

Mapping the Milky Way with Gaia Bp/Rp spectra

II: The inner stellar halo traced by a large sample of blue horizontal branch stars

Wenbo Wu^{1,3,4}, Xianhao Ye^{1,3,4}, C. Allende Prieto^{1,3,4}, Yuqin Chen^{1,2}, Xiang-Xiang Xue^{1,5}, Gang Zhao^{1,2}, Jingkun Zhao^{1,2}, David S. Aguado^{3,4}, Jonay I. González Hernández^{3,4}, Rafael Rebolo^{3,4}

¹ National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

² School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

³ Instituto de Astrofísica de Canarias, C/ Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain

⁴ Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain

⁵ Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206, People's Republic of China

Received ? xx, xxxx

ABSTRACT

Context. Due to their nearly constant absolute magnitudes and old ages, blue horizontal branch (BHB) stars are frequently used as standard candles to study the kinematics and structures of our galaxy. The number of identified BHB stars has significantly increased due to the advent of large scale surveys in the last two decades. Recently, Gaia DR3 was released including a catalog of around 220 million low-resolution spectra (Bp/Rp, or XP hereafter). These data have great potential for identifying many interesting stellar objects including BHB stars.

Aims. We construct a full-sky BHB catalog from Gaia Bp/Rp spectra and use it to explore the shape of the inner stellar halo.

Methods. We selected BHB stars based on synthetic photometry and stellar atmosphere parameters inferred from Gaia Bp/Rp spectra. We generated the synthetic SDSS broad-band ugr and Pristine narrow-band CaHK magnitudes from Gaia Bp/Rp data. A photometric selection of BHB candidates was made in the $(u - g, g - r)$ and $(u - \text{CaHK}, g - r)$ color-color spaces. A spectroscopic selection in $T_{\text{eff}} - \log g$ space was applied to remove stars with high surface gravity. The selection function of BHB stars was obtained by using the Gaia DR3 photometry. A non-parametric method that allows the variation in the vertical flattening q with the Galactic radius, was adopted to explore the density shape of the stellar halo.

Results. We present a catalog of 44,552 high latitude ($|b| > 20^\circ$) BHB candidates chosen with a well-characterized selection function. The stellar halo traced by these BHB stars is more flattened at smaller radii ($q = 0.4$ at $r \sim 8$ kpc), and becomes nearly spherical at larger radii ($q = 0.8$ at $r \sim 25$ kpc). Assuming a variable flattening and excluding several obvious outliers that might be related to the halo substructures or contaminants, we obtain a smooth and consistent relationship between r and q , and the density profile is best fit with by a single power law with an index $\alpha = -4.65 \pm 0.04$.

Key words. Stars: horizontal-branch-Galaxy: halo-Galaxy: stellar content-Galaxy: structure

1. Introduction

Blue horizontal branch stars are old ($\text{age} \geq 10$ Gyr), metal-poor ($[M/H] < -1$), and low-mass stars ($\sim 0.5 - 1 M_\odot$) that have evolved past the helium flash at the end of the red giant phase. BHB stars were first identified by a photometric study of the globular clusters (GCs) of M92 and M3 in the 1950s (Arp et al. 1952). These core-helium burning stars are photometrically A or B-types, and form a clear horizontal branch extending from the blue side of the instability strip to the location of the hot subdwarfs in the Hertzsprung–Russell (HR) diagram. Due to their nearly constant absolute magnitudes, BHB stars are frequently used as standard candles to trace distant Galactic structures and study the mass assembly history of the Milky Way: determining the distance and the age of the stellar populations in the Galactic halo and the GCs (e.g., Preston et al. 1991; Jofré & Weiss 2011; VandenBerg et al. 2013; Das et al. 2016; Subramaniam et al. 2017; Whitten et al. 2019), measuring the radial density profile and the velocity distribution of the stellar halo

(e.g., Sluis & Arnold 1998; Deason et al. 2011; Das et al. 2016; Thomas et al. 2018; Fukushima et al. 2018, 2019; Bird et al. 2021; Li & Binney 2022; Fukushima et al. 2025; Amarante et al. 2024; Yu et al. 2024), studying the kinematic substructures and the stellar streams (e.g., Xue et al. 2011; Ruhland et al. 2011; Belokurov & Koposov 2016; Deason et al. 2018; Yang et al. 2019b,a; Gryniewicz et al. 2021), and estimating the enclosed mass of the Milky Way (e.g., Xue et al. 2008; Kafle et al. 2012; Deason et al. 2021; Bird et al. 2022).

Photometrically, BHB stars reside in a narrow color range that facilitates their identification in specific color-color diagrams. Pier (1982) published a small catalog of A and B-type stars selected from the $(U - B, B - V)$ diagram. Sommer-Larsen et al. (1989) constructed a sample of 126 faint A-type stars (mainly BHB stars) satisfying $0 \leq (B - V)_0 \leq 0.2$ and $V \leq 19$. Sommer-Larsen et al. (1989) selected 4,175 candidates of field horizontal-branch and A-type stars using objective-prism/interference-filter data. However, BHB catalogs from these early studies suffer from heavy contamination of high surface gravity stars such as blue stragglers (BS) and A-type main sequence (MS) stars. In the last two decades, the study of

* Corresponding author, e-mail: carlos.allende.prieto@iac.es

** Corresponding author, e-mail: gzhao@nao.cas.cn

BHB stars has been largely advanced by large-scale sky surveys. Yanny et al. (2000) demonstrated that a filter cut using the dereddened color-color diagram ($u_0 - g_0, g_0 - r_0$) can separate the high surface gravity stars from the BHB stars to a certain extent in the Sloan Digital Sky Survey (SDSS; York et al. 2000). Smith et al. (2010) identified 27,074 probable BHB stars from SDSS Data Release 7 (DR7) photometric data through a support vector machine (SVM) method. Vickers et al. (2012) found that the SDSS z band photometry can also be used as a surface gravity indicator to distinguish BHB stars from contaminants. Beers et al. (2007) provided the coordinates and photometric information of 12,056 stars identified in the HK Survey as field horizontal-branch or A-type stars. Montenegro et al. (2019) detect 12,544 probable BHB stars in the Milky Way bulge-halo transition region by applying specific cuts in the $ZYJHK_s$ near-infrared bands from the Vía Láctea (VVV) ESO Public Survey data (Minniti et al. 2010). Starkenburg et al. (2019) find that a combination of the Pristine survey CaHK narrow-band photometry and the SDSS ugr broad-band photometry can achieve a high completeness (91%) and purity (93%) in the selection of BHB stars. Culpan et al. (2021) present a catalog of 57,377 BHB candidates selected from the color-magnitude diagram (CMD) of Gaia Early Data Release 3 (EDR3; Gaia Collaboration et al. 2021) with a purity of $\sim 70\%$. Their follow-up study selects 22,335 BHB stars from the CMD of Gaia DR3 and shows a contamination level of less than 12% by the synthetic spectral energy distribution (SED) (Culpan et al. 2024). Amarante et al. (2024) identify 95,466 possible BHB stars from the Legacy Survey photometry data (Dey et al. 2019) using a probabilistic methodology.

A photometric selection will inevitably involve some contamination due to uncertainties in the dust maps and the mixture of BHB and BS stars in color-color space. The most accurate way to construct a clean catalog of BHB stars is via spectral features. The Balmer-line profiles have proven to be an efficient indicator to distinguish between BHB and BS stars, since their depth (f_m) and broadening ($D_{0.2}$) can provide indirect information of T_{eff} and $\log g$ (Pier 1983). Clewley et al. (2002) developed a scale-width-shape method that discriminates the BHB stars from the contaminants based on a fitting of the Balmer-lines with a Sérsic profile. Through an analysis of the Balmer-line profiles and a color cut, previous studies selected 4,985 BHB stars from SDSS DR8 data (Xue et al. 2011) and 5,436 BHB stars from the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) DR5 data (Ju et al. 2024). As demonstrated in Barbosa et al. (2022), the BHB and MS/BS stars can be distinguished by applying a Gaussian Mixture model to their $\log g$ distributions.

Gaia Data Release 3 (DR3; Gaia Collaboration et al. 2023b) has released around 220 million low-resolution spectra obtained from the blue ($330 \leq \lambda \leq 680$ nm) and red ($640 \leq \lambda \leq 1050$ nm) Gaia slitless spectrophotometers (Carrasco et al. 2021; Montegriffo et al. 2023; De Angeli et al. 2023). These spectra (hereafter referred to as ‘XP spectra’) have a variable resolution ranging from 20 to 90 as a function of wavelength. Through spectral shape and synthetic photometry, previous studies have identified many interesting stellar objects such as white dwarfs (García-Zamora et al. 2023; Vincent et al. 2024), ultra-cool dwarfs (Cooper et al. 2024), carbon-enhanced metal-poor stars (Sanders & Matsunaga 2023), stars at the tip of the red giant branch (Li et al. 2023), and magnetic chemically peculiar stars (Paunzen & Prišegen 2022) from XP spectra.

In this study, we aim to construct a clean, full-sky catalog of BHB stars selected from XP spectra. This paper is organized as follows. In Section 2.1, we introduce the photometric and spectroscopic criteria for selecting BHB stars from the XP spectra. In Section 2.2 and 2.3, we present a clean BHB catalog with the heliocentric distance and the selection function provided. We analyze the density shape of the inner stellar halo traced by the obtained BHB catalog in Section 3. Finally, a summary is made in Section 4.

2. Construction of the BHB catalog

2.1. Identification

In this study, we will use synthetic photometry and stellar atmospheric parameters inferred from XP spectra to select probable BHB candidates. Gaia XP spectra are represented by a linear combination of Hermite basis functions (Montegriffo et al. 2023). We used the Python library GaiaXPpy¹, developed by the Gaia Data Processing and Analysis Consortium (DPAC), to transfer these coefficients into calibrated photometry and spectra.

2.1.1. Photometric cuts

GaiaXPpy allows the generation of synthetic photometry for multiple photometric systems from the input internally calibrated XP spectra (Gaia Collaboration et al. 2023a). Inspired by previous studies of Sirko et al. (2004) and Starkenburg et al. (2019), we used the synthetic SDSS and Pristine CaHK photometry to make a rough photometric selection of BHB stars. We chose the photometric system of SDSS_Std in the generation of ugr broad-band photometry because the procedure in GaiaXPpy benefits from a standardization that minimizes zero-point differences and/or trends as a function of colors in this system. The extinction coefficients ($R_u = 4.239, R_g = 3.303$, and $R_r = 2.285$) are provided by Schlafly & Finkbeiner (2011), and the CaHK extinction $R_{\text{CaHK}} = 3.918$ is obtained from Martin et al. (2024). The dust reddening $E(B-V)$ is provided by the two-dimensional dust map `dustmaps.sfd.SFDQuery` (Schlegel et al. 1998; Schlafly & Finkbeiner 2011; Green 2018).

We first applied a color cut $0.8 < u_{\text{std};0} - g_{\text{std};0} < 1.4$ and $-0.3 < g_{\text{std};0} - r_{\text{std};0} < 0.1$, and tagged stars satisfying this criterion as ugr -cut candidates. In this study we used the subscript ‘;0’ or ‘0’ to indicate dereddened stellar color or extinction corrected magnitude. This photometric selection effectively removes some undesired objects, such as white dwarfs, quasars, and other stars (Sirko et al. 2004; Vickers et al. 2012). However, there are still some contaminants remaining in these ugr -cut candidates, particularly BS stars.

The color-color space ($u_0 - \text{CaHK}_0, g_0 - r_0$) has been proven to efficiently separate BHB stars from BS stars by Starkenburg et al. (2019). However, our synthetic *CaHK* photometry is not a standardized system like SDSS_Std that may have some zero point offsets and trends as a function of color. Therefore, we need to carry out a test before applying this selection criterion with confidence. Among these ugr -cut candidates, 2,953 of them are previously confirmed as BHB stars by Barbosa et al. (2022), and 7,781 of them are possible high surface gravity stars with

¹ <https://gaia-dpci.github.io/GaiaXPpy-website/>

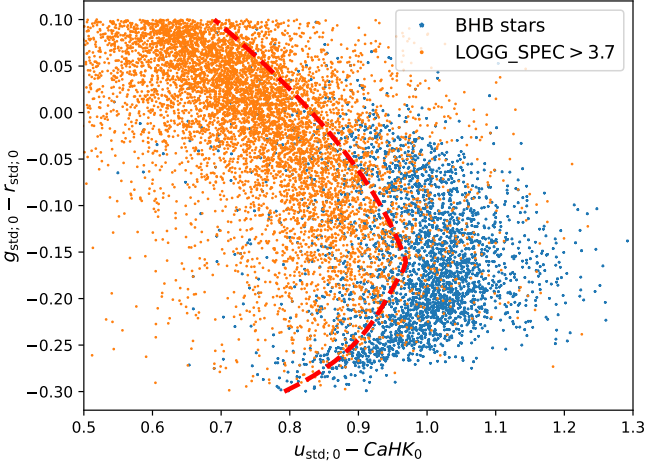


Fig. 1. $u_{\text{std},0} - \text{CaHK}_0$ versus $g_{\text{std},0} - r_{\text{std},0}$ for previously spectrally confirmed BHB stars (blue points) and high surface gravity stars (orange points). Although there are some mixtures of these two kinds of stars in this color-color diagram, we can still make a rough separation of them by the dashed red lines.

LOGG_SPEC > 3.7 in the Sloan Extension for Galactic Understanding and Exploration survey (SEGUE; Yanny et al. 2009). Figure 1 shows the distribution of these BHB and high surface gravity stars in $(u_{\text{std},0} - \text{CaHK}_0, g_{\text{std},0} - r_{\text{std},0})$ space. Although there is some overlap in this color-color diagram, we can see that these two kinds of stars are mainly located in two different regions. A dashed ridgeline $f(g_{\text{std},0} - r_{\text{std},0})$ was used to separate the BHB stars from the contaminants roughly. We applied this ridgeline to the whole *ugr*-cut sample and obtained 687,949 probable photometric-cut BHB candidates satisfying $0.8 < u_{\text{std},0} - g_{\text{std},0} < 1.4$, $-0.3 < g_{\text{std},0} - r_{\text{std},0} < 0.1$, and $u_{\text{std},0} - \text{CaHK}_0 > f(g_{\text{std},0} - r_{\text{std},0})$, where the ridge line is approximately defined as:

$$f(g_{\text{std},0} - r_{\text{std},0}) = 0.83 - 1.32(g_{\text{std},0} - r_{\text{std},0}) - 1.45(g_{\text{std},0} - r_{\text{std},0})^2 + 11.24(g_{\text{std},0} - r_{\text{std},0})^3. \quad (1)$$

2.1.2. Spectroscopic cuts

Although photometric selection can remove a large number of undesired objects, there are still some remaining contaminants as illustrated in Figure 1. To obtain a cleaner catalog, we fit the XP spectra of these photometric-cut BHB stars and inferred their stellar atmospheric parameters (T_{eff} , $\log g$, $[\text{M}/\text{H}]$).

Our sampled XP spectra have the same wavelength range of $360 \text{ nm} \leq \lambda \leq 990 \text{ nm}$ as Paper-I of Ye et al. (2025), and contain 330 points that are spaced evenly on a logarithmic scale. To remove the effect of interstellar extinction, we corrected the XP data with the two-dimensional dust map `dustmaps.sfd.SFDQuery` and the extinction curve of `ccm89` (Sommer-Larsen et al. 1989). The FORTRAN-90 code `FERRE`¹ (Allende Prieto et al. 2006), which matches the input spectra to best-fitting templates from a grid of synthetic spectra by χ^2 minimization, is used to infer stellar parameters from XP spectra

(T_{eff} , $\log g$, $[\text{M}/\text{H}]$). Since BHB stars are mainly in the temperature range of $7000 < T_{\text{eff}} < 12000 \text{ K}$, we adopted the `nsc3` library of Allende Prieto et al. (2018) as model spectra in the fitting. To match the low resolution of the XP spectra, we degraded the synthetic spectra to a constant resolving power of $R = 104$. Considering the lower temperature boundary of the `nsc3` library, we need to ensure that the input XP spectra belong to stars of high temperature. In Paper-I we fit each whole XP spectrum with a library of spectra covering the temperature range of $3500 < T_{\text{eff}} < 8000 \text{ K}$. We required that T_{eff} estimated from Paper-I is larger than 6900 K .

We only kept stars with $G > 11.5$ since Andrae et al. (2023) mentioned that the internal-calibration process of XP spectra may have some instrumental effects for stars brighter than $G = 11.5$. There is a high density of MS stars in the Magellanic clouds, potentially introducing significant contamination of MS stars to our selected BHB catalog. To correct for the effect of dust reddening in the XP spectra, we adopted $R_v = 3.1$ in the extinction law of `ccm89` for the Milky Way stars, which is not appropriate for stars in the Magellanic clouds. Therefore, following Culpán et al. (2021), objects were removed by excluding stars in the region of the Magellanic clouds where the local apparent stellar population density exceeds 50,000 stars per deg^2 : (1) the Large Magellanic Cloud at $274.5^\circ < l < 286.5^\circ$ and $-37.9^\circ < b < -27.9^\circ$. (2) the Small Magellanic Cloud at $299.8^\circ < l < 305.8^\circ$ and $-46.3^\circ < b < -42.3^\circ$.

After fitting, we defined the residual between the Normalized XP spectra (Flux_{XP}) and the fitting results ($\text{Flux}_{\text{fitting}}$) as ΔFlux . Our sampled XP spectrum consists of 330 data points that are spaced evenly on a logarithmic scale within the wavelength range. To remove stars with poor fittings, we required that the number of points with $|\Delta\text{Flux}| > 0.1$ is less than 60 in any given spectrum.

A metallicity cut $[\text{M}/\text{H}] \leq -1$ is frequently used to remove metal-rich disk stars in a BHB sample. However, metallicity determined from such low-resolution spectra is possibly unreliable for stars with low ($T_{\text{eff}} < 4000 \text{ K}$) and high ($T_{\text{eff}} > 7000 \text{ K}$) temperatures. Therefore, we chose an alternative cut in galactic latitude $|b| > 20^\circ$; 67,055 stars remain after the selection.

Previous studies usually distinguish between BHB and BS stars by the shape of their Balmer-lines (Xue et al. 2008, 2011; Ju et al. 2024). However, systematic errors that depend on the stellar color and magnitude are found in Gaia XP spectra (Montegriffo et al. 2023; Huang et al. 2024), which could directly distort the shape of spectral lines and introduce a large uncertainty in the fitting of Balmer-lines. Paper-I introduces a method to correct for the effects of systematic errors, but it is not valid for stars with high temperatures. Since many previous studies have already obtained a reliable estimation of the stellar parameters without the correction of systematic errors in the XP spectra, we will rely on our estimated parameters rather than the profile of Balmer-lines in the selection.

The top panel of Figure 2 shows the density distribution of these high-latitude photometric-cut BHB candidates in $T_{\text{eff}} - \log g$ diagram. We find that there are mainly three groups in Figure 2, of which the major group has the smallest $\log g$ and shows an increasing $T_{\text{eff}} - \log g$ relationship, and the other two minor groups are composed of stars with higher $\log g$. According to the trends of $\log g$ with T_{eff} , we suspect that stars in the major group are true BHB stars, while the other two minor groups are dwarfs or BS stars that are falsely identified as BHB stars in the pho-

¹ <https://github.com/callendeprieto/ferre>

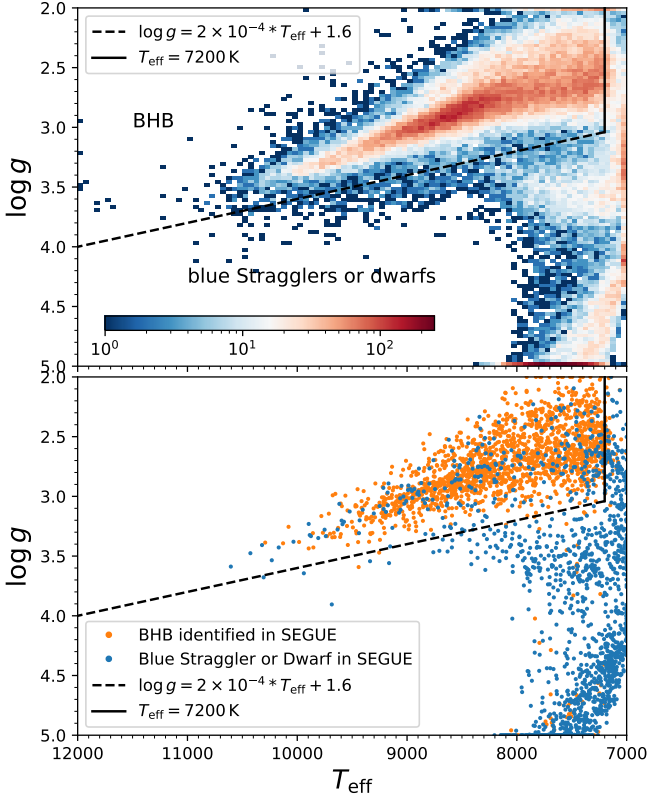


Fig. 2. Separation of BHB stars from high $\log g$ contaminants in the $T_{\text{eff}} - \log g$ space. The top panel shows the density distribution of the high-latitude photometric-cut BHB candidates in $T_{\text{eff}} - \log g$ space. The bottom panel shows a small fraction of our candidates that have been previously identified as BHB (blue points) or high surface gravity stars (orange points) in the SEGUE data by Barbosa et al. (2022). Two black lines are employed to separate these two types of stars, and we apply this cut to the whole candidates as shown in the top panel.

tometric selection. To verify our assumption, we cross-matched our candidates with the SEGUE survey and found 3,292 common stars. The condition in the bottom panel of Figure 2 is consistent with our assumption, as the majority of the previously spectrally confirmed BHB stars belong to the main group, while the other two groups are mainly composed of high surface gravity stars. A simple line of $\log g = 2 \times 10^{-4} \times T_{\text{eff}} + 1.6$ was employed to separate the major group from the other two minor ones. We further removed stars with $T_{\text{eff}} < 7200 \text{ K}$, as this selection line performs poorly at eliminating contaminants in the low-temperature regime. Although this separation is imperfect, as we can still see some overlap, it works well for most SEGUE stars and we applied it to all BHB candidates.

The selection criteria in this section are summarized as Equation 2:

$$\begin{aligned}
 &T_{\text{eff}}; Y_e \geq 6900 \text{ K} \\
 &\text{Stars in the region of the Magellanic Clouds are removed} \\
 &G > 11.5 \\
 &|b| > 20^\circ \\
 &12000 > T_{\text{eff}} > 7200 \text{ K} \\
 &\log g < 2 \times 10^{-4} \times T_{\text{eff}} + 1.6.
 \end{aligned} \tag{2}$$

We defined these 44,552 selected stars as XP spectrally confirmed BHB candidates.

BHB stars projected within six half-mass radii from the center of known globular clusters within 25 kpc of the Sun were selected as potential globular cluster members. Parameters of the globular clusters were obtained from Baumgardt & Vasiliev (2021) and Vasiliev & Baumgardt (2021). Our BHB catalog is composed of 44,032 field BHB stars and 520 potential globular cluster members.

2.2. Distance estimation

Although BHB stars are known to have almost constant absolute magnitudes, there are still some variations as a function of color. Deason et al. (2011) fit the absolute magnitude M_g of BHB stars as a function of $g - r$ color with a quartic polynomial. Gryniewicz et al. (2021) extend the fitting to the blue tail of the Horizontal Branch stars of $g - r < -0.25$. Due to the potential uncertainties in the synthetic *ugr* photometry, we will rely on the Gaia broad band photometry to establish a relationship between the absolute magnitude M_G and dereddened stellar color $G_{\text{BP},0} - G_{\text{RP},0}$. The extinction coefficients of Gaia Broad band photometry ($R_G = 2.364, R_{G_{\text{BP}}} = 2.998, R_{G_{\text{RP}}} = 1.737$) are provided by Zhang & Yuan (2023). The absolute magnitude M_G was obtained from $M_G = G_0 - 5 \log_{10}(d_{\text{parallax}}) + 5$, where distance d_{parallax} is given in units of parsecs (pc) and derived from the Gaia DR3 parallax for 6,065 BHB stars with a reliable parallax ($\text{parallax_over_error} > 10$). We took into account the parallax zero point offset of -0.017 mas as discussed by Lindegren et al. (2021). These BHB stars exhibit a clear horizontal branch structure in the $(G_{\text{BP},0} - G_{\text{RP},0}, M_G)$ diagram as is shown in Figure 3. Three dashed red curves that represent the P_{16}, P_{50} , and P_{84} percentiles distributions were obtained after the removal of possible contaminants. We applied a quartic polynomial fitting to the three curves as Equation 3:

$$\begin{aligned}
 M_G(P_{84}) &= 0.68 - 1.11(G_{\text{BP},0} - G_{\text{RP},0}) - 26.49(G_{\text{BP},0} - G_{\text{RP},0})^2 \\
 &\quad + 182.44(G_{\text{BP},0} - G_{\text{RP},0})^3 - 326.44(G_{\text{BP},0} - G_{\text{RP},0})^4 \\
 M_G(P_{50}) &= 0.93 - 3.12(G_{\text{BP},0} - G_{\text{RP},0}) + 6.61(G_{\text{BP},0} - G_{\text{RP},0})^2 \\
 &\quad + 9.74(G_{\text{BP},0} - G_{\text{RP},0})^3 - 45.67(G_{\text{BP},0} - G_{\text{RP},0})^4 \\
 M_G(P_{16}) &= 1.17 - 3.28(G_{\text{BP},0} - G_{\text{RP},0}) + 16.95(G_{\text{BP},0} - G_{\text{RP},0})^2 \\
 &\quad - 86.45(G_{\text{BP},0} - G_{\text{RP},0})^3 + 181.85(G_{\text{BP},0} - G_{\text{RP},0})^4.
 \end{aligned} \tag{3}$$

The heliocentric distance, d_{helio} , was calculated by substituting $M_G(P_{50})$ into $d_{\text{helio}} = 10^{0.2(G_0 - M_G + 5)}$.

As is shown in Figure 4, our obtained d_{helio} is generally in line with d_{parallax} . Figure 5 shows the normalized apparent magnitude distribution of our selected BHB sample, with G magnitudes ranging from 11.5 to approximately 17.5, peaking at around 15.5. Due to the limiting magnitude of the XP spectra ($G \sim 17.65$), d_{helio} of these BHB candidates is within 25 kpc, indicating that most of them are located in the inner halo (at a Galactocentric radius $r_{\text{gc}} < 30 \text{ kpc}$).

2.3. Completeness and purity

To check the completeness and purity of our sample, we used the SEGUE spectrally confirmed BHB catalog of Barbosa et al. (2022) as the reference library. Many previous studies selected BHB stars from the SEGUE survey using different methods.

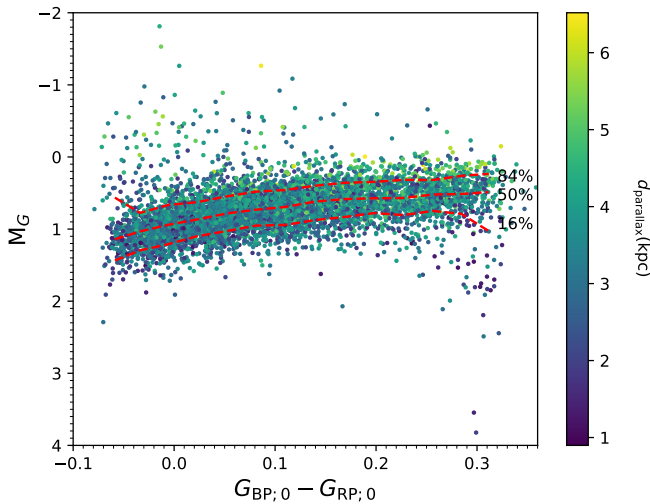


Fig. 3. Gaia DR3 CMD of 6,065 selected BHB stars that have a reliable parallax. These BHB stars display a clear horizontal branch structure, and three dashed red curves that represent the P_{16} , P_{50} , and P_{84} percentiles distributions of M_G are obtained.

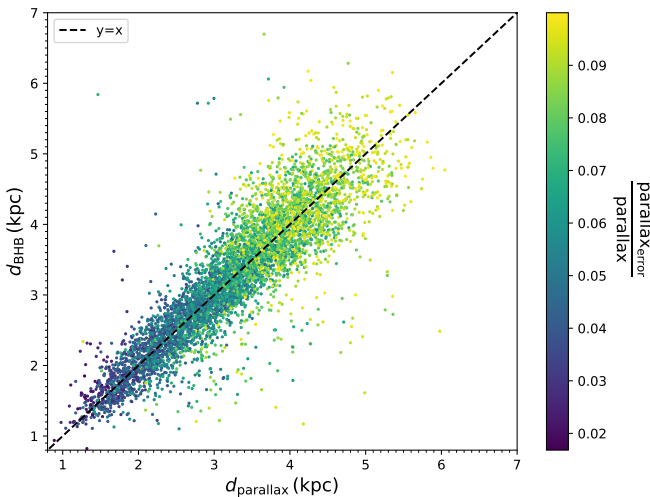


Fig. 4. Comparison of the distance calculated from the parallax ($d_{\text{parallelax}}$) and the theoretical absolute magnitude (d_{helio}) for BHB stars that have a reliable measurement of the parallax. An increasing dispersion with the increasing parallax error is expected.

Here we adopted the BHB sample of Barbosa et al. (2022) because their selection is also mainly based on the $\log g$ distribution like us. Starkenburg et al. (2019) propose two different ways to estimate the completeness of their photometrically confirmed BHB sample selected from the Pristine survey. One approach consists of a direct comparison to the SEGUE spectrally confirmed BHB catalog, while the other is based on a test of Monte Carlo simulated mock data sets. The completeness estimated from the two methods is in good agreement for stars brighter than $g = 19$. Although the SEGUE spectrally confirmed BHB catalog may not be perfectly complete, this consistency suggests that its completeness, $C(\text{SEGUE})$, is likely close to 1 and does not vary significantly with the magnitude for stars of $g < 19$. The limiting magnitude of Gaia XP spectra is around $G = 17.65$, which is much brighter than $g = 19$ according to the empirical transformation equations derived by Evans et al. (2018).

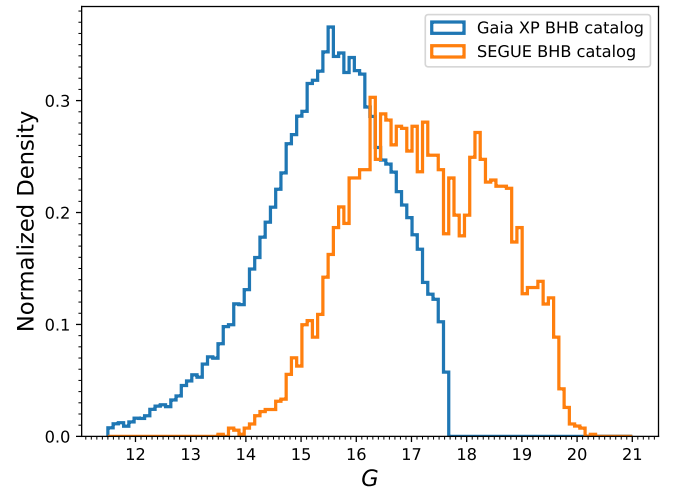


Fig. 5. Normalized apparent magnitude distribution in G band for XP spectrally confirmed BHB sample (blue hist) and SEGUE spectrally confirmed BHB sample of Barbosa et al. (2022) (orange hist). Our selected BHB sample is more focused on the inner stellar halo than the SEGUE BHB sample due to the limiting magnitude of Gaia XP spectra.

To explore the completeness and purity of our sample, we constructed a combined catalog by cross-matching the Gaia XP spectra with the SEGUE survey. Within this catalog, 1,906 stars are identified as BHB candidates by Gaia XP spectra in this study. Among them, 1,815 candidates are confirmed to be real low $\log g$ stars with $\text{LOGG_SPEC} < 3.7$ according to the SEGUE data, which suggests an overall purity of around $1,815/1,906 \sim 95\%$ in this subsample. Cross-matching the SEGUE BHB sample of Barbosa et al. (2022) and the combined catalog yields 3,051 common sources. Among them, 1,643 are also identified as BHB stars by Gaia XP spectra in this study, while the rest are excluded by our photometric and spectroscopic cuts. We divided the 1,815 real low $\log g$ candidates into 40 bins of color and magnitude, and the bin edges were selected to ensure that each bin contains a similar number of around 45 stars. For each each bin, the top panel of Figure 6 shows the ratio of BHB stars found from the same combined catalog by us (the 1,815 real low $\log g$ candidates) and by Barbosa et al. (2022) (the subsample containing 3,051 common sources). The completeness is defined as the multiplicity of this ratio and $C(\text{SEGUE})$. The purity is defined as the proportion of real low $\log g$ stars in the 1,903 BHB candidates and shown in the middle panel of Figure 6. We can see that the completeness declines quickly as stars become fainter, while for the purity it shows little dependence on the magnitude and fluctuates randomly from 0.85 to 1.

Ju et al. (2024) propose a different method to estimate the purity of a BHB sample by counting the contribution of stars with $[\text{M}/\text{H}] < -1$. Following their method, we cross-matched the XP spectrally confirmed BHB catalog with the LAMOST OBA-type stars catalog of Xiang et al. (2022) and obtained 4,332 common sources. The LAMOST OBA-type stars catalog spans a broader range of magnitudes than the SEGUE BHB catalog, with G spanning approximately from 4.5 to 21.5. Among these common sources, 3,837 stars are classified as metal-poor with $[\text{Fe}/\text{H}]_{\text{LAMOST}} < -1$ by Xiang et al. (2022), which also suggests a similar overall purity of $3,837/4,332 \sim 89\%$. We also divided these common sources into different bins to investigate the dependence of the purity on stellar color and magnitude. As shown in the bottom panel of Figure 6, the purity does not exhibit sig-

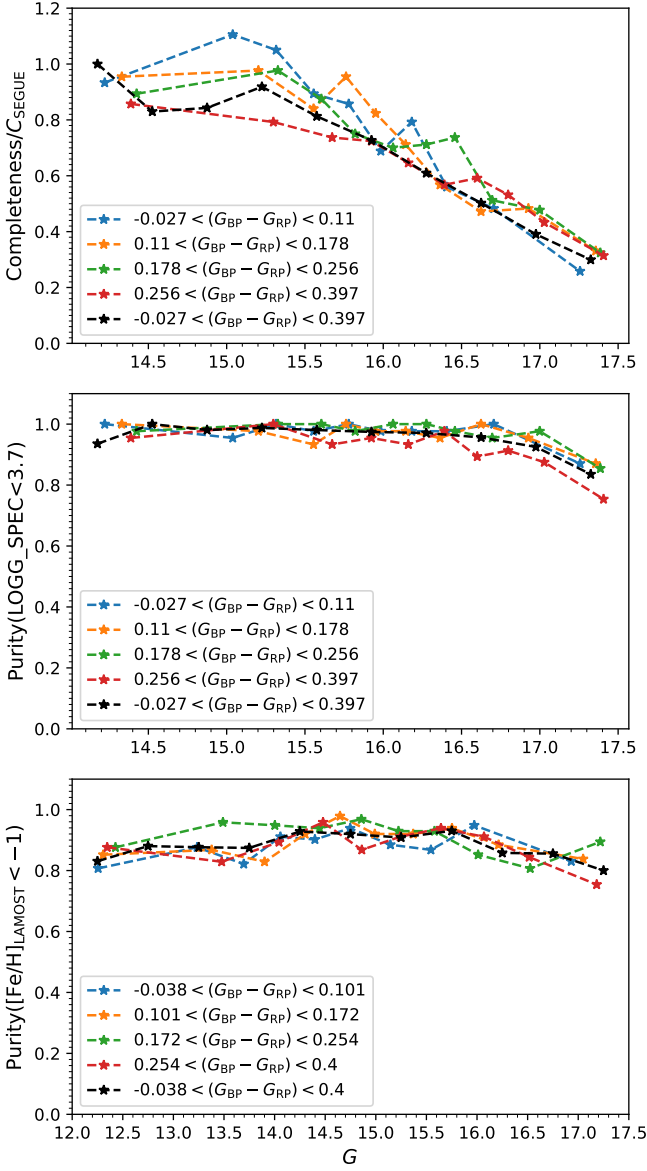


Fig. 6. Completeness and purity of the overlapping BHB sample as a function of G for BHB stars of different stellar colors. The top panel shows the ratio of BHB stars found from the same combined catalog (Gaia XP spectra \times SEGUE survey) by us and by Barbosa et al. (2022). The middle panel presents the purity defined as the proportion of low $\log g$ stars in our 1,903 BHB candidates that have a measurement of LOGG_SPEC in the SEGUE survey. The bottom panel displays the purity defined as the proportion of metal-poor stars in our 4,332 BHB candidates that have a measurement of $[\text{Fe}/\text{H}]_{\text{LAMOST}}$ in the LAMOST OBA-type stars catalog. The dashed black lines describe the completeness and purity as a function of G without the consideration of stellar color. The completeness declines rapidly as stars become fainter, while for the purity it remains almost constant with minimal variation.

nificant variation with $G_{BP} - G_{RP}$ or G , which is very similar to the purity defined by the proportion of low $\log g$ stars in the middle panel.

2.4. Selection function

The selection function of the BHB sample is defined as Eq 4:

$$S_{\text{BHB}} = \frac{n_{\text{BHB};\text{sub}}(l, b, c, m)}{n_{\text{BHB};\text{ph}}(l, b, c, m)} = \frac{n_{\text{BHB};\text{XP}}(l, b, c, m)}{n_{\text{BHB};\text{ph}}(l, b, c, m)} \times \frac{n_{\text{BHB};\text{sub}}(l, b, c, m)}{n_{\text{BHB};\text{XP}}(l, b, c, m)}, \quad (4)$$

where $n_{\text{BHB};\text{sub}}(l, b, c, m)$, $n_{\text{BHB};\text{XP}}(l, b, c, m)$, and $n_{\text{BHB};\text{ph}}(l, b, c, m)$ are BHB star counts in our sample, Gaia XP spectra, and the complete photometric data in the color-magnitude diagram, respectively. We used the subscript ‘sub’ to indicate that our BHB sample is only a subsample of BHB stars in the XP spectra. The first factor of S_{BHB} can be written as follows:

$$\frac{n_{\text{BHB};\text{XP}}(l, b, c, m)}{n_{\text{BHB};\text{ph}}(l, b, c, m)} = \frac{n_{\text{XP}}(l, b, c, m) \times r_{\text{BHB};\text{XP}}(l, b, c, m)}{n_{\text{ph}}(l, b, c, m) \times r_{\text{BHB};\text{ph}}(l, b, c, m)} = S_{\text{XP}}(l, b, c, m) \times \frac{r_{\text{BHB};\text{XP}}(l, b, c, m)}{r_{\text{BHB};\text{ph}}(l, b, c, m)}, \quad (5)$$

where $n_{\text{XP}}(l, b, c, m)$ and $n_{\text{ph}}(l, b, c, m)$ are star counts in Gaia XP spectra and the complete photometric survey in the color-magnitude diagram. For stars in a certain (l, b, c, m) bin, they might be BHB stars or some other kinds of stars like nearby Blue Stragglers. We defined $r_{\text{BHB};\text{XP}}(l, b, c, m)$ and $r_{\text{BHB};\text{ph}}(l, b, c, m)$ as the ratio of BHB stars to all stellar objects in Gaia XP spectra and a complete photometric survey. We assumed that the sample of Gaia XP spectra survey does not favor one type of star over another in a certain (l, b, c, m) bin, which means $r_{\text{BHB};\text{XP}}(l, b, c, m) \approx r_{\text{BHB};\text{ph}}(l, b, c, m)$. We adopted the Gaia DR3 photometric survey as the reference since it is almost complete for stars of $G < 19$ (Cantat-Gaudin et al. 2023). Using the Python package *gaiaunlimited* (Castro-Ginard et al. 2023), we generated the selection function $S_{\text{XP}}(l, b, c, m)$ at the resolution of the HEALPix¹ level 5 in the Sky positions (12,288 equally sized areas, referred to as the 12,288 base pixels), $G \in [3, 20]$ in steps of 0.2, and $G - G_{\text{RP}} \in [-0.05, 0.45]$ in bins of 0.25.

The second factor $n_{\text{BHB};\text{sub}}(l, b, c, m)/n_{\text{BHB};\text{XP}}(l, b, c, m)$ is the probability P_{BHB} that a BHB star can be correctly selected from XP spectra. $n_{\text{BHB};\text{XP}}$ is an unknown component because we cannot identify all BHB stars from XP spectra. However, a fraction of them have been identified by the SEGUE data, which are the common 3,051 sources mentioned above, and P_{BHB} estimated from this subsample is used as an approximation of the second factor. The small number of common sources prevents us from exploring P_{BHB} for different sky positions. Assuming that P_{BHB} is irrelevant to the positions in the sky, it corresponds only to the completeness mentioned above in Figure 6. We ignored the effect of stellar color since it is only a secondary determinant of the completeness. We assumed $C(\text{SEGUE})$ as a constant since it should not vary significantly with magnitude. For stars of $G \geq 14$, the probability P_{BHB} was obtained by performing a quadratic 1-D interpolation to the dashed black line in the top panel of Figure 6. The completeness increases as BHB stars become brighter and approaches 1 for stars of $G = 14.2$. Therefore, for stars brighter than $G = 14.2$ we assumed a probability

¹ HEALPix, an acronym for Hierarchical Equal Area isoLatitude Pixelization of a sphere, is an algorithm that describes the Sky positions (l, b) . Stars with similar coordinates (l, b) have the same HEALPix index number. Further information about HEALPix is available in the Gaia archive documentation.

$P_{\text{BHB}} = 1$. The selection function S_{BHB} is defined as:

$$S_{\text{BHB}}(l, b, G - G_{\text{RP}}, G) = S_{\text{XP}}(l, b, G - G_{\text{RP}}, G)P_{\text{BHB}}(G), \quad (6)$$

where c is the stellar color of $G - G_{\text{RP}}$, and m is the apparent magnitude of G . The selection function is also provided as a parameter in our catalog.

3. Density shape of the inner stellar halo

Understanding the spatial distribution of halo stars, specifically the radial density profile and the vertical flattening, is critical to disentangling the mass assembly history of the Milky Way. The radial density profile of the stellar halo is typically described by a broken power law (BPL) with a constant flattening $q \sim 0.7$ in previous studies (e.g., Watkins et al. 2009; Deason et al. 2011; Faccioli et al. 2014; Xue et al. 2015; Pila-Díez et al. 2015; Das et al. 2016; Medina et al. 2024). A break radius r_{break} has been identified between 15 and 30 kpc. The radial density profile inside r_{break} follows a power law with an index of about 2.5 – 3, while stars outside r_{break} follow a much steeper one with an index of approximately 4 – 5. Some studies suggest that there might be two break radii at the inner ($r_{\text{break}} \sim 10 - 15$ kpc) and outer ($r_{\text{break}} \sim 20 - 30$) parts of the stellar halo (Yang et al. 2022; Han et al. 2022). Recently, several studies find that r_{break} is not necessarily needed assuming a varying flattening, in which the radial density profile can be described by a single power law (SPL) with an index of 4 – 5 using an oblate ellipsoid model (Das et al. 2016; Hernitschek et al. 2018; Xu et al. 2018; Wu et al. 2022b; Chen et al. 2023) or an index of 2.96 using a triaxial ellipsoid model (Iorio et al. 2018).

Liu et al. (2017) and Xu et al. (2018) describe a method to derive the stellar density profile along each line of sight (l – b plate) for the LAMOST spectroscopic survey. Using their method, we can explore the radial density profile of the stellar halo with varying flattening through our identified BHB sample. In their works, the ratio between the stellar density obtained from the photometric (ν_{ph}) and spectroscopic (ν_{sp}) surveys is defined as:

$$\nu_{\text{ph}}(D|l, b, c, m) = \nu_{\text{sp}}(D|l, b, c, m)S^{-1}(l, b, c, m), \quad (7)$$

where D is the distance along a given line of sight, and S is the selection function of BHB stars (S_{BHB}) in this study.

After integrating over c and m , the stellar density profile for a given line of sight is defined as:

$$\nu_{\text{ph}}(D|l, b) = \iint \nu_{\text{sp}}(D|l, b, c, m)S^{-1}(l, b, c, m)dc dm. \quad (8)$$

To account for errors arising from distance, we employed a kernel density estimation (KDE) method to derive ν_{sp} along a given line of sight. The contribution of a star i is treated as a probability density function p_i extending along the corresponding line of sight, and p_i is defined as:

$$p_i(D) = \frac{\mathcal{N}(D|D_i, \sigma_{D_i}^2)}{\int_{D_{\min}}^{D_{\max}} \mathcal{N}(D_x|D_i, \sigma_{D_i}^2)dD_x}, \quad (9)$$

where \mathcal{N} is a normal function. D_i is the estimated distance of star i , and σ_{D_i} is the uncertainty of the estimated distance. In this study, σ_{D_i} is defined as:

$$\sigma_{D_i} = \frac{D_i(84\%) - D_i(16\%)}{2} \quad (10)$$

where $D_i(16\%)$ and $D_i(84\%)$ are the 16% and 84% distributions of d_{helio} obtained by substituting $M_G(P_{16})$ and $M_G(P_{84})$ into Equation 3. We set $D_{\min} = 0$ kpc and $D_{\max} = 200$ kpc.

We took into account the contribution of all stars around c and m . The density profile ν_{sp} is defined as:

$$\nu_{\text{sp}}(D|l, b, c, m) = \frac{1}{\Omega D^2} \sum_i^{n_{\text{sp}}(l, b, c, m)} p_i(D), \quad (11)$$

where Ω is the solid angle of the given line of sight, and $n_{\text{sp}}(l, b, c, m)$ is the number of spectroscopic stars within a given (l, b, c, m) bin. Note that in this form, the derived ν_{ph} is a continuous function of D .

In the LAMOST survey, stars are observed in thousands of plates (lines of sight) with a constant solid angle. To achieve a similar situation, we divided the sky positions into 3,072 plates at the resolution of the HEALPix level 4, and BHB stars (potential globular cluster members are removed) were distributed in different lines of sight according to their Galactic coordinates. Since Ω , dc , dm , and $C(\text{SEGUE})$ are constant values that have no impact on the obtained density shape, we normalized them to 1 for computational convenience. ν_{ph} is obtained by combing Equation 8 and 11. We only adopted ν_{ph} at the positions where we can find a star, since the stellar density derived at the positions without sampled stars may suffer from large uncertainties. In other words, every BHB star in our sample is assigned a ν_{ph} that represents an estimate of the stellar density at its position.

Figure 7 shows the median value of ν_{ph} in the $|R|$ – Z pixel map. R and Z are cylindrical Galactocentric coordinates obtained from the conventions in astropy (Astropy Collaboration et al. 2018). We used the default values of the Solar Galactocentric distance $r_{\text{gc}, \odot} = 8.122$ kpc (GRAVITY Collaboration et al. 2018), and height $Z_{\odot} = 20.8$ pc (Bennett & Bovy 2019). We can clearly see a vertically flattened shape of the stellar halo that shows a larger flattening at a smaller radius. To characterize the vertical flattening quantitatively, we fit the density profile as a function of the flattened Galactic radius r with an ellipsoid model as follows:

$$\nu(r) = \nu_0 r^{-\alpha}, \quad (12)$$

$$r = \sqrt{R^2 + (Z/q(r))^2}, \quad (13)$$

where $q(r)$ is the flattening parameter.

As noted by Xu et al. (2018), constraining r and q is challenging in the Galactocentric cylindrical coordinate because q is very sensitive to the slope of the iso-density surface, which becomes too steep at low latitudes. Therefore, following Xu et al. (2018) and Wu et al. (2022a), we will conduct our analysis in the Galactocentric polar coordinates where R and Z can be defined as follows:

$$R = r \cos(\eta), \quad (14)$$

$$Z = r q \sin(\eta), \quad (15)$$

$$r_{\text{gc}} = \sqrt{R^2 + Z^2} \quad (16)$$

$$\sin(\theta) = Z/r_{\text{gc}}. \quad (17)$$

Combining Equation 14, 15, and 16, we rewrite r_{gc} as:

$$r_{\text{gc}} = r \sqrt{1 + (q^2 - 1)\sin^2(\eta)}. \quad (18)$$

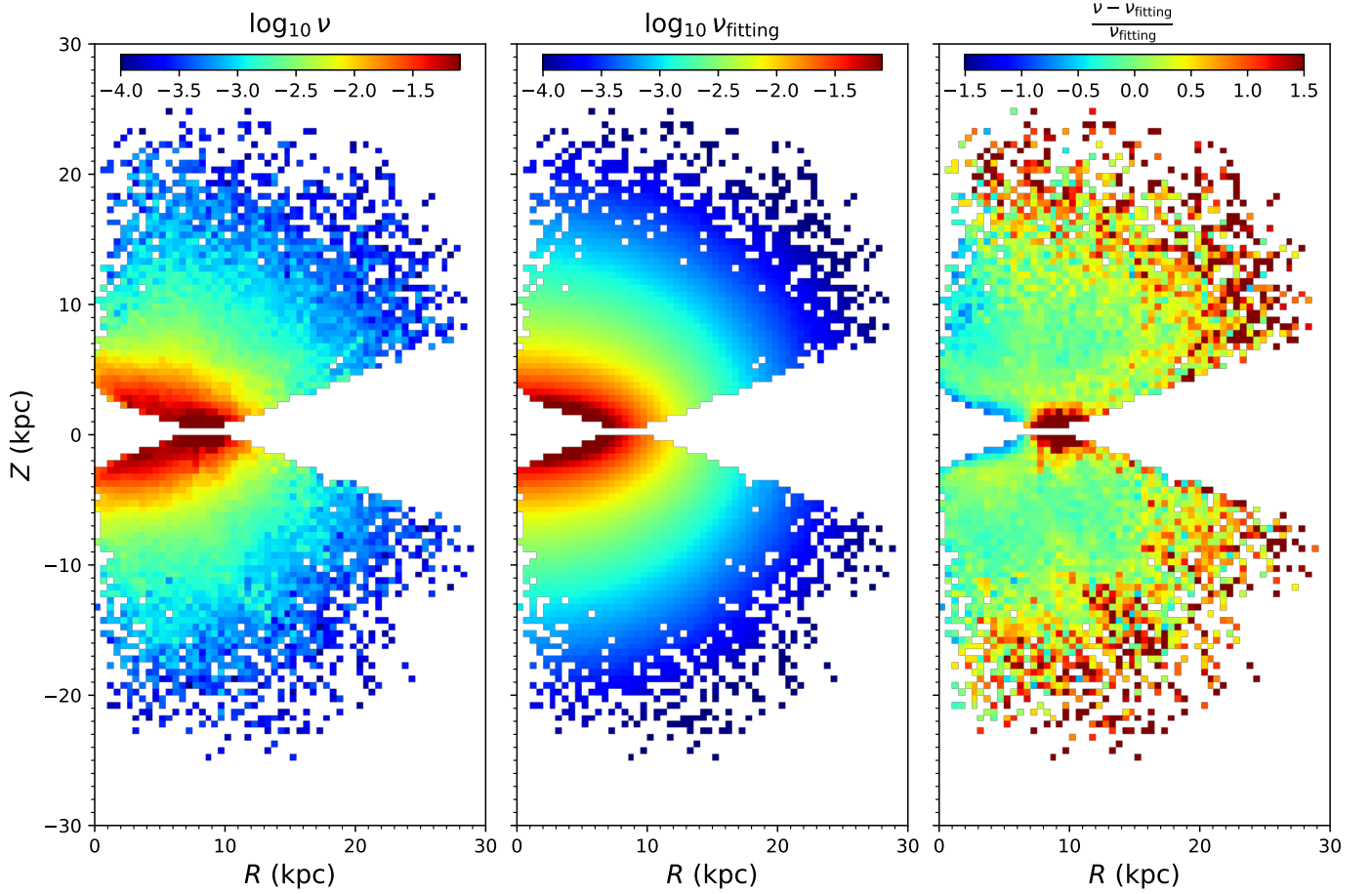


Fig. 7. Number density map of the stellar halo in $R - Z$ space for the BHB data ($\log_{10} \nu$, left panel) and the model ($\log_{10} \nu_{\text{fitting}}$, middle panel). The pixel size is 0.5×0.5 (kpc). The color in each pixel represents the mean density on a logarithmic scale. The right panel shows the residual map $((\nu - \nu_{\text{fitting}})/\nu_{\text{fitting}})$, where the most prominent difference is seen in the solar neighborhood and the outer region ($r_{\text{gc}} > 15$ kpc).

Using Equation 17, we write $\sin(\eta)$ as:

$$\sin(\eta) = \frac{r_{\text{gc}} \sin(\theta)}{rq}. \quad (19)$$

Substituting Equation 19 into Equation 18, we can fit the iso-density contour in $r_{\text{gc}} - \sin(\theta)$ space with Equation 20,

$$r_{\text{gc}} = rq \sqrt{\frac{1}{q^2 - (q^2 - 1) \sin^2(\theta)}}. \quad (20)$$

We divided the BHB stars into different bins of $\log_{10}(\nu)$ with a bin size of 0.1 in the range $-1.2 < \log_{10}(\nu_{\text{ph}}) < -3.8$. We fit the spatial distribution ($r_{\text{gc}}, \sin(\theta)$) of the BHB stars with an ellipse of Equation 20 in each $\log_{10}(\nu)$ bin by `scipy.optimize.curve_fit`. Our fitting results for all bins are shown in Figure 8 and the specific values are recorded in Table 1. The variation in q with increasing r shows a clear tendency that the stellar halo becomes more vertically flattened in the inner part. In Figure 9, we find that the radial density profile of the stellar halo can be simply described by a SPL of an index of $\alpha = -4.80 \pm 0.05$ with varying flattening. Our result is almost the same as the radial density profile ($\alpha = -5.03 \pm 0.64$) derived from the LAMOST K-giant sample in Xu et al. (2018), but has a smaller dispersion that might be related to the larger dataset and the accurate distance estimation of the BHB sample.

The number density map (ν) and the model (ν_{fitting}) fit are shown in Figure 7. In the spatial positions of each BHB star, ν_{fitting} was generated based on parameters derived from the binned data in Figure 8 and 9. We applied a second-order polynomial function ($q = f(r)$) to describe the $r - q$ relationship and q is extrapolated as a constant of 0.81 beyond $r_q = 23$ kpc. We can see that the number density map of the BHB sample is generally consistent with the model except for some prominent overdensities in the solar neighborhood and the outer region ($r_{\text{gc}} > 15$ kpc). From Figure 10, we find that the density distribution of $\frac{\nu - \nu_{\text{fitting}}}{\nu_{\text{fitting}}}$ can be well described by a Gaussian mixture model of two components, of which the primary is centered at -0.12 with a standard deviation of 0.30, and the minor is centered at around 0.53 with a standard deviation of 0.50. BHB stars in the minor component shows an underestimated ν_{fitting} , which is likely to be related to the prominent overdensities found in the residual map.

Several contributing factors may account for the presence of the minor component. The selection function correction may not be suitable for all BHB stars due to the assumptions made in section 2.3. The error in the distance estimation (σ_{D_i}) increases for more distant BHB stars, potentially leading to an increased scatter relative to the fitting line for stars in the outer halo bins of Figure 7. Considering the existence of multiple stellar substructures in the Galactic halo, it is natural that a fraction of our selected BHB stars belong to some unresolved substructures. The pres-

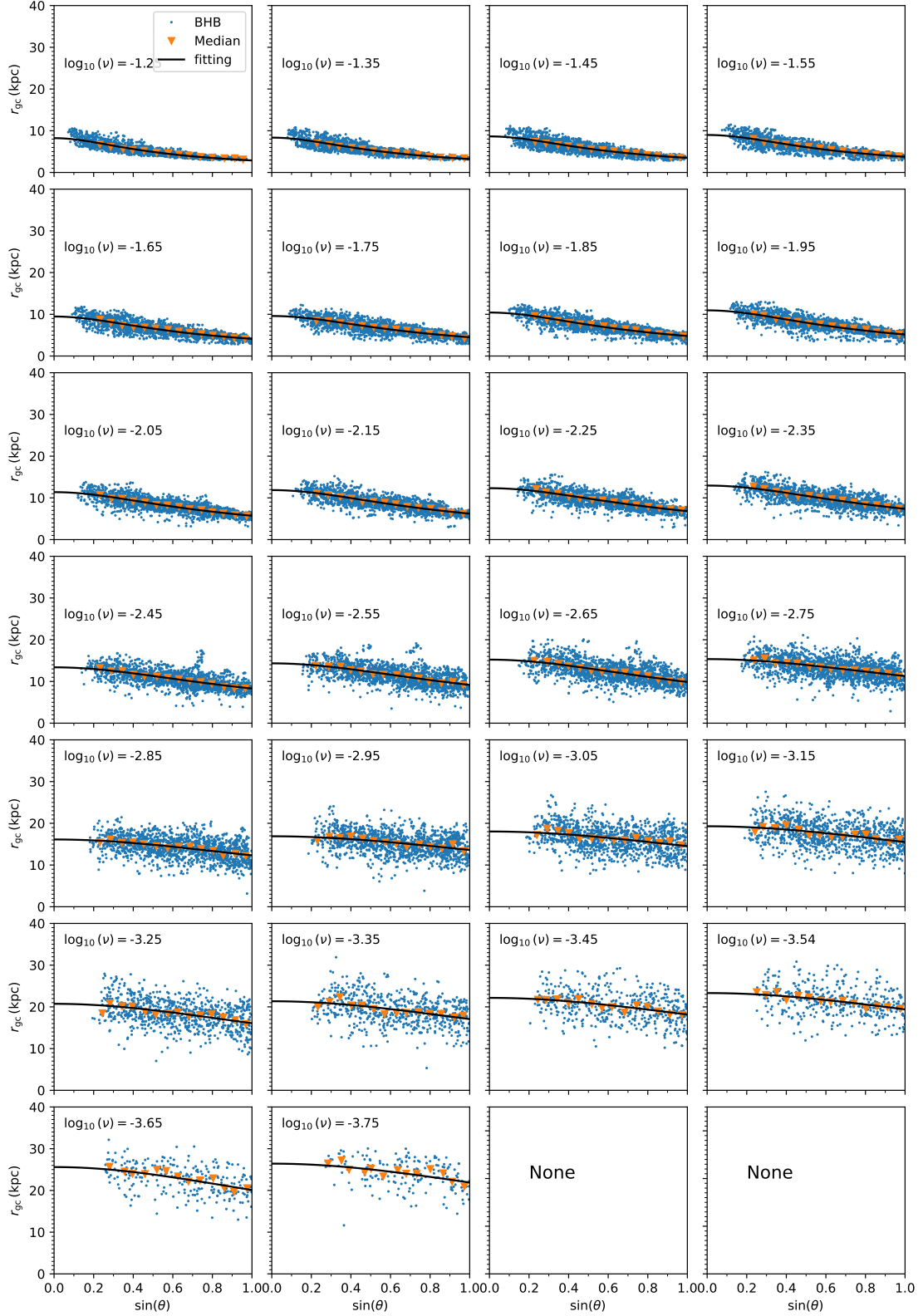


Fig. 8. Fitting results of all BHB stars in different bins of $\log_{10}(\nu)$. The blue points are BHB stars in a certain bin and the orange inverted triangles represent the median value of r_{gc} . The black lines are the best fitting results with the specific parameters recorded in Table 1.

ence of non-axisymmetric substructures may introduce a bias to the fitting, since the stellar halo is viewed as a smooth component in our model. The inclusion of contaminants can cause an overestimation of ν , and may explain the most prominent over-density in the solar neighborhood considering the large numbers

of metal-rich stars in this region. Recent results from nearby stars suggest an early formation of the disk component within the first few billion years (Belokurov & Kravtsov 2022; Conroy et al. 2022; Zhang et al. 2024), and a disk-like geometry could exist even for a very old stellar population of ~ 13 Gyr (Xiang et al.

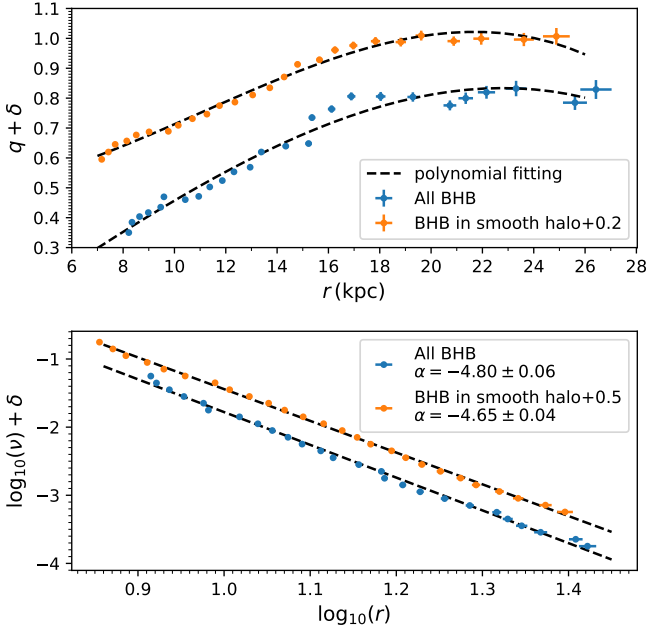


Fig. 9. Density shape and profile of the stellar halo traced by the BHB sample. The top panel shows the variation in q with increasing r for all BHB stars (blue points) and the subsample (orange points). To better distinguish the results of the two samples, a constant of 0.2 is added to q of the subsample. The stellar halo is largely vertically flattened in the inner part and transformed into a nearly spherical shape at the outer radii. The bottom panel shows the relationship between the radial stellar density $\log(\nu)$ and the flattened radius r . Assuming a variable flattening, the radial density profile of the stellar halo derived from all BHB stars (blue points) can be well described by a SPL with an index $\alpha = -4.80 \pm 0.06$, and for the subsample (orange points, a constant of 0.5 is added to $\log(\nu)$) the best result is $\alpha = -4.65 \pm 0.04$.

2025). It is hard for us to remove all disk stars due to the lack of metallicity and kinematic information. Therefore, it is possible that a small fraction of our identified BHB stars with low $|Z|$ actually belong to the old disk.

To reduce the influence of the minor component, we applied a rough cut of $\frac{\nu - \nu_{\text{fitting}}}{\nu_{\text{fitting}}} < 0.6$ and defined the remaining stars as the subsample of the smooth stellar halo. We then repeated the fitting process and presented the results for two representative bins in Figure 11. As is shown in Figure 9, the main tendency that the stellar halo is more vertically flattened in the inner radii remains unchanged for the subsample, but the $r - q$ relationship is now better described by a second-order polynomial function without evident outliers at $r \sim 17$ kpc. Overall, the increasing trend of q with r is smoother and more consistent for the subsample. The density profile exhibits only a slight change and is well characterized by a SPL of an index of $\alpha = -4.65 \pm 0.04$. For each BHB star in the subsample, we generate its ν_{fitting} using the updated $r - q$ relationship and density profile, where q is extrapolated as a constant of 0.80 beyond $r_q = 23$ kpc. In Figure 10, the density distribution of $\frac{\nu - \nu_{\text{fitting}}}{\nu_{\text{fitting}}}$ for the smooth halo is well fit by a single Gaussian function, with a mean of 0.05 and a standard deviation of 0.38. A significant improvement in consistency between the BHB data and the model is clearly shown in the residual map of Figure 12. The most prominent overdensity in the solar neighborhood disappears, and the discrepancies in the outer part of the stellar halo are also largely improved.

Table 1. Best estimated parameters of the stellar halo shape derived from the BHB catalog.

$\log_{10}(\nu)$	N^a	r^a	q^a	N^b	r^b	q^b
-1.25	1,701	8.22	0.35	1,220	7.17	0.40
-1.35	2,023	8.35	0.38	1,527	7.43	0.42
-1.45	2,561	8.64	0.40	2,078	7.69	0.45
-1.55	2,500	8.98	0.42	2,217	8.14	0.46
-1.65	2,469	9.46	0.44	2,189	8.51	0.48
-1.75	2,350	9.59	0.47	2,224	9.01	0.49
-1.85	2,384	10.42	0.46	2,257	9.76	0.49
-1.95	2,212	10.95	0.47	2,097	10.15	0.51
-2.05	2,177	11.38	0.50	2,070	10.70	0.53
-2.15	2,046	11.86	0.52	1,919	11.26	0.55
-2.25	1,987	12.32	0.55	1,881	11.76	0.57
-2.35	2,028	12.95	0.57	1,887	12.36	0.59
-2.45	1,853	13.39	0.62	1,707	13.05	0.61
-2.55	1,808	14.34	0.64	1,593	13.72	0.63
-2.65	1,818	15.23	0.65	1,608	14.27	0.67
-2.75	1,727	15.36	0.73	1,453	14.80	0.71
-2.85	1,477	16.12	0.76	1,249	15.65	0.73
-2.95	1,356	16.89	0.81	1,091	16.25	0.76
-3.05	1,194	18.04	0.81	937	16.97	0.78
-3.15	1,018	19.29	0.80	789	17.83	0.79
-3.25	832	20.73	0.78	650	18.82	0.79
-3.35	654	21.35	0.80	549	19.62	0.81
-3.45	496	22.16	0.82	439	20.88	0.79
-3.54	387	23.31	0.83	328	21.95	0.80
-3.65	263	25.61	0.78	224	23.61	0.80
-3.75	182	26.42	0.83	147	24.88	0.81

– N is the number of stars in the bin, a indicates the BHB sample, and b indicates the BHB subsample of the smooth stellar halo.

4. Summary

In this study, we constructed a catalog of 44,524 BHB stars selected from the Gaia DR3 XP spectra by combining photometric and spectroscopic data. Using a fraction of BHB stars with a reliable parallax, we derived a relationship between the absolute magnitude M_G and the dereddened stellar color $G_{BP,0} - G_{RP,0}$. A purity of around 90% was obtained by adopting the SEGUE spectroscopic-confirmed BHB catalog of Barbosa et al. (2022) and the hot stars catalog of Xiang et al. (2022) as the reference datasets. We estimated the selection function of these BHB stars by comparing our catalog to the complete photometric survey of Gaia DR3. Assuming that the density shape varies with the radius, the radial density profile of the Galactic stellar halo traced by the BHB catalog can be well defined by a single power law with $\alpha = -4.80 \pm 0.06$. A comparison between the BHB data (ν) and model (ν_{fitting}) suggests a possible bias introduced by contaminants or stars from the stellar substructures. After removing these obvious outliers ($\frac{\nu - \nu_{\text{fitting}}}{\nu_{\text{fitting}}} > 0.6$), we obtained a smoother and more consistently positive relationship of $r - q$, and the density profile only shows a slight change with the best-estimated index $\alpha = -4.65 \pm 0.04$.

This study is an attempt to construct a full-sky covered complete BHB catalog of the Milky Way from the low-resolution XP spectra. Although the low-resolution XP spectra prevent us from obtaining an extremely accurate estimation of the stellar parameters, we are still able to achieve a purity of around 87%, similar to the BHB catalog of Ju et al. (2024), which was se-

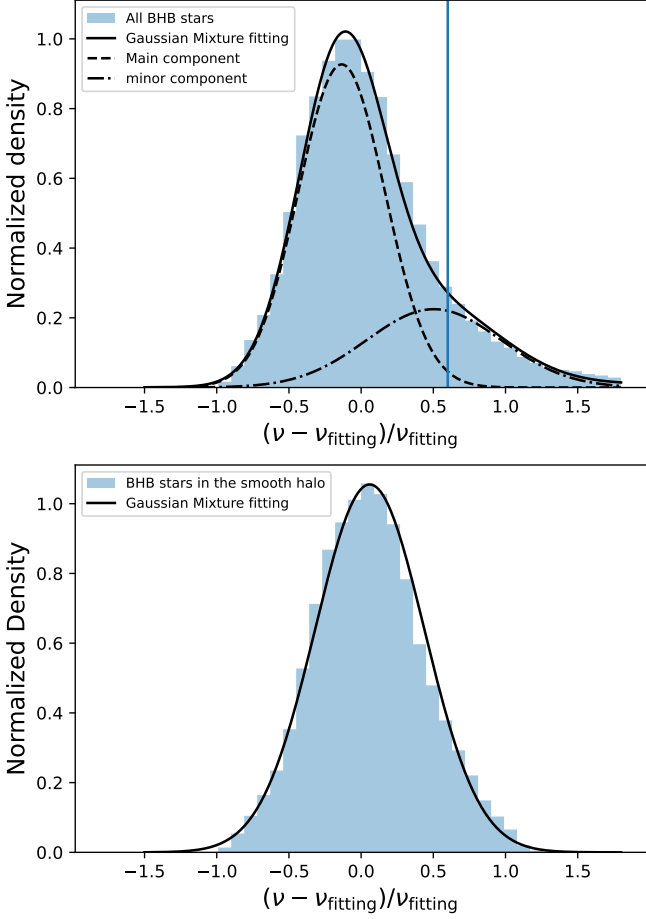


Fig. 10. Normalized density distribution of $(\nu - \nu_{\text{fitting}})/\nu_{\text{fitting}}$ obtained from the whole BHB stars (top panel) and the subsample of the smooth halo (bottom panel). The blue vertical line represents the position of $(\nu - \nu_{\text{fitting}})/\nu_{\text{fitting}} = 0.6$.

lected from the higher resolution spectra of LAMOST. However, the completeness, which drops from about 90% at $G = 14$ mag to around 40% at $G = 17$ mag, suggests that there are still a large number of BHB stars that wait to be identified from the XP spectra, especially in the faint magnitudes. The lack of radial velocity measurements in this BHB catalog makes exploring the phase-space distribution of the stellar halo difficult. The upcoming Chinese Space Station Telescope (CSST) spectroscopic survey will release hundreds of millions of high-quality low-resolution ($R > 200$) slitless spectra. We believe that our selection method will also be appropriate for the CSST low-resolution spectra survey (a limiting magnitude of around 23 mag in U, V, and I bands), which is much deeper than the Gaia XP spectra (a limiting magnitude of around 17.5 mag in G band) and will allow us to explore the outer regions of the stellar halo.

5. Data availability

Tables of the selected BHB members are only available in electronic form at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](ftp://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/>.

Acknowledgements. This study is supported by the National Key R&D Program of China under grant Nos. 2023YFE0107800, 2024YFA1611900, and

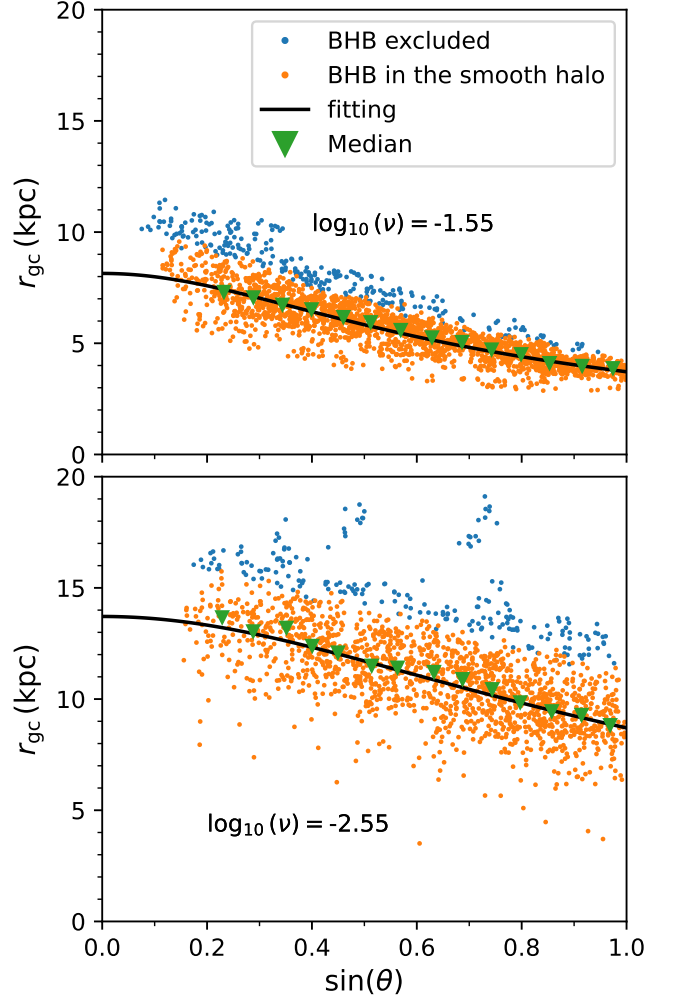


Fig. 11. Fitting results in two bins of $\log_{10}(\nu) = -1.55$ (top panel) and -2.55 (bottom panel) for the BHB subsample of the smooth halo. In each panel, the orange points are BHB stars within the specified bin, while the green inverted triangles indicate the median value of r_{gc} . The black lines depict the best fitting results of the subsample. The blue points denote stars that were previously included in Figure 8 but have been excluded here due to their large residual of $\frac{\nu - \nu_{\text{fitting}}}{\nu_{\text{fitting}}} \geq 0.6$. In the top panel, the excluded BHB stars are responsible for the prominent overdensity in the solar neighborhood in Figure 7. In the bottom panel, these excluded stars are likely associated with the overdensity in the outer regions of the halo.

the National Natural Science Foundation of China under grant Nos. 12588202, 12273055, 12373020. This study is also supported by International Partnership Program of Chinese Academy of Sciences. Grant No. 178GJHZ2022040GC, and CAS Project for Young Scientists in Basic Research grant No. YSBR-062 and YSBR-092. We also thank the support by the China Manned Space Program with grant nos. CMS-CSST-2025-A11 and CMS-CSST-2025-A12. Xi-anhao Ye and Wenbo Wu acknowledge the support from the China Scholarship Council. CAP, DA, JIGH, and RRL acknowledge financial support from the Spanish Ministry of Science, Innovation and Universities (MICIU) projects PID2020-117493GB-I00 and PID2023-149982NB-I00. This research made use of computing time available on the high-performance computing systems at the Instituto de Astrofísica de Canarias. This work presents results from the European Space Agency (ESA) space mission Gaia. Gaia data are being processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the Gaia MultiLateral Agreement (MLA). The Gaia mission website is <https://www.cosmos.esa.int/gaia>. The Gaia archive website is <https://archives.esac.esa.int/gaia>. This job has made use of the Python package GaiaXPpy, developed and maintained by members of the Gaia Data Processing

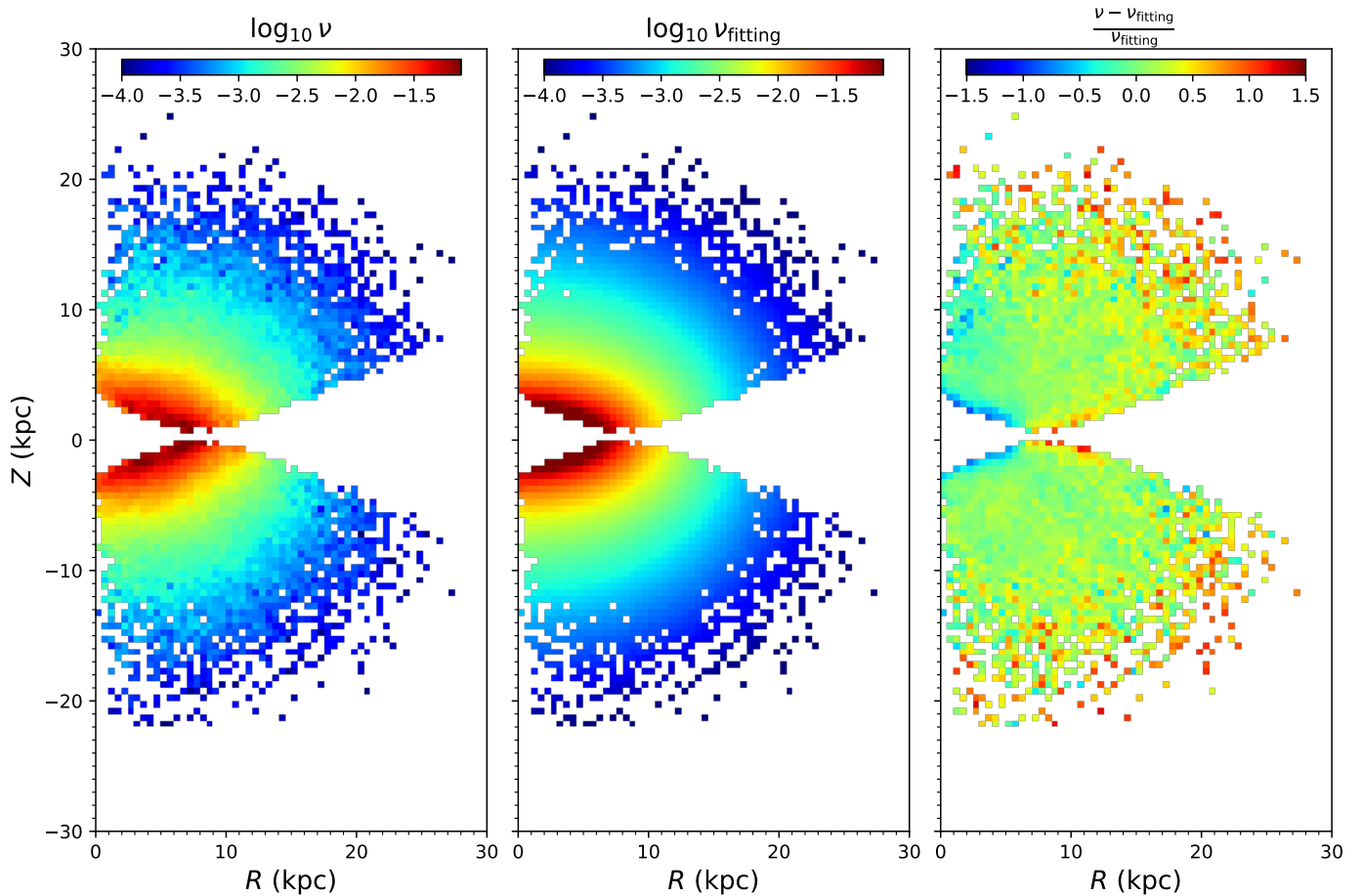


Fig. 12. Number density map of the stellar halo in $R - Z$ space for the BHB stars in the subsample ($\log_{10} \nu$, left panel) and the model ($\log_{10} \nu_{\text{fitting}}$, middle panel). The pixel size is 0.5×0.5 (kpc). The color in each pixel represents the mean density on a logarithmic scale. The residual map $((\nu - \nu_{\text{fitting}})/\nu_{\text{fitting}})$ presented in the right panel indicates a good consistency between the data and the model, where no prominent overdensity exists in the solar neighborhood and the outer parts of the stellar halo.

and Analysis Consortium (DPAC), and in particular, Coordination Unit 5 (CU5), and the Data Processing Centre located at the Institute of Astronomy, Cambridge, UK (DPCI). Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

References

- Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, *ApJ*, 636, 804
Allende Prieto, C., Koesterke, L., Hubeny, I., et al. 2018, *A&A*, 618, A25
Amarante, J. A. S., Koposov, S. E., & Laporte, C. F. P. 2024, *A&A*, 690, A166
Andrae, R., Fouesneau, M., Sordo, R., et al. 2023, *A&A*, 674, A27
Arp, H. C., Baum, W. A., & Sandage, A. R. 1952, *AJ*, 57, 4
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123
Barbosa, F. O., Santucci, R. M., Rossi, S., et al. 2022, *ApJ*, 940, 30
Baumgardt, H. & Vasiliev, E. 2021, *MNRAS*, 505, 5957
Beers, T. C., Almeida, T., Rossi, S., Wilhelm, R., & Marsteller, B. 2007, *ApJS*, 168, 277
Belokurov, V. & Koposov, S. E. 2016, *MNRAS*, 456, 602
Belokurov, V. & Kravtsov, A. 2022, *MNRAS*, 514, 689
Bennett, M. & Bovy, J. 2019, *MNRAS*, 482, 1417
Bird, S. A., Xue, X.-X., Liu, C., et al. 2022, *MNRAS*, 516, 731
Bird, S. A., Xue, X.-X., Liu, C., et al. 2021, *ApJ*, 919, 66
Cantat-Gaudin, T., Fouesneau, M., Rix, H.-W., et al. 2023, *A&A*, 669, A55
Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, *A&A*, 652, A86
Castro-Ginard, A., Brown, A. G. A., Kostrzewa-Rutkowska, Z., et al. 2023, *A&A*, 677, A37
Chen, A., Li, Z., Wang, Y., et al. 2023, *MNRAS*, 525, 3075
Clewley, L., Warren, S. J., Hewett, P. C., et al. 2002, *MNRAS*, 337, 87
Conroy, C., Weinberg, D. H., Naidu, R. P., et al. 2022, *arXiv e-prints*, arXiv:2204.02989
Cooper, W. J., Smart, R. L., Jones, H. R. A., & Sarro, L. M. 2024, *MNRAS*, 527, 1521
Culpan, R., Dorsch, M., Geier, S., et al. 2024, *A&A*, 685, A134
Culpan, R., Pelisoli, I., & Geier, S. 2021, *A&A*, 654, A107
Das, P., Williams, A., & Binney, J. 2016, *MNRAS*, 463, 3169
De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, *A&A*, 674, A2
Deason, A. J., Belokurov, V., & Evans, N. W. 2011, *MNRAS*, 416, 2903
Deason, A. J., Belokurov, V., Koposov, S. E., & Lancaster, L. 2018, *ApJ*, 862, L1
Deason, A. J., Erkal, D., Belokurov, V., et al. 2021, *MNRAS*, 501, 5964
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, 157, 168
Evans, D. W., Riello, M., De Angeli, F., et al. 2018, *A&A*, 616, A4
Faccioli, L., Smith, M. C., Yuan, H. B., et al. 2014, *ApJ*, 788, 105
Fukushima, T., Chiba, M., Homma, D., et al. 2018, *PASJ*, 70, 69
Fukushima, T., Chiba, M., Tanaka, M., et al. 2019, *PASJ*, 71, 72
Fukushima, T., Chiba, M., Tanaka, M., et al. 2025, *PASJ*, 77, 178

- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, *A&A*, 649, A1
- Gaia Collaboration, Montegriffo, P., Bellazzini, M., et al. 2023a, *A&A*, 674, A33
- Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023b, *A&A*, 674, A1
- García-Zamora, E. M., Torres, S., & Rebassa-Mansergas, A. 2023, *A&A*, 679, A127
- GRAVITY Collaboration, Abuter, R., Amorim, A., et al. 2018, *A&A*, 615, L15
- Green, G. 2018, *The Journal of Open Source Software*, 3, 695
- Grynciewicz, R., Newberg, H. J., Martin, C., Donlon, Thomas, I., & Amy, P. M. 2021, *ApJ*, 910, 102
- Han, J. J., Conroy, C., Johnson, B. D., et al. 2022, *AJ*, 164, 249
- Hernitschek, N., Cohen, J. G., Rix, H.-W., et al. 2018, *ApJ*, 859, 31
- Huang, B., Yuan, H., Xiang, M., et al. 2024, *ApJS*, 271, 13
- Iorio, G., Belokurov, V., Erkal, D., et al. 2018, *MNRAS*, 474, 2142
- Jofré, P. & Weiss, A. 2011, *A&A*, 533, A59
- Ju, J., Cui, W., Huo, Z., et al. 2024, *ApJS*, 270, 11
- Kafle, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, *ApJ*, 761, 98
- Li, C. & Binney, J. 2022, *MNRAS*, 510, 4706
- Li, S., Casertano, S., & Riess, A. G. 2023, *ApJ*, 950, 83
- Lindgren, L., Bastian, U., Biermann, M., et al. 2021, *A&A*, 649, A4
- Liu, C., Xu, Y., Wan, J.-C., et al. 2017, *Research in Astronomy and Astrophysics*, 17, 096
- Martin, N. F., Starkenburg, E., Yuan, Z., et al. 2024, *A&A*, 692, A115
- Medina, G. E., Muñoz, R. R., Carlin, J. L., et al. 2024, *MNRAS*, 531, 4762
- Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, *New A*, 15, 433
- Montegriffo, P., De Angeli, F., Andrae, R., et al. 2023, *A&A*, 674, A3
- Montenegro, K., Minniti, D., Alonso-García, J., et al. 2019, *ApJ*, 872, 206
- Paunzen, E. & Prišegen, M. 2022, *A&A*, 667, L10
- Pier, J. R. 1982, *AJ*, 87, 1515
- Pier, J. R. 1983, *ApJS*, 53, 791
- Pila-Díez, B., de Jong, J. T. A., Kuijken, K., van der Burg, R. F. J., & Hoekstra, H. 2015, *A&A*, 579, A38
- Preston, G. W., Shectman, S. A., & Beers, T. C. 1991, *ApJ*, 375, 121
- Ruhland, C., Bell, E. F., Rix, H.-W., & Xue, X.-X. 2011, *ApJ*, 731, 119
- Sanders, J. L. & Matsunaga, N. 2023, *MNRAS*, 521, 2745
- Schlafly, E. F. & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Sirko, E., Goodman, J., Knapp, G. R., et al. 2004, *AJ*, 127, 899
- Sluis, A. P. N. & Arnold, R. A. 1998, *MNRAS*, 297, 732
- Smith, K. W., Bailer-Jones, C. A. L., Klement, R. J., & Xue, X. X. 2010, *A&A*, 522, A88
- Sommer-Larsen, J., Christensen, P. R., & Carter, D. 1989, *MNRAS*, 238, 225
- Starkenburg, E., Youakim, K., Martin, N., et al. 2019, *MNRAS*, 490, 5757
- Subramaniam, A., Sahu, S., Postma, J. E., et al. 2017, *AJ*, 154, 233
- Sun, S., Wang, F., Zhang, H., et al. 2025, *ApJ*, 979, 213
- Thomas, G. F., McConnachie, A. W., Ibata, R. A., et al. 2018, *MNRAS*, 481, 5223
- VandenBerg, D. A., Brogaard, K., Leaman, R., & Casagrande, L. 2013, *ApJ*, 775, 134
- Vasiliev, E. & Baumgardt, H. 2021, *MNRAS*, 505, 5978
- Vickers, J. J., Grebel, E. K., & Huxor, A. P. 2012, *AJ*, 143, 86
- Vincent, O., Barstow, M. A., Jordan, S., et al. 2024, *A&A*, 682, A5
- Watkins, L. L., Evans, N. W., Belokurov, V., et al. 2009, *MNRAS*, 398, 1757
- Whitten, D. D., Beers, T. C., Placco, V. M., et al. 2019, *ApJ*, 884, 67
- Wu, W., Zhao, G., Xue, X.-X., Bird, S. A., & Yang, C. 2022a, *ApJ*, 924, 23
- Wu, W., Zhao, G., Xue, X.-X., Pei, W., & Yang, C. 2022b, *AJ*, 164, 41
- Xiang, M., Rix, H.-W., Ting, Y.-S., et al. 2022, *A&A*, 662, A66
- Xiang, M., Rix, H.-W., Yang, H., et al. 2025, *Nature Astronomy*, 9, 101
- Xu, Y., Liu, C., Xue, X.-X., et al. 2018, *MNRAS*, 473, 1244
- Xue, X.-X., Rix, H.-W., Ma, Z., et al. 2015, *ApJ*, 809, 144
- Xue, X.-X., Rix, H.-W., Yanny, B., et al. 2011, *ApJ*, 738, 79
- Xue, X. X., Rix, H. W., Zhao, G., et al. 2008, *ApJ*, 684, 1143
- Yang, C., Xue, X.-X., Li, J., et al. 2019a, *ApJ*, 886, 154
- Yang, C., Xue, X.-X., Li, J., et al. 2019b, *ApJ*, 880, 65
- Yang, C., Zhu, L., Tahmasebzadeh, B., Xue, X.-X., & Liu, C. 2022, *AJ*, 164, 241
- Yanny, B., Newberg, H. J., Kent, S., et al. 2000, *ApJ*, 540, 825
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, 137, 4377
- Ye, X., Wu, W., Allende Prieto, C., et al. 2025, *A&A*, 695, A75
- York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, *AJ*, 120, 1579
- Yu, F., Li, T. S., Speagle, J. S., et al. 2024, *ApJ*, 975, 81
- Zhang, H., Ardem-Arentsen, A., & Belokurov, V. 2024, *MNRAS*, 533, 889
- Zhang, R. & Yuan, H. 2023, *ApJS*, 264, 14

Appendix A: Necessity of a combined photometric and spectroscopic cut

In this appendix, we conducted a test by applying only photometric cuts, only spectroscopic cuts, and combined photometric/spectroscopic cuts to Gaia XP spectra. Several high-latitude ($|b| > 20^\circ$) selected BHB candidates are also available in the SEGUE spectroscopic catalog. Figure A.1 shows their distributions in the $\text{TEFF_SPEC} - \text{LOGG_SPEC}$ diagram ($T_{\text{eff}} - \log g$ provided by the SEGUE survey). Spectroscopic cuts perform better in the selection than photometric cuts, but there are still some contaminants that can only be removed by the photometric cuts. The combined photometric/spectroscopic cuts produce the cleanest sample, as indicated by the smallest rate (around 5%) of contaminants from high surface gravity stars ($\text{LOGG_SPEC} > 3.7$).

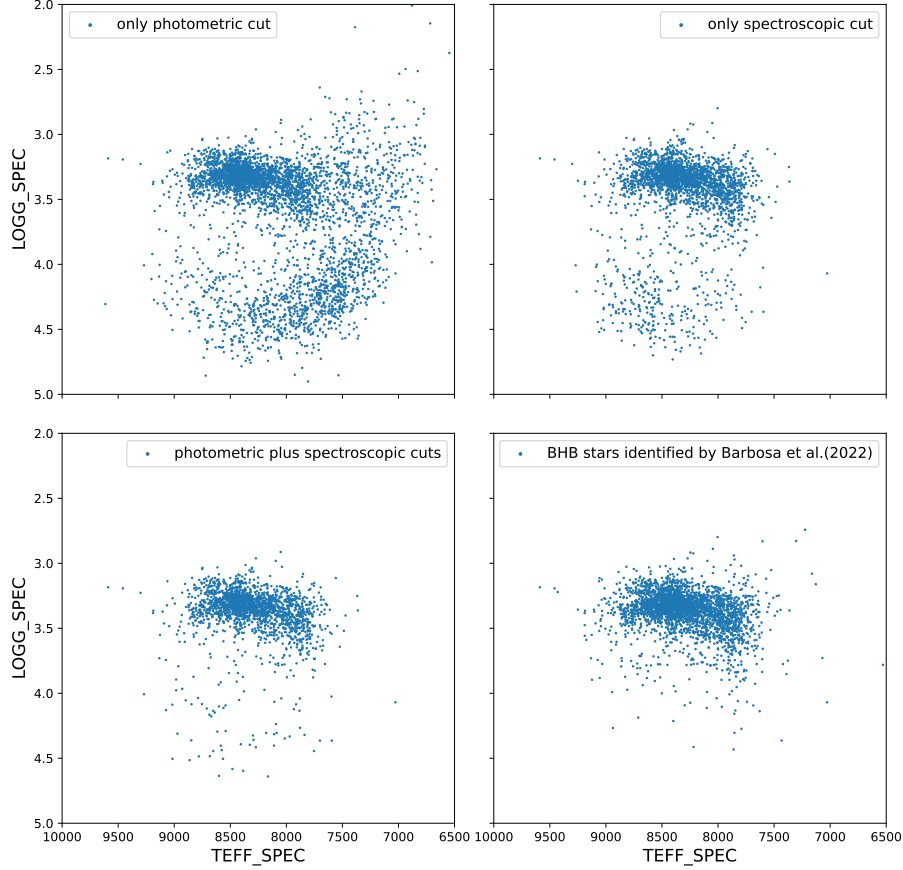


Fig. A.1. $\text{TEFF_SPEC} - \text{LOGG_SPEC}$ diagram of the selected BHB candidates obtained by different cuts, and the BHB sample of Barbosa et al. (2022) is used as a reference. The proportion of contaminants from high surface gravity ($\text{LOGG_SPEC} > 3.7$) stars is 28% for only photometric cuts, 14% for only spectroscopic cuts, and 5% for combined photometric/spectroscopic cuts.

Appendix B: Influence of the stellar substructures

In this appendix, we compared our sample with RR Lyrae members of several known halo substructures identified in Sun et al. (2025), and selected BHB stars that are similar to these members in both spatial positions (within 1 kpc) and proper motions (within 1 mas yr^{-1}). Here we note that this method can only identify a small fraction of BHB stars that may belong to some distinct halo substructures. Figure B.1 exhibits the distributions of these halo substructures in $l - b$ diagram, along with the normalized distributions of $(\nu - \nu_{\text{fitting}})/\nu_{\text{fitting}}$ for BHB stars associated with the two most obvious substructures of the Hercules-Aquila Cloud and the Sagittarius stream. The stellar density ν does not show a large deviation from ν_{fitting} in the spatial positions corresponding to the Hercules-Aquila Cloud. However, this is not the case in the spatial positions containing BHB stars from the Sagittarius stream, where the mean value of $(\nu - \nu_{\text{fitting}})/\nu_{\text{fitting}}$ reaches as high as 1.17. Most BHB stars of the Hercules-Aquila Cloud are located in the inner halo $10 < r_{\text{gc}} < 20 \text{ kpc}$, while the BHB members of the Sagittarius stream are mainly found in the outer halo $r_{\text{gc}} > 20 \text{ kpc}$. The stellar density is much lower in the outer halo compared to the inner halo. Therefore, if we assume the same increase in stellar density caused by halo substructures, their impact on the derived density profile and shape is likely to be more pronounced in the outer halo. This interpretation is consistent with the residual map shown in Figure 7, where the most prominent difference is typically seen in the outer halo.

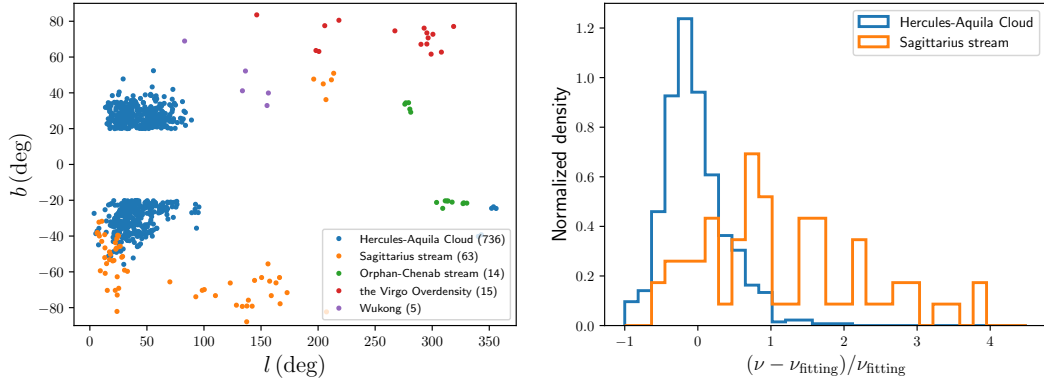


Fig. B.1. Sky positions in l - b space (left panel) are shown for BHB stars that may belong to several halo substructures, along with the normalized distribution of $(\nu - \nu_{\text{fitting}})/\nu_{\text{fitting}}$ (right panel) for BHB stars associated with the Hercules-Aquila Cloud and the Sagittarius stream. The numbers in parentheses indicate the number of BHB candidates associated with each substructure.