log N_{\mathrm{CaII}} = 11.93 \pm 0.04$. The strengths of both lines seem to not perfectly match, as indicated by the simultaneous fit of both lines. Na I D1&D2 lines are not present and hence, not fitted.

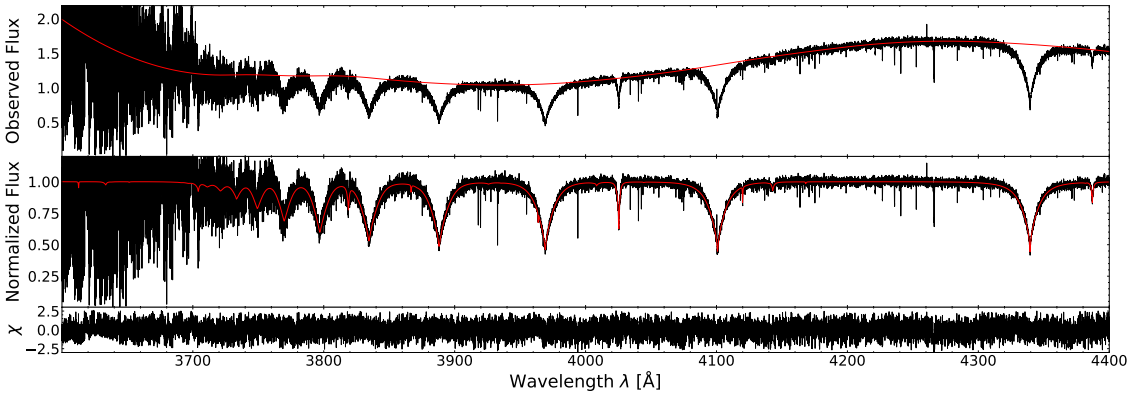


Figure 6.9: Blue part of one of the recorded FEROS spectra of HD 205805, which contains the Ca II H&K lines, together with the spectral fit (Subsection 3.3.3). At shorter wavelengths it is heavily dominated by noise, as all FEROS spectra are. Apart from the usual and properly fitted hydrogen and helium lines, it shows a lot of metal lines that are not included in the synthetic spectra (see Section 3.1). Effective temperature and surface gravity are $T_{\rm eff} = 24,590 \pm 60$ K and $\log g = 5.054 \pm 0.009$, respectively.

were recorded two years apart from each other, the star is part of a wide binary with a faint and outshone companion, e.g. a white dwarf. Then the radial velocity would change significantly only in the course of months or years and not within one or a few observation nights. This would mean the whole system was formerly a triple system, one single star plus the binary that merged during common envelope evolution and formed the hot subdwarf star. Apart from the UV, the SED (Figure 6.10) is well covered by photometric data and also does not show signs for a cooler companion. The derived reddening is $E(44-55) = 0.023 \pm 0.003$ mag, almost five times higher than the predicted line of sight reddenings from Schlegel et al. (1998) and Schlafly & Finkbeiner (2011) of about 0.005 mag. This reddening excess might come from dense localized structures in the ISM that are not resolved by the 2D dust maps, or indeed from dust in circumstellar material. The spectra support the latter scenario: As discussed in Section 5.5, the content of interstellar dust and neutral sodium is much more closely linked than that of dust and ionized calcium. But despite the non-negligible reddening from the SED, no Na I D lines are found at all in the spectrum (Figure 6.12). This speaks for ionized sodium as expected for the CSM in the vicinity of a hot star. On the other hand, the Ca II K line (Figure 6.11) as well as the H line (Figure 6.12) are clearly visible. The radial velocity of the K line of $v_{\rm rad,CaII\,K} = 50.5 \pm 0.9\,{\rm km\,s^{-1}}$ perfectly matches with the star and its equivalent width is $W_{\rm CaII\,K} = 0.080 \pm 0.005$ Å. The latter one might be slightly inflated, because in the X-Shooter spectra it is partly integrated over the feature in the left wing of the line (see left panel of Figure 6.11). The average calcium column density from the Voigt profile fits (Figure 6.12) is $\log N_{\rm CaII} = 11.67 \pm 0.03$. Again the strengths of the H&K lines do not match with each other very well. Figure 6.13 shows the blue part of one of the FEROS spectra for EC 03591-3232, where the calcium lines are located, together with the spectral fit including hydrogen and helium lines. Apart from those, a lot of metal lines are visible, just like for HD 205805. The effective temperature of the star is $T_{\rm eff} = 28,850 \pm 60$ K and the surface gravity is $\log g = 5.786 \pm 0.010$. With a distance of $d = 184 \pm 2$ pc it is even closer than HD 205805, another argument for the circumstellar origin of the Ca II lines and the reddening excess. More information and parameters of the star, like the stellar parameters, are provided in Table E.2 and one whole spectrum is shown in Figure E.2.

UVO 0825+15

The third CSM candidate is UVO 0825+15 (or GALEX J0828+1452, Gaia DR3 651745279826458112). It has an radial velocity of $v_{\rm rad} = 31.23 \pm 0.20\,{\rm km\,s^{-1}}$, derived from its seven X-Shooter and six FEROS spectra. The scattering of all measured radial velocities is below $\pm 1\,{\rm km\,s^{-1}}$, even across

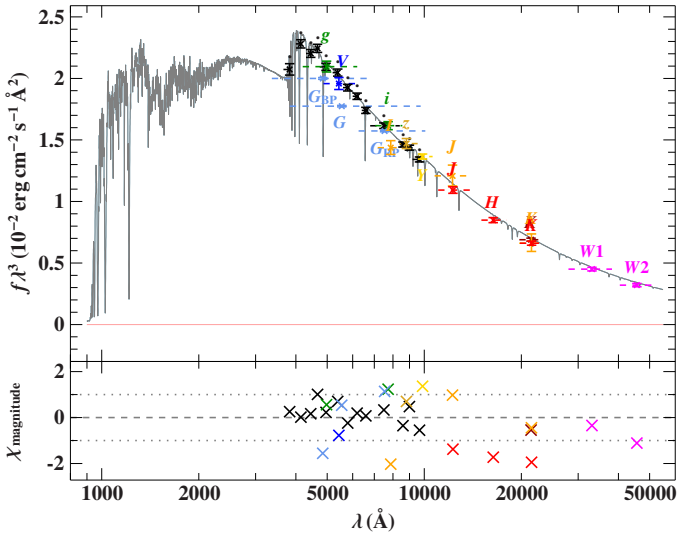


Figure 6.10: Observed SED for EC 03591-3232. Apart from the UV, it is well covered by photometric data and does not show any peculiarities, especially no indication for a cool companion. The averaged reddening from all observations is $E(44-55) = 0.023 \pm 0.003$ mag, much higher than predicted by the 2D line of sight dust maps of Schlegel et al. (1998) and Schlafly & Finkbeiner (2011).

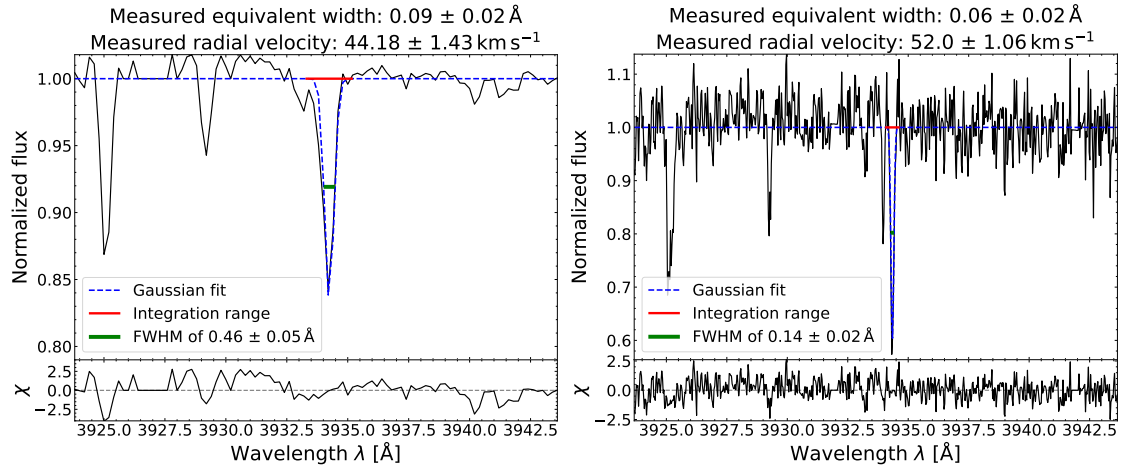


Figure 6.11: Example Gaussian fits to the Ca II K line present in the spectra of EC 03591-3232. The averaged radial velocity of $v_{\mathrm{rad,CaII\,K}} = 55.5 \pm 0.9$ km s^{-1} from all spectra perfectly matches with that of the star ($v_{\mathrm{rad}} = 52.0 \pm 1.1$ km s^{-1}). The integration over the 5σ range gives an averaged equivalent width of $W_{\mathrm{CaII\,K}} = 0.080 \pm 0.005$ Å. *Left:* Example line observed by X-Shooter. The equivalent widths might be slightly inflated, since it is partly integrated over the spectral feature in the left wing of the line. *Right:* Example line observed by FEROS.

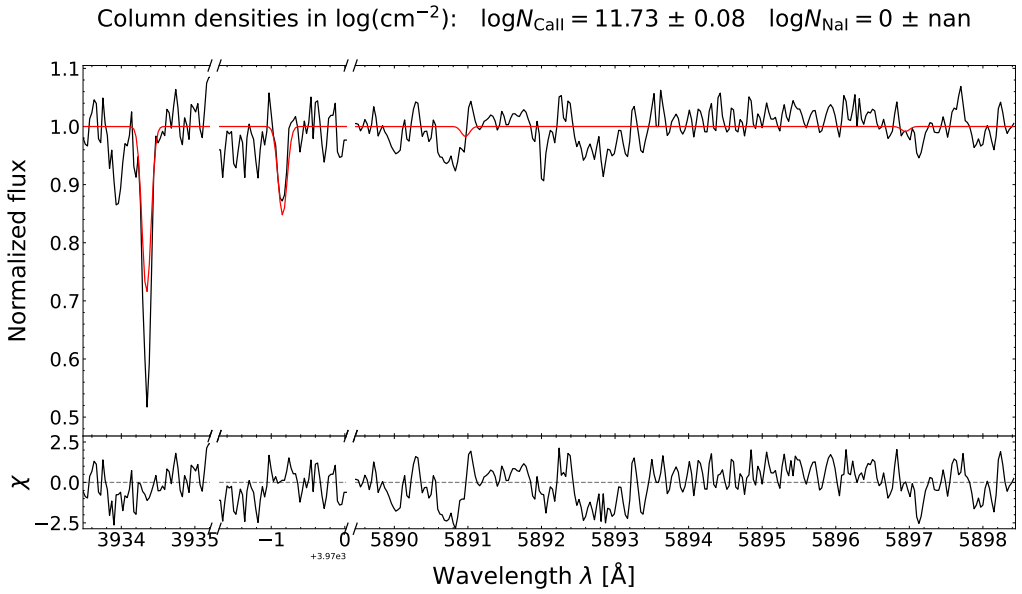


Figure 6.12: Example Voigt profile fit to the Ca II H&K lines present in one of the FEROS spectra of EC 03591-3232, resulting into an averaged ionized calcium column density of $\log N_{\rm CaII} = 11.67 \pm 0.03$. The strengths of both lines seem to not perfectly match, as indicated by the simultaneous fit of both lines. Na I D1&D2 lines are not present and hence, not fitted.

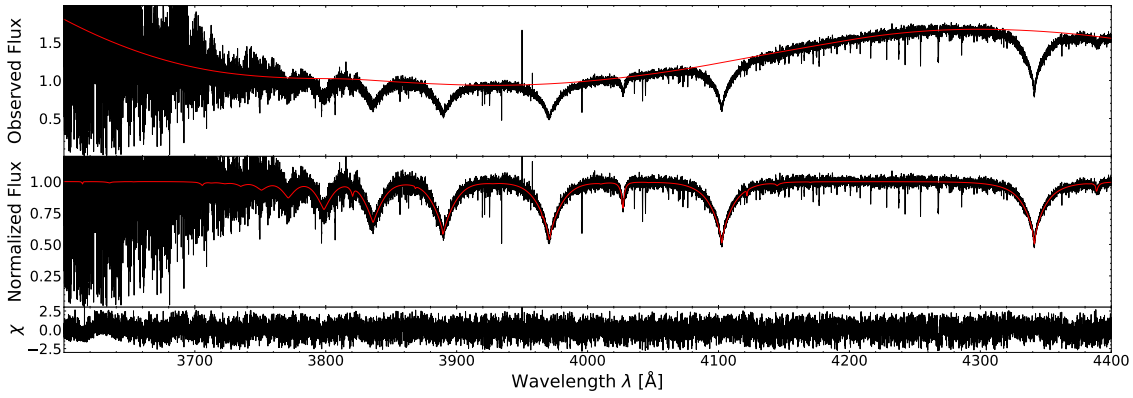


Figure 6.13: Blue part of one of the recorded FEROS spectra of EC 03591-3232, which contains the Ca II H&K lines, together with the spectral fit. Apart from the hydrogen and helium lines included in the models, it shows a lot of metal lines similar to HD 205805. Effective temperature and surface gravity are $T_{\rm eff} = 28,850 \pm 60$ K and $\log g = 5.786 \pm 0.010$, respectively.

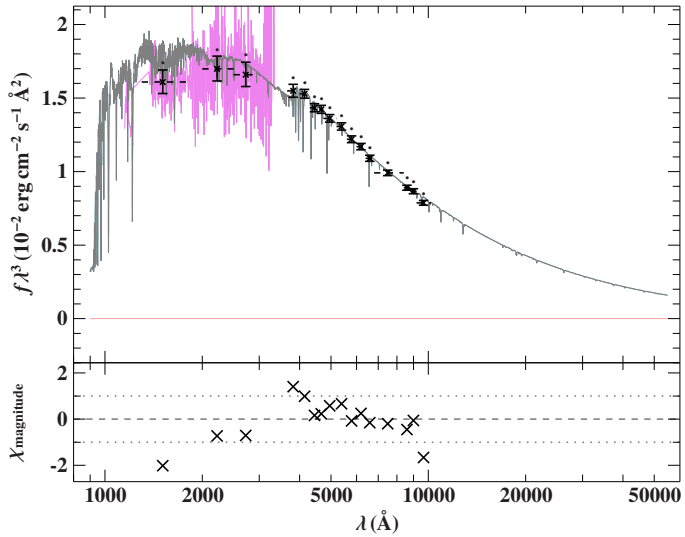


Figure 6.14: Observed SED for UVO 0825+15. It is far less well covered than the other two candidates, especially in the infrared, but the amount of data is still sufficient to constrain the derived parameters. The averaged reddening from all observations is $E(44-55) = 0.029 \pm 0.005$ mag.

the two spectrographs, clearly confirming that the star is a single object. This also suggests that the reason for the deviation for EC 03591-3232 is indeed of physical rather than technical nature. Unfortunately, for UVO 0825+15 only photometric data reconstructed from IUE spectra in the UV and *Gaia* XP spectra in the optical (see Section 3.4) is available. Therefore, the SED in Figure 6.14 is not as well covered as those for the other two candidates, in particular in the infrared. Nevertheless, the amount of data is still sufficient to constrain the derived parameters, like an reddening of $E(44-55) = 0.029 \pm 0.005$ mag. This time the SED reddening is slightly lower than the predictions by Schlegel et al. (1998) and Schlafly & Finkbeiner (2011) of about 0.035 mag. The findings concerning the calcium and sodium lines in principle exactly reproduce those of the other two candidates (Figure 6.15 and Figure 6.16): The Ca II K radial velocity of 30.0 ± 0.6 km s^{-1} agrees well with the stellar one, the equivalent width is $W_{\mathrm{CaII\,K}} = 0.061 \pm 0.007$ Å, and the column density is $\log N_{\mathrm{CaII}} = 11.66 \pm 0.04$, but the line strengths of the Ca II H&K lines do not match well[11]. Na I D lines are again not present. In Figure 6.17 the blue part of one of the FEROS spectra is shown, where again metal lines are visible. UVO 0825+15 has a rather high effective temperature of $39,200 \pm 100$ K and a surface gravity of $\log g = 5.947 \pm 0.004$. With a distance of 270 ± 4 pc it is still very close-by, but slightly further away than the other two candidates. More information and parameters of the star, like the stellar parameters, are provided in Table E.3 and one whole spectrum is shown in Figure E.3.

6.3.2 Close Binary CSM Candidates

The search for CSM candidates in close binary systems is also reasonable, since according to our current understanding the formation of such short period systems certainly includes common envelope evolution. There is right now no other mechanism known that lets the orbits of binaries shrink that drastically.

Fortunately, the search for CSM around close binaries is also possible with the same methods as before, because the cool companions are not expected to contribute with photospheric Ca II H&K lines, unlike warmer F/G/K-type companions in wide binaries. The faint components in close binary systems containing hot subdwarf stars are mostly white dwarfs or M dwarfs. White dwarfs usually only show broad Balmer or helium lines in their optical spectra due to gravitational settling of heavier elements (Gray & Corbally, 2009). On the other hand, M-type main sequence stars are too cool to show strong Ca II lines (also Gray & Corbally, 2009). The lines are usually very weak and additionally, the flux in the blue regime is much too small for M dwarfs compared to hot subdwarfs to contribute significantly in the overall spectrum of the system.

However, if the potential CSM persisting after CE ejection is thought to remain stationary and gravitationally bound to the system, its radial velocity is not expected to align with the stellar radial

[11] Actually, the Ca II H line is almost not visible due to the noise in the spectrum.

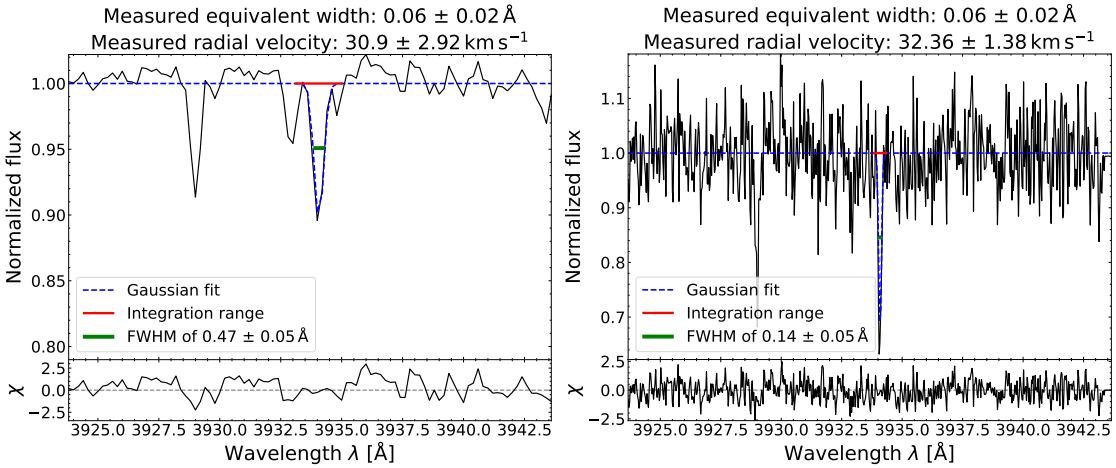


Figure 6.15: Example Gaussian fits to the Ca II K line present in the spectra of UVO 0825+15. The averaged radial velocity of $v_{\mathrm{rad,CaII\,K}} = 30.0 \pm 0.6\,\mathrm{km\,s^{-1}}$ from all spectra perfectly matches with that of the star ($v_{\mathrm{rad}} = 31.23 \pm 0.20\,\mathrm{km\,s^{-1}}$). The integration over the 5σ range gives an averaged equivalent width of $W_{\mathrm{CaII\,K}} = 0.061 \pm 0.007$ Å. *Left:* Observation by X-Shooter. *Right:* Observation by FEROS.

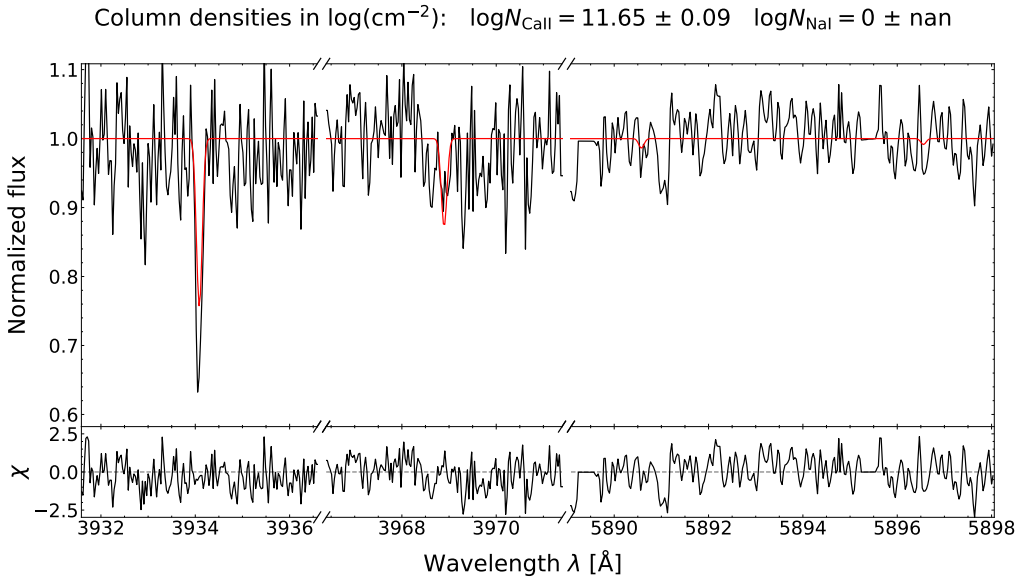


Figure 6.16: Example Voigt profile fit to the Ca II H&K lines present in one of the FEROS spectra of UVO 0825+15, resulting into an averaged ionized calcium column density of $\log N_{\mathrm{CaII}} = 11.66 \pm 0.04$. The strengths of both lines again seem to not perfectly match, as indicated by the simultaneous fit of both lines. Na I D1&D2 lines are not present and hence, not fitted.

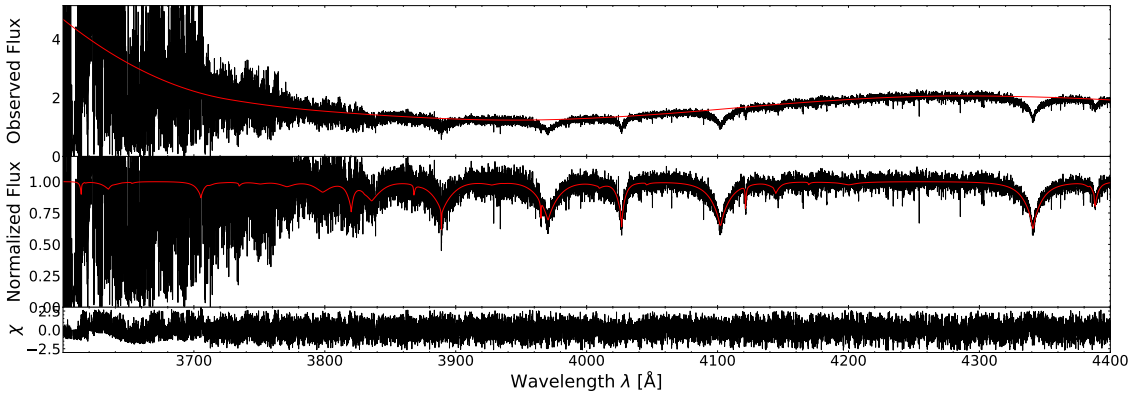


Figure 6.17: Blue part of one of the recorded FEROS spectra of UVO 0825+15, which contains the Ca II H&K lines, together with the spectral fit. Apart from the hydrogen and helium lines included in the models, it shows a lot of metal lines similar to the other two candidates. Effective temperature and surface gravity are $T_{\rm eff} = 39,200 \pm 100$ K and $\log g = 5.947 \pm 0.004$, respectively.

velocity of the hot subdwarf, but with the systemic velocity of the binary system. This systemic velocity can only be determined by modeling the radial velocity curve. This means taking all radial velocity measurements, finding the period of the system, phase folding the radial velocities over time, fitting a sine function (in the ideal case of a circular orbit), and obtaining the zero point of the sine curve, which gives the systemic velocity.

Unfortunately, for many objects in the higher resolution sample the number of observed epochs is not sufficient to precisely model the radial velocity curve. Furthermore, for the remaining objects this approach requires a lot of effort and is unfortunately out of the scope of this work. Therefore, the attempt is left as an outlook for the near future.

6.3.3 Discussion

With HD 205805, EC 03591-3232, and UVO 0825+15 three promising single candidates for CSM have been identified. With distances of less than 300 pc they are all close-by and hence, well known and well studied hot subdwarf stars. For example, all of them are included in the 500 pc volume complete sample of Dawson et al. (2024). All the findings, observed parameters, and connections between the three candidates are summarized and discussed in the following. Summary tables for the measured parameters can be found in Table 6.2 and Table 6.1.

- **Radial velocities:** A central criterion for single CSM candidates was that the radial velocity of Ca II lines present in their spectra aligns with that of the star. To avoid confusion with a possible ISM component, which typically shows radial velocities close to zero, the search was restricted to higher values of $|v_{\rm rad}| > 30\,{\rm km\,s^{-1}}$. In all the three cases both radial velocities are in astonishing good agreement. However, this criterion assumes stationary CSM gravitationally bound to the objects. It does not account for material that is still expanding or falling back on the star, which would show a different radial velocity. To search for moving CSM is even more difficult, because no constraints on the position of the Ca II lines can be made a priori.

- **Reddening:** Li et al. (2025) argued that a SED reddening exceeding the predictions by dust maps indicates CSM. However, as discussed in Subsection 6.2.4, this criterion is quite uncertain due to the finite spatial resolution of dust maps. Still, it appears worthwhile to have a look at the measured reddenings: For EC 03591-3232 and UVO 0825+15 the SED reddening is similar (Table 6.1). While EC 03591-3232 is strongly exceeding the prediction by the dust maps of Schlegel et al. (1998) and Schlafly & Finkbeiner (2011), UVO 0825+15 is slightly below the prediction. However, these maps give 2D line of sight reddenings, which can be considered as maximally expected values. Since the stars are very close, only little

intervening interstellar material is expected. Hence, it is surprising that the SED reddening is so close to or even exceeding the predictions. This could indicate dust in a circumstellar environment, but represents a much weaker argument than the radial velocity one. On the other hand, HD 205805 shows only little reddening, which is almost a factor of three lower than the prediction. The small value could possibly originate from circumstellar dust, but from little amounts of interstellar one as well. Thus, it does not allow for a conclusion.

- **Ca II H&K lines:** The spectra of all three candidates show weak, but clearly visible Ca II K lines as well as even weaker Ca II H lines close to the detection limit. The equivalent widths of the K line are all about $W_{\mathrm{CaII\,K}} \sim 0.06$ Å (Table 6.1), only that of EC 03591-3232 is slightly larger, but most likely inflated by a small spectral feature in the wing of the line (Figure 6.11). The overall ionized calcium column densities $\log N_{\mathrm{CaII}}$ from combined Voigt profile fits to the two lines are about 11.9 for HD 205805 and 11.7 for EC 03591-3232 as well as UVO 0825+15. These correspond to a number of Ca II ions in the line of sight to the stars between about 4.5×10^{11} cm^{-2} and 8.5×10^{11} cm^{-2}. Compared to average ISM column densities (Figure 5.2), the values are quite low, indicating rather thin and diffuse material. However, this depends on the initial chemical composition of the progenitor system. If the evolved star, from which the material came, was metal poor, the CSM could also be dense, just with little calcium content. To determine the hydrogen column densities, which would be a more proper proxy for the overall density of the material, one would need to know detailed chemical abundances to scale the calcium column density with them. Unfortunately, hot subdwarfs show metallicities that can vary drastically (Heber, 2016) and hence, it is not reasonable to assume a certain generic chemical composition. Moreover, one little discrepancy concerning the Ca II lines is that the line strengths of the H and K lines in all three cases seem to not fully match as seen from the combined Voigt profile fits. Intrinsically, the lines originate from the same particles and therefore, it is not possible that environmental conditions influence the shape of the lines independently from each other. This issue remains unresolved, but due to the weakness of the Ca II H lines the reason could be just the noise in the spectra.

- **Na I D1&D2 lines:** Since interstellar neutral sodium and interstellar dust are both predominantly found in cold neutral clouds, their content in the ISM is correlated (see Section 5.2). But despite the intermediate reddening values compared to the ISM measurements (Figure 5.6), none of the three candidates showed Na I D1&D2 lines in their spectra. This speaks once more for CSM rather than ISM, because in the vicinity of hot stars sodium is expected to become entirely ionized, while some newly formed dust grains from the CE ejection could possibly still survive and cause the reddening.

- **Metal lines:** Interestingly, all three candidates clearly exhibit a lot of metal lines in their spectra. While most of the hot subdwarf stars show mostly hydrogen and helium lines, some also show peculiar chemical compositions with considerable scatter from star to star (Geier, 2013; Heber, 2016). Such ”forests“ of metal lines have already been observed before, but it is intriguing that all three CSM candidates have this pattern. An explanation could be that these stars are indeed young objects, recently formed from CE evolution. Usually, due to their compactness hot subdwarfs are also affected by gravitational settling, similar to white dwarfs, but not as strongly. The gravitational pull lets heavier elements sink down in the atmosphere, while lighter ones rise. Now, if such an object was recently formed, it might not have had enough time to let the heavier elements sink deep enough, such that still a lot of metal lines arise from them in the spectra. Unfortunately, it was not in the scope of this work to investigate the forests of metal lines and it is just left as outlook.

- **Origin of the Ca II lines:** The most central argument for CSM is the presence of Ca II lines observed exactly at the radial velocity of the star. Therefore, it is important to rule out all other possible origins for these lines like an interstellar or photospheric origin or the confusion with another line (e.g., metal line) exactly at the same position. The photospheric

one can be excluded, because the three stars themselves are just too hot to show photospheric Ca II lines (Gray & Corbally, 2009) and there are no signs at all for a binary companion that could contribute with them. The confusion with a metal line seems to be possible, since all spectra show plenty of them, but at the moment there is no other line exactly at the position of the Ca II K line known. Lastly, the stars are pretty close to us and not a lot of intervening interstellar gas is expected. This is also confirmed by the absence of Na I D lines. Of course, it is imaginable that there is only one warm partially ionized cloud in the line of sight to each star, where Ca II is predominantly found, but Na I is already ionized. However, it is highly unlikely that this cloud furthermore moves by chance exactly with the same velocity as each star, especially at such extreme radial velocities unusual for interstellar clouds. In conclusion, CSM is the most probable origin for these lines.

- **Atmospheric and stellar parameters:** Finally, it appears worthwhile to have a closer look on the objects hosting the potential CSM. Their atmospheric and stellar parameters are listed in Table 6.2. The masses are close to the canonical mass for sdB stars of $0.47\,M_{\odot}$ (Heber, 2016) and also their radii and luminosities are well within the usual range for hot subdwarf stars (Figure C.5). Their atmospheric parameters as well as their memberships to different hot subdwarf populations can be evaluated best using the same diagnostic diagrams like in Section 4.2. In Figure 6.18 the helium abundance is plotted over the temperature. Shown is the whole sample of this work and the CSM candidates are marked with big asterisks. Obviously, the three stars are located far from each other, but still close to the main clusters. According to Figure 4.4, HD 205805 and EC 03591-3232 seem to belong to the sdB population, but are slightly more helium enhanced than average sdBs. An even higher helium abundance together with a much higher temperature shows UVO 0825+15, which places it close to the iHe-sdOBs. The Kiel diagram (Figure 6.19) shows a similar picture. While UVO 0825+15 is located near a main cluster on the helium main sequence (He-MS), HD 205805 and EC 03591-3232 are located close to the sdB cluster near the extreme horizontal branch (EHB). Apparently, HD 205805 has a $\log g$ that is slightly lower than for average sdB stars, while that of EC 03591-3232 is slightly higher. Overall, the positions of the three candidates close to, but not exactly on the main clusters, could indicate that they are currently evolving towards them, as expected for young objects that were formed recently through common envelope evolution. However, there is no clear evidence for this assumption. It is also worth to note, that the derived population memberships of the three candidates match with that of Dawson et al. (2024). Moreover, they are all well known objects that were already studied before and their measured parameters can be compared to the literature. For example, the atmospheric parameters of HD 205805 are in great agreement with Przybilla et al. (2006) and that of EC 03591-3232 and UVO 0825+15 are in good agreement with Geier et al. (2013) and Németh et al. (2012), respectively.

- **Rotational velocity:** Due to angular momentum conservation merger products should show strong rotation. The orbital angular momentum of the progenitor system is converted into rotation of the remnant. This is sometimes observed for eHe-sdOs, which form from the merging of two He-WDs (e.g., Schwab, 2018). Therefore, it appears worthwhile to have a look on the projected rotational velocities of the three CSM candidates, which might have formed through a CE merger event. The measured values are as follows: $v_{\rm rot}\sin i = 10.2 \pm 0.3\,{\rm km\,s^{-1}}$ for HD 205805, $v_{\rm rot}\sin i = 19.4 \pm 1.6\,{\rm km\,s^{-1}}$ for EC 03591-3232, and $v_{\rm rot}\sin i = 0 \pm 0\,{\rm km\,s^{-1}}$ for UVO 0825+15. While HD 205805 and EC 03591-3232 show little signs for rotation, UVO 0825+15 does not rotate at all. This slow rotation of sd(O)B stars is in consistency with for example Geier & Heber (2012). An explanation could be that, if the stars merged inside of a CE before its ejection, the viscosity of the surrounding material and the corresponding friction could have reduced the angular momentum of the object leading to larger rotation periods. However, this conclusion is purely speculative and the small rotation rates of apparently single sdB stars remain a puzzle for the current understanding of hot subdwarf formation (Geier & Heber, 2012).

Table 6.1: SED reddening, dust map reddening prediction (mean value from both maps of Schlegel et al. (1998) and Schlafly & Finkbeiner (2011)), Ca II K line equivalent width, and Ca II column density for the three CSM candidates. The extremely small statistical uncertainties on the line parameters are not provided, but can be reviewed in Appendix E.

	$E(44-55)$ [mag]	$E(B-V)_{\mathrm{Map}}$ [mag]	$W_{\mathrm{CaII\,K}}$ [Å]	$\log N_{\mathrm{CaII}}$ [cgs]
HD 205805	0.009 ± 0.003	0.0225 ± 0.0007	0.065	11.93
EC 03591-3232	0.023 ± 0.003	0.0049 ± 0.0007	0.080	11.67
UVO 0825+15	0.029 ± 0.005	0.036 ± 0.002	0.061	11.66

Table 6.2: Atmospheric and stellar parameters of the three CSM candidates. The extremely small statistical uncertainties are not provided, but can be reviewed in Appendix E.

	T_{eff} [K]	$\log g$ [cgs]	$\log y$	M $[M_\odot]$	R $[R_\odot]$	L $[L_\odot]$
HD 205805	24,590	5.054	-1.994	0.49	0.35	40
EC 03591-3232	28,850	5.786	-1.881	0.52	0.15	13
UVO 0825+15	39,200	5.947	-0.694	0.54	0.13	36

- **Further peculiarities:** It is not clear, whether there is a direct connection with common envelope evolution/ejection and circumstellar material, but two of the three stars are known for intriguing peculiarities: On the one hand, EC 03591-3232 shows a ^{3}He anomaly and on the other hand, UVO 0825+15 is well known for its heavy metal enrichment and pulsations. This should be kept in mind, when the stars are investigated further, for example by follow-up observations. Perhaps a link to CSM that is not apparent yet can be found.

In conclusion it can be said that with HD 205805, EC 03591-3232, and UVO 0825+15 three promising candidates for circumstellar material persistent after common envelope ejection have been identified. However, the most convincing argument is the presence of Ca II H&K lines in their spectra, for which at the moment just no other explanation is apparent than CSM. Therefore, they remain only candidates and it can not be concluded with a clear detection of CSM. Follow-up observations are required, for example with even higher resolution spectrographs or in other wavelength bands. For instance, radio observations utilizing the 21 cm line of neutral hydrogen could try to constrain the hydrogen column densities of material around the stars. With brightnesses of about tenth magnitude, eleventh magnitude, and twelfths magnitude, respectively, they are fairly bright objects (in the optical) and offer best opportunities for various kinds of follow-up.

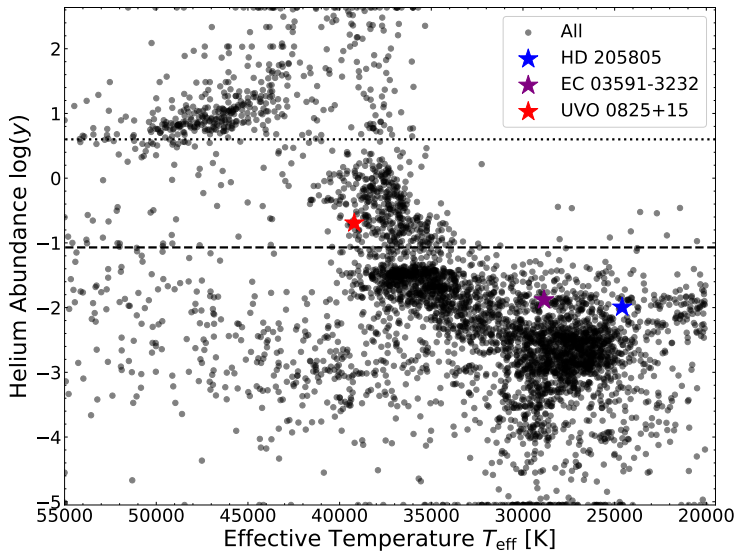


Figure 6.18: $\log y$ over T_{eff} diagram similar to Figure 4.4. Shown is the whole sample of this work in black together with the three CSM candidates depicted as big asterisks. HD 205805 and EC 03591-3232 seem to belong to the sdB population and UVO 0825+15 to the iHe-sdOB one.

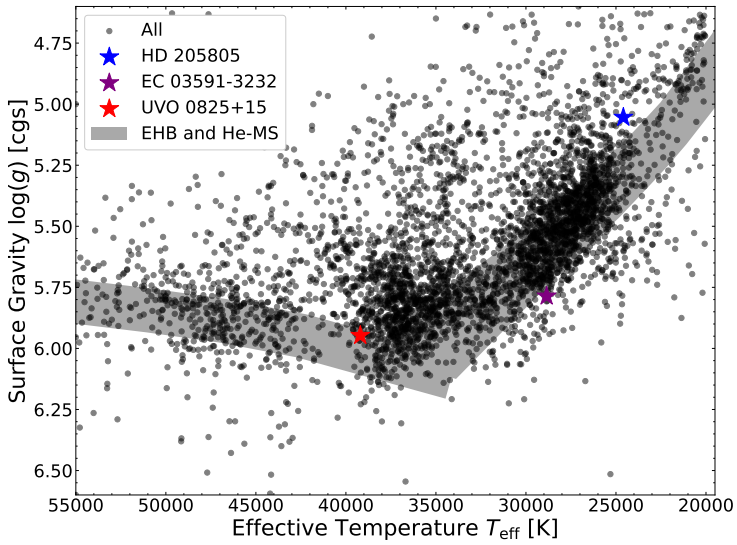


Figure 6.19: Kiel diagram similar to Figure 4.3. Shown is the whole sample of this work in black together with the three CSM candidates depicted as big asterisks. UVO 0825+15 is located on the helium main sequence (He-MS) of Paczyński (1971), while HD 205805 and EC 03591-3232 are located close to the extreme horizontal branch (EHB) of Dorman et al. (1993).

Chapter 7

Summary and Outlook

The interstellar medium (ISM) comprises many different distinct phases and environments. In particular, the ionization states of elements vary depending on these phases as well as the ratio of interstellar gas to dust. While the conditions in the Galactic disk, especially in the solar neighborhood, are well studied and mapped, the ISM structure of the Galactic halo is poorly understood, mainly due to a lack of available background sources observable out to the halo. This thesis presents hot subdwarf stars as possible light bulbs to trace the ISM at high Galactic latitudes. This population of stars represents late stages of stellar evolution that formed most likely via binary interaction including common envelope evolution. Hot subdwarfs are hot and bright enough to be observed out to the Galactic halo.

Characteristic interstellar absorption lines in optical spectra are the Ca II H&K as well as Na I D1&D2 lines and their properties can be used as proxy for interstellar gas. Another advantage of hot subdwarf stars is that they do not intrinsically show these lines in their spectra. However, the ISM lines are superimposed on the photospheric spectrum. In order to measure their shapes, it is necessary to model the photospheric spectrum as well. To this end, large grids of synthetic spectra are used. The derived stellar model flux is taken as continuum for the ISM lines to resolve the blends and to isolate the interstellar contribution. As side product, this yields the atmospheric parameters of the stars themselves. Furthermore, the reddening of starlight observed in stellar spectral energy distributions (SEDs) represents a proxy for the amount of interstellar dust along the line of sight. Once more, the results from the spectral analysis of the stellar spectrum are required to model these SEDs. As side product, stellar parameters are obtained. Combining both analyses, the spectral analyses as well as the SED modeling, allows to comprehensively trace and investigate the ISM. To this end, in total 6,400 stellar spectra for about 3,500 unique stars were analyzed in the course of this work. The spectra were queried either from large spectroscopic databases (SDSS, LAMOST, and DESI; all low resolution) or from the ESO archive (FEROS, SPY/UVES, and X-Shooter; higher resolution spectra). Subsequently, the SEDs of the corresponding objects were modeled in order to derive information about interstellar dust.

Besides the investigation of the interstellar medium, the second major aim of this study was to search for circumstellar material (CSM) around hot subdwarf stars that possibly remained after common envelope ejection. It is expected to cause similar absorption lines than the ISM and therefore, the same methods as before could be used for this search. The analysis techniques and results for both, the investigation of the ISM as well as the search for CSM, are summarized in the following.

7.1 A New Set of Python Scripts for Spectral Analyses

Some methods and tools, like synthetic model spectra for hot subdwarf stars or a SED fitting routine, were already available ready to use and just needed to be applied. However, the spectral analyses of this work required specific needs and hence, a new set of stellar Python scripts was developed in order to carry out the analyses. The central components are:

- `read_observation.py`: Reads and processes recorded spectra. The file formats of many

common surveys and databases are implemented.

- `interpolate_in_grid.py`: Auxiliary script to linearly interpolate within a given grid of model spectra, for instance necessary for spectral fits. Can also be used independently to extract synthetic spectra for a given combination of atmospheric parameters from a grid of model spectra.

- `spectral_fit.py`: Executes the fitting of synthetic spectra to observed ones resulting into the determination of atmospheric parameters.

- `ism_analysis.py`: First extracts and processes the regions around the interstellar Ca II and Na I lines. The synthetic photospheric stellar spectrum is used as continuum for the ISM. This, for example, resolves the blend of the weak interstellar Ca II H line with the in general strong photospheric Hε Balmer line. Subsequently, comprehensively analyzes the interstellar absorption lines, including the treatment of low spectral resolution and noise. Resulting parameters are for example equivalent widths, column densities, and radial velocities.

Since the scripts are still not purely restricted to the aims of this work, they could be of general usage for spectral analyses. Section 3.3 can be considered as kind of manual for these programs.

Simulated synthetic absorption lines for various spectral resolutions and signal to noise ratios (SNRs) were generated, in order to quantify these effects on the measured parameters. Both, the resolution and the SNR, determine the amount of spectral information contained in a recorded spectrum. Because interstellar absorption lines are often weak spectral features, low values of both were a major issue for the analysis of this work. The simulations demonstrated that the lower the spectral resolution, the higher the SNR must be to obtain reliable results. This result lead to strict SNR cuts applied to the observed data. Furthermore, detailed interstellar line profile fits intrinsically suffer from low spectral resolutions. Therefore, it was agreed on a systematic uncertainty of 0.5 on the derived column densities $\log N$. Also interstellar line radial velocities are poorly constrained in the low resolution case.

7.2 New Atmospheric and Stellar Parameters for Hot Subdwarf Stars

As valuable side-effect of fitting the spectra and modeling the SEDs of hot subdwarf stars, new atmospheric and stellar parameters are determined for all stars of the sample. Since it is not the purpose of this work to study the background sources themselves, the results were only briefly checked for convenience. First, diagnostic diagrams showed that there are no systematic deviations between the results from spectra of different origin and all parameters lie in the expected range for hot subdwarf stars. Many findings found in the literature were reproduced. For example, the stars are exactly located on their associated evolutionary branches, like the extreme horizontal branch (EHB) or the helium main sequence (He-MS), in the Kiel or Hertzsprung-Russell diagram. Furthermore, the statistical evaluation of the mass distribution gave a mode of $0.43^{+0.15}_{-0.16} M_{\odot}$ and a median of $0.45^{+0.20}_{-0.13} M_{\odot}$, both consistent with the canonical mass for sdB stars of about $0.47 M_{\odot}$. The distribution of radial velocities on the sky showed predominantly negative values on the Western Galactic side and predominantly positive values on the Eastern side, consistent with differential Galactic rotation.

Only since the advent of *Gaia* and its measured parallaxes, distances are available for a large number of hot subdwarf stars. Those distances are required for the determination of stellar parameters like mass, radius, and luminosity. While the atmospheric parameters of hot subdwarfs derived from spectroscopy have been studied for decades, stellar parameters for large samples of stars are only determinable since a few years. Therefore, the collected data set of this work readily allows for comprehensive follow-up on the properties of hot subdwarf stars.

7.3 Investigation of the Interstellar Medium

The core of this study is the investigation of the interstellar medium using hot subdwarf stars as background sources. This comprises not only to trace the distribution of the ISM in the Milky Way, but also to examine the abundances of neutral and ionized gas as well as that of interstellar dust. A further topic is the connection between interstellar gas and dust in the form of gas to dust ratios. The results are the following:

Interstellar parameter distributions. First, the raw resulting parameter distributions were considered and statistically evaluated. For example, the median of the Ca II column density is $\log \bar{N}_{\mathrm{Ca\,II}} = 12.7^{+0.6}_{-0.5}$, while that of Na I is $\log \bar{N}_{\mathrm{Na\,I}} = 12.4^{+0.9}_{-0.5}$. The confidence intervals rather represent the dispersion within the ISM than an uncertainty. The median value for the interstellar radial velocity is $\bar{v}_{\mathrm{rad,ISM}} = -8^{+29}_{-28}\,\mathrm{km\,s^{-1}}$, where the slight trend towards negative radial velocities can be explained with an observational bias due to the incomplete coverage of the sky with sight lines. The median of the observed reddening parameters is $\bar{E}(44-55) = 0.05^{+0.09}_{-0.04}$ mag.

Interstellar parameter correlations. The column density ratio of calcium to sodium gives conclusion about the respective phases of the ISM. On average the ratio was determined to $N_{\mathrm{Ca\,II}}/N_{\mathrm{Na\,I}} \approx 0.4$. This is a typical value for warm partially ionized clouds with temperatures between about 2,500-13,000 K, where predominantly ionized calcium is found rather than neutral sodium. With a covered volume of about 4 kpc around the Sun by hot subdwarf stars, also parts of the nearby Galactic halo seem to contain warm medium or at least more Ca II than Na I.

Furthermore, since column densities as well as equivalent widths have been measured, the curve of growth behavior of the spectral lines could be tested, which determines the growth of both parameters with respect to each other. The theoretical course of the curve of growth could be perfectly reproduced. It showed that the interstellar calcium and sodium densities are small to moderate, because they primarily lie in the saturated line core regime of the curve of growth.

Interstellar sky maps. Two dimensional interstellar sky maps were created for all the measured parameters. Unfortunately, the coverage of the sky was too poor to resolve a lot of structures. The only major structure that was clearly visible is the Taurus–Perseus–Auriga–California complex as part of the Gould belt. Still, some meaningful trends could be observed concerning the distribution of interstellar gas and dust on the sky. Apparently, the content of Na I and dust seems to be much more closely confined to the Galactic disk than that of Ca II, which is found more uniformly even at high Galactic latitudes.

Variations with distance, direction, and height from the Galactic plane. Owing to *Gaia* also the third dimension is accessible. Therefore, the findings from the sky maps were more thoroughly investigated. It was found that as expected, most of the interstellar gas and dust is concentrated in the Galactic disk. This pattern is even stronger for Na I and dust, where the contents for the Galactic halo fall well below the average values. For Ca II, however, they stay roughly at or are even still above average. This confirms that gaseous Ca II is more uniformly distributed towards the halo than gaseous Na I and interstellar dust.

Interstellar gas to dust ratios. Interstellar dust is expected to coexist with gas in interstellar clouds and hence, the amounts of gas and dust should be correlated. Poznanski et al. (2012) found an empirical relation between interstellar line strengths and reddening, which is constrained best for Na I equivalent widths. This relation has been tested using the parameters measured in this work. While the overall trend seems to match, a slight systematic offset was found. This is most likely explained by the different kinds of background sources. Poznanski et al. (2012) exclusively took extragalactic sources, quasars and galaxies, and therefore, their relation accounts for the overall ISM in the line of sight instead of just the local one, like in this work.

It is known that the average gas to dust ratio differs between the Galactic disk and the halo. However, in other studies the gas to dust ratio is given as hydrogen column density over reddening and the former one is not directly accessible in this work. It was tried to apply the relation of Murga et al. (2015) for a conversion of Ca II and Na I equivalent widths to hydrogen column densities, but

this procedure was highly uncertain. The resulting average gas to dust ratios are in the possible range for both, the disk and the halo, and hence, do not provide a conclusion.

The investigation of calcium and sodium gas to dust ratios in dependency on distance, direction, and height from the Galactic plane showed at least little trends. For example, that of Ca II clearly increases with distance, while that of Na I does not. Also a little dip in the calcium gas to dust ratio appears close to the disk, which is not apparent for sodium. This represents another indicator that neutral sodium and dust are much more closely linked in the interstellar medium than ionized calcium and dust. Na I coexists with dust in the cold neutral medium, while Ca II is predominantly found in warm partially ionized medium, where the abundances of Na I and dust are much lower. Moreover, Ca II extends more uniformly to much greater distances out to the halo, where only little dust is found.

7.4 Search for Circumstellar Material Around Hot Subdwarf Stars

Many formation channels for hot subdwarf stars include common envelope (CE) evolution. Eventually, the CE is ejected and some material is expected to remain bound to the host system. Direct detections of such circumstellar material (CSM) have been elusive yet and hence, as second aim of this work it was searched for such CSM using the same methods as before.

Recent detections of CSM. Out of a sample of 727 sdB stars, Li et al. (2025) identified 145 as CSM candidates based on the appearances of Ca II K lines and reddening. However, the analysis was based on low resolution LAMOST spectra, which are also part of this sample. The simulations of interstellar parameters have shown that low resolution comes with major issues and uncertainties on the derived parameters. Li et al. (2025) mostly neglects these uncertainties. Moreover, some of the used criteria for the identification of CSM candidates by Li et al. (2025) are intrinsically uncertain, which becomes apparent from the analysis of this work. By a comparison of the results from both analyses and a more careful treatment, the sample of CSM candidates was reduced to 49 objects. Still, the uncertainties on their parameters are just too high to claim them as detection. High resolution follow-up is required to confirm or falsify them.

Search for new candidates. The most convincing argument for CSM bound to the host system is, if the radial velocity of the Ca II lines exactly matches with that of the object, when at the same time an interstellar or photospheric origin can be excluded. Therefore, it was searched among the sample of higher resolution spectra (918 spectra for 170 unique stars) for objects fulfilling this criterion. The search was further restricted to single stars, since wide binary systems are not expected to be formed via common envelope channels and for close binaries the radial velocity of the lines would need to match with the systemic velocity instead of the apparent stellar velocity, which is in general unknown. Those single hot subdwarf stars are thought to be formed in merger events, like common envelope mergers or a merger event without CE evolution including at least one helium white dwarf component. Even in the latter case a certain circumstellar environment can be expected.

In total, three CSM candidates were found: The two sdB stars HD 205805 and EC 03591-3232 and the iHe-sdOB UVO 0825+15. All of them show Ca II lines exactly at the stellar radial velocity and at the same time exceeding values of $|v_{\mathrm{rad}}| = 30\,\mathrm{km\,s^{-1}}$ unusual for the interstellar medium. The resulting Ca II column densities are well below $\log N = 12$, indicating rather diffuse material. Na I lines are completely absent in the spectra of the three candidates, which is expected for material close to hot stars, where sodium is most likely ionized. Interestingly, all of the three candidates show many metal lines in their spectra. This pattern could be explained with the age of the stars. If they were recently formed by CE evolution and ejection, heavier elements perhaps had not enough time yet to sink down due to gravitational settling.

The atmospheric and stellar parameters of the candidates are consistent with the expected range for hot subdwarf stars and do not show peculiarities. However, it is known that EC 03591-3232 shows a ^{3}He isotope anomaly and UVO 0825+15 is enriched with heavy elements and pulsates. The helium-poor stars HD 205805 and EC 03591-3232 are located on the EHB, while the intermediate

helium-rich star UVO 0825+15 resides on the He-MS. A merger formation scenario seems possible for the three single stars, even though no significant rotation was observed for them. The slow rotation rates of hot subdwarf stars are still an unresolved puzzle in general (Geier & Heber, 2012), but could be maybe explained by friction inside of the common envelope after the merging of the two components. However, this conclusion is purely speculative.

7.5 Outlook

The methods and findings of this work laid a foundation for the investigation of the interstellar medium using hot subdwarf stars as background light bulbs as well as for the search for circumstellar material around these stars. They are an excellent starting point for extensions of these analyses and for possible follow-up. Some future prospects are the following:

Investigation of the ISM. Even though the investigation of the ISM yielded promising results, there are two major limitations: One the one hand, low resolution spectra introduce large statistical and systematic uncertainties on the measured parameters. On the other hand, the coverage of sight lines on the sky is poor and inhomogeneous. Interstellar structures can mostly not be resolved and a little observational bias is introduced. Both issues could be reduced by large spectroscopic data releases in the future. Such will be for example SDSS-V (Kollmeier et al., 2017) or 4MOST (de Jong et al., 2022), which will not only enhance the number of sight lines, especially in the Southern hemisphere, but also provide slightly higher resolution spectra. If the number of objects increases strong enough, it might become possible to compile comprehensive gas and dust maps as well as to refine the results of this work.

Search for CSM. The reduced sample of 49 CSM candidates from Li et al. (2025) can only be confirmed or falsified by high resolution follow-up. This might be possible by future public spectroscopic data releases from large surveys or by the granting of observation time for high resolution spectrographs.

With HD 205805, EC 03591-3232, and UVO 0825+15 three CSM candidates were identified already based on high resolution spectroscopy. It is likely that they are surrounded by circumstellar material, but still follow-up observations are required to corroborate this assumption. Those could be either the recording of even higher resolution spectra or observations in other wavelength regimes than the optical. For example, UV spectra would show many more interstellar resonance lines from metals helping to distinguish between interstellar and circumstellar components. Furthermore, radio observations could look for local enhancements of neutral hydrogen near the candidates, which would indicate CSM.

Moreover, the search for CSM candidates will be extended to binary systems. Already with the available data it is for many objects possible to model the radial velocity curves in order to determine the systemic velocity. Then it can be searched for Ca II lines in the spectra, where the radial velocities exactly match with that systemic velocity. If other explanations, like an interstellar or photospheric origin of the lines, can be excluded, this would clearly indicate circumstellar material and lead to new candidates.

Chapter 8

Bibliography

Abbott T. M. C., et al., 2021, ApJS, 255, 20
Alam S., et al., 2015, ApJS, 219, 12
Almeida A., et al., 2023, ApJS, 267, 44
Althaus L. G., Miller Bertolami M. M., Córsico A. H., 2013, A&A, 557, A19
Alves J., et al., 2020, Nature, 578, 237
Anders F., et al., 2022, A&A, 658, A91
Arenou F., et al., 2018, A&A, 616, A17
Asplund M., Amarsi A. M., Grevesse N., 2021, A&A, 653, A141
Babusiaux C., et al., 2023, A&A, 674, A32
Barlow B. N., Wade R. A., Liss S. E., Østensen R. H., Van Winckel H., 2012, ApJ, 758, 58
Baron D., Poznanski D., Watson D., Yao Y., Prochaska J. X., 2015, MNRAS, 447, 545
Bartelmann M., 2021, Theoretical Astrophysics: An Introduction. Lecture Notes, Heidelberg University Publishing, doi:10.17885/heiup.822, https://heiup.uni-heidelberg.de/catalog/book/822
Baschek B., Norris J., 1975, ApJ, 199, 694
Bianchi L., Shiao B., Thilker D., 2017, ApJS, 230, 24
Bohlin R. C., Savage B. D., Drake J. F., 1978, ApJ, 224, 132
Butler K., Giddings J., 1985, Newsletter on the analysis of astronomical spectra
Carnall A. C., 2017, arXiv e-prints, p. arXiv:1705.05165
Carroll B. W., Ostlie D. A., 2017, An introduction to modern astrophysics, Second Edition. Cambridge University Press
Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102
Clausen K. F., Perryman M. A. C., 1990, Acta Astronautica, 22, 229
Clayton D. D., 1983, Principles of stellar evolution and nucleosynthesis. University of Chicago press
Clayton M., Podsiadlowski P., Ivanova N., Justham S., 2017, MNRAS, 470, 1788
Cui X.-Q., et al., 2012, Research in Astronomy and Astrophysics, 12, 1197
Culpan R., Geier S., Reindl N., Pelisoli I., Gentile Fusillo N., Vorontseva A., 2022, A&A, 662, A40
Culpan R., Dorsch M., Geier S., Pelisoli I., Heber U., Kubátová B., Cabezas M., 2024, A&A, 685, A134
Cutri R. M., et al., 2003, VizieR Online Data Catalog: 2MASS All-Sky Catalog of Point Sources (Cutri+ 2003), VizieR On-line Data Catalog: II/246. Originally published in: University of Massachusetts and Infrared Processing and Analysis Center, (IPAC/California Institute of Technology) (2003)
D'Cruz N. L., Dorman B., Rood R. T., O'Connell R. W., 1996, ApJ, 466, 359
DESI Collaboration et al., 2016a, arXiv e-prints, p. arXiv:1611.00036
DESI Collaboration et al., 2016b, arXiv e-prints, p. arXiv:1611.00037
DESI Collaboration et al., 2024, AJ, 168, 58
DESI Collaboration et al., 2025, arXiv e-prints, p. arXiv:2503.14745
D'Odorico S., et al., 2006, in McLean I. S., Iye M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy. p. 626933, doi:10.1117/12.672969
Dame T. M., Hartmann D., Thaddeus P., 2001, ApJ, 547, 792
Dawson H., et al., 2024, A&A, 686, A25
Dekker H., D'Odorico S., Kaufer A., Delabre B., Kotzlowski H., 2000, in Iye M., Moorwood A. F., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4008, Optical and IR Telescope Instrumentation and Detectors. pp 534–545, doi:10.1117/12.395512
Demyk K., 2011, in European Physical Journal Web of Conferences. p. 03001, doi:10.1051/epjconf/20111803001
Dewi J. D. M., Tauris T. M., 2000, A&A, 360, 1043
Dorman B., Rood R. T., O'Connell R. W., 1993, ApJ, 419, 596
Dorsch M., 2023, PhD thesis, Friedrich Alexander University of Erlangen-Nuremberg, Germany
Draine B. T., 2011, Physics of the interstellar and intergalactic medium. Vol. 19, Princeton University Press
Driebe T., Schoenberner D., Bloeckera T., Herwig F., 1998, VizieR Online Data Catalog: PSR J1012+5307 evolutionary tracks (Driebe+ 1998), VizieR On-line Data Catalog: J/A+A/339/123. Originally published in: 1998A&A...339..123D, doi:10.26093/cds/vizier.33390123

Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, The Messenger, 154, 32
Fabricius C., et al., 2016, A&A, 595, A3
Fabricius C., et al., 2021, A&A, 649, A5
Fitzpatrick E. L., Massa D., Gordon K. D., Bohlin R., Clayton G. C., 2019, ApJ, 886, 108
Gaia Collaboration et al., 2016a, A&A, 595, A1
Gaia Collaboration et al., 2016b, A&A, 595, A2
Gaia Collaboration et al., 2018, A&A, 616, A1
Gaia Collaboration et al., 2021, A&A, 650, C3
Gaia Collaboration et al., 2023a, A&A, 674, A1
Gaia Collaboration et al., 2023b, A&A, 674, A33
Geier S., 2013, A&A, 549, A110
Geier S., Heber U., 2012, A&A, 543, A149
Geier S., Heber U., Edelmann H., Morales-Rueda L., Kilkenny D., O'Donoghue D., Marsh T. R., Copperwheat C., 2013, A&A, 557, A122
Geier S., Raddi R., Gentile Fusillo N. P., Marsh T. R., 2019, A&A, 621, A38
Geier S., Dorsch M., Pelisoli I., Reindl N., Heber U., Irrgang A., 2022, A&A, 661, A113
Glanz H., Perets H. B., 2021, MNRAS, 507, 2659
González-Bolívar M., De Marco O., Lau M. Y. M., Hirai R., Price D. J., 2022, MNRAS, 517, 3181
González-Riestra R., Cassatella A., Wamsteker W., 2001, A&A, 373, 730
Gray D. F., 2021, The observation and analysis of stellar photospheres. Cambridge university press
Gray R. O., Corbally J. C., 2009, Stellar Spectral Classification. Princeton university press
Green G. M., et al., 2015, ApJ, 810, 25
Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, ApJ, 887, 93
Green G. M., Zhang X., Zhang R., 2025, ApJ, 988, 5
Gudennavar S. B., Bubbly S. G., Preethi K., Murthy J., 2012, ApJS, 199, 8
Gupta A., Sahijpal S., 2020, MNRAS, 494, 4149
Han Z., Podsiadlowski P., Eggleton P. P., 1995, MNRAS, 272, 800
Han Z., Podsiadlowski P., Maxted P. F. L., Marsh T. R., 2003, MNRAS, 341, 669
Hastings W. K., 1970, Biometrika, 57, 97
He R., Meng X., Lei Z., Yan H., Lan S., 2025, A&A, 693, A121
Heber U., 2009, ARA&A, 47, 211
Heber U., 2016, PASP, 128, 082001
Heber U., Irrgang A., Schaffenroth J., 2018, Open Astronomy, 27, 35
Henden A. A., Levine S., Terrell D., Welch D. L., 2015, in American Astronomical Society Meeting Abstracts #225. p. 336.16
Houck J. C., Denicola L. A., 2000, in Manset N., Veillet C., Crabtree D., eds, Astronomical Society of the Pacific Conference Series Vol. 216, Astronomical Data Analysis Software and Systems IX. p. 591
Hubeny I., Mihalas D., 2014, Theory of Stellar Atmospheres. Princeton University Press
Humason M. L., Zwicky F., 1947, ApJ, 105, 85
Husser T. O., Wende-von Berg S., Dreizler S., Homeier D., Reiners A., Barman T., Hauschildt P. H., 2013, A&A, 553, A6
Iaconi R., De Marco O., Passy J.-C., Staff J., 2018, MNRAS, 477, 2349
Iben Jr. I., Livio M., 1993, PASP, 105, 1373
Irrgang A., 2014, PhD thesis, Friedrich Alexander University of Erlangen-Nuremberg, Germany
Irrgang A., Wilcox B., Tucker E., Schiefelbein L., 2013, A&A, 549, A137
Istrate A. G., Marchant P., Tauris T. M., Langer N., Stancliffe R. J., Grassitelli L., 2016, A&A, 595, A35
Ivanova N., 2011, in Schmidtobreick L., Schreiber M. R., Tappert C., eds, Astronomical Society of the Pacific Conference Series Vol. 447, Evolution of Compact Binaries. p. 91 (arXiv:1108.1226), doi:10.48550/arXiv.1108.1226
Ivanova N., Chaichenets S., 2011, ApJL, 731, L36
Ivanova N., Nandez J. L. A., 2016, MNRAS, 462, 362
Ivanova N., Podsiadlowski P., 2002, Ap&SS, 281, 191
Ivanova N., et al., 2013a, A&AR, 21, 59
Ivanova N., Justham S., Avendano Nandez J. L., Lombardi J. C., 2013b, Science, 339, 433
Ivanova N., Justham S., Ricker P., 2020, Common Envelope Evolution. IOP Publishing, doi:10.1088/2514-3433/abb6f0
Izzard R. G., Hall P. D., Tauris T. M., Tout C. A., 2012, in Planetary Nebulae: An Eye to the Future. pp 95–102, doi:10.1017/S1743921312010769
Kalberla P. M. W., Kerp J., 2009, ARA&A, 47, 27
Kaper L., et al., 2009, in Moorwood A., ed., Astrophysics and Space Science Proceedings Vol. 9, Science with the VLT in the ELT Era. p. 319 (arXiv:0803.0609), doi:10.1007/978-1-4020-9190-2_53
Kaufer A., Pasquini L., 1998, in D'Odorico S., ed., Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 3355, Optical Astronomical Instrumentation. pp 844–854, doi:10.1117/12.316798
Kawka A., Vennes S., O'Toole S., Németh P., Burton D., Kotze E., Buckley D. A. H., 2015, MNRAS, 450, 3514
Kollmeier J. A., et al., 2017, arXiv: Astrophysics of Galaxies
Kramer M., Schneider F. R. N., Ohlmann S. T., Geier S., Schaffenroth V., Pakmor R., Röpke F. K., 2020,

A&A, 642, A97
Kramida A., Yu. Ralchenko Reader J., and NIST ASD Team 2024, NIST Atomic Spectra Database (ver. 5.12), [Online]. Available: https://physics.nist.gov/asd [2025, September 22]. National Institute of Standards and Technology, Gaithersburg, MD.
Kreuzer S., 2021, PhD thesis, Friedrich Alexander University of Erlangen-Nuremberg, Germany
Kupfer T., et al., 2015, A&A, 576, A44
Kurucz R. L., 1996, in Adelman S. J., Kupka F., Weiss W. W., eds, Astronomical Society of the Pacific Conference Series Vol. 108, M.A.S.S., Model Atmospheres and Spectrum Synthesis. p. 160
Lallement R., Welsh B. Y., Vergely J. L., Crifo F., Sfeir D., 2003, A&A, 411, 447
Lan T.-W., Ménard B., Zhu G., 2014, ApJ, 795, 31
Lawrence A., et al., 2007, MNRAS, 379, 1599
Lenz D., Hensley B. S., Doré O., 2017, ApJ, 846, 38
Levenberg K., 1944, Quarterly of applied mathematics, 2, 164
Li J., et al., 2022, MNRAS, 515, 3370
Li J., et al., 2025, MNRAS, 537, 1950
Lindegren L., et al., 2016, A&A, 595, A4
Lindegren L., et al., 2021, A&A, 649, A2
Linsky J. L., Redfield S., Ryder D., Chasan-Taber A., 2022, AJ, 164, 106
Lisker T., Heber U., Napiwotzki R., Christlieb N., Han Z., Homeier D., Reimers D., 2005, A&A, 430, 223
Liszt H., 2014a, ApJ, 780, 10
Liszt H., 2014b, ApJ, 783, 17
Liszt H., Gerin M., 2023, ApJ, 943, 172
Loveday J., 2002, Contemporary Physics, 43, 437
Loveridge A. J., van der Sluys M. V., Kalogera V., 2011, ApJ, 743, 49
Luo Y., Németh P., Wang K., Wang X., Han Z., 2021, ApJS, 256, 28
Ma X.-Y., Zhu Y.-Y., Yan Q.-B., You J.-Y., Su G., 2020, MNRAS, 497, 2190
Magnier E. A., et al., 2020, ApJS, 251, 6
Marasco A., Fraternali F., Lehner N., Howk J. C., 2022, MNRAS, 515, 4176
Marquardt D. W., 1963, Journal of the society for Industrial and Applied Mathematics, 11, 431
Mascetti L., Delmotte N., Arnaboldi M., Forchi V., Micol A., Barbieri M., Romaniello M., Coccato L., 2024, in Benn C. R., Chrysostomou A., Storrie-Lombardi L. J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 13098, Observatory Operations: Strategies, Processes, and Systems X. p. 130980M, doi:10.1117/12.3017760
McMahon R. G., Banerji M., Gonzalez E., Koposov S. E., Bejar V. J., Lodieu N., Rebolo R., VHS Collaboration 2013, The Messenger, 154, 35
Megier A., Strobel A., Bondar A., Musaev F. A., Han I., KreŁowski J., Galazutdinov G. A., 2005, ApJ, 634, 451
Megier A., Strobel A., Galazutdinov G. A., Krełowski J., 2009, A&A, 507, 833
Mengel J. G., Norris J., Gross P. G., 1976, ApJ, 204, 488
Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, JCP, 21, 1087
Miller Bertolami M. M., 2016, A&A, 588, A25
Molina F., Vos J., Németh P., Østensen R., Vuković M., Tkachenko A., van Winckel H., 2022, A&A, 658, A122
Morton D. C., 2000, ApJS, 130, 403
Munari U., Zwitter T., 1997, A&A, 318, 269
Murga M., Zhu G., Ménard B., Lan T.-W., 2015, MNRAS, 452, 511
Napiwotzki R., et al., 2001, Astronomische Nachrichten, 322, 411
Napiwotzki R., et al., 2003, The Messenger, 112, 25
Napiwotzki R., et al., 2005, in Koester D., Moehler S., eds, Astronomical Society of the Pacific Conference Series Vol. 334, 14th European Workshop on White Dwarfs. p. 375
Napiwotzki R., et al., 2020, A&A, 638, A131
Nelder J. A., Mead R., 1965, The computer journal, 7, 308
Németh P., Kawka A., Vennes S., 2012, MNRAS, 427, 2180
Newville M., Stensitzki T., Allen D. B., Ingargiola A., 2014, LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python, doi:10.5281/zenodo.11813
Nieva M. F., Przybilla N., 2007, A&A, 467, 295
Onken C. A., et al., 2019, PASA, 36, e033
Paczyński B., 1971, Acta Astron., 21, 1
Paczynski B., 1976, in Eggleton P., Mitton S., Whelan J., eds, IAU Symposium Vol. 73, Structure and Evolution of Close Binary Systems. p. 75
Passy J.-C., et al., 2012, ApJ, 744, 52
Pastorello A., et al., 2019, A&A, 630, A75
Pelisoli I., Vos J., Geier S., Schaffenroth V., Baran A. S., 2020, A&A, 642, A180
Podsiadlowski P., 2001, in Podsiadlowski P., Rappaport S., King A. R., D'Antona F., Burderi L., eds, Astronomical Society of the Pacific Conference Series Vol. 229, Evolution of Binary and Multiple Star Systems. p. 239
Podsiadlowski P., Han Z., Lynas-Gray A. E., Brown D., 2008, in Heber U., Jeffery C. S., Napiwotzki R., eds, Astronomical Society of the Pacific Conference Series Vol. 392, Hot Subdwarf Stars and Related

Objects. p. 15 (arXiv:0808.0574), doi:10.48550/arXiv.0808.0574
Podsiadlowski P., Ivanova N., Justham S., Rappaport S., 2010, MNRAS, 406, 840
Politano M., Taam R. E., van der Sluys M., Willems B., 2008, ApJL, 687, L99
Powell M. J., 1964, The computer journal, 7, 155
Poznanski D., Prochaska J. X., Bloom J. S., 2012, MNRAS, 426, 1465
Przybilla N., Nieva M. F., Edelmann H., 2006, Baltic Astronomy, 15, 107
Przybilla N., Nieva M.-F., Butler K., 2011, in Journal of Physics Conference Series. p. 012015 (arXiv:1111.1445), doi:10.1088/1742-6596/328/1/012015
Rachford B. L., et al., 2009, ApJS, 180, 125
Redfield S., Linsky J. L., 2002, ApJS, 139, 439
Retzlaff J., Arnaboldi M., Delmotte N. A. R., Mascetti L., Micol A., 2016, in Peck A. B., Seaman R. L., Benn C. R., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9910, Observatory Operations: Strategies, Processes, and Systems VI. p. 991009, doi:10.1117/12.2231664
Richmond M. W., Treffers R. R., Filippenko A. V., Paik Y., Leibundgut B., Schulman E., Cox C. V., 1994, AJ, 107, 1022
Richter P., et al., 2017, A&A, 607, A48
Riello M., et al., 2021, A&A, 649, A3
Robertson J. G., 2017, PASA, 34, e035
Röpke F. K., De Marco O., 2023, Living Reviews in Computational Astrophysics, 9, 2
Sand C., Ohlmann S. T., Schneider F. R. N., Pakmor R., Röpke F. K., 2020, A&A, 644, A60
Savage B. D., Bohlin R. C., Drake J. F., Budich W., 1977, ApJ, 216, 291
Schaffenroth V., Pelisoli I., Barlow B. N., Geier S., Kupfer T., 2022, A&A, 666, A182
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlafly E. F., Meisner A. M., Green G. M., 2019, ApJS, 240, 30
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Schwab J., 2018, MNRAS, 476, 5303
Shull J. M., Panopoulou G. V., 2024, ApJ, 961, 204
Shull J. M., Danforth C. W., Anderson K. L., 2021, ApJ, 911, 55
Smee S. A., et al., 2013, The Astronomical Journal, 146, 32
Sodroski T. J., Odegard N., Arendt R. G., Dwek E., Weiland J. L., Hauser M. G., Kelsall T., 1997, ApJ, 480, 173
Speagle J. S., 2020, A Conceptual Introduction to Markov Chain Monte Carlo Methods (arXiv:1909.12313), https://arxiv.org/abs/1909.12313
Steinmetz T., Kamiński T., Schmidt M., Kiljan A., 2024, A&A, 682, A127
Stoehr F., et al., 2008, in Argyle R. W., Bunclark P. S., Lewis J. R., eds, Astronomical Society of the Pacific Conference Series Vol. 394, Astronomical Data Analysis Software and Systems XVII. p. 505
Stroeer A., Heber U., Lisker T., Napiwotzki R., Dreizler S., Christlieb N., Reimers D., 2007, A&A, 462, 269
Szalay A., 1999, Computing in Science & Engineering, 1, 54
Taam R. E., Ricker P. M., 2010, NewAR, 54, 65
Tauris T. M., Dewi J. D. M., 2001, A&A, 369, 170
Tout C. A., Regos E., 2003, in Turcotte S., Keller S. C., Cavallo R. M., eds, Astronomical Society of the Pacific Conference Series Vol. 293, 3D Stellar Evolution. p. 100
Ungerechts H., Thaddeus P., 1987, ApJS, 63, 645
Van Parys A., n.d., Measures of central tendency: Mean, median and mode, https://ledidi.com/academy/measures-of-central-tendency-mean-median-and-mode
Vernet J., et al., 2011, A&A, 536, A105
Virtanen P., et al., 2020, scipy/scipy: SciPy 1.6.0, doi:10.5281/zenodo.4406806
Vos J., Vučković M., Chen X., Han Z., Boudreaux T., Barlow B. N., Østensen R., Németh P., 2019, MNRAS, 482, 4592
Wakker B. P., van Woerden H., 1997, ARA&A, 35, 217
Wales D. J., Doye J. P. K., 1997, The Journal of Physical Chemistry A, 101, 5111–5116
Webbink R. F., 1984, ApJ, 277, 355
Weingartner J. C., Draine B. T., 2001, ApJ, 563, 842
Welsh B. Y., Vedder P. W., Vallerga J. V., Craig N., 1991, ApJ, 381, 462
Welsh B. Y., Lallement R., Vergely J. L., Raimond S., 2010, A&A, 510, A54
Welty D. E., Hobbs L. M., Kulkarni V. P., 1994, ApJ, 436, 152
Werk J. K., et al., 2019, ApJ, 887, 89
West J. L., Gaensler B. M., Miville-Deschênes M. A., Mahajan N., Dechant J., Boulanger F., Martin P. G., Zelko I. A., 2023, ApJ, 959, 115
Xu X.-J., Li X.-D., 2010, ApJ, 716, 114
York D. G., et al., 2000, The Astronomical Journal, 120, 1579
Yu J., Zhang X., Lü G., 2021, MNRAS, 504, 2670
Yuan H. B., et al., 2015, MNRAS, 448, 855
Zasowski G., et al., 2015, ApJ, 798, 35
Zhang X., Green G. M., 2025, Science, 387, 1209
Zhang R., Yuan H., Chen B., 2023, ApJS, 269, 6
Zhao G., Zhao Y.-H., Chu Y.-Q., Jing Y.-P., Deng L.-C., 2012, Research in Astronomy and Astrophysics, 12, 723

Zhukovska S., Henning T., Dobbs C., 2018, ApJ, 857, 94
de Jong R. S., et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V. p. 91470M, doi:10.1117/12.2055826
de Jong R. S., et al., 2022, in Evans C. J., Bryant J. J., Motohara K., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 12184, Ground-based and Airborne Instrumentation for Astronomy IX. p. 1218414, doi:10.1117/12.2628965
van Leeuwen F., 1997, SSR, 81, 201
van Leeuwen F., 2009, A&A, 500, 505

Appendix A

The Statistical Treatment of Parameter Distributions

In general, if a parameter is measured multiple times, it does not always result into the same measured value. From pure statistical uncertainties, the results will scatter in form of a Gaussian distribution around the true value. However, in reality often more factors than only statistical uncertainties play a role. The resulting distribution might be oddly shaped, skewed, or show more than one distinct peak. Then, there are in principle three different choices to quantify the average or most probable value as well as its uncertainty. Each measure captures different aspects of the distribution and remains informative under different assumptions about its shape. The three approaches are briefly discussed in the following and are illustrated for unimodal distributions in Figure A.1.

Mean and Standard Deviation

The arithmetic mean μ of the distribution for a measured quantity X with measurements x is defined as

$$\mu(X) = \int_{-\infty}^{\infty} x\, p(x)\, \mathrm{d}x\,, \tag{A.1}$$

where $p(x)$ is called the probability density function (PDF). It provides a natural measure of the central value if the distribution is symmetric and unimodal (one distinct peak). The spread around the mean is quantified by the standard deviation

$$\sigma = \sqrt{\left(\int_{-\infty}^{\infty} (x-\mu)^2\, p(x)\, \mathrm{d}x \right)}\,. \tag{A.2}$$

For finite samples, the mean and the standard deviation are given by the discrete expressions

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad \text{and} \qquad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}\,, \tag{A.3}$$

respectively, where N is the number of measurements. The mean and the standard deviation are widely used, but they are sensitive to outliers and might provide a misleading picture if the distribution is skewed or multimodal (more than one distinct peak).

Median and Quantiles

An alternative, more robust, characterization is given by the median m, which is defined as the value for which half of the distribution lies below and half above:

$$P(X \leq m) = \int_{-\infty}^{m} p(x)\, \mathrm{d}x = 0.5. \tag{A.4}$$

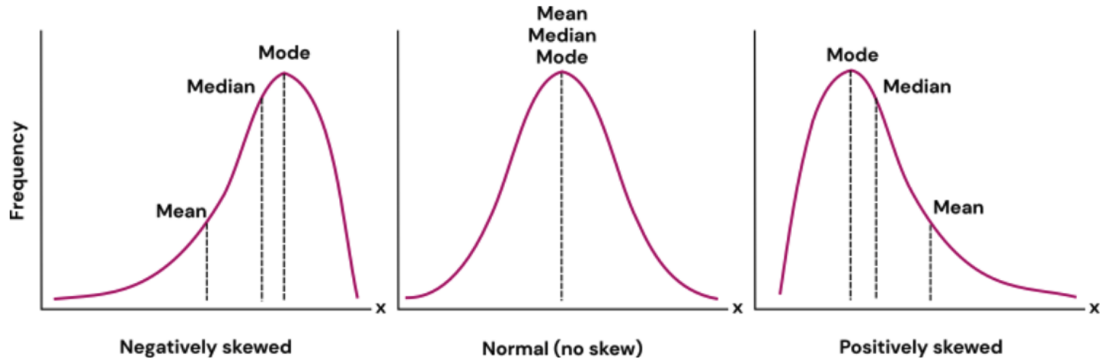


Figure A.1: Illustration of different choices for average values of unimodal distributions (Van Parys, nd). For a Gaussian (normal) distribution mean, median, and mode are identical. For skewed ones, the mode corresponds to the maximum and the medium to the point, where the number of datapoints towards both sides is equal. The mean is only informative for Gaussian-like shaped distributions.

In more common words, the median is exactly the value in the middle of a sorted dataset. As measure for its uncertainty can serve the so-called α-quantiles q_α. They are defined as

$$P(X \leq q_\alpha) = \int_{-\infty}^{q_\alpha} p(x)\,\mathrm{d}x = \alpha. \tag{A.5}$$

For example, the central $(1-\alpha)$-interval based on quantiles is given by

$$\left[q_{\frac{\alpha}{2}},\, q_{1-\frac{\alpha}{2}}\right], \tag{A.6}$$

which covers a fraction $(1-\alpha)$ of the parameter distribution. The commonly used 68% (~1σ) interval corresponds to the quantiles at 16% and 84%. Unlike the mean and standard deviation, quantile-based intervals remain more robust under skewness, multimodality, or heavy-tailed distributions. However, the upper and lower uncertainties (quantiles) can become asymmetric.

Mode and Highest Density Interval (HDI)

The mode x_{mode} is defined as the value of x at which the distribution has its maximum. While the mode captures the most probable parameter value, it is often complemented by an interval that reflects the uncertainty of the maximum. A commonly used confidence interval is the highest density interval (HDI). For a given probability p, e.g. 0.68 (~1σ) or 0.95 (~2σ), the HDI is the narrowest (and possibly asymmetric) interval $[a,b]$ such that

$$\int_a^b p(x)\,\mathrm{d}x = p\,, \tag{A.7}$$

and for all $x \in [a,b]$ and $y \notin [a,b]$, it holds that $p(x) \geq p(y)$. In other words, every point inside the HDI has a higher probability than any point outside. This makes the HDI the most credible region of the distribution. Unlike quantile-based intervals, which always exclude the same amount of probability from each side of the distribution, the HDI automatically adapts to the shape of the distribution and always contains its mode. However, if the distribution is either multimodal or does not show a distinct peak, i.e. it is almost flat, the mode and HDI is no longer reasonable

Appendix B

Supplementary Material for Chapter 3

B.1 Details About The Spectral Interpolation Routine

This Appendix provides a few more technical details about the interpolation routine of the `interpolate_in_grid.py` script introduced in Subsection 3.3.2: First, a information file for the selected grid is read and the parameters are separated into interpolatable and fixed values. All parameters that have at least two grid points can be interpolated, but those having only one are fixed to that value. If the input value for a fixed parameter differs from the grid point, the input value is reset to the fixed value and if the input value for an interpolatable parameter lies outside of the grid border, it is set to the nearest valid grid point with a small buffer. Then, for every interpolatable parameter the nearest grid points to the given input values are determined and the corresponding models are stored. The list of models is then reshaped as needed for the `RegularGridInterpolator`[1] from `scipy.interpolate`, which is the object from the scipy library that finally carries out the linear multidimensional interpolation. The outcome is an interpolated model spectrum (flux over wavelength) for exactly the given atmospheric parameters. This spectrum can then be plotted as well as saved (both, the plot and a table containing the spectrum) as requested by the user.

B.2 The `get_minimizer` Function of the Spectral Fitting Routine

This Appendix provides further information about the setting up of the `Minimizer` object by the `get_minimizer` function for the spectral fits (see Subsection 3.3.3): When `get_minimizer` is called, first an empty `Parameters` object is created. Then, the `GridInterpolator` object from `interpolate_in_grid.py` is initialized to check, which atmospheric parameters shall be varied in the respective model grid. Those are passed to the `Parameters` object with the grid borders as upper and lower fit limits. Parameters that can not be varied are added to it as well together with the non-grid parameters like radial velocity, resolution parameters, and projected rotational velocity. Of course, only those parameters specified as free parameters by the user in the `setup_fitting_parameters` function are used as free parameters during the fit and all others are fixed. If a parameter that can not be varied in the chosen model grid is given as free parameter, it is fixed nevertheless. The limits for the non-grid parameters are also set to reasonable physical values, e.g., to $\pm 1000\,\mathrm{km\,s^{-1}}$ for the radial velocity. Additionally, if the helium abundance given as $\log n(\mathrm{He})$ is a variable fit parameter, also the helium abundance given as $\log y$ is added as fixed parameter in dependency on $\log n(\mathrm{He})$ according to Equation 3.4. The last fit parameters needed are the anchor points for the normalization process, which works as follows: The aim is to find a continuum estimate for an observed spectrum using synthetic models. All fluxes are then divided by this continuum to end up with a spectrum normalized to unity. Technically, this is achieved by fitting a spline through a set of fixed x-anchor points (wavelengths) and variable y-anchor points (flux scaling factors), which are treated as free parameters during the fitting process. The spline

[1] https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.RegularGridInterpolator.html

then represents an estimate of the stellar continuum. Therefore, the x-values are given by the set of anchor points defined in the `setup_cspline_anchorpoints` function and added to the `Parameters` object as fixed parameters, while the y-values are added as free fit parameters. As initial values the median flux of the 10 Å neighborhood around each x-anchor point is taken. A bit of caution is needed for the choice of the x-anchor points in the `setup_cspline_anchorpoints` function, because the normalization process is very sensitive to it. They should always, if possible, be placed in featureless regions of a spectrum, i.e. not within spectral lines. Otherwise, it can lead to unnatural ”wiggles“ in the resulting continuum estimate. In some cases, it might be helpful to adjust the default anchor points, for example to reduce the spacing between them in order to account for irregularities in the spectrum. As last step, the created `Parameters` object and the fit function `fit_func` are used to create the `Minimizer` object needed for the minimization process.

Appendix C

Supplementary Material for Chapter 4

C.1 Atmospheric and Stellar Parameters

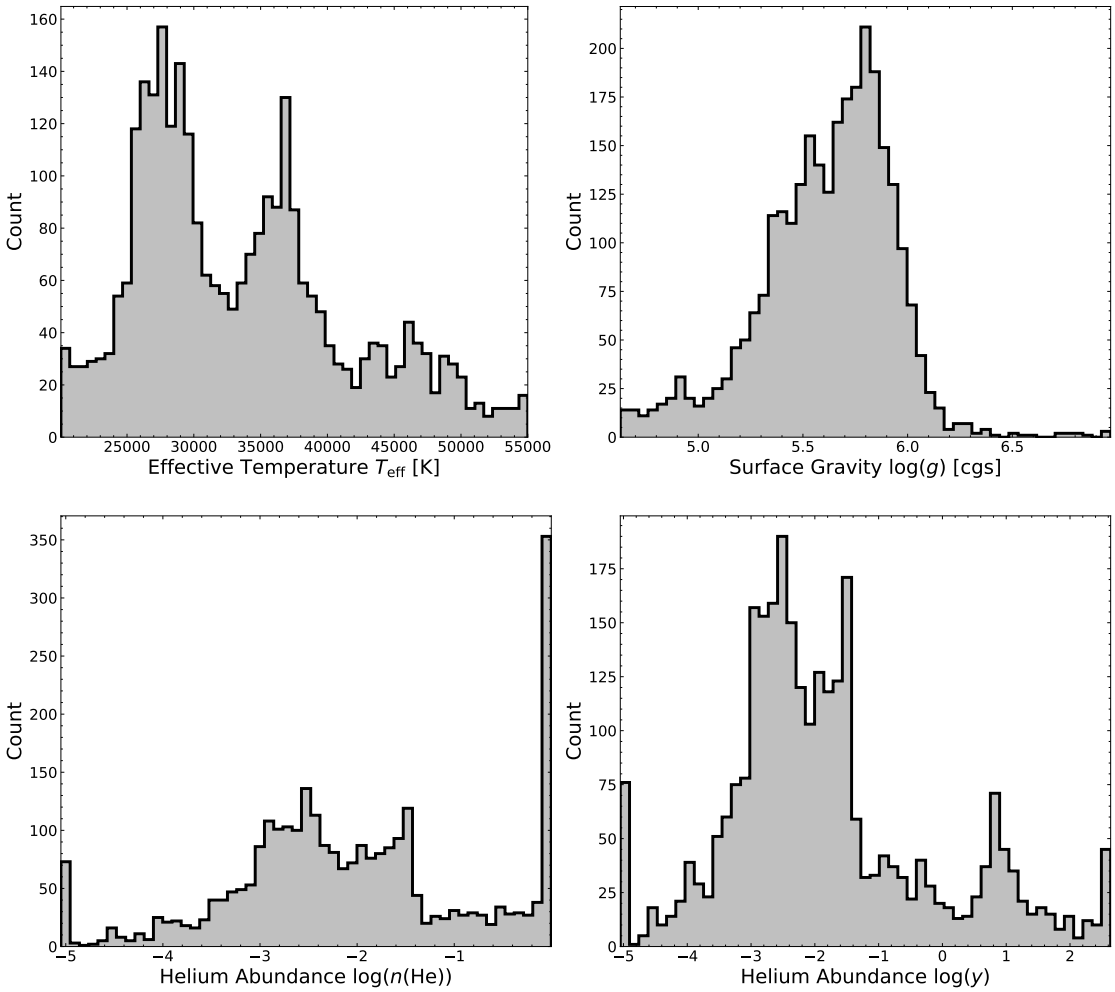


Figure C.1: Raw atmospheric parameter distribution for the sample of analyzed hot subdwarf stars in this work. The distinct peaks correspond to the different hot subdwarf populations as outlined in Section 4.2.

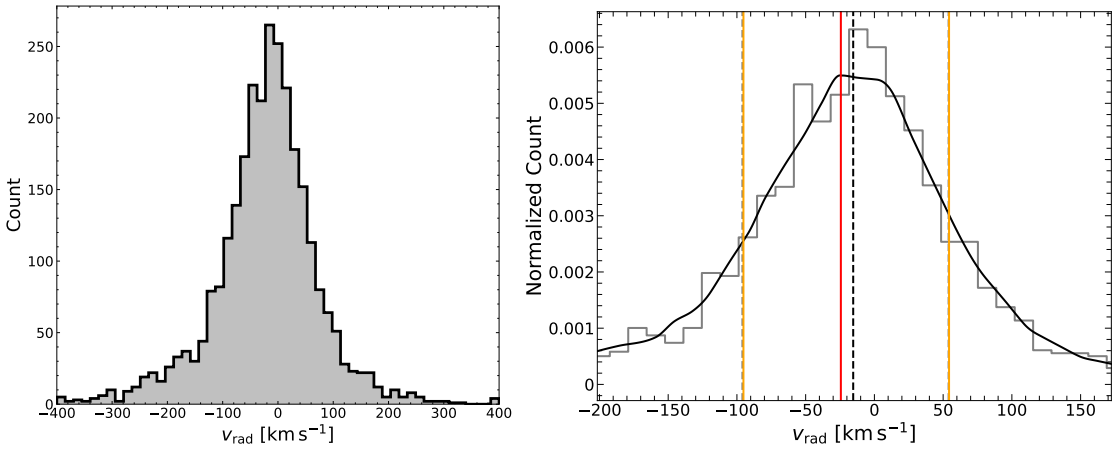


Figure C.2: Stellar radial velocity distribution (left) and its statistical evaluation (right). The mode and HDI is $v_{\mathrm{rad}} = -25^{+79}_{-71}\,\mathrm{km\,s^{-1}}$, while the median with quantiles is $v_{\mathrm{rad}} = -16^{+70}_{-81}\,\mathrm{km\,s^{-1}}$. The confidence intervals reflect the dispersion of the radial velocities and the skewness towards negative radial velocities is most likely an observational bias, because the Eastern Galactic hemisphere, which shows mostly positive radial velocities due to differential Galactic rotation, is far less well covered than the Western side, which mostly shows negative radial velocities (see Figure 4.10).

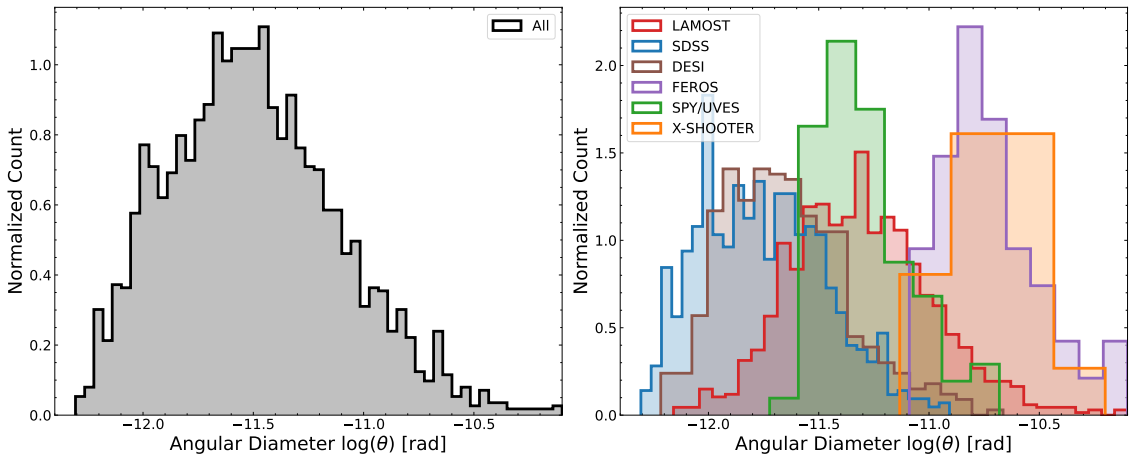


Figure C.3: Angular diamater distribution as derived from the SEDs. The offsets between the different surveys/spectrographs directly reflects the distances to the observed objects. Higher angular diameter on average means closer objects and vice versa.

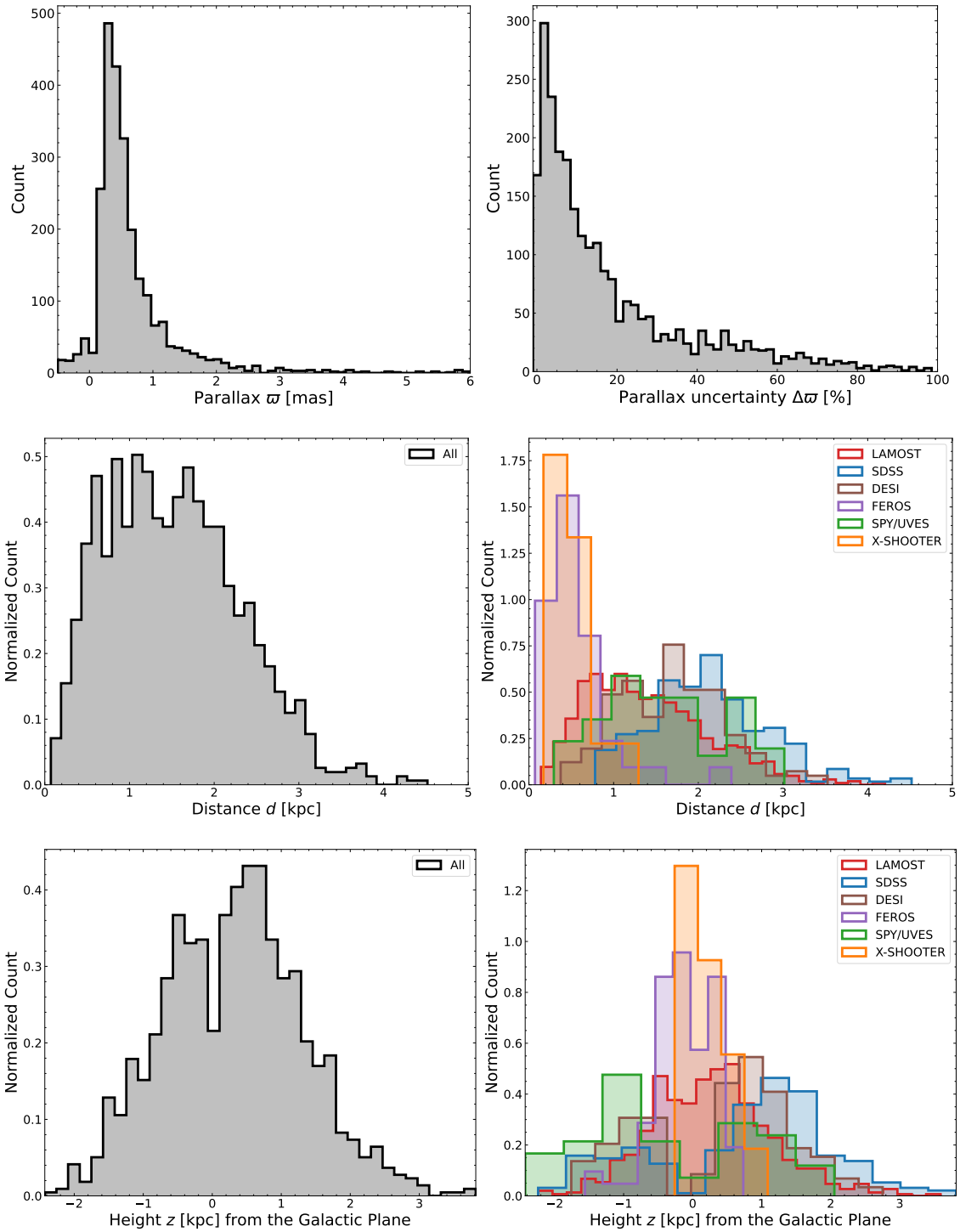


Figure C.4: Parallax, parallax uncertainty, distance, and height from the Galactic plane distribution. It is clearly visible that the objects observed with FEROS and X-Shooter are on average much closer than the others.

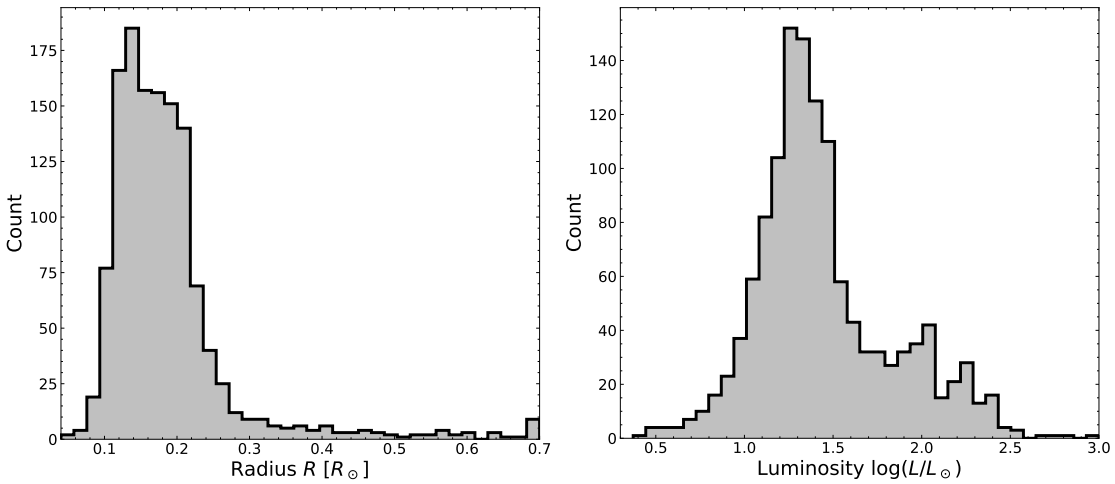


Figure C.5: Radius and luminosity distributions. The typical ranges for hot subdwarf stars of $0.1-0.3\,R_\odot$ and $10-1,000\,L_\odot$ are clearly reproduced.

C.2 Simulated Interstellar Line Parameters

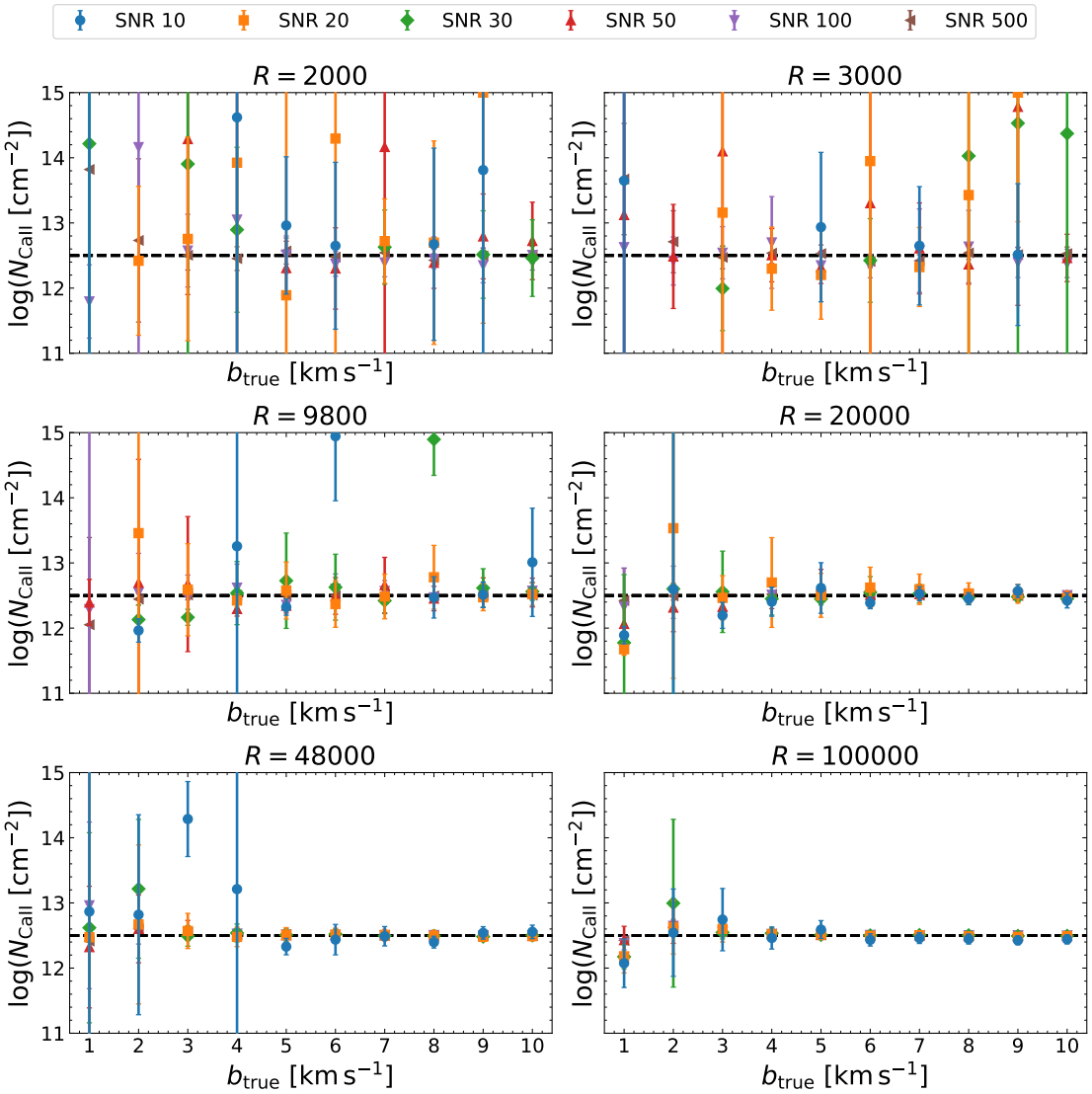


Figure C.6: Analog plot to Figure 4.11 for Ca II with similar behavior.

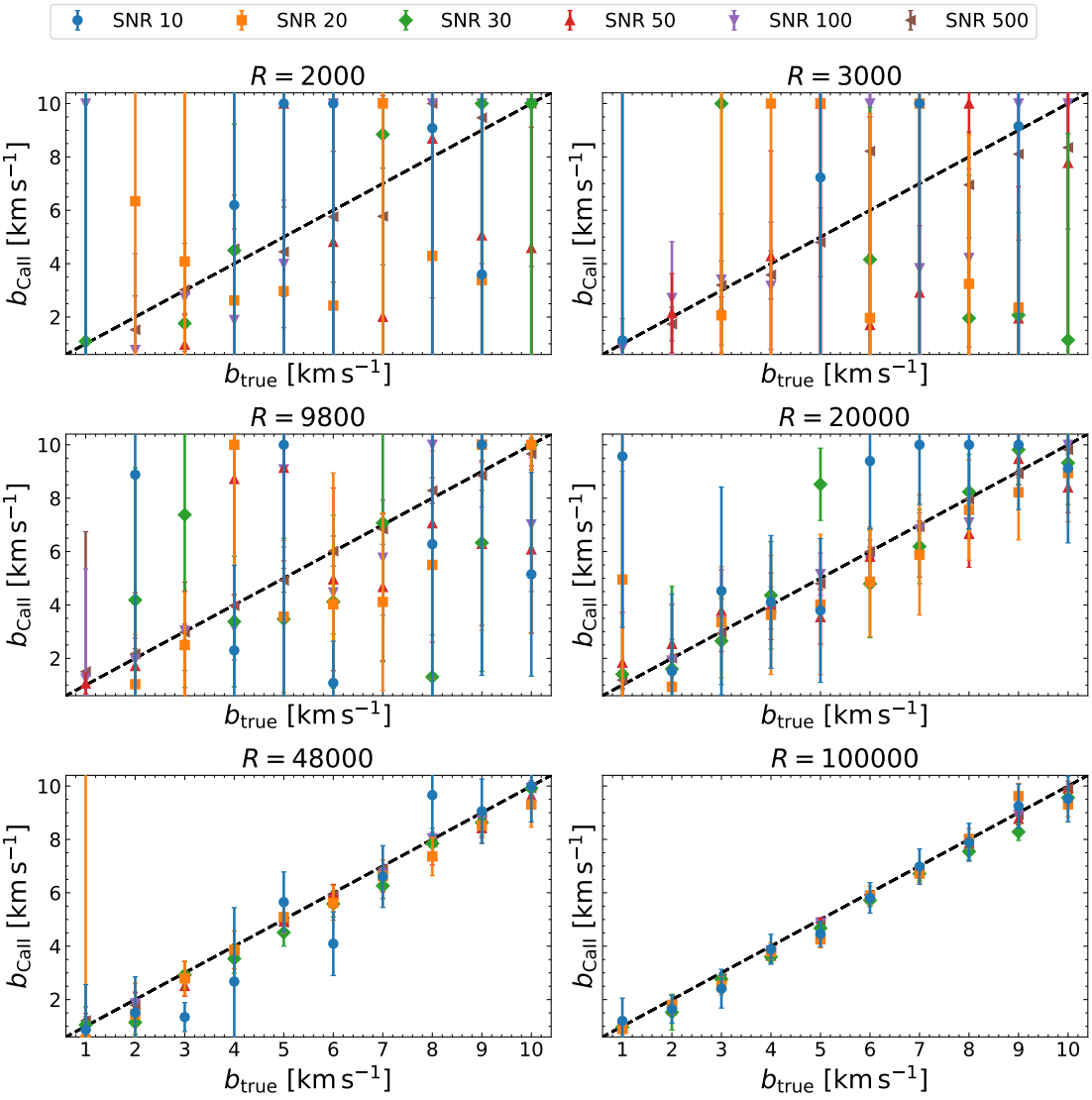


Figure C.7: Analog plot to Figure 4.12 for Ca II with similar behavior.

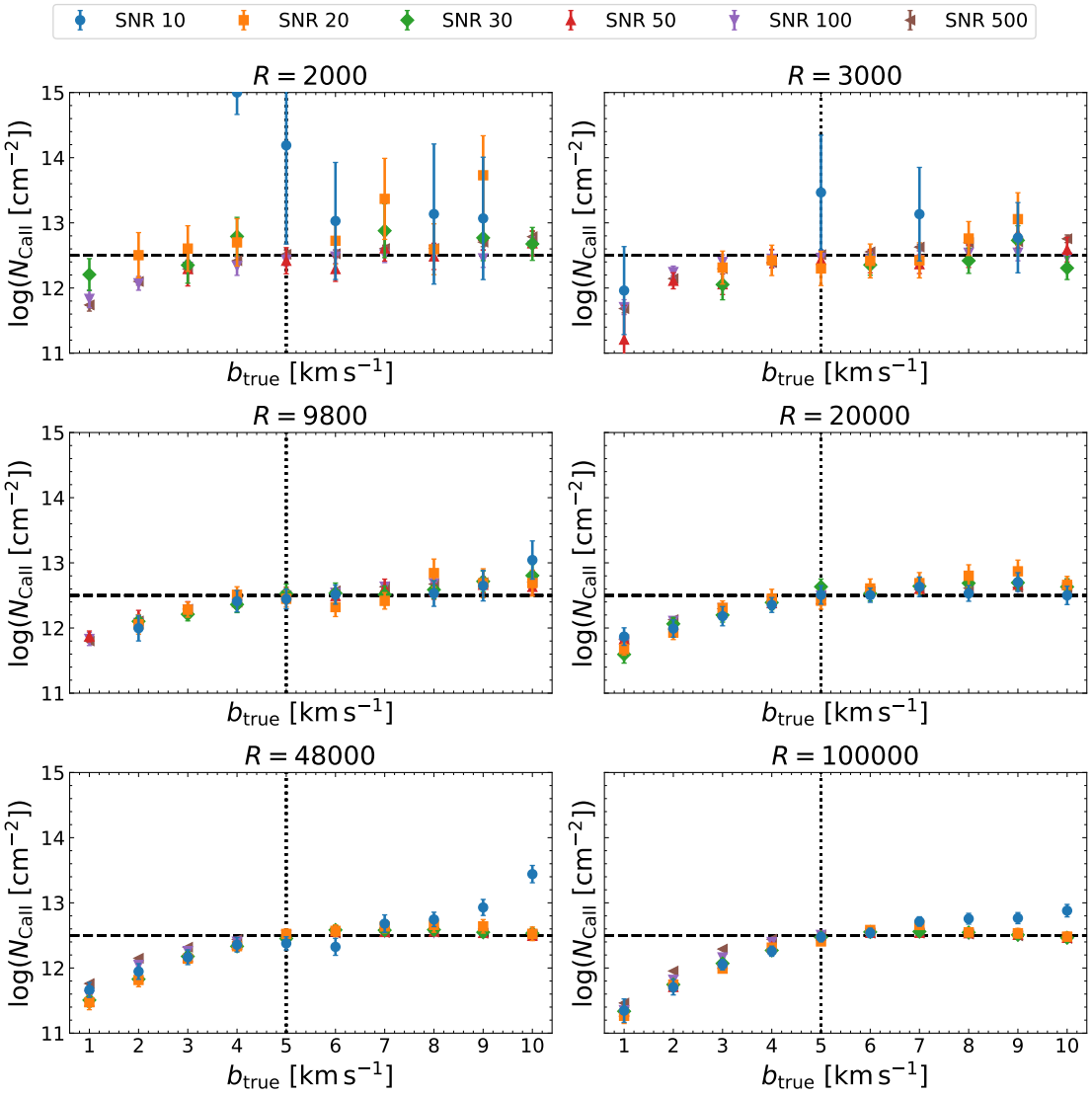


Figure C.8: Analog plot to Figure 4.14 for Ca II with similar behavior.

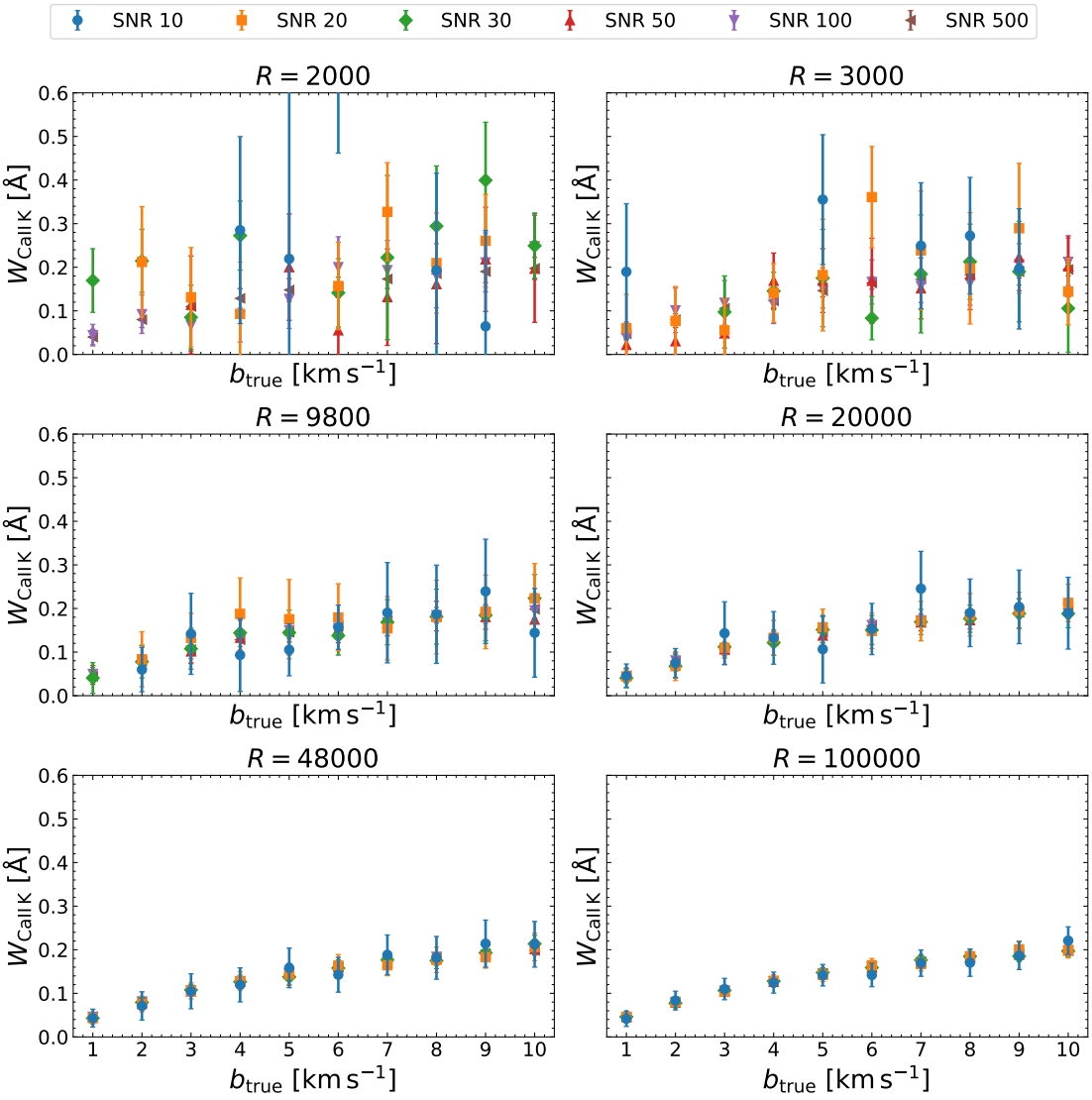


Figure C.9: Analog plot to Figure 4.16 for the Ca II K line with similar behavior.

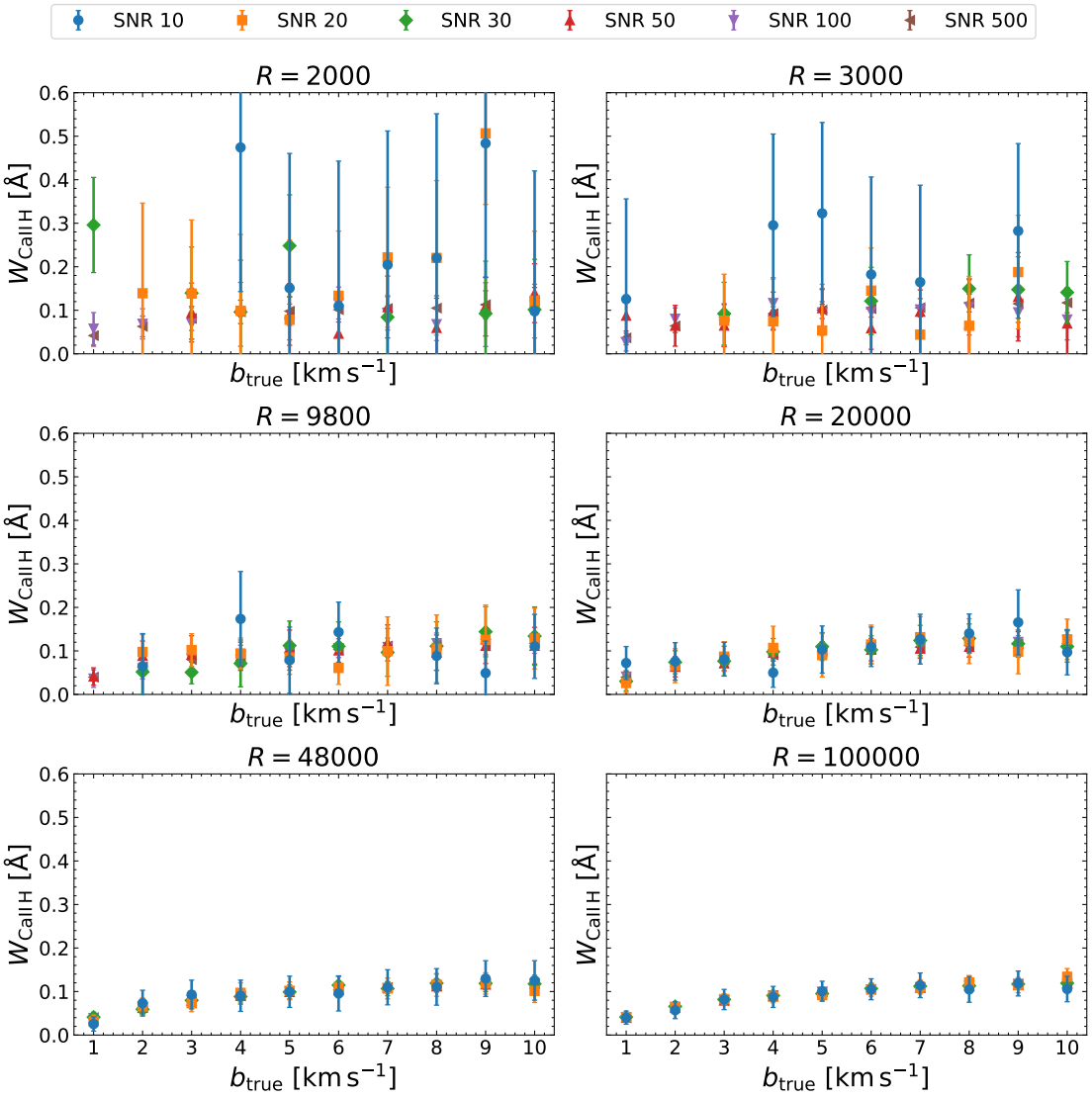


Figure C.10: Analog plot to Figure 4.16 for the Ca II H line with similar behavior.

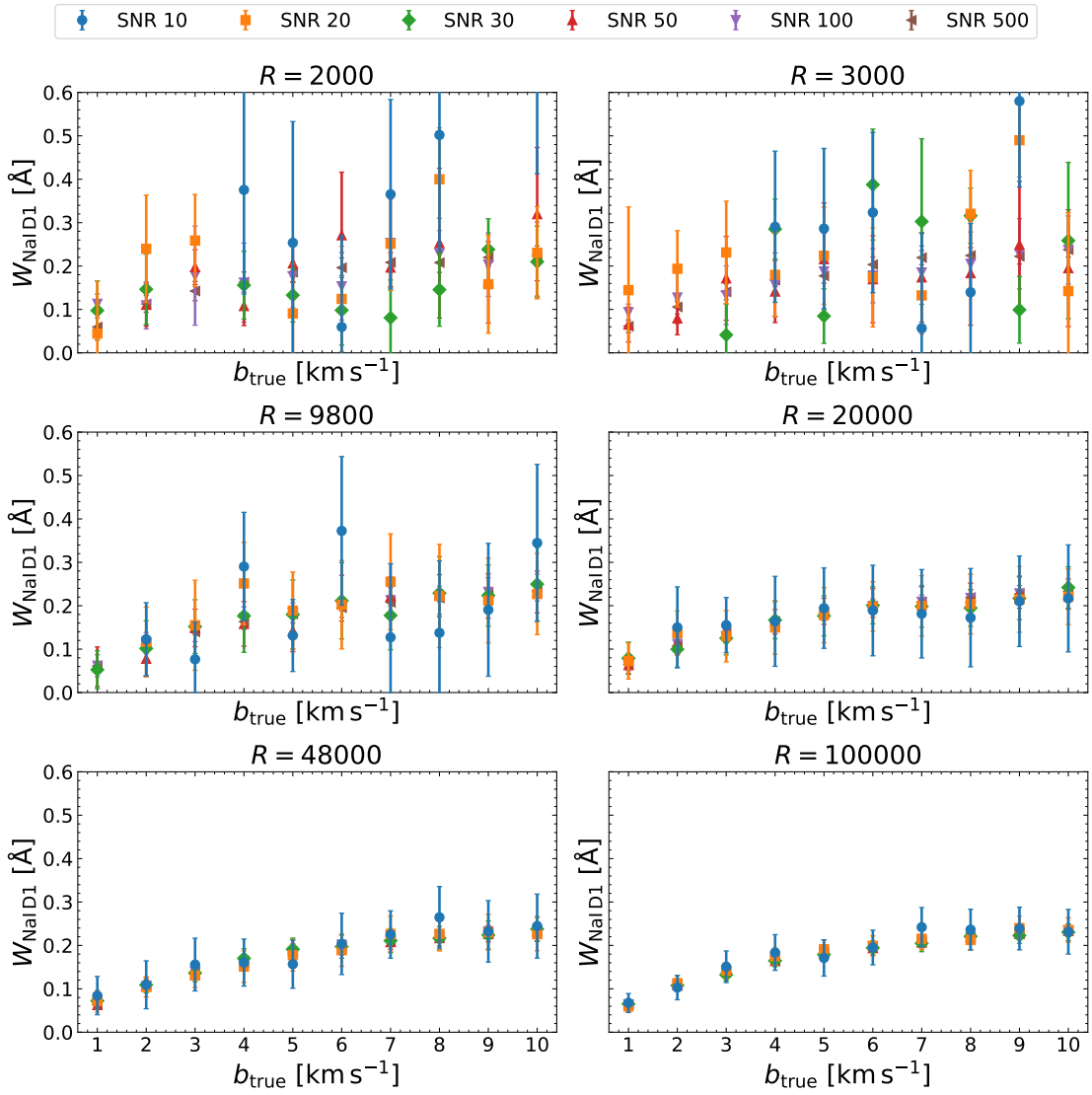


Figure C.11: Analog plot to Figure 4.16 for the Na I D1 line with similar behavior.

Appendix D

Supplementary Material for Chapter 5

D.1 Observed Interstellar Line Parameters and Reddening

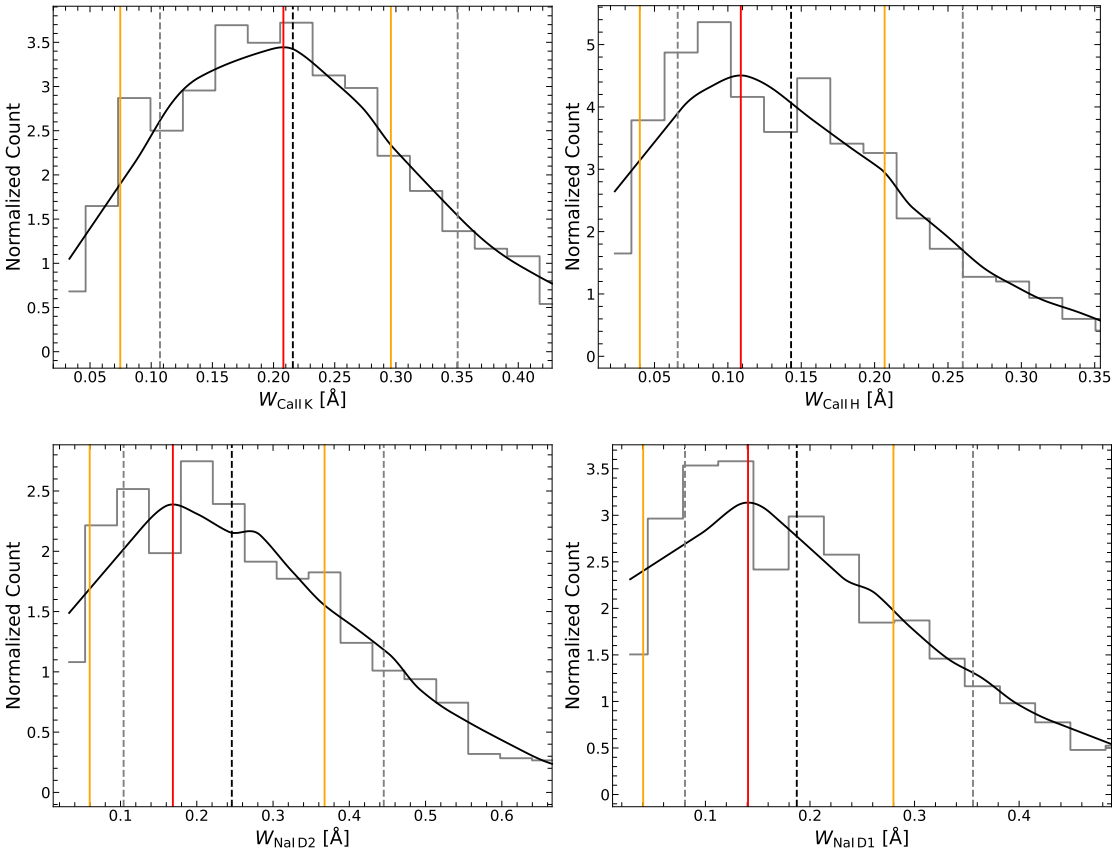


Figure D.1: Statistical evaluations of the equivalent width distributions shown in Figure 5.4. The results are the following: $\bar{W}_{\mathrm{CaII\,K,mode}} = 0.21^{+0.09}_{-0.14}$ Å and $\bar{W}_{\mathrm{CaII\,K,median}} = 0.22^{+0.14}_{-0.11}$ Å, $\bar{W}_{\mathrm{CaII\,H,mode}} = 0.11^{+0.10}_{-0.07}$ Å and $\bar{W}_{\mathrm{CaII\,H,median}} = 0.14^{+0.12}_{-0.08}$ Å, $\bar{W}_{\mathrm{NaI\,D2,mode}} = 0.17^{+0.20}_{-0.11}$ Å and $\bar{W}_{\mathrm{NaI\,D2,median}} = 0.25^{+0.20}_{-0.15}$ Å, as well as $\bar{W}_{\mathrm{NaI\,D1,mode}} = 0.14^{+0.14}_{-0.11}$ Å and $\bar{W}_{\mathrm{NaI\,D1,median}} = 0.19^{+0.17}_{-0.11}$ Å.

D.2 Distribution of Interstellar Gas and Dust on the Sky

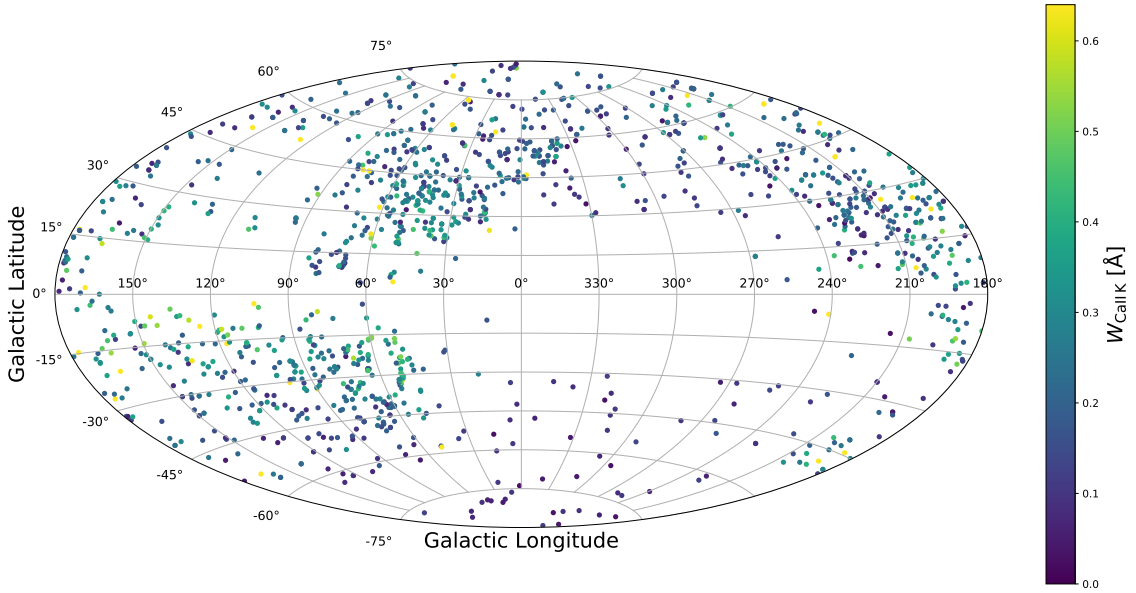


Figure D.2: Similar to Figure 5.11, but for the Ca II K line equivalent width.

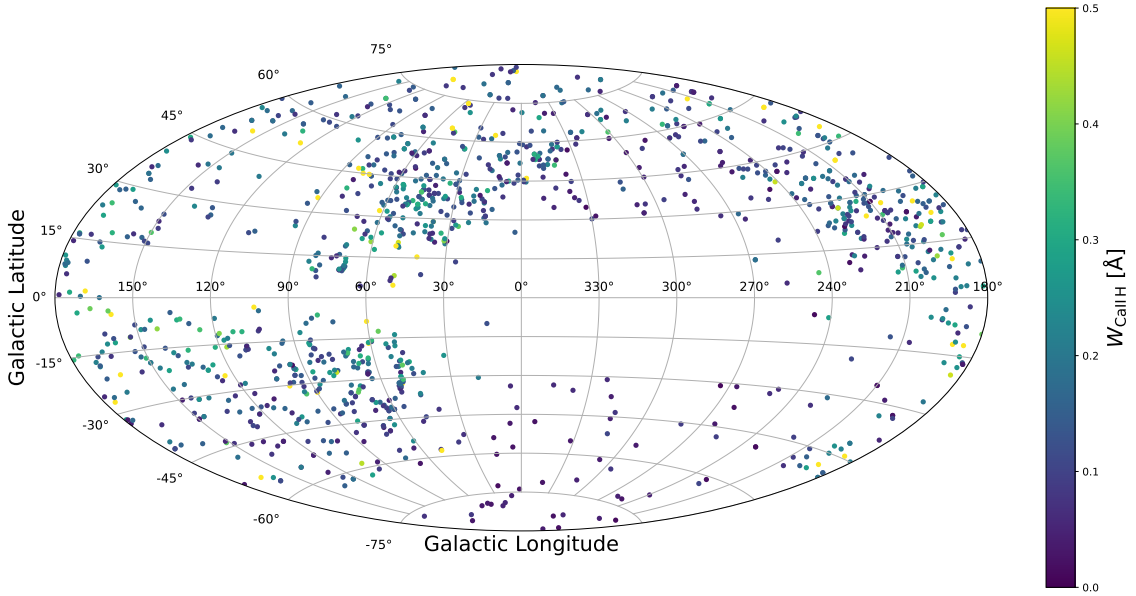


Figure D.3: Similar to Figure 5.11, but for the Ca II H line equivalent width.

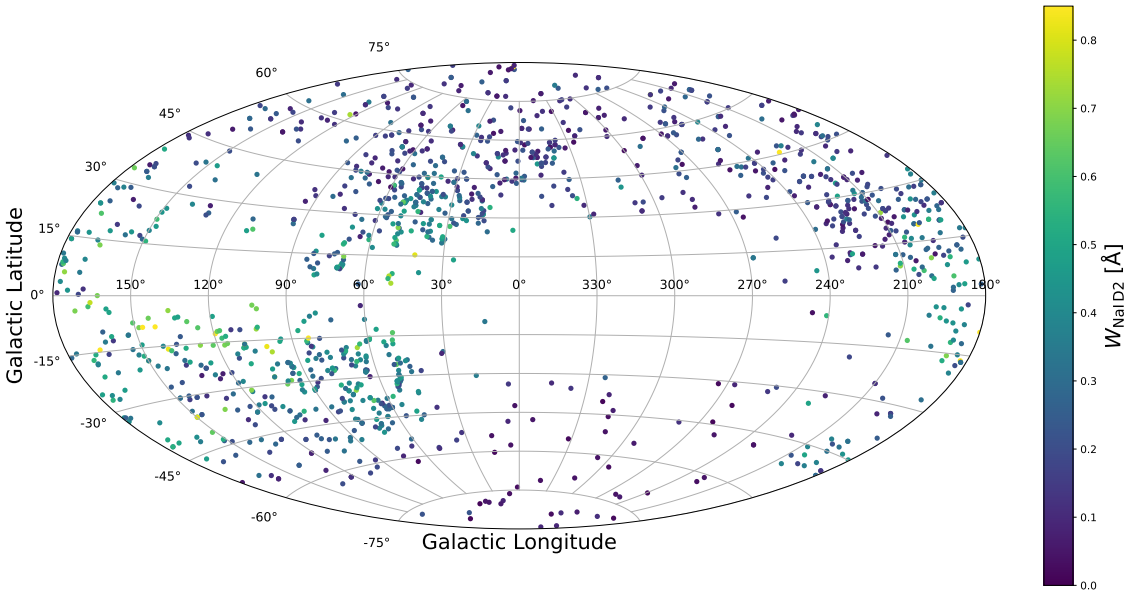


Figure D.4: Similar to Figure 5.12, but for the Na I D2 line equivalent width.

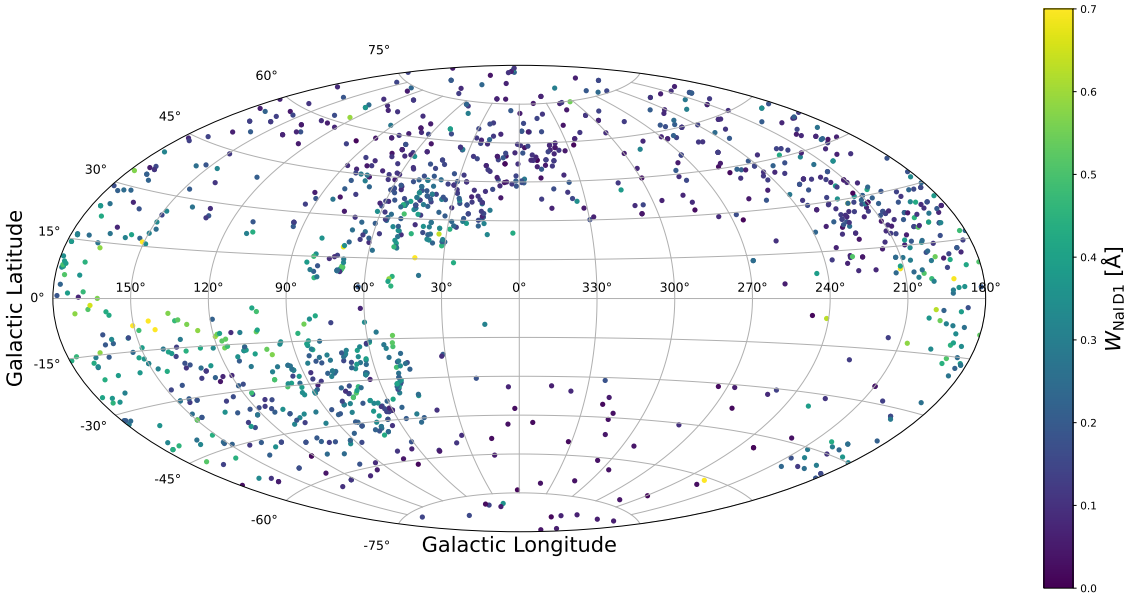


Figure D.5: Similar to Figure 5.12, but for the Na I D1 line equivalent width.

D.3 Variation of Interstellar Properties with Distance, Direction, and Height from the Galactic Plane

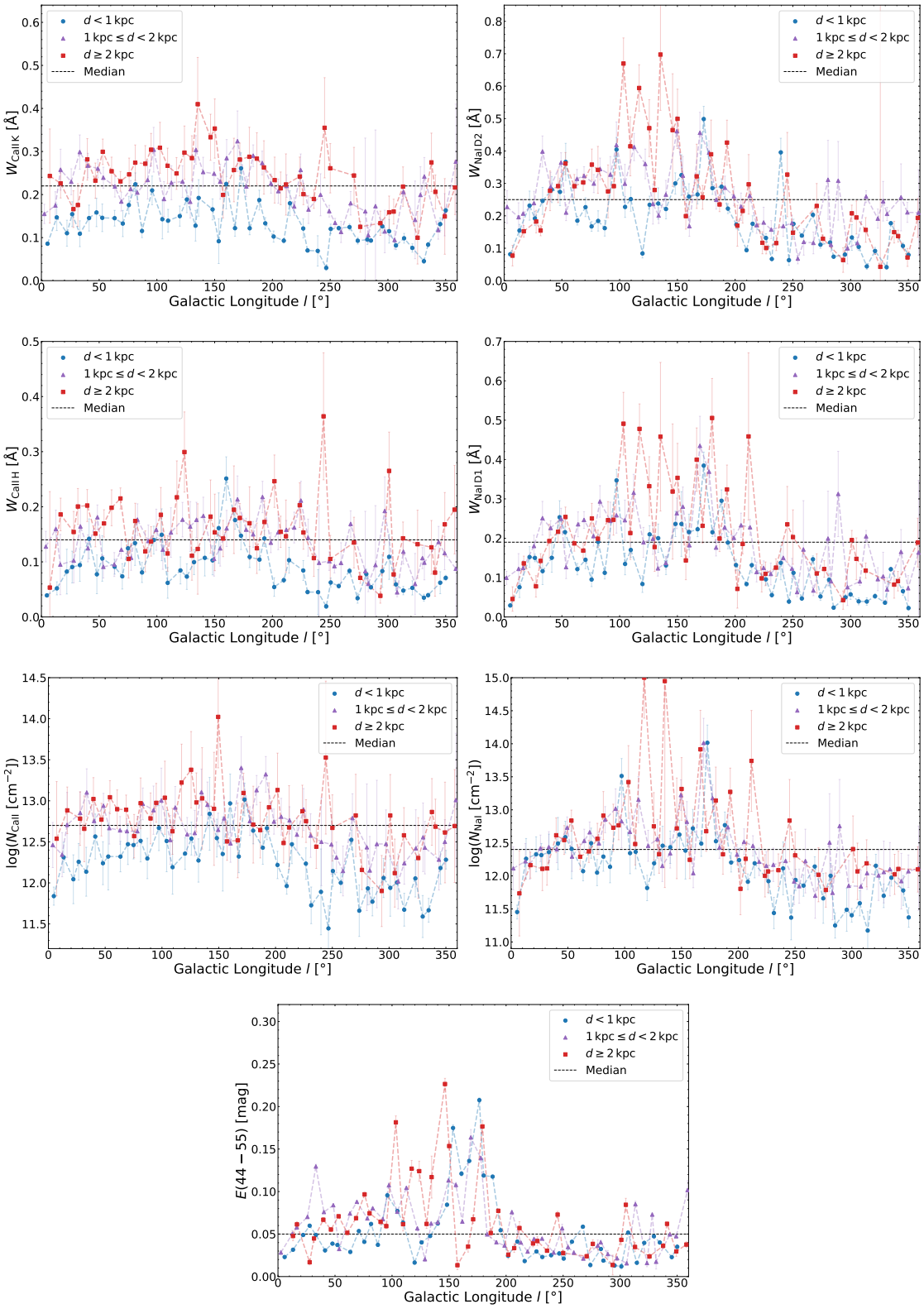


Figure D.6: Analog plot to Figure 5.16, but for Galactic longitude. Except of the Taurus–Perseus–Auriga–California complex at about $100^\circ < l < 210^\circ$ almost no feature is visible, which is more pronounced for Na I and dust (cold neutral medium) than for Ca II (warm medium).

Appendix E

Supplementary Material for Chapter 6.3

Table E.1: Information and measured parameters for the CSM candidate HD 205805 from one of the in total five observations. *Left:* Results from the spectral fit and the ISM line analysis. *Right:* Results from the SED fitting. As described in Section 3.4, the fit always assumes a second component. But if such a component is not present, the surface ratio becomes zero as in this case.

Parameter	**1σ confidence interval**
Information	
Object name	HD-205805
Right ascension	324.75985°
Declination	-46.10152°
Galactic longitude l	353.12160°
Galactic latitude b	-47.79040°
Survey/spectrograph	FEROS
Estimated SNR	45.4
Effective SNR	37.2
Spectral fit	
Model grid	sdB
Resolution offset $R = \lambda/\Delta\lambda$ (fixed)	48000
Resolution slope $1/\Delta\lambda$ (fixed)	0.000 Å^{-1}
Projected rotational velocity $v\sin(i)$	9 ± 1 km s^{-1}
Effective temperature $T_{\rm eff}$	24457 ± 68 K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$	5.001 ± 0.009
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0.00 dex
He abundance $\log(n(\text{He}))$	-1.997 ± 0.006
Radial velocity $v_{\rm rad}$	-69.4 ± 0.1 km s^{-1}
$\chi^2_{\rm red,0}$	3.545
$\chi^2_{\rm red,initial}$	1.469
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	1.212
ISM line analysis	
Equivalent width $W_{\rm CaII\,K}$	0.06 ± 0.00 Å
Equivalent width $W_{\rm CaII\,H}$	0.04 ± 0.01 Å
Equivalent width $W_{\rm NaI\,D2}$ (fixed)	0.00 Å
Equivalent width $W_{\rm NaI\,D1}$ (fixed)	0.00 Å
Column density $\log(N_{\rm CaII}\,(\text{cm}^{-2}))$	11.95 ± 0.02
Column density $\log(N_{\rm NaI}\,(\text{cm}^{-2}))$ (fixed)	0.00
Doppler parameter $b_{\rm CaII}$ (fixed)	5.00 km s^{-1}
Doppler parameter $b_{\rm NaI}$ (fixed)	0.00 km s^{-1}
Radial velocity $v_{\rm rad}$ (estimated)	-69.00 km s^{-1}
$\chi^2_{\rm red,initial}$	1.390
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	1.179

Object: ADP.2016-09-20T12:22:52.050	68% confidence interval
Color excess $E(B-V)$ from SFD (1998)	0.0242 ± 0.0007 mag
Color excess $E(B-V)$ from S&F (2011)	0.0208 ± 0.0006 mag
Color excess $E(44-55)$	0.0072 ± 0.0021 mag
Extinction parameter $R(55)$ (fixed)	3.02
Angular diameter $\log(\Theta\,(\text{rad}))$	-10.1903 ± 0.0016
Parallax ϖ (*Gaia*, RUWE = 1.15)	4.11 ± 0.06 mas
Distance d (*Gaia, mode*)	$243.1^{+2.9}_{-3.0}$ pc
Distance d (*Gaia, median*)	243.1 ± 3.0 pc
Component 1:	
Effective temperature $T_{\rm eff}$ (prescribed)	24460 ± 80 K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$ (prescribed)	5.001 ± 0.014
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0 dex
Helium abundance $\log(n(\text{He}))$ (fixed)	-1.99728
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, mode)	$0.252^{+0.021}_{-0.023}$ kpc
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, median)	$0.250^{+0.021}_{-0.023}$ kpc
Radius $R = \Theta/(2\varpi)$ (mode)	$0.348 \pm 0.005\,R_\odot$
(median)	$0.348 \pm 0.005\,R_\odot$
Mass $M = gR^2/G$ (mode)	$0.442 \pm 0.018\,M_\odot$
(median)	$0.443^{+0.019}_{-0.018}\,M_\odot$
Luminosity $L/L_\odot = (R/R_\odot)^2(T_{\rm eff}/T_{\rm eff,\odot})^4$ (mode)	$38.9^{+1.2}_{-1.1}$
(median)	$39.0^{+1.2}_{-1.1}$
Gravitational redshift $v_{\rm grav} = GM/(Rc)$	$0.809^{+0.028}_{-0.027}$ km s^{-1}
Component 2:	
Effective temperature $T_{\rm eff}$	4100^{+1700}_{-1800} K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$ (prescribed)	4.50 ± 0.10
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0 dex
Helium abundance $\log(n(\text{He}))$ (fixed)	-1.05
Surface ratio $A_{\rm eff}/A_{\rm eff,1}$	$0.00^{+0.09}_{-0.00}$
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, mode)	$0.44^{+0.07}_{-0.06}$ kpc
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, median)	0.44 ± 0.07 kpc
Radius $R = (A_{\rm eff}/A_{\rm eff,1})^{1/2}\Theta/(2\varpi)$ (mode)	$0.08 \pm 0.04\,R_\odot$
(median)	$0.08 \pm 0.04\,R_\odot$
Mass $M = gR^2/G$ (mode)	$\left(2.3^{+114.3}_{-1.8}\right) \times 10^{-4}\,M_\odot$
(median)	$0.008^{+0.010}_{-0.006}\,M_\odot$
Luminosity $L/L_\odot = (R/R_\odot)^2(T_{\rm eff}/T_{\rm eff,\odot})^4$ (mode)	$\left(1.9^{+379.4}_{-2.7}\right) \times 10^{-5}$
(median)	$0.0013^{+0.0064}_{-0.0012}$
Gravitational redshift $v_{\rm grav} = GM/(Rc)$	$0.053^{+0.033}_{-0.028}$ km s^{-1}
Generic excess noise $\delta_{\rm excess}$	0.000 mag
Reduced χ^2 at the best fit	0.83
Parameter reference	Spectroscopy

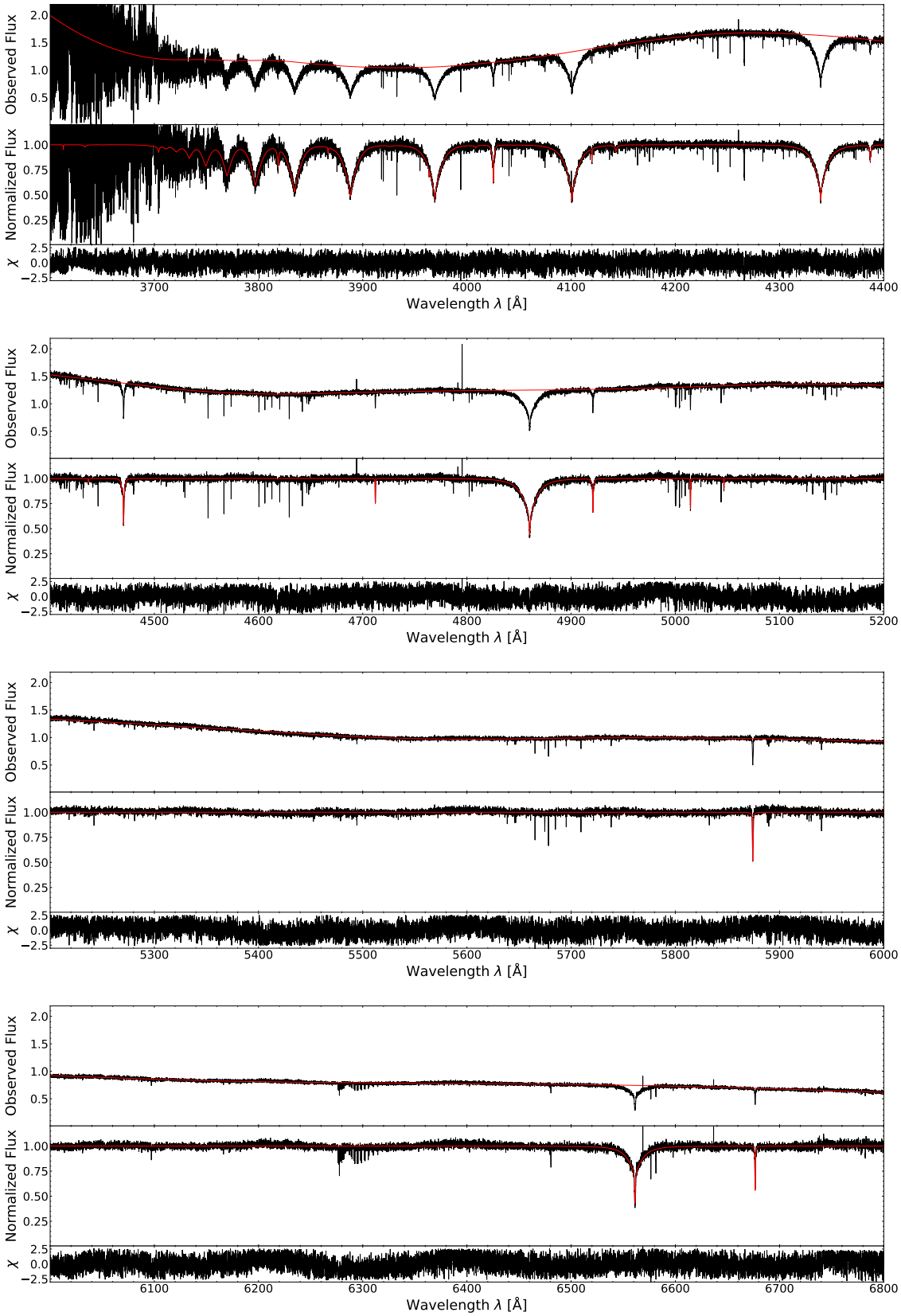


Figure E.1: One full FEROS spectrum of HD 205805 together with the spectral fit. Besides the hydrogen and helium lines included in the synthetic spectrum, it shows a lot of metal lines. At longer wavelengths some ”wiggles“ are visible in the residuals, indicating issues with the calibration. However, it most likely affects the outcomes negligibly.

Table E.2: Information and measured parameters for the CSM candidate EC 03591-3232 from one of the observations. *Left:* Results from the spectral fit and the ISM line analysis. *Right:* Results from the SED fitting.

Parameter	**1σ confidence interval**
Information	
Object name	EC03591-3232
Right ascension	60.24858°
Declination	-32.39508°
Galactic longitude l	232.11087°
Galactic latitude b	-48.85624°
Survey/spectrograph	FEROS
Estimated SNR	49.5
Effective SNR	42.9
Spectral fit	
Model grid	sdB
Resolution offset $R = \lambda/\Delta\lambda$ (fixed)	48000
Resolution slope $1/\Delta\lambda$ (fixed)	0.000 Å^{-1}
Projected rotational velocity $v\sin(i)$	27 ± 1 km s^{-1}
Effective temperature $T_{\rm eff}$	28424 ± 46 K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$	5.875 ± 0.007
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0.00 dex
He abundance $\log(n(\text{He}))$	-1.969 ± 0.005
Radial velocity $v_{\rm rad}$	56.2 ± 0.3 km s^{-1}
$\chi^2_{\rm red,0}$	3.572
$\chi^2_{\rm red,initial}$	1.313
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	1.146
ISM line analysis	
Equivalent width $W_{\rm CaII\,K}$	0.07 ± 0.02 Å
Equivalent width $W_{\rm CaII\,H}$	0.02 ± 0.01 Å
Equivalent width $W_{\rm NaI\,D2}$ (fixed)	0.00 Å
Equivalent width $W_{\rm NaI\,D1}$ (fixed)	0.00 Å
Column density $\log(N_{\rm CaII}\,(\text{cm}^{-2}))$	11.73 ± 0.08
Column density $\log(N_{\rm NaI}\,(\text{cm}^{-2}))$ (fixed)	0.00
Doppler parameter $b_{\rm CaII}$ (fixed)	5.00 km s^{-1}
Doppler parameter $b_{\rm NaI}$ (fixed)	0.00 km s^{-1}
Radial velocity $v_{\rm rad}$	51.81 ± 0.00 km s^{-1}
$\chi^2_{\rm red,initial}$	0.827
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	0.910

Object: ADP.2016-09-21T13:54:48.669	68% confidence interval
Color excess $E(B-V)$ from SFD (1998)	$(5.2 \pm 0.7) \times 10^{-3}$ mag
Color excess $E(B-V)$ from S&F (2011)	$(4.5 \pm 0.6) \times 10^{-3}$ mag
Color excess $E(44-55)$	0.0217 ± 0.0022 mag
Extinction parameter $R(55)$ (fixed)	3.02
Angular diameter $\log(\Theta\,(\text{rad}))$	$-10.4409^{+0.0015}_{-0.0014}$
Parallax ϖ (*Gaia*, RUWE = 1.13)	5.43 ± 0.04 mas
Distance d (*Gaia, mode*)	$184.2^{+1.8}_{-1.7}$ pc
Distance d (*Gaia, median*)	$184.2^{+1.8}_{-1.7}$ pc
Component 1:	
Effective temperature $T_{\rm eff}$ (prescribed)	28420 ± 60 K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$ (prescribed)	5.875 ± 0.013
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0 dex
Helium abundance $\log(n(\text{He}))$ (fixed)	-1.96917
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, mode)	$0.164^{+0.014}_{-0.015}$ kpc
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, median)	$0.163^{+0.014}_{-0.015}$ kpc
Radius $R = \Theta/(2\varpi)$ (mode)	$0.1481^{+0.0014}_{-0.0015}\,R_\odot$
(median)	$0.1480 \pm 0.0015\,R_\odot$
Mass $M = gR^2/G$ (mode)	$0.599 \pm 0.021\,M_\odot$
(median)	$0.599^{+0.022}_{-0.021}\,M_\odot$
Luminosity $L/L_\odot = (R/R_\odot)^2(T_{\rm eff}/T_{\rm eff,\odot})^4$ (mode)	$12.88^{+0.28}_{-0.27}$
(median)	$12.89^{+0.28}_{-0.27}$
Gravitational redshift $v_{\rm grav} = GM/(Rc)$	2.57 ± 0.08 km s^{-1}
Component 2:	
Effective temperature $T_{\rm eff}$	4100^{+1700}_{-1800} K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$ (prescribed)	4.50 ± 0.10
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0 dex
Helium abundance $\log(n(\text{He}))$ (fixed)	-1.05
Surface ratio $A_{\rm eff}/A_{\rm eff,1}$	$0.00^{+0.10}_{-0.00}$
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, mode)	$0.78^{+0.12}_{-0.11}$ kpc
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, median)	$0.79^{+0.13}_{-0.11}$ kpc
Radius $R = (A_{\rm eff}/A_{\rm eff,1})^{1/2}\Theta/(2\varpi)$ (mode)	$0.037 \pm 0.017\,R_\odot$
(median)	$0.037 \pm 0.017\,R_\odot$
Mass $M = gR^2/G$ (mode)	$\left(6^{+227}_{-}\right) \times 10^{-5}\,M_\odot$
(median)	$0.0016^{+0.0019}_{-0.0011}\,M_\odot$
Luminosity $L/L_\odot = (R/R_\odot)^2(T_{\rm eff}/T_{\rm eff,\odot})^4$ (mode)	$\left(5^{+739}_{-6}\right) \times 10^{-6}$
(median)	$\left(2.6^{+12.9}_{-2.4}\right) \times 10^{-4}$
Gravitational redshift $v_{\rm grav} = GM/(Rc)$	$0.024^{+0.014}_{-0.013}$ km s^{-1}
Generic excess noise $\delta_{\rm excess}$	0.002 mag
Reduced χ^2 at the best fit	1.00
Parameter reference	Spectroscopy

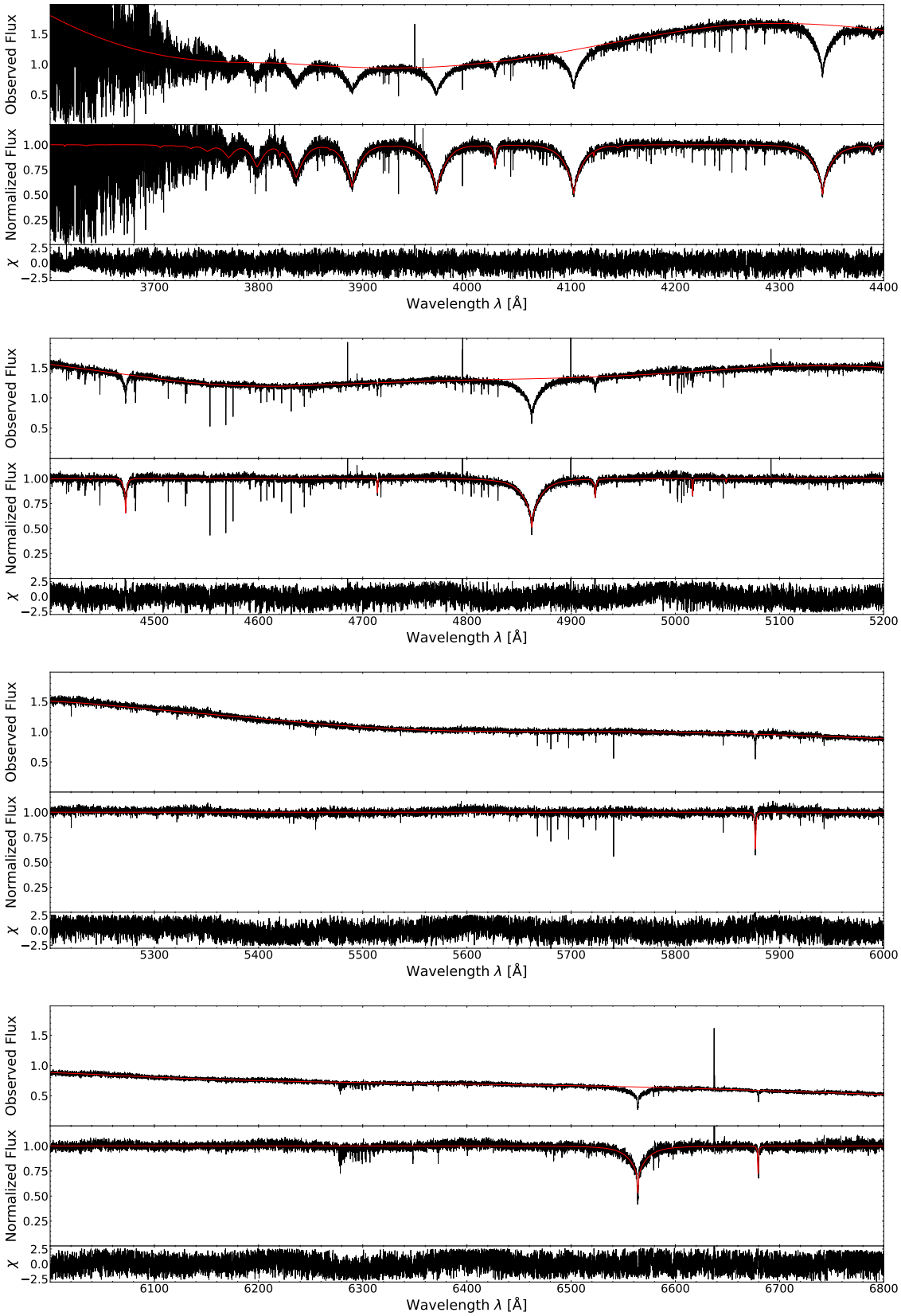


Figure E.2: One full FEROS spectrum of EC 03591-3232 together with the spectral fit. Besides the hydrogen and helium lines included in the synthetic spectrum, it shows a lot of metal lines, similar to HD 205805. Some ”wiggles“ are visible in the residuals, indicating issues with the calibration. However, it most likely affects the outcomes negligibly.

Table E.3: Information and measured parameters for the CSM candidate UVO 0825+15 from one of the observations. *Left:* Results from the spectral fit and the ISM line analysis. *Right:* Results from the SED fitting.

Parameter	**1σ confidence interval**
Information	
Object name	GALEXJ0828+1452
Right ascension	127.14405°
Declination	14.86950°
Galactic longitude l	209.86937°
Galactic latitude b	28.08501°
Survey/spectrograph	FEROS
Estimated SNR	36.1
Effective SNR	31.8
Spectral fit	
Model grid	sdB
Resolution offset $R = \lambda/\Delta\lambda$ (fixed)	48000
Resolution slope $1/\Delta\lambda$ (fixed)	0.000 Å^{-1}
Projected rotational velocity $v\sin(i)$	0 ± 19 km s^{-1}
Effective temperature $T_{\rm eff}$	38476 ± 52 K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$	5.930 ± 0.011
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0.00 dex
He abundance $\log(n(\text{He}))$	-0.783 ± 0.005
Radial velocity $v_{\rm rad}$	31.7 ± 0.2 km s^{-1}
$\chi^2_{\rm red,0}$	3.546
$\chi^2_{\rm red,initial}$	1.271
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	1.127
ISM line analysis	
Equivalent width $W_{\rm CaII\,K}$	0.06 ± 0.02 Å
Equivalent width $W_{\rm CaII\,H}$	0.05 ± 0.01 Å
Equivalent width $W_{\rm NaI\,D2}$ (fixed)	0.00 Å
Equivalent width $W_{\rm NaI\,D1}$ (fixed)	0.00 Å
Column density $\log(N_{\rm CaII}\,(\text{cm}^{-2}))$	11.65 ± 0.09
Column density $\log(N_{\rm NaI}\,(\text{cm}^{-2}))$ (fixed)	0.00
Doppler parameter $b_{\rm CaII}$ (fixed)	5.00 km s^{-1}
Doppler parameter $b_{\rm NaI}$ (fixed)	0.00 km s^{-1}
Radial velocity $v_{\rm rad}$	32.22 ± 1.01 km s^{-1}
$\chi^2_{\rm red,initial}$	0.852
$\chi^2_{\rm red,final}$	1.000
Noise scale factor	0.923

Object: Gaia-DR2-6517452798264581 12-2	68% confidence interval
Color excess $E(B-V)$ from SFD (1998)	0.0391 ± 0.0013 mag
Color excess $E(B-V)$ from S&F (2011)	0.0336 ± 0.0011 mag
Color excess $E(44-55)$	0.028 ± 0.005 mag
Extinction parameter $R(55)$ (fixed)	3.02
Angular diameter $\log(\Theta\,(\text{rad}))$	-10.6538 ± 0.0029
Parallax ϖ (*Gaia*, RUWE = 1.18)	3.71 ± 0.06 mas
Distance d (*Gaia, mode*)	269 ± 4 pc
Distance d (*Gaia, median*)	270 ± 4 pc
Component 1:	
Effective temperature $T_{\rm eff}$ (prescribed)	38480 ± 70 K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$ (prescribed)	5.930 ± 0.015
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0 dex
Helium abundance $\log(n(\text{He}))$ (fixed)	-0.783132
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, mode)	$0.251^{+0.021}_{-0.023}$ kpc
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, median)	$0.250^{+0.021}_{-0.023}$ kpc
Radius $R = \Theta/(2\varpi)$ (mode)	$0.1326 \pm 0.0020\,R_\odot$
(median)	$0.1327^{+0.0021}_{-0.0020}\,R_\odot$
Mass $M = gR^2/G$ (mode)	$0.546 \pm 0.025\,M_\odot$
(median)	$0.547^{+0.026}_{-0.025}\,M_\odot$
Luminosity $L/L_\odot = (R/R_\odot)^2(T_{\rm eff}/T_{\rm eff,\odot})^4$ (mode)	34.7 ± 1.1
(median)	34.8 ± 1.1
Gravitational redshift $v_{\rm grav} = GM/(Rc)$	2.62 ± 0.10 km s^{-1}
Component 2:	
Effective temperature $T_{\rm eff}$	4100^{+1700}_{-1800} K
Surface gravity $\log(g\,(\text{cm s}^{-2}))$ (prescribed)	4.50 ± 0.10
Microturbulence ξ (fixed)	0 km s^{-1}
Metallicity z (fixed)	0 dex
Helium abundance $\log(n(\text{He}))$ (fixed)	-1.05
Surface ratio $A_{\rm eff}/A_{\rm eff,1}$	0^{+5}_{-0}
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, mode)	$1.27^{+0.20}_{-0.18}$ kpc
Spec. distance $d_{\rm spec}$ ($M = 0.47 \pm 0.08\,M_\odot$, median)	$1.29^{+0.20}_{-0.18}$ kpc
Radius $R = (A_{\rm eff}/A_{\rm eff,1})^{1/2}\Theta/(2\varpi)$ (mode)	$0.24 \pm 0.11\,R_\odot$
(median)	$0.23 \pm 0.11\,R_\odot$
Mass $M = gR^2/G$ (mode)	$\left(6.0^{+907.1}_{-2.9}\right) \times 10^{-4}\,M_\odot$
(median)	$0.06^{+0.08}_{-0.05}\,M_\odot$
Luminosity $L/L_\odot = (R/R_\odot)^2(T_{\rm eff}/T_{\rm eff,\odot})^4$ (mode)	$\left(3^{+311}_{-5}\right) \times 10^{-4}$
(median)	$0.010^{+0.052}_{-0.010}$
Gravitational redshift $v_{\rm grav} = GM/(Rc)$	0.16 ± 0.09 km s^{-1}
Generic excess noise $\delta_{\rm excess}$	0.014 mag
Reduced χ^2 at the best fit	1.00
Parameter reference	Spectroscopy

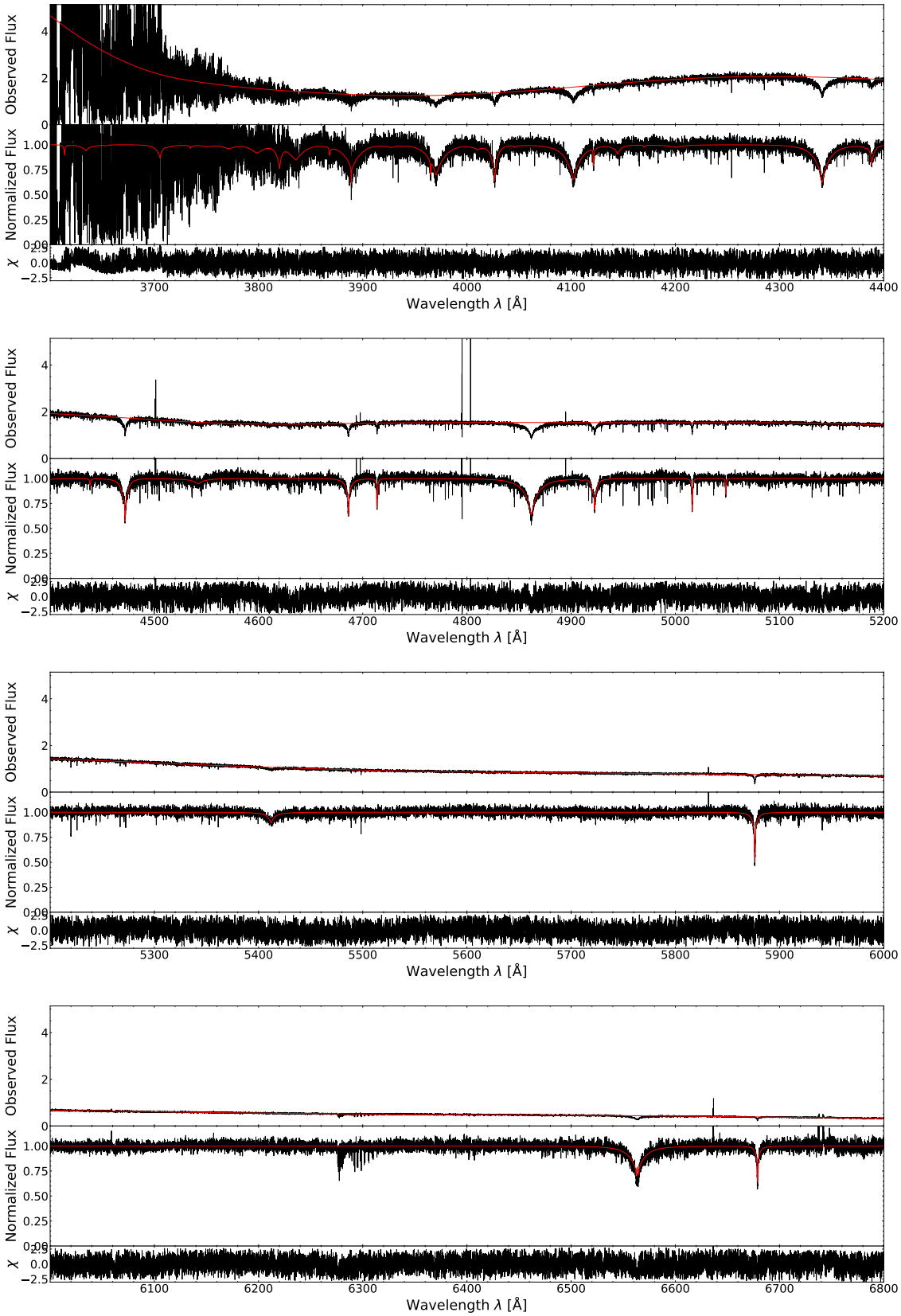


Figure E.3: One full FEROS spectrum of UVO 0825+15 together with the spectral fit. Besides the hydrogen and helium lines included in the synthetic spectrum, it shows metal lines, similar to the other two CSM candidates.

Acknowledgements

First and foremost, I want to thank Uli Heber for giving me the opportunity to work on this project and his exceptional support. Whether in our weekly Zoom meetings, in person at the observatory, or via email, he was always approachable and answered all my questions in great detail. This broadened my understanding of astrophysics not only in context of this work, but also in general.

Furthermore, I want to thank my supervisor for the more technical concerns, Matti Dorsch. In particular at the beginning of my project, he extensively helped me to lay the foundation for the developed set of Python programs and provided several key functions. Also later on he was open to any issue I had and truly everything got resolved.

A particular thanks to Stephan Geier and to all the Potsdam people with whom we closely collaborated. Thanks for the weekly Zoom meetings, the fruitful discussions, and the opportunities for test talks. Additionally, I would like to thank my fellow students in Uli's research group Anna Gebhardt and Lennard Kufleitner for the cooperation and the friendly atmosphere.

I gratefully thank Uli Heber again for letting me speak at the FRANCI meeting at ECAP in Erlangen, at the yearly meeting of the Astronomische Gesellschaft (AG) in Cologne, and especially at the 12th international meeting on hot subdwarfs and related objects (sdOB12) in Little Switzerland, North Carolina, USA. I would like to thank Brad Barlow for organizing this unique sdOB12 conference somewhere in the nowhere of the Appalachian mountains as well as all the fellow delegates from Potsdam and from everywhere else for the valuable scientific discussions and the great time there.

An important aspect for my time at the observatory was the employment as HiWi. Therefore, thanks to Manami Sasaki for continuously extending my contract for the whole time from the Bachelor to the end of my Master. Also a special thanks to Uli Heber for making it possible to extend my HiWi from his own resources in the last month of my project, when the budget for HiWi contracts got drastically shortened by the university.

Moreover, I would like to thank Marina Brinkmann for her constant support and her special care, especially in the last few weeks of this project, when the approaching deadline caused an exponential rise of the stress level.

Last but not least, a big thanks to my family and friends for all the support throughout this project!

Eigenständigkeitserklärung

Hiermit versichere ich, Sebastian Weich (22527340), die vorgelegte Arbeit selbstständig und ohne unzulässige Hilfe Dritter sowie ohne die Hinzuziehung nicht offengelegter und insbesondere nicht zugelassener Hilfsmittel angefertigt zu haben. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen und wurde auch von keiner anderen Prüfungsbehörde bereits als Teil einer Prüfung angenommen.

Die Stellen der Arbeit, die anderen Quellen im Wortlaut oder dem Sinn nach entnommen wurden, sind durch Angaben der Herkunft kenntlich gemacht. Dies gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen sowie für Quellen aus dem Internet.

Mir ist insbesondere bewusst, dass die Nutzung künstlicher Intelligenz verboten ist, sofern diese nicht ausdrücklich als Hilfsmittel von dem Prüfungsleiter bzw. der Prüfungsleiterin zugelassen wurde. Dies gilt insbesondere für Chatbots (insbesondere ChatGPT) bzw. allgemein solche Programme, die anstelle meiner Person die Aufgabenstellung der Prüfung bzw. Teile derselben bearbeiten könnten.

Des Weiteren ist mir bekannt, dass die gemeinsame Bearbeitung der Aufgabenstellung mit anderen Personen in einem Raum oder mithilfe sozialer Medien eine unzulässige Hilfe Dritter im o.g. Sinne darstellt, wenn nicht ausdrücklich Gruppenarbeit vorgesehen ist. Jeder Austausch mit anderen Personen mit Ausnahme von Prüfenden und Aufsichtführenden während der Prüfungszeit über Aufbau oder Inhalte der Prüfung oder Informationen (z.B. Quellen) ist unzulässig. Gleiches gilt für den Versuch der jeweiligen Handlung.

Verstöße gegen die o.g. Regeln sind als Täuschung bzw. Täuschungsversuch zu qualifizieren und führen zu einer Bewertung der Prüfung mit „nicht bestanden“.

Erlangen, den 30.09.2025

Ort, Datum	Sebastian Weich