Rijksuniversiteit Groningen

Bachelor Thesis

Tracing the Inner Galactic Halo with Blue Horizontal Branch Stars

Using Spectrophotometry from *Gaia* Data Release 3

rijksuniversiteit groningen

Author:
Annemarijn Zwerver

Supervisors:
Else Starkenburg, Manuel Bayer

Date:
29/06/2023

Second reader:
Eline Tolstoy

Abstract

In this thesis we draw synthetic photometry from novel low-resolution spectra provided by *Gaia* Data Release 3 to discriminate blue horizontal branch stars from the contamination population of blue stragglers. We investigate the use of several filter combinations and use synthetic *Pristine* CaHK and synthetic SDSS *ugr* photometry to create an all-sky sample of blue horizontal branch star candidates with a purity of 99% and a completeness of 75%. To pick-up fainter magnitudes, we use SDSS DR18 *ugr* photometry in combination with synthetic Pristine CaHK photometry, improving the completeness to 81%, but limiting us to the SDSS footprint. We also investigate the use of J-PAS 4200 photometry to compensate for the relatively large uncertainties in CaHK and suggest follow-up. We subsequently use these samples of standard candles to create 3D maps of the inner halo of the Milky Way and investigate spatial overlap with stellar streams, including the Sagittarius stream. With the currently available XP spectra of roughly 200 million stars we are restricted to 17.65 mag, but we foresee large potential in future Gaia releases, where the XP spectra of 2 billion sources (down to 20.5 mag) will become available, and by applying the methods we present in this thesis we might be able to significantly increase the number of BHB candidates, as well as mapping the halo out to distances of 100 kpc.

Acknowledgements

I would like to thank my supervisors Else Starkenburg and Manuel Bayer for their support during my thesis. Besides our weekly meeting, they were always available, whether for questions, feedback or to provide important insights. My supervisors carry an enthusiasm for their research, which has been very contagious. I have learned a great deal the past weeks and it has been a pleasure to work with these people.

In addition, I would like to thank Eduardo Balbinot for providing me with synthetic Pristine CaHK and synthetic J-PAS photometry for A-type candidate stars, as this data is not included in the GSCP.

I would also like to thank my fellow students Abel Kersten and Berber Schuurman for their company and support while working on this thesis.

This thesis marks the end of my Bachelor in Astronomy and it has not always been an easy journey. Therefore I would also like to thank my wonderful parents, my brother and sister, my boyfriend, family and friends for all their love and support throughout these years.

This work has made use of data from the European Space Agency (ESA) mission Gaia (*https://www.cosmos.esa.int/gaia*), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, *https://www.cosmos.esa.int/web/gaia/dpac/consortium*).

This work has made use of data from the Sloan Digital Sky Survey (SDSS) data release 18 (*http://www.sdss.org/*).

Contents

1 Introduction

1.1 Galactic archaeology in the stellar halo of the Milky Way

Over the past 20 years the stellar halo component of our Milky Way galaxy has become an increasing field of interest. The halo stretches far beyond the Galaxy's bright central regions, encompassing both the bulge and the disk of our Milky Way. The Galactic halo could be considered the most important component in unravelling the formation history and evolution of the Milky Way. From simulations based on the ΛCDM model we know that structure formation happens hierarchically, implying that the Milky Way galaxy was likely formed via merging processes and accretion of smaller structures, such as galaxies. In this merging process, smaller galaxies can become tidally disrupted and form stellar streams (Helmi, 2020).

One reason to search the halo for clues about our Galaxy's past is that it contains some of the oldest and most metal-poor stars. Some of these stars have lifetimes comparable to the age of the Universe and are thought to have formed during the formation of the Milky Way. For this reason the chemical abundances, ages, and kinematics of these stars form a fossil-record for astronomers, containing information about the formation history of the Milky Way Galaxy over the past ~ 10 Gyr. The implementation of this idea of Galactic archaeology has been accelerated specifically by the second data release of *Gaia* in combination with high resolution spectroscopy, most notably APOGEE. Using the concept of Galactic archaeology we now know that the stellar halo component contains both stars from the original disk of the Milky Way and from destroyed satellites. In fact, most of the mass in the stellar halo is thought to originate from a few massive merging events. See for a more comprehensive review on Galactic archaeology the work of Helmi (2020) and references therein.

Another interesting feature of the halo component of our Galaxy is its in-homogeneity. The halo shows a numerous amount of over-densities and kinematical substructures of different forms and sizes (Mateu, 2023). One large kinematical substructure is the Sagittarius stellar stream ($\sim 10^8 M_\odot$) which originates from the Sagittarius dwarf galaxy and is the result of billions of years of merging (Helmi, 2020). The Sagittarius stream has been widely studied, because it wraps around the Milky Way, hereby covering a large distance range. For this reason, the Sagittarius stream can be used to constrain the dark matter potential of the Milky Way (Law and Majewski, 2010). More recently, an even larger substructure ($\sim 10^9 M_\odot$), named *Gaia*-Enceladus, has been discovered which is thought to be responsible for a large fraction of the inner halo (Helmi, 2020). When stellar streams are destroyed completely, they start to become a part of the smooth stellar halo (Helmi, 2020).

Mapping the halo and its substructures allows astronomers, among other things, to distinguish the smooth component of the Milky Way and calculate the halo's density profile (see for example Thomas et al. 2018). A steep stellar density profile implies that the galaxy has gone through a more active accretion history (Libeskind et al., 2011). Consequently, knowing the density profile of our galactic halo provides information about the formation of the Milky Way. Additionally, it also helps on putting constraints on the models that describe the amount of dark matter, the dark matter profile, and the mass of the Milky Way (Helmi, 2020).

Standard candles form an essential part in tracing the outer parts of our Milky Way (see for example Starkenburg et al. 2019). Parallax measurements start to become less accurate beyond the Solar neighborhood, but the consistent absolute magnitudes of variable stars like

RR Lyrae, M-giants, carbon stars, and blue horizontal branch stars (BHB) stars form a reliable way to determine distances (Starkenburg et al., 2019). Standard candles can be used to map the Galactic halo and to study various (kinematical) substructures in the outer halo (Starkenburg et al., 2019). Comparing the spatial distribution of different standard candles -for instance RR Lyrae and BHB stars- can provide us with information about progenitor galaxies, such as their metallicity and dominant stellar population (Starkenburg et al., 2019).

1.2 Blue horizontal branch stars

BHB stars are hot and bright A-type stars ($7500K \leq T_{eff} \leq 9000K$) (Thomas et al., 2018), with longer lifetimes and higher masses than main-sequence (MS) A-type stars (Yanny et al., 2000). BHB stars can be used as standard candles, because their low consistent absolute magnitudes ($M_g \sim 0.5$ in the Sloan Digital Sky Survey (SDSS) g band) make them identifiable in the outer halo ($R_{GC} > 20$ kpc), up on till distances bigger than 100 kpc (see for exampleThomas et al. 2018). In the work of Fukushima et al. (2018) the halo is traced with BHB stars between r = 50 kpc to 300 kpc, which is well below the virial radius of the dark matter halo (200-300 kpc). A clear boundary of the stellar halo could not be identified. BHB stars form a part of the old stellar population in the halo (Thomas et al., 2018), where their color space is not contaminated by many other stars with of A-type colors (Yanny et al., 2000). Hot and young MS stars are very rare in this environment. For one, no significant star formation has taken place in the outer halo over the last billion years, which constitutes a greater timescale than the average lifetime of an A-type MS stars (Yanny et al., 2000). Second, A-type MS stars that are formed in star-forming regions in the Galactic disk do not have the velocities required to travel towards the far halo (Yanny et al., 2000).

Blue straggler (BS), however, can be found in any stellar population (Starkenburg et al., 2019), including the outer halo environment. They have similar temperatures and optical colors as BHB stars, thereby contaminating their color space (Starkenburg et al., 2019). BS are MS stars that go beyond the MS turn-off point, appearing more luminous and bluer in the Hertzsprung Russel Diagram (Starkenburg et al., 2019). According to the theory for standard evolution of stars BS should have left the MS. The most probable explanation that these stars are still on the main-sequence is that they exist in binary systems, where they gain mass by accretion from the other member to stay longer on the MS (LeBlanc, 2010). As BS are intrinsically fainter than BHB stars and do not have a consistent absolute magnitude as standard candles, they can give a crooked picture of the halo when not making a proper distinction (Starkenburg et al., 2019).

Even though BHB stars are similar to BS, underlying physical differences between them result in spectral differences. Most notably, BS are more compact than BHB stars, and consequently have larger surface gravities (Starkenburg et al., 2019). This is because BS originate from MS hydrogen burning, whereas BHB stars are giant helium burning stars (Starkenburg et al., 2019). In general, BHB stars have log $g = 2.8 - 3.75$ and blue stragglers have log g $= 3.75 - 5.0$ (Starkenburg et al., 2019).

By comparing synthetic spectra of BHB stars, with those of BS it can be seen that certain features are sensitive to surface gravity (see Figure 1). Surface gravity, which correlates to the pressure broadening by the hydrostatic equilibrium, widens spectral lines. In A-type stars the Balmer lines, which is the set of hydrogen transitions emanating from n = 2 (LeBlanc, 2010), are very strong. The wings of the Balmer lines of hydrogen are wider in an atmosphere with a higher

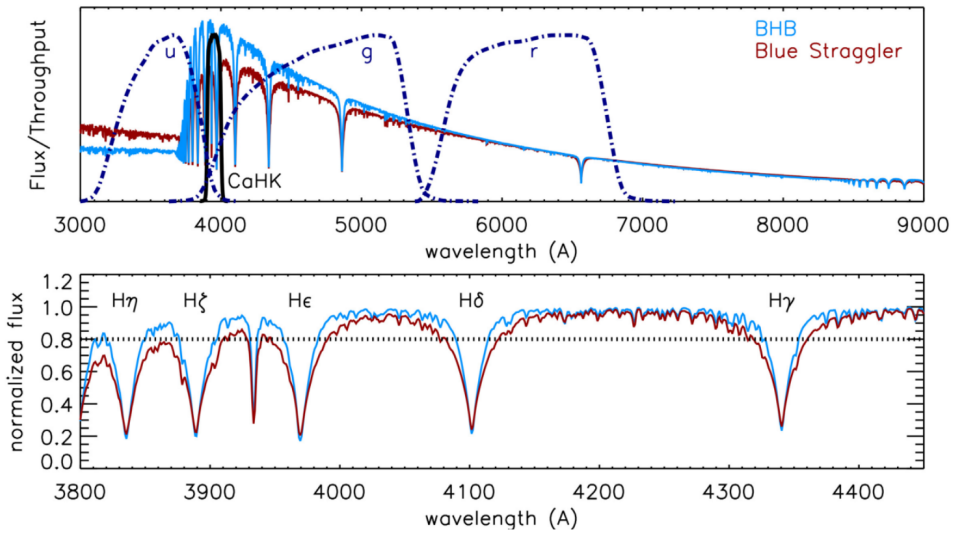

Figure 1: Synthetic stellar spectra from Munari et al. (2005) for stars with similar A-type star temperatures and metallicities, but different surface gravities which expose the difference between blue horizontal branch stars and blue stragglers. The blue dotted curves in the top panel are the transmission curves for the Sloan Digital Sky Survey *ugr* filter bands and the black solid curve is the transmission curve of the CaHK filter from the Pristine survey (Starkenburg et al., 2017). The bottom panel shows a narrower-wavelength range such that the Balmer series of hydrogen are visible. Figure from Starkenburg et al. (2019).

surface gravity and thus wider in BS (see bottom panel of Figure 1) (LeBlanc, 2010). The Balmer jump, a large flux decrease due to ionization from n = 2 (near 3800 Å), is steeper for stars of lower surface gravity (Starkenburg et al., 2019). Before the Balmer jump BS are more luminous, while after the Balmer jump BHB stars are more luminous. To a lesser extent a similar effect can be seen at the Passchen series and jump, which is the set of hydrogen transitions emanating from n = 3 around 8700 Å. At redder wavelengths the temperature difference starts to dominate over the gravity difference due to the sensitivity of the black-body curve (Starkenburg et al., 2019).

1.3 Previous work

For the reasons mentioned above, spectroscopic measurements can be used to distinguish BHBs from BS stars. For example, in the work of Xue et al. (2008) the gravity sensitivity of the width of the Balmer lines is employed to select BHBs from spectral libraries of A-type stars. The disadvantage of spectroscopy is that it is a time-consuming process and one that involves difficulties for fainter stars. For this reason, photometric methods have been developed. Spectral differences between the two stars can be detected as band color difference when opting the right filters.

Previous works generally make use of broad-band photometry information to select BHB stars. In the work of Deason et al. (2011) a method is developed to select BHB stars based on *ugr* SDSS broad-band photometry. They define the probability of BHB or BS membership by the locus of these stars in $u-g$, $g-r$ space. In the work of Fukushima et al. (2018) photometry from Hyper Suprime Cam (HSC) is used and BHB stars are then selected by performing color cuts to

select on stars with A-type-like temperatures. The success of their method can be expressed in a completeness of 67 per cent and a purity of 62 percent. In the work of Thomas et al. (2018) information in the *u*-band from CFIS and the *griz* band from Pan-STARRS are combined and given to a discrete classification algorithm to separate BHBs and BSs. This creates a sample with a completeness of 71 % and a contamination of 24 %. All in all, using information from broad-band photometry alone has lead to samples with an average completeness of 50-70 % and contamination rates as high as 30 %, where these success rates decrease at fainter magnitudes (see for example Starkenburg et al. 2019).

In the work of Starkenburg et al. (2019) *CaHK* narrow-band photometry of the Pristine survey (Starkenburg et al., 2017) together with SDSS *ugr* photometry is used to create a sample of BHB stars with a purity of 93 % and a completeness of 91 %. Although the Pristine survey is designed to be sensitive for metals in FGK stars, its narrow wavelength window coincides with the wavelength range where the flux difference between the BHB and BS stars is very large, as is visible in Figure 1. Their unprecedented clean and pure sample is used to investigate the Galactic halo for substructures over the Pristine footprint. Multiple parts of the Sagittarius stream have been traced and suggest that the apocentre of the trailing arm extends even further than previously measured. Furthermore, their sample has demonstrated that the halo follows a density profile with a negative power law of 3.5-4.

Besides the completeness and purity of a sample, it is also of importance to consider the limiting magnitudes of the filters, or in other words, how deep or shallow the photometry is in the filters. In addition, the sky coverage of the surveys is also of importance. In the technique of Starkenburg et al. (2019) the relatively small Pristine Survey footprint of 1000 deg^2 (which currently covers about 6000 deg^2) and the relatively shallow SDSS *u* band are the limiting factors. The uncertainties in the SDSS *u* band photometry become significantly large for $g \geq 20$.

1.4 *Gaia* low-resolution spectrophotometry

At the in-between of broad-band photometry surveys and time-costly spectroscopy, novel spectrophotometry is provided by the recently published third data release from the all-sky, space-based *Gaia* observatory (Gaia Collaboration et al., 2016) from European Space Agency (ESA) (Gaia Collaboration et al., 2022). *Gaia* has measured novel low-resolution spectrophotometry ($\lambda/\Delta\lambda \approx 25 - 100$) for about 220 million sources. These low resolution spectra are measured in the wavelength range of *Gaia*'s *BP* and *RP* color bands, otherwise known as the XP spectra : 330 $nm \leq \lambda \leq$1050 nm (see Figure 4). The XP spectra are continuously defined as an array of coefficients to be applied to a set of basis functions. The upper panel in Figure 3 includes an example of how the XP spectra of a BHB star looks like, compared to the spectra of a BHB star measured with high-resolution spectrometry from SDSS DR17 (Abdurro'uf et al., 2022). XP spectra offer the ability to obtain synthetic photometry for any passband that is fully enclosed in this wavelength range. The *Gaia* Synthetic Photometry Catalog (GSPC) provides such synthetic photometry for several widely used photometric systems. XP spectra will be published for the entire *Gaia* data in future releases, which includes $\sim$ 2 billion sources. In this release only source mean spectra are available, which are spectra that have been generated from multiple observations of the same source. In future releases single-observation spectra (epoch spectra) will become available. In the release paper Gaia Collaboration et al. (2022) the performance of the XP synthetic photometry is explored and, among other things, the medium-width passbands from the *Gaia* C1 system (Jordi et al., 2006) are used to roughly separate giants from main

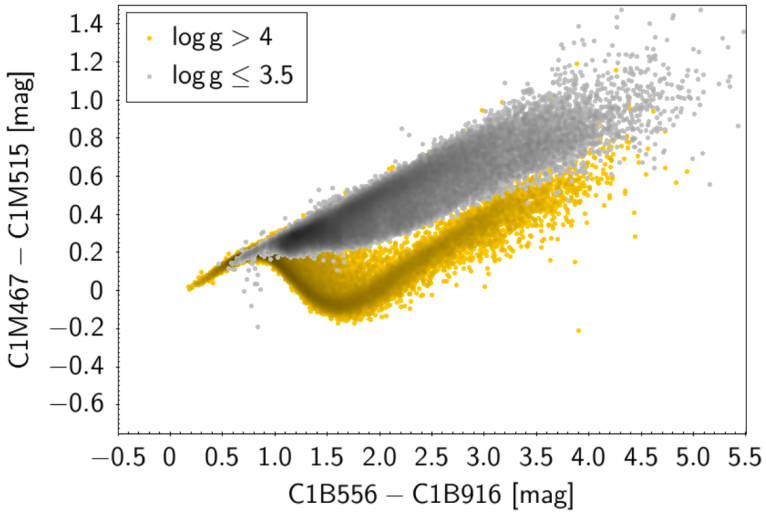

Figure 2: Color-color diagram using synthetic photometry in the C1 photometric system from Jordi et al. (2006) to separate giants (grey) from main sequence stars (orange). The (C1B556 - C1B916) color is gravity sensitive, whereas the (C1M467 - C1M515) color is sensitive to effective temperature. Figure from Gaia Collaboration et al. (2022)

sequence stars (see Figure 2) which have different ranges of surface gravity. As blue horizontal branch stars and blue stragglers have different surface gravity ranges it would be interesting to see whether the synthetic photometry from *Gaia* DR3 can be used to discriminate between the two. The right panel Figure 3 includes the XP spectra of a single BHB and BS star, which are spectroscopically confirmed by Xue et al. (2008). The flux difference between the two stars before and after 400 nm (around the Balmer Jump) gives us a motivation to find out whether this is possible. The bottom panel in Figure 3 shows a SDSS DR17 spectra (Abdurro'uf et al., 2022) of a BHB star to illustrate the difference in resolution when comparing this spectra with low-resolution XP spectra.

1.5 Thesis description

The objective of this thesis is to build a 3D map of the Galactic halo by creating a pure and complete sample of BHB stars using the newly available XP spectra from the third *Gaia* Data Release. In Section 2 we shortly describe the datasets and surveys that are employed in this thesis. In Section 3 we outline our methodology, where we first use information from synthetic stellar spectra (see Section 3.1) to determine which filters can be used to discriminate BHB stars from BS (see Section 3.2). Then, in Section 3.3 and 3.4 we continue to work with two methods, one which employs Pristine CaHK and SDSS photometry based on *Gaia* DDR3 spectrophotometry, the other using SDSS DR18 photometry instead. During the course of this project we also shortly explored the use of synthetic J-PAS 4200 photometry, which is described in Section 3.5. In the analysis (Section 4) we create two samples of candidate BHB stars, one from each method. We use the fact that BHB stars are standard candles to create a 3D maps of the inner halo. These 3D maps are then used to investigate the overlap with stellar streams (see Section 4.3), and separately investigate the Sagittarius stream (see Section 4.4). A discussion and conclusion can be found in Section 5 and 6 respectively.

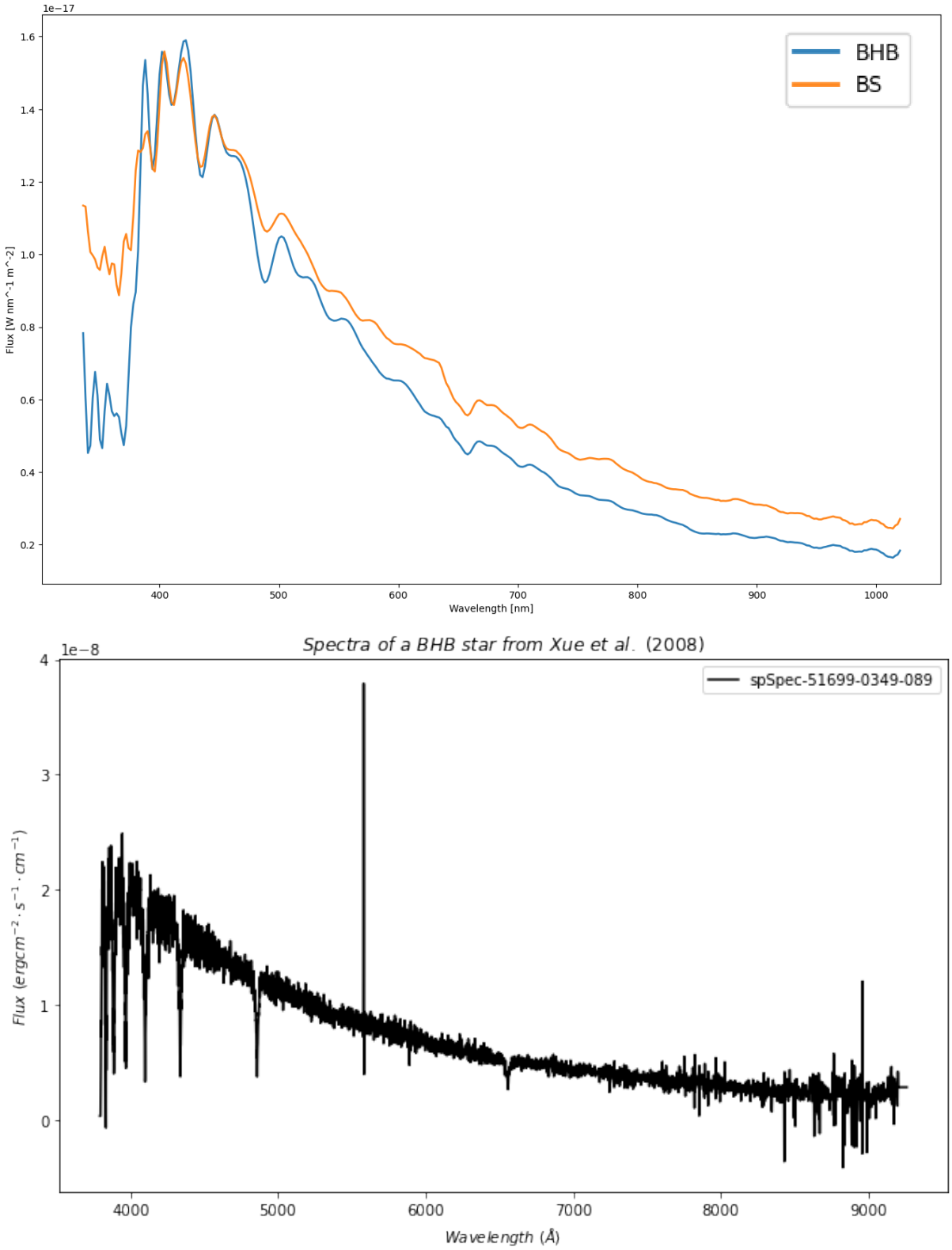

Figure 3: Top panel: XP spectra for a BHB star (in blue) and a BS star (in orange) in the Xue et al. (2008) sample, created with the `gaiaxpy` Python package, having *Gaia* source ids of 743155545527726720 and 922990709589259776 respectively. The XP spectra are measured in the external/absolute system where the spectrum is defined in units of $Wnm^{-1}m^{-2}$ on a scale of absolute wavelengths. Bottom panel: Sloan Digital Sky Survey Data Release 7 spectra of a BHB star from the Xue et al. (2008) sample.

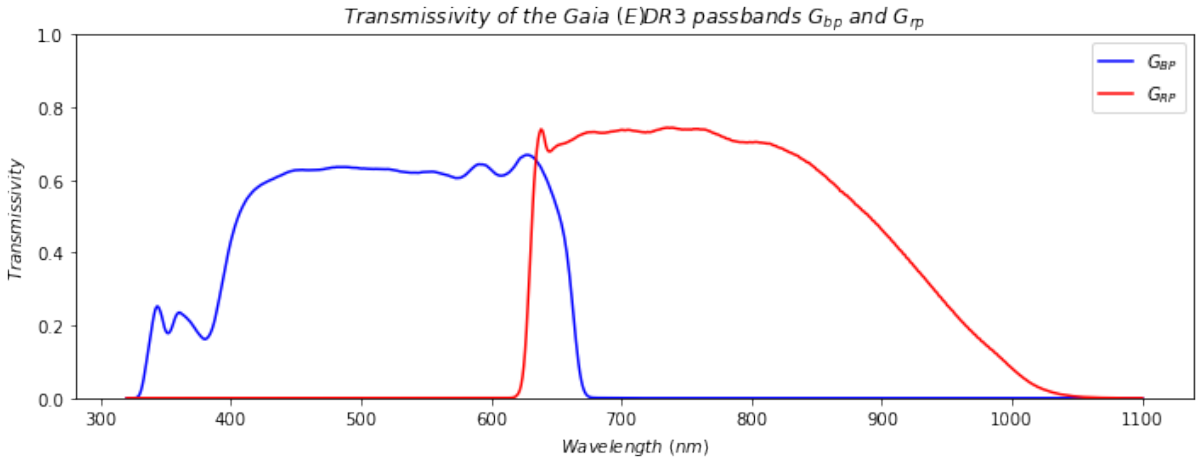

Figure 4: The transmissivity of the *Gaia* (E)DR3 passbands G_{BP} (blue curve) and G_{RP} (red curve). *Gaia* (E)DR3 passbands data provided by ESA/*gaia*/DPAC, P. Montegriffo, F. De Angeli, M. Bellazzini, E. Pancino, C. Cacciari, D. W. Evans, and CU5/PhotPipe team [Published: 29 October 2020]

2 Datasets

2.1 Spectroscopic sample of blue horizontal branch stars and blue stragglers from Xue et al. (2008)

A spectroscopic confirmed dataset of 4895 BHB stars and 3689 BS stars from Xue et al. (2008) will be used to function as a training sample for the methods used. Within this sample 2148 BHB stars and 1151 BS stars have *Gaia* DR3 XP spectra. The Xue et al. (2008) sample is created by using photometry and spectra from SDSS DR6 and by applying color cuts as well as cuts on the Balmer line profiles. Since BS stars have higher surface gravities, the gravity-sensitive Balmer lines are slightly wider at 20 per cent below the local continuum than those of a BHB star. For this work the full list of candidates contained in Table 1 from Xue et al. (2008) is used, which has a contamination of BS stars and MS stars of about 50 percent. The final sample of BHB stars in Xue et al. (2008) is relatively pure, having a contamination well below 10 percent.

2.2 *Gaia* Data Release 3

Gaia is a spacecraft from the European Space Agency (ESA) (Gaia Collaboration et al., 2016), which specializes in measuring all-sky astrometry (parallaxes, proper motion and positions) and photometry with unprecedented precision, reaching submilli-magnitude precision in the range $10.0 \leq G \leq 17.0$ mag. Its objective is to create the largest and most precise 3D map of the Galaxy by measuring nearly two billion sources. It is launched in 2013 and is expected to operate until 2025. In this work we use data from *Gaia* Data Release 3 (DR3) (Gaia Collaboration and Brown, 2021). More specifically we will use the 220 million sources within DR3 for which *Gaia* has measured low-resolution blue and red prism photometer BP/RP mean spectra (Gaia Collaboration et al., 2022) (see Section 1.4 for more information on XP spectra). *Gaia* integrates photometry in three passbands: G, G_{BP}, G_{RP}. The transmission curves of G_{BP}, G_{RP} passbands are shown in Figure 4.

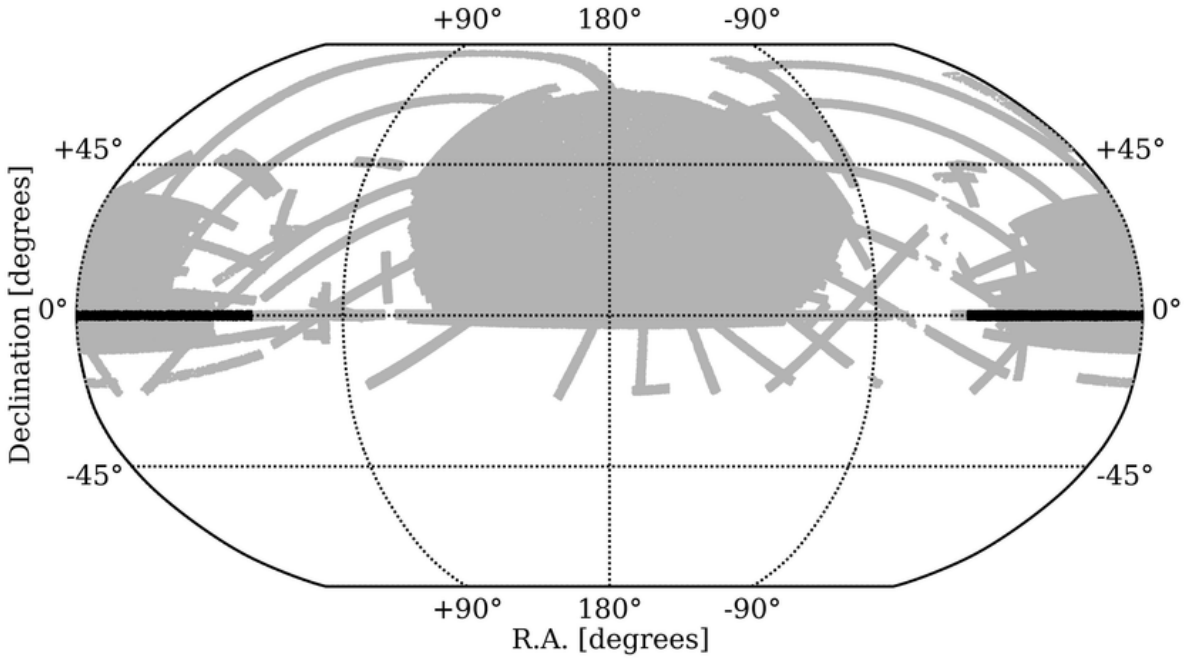

Figure 5: The photometric footprint (in grey, $\sim$ 14.000 deg^2) of the Sloan Digital Sky Survery (SDSS) data release 12 (Alam et al., 2015). Figure from Gentile Fusillo et al. (2015).

2.3 Sloan Digital Sky Survey (SDSS) photometry

The Sloan Digital Sky Survey (SDSS) (Stoughton et al., 2002) is a series of sky surveys, which started in 2000 and focuses on wide-field optical and infrared imaging as well as spectroscopy. A 2.5 m optical telescope is used, which is located in Apache Point Observatory in New Mexico. In this work we use SDSS data release 18 (DR18) photometry (Almeida et al., 2023) in the SDSS *ugr* bands, which transmission curves are also shown in Figure 1. The SDSS footprint can be seen in Figure 5. Note that this is the SDSS DR12 footprint and that we use SDSS DR18 in this work. The characteristics of these footprint however remain similar.

3 Methodology: Selection of blue horizontal branch stars

3.1 Synthetic stellar spectra

The first step towards finding a method to select BHB stars with the use of XP spectra data is to know which wavelength regions can be used to discriminate BHB from BS stars. This knowledge can then help in opting combinations of available synthetic pass-bands, that make the two populations stand out in color. As previously mentioned, we know that the main difference between BHB and BS stars is their surface gravity. To further understand how this underlying physical difference and how other stellar parameters influence the shape of the spectra synthetic stellar, we employ synthetic stellar spectra. Given a set of effective temperature, metallicity, surface gravity and bolometric luminosities ($T_{eff}, [Fe/H], log\ g, L/L_{\odot}$) synthetic spectra can be created. With the implemented Basel Stellar Library (Lejeune et al., 1998) within `pystellibs` [1], models for BHBs and BS are created with effective temperatures for A-type stars, ranging from 7500 to 9000 K (Thomas et al., 2018). Furthermore, the models for BHBs are selected to have $2.8 \leq log\ g\ [g\ cm^{-1}s^2] \leq 3.75$, $-2.5 \leq [Fe/H] \leq -1.5$ (Xue et al., 2008) and $L/L_{sun} \approx 30-100$, where the relevant luminosity range is an estimate for BHBs in the old halo ($>$ 8 Gyr) (Binney and M.Merrifield, 1998). The models for BS are selected to have $3.75 \leq log\ g \leq 5.0$, $1.0 \leq [Fe/H] \leq 0.5$ (Xue et al., 2008), and $L/L_{sun} \approx 4 - 40$ (Raso et al., 2019). For the purpose of this study, only the wavelength range in which *Gaia* has measured XP spectra will be considered ($330\ nm \leq \lambda \leq 1050\ nm$). Additionally, we will not consider the width of the Balmer lines in this work as a means of classification, because with the low-resolution of the XP spectra ($\lambda/\Delta\lambda \approx 25 - 100$) this subtle difference is not easily detectable. Instead, we will look at flux differences that might be detectable using broad, medium, and narrow band photometry.

We confirm from synthetic stellar spectra that only the surface gravity and temperature noticeably affect the spectral shape. The metallicity has a negligible effect on the spectral shape of both stars, which is expected when considering that spectral lines in these hot A-type stars are mostly due to hydrogen LeBlanc (2010). The bolometric luminosity only shifts the spectrum up and down in flux and does not affect the spectral shape. As BHBs and BSs have similar effective temperatures ranges, the temperature-sensitivity of the spectra might not be of use in their classification. In Figure 22 and 23 in the Appendix A one can find examples of synthetic spectra for stars with different temperatures and typical surface gravities for BHB and BS stars. From these spectra it can be inferred that an increase in T_{eff} steepens the spectral shape, hereby creating a larger color difference for stars with a higher difference in T_{eff}. As a consequence, the intersection of both curves after the Balmer Jump shift to redder parts of the spectrum for smaller difference in T_{eff}. In Figure 6 two synthetic spectra are shown with the typical surface gravities for a BHB star ($log\ g = 3$) and a BS star ($log\ g = 5$), both with the same effective temperature (8500 K), solar metallicity and solar luminosity. The fact that blue stragglers are more compact than blue horizontal branch stars results in a color difference which is most visible around the Balmer jump ($\sim$ 3800 Å), which is steeper for the less compact BHBs. This effect of the gravity sensitivity of the Balmer jump is visible to a lesser extent at the Passchen jump, which is better visible in logarithmic scale (see Appendix 3 Figure 24). Even though the color difference around the Balmer jump is relatively larger, the transmissivity in the G_{BP} band drops significantly below 400 nm (as visible from the transmissivity curves from the *Gaia* XP spectra in Figure 4), which means that the uncertainties in the XP spectra around the Balmer jump are expected to be relatively large. In comparison, the color difference around the Passchen jump is

[1] *https://mfouesneau.github.io/pystellibs/*

much smaller, but resides in the wavelength region where the transmissivity of the *Gaia* passbands peaks, meaning that small flux differences between BHBs and BSs might be detectable. From these synthetic spectra it can be concluded that the gravity-sensitivity of the Balmer jump and the Passchen jump could be used to discriminate BHB stars from more compact BS.

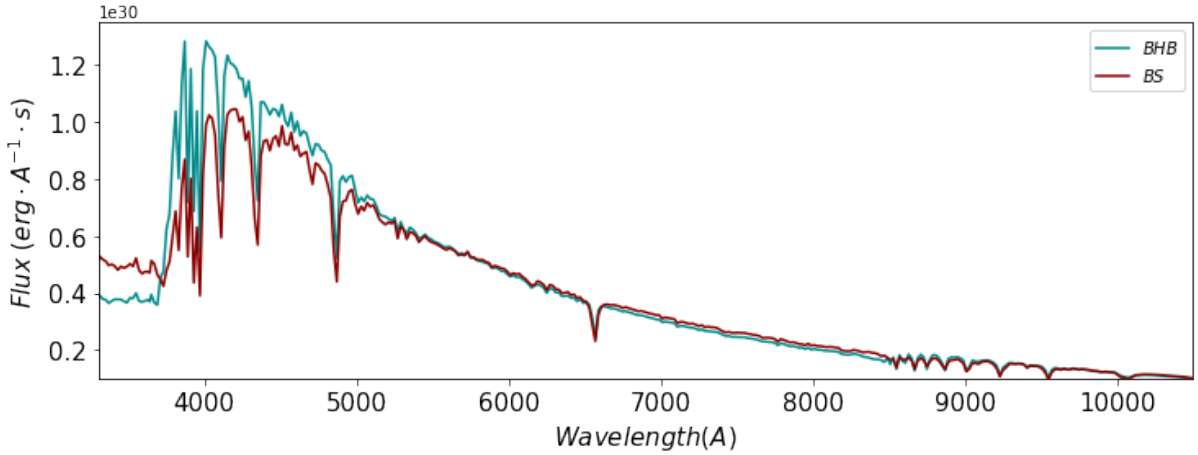

Figure 6: Basel stellar library (Lejeune et al., 1998) synthetic stellar spectra generated with the Python tool `pystellibs`. For stars with $T_{eff} = 8500K$, $[Fe/H] = -0.5$ and solar luminosity. One of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot). The other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot).

3.2 Selecting filters

The `gaiaxpy` Python package[2] is used to get synthetic magnitudes, fluxes and flux uncertainties from XP spectra by applying a synthetic filter. By subtracting the magnitude in two synthetic filters, a color can be obtained. The aim is to create a color space, in which one axis shows a gravity-sensitive color while the other axis shows a temperature-sensitive color, which is a similar approach to the one adopted in Starkenburg et al. (2019). With the information we got from synthetic stellar spectra in the previous Section, we tested several combinations of available synthetic filters on a literature sample of 4895 BHB and 3689 BS stars from Xue et al. (2008), of which respectively 2148 and 1151 have XP spectra. The *Gaia* Synthetic Photometry Catalogue (GSPC) contains standardised photometry for a select few photometric systems, to correct for the systematic errors that affect the XP spectra. The original transmission curves of these systems are slightly altered (see Figure 4 in Gaia Collaboration et al. 2022). In this work, we only use standardised SDSS photometry. The other photometric systems we use, are not standardised. All the filters that are used have a full width at half maximum (FWHM) much bigger than the Doppler shift ($1 - 3$ Å) that could be created when considering an average radial velocity dispersion of $\sim$ 100 km/s of blue stragglers and BHB stars in the halo (Brown et al., 2009). As suggested in the *Gaia* XP spectra release paper (Gaia Collaboration et al., 2022), only stars with a signal-to-noise ratio (SNR) > 30 in the used filters are considered and for filters in the GSPC an additional selection of `Xflag = 1` is used to select on stars where standardisation and validation have been performed withing the ranges of the GBP, GRP colours and G magnitudes. Additionally, we only considered stars with uncertainties $\leq 0.02 - 0.03$ in the colors we selected.

[2]https://gaia-dpci.github.io/GaiaXPy-website/

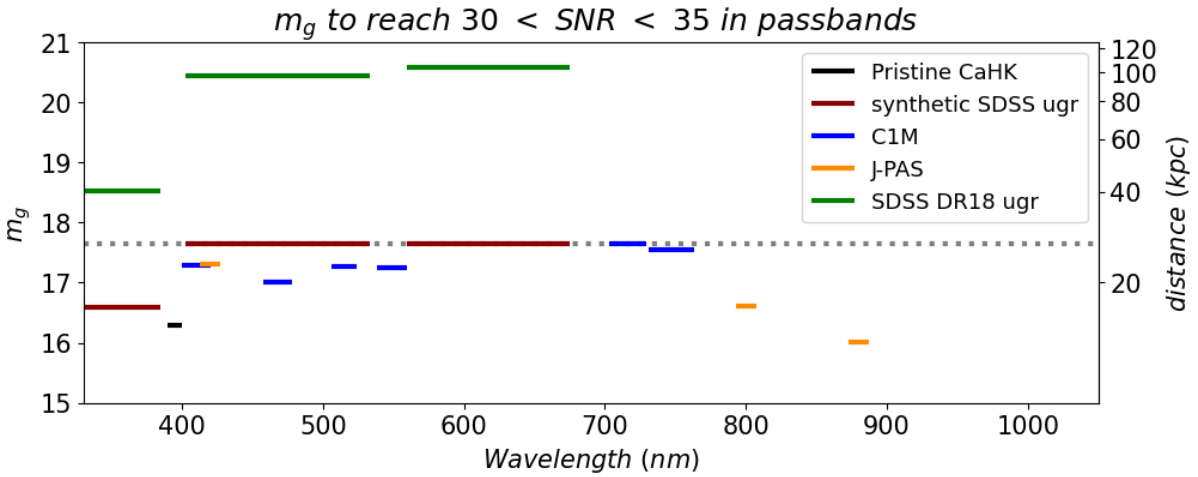

Figure 7: The average apparent magnitude in the SDSS g band (m_g) needed to reach a signal-to-noise ratio between 30 and 35 in any of the tested passbands. The lines represent the wavelength window of each passband and are shown at FWHM from the filter centres. The average m_g and corresponding distances are calculated for BHB stars in the Xue et al. (2008) sample that have synthetic photometry. The m_g is shown for synthetic Pristine CaHK (black line) (Starkenburg et al., 2017), the synthetic C1M410, C1M467, C1M515,C1M549, C1M716 and C1M825 bands (Jordi et al., 2006) (in blue), the J-PAS 4200, J-PAS 8000 and J-PAS 8800 (in orange) (Benitez et al., 2014) and the synthetic SDSS *ugr* (in red) (Doi et al., 2010). The dashed grey line is the limiting magnitude of published sample of 220 million sources with XP spectra Gaia Collaboration et al. (2022). The m_g is also shown for SDSS DR 18 *ugr* bands (in green).

For each narrow -, medium -, or broad-band filter we select BHB stars in the Xue et al. (2008) sample with a SNR between 30 and 35 in that band, as representatives for the faintest A-stars we can use. We then calculate their mean SDSS g mag and convert that into distances using $M_{SDSSg} \approx 0.5$ for a BHB star, to have an indication of how deep one can look into the halo with each filter. The results are shown in Figure 7 and allow use to asses what filter combination allows us to look the furthest into the halo. The published sample of about 220 million sources with XP spectra is limited to sources brighter than $G = 17.65$ (Gaia Collaboration et al., 2022), which sets the upper limit to a distance of $\sim$ 27 kpc when using BHB stars.

Using Figure 7 and the information we got from synthetic stellar spectra, we set up three samples. Firstly, we use synthetic Pristine CaHK combined with synthetic *ugr* SDSS photometry to find BHB stars, hereby building on the success of the method used in the work of Starkenburg et al. (2019). From the synthetic stellar spectra in Figure 1 it can be seen how a BHB star will appear redder in $u - CaHK$ than a BS star, due to the gravity sensitivity of the Balmer jump, and that $g - r$ might be an indicator of the temperature, because the temperature sensitivity of the black body curve increases for redder wavelengths. Due to the transmissivity drop around 400 nm in the G_{BP} band, the uncertainties in the SDSS u band and Pristine CaHK band are larger than those in the g and r band. As a consequence the synthetic SDSS u band and the Pristine CaHK filter are quite shallow: $m_g \approx 16.3$ and 16.6 respectively for $30 < \text{SNR} < 35$.

Therefore, we create a second sample, were we employ *ugr* photometry from SDSS DR18 (Almeida et al., 2023) (not from XP spectra) combined with synthetic CaHK photometry, to

look further into the halo, up till 18.5 kpc with SDSS u and even further with the SDSS g and r band (see Figure 7) . When we employ SDSS DR18 photometry, only BHB stars with a magnitude uncertainty < 0.036 in the SDSS ugr filters are considered to select on good photometry, which corresponds to the same order of magnitude as considering SNR > 30.

Thirdly, we aim to explore the use of the J-PAS 4200 filter in combination with *ugr* SDSS DR18 photometry. J-PAS 4200 is a narrow pass-band which peaks at 420 nm and which wavelength range is slightly redder than CaHK, still picking up the flux difference between BHB and BS stars around the Balmer Jump. Whereas the Pristine CaHK filter coincides with the transmissivity drop of the *Gaia* BP/RP spectra, the transmissivity in the wavelength window of the J-PAS 4200 filter peaks again (as visible from the transmissivity curves of the Gaia passbands in Figure 4). As a consequence, the use of J-PAS 4200 allows us to look further into the halo with BHB stars (about 1 mag fainter) than the CaHK filter (see Figure 7).

For a more extensive overview of the other filter combinations that were tested, one can take a look at the color-color plots of the Xue et al. (2008) sample in Appendix B.

3.3 All-sky sample of blue horizontal branch with Pristine CaHK and SDSS photometry based on *Gaia* DR3 spectrophotometry

In order to select BHB stars from synthetic SDSS and synthetic Pristine CaHK photometry, we first adopt a similar selection to the one used in the method by Deason et al. (2011), by only using stars with $-0.25 < g_0 - r_0 < 0.0$, corresponding to the $g_0 - r_0$ colors of A-type stars. As the uncertainties in the u-band are relatively high, no cut on the $u_0 - g_0$ color is performed. Here the u_0, g_0, r_0 and $CaHK_0$ magnitudes refer to the extinction-corrected magnitudes. To correct for interstellar reddening a first-order extinction correction to the synthetic magnitudes was performed by using the dust map from Schlegel et al. (1998). In Appendix C one can find a more detailed description of how these extinction corrected magnitudes were obtained. Here we assume no uncertainties in the extinction corrected magnitudes. The synthetic CaHK magnitudes are provided in the Vega magnitude system and the synthetic SDSS magnitudes in the AB magnitude system. To convert from Vega to AB, we subtract 0.066 mag from the CaHK magnitudes, as suggested by the Pristine Survey wiki. No selection on variable stars is performed (see Section 5 for further explanation). Additional to selecting on stars with SNR > 30 in the filters and `u,g,r flag = 1`, an uncertainty cut of $g_0 - r_0$ and $u_0 - CaHK_0 < 0.03$ is performed to avoid clutter from low SNR observations. Based on this dataset a slightly stricter color cut is performed: $-0.225 < g_0 - r_0 < -0.05$, $0.6 < u_0 - CaHK_0 < 1.4$.

These selections leave us with 60052 sources, which distribution in $g_0 - r_0$ and $u_0 - CaHK_0$ color space can be seen in the top right panel of Figure 8. We calculate the color uncertainties through Gaussian error propagation of the flux uncertainties. The mean uncertainties in $g_0 - r_0$ are 0.0035 mag, which is considerably smaller than the mean uncertainties in $u_0 - CaHK_0$ which is 0.024 mag (as visible from the right panel in Figure 9). From these distribution in colorspace two clear bands can be identified, although they seem to touch for lower $g_0 - r_0$ and partly overlap. When applying the same selections on the Xue et al. (2008) sample and overplotting it (see top left panel of Figure 8), it becomes clear that the rightmost sequence corresponds to BHB stars and that the leftmost distribution belongs to BS. There are however some BHB stars from the Xue et al. (2008) sample that overlap with the BS distribution, especially at higher $g_0 - r_0$.

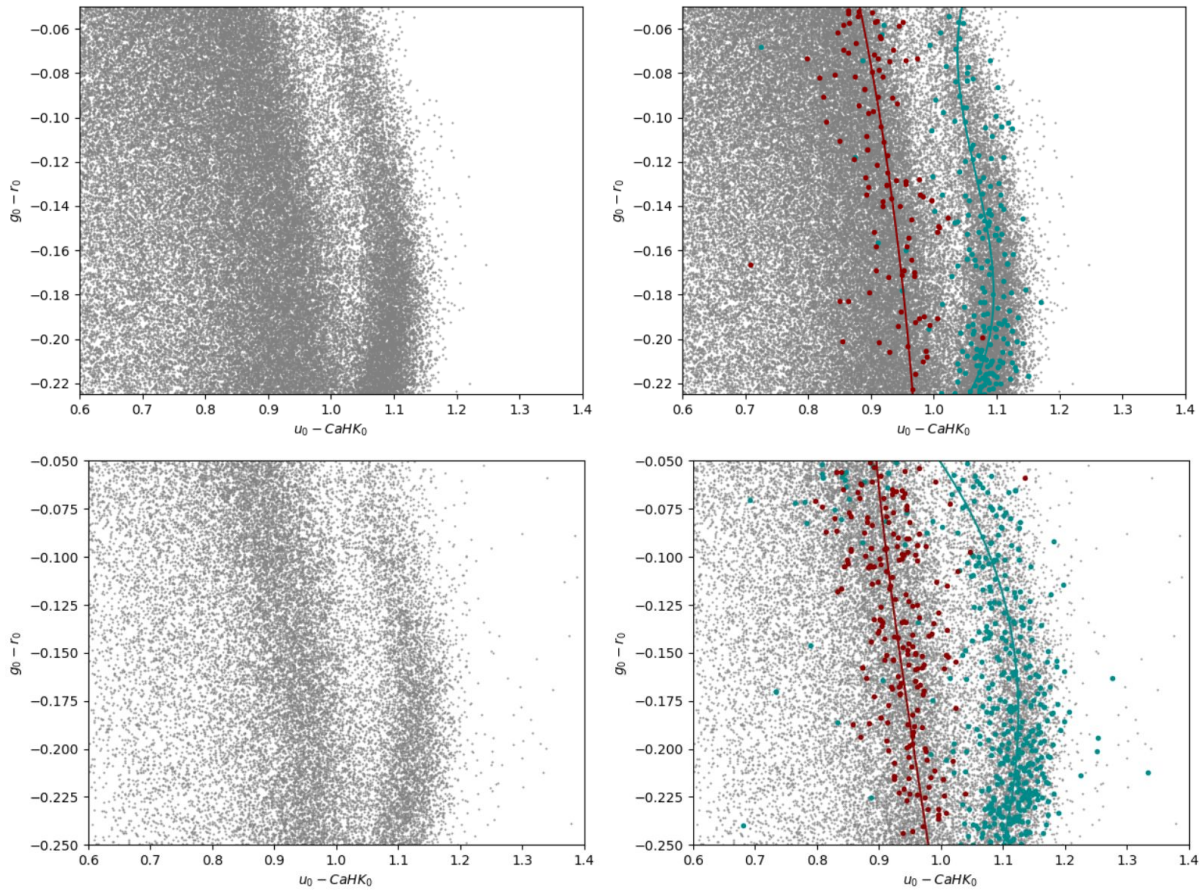

Figure 8: Top right panel: Color-color space using synthetic photometry from *Gaia* XP spectra in the SDSS *ugr* bands and Pristine CaHK, for a selected sample of A-type colored stars with high quality photometry (see Section 3.3 for selection criteria). The bottom right panel is created using a similar procedure, but instead employs *ugr* photometry from SDSS DR18 (see section 3.4 for the used selection criteria). Left panels: Overplotted are the ridgelines for blue horizontal branch stars and blue stragglers (blue and red curve respectively) defined in equation 1 and 2 for the top left panel and in equation 3 and 4 for the bottom left panel. Blue horizontal branch stars and blue stragglers (the blue and red dots respectively) stars from the Xue et al. (2008) sample that survive the same selection criteria are overplotted as well.

There are a lot of sources on the left-hand side of the BS distribution, which do not seem to belong to either distribution. These stars might be young, hot main sequence stars from the disk as these stars contaminate the color space of BHB stars.

Deason et al. (2011) have defined third-degree polynomial ridgelines to separate BHB and BS stars based on their classifications from stellar spectra. Even though these ridgelines do not neatly fit our distributions based on synthetic SDSS and CaHK, we do use a similar approach in selecting BHBs, by fitting a third degree polynomial to the BS and BHB stars from the clipped Xue et al. (2008) sample. Fitting using iterative sigma clipping ($< 3\sigma$) was used to exclude outliers from the Xue et al. (2008), which most visually includes the cloud of BHB stars that overlap with the BS distribution for higher $g_0 - r_0$. The then obtained ridgelines are described by:

$$(u_0 - CaHK_0)_{BHB} = 1.14 + 3.47 \cdot (g_0 - r_0) + 34.08 \cdot (g_0 - r_0)^2 + 90.68 \cdot (g_0 - r_0)^3 \quad (1)$$

$$(u_0 - CaHK_0)_{BS} = 0.84 - 0.88 \cdot (g_0 - r_0) - 1.71 \cdot (g_0 - r_0)^2 - 1.07 \cdot (g_0 - r_0)^3 \quad (2)$$

These ridgelines are shown in the top right panel of Figure 8. In Section 4.1 these ridgelines are used to calculate the probability that a star in the sample is either a BHB star or a BS star.

3.4 Deeper into the halo with SDSS DR18 photometry

The relatively high uncertainties in the SDSS u magnitudes from XP spectra (~ 0.018 mag) limit us to look further into the halo. Therefore we employ SDSS DR18 *ugr* photometry (Almeida et al., 2023), where the mean error in the u band is much smaller (~ 0.006 mag). This results in uncertainties in $u_0 - CaHK_0$ which are generally lower (see Figure 9), allowing us to look at fainter magnitudes where the SNRs are usually lower. We create another sample of BHB stars based on u_0, g_0, r_0 photometry from SDSS DR18 combined with synthetic CaHK, with the aim to look further into the halo. Here we follow a similar selection procedure as outlined in the previous section.

The extinction corrected magnitudes and their uncertainties are given in the SDSS DR18. From the SDSS DR18 only sources classified as stars are considered. Then we apply the same color-cut as described in Deason et al. (2011) to select on stars with A-type colors: $-0.25 < g_0 - r_0 < 0.0$ and $0.9 < u_0 - g_0 < 1.4$. This color cut should also ensure that there is little contamination from QSOs and white dwarfs, as demonstrated in Deason et al. (2011). Additionally, stars are removed that do not have the SDSS clean photometry flag. Stars that are flagged to have saturated photometry in the SDSS *ugr* bands, de-blending or interpolation problems, or are close to the edge of the frame are removed from the sample as well. To remove contamination from variable stars, stars with a `HERN_VARIABLE` flag in PANSTARRS-1 DR2 are deselected, removing only an additional 0.8% percent of the sample (Chambers et al., 2019). Lastly, only stars that have SNR > 30 in the synthetic CaHK and u_0, g_0, r_0 uncertainties < 0.036 are used. Only 15 % of this sample of SDSS based on A-type coloured stars have XP spectra and can be used to obtain synthetic Pristine CaHK. This is due to the relatively small SDSS DR18 footprint, which only covers the Northern Hemisphere, whereas *Gaia* DR3 provides all-sky data. The same cuts on the $g_0 - r_0$ and $u_0 - CaHK_0$ colors and color uncertainties were applied as in Section 3.3, with the exception of using a slightly more lenient cut in the $u_0 - CaHK_0$ by selecting on $-0.25 < u_0 - CaHK_0 < -0.05$. The same selections are applied to the Xue et al. (2008) sample, leaving 213 BS and 428 BHB stars. After applying the selection criteria described above, we are

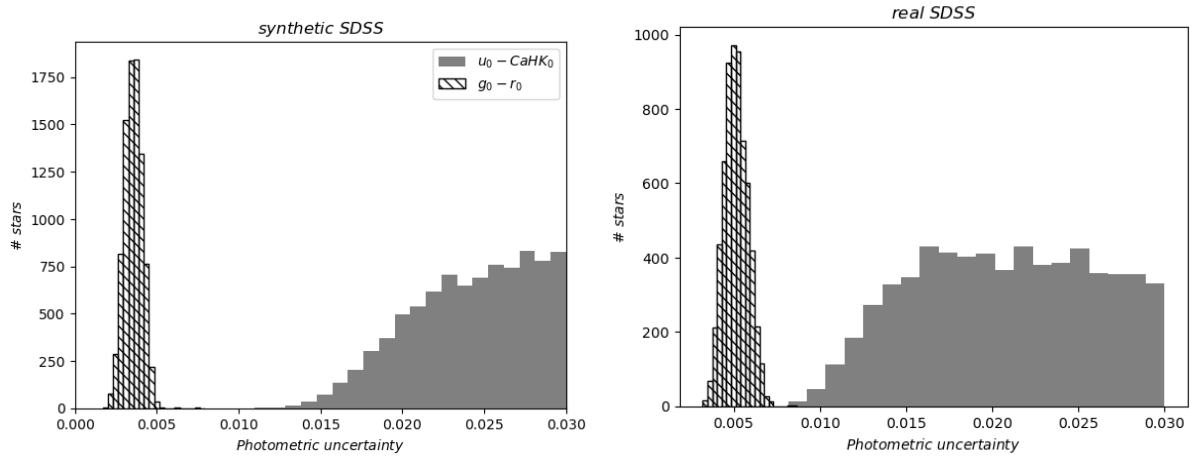

Figure 9: Photometric uncertainties in the $g_0 - r_0$ and $u_0 - CaHK_0$ colors for the sample of A-type colored stars using synthetic SDSS photometry from XP spectra (right panel) and SDSS DR18 photometry (left panel).

left with a sample of 22739 stars. These stars are plotted in $g_0 - r_0$ and $u_0 - CaHK_0$ color-space, visible in the bottom right panel of Figure 8. From Figure 8 two bands can be clearly identified and can be confirmed to belong to the BHB and BS populations by overplotting the clipped Xue et al. (2008) sample. Apart from the number of sources, this color-color plot looks visually similar to the one employing SDSS photometry from XP spectra. Fitting the ridgelines along the same procedure as described in section 3.3 now gives:

$$(u_0 - CaHK_0)_{BHB} = 0.86 - 3.22 \cdot (g_0 - r_0) - 11.85 \cdot (g_0 - r_0)^2 - 10.92 \cdot (g_0 - r_0)^3 \quad (3)$$

$$(u_0 - CaHK_0)_{BS} = 0.88 - 0.17 \cdot (g_0 - r_0) \quad 1.36 \cdot (g_0 - r_0)^2 2.08 \cdot (g_0 - r_0)^3 \quad (4)$$

These ridgelines are visible in the bottom right panel of Figure 8. In Section 4.1 these ridgelines are used to calculate the probability that a star in this sample is either a BHB star or a BS star.

3.5 Using J-PAS 4200 to solve the disadvantage of relatively large uncertainties in Pristine CaHK

As pointed out in Section 3.2, we aim to use synthetic photometry from the J-PAS 4200 filter in combination with *ugr* SDSS DR18 photometry to solve the disadvantage of the relatively large uncertainties and shallowness of synthetic Pristine CaHK photometry from XP spectra. Due to time constraints, we solely explored the use of the J-PAS 4200 filter in combination with *ugr* photometry on the Xue et al. (2008) sample, which seems to some degree successfully discriminate BHB and BS stars in color space (see Figure 10). Note that we did not use the extinction corrected magnitudes here. A significant number of BHB stars (1221) from the Xue et al. (2008) have SNR $>$ 30 in J-PAS 4200, *ugr* SDSS mag uncertainties $<$ 0.036 and (J-PAS 4200 - SDSS u) uncertainties $<$ 0.2. The overlap between the two sequences in colorspace is bigger than when the Pristine CaHK filter is used instead of J-PAS 4200. We have already obtained extinction corrected magnitudes for the J-PAS filters for A-type candidate stars and by following a similar procedure as outlined in the previous two sections, a new sample of BHB candidates based on J-PAS 4200 photometry could be created.

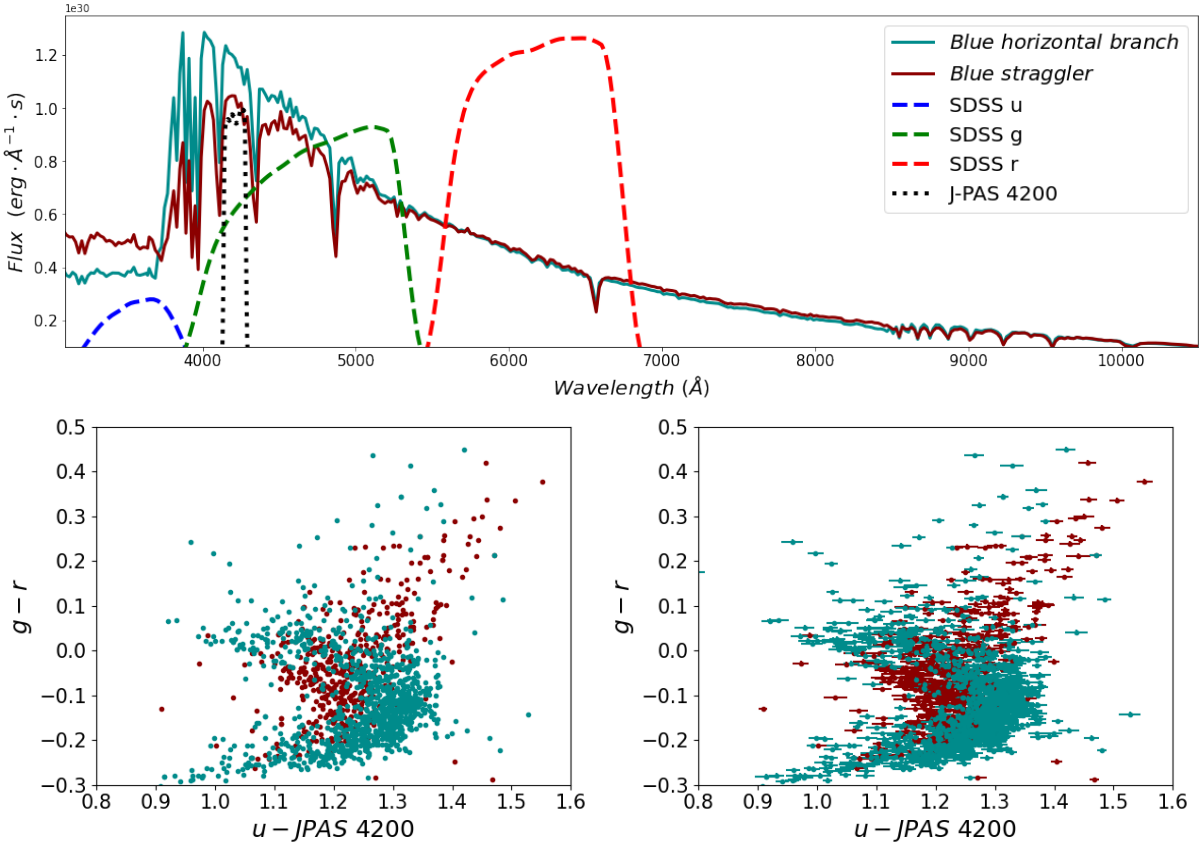

Figure 10: Top panel: Basel stellar library Lejeune et al. (1998) synthetic stellar spectra generated with the Python tool `pystellibs`. One of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot). The other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). The transmission curves of the SDSS *ugr* bands (Doi et al., 2010) and the J-PAS 4200 (Benitez et al., 2014) are shown and scaled to be visible. Please note that we show here the non-standardised SDSS transmission curves, while in fact we use the slightly altered standardised versions. Bottom panels: Colorspace of blue horizontal branch stars (in blue) and blue stragglers (in red) from the Xue et al. (2008) sample that survive the selection criteria described in Section 3.5. The uncertainties in the colours are shown as error bars in the left panel.

3.6 Removing contamination from the disk

As previously mentioned, young and hot MS stars are known to contaminate the colorspace of BHB stars. Although they are very rare in the halo, there presence in the Galaxy's disk is significant. The first-order extinction corrections we perform are also less trustworthy for areas in the disk where gas and dust is much more prominent. For these reasons, and since our interest lies only within the halo, a galactic latitude (b) cut ($-25° < b < 25°$) is performed to get rid of stars that might belong in the disk. This significantly reduces the samples of A-type colored stars, leaving 8766 stars in the synthetic SDSS sample and 6348 in the SDSS DR18 sample. This significant reduction is most probably due to fact that the stellar density in the disk is much higher compared to that of the halo. If we now look at the color-color plots again after this galactic latitude cut is performed (see Figure 11), the sequences stand out even more clearly. The cloud of stars on the right-hand-side of the BS sequence is not visible anymore in both the synthetic

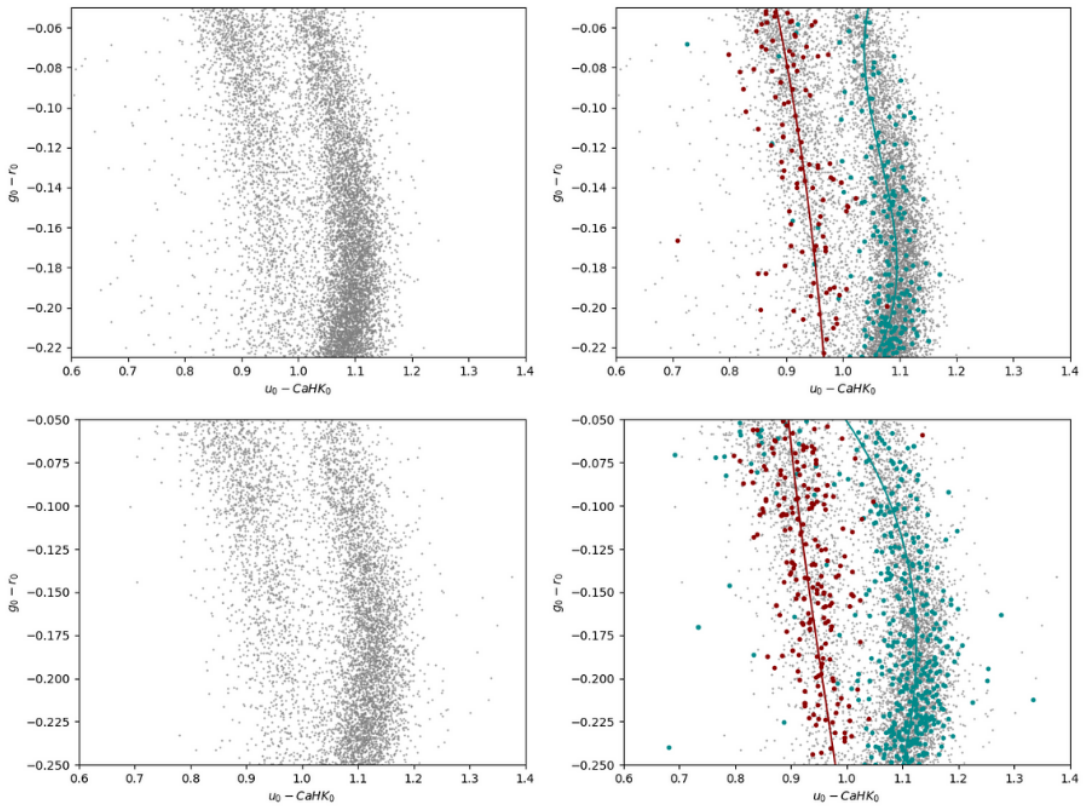

Figure 11: In addition to Figure 8 a galactic latitude cut is performed (see section 3.6 for details). Top right panel: Color-color space using synthetic photometry from *Gaia* XP spectra in the SDSS u,g,r bands and Pristine CaHK, for a selected sample of A-type colored stars with high quality photometry (see section 3.3 for selection criteria). The bottom right panel is created using a similar procedure, but instead employs real u,g,r photometry from SDSS DR18 (see section 3.4 for the used selection criteria). Left panels: Overplotted are the ridgelines for BHB and BS stars (blue and red curve respectively) defined in equation 1 and 2 for the top left panel and in equation 3 and 4 for the bottom left panel. BHB and BS (the blue and red dots respectively) stars from the Xue et al. (2008) sample that survive the same selection criteria are overplotted as well.

SDSS plot and the SDSS DR18 plot, suggesting that these stars with lower $u_0 - CaHK_0$ might be young and hot main sequence stars. This cloud feature is not visible in the color-color plots in the work of Starkenburg et al. (2019) (see the bottom panels of Figure 2 in their work). As the Pristine footprint at that time only covered areas with high galactic latitude, this confirms our suspicion that the stars in this cloud are disk stars.

4 Results: Investigating the inner halo with candidate blue horizontal branch stars

4.1 Two pure and complete samples of blue horizontal branch stars

In this section we use the synthetic SDSS sample and the SDSS DR18 sample (described in Section3.3 and 3.4 respectively) of A-type colored stars, for which the galactic latitude cut has been performed (see Section 3.6), to create two clean samples of BHB candidates. The performance of these samples are compared by determining the purity and completeness of their BHB selection.

To calculate the probability that a source in our remaining synthetic SDSS and SDSS DR18 sample is either a BHB star or BS, the method from Deason et al. (2011) , which is also employed in the work of Starkenburg et al. (2019), is closely followed. We assume here the same measured standard deviations for the BHB and BS populations as in the work of Deason et al. (2011): ($\sigma_{BHB,0}(u_0 - CaHK_0) = 0.04$ and $\sigma_{BS,0}(u_0 - CaHK_0) = 0.045$). Although, in their work they consider $u_0 - g_0$ instead. We will calculate the likelihood based on the defined ridgelines in Section 3.3 and 3.4:

$$p(ugrCaHK|BHB) \propto exp\left(-\frac{[(u_0 - CaHK_0) - (u_0 - CaHK_0)_{BHB}]^2}{2\sigma^2_{BHB}}\right), \tag{5}$$

$$p(ugrCaHK|BS) \propto exp\left(-\frac{[(u_0 - CaHK_0) - (u_0 - CaHK_0)_{BS}]^2}{2\sigma^2_{BS}}\right).$$

Where σ_{BHB} and σ_{BHB} include both the intrinsic width of the populations, as well as the uncertainty on our measured colors $\sigma_{BHB,0}$ and $\sigma_{BS,0}$:

$$\sigma_{BHB} = \sqrt{\sigma^2_{BHB,0} + \sigma^2_{BHB,0(u_0-CaHK_0)}} \quad , \quad \sigma_{BS} = \sqrt{\sigma^2_{BS,0} + \sigma^2_{BS,0(u_0-CaHK_0)}} \tag{6}$$

Then the posterior probability of a star being a BHB or BS star based on their position in colorspace is given by:

$$P(BHB|ugrCaHK) = \frac{p(ugrCaHK|BHB)N_{BHB}}{p(ugrCaHK|BHB)N_{BHB} + p(ugrCaHK|BS)N_{BS}} \tag{7}$$

$$P(BS|ugrCaHK) = \frac{p(ugrCaHK|BS)N_{BS}}{p(ugrCaHK|BHB)N_{BHB} + p(ugrCaHK|BS)N_{BS}}$$

Where we set N_{BHB} and N_{BS} equal to 1. We remove stars in both samples that are further than 3σ from the defined ridgelines, to try to remove any contamination.

Calculating the probabilities of a star in the Xue et al. (2008) sample being either a BHB or BS star using equations 7, with ridgelines defined in equations 1 and 2 for the synthetic SDSS sample and equations 3 and 4 for the SDSS DR18 sample, results in two clearly divided probability distributions, shown in Figure 12. Notable is that peak of the P_{BHB} distribution is slightly more narrow, when employing the SDSS DR18 sample and that this sample selects relatively more BHB stars. We use these definitions to calculate the purity and completeness:

$$purity = \frac{N_{PBHB,bhb>0.9}}{N_{PBHB,bhb>0.9} + N_{PBHB,bs>0.9}}$$

$$completeness = \frac{N_{PBHB,bhb>0.9}}{N_{BHB,Xue\ et\ al.\ (2008)}}$$

Success rates of our BHB selection		
BHB candidate sample	purity	completeness
synthetic SDSS *ugr*, synthetic Pristine CaHK	99 %	75 %
SDSS DR18 *ugr*, synthetic Pristine CaHK	99 %	81 %

Table 1: Purity and completeness of our blue horizontal branch star selection.

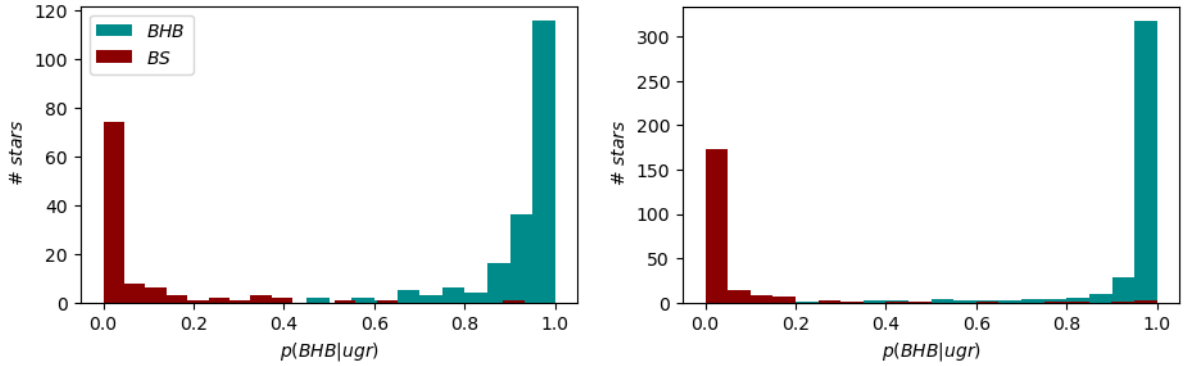

Figure 12: The probability that a star from the Xue et al. (2008) sample is a BHB star, according to the methods described in section 3.3 and 3.4. The right panel includes the probabilities for the synthetic SDSS sample and the left panel shows them for the sample created using SDSS DR18 data.

where $N_{PBHB,bhb>0.9}$ is the number of BHB stars in the Xue et al. (2008) sample that have been correctly identified, $N_{PBHB,bs>0.9}$ is the number of BS stars that have been misclassified as a BHB star, both with a probability higher than 90% of being a BHB star, and $N_{BHB,Xue\ et\ al.\ (2008)}$ is the total number of BHB stars in the Xue et al. (2008) sample. Considering stars from the Xue et al. (2008) sample with $P_{BHB} > 0.9$ results in the success rates given in Table 1. We calculate a purity of 99% and a completeness of 75% of BHB selection when employing synthetic SDSS photometry and a purity of 99% and an ever higher completeness of 81% with SDSS DR18 photometry. This implies that both methods produce samples with a contamination of only 1%.

Selecting on stars with $P_{BHB} > 0.9$ results in a sample of 5315 blue horizontal branch star candidates in the synthetic SDSS sample and 4071 blue horizontal branch star candidates in the SDSS DR18 sample. The two samples have 1432 blue horizontal branch star candidates in common (these stars have the same Gaia source ID). Please note that we use the term blue horizontal branch star candidates here to indicate that spectroscopic observations will be needed to fully confirm that these are true blue horizontal branch stars.

4.2 Towards a 3D map of the inner halo

As mentioned in the Introduction, BHB stars are standard candles, meaning that their consistent absolute magnitudes can be used to compute their distances. Here we employ the relation given in the work of Deason et al. (2011), which is also used in the methods of Starkenburg et al. (2017):

$$M_{g,BHB} = 0.434 - 0.169(g_0 - r_0) + 2.319(g_0 - r_0)^2 + 20.449(g_0 - r_0)^3 + 94.517(g_0 - r_0)^4 \quad (8)$$

Where $M_{g,BHB}$ is the absolute magnitude of the BHB in the SDSS g band. As the SDSS photometry in the GSPC is standardised, this relation should hold for synthetic SDSS photometry as well. The relation $d = 10^{\frac{m_{g_0} - M_{g,BHB} + 5}{5}}$ is used to calculate the distances. Using a Monte Carlo simulation statistical uncertainties in the distance are calculated to be ~ 0.02 kpc for the SDSS DR18 sample and ~ 0.01 kpc for the synthetic SDSS photometry sample on average. No systematic uncertainties are considered, although the relation between $M_{g,BHB}$ and $(g_0 - r_0)$ we use (see equation 8) does include such uncertainties. With this information a 3D map of the inner halo of the Milky Way is created using our samples of BHB candidates.

Figure 13 shows the right ascension versus distance, and declination versus distance maps for the two samples. Both samples cover distances which are still within the inner halo ($R_{GC} < 20$ kpc). The BHB candidate sample employing SDSS DR18 clearly looks further into the halo, with distances reaching $\sim$ 17 kpc, whereas the BHB sample employing SDSS photometry from XP spectra covers distances in between $\sim$ 5 kpc and ~ 12.5 kpc.

High concentrations of stars at a specific RA, visible as stripe features, can be seen in the left panels of Figure 13. Matching the `astrometric_matched_transits` parameter, which is the number of field–of–view transits made by *Gaia* for a certain source, with our synthetic SDSS sample, shows that these stripes appear at similar RA where the `astrometric_matched_transits` number is high. These stripes are thus likely a result of the scanning pattern of *Gaia* (see left panel in Figure 14). Globular clusters might form another cause of these stripe features. Photometry from globular clusters are less reliable due to high crowding in these areas. Therefore we removed the vicinity of globular clusters from our sample. We use the Harris catalogue (Harris, 1996) to get the coordinates of 127 globular clusters that are found within 20 kpc from the Sun, most of which are located in the disk (see Figure 15), and remove BHB candidates that are cross-matched within a radius of 1° from them. This removes 103 BHB candidates from the synthetic SDSS sample and 41 BHB candidates from the SDSS DR18 sample, which compared to the samples sizes is not a substantial amount.

The right and left panels in Figure 15 show the Galactic coordinates and distances of BHB candidates in the two samples, hereby proving a 3D view of BHB candidates in the inner halo of the Milky Way. Several features can be traced back to the *Gaia* scanning pattern (see Figure 14) and the SDSS footprint (see Figure 5). Holes in the sky-coverage of our samples coincide with the location of similar features in the *Gaia* scanning pattern with low number of observations. The fact that the 3D map of the SDSS DR18 sample only covers the Northern Hemisphere and shows some stripe features close to the disk, is the result of the SDSS DR18 footprint.

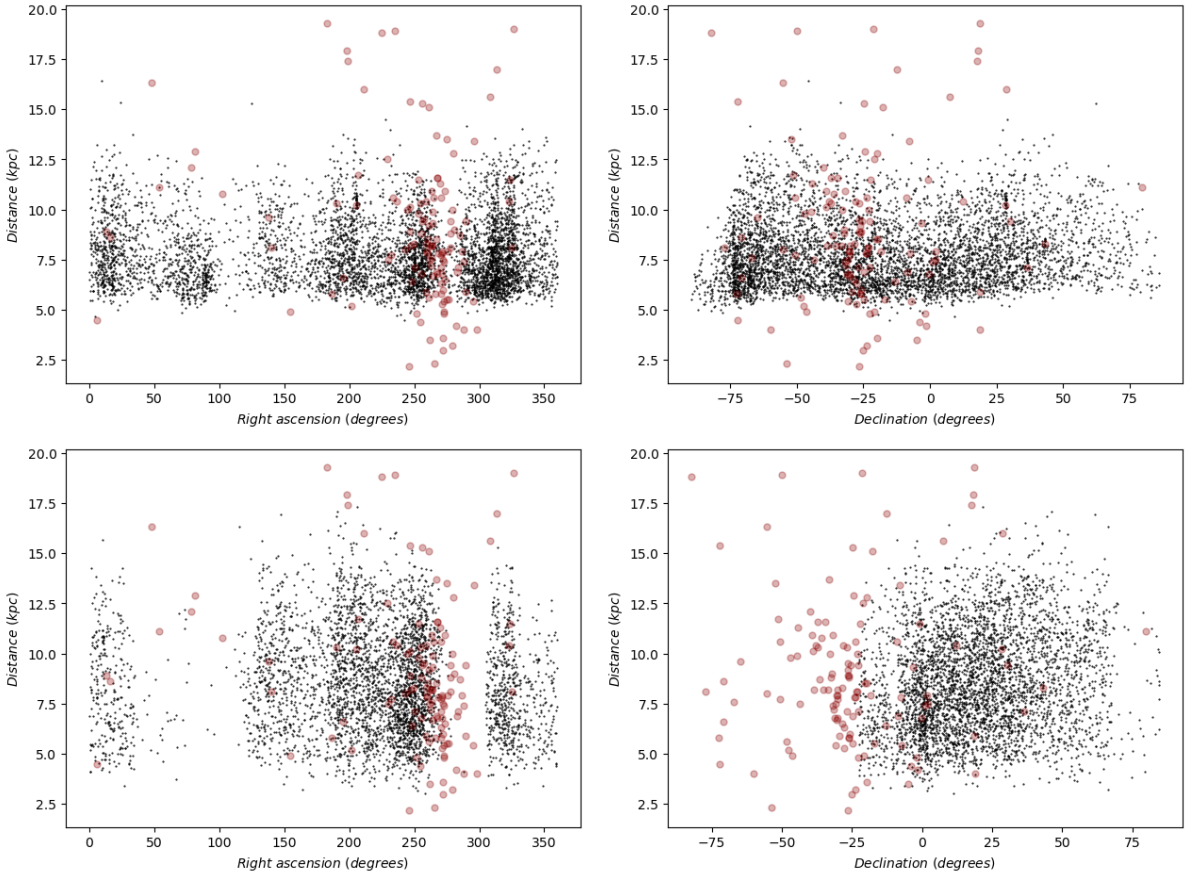

Figure 13: Right ascension versus distance maps (right panels) and declination versus distance maps (left panels) for the blue horizontal branch star samples (shown in black) utilizing synthetic SDSS + synthetic CaHK photometry (upper panels) and SDSS DR18 + synthetic CaHK photometry (bottom pannels). Additionally, globular clusters from the the Harris catalogue (Harris, 1996) within 20 kpc from the Sun are visible as red circles.

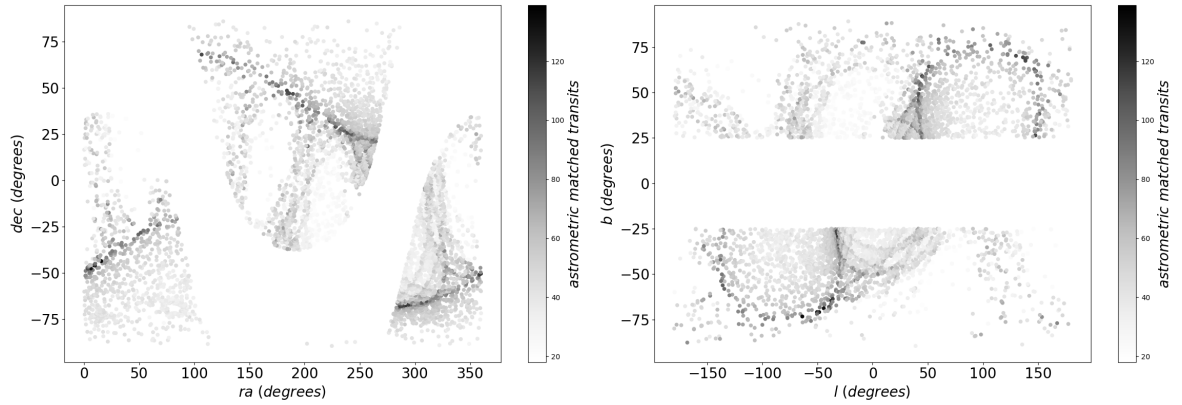

Figure 14: The number of field–of–view transits made by *Gaia* (`astrometric_matched_transits`), matched for the sample of blue horizontal branch stars candidates, that is found using synthetic SDSS and synthetic CaHK photometry. Features from the *Gaia* scanning pattern are made visible when these stars are plotted in ICRS (left panel) and galactic (right panel) coordinates.

Additionally, one might notice a "stream"-like feature of BHB candidates at further distances, going "into" the disk and "coming" out of the disk. Again, this is an effect of the *Gaia* scanning pattern. This "stream"-feature exactly traces the shape in the right panel of figure 14 for stars with a high `astrometric_matched_transits` value. As the *Gaia* telescope observes the locations of these BHB candidates more often, photometry of higher quality of these stars can be obtained, making these more distant stars a part of our samples.

Figure 16 shows how our two samples of BHB candidates are distributed over the *Gaia* DR2 sky map (notice that the emptiness in the disk is due to our cut in Galactic latitude). As previously described, some part of the sky distribution of our BHB candidates seems to be a consequence of the *Gaia* scanning pattern and the SDSS footprint. A high concentration of BHB candidates in the synthetic SDSS sample around near $l, b = (60°, -40°)$ and $l, b = (80°, -30°)$ can be confirmed to coincide with the sky coordinates of the Small -and Large Magellanic Clouds (SMC and LMC) respectively when zooming in (see Figure 17). As the distances to our BHB candidates are far smaller than the distances to LMC and the SMC (about 50 kpc and 60 kpc respectively Subramanian, S. and Subramaniam, A. 2009), these stars are likely misclassified. This might be a consequence of unreliable photometry due to high crowding in these areas or because these are other blue and bright sources within these satellite galaxies such as blue supergiants. The BHB star candidates around the on-sky location of the LMC and SMC are not removed from the sample due to time constraints. These stars are adding some contamination which is not accounted for in the calculation of the purity.

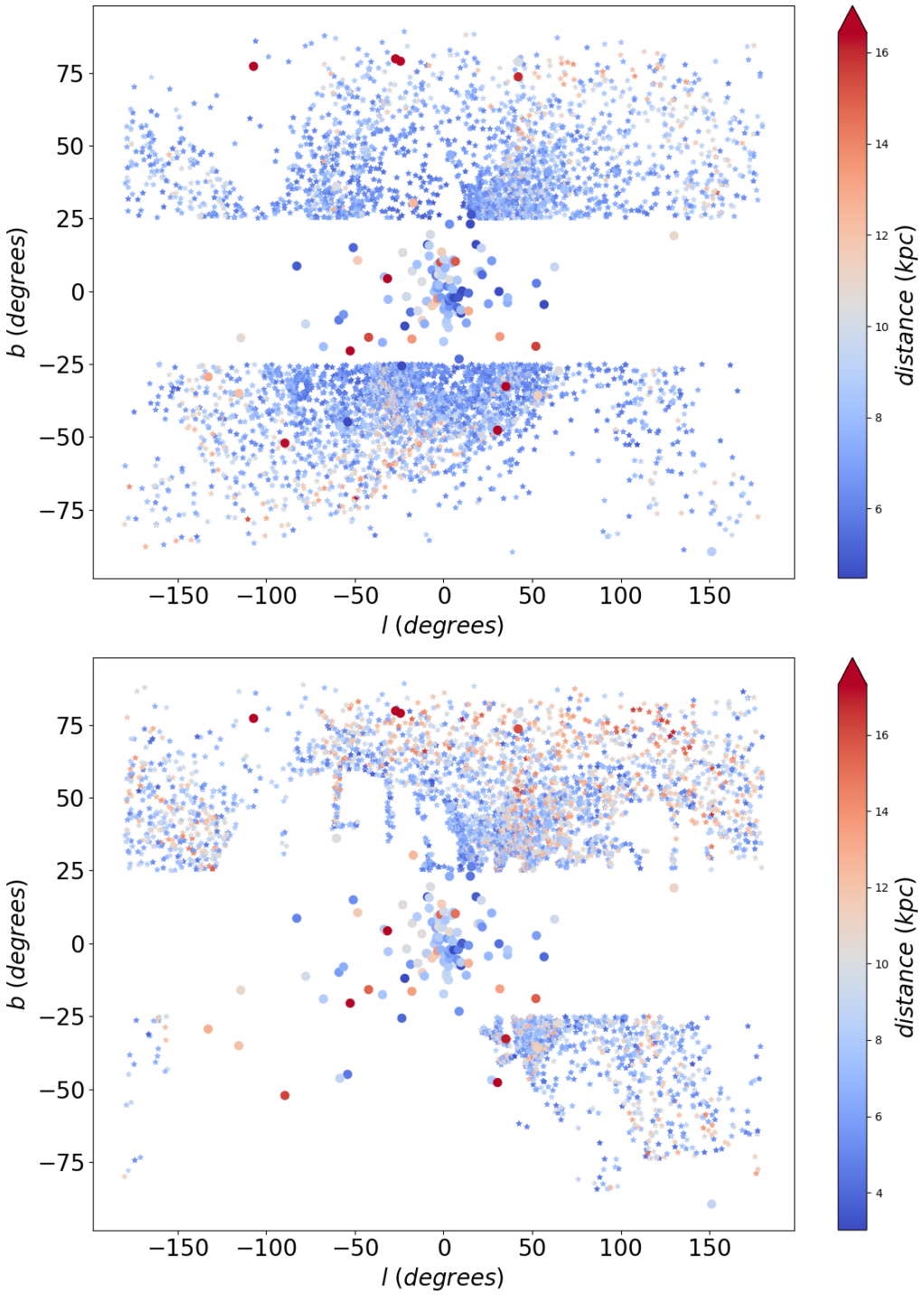

Figure 15: 3D maps of the samples of blue horizontal branch candidates (viewed as star-shaped dots) from synthetic SDSS combined with synthetic CaHK photometry (upper panel) and SDSS DR18 combined with synthetic CaHK photometry (bottom panel). Distances are calculated using relation 8 from Deason et al. (2011). Additionally, globular clusters from the Harris catalogue (Harris, 1996) within 20 kpc from the Sun are visible as large dots. Stars that are within 1° from these globular clusters are removed from the sample. Note that the distance bars are shown separately for the distance range of each sample.

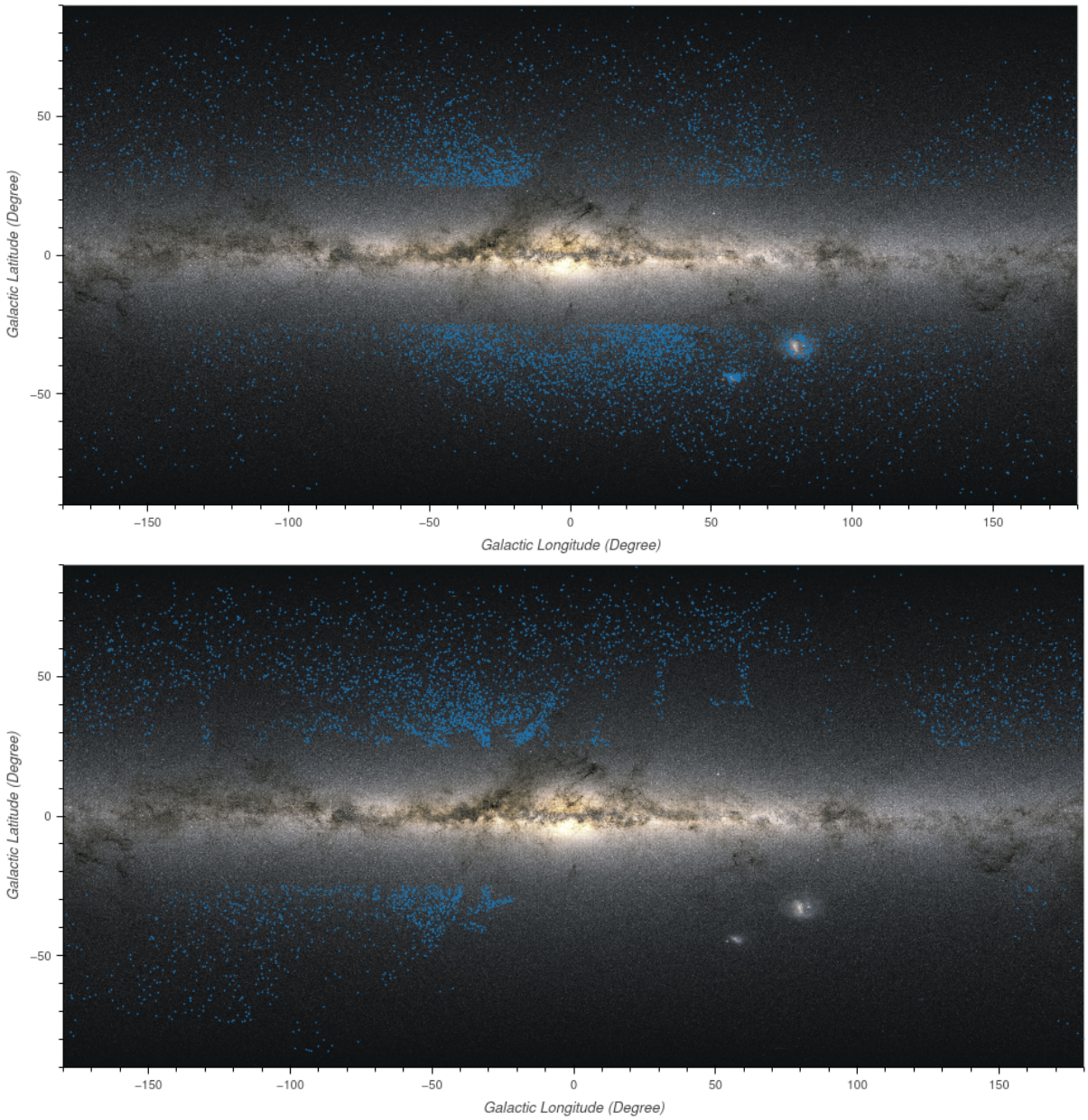

Figure 16: All-sky samples of blue horizontal branch star candidates with (blue data points), found using synthetic CaHK and synthetic SDSS photometry from *Gaia* XP spectra (top panel), and synthetic CaHK combined with SDSS DR18 photometry (bottom panel). These samples are overlay-ed on a sky map that show the brightness of the almost 1.7 billion sources in *Gaia* DR2, measured in the *Gaia* G, G_{BP} and G_{RP} photometric bands. A cut in Galactic latitude ($-25^\circ < b^\circ < 25$) is performed to avoid contamination from young, hot main sequence stars in the disk. The stripe features that appear for the SDSS DR18 sample are a consequence of the SDSS footprint. Credits for the *Gaia* DR2 sky map: *ESA/Gaia/DPAC, A. Moitinho / A. F. Silva / M. Barros / C. Barata, University of Lisbon, Portugal; H. Savietto, Fork Research, Portugal*. These figures are made by using the `mw_plot` Python package (for more information see `https://milkyway-plot.readthedocs.io/en/latest/index.html`)

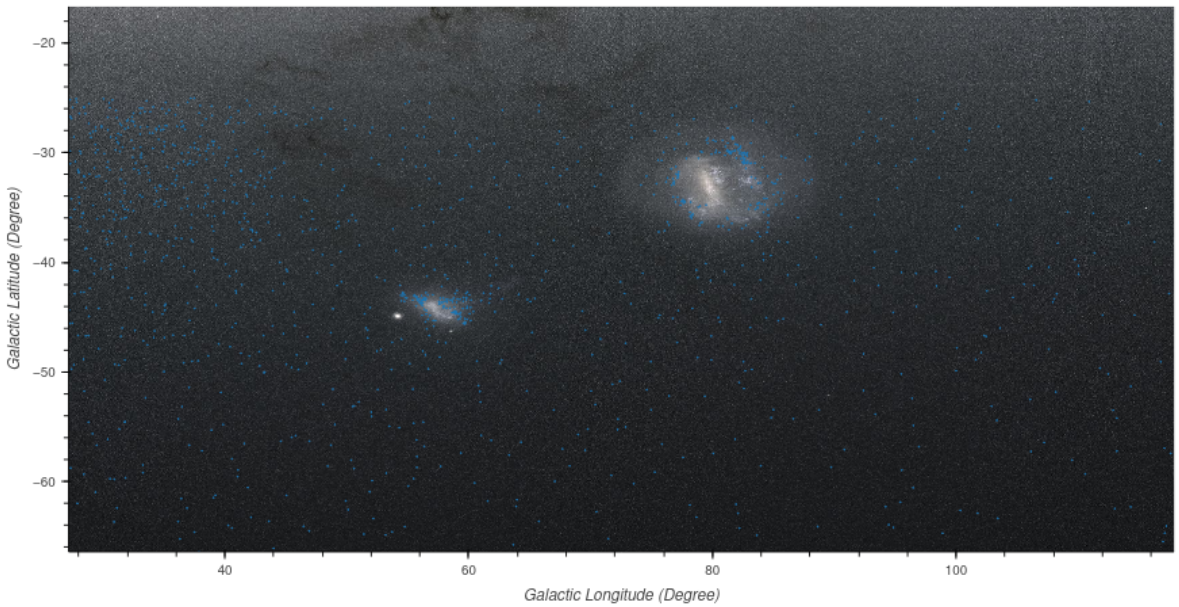

Figure 17: A zoomed-in version of the top panel from Figure 16 to show that the high concentration of this sample of blue horizontal branch star candidates near l, b = (60°, −40°) and l,b = (80°, −30°) coincide with the Small -and Large Magellanic Clouds (SMC and LMC) respectively. Credits for the *Gaia* DR2 sky map: *ESA/Gaia/DPAC, A. Moitinho / A. F. Silva / M. Barros / C. Barata, University of Lisbon, Portugal; H. Savietto, Fork Research, Portugal.* This Figure were made by using the `mw_plot` python package.

4.3 Tracing stellar streams

In this Section we investigate spatial overlap between our two samples of BHB candidates and stellar streams. Here we stress that additional proper motion and/or radial velocity information is needed to further confirm that stars are actual part of stellar streams (Helmi, 2020). We study this spatial overlap by utilizing the `galstreams` Python package[3], which includes the extensive galstreams library from Mateu (2023), which provides, among other things, the distances and galactic coordinates of 95 Galactic stellar streams. We separately investigate the Sagittarius stream in the next Section. Even though most stellar streams are found in the outer halo, some stellar streams, or parts of them, can be found within the distance range of our BHB candidates. Figure 18 encapsulates the spatial information (Galactic coordinates and heliocentric distances) of the stellar streams and our BHB candidates. Figure 19 includes polar plots, which show the galactic longitude (circularly) and the distance (radially), which makes it easier to spot spatial overlap between streams and our samples. Note that the distance uncertainties cannot be read from these plots. We look for BHB candidates stars that are within a 3D radius of 0.6 kpc (which is roughly three times the average distance uncertainty) from stellar streams, using the distances and coordinates provided by `galstreams` and find six BHB candidates that are located within the C-7-I21, Cocytos-G09, GD-1-I21, M92-I21, OmegaCen-I21 and Svol-I21 stellar streams (see Table 2 for more details). Here we did not account for the fact that stellar streams have a certain width and that there are some systematic uncertainties in the distances. Therefore we suspect that more BHB candidates can be located in some of the stellar streams and suggest further investigation. By carefully studying Figure 18 and 19 we think that some of our BHB candidates

[3]See https://github.com/cmateu/galstreams

might be in the Svol, *Gaia*-6, M68-Fjorm, ACS and the M2 stellar streams and suggest follow-up on this.

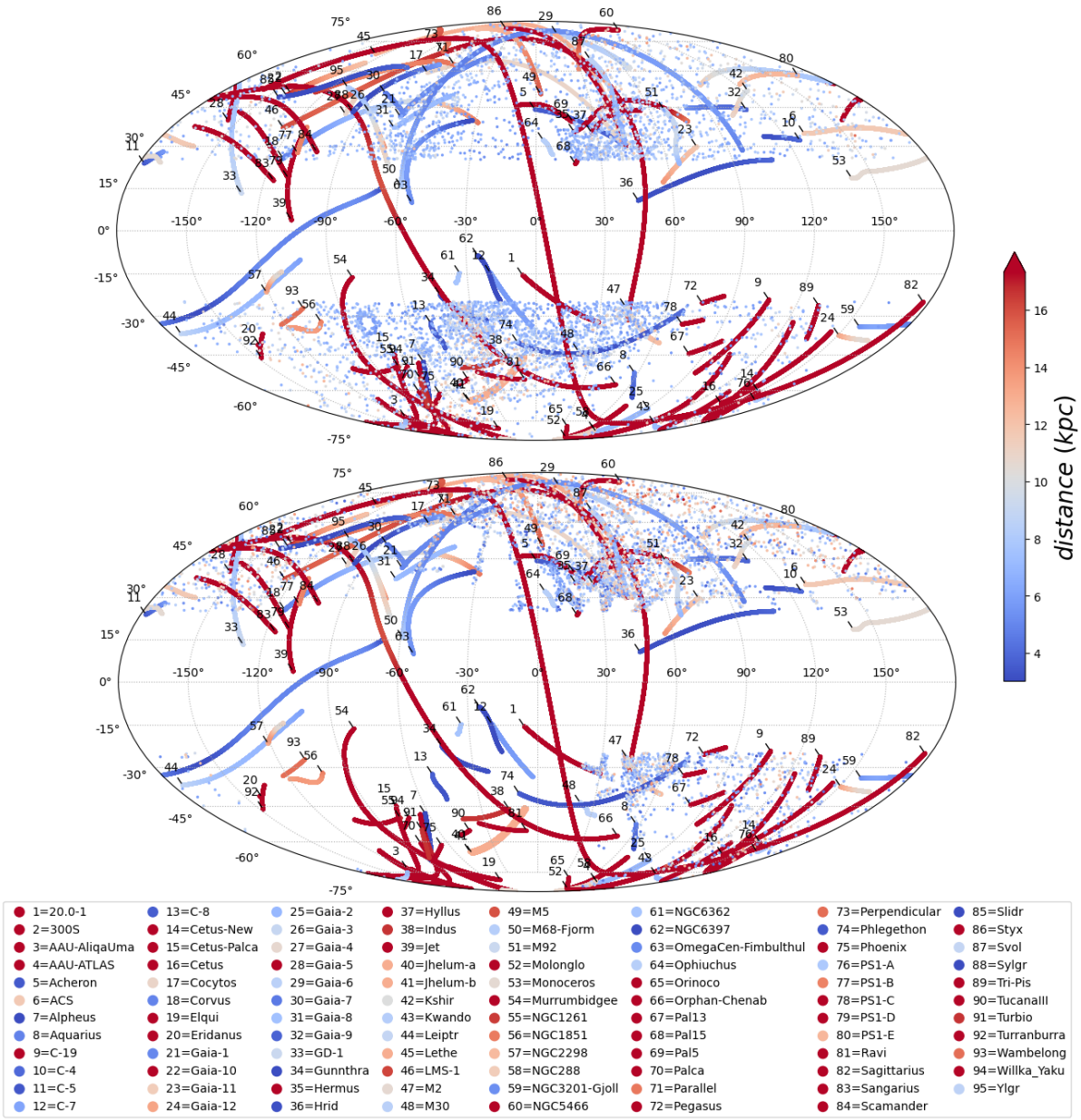

Figure 18: This figure shows the spatial overlap of our two samples of blue horizontal branch stars with stellar stream models from the galstreams library (Mateu, 2023). The synthetic SDSS sample (top panel) and the SDSS DR18 sample (bottom panel) are color-coded by their heliocentric distances and shown using a Mollweide projection. Additionally, the distance and galactic coordinates of 95 Galactic stellar streams are plotted and labelled at their end-points.

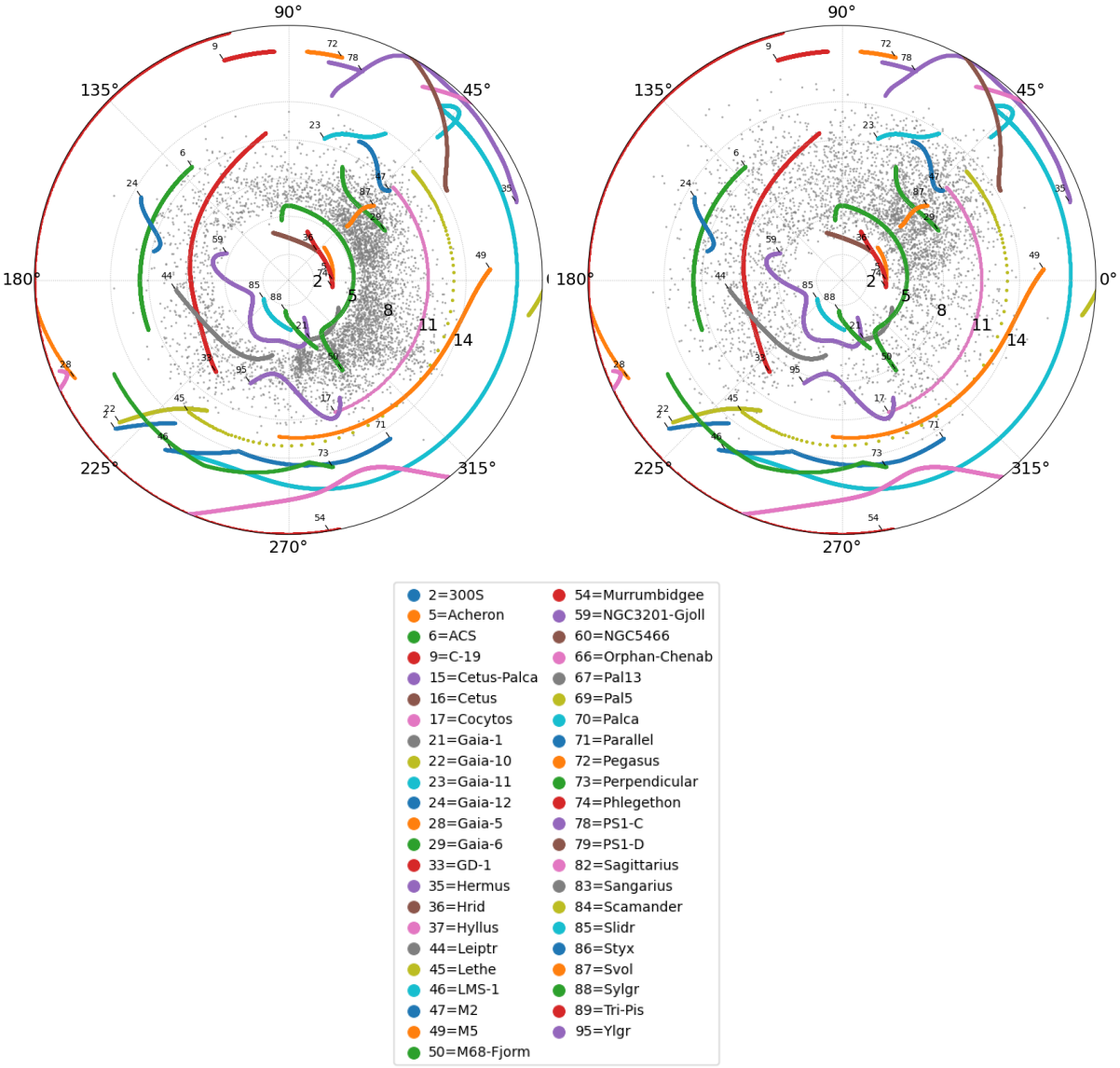

Figure 19: A polar plot showing the the galactic longitude l and the heliocentric distances in kpc (shown radially), for our synthetic SDSS (right panel) and SDSS DR18 sample (left panel) of blue horizontal branch stars (shown as grey dots). Additionally, stellar streams from the galstreams library (Mateu, 2023) are shown and labelled at their end-points.

4.4 Tracing the Sagittarius stream

As detailed in the Introduction, the Sagittarius stream is one of the most important stellar streams in unravelling the history of the Milky Way. Its vastness can be used to constrain the dark matter potential of our Galaxy. To investigate the spatial overlap with the Sagittarius stream, we use the Sagittarius stream model from Law and Majewski (2010). The Sagittarius stream coordinate system from Law and Majewski (2010) is defined in a plane almost perpendicular to the plane of the Milky Way, with its origin placed in the Sagittarius dwarf galaxy.

To convert from the Equatorial coordinate system (right ascension: α, declination: δ) to the Sagittarius stream coordinate system (Λ, B), we use the coordinate transformation equations from Belokurov et al. (2014):

$$\begin{aligned}\Lambda = &\ \text{atan}\,2(-0.93595354\cos(\alpha)\cos(\delta) - 0.31910658\sin(\alpha)\cos(\delta) + 0.14886895\sin(\delta),\\ &\ 0.21215555\cos(\alpha)\cos(\delta) - 0.84846291\sin(\alpha)\cos(\delta) - 0.48487186\sin(\delta))\\ \text{B} = &\ \arcsin(0.28103559\cos(\alpha)\cos(\delta) - 0.42223415\sin(\alpha)\cos(\delta) + 0.86182209\sin(\delta))\end{aligned} \tag{9}$$

Here Λ increases in the direction of Sagittarius motion. The definitions above differ slightly from the $(\Lambda_\odot, B_\odot)$ given by Law and Majewski (2010), but can be converted with $\Lambda_\odot = 360° - \Lambda$ and $B_\odot = -B$. The Sagittarius stream model in the Λ, B coordinate system is showed in Figure 20. The model covers distances from 0 kpc to above 100 kpc. To investigate the overlap between this model and our BHB candidate samples, only stars with $-20° < B < 20°$ are considered, which corresponds to the coordinates where most points in the model are located (as visible from Figure 20). Focusing only on the parts of the stream that have a similar distance range as the BHB candidates ($\sim 0 - 20$ kpc), allows us to compare the overlap of the stream and the BHB candidate samples.

From Figure 21 it can be seen that there is spatial overlap with the Sagittarius stream model in both samples. Most notably are the BHB candidates in the SDSS DR18 sample around $\Lambda = -100°$, which trace the Sagittarius stream from around 5 kpc up till distances up till 15 kpc. To a lesser extent this overlap is also visible with stars in the synthetic SDSS sample. There is also some spatial overlap around $\Lambda = 100°$. Additionally, there are some BHB candidates around $\Lambda \approx -150°$, which fall a little outside this particular model of the Sagittarius stream model and lie close to the anti-centre of the Milky Way. More investigation is needed to see whether these stars belong to the disk or that they might be part of the Sagittarius stream. There also some BHB candidates around $\Lambda = 0$, which corresponds to the space around the Galactic centre. In Appendix D we provide a table which includes a selection of BHB stars candidates within the SDSS DR18 sample that spatially overlap with the Sagittarius stream.

BHB candidates that spatially overlap stellar streams					
stellar stream	Gaia source ID	ra (deg)	dec (deg)	distance (kpc) ± 0.2	sample
C-7-I21	667344318864808627	301.57	-44.98	5.91	synthetic SDSS
Cocytos-G09	4561190092227313792	255.34	19.28	10.99	both
GD-1-I21	1566732502541484928	200.54	58.09	10.12	synthetic SDSS
M92-I21	4611010578987698816	268.60	39.21	8.65	both
OmegaCen-I21	6155408113422464512	192.59	-36.16	5.72	synthetic SDSS
Svol-I21	1315570442934606208	241.50	26.66	6.97	both

Table 2: Blue horizontal branch star candidates from our samples that are cross-matched within 0.6 kpc with stellar streams from `galstreams` (Mateu, 2023) based on 3D spatial information.

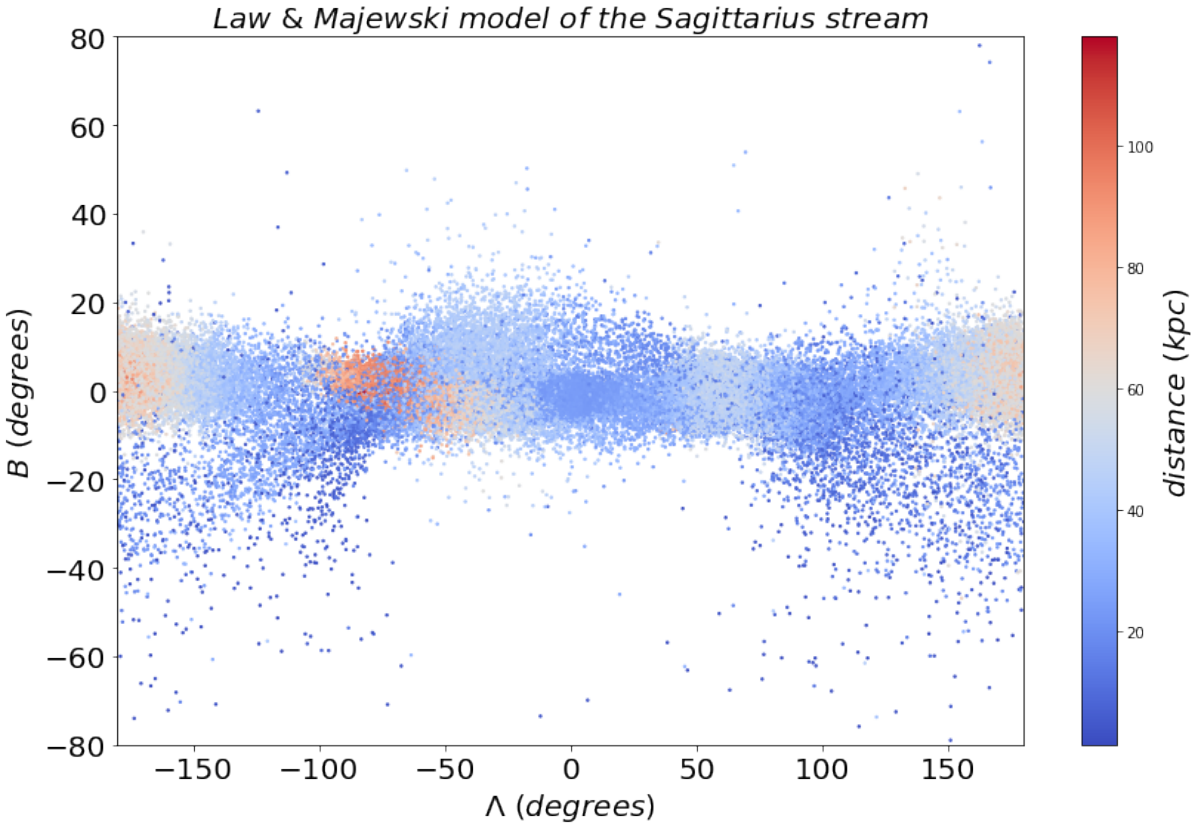

Figure 20: The Saggitarius stream modelled by Law and Majewski (2010), shown in the coordinate system (Λ, B) of the orbital plane of the Sagittarius stream and color-coded by predicted distances.

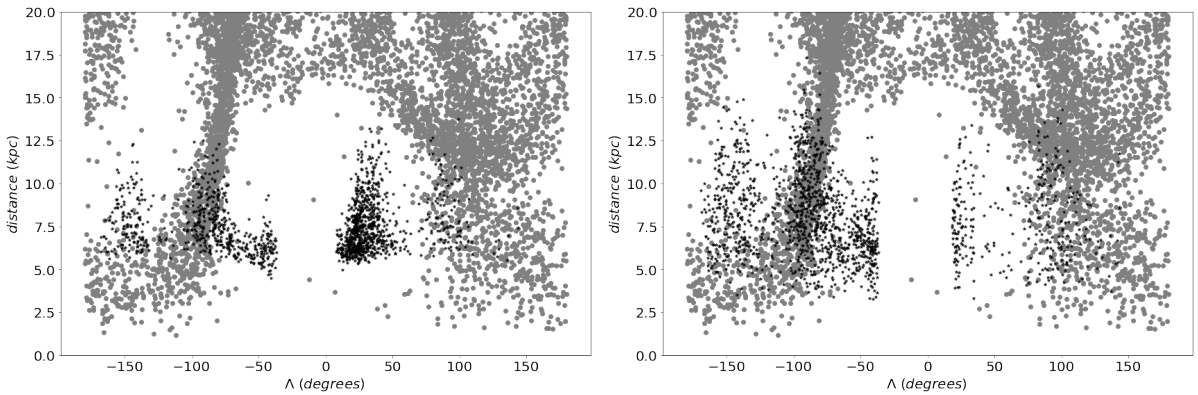

Figure 21: The heliocentric distances against Λ in the coordinate system of the orbital plane of the Sagittarius stream (Λ, B) for a sample of blue horizontal branch (black data points) with $-20° < B < 20°$ and $-25° < b < 25°$. The Sagittarius stream model by Law and Majewski (2010) is overplotted in grey. The sample of candidate blue horizontal branch stars in the right panel is obtained using synthetic SDSS data and synthetic Pristine CaHK photometry, while the sample on the left panel is obtained using SDSS DR18 and synthetic Pristine CaHK photometry.

5 Discussion: Improvements and future outlook

5.1 Comparison with similar works

Our two samples of BHB candidate stars present an excellent case for spectroscopic follow-up to further confirm their nature. The purity of our two BHB candidate samples (99% in both cases) are unprecedentedly high, comparable to the purity of the BHB selection in the work of Starkenburg et al. (2017) which is 91%. Our measured completeness (75% for the sample employing synthetic SDSS photometry and 81% for the sample employing SDSS DR18 photometry) are higher than the average completeness obtained in previous works using broad-band photometry, which is about 50-70% (Starkenburg et al., 2019). Additionally, our samples have an uncommonly high sky coverage compared to those of other works which use similar methods. Whereas the work of Starkenburg et al. (2019) is limited to the Pristine footprint (which has grown since their publication to 6000 deg^2), the use of *Gaia* synthetic Pristine CaHK photometry allows us to look all-sky. The limiting factor in our work is the distance range of our samples, which only includes candidate BHB stars in the inner halo, even when employing SDSS DR18 *ugr* photometry (see Figure 7). In the works of Fukushima et al. (2018), Starkenburg et al. (2019) and Gentile Fusillo et al. (2015) the high luminosity of BHB stars is used to look far into the outer halo. The shallowness of our samples is a consequence of the limiting magnitude of the Pristine CaHK filter, which is $m_g \approx 16.3$ for $30 < SNR < 35$, which further restricts our sample, which is already limited to 17.65 mag, which corresponds to a distance of ~ 27 kpc for a BHB star. In addition, we applied a quite strict SNR selection of SNR > 30, as suggested in the work of Gaia Collaboration et al. (2022). Applying a more lenient SNR selection might allow us to look further into the halo.

5.2 Alternative literature sample of blue horizontal branch stars and blue stragglers

The Xue et al. (2008) sample is used as a training sample in this work but is also used to fit the ridgelines and to determine the purity and completeness of our BHB candidate samples. In the work of Xue et al. (2008) the widths of the Balmer lines are measured to to discriminate BHBs from BS, whereas recent works - such as Barbosa et al. (2022) - employ newer techniques as a means of classification, such as fitting spectral line profiles or using log g information. There are some BHB stars in this sample with higher $g_0 - r_0$ which fall more into the BS sequence. In the work of Starkenburg et al. (2019) it is also pointed out that the spectroscopic classification in this regime becomes less trustworthy and they consider it likely that these stars are misclassified. Using a sample that is based on newer techniques, such as the sample from Barbosa et al. (2022) might therefore be favorable. We also use the Xue et al. (2008) sample to fit the ridgelines, from which we subsequently draw a purity and completeness estimate. As a consequence, our success rates might represent a best case scenario and using an independent literature sample to calculate the completeness and purity from these ridgelines might give a more realistic view. The size of our clipped Xue et al. (2008) sample is also too small to get an idea of how purity and completeness might change with magnitude. Using a Monte Carlo simulation to create mock datasets, as is done in the methods of Starkenburg et al. (2019), might help to get more insight in the magnitude dependence of the sample performances.

5.3 Attempt to remove variable stars

In the methodology described in Section 3.3, no cuts are performed to remove contamination from variable stars. As the published sample with XP spectra by Gaia Collaboration et al. (2022) only contains mean spectra (spectra generated from several observations) and only contains sources with high quality XP spectra, we do not expect a lot of contamination from variable stars. We tried to exclude variable stars from the synthetic SDSS sample, by using the `pvar` < 0.3 selection from the work of Martin, N.F., et al. (2023) [in preparation] - where `pvar` defines the probability of a source to be variable - but this left only a handful of BHB stars in the Xue et al. (2008) sample. This suggests that perhaps BHB stars have some unknown intrinsic low-level variability or that this particular `pvar` selection only works well for FGK type stars, which are the stellar types considered in their work. Further investigation is needed.

5.4 Future investigation in the use of J-PAS filters

In Section 3.5 we shortly demonstrated how synthetic photometry from the J-PAS 4200 filter in combination with SDSS DR18 *ugr* photometry can be used to discriminate BHB from BS stars in colorspace. Due to time constraints we did not manage to apply this method to our sample of A-type candidates and obtain a sample of candidate BHB stars based on this. In future work we therefore intend to further explore the use of the J-PAS 4200 filter in BHB and BS classification. We think that the application of this method on the Xue et al. (2008) is promising, as relatively more BHB stars survive the selection procedure (see Figure 10), suggesting a higher completeness of BHB selection compared to the other two methods which employ synthetic Pristine CaHK. Additionally, we suspect that the use of J-PAS 4200 allows us to look further into the halo, but that the relatively larger overlap between the BHB and BS sequence in color space might result in more contamination from BS.

Furthermore, we tested the use of J-PAS 8000 and J-PAS 8800 and think that these filters might be able to stress the flux difference between BHB and BS stars around the Passchen jump (see Figure 24). In addition, we suggest future investigation in the use of clustering algorithms, instead of ridgelines, as a means to find any groupings in color space. This method might present a much more general case to identify BHB and BS stars, which does not depend on the filters one uses.

5.5 Further investigation of stellar stream membership

In Section 4.3 we find that six of our BHB candidates are located in a selection of stellar streams (see Table 2) and suspect that further investigation will allow more BHB candidates to be located in some of the stellar streams. In Section 4.4, we confirm spatial overlap between our BHB candidates and the Sagittarius stream. In both cases, the addition of proper motion and/or radial velocity information can further confirm the BHB's stellar stream membership. From another point of view, BHB stars and other standard candles provide more accurate distance measurements and can therefore help in providing more accurate distances to stellar streams. On a side note, putting a constraint on the proper motion might also help in further cleaning our BHB sample from potential BS contamination. In the work of Lancaster et al. (2019) it is pointed out that, as BS are intrinsically fainter, BS that are misclassified as a BHB star will on average be placed much further out. This will then result in a tangential velocity (calculated from the proper motion) that might be too high.

If some of our BHB candidates can be confirmed to be members of stellar streams, this provides us with knowledge of a stream's progenitor galaxy (Starkenburg et al., 2017). BHB stars are thought to trace old stellar populations and if we find, for instance, that a stream has many BHB members, this might indicate that the progenitor galaxy was quite old and metal-poor. Additionally, it would be interesting to compare the distribution of our BHB candidates with the distribution of other stellar populations, to get more information about the stellar populations that were present in the progenitor galaxies of the stellar streams. One could use, for example, a sample of main sequences stars (see for example Viswanathan et al. 2023) or RR Lyrea stars (see for example Xin-Yi Li et al. 2023).

5.6 XP spectra for about 2 billion sources

At the moment XP spectra is available for about 220 million sources but will amount to the entire Gaia data set in future releases, which will contain about 2 billion sources (Gaia Collaboration et al., 2022). Not before the end of 2025, the fourth *Gaia* data release (DR4) will be published and Gaia DR5, which will be based on all mission data, will not be published before the end of 2030[4]. The methods we present in this thesis hold the potential to significantly increase the number of BHB candidates by using the XP spectra of these 2 billion sources. Moreover, future releases will contain sources down to G $\sim$ 20.5 mag. This will enable us to use BHB stars to trace the outer parts of our Milky Way, up till distances as large as 100 kpc and to map the outer halo's substructures, including the outer parts of the Sagittarius stellar stream.

[4]For more information see *https://www.cosmos.esa.int/web/gaia/release.*

6 Conclusion

We demonstrate for the first time that synthetic photometry from *Gaia* spectrophotometry (Gaia Collaboration et al., 2022) can be used to discriminate standard candle BHB stars from the contamination population of BS. With information from synthetic stellar spectra and by employing the difference in surface gravity between BHB and BS stars, we tested several filter combinations on their BHB selection. We build on the success of the methods of Starkenburg et al. (2019) by using synthetic Pristine CaHK and synthetic SDSS *ugr* photometry to create an all-sky sample of $\sim$ 5300 BHB candidates with an unprecedentedly high purity of 99% and a completeness of 75%. To pick-up fainter magnitudes, SDSS DR18 *ugr* photometry in combination with synthetic Pristine CaHK photometry is used to create a sample of $\sim$ 4000 BHB candidates, improving the completeness to 81%, but limiting us to the SDSS footprint. We demonstrate that J-PAS filters could be used to discriminate BHB and BS stars from flux differences around the Balmer and Passchen jump and suggest follow-up. Using these new and pure samples of BHB candidates we create 3D maps of the inner halo of the Galaxy, up until distances of $\sim$ 12.5 kpc when employing synthetic SDSS photometry and $\sim$ 17.5 kpc when employing SDSS DR18 data. Spectroscopic follow-up could provide additional proof that the stars in our sample are indeed BHB stars. We investigate spatial overlap between our BHB candidates and stellar streams, hereby locating a few BHB candidates in for instance the Svol-I21 and M92-I21 stellar streams. However, we suspect that more stars in our samples can be located to some of the stellar streams and suggest further investigation. We locate many BHB candidates within the Sagittarius stream and encourage to further look into the origin of the stars around $\Lambda = -150°$. Proper motion or radial velocity information can further confirm stream membership of BHB stars. For future work, it might be of interest to compare the number of BHB stars within a stellar stream with those other stellar populations to find out more about the properties of the progenitor galaxy. In future *Gaia* releases, the XP spectra of 2 billion sources will become available and by applying the methods we present in this thesis we might be able to significantly increase the number of BHB candidates as well as mapping the halo out to distances of 100 kpc, hereby further unravelling the history of our Milky Way.

References

Abdurro'uf, Accetta, et al. (2022). The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data. *ApJS*, 259(2):35.

Alam, S. et al. (2015). The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III. *ApJS*, 219(1):12.

Almeida, A. et al. (2023). The eighteenth data release of the sloan digital sky surveys: Targeting and first spectra from sdss-v.

Barbosa, F. O. et al. (2022). The SDSS-Gaia View of the Color-Magnitude Relation for Blue Horizontal-branch Stars. *ApJ*, 940(1):30.

Belokurov, V. et al. (2014). Precession of the sagittarius stream. *Monthly Notices of the Royal Astronomical Society*, 437(1):116–131.

Benitez, N. et al. (2014). J-pas: The javalambre-physics of the accelerated universe astrophysical survey.

Binney, J. and M.Merrifield, editors (1998). *Galactic Astronomy*. Princeton university press.

Brown, W. R. et al. (2009). Velocity dispersion profile of the milky way halo. *The Astronomical Journal*, 139(1):59.

Chambers, K. C. et al. (2019). The pan-starrs1 surveys.

Deason, A. J., Belokurov, V., and Evans, N. W. (2011). The Milky Way stellar halo out to 40 kpc: squashed, broken but smooth. *Monthly Notices of the Royal Astronomical Society*, 416(4):2903–2915.

Doi, M. et al. (2010). PHOTOMETRIC RESPONSE FUNCTIONS OF THE SLOAN DIGITAL SKY SURVEY IMAGER. *The Astronomical Journal*, 139(4):1628–1648.

Draine, B., editor (2011). *Physics of the Interstellar and Intergalactic Medium*. Princeton university press.

Fukushima, T. et al. (2018). Structure of the milky way stellar halo out to its outer boundary with blue horizontal-branch stars. *Publications of the Astronomical Society of Japan*, 70(4).

Gaia Collaboration and Brown, A. G. A., a. o. (2021). Gaia Early Data Release 3. Summary of the contents and survey properties. , 649:A1.

Gaia Collaboration, P. et al. (2016). The Gaia mission. , 595:A1.

Gaia Collaboration, Montegriffo, P. et al. (2022). Gaia Data Release 3: The Galaxy in your preferred colours. Synthetic photometry from Gaia low-resolution spectra. *arXiv e-prints*, page arXiv:2206.06215.

Gentile Fusillo, N., Hermes, J., and Gänsicke, B. (2015). A search for variable white dwarfs in large area time domain surveys: a pilot study in sdss stripe 82. *Monthly Notices of the Royal Astronomical Society*, 455.

Harris, W. E. (1996). A Catalog of Parameters for Globular Clusters in the Milky Way. , 112:1487.

Helmi, A. (2020). Streams, substructures, and the early history of the milky way. *Annual Review of Astronomy and Astrophysics*, 58(1):205–256.

Jordi, C. et al. (2006). The design and performance of the gaia photometric system. *Monthly Notices of the Royal Astronomical Society*, 367(1):290–314.

Lancaster, L. et al. (2019). The halo's ancient metal-rich progenitor revealed with BHB stars. *Monthly Notices of the Royal Astronomical Society*, 486(1):378–389.

Law, D. R. and Majewski, S. R. (2010). The Sagittarius Dwarf Galaxy: A Model for Evolution in a Triaxial Milky Way Halo. *ApJ*, 714(1):229–254.

LeBlanc, F., editor (2010). *An Introduction to Stellar Astrophysics*. John Wiley and Sons, Ltd.

Lejeune, T., Cuisinier, F., and Buser, R. (1998). VizieR Online Data Catalog: A standard stellar library. II. (Lejeune+ 1998). *VizieR Online Data Catalog*, pages J/A+AS/130/65.

Libeskind, N. I., Knebe, A., et al. (2011). Disentangling the dark matter halo from the stellar halo. *Monthly Notices of the Royal Astronomical Society*, 418(1):336–345.

Mateu, C. (2023). galstreams: A library of Milky Way stellar stream footprints and tracks. *MNRAS*, 520(4):5225–5258.

Munari, U. et al. (2005). An extensive library of 2500 10 500 Å synthetic spectra. , 442(3):1127–1134.

Raso, S. et al. (2019). Spectral Energy Distribution of Blue Stragglers in the Core of 47 Tucanae. *ApJ*, 879(1):56.

Schlafly, E. F. and Finkbeiner, D. P. (2011). MEASURING REDDENING WITH SLOAN DIGITAL SKY SURVEY STELLAR SPECTRA AND RECALIBRATING SFD. *The Astrophysical Journal*, 737(2):103.

Schlegel, D. J. et al. (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *ApJ*, 500(2):525–553.

Starkenburg, E. et al. (2017). The pristine survey – i. mining the galaxy for the most metal-poor stars. *Monthly Notices of the Royal Astronomical Society*, 471(3):2587–2604.

Starkenburg, E. et al. (2019). The Pristine survey - VII. A cleaner view of the Galactic outer halo using blue horizontal branch stars. *MNRAS*, 490(4):5757–5769.

Stoughton, Chris, L. R. H. et al. (2002). Sloan Digital Sky Survey: Early Data Release. , 123(1):485–548.

Subramanian, S. and Subramaniam, A. (2009). Depth estimation of the large and small magellanic clouds. *A&A*, 496(2):399–412.

Thomas, G. F. et al. (2018). A-type stars in the Canada-France Imaging Survey I. The stellar halo of the Milky Way traced to large radius by blue horizontal branch stars. *MNRAS*, 481(4):5223–5235.

Viswanathan, A., Starkenburg, E., Koppelman, H. H., Helmi, A., Balbinot, E., and Esselink, A. F. (2023). Hidden deep in the halo: selection of a reduced proper motion halo catalogue and mining retrograde streams in the velocity space. *Monthly Notices of the Royal Astronomical Society*, 521(2):2087–2102.

Xin-Yi Li, Y. H. et al. (2023). Photometric metallicity and distance estimates for ∼136,000 RR lyrae stars from gaia data release 3. *The Astrophysical Journal*, 944(1):88.

Xue, X. X. et al. (2008). The Milky Way's Circular Velocity Curve to 60 kpc and an Estimate of the Dark Matter Halo Mass from the Kinematics of ~2400 SDSS Blue Horizontal-Branch Stars. *ApJ*, 684(2):1143–1158.

Yanny, B. et al. (2000). Identification of a-colored stars and structure in the halo of the milky way from sloan digital sky survey commissioning data*. *The Astrophysical Journal*, 540(2):825.

Appendix

A Temperature sensitivity of blue horizontal branch stars and blue stragglers from synthetic stellar spectra

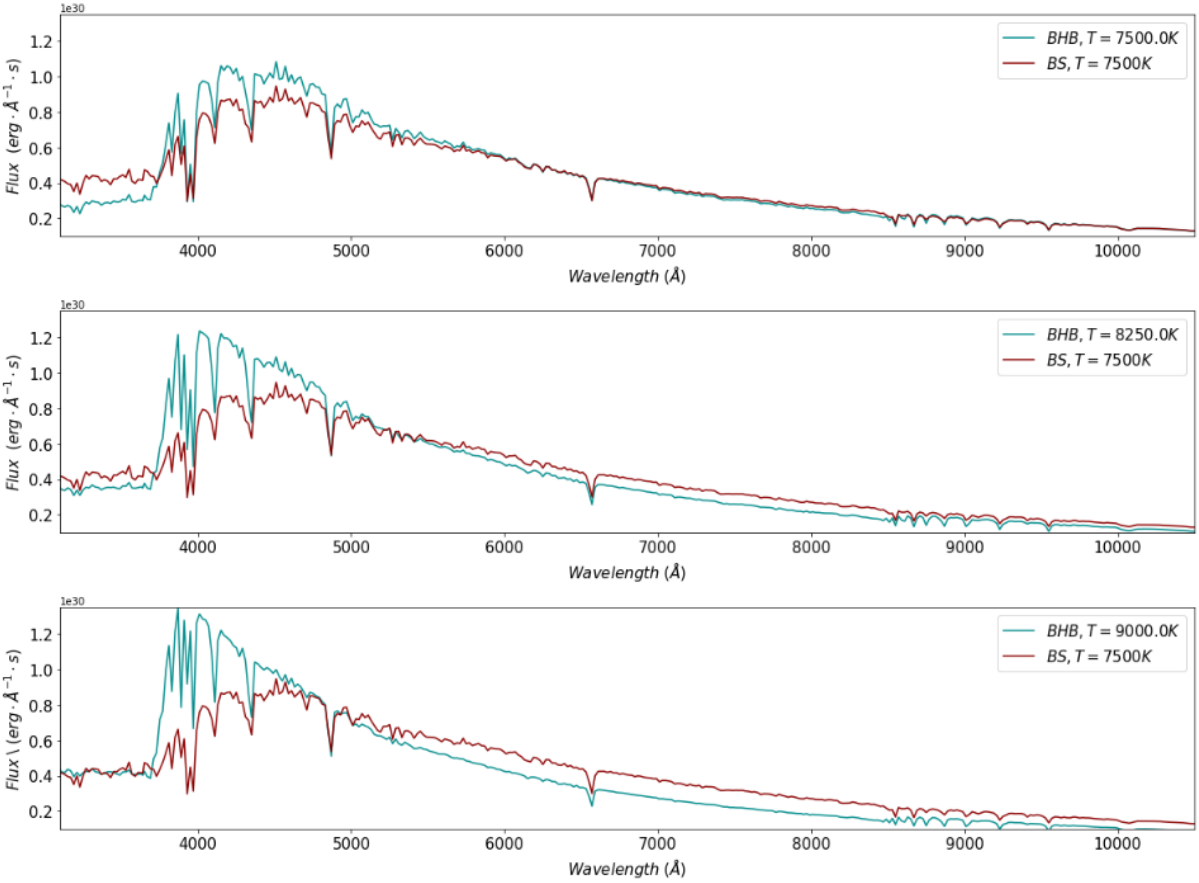

Figure 22: Synthetic stellar spectra from Lejeune et al. (1998) for stars with solar metallicity and $[Fe/H] = -0.5$, where one of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot) and the other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). To demonstrate the temperature sensitivity of the spectra, spectra are plotted for different T_{eff} for the blue horizontal branch star and at constant T_{eff} for the BS.

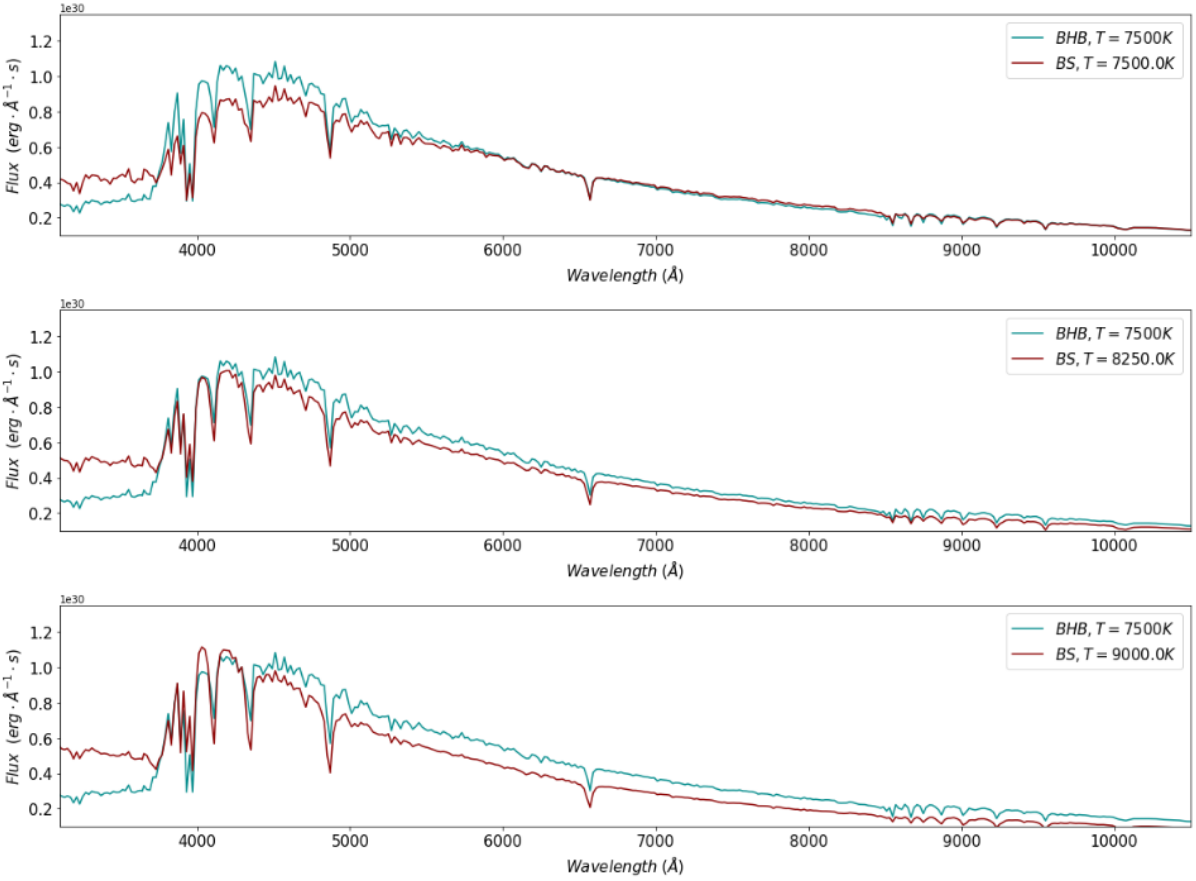

Figure 23: Synthetic stellar spectra from Lejeune et al. (1998) for stars with solar metallicity and $[Fe/H] = -0.5$, where one of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot) and the other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). To demonstrate the temperature sensitivity of the spectra, spectra are plotted for different T_{eff} for the blue straggler and at constant T_{eff} for the blue horizontal branch star.

B Filter combinations tested on the Xue et al. (2008) sample

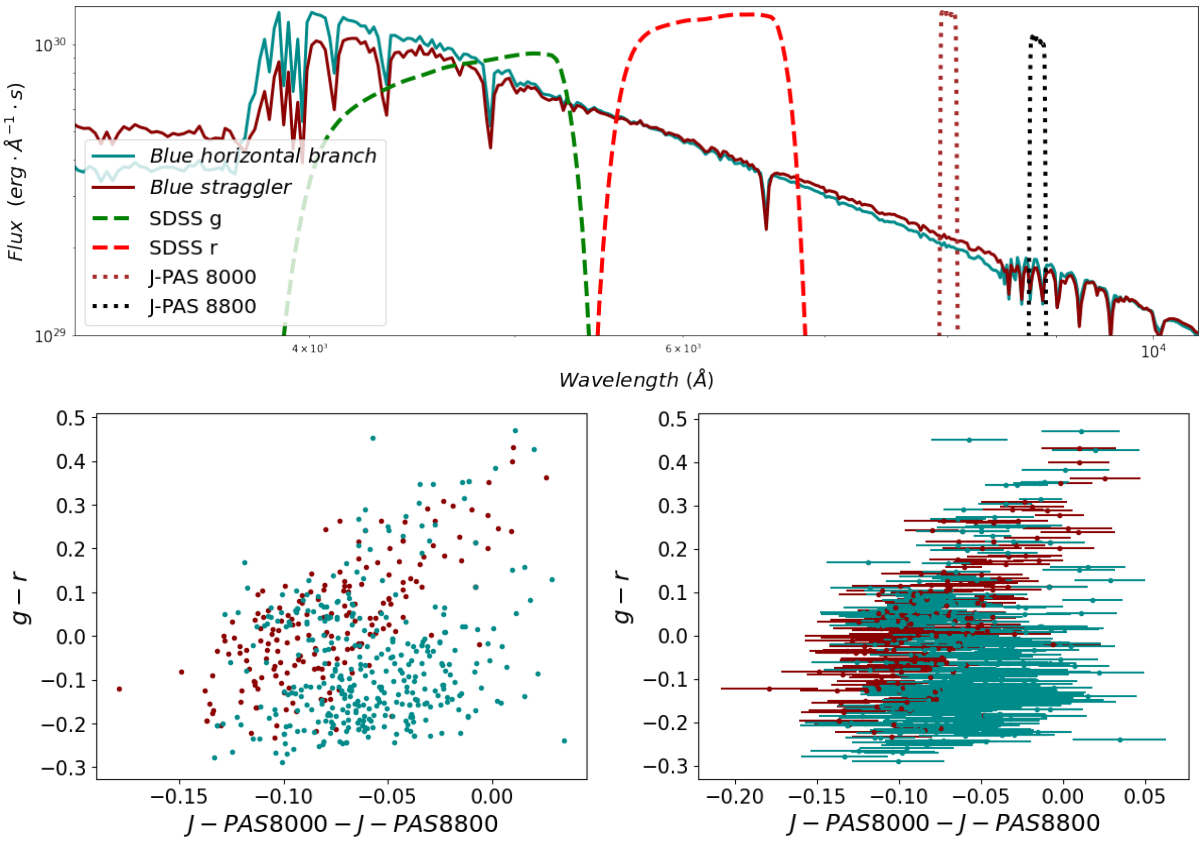

Figure 24: Top panel: Basel stellar library Lejeune et al. (1998) synthetic stellar spectra generated with the Python tool `pystellibs`. One of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot). The other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). The transmission curves of the SDSS gr bands (Doi et al., 2010) and the J-PAS 8000 and J-PAS 8800 (Benitez et al., 2014) are shown and scaled to be visible. Please note that we show here the non-standardised SDSS transmission curves, while in fact we use the slightly altered standardised versions. Note that the filters and spectra are plotted in logspace to bring out the flux difference around the Passchen jump. Bottom panels: Color-color space of blue horizontal branch stars (in blue) and blue stragglers (in red) from the Xue et al. (2008) sample that survive the selection criteria described in Section 3.2 using synthetic SDSS ug, J-PAS 8000 and J-PAS 8800 photometry. The uncertainties in the colours are shown as error bars in the left panel.

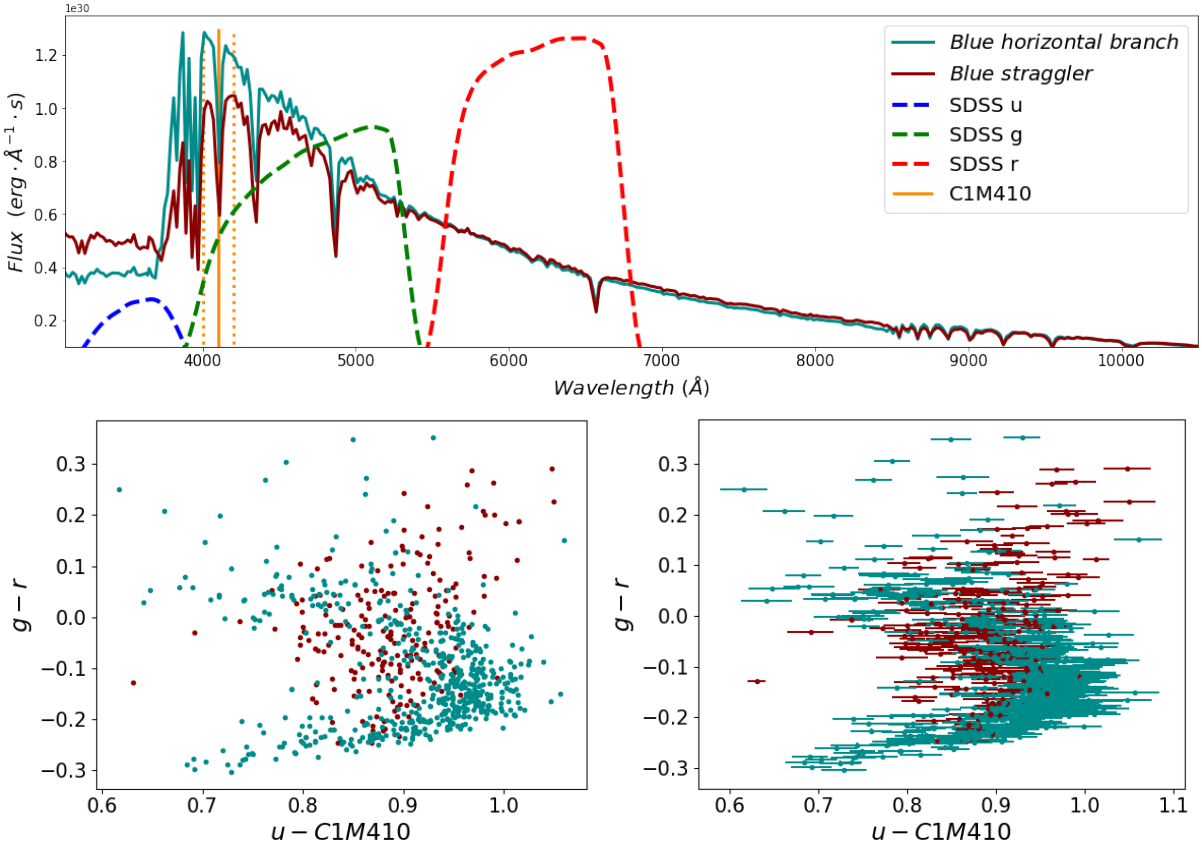

Figure 25: Top panel: Basel stellar library Lejeune et al. (1998) synthetic stellar spectra generated with the Python tool `pystellibs`. One of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot). The other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). The transmission curves of the SDSS *ugr* bands (Doi et al., 2010) are shown and scaled to be visible. For the C1M410 filter, the center of the band (solid vertical line) and the FWHM of the filter (dashed vertical lines) are shown (Jordi et al., 2006). Please note that we show here the non-standardised SDSS transmission curves, while in fact we use the slightly altered standardised versions. Bottom panels: Color-color space of blue horizontal branch stars (in blue) and blue stragglers (in red) from the Xue et al. (2008) sample that survive the selection criteria described in Section 3.2. The uncertainties in the colours are shown as error bars in the left panel.

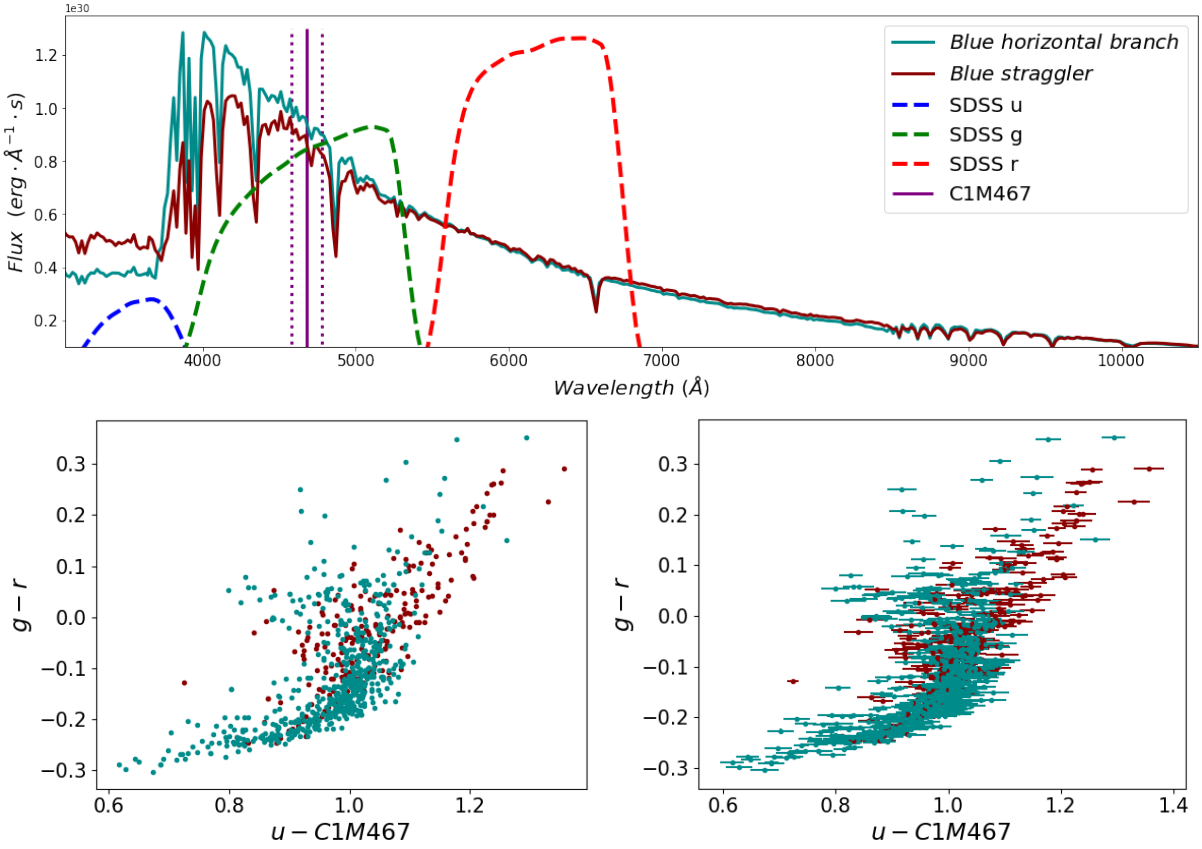

Figure 26: Top panel: Basel stellar library Lejeune et al. (1998) synthetic stellar spectra generated with the Python tool `pystellibs`. One of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot). The other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). The transmission curves of the SDSS *ugr* bands (Doi et al., 2010) shown and scaled to be visible. For the C1M467 filter, the center of the band (solid vertical line) and the FWHM of the filter (dashed vertical lines) are shown (Jordi et al., 2006). Bottom panels: Color-color space of blue horizontal branch stars (in blue) and blue stragglers (in red) from the Xue et al. (2008) sample that survive the selection criteria described in Section 3.2. The uncertainties in the colours are shown as error bars in the left panel.

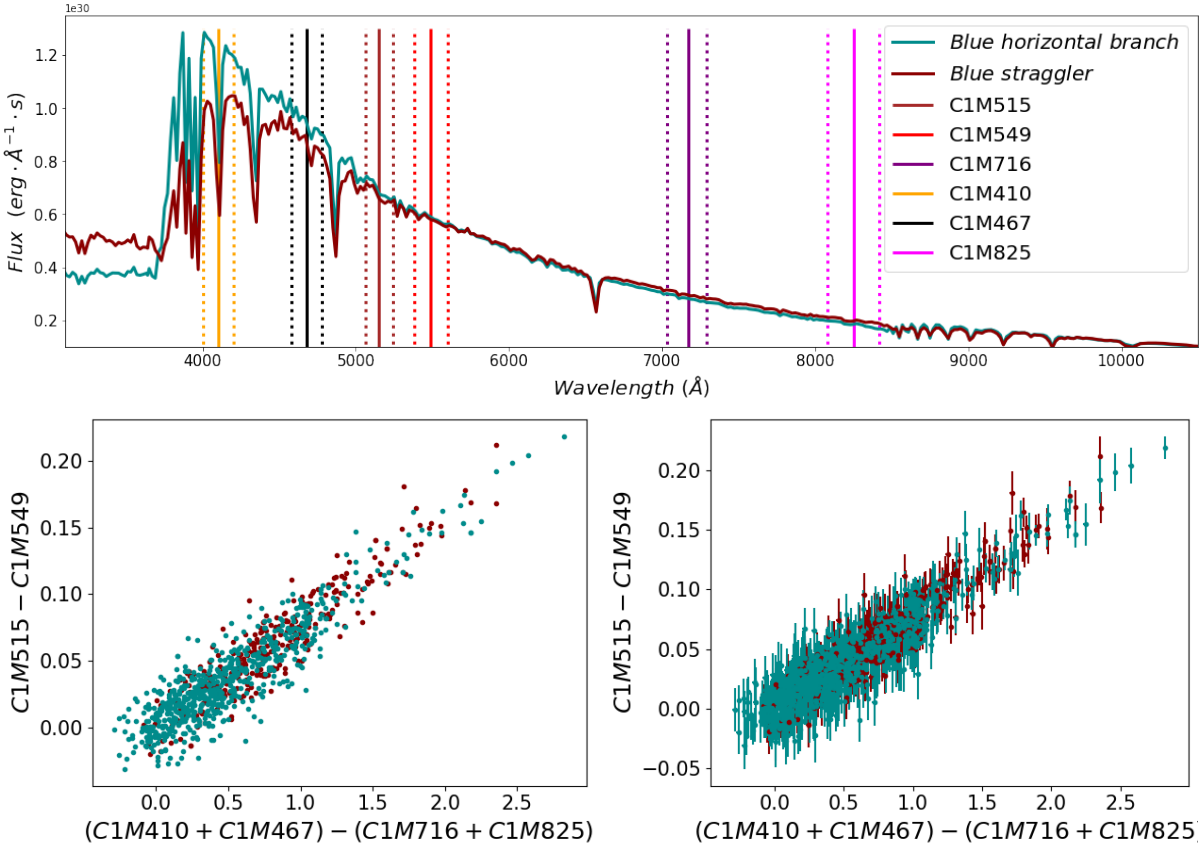

Figure 27: Top panel: Basel stellar library Lejeune et al. (1998) synthetic stellar spectra generated with the Python tool `pystellibs`. One of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot). The other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). For the C1M filters, the center of the bands (solid vertical line) and their FWHM (dashed vertical lines) are shown (Jordi et al., 2006). Bottom panels: Color space of blue horizontal branch stars (in blue) and blue stragglers (in red) from the Xue et al. (2008) sample that survive the selection criteria described in Section 3.2. The uncertainties in the colours are shown as error bars in the left panel.

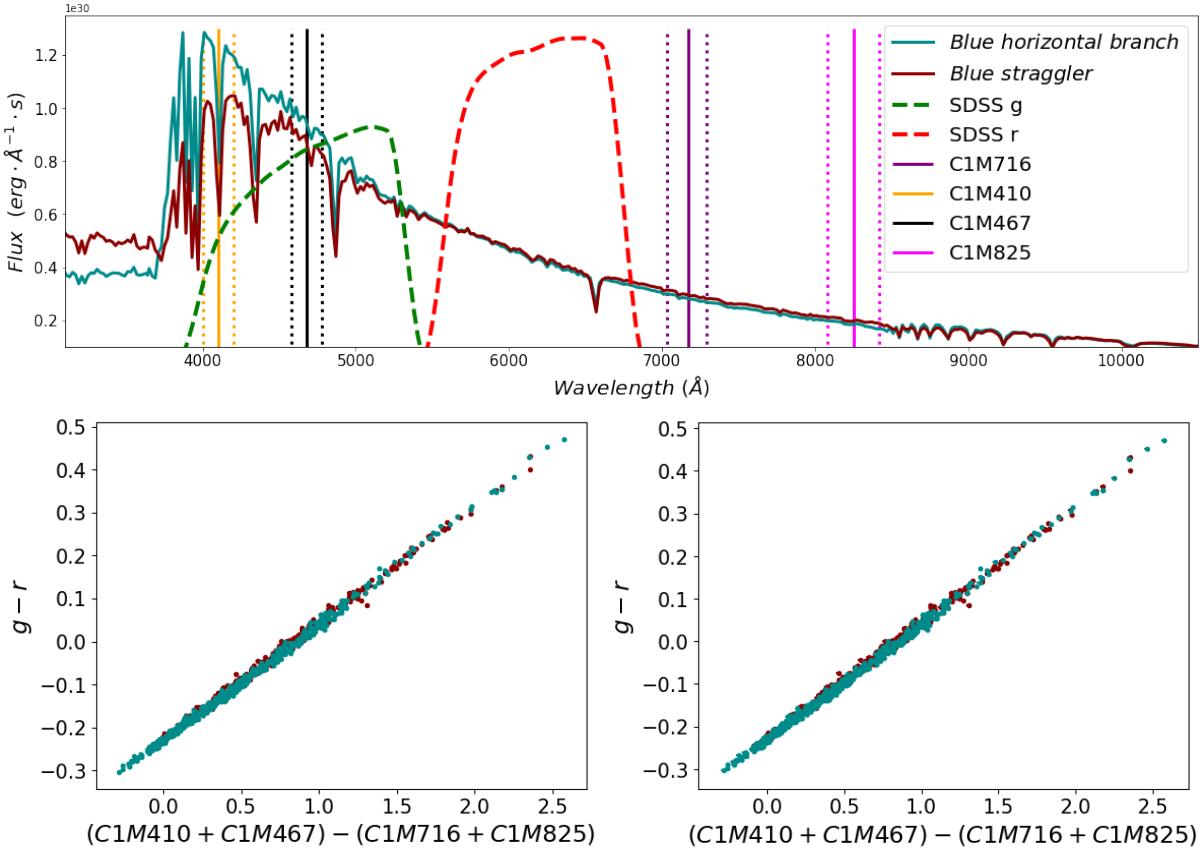

Figure 28: Top panel: Basel stellar library Lejeune et al. (1998) synthetic stellar spectra generated with the Python tool `pystellibs`. One of them has $log\ g = 3$ which is typical for a blue horizontal branch star (the blue line in the plot). The other has $log\ g = 5$ which is typical for a blue straggler (the red line in the plot). For the C1M filters, the center of the bands (solid vertical line) and their FWHM (dashed vertical lines) are shown (Jordi et al., 2006). The transmission curves of the SDSS *gr* bands (Doi et al., 2010) are shown and scaled to be visible. Please note that we show here the non-standardised SDSS transmission curves, while in fact we use the slightly altered standardised versionsBottom panels: Color space of blue horizontal branch stars (in blue) and blue stragglers (in red) from the Xue et al. (2008) sample that survive the selection criteria described in Section 3.2. The uncertainties in the colours are shown as error bars in the left panel.

C Extinction correction

In Section 3.3 and 3.4 the magnitudes in the synthetic $CaHK$ and synthetic SDSS ugr photometric bands are corrected for interstellar reddening, which is caused by dust along the line-of-sight between the observer and the source (Draine, 2011). To obtain the extinction correction $E(B-V)$ and calculate the extinction corrected magnitudes $(u_0, g_0, r_0, CaHK_0)$, the dust map from Schlegel et al. (1998) is used, which is provided by the `dustmaps` Python package. Since our interest lies within the halo, where dust and gas are less prominent, a first-order correction will be sufficient, in which dust and gas along the line-of-sight is integrated. Even though we look at the halo, performing an extinction correction is important, especially because the blue starlight of A-type stars is more prone to absorption and scattering by dust and gas Draine (2011). We use the reddening values $E(B-V)_{SFD}$ from Schlegel et al. (1998) provided by `dustmaps` at the coordinates of each source and use the constants C_X from Schlafly and Finkbeiner (2011) for Rv = 1, which are 4.239 for SDSSu, 3.303 for SDSSg, 2.285 for SDSSr and 3.924 for CaHK (Starkenburg et al., 2017), the reddening A_X in a filter X are then calculated to be:

$$A_X = C_X \cdot E(B-V)_{SFD}$$

The extinction corrected magnitudes are then:

$$m_{X,0} = m_X - A_X$$

D Blue horizontal branch star candidates in the Sagittarius stream

Gaia DR3 source ID	RA [deg]	DEC [deg]	distance ±0.2 [kpc]
1229515894074964480	210.94461564	12.5254136	10.59
1229582689406367872	210.71163557	12.8063017	4.07
1229767785316937088	210.27014109	12.7849732	9.01
1230229064804582912	213.40675516	14.4072918	15.18
1230432886772520704	211.6529724	14.5131572	13.70
1230551221711640832	212.22204556	15.1287753	13.41
1230590628036049152	212.55237717	15.5228743	10.29
1231241302696913792	211.23592516	15.3383626	10.74
1231430246898140160	211.84545691	16.0185210	6.83
1232013228580345344	214.36464463	16.1849573	9.28
1232048928347240832	212.95663841	15.6234396	5.15
1232784604705511168	215.54748435	17.7195573	8.61
1233177817551475456	212.04869897	16.7205993	10.45
1233433419645024256	213.92150088	17.1898138	9.57
1233630644543627264	214.77430303	18.3495022	8.12
1233675277843748224	214.34086243	18.3563984	7.94
1238906788528149888	215.94060288	18.678467	6.94
1238994642083847680	216.41928405	18.8188633	1.63

Table 3: A selection of BHB candidates from the sample employing synthetic Pristine CaHK and SDSS DR18 photometry (see Section 3.4 for details), which we located within the Sagittarius stream. For more details see Section 20.