The Astrophysical Journal, 940:30 (9pp), 2022 November 20 https://doi.org/10.3847/1538-4357/ac983f
© 2022. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

The SDSS-Gaia View of the Color–Magnitude Relation for Blue Horizontal-branch Stars

Fabrícia O. Barbosa[1], Rafael M. Santucci[2,3], Silvia Rossi[1], Guilherme Limberg[1], Angeles Pérez-Villegas[4], and Hélio D. Perottoni[1,5]
[1] Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Departamento de Astronomia, SP 05508-090, São Paulo, Brazil
fabriciaob@usp.br
[2] Universidade Federal de Goiás, Instituto de Estudos Socioambientais, Planetário, Goiânia, GO 74055-140, Brazil
[3] Universidade Federal de Goiás, Campus Samambaia, Instituto de Física, Goiânia, GO 74001-970, Brazil
[4] Instituto de Astronomía, Universidad Nacional Autónoma de México, Apartado Postal 106, C.P. 22800, Ensenada, B.C., Mexico
[5] Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, E08028 Barcelona, Spain
Received 2022 July 25; revised 2022 September 23; accepted 2022 October 4; published 2022 November 17

Abstract

We present an updated sample of blue horizontal-branch (BHB) stars selected from the photometric and spectroscopic data from Sloan Digital Sky Survey and its associated project Sloan Extension for Galactic Understanding and Exploration (SEGUE). With these data, we selected candidates for A-type stars in the color–color space and then a mixture modeling technique was implemented in order to distinguish between BHB and main-sequence/blue-straggler stars based on their surface gravity values ($\log g$) estimated by the SEGUE Stellar Parameter Pipeline. Our robust approach allows us to attribute individual probabilities of each star truly being in the BHB stage. Hence, our method is advantageous in comparison to previous SEGUE BHB selections that adopted simple $\log g$ cuts. We also revisit the color–magnitude relation for these stars and propose two calibrations, based on updated distances for Galactic globular clusters, to estimate absolute magnitudes with $(g-r)_0$ and $(u-r)_0$ colors.

Unified Astronomy Thesaurus concepts: Horizontal branch stars (746); Stellar distance (1595)

1. Introduction

The Gaia mission (Gaia Collaboration et al. 2016) has provided a better understanding of the Galaxy, in particular regarding the field of Galactic archaeology (Helmi 2020; Brown 2021). The astrometric information provided for an unprecedented number of objects has dramatically changed the way we study the Galactic halo (e.g., Belokurov et al. 2018; Koppelman et al. 2018; Malhan et al. 2018; Myeong et al. 2018).

Despite the huge amount of direct measurements supplied by Gaia, distances inferred from brightness are still of great value. At magnitude $G < 15$, the early third data release (EDR3) presents parallax uncertainties of ∼0.02 mas (Gaia Collaboration et al. 2021), and they increase significantly for fainter stars. To overcome this limitation, we can use various well-known distance tracers such as RR Lyrae (Shapley 1916), Cepheids (Leavitt & Pickering 1912), and blue horizontal-branch (BHB) stars (Cacciari 1999).

BHBs are metal-poor ([Fe/H][6] $\lesssim$ –0.5; Santucci et al. 2015a) A- or B-type stars that burn helium in their cores. These evolved stars present a high and nearly constant luminosity, making them perfect for investigating the outer regions of the halo and the assembly history of our Galaxy (Deason et al. 2011, 2017; Xue et al. 2011, 2008; Belokurov et al. 2014; Santucci et al. 2015a). In recent works, BHBs were used to study dynamical substructures and stellar streams (Yuan et al. 2019, 2020, 2022; Peñarrubia & Petersen 2021; Li et al. 2022; Wu et al. 2022), the connection between the apocenter pileup of orbits and the so-called "break-radius" of the stellar halo density profile (Deason et al. 2018), the anisotropy of the halo velocity distribution (Lancaster et al. 2019), the age gradient of the halo out to ∼35 kpc (Whitten et al. 2019), to estimate the total dynamical mass of the Milky Way (Deason et al. 2021; Bird et al. 2022), and even to demonstrate the influence of the Large Magellanic Cloud in our Galaxy's halo (Erkal et al. 2021; Petersen & Peñarrubia 2021).

The well-defined structure of the horizontal branch in the color–magnitude diagram, a roughly constant luminosity, permits the development of a distance calibration for these BHBs. The first approximation developed was a linear fit, using the $(B-V)$ color and absolute magnitude in the V band, for stars in globular clusters (Hayes & Philip 1979). Likewise, Preston et al. (1991) defined a smoother relation, a fourth degree polynomial, for the same color–magnitude space. Two decades later, a widely used calibration was presented by Deason et al. (2011, hereafter D11) based on magnitudes in the *ugriz* system for the Sloan Digital Sky Survey (SDSS; York et al. 2000) eighth data release (DR8; Aihara et al. 2011), which had its color range extended by Belokurov & Koposov (2016) afterwards. In the meantime, Fermani & Schönrich (2013) argued that it is extremely important to take into account the effect of the metallicity on the absolute magnitude estimation, proposing a new calibration based on a statistical method. However, Santucci et al. (2015b) and Utkin & Dambis (2020) showed that the differences between considering and not considering the metallicity in the relations are negligible, with D11's estimates being 2.5% higher, within (1σ) errors of both calibrations.

D11's relation still remains the most used calibration for BHB stars (Santucci et al. 2015a, 2015b; Thomas et al. 2018; Whitten et al. 2019; Donlon et al. 2020; Martin et al. 2022) even though photometric data have been updated several times

[6] $[A/B] = \log(N_A/N_B)_\star - \log(N_A/N_B)_\odot$, where N_A and N_B are the number density of atoms of the elements A and B, respectively. $\star$ refers to the considered star, and $\odot$ refers to the Sun.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

since then. Moreover, we can now compare photometric distances of BHB stars with purely geometric estimates from Gaia's parallaxes (e.g., Bailer-Jones et al. 2021) as well as new measurements for Galactic globular clusters (Vasiliev & Baumgardt 2021). These facts bring to light the relevance of reviewing the D11's calibration with recent data.

This paper is organized as follows. In Section 2, we describe the photometric selection and revise a previous method to identify BHB stars. Section 3 presents the selection of stars in globular clusters and the method used to define the absolute magnitude calibration. Finally, in Section 4 we discuss our results.

2. Data

2.1. A-type Stars

The initial selection of A-type stars was made using the photometry from the sixteenth data release (DR16) of SDSS (Ahumada et al. 2020). For the selection of BHB stars, we were specially interested in the spectroscopic data obtained by the Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009) processed by the SEGUE Stellar Parameter Pipeline (SSPP; Lee et al. 2008a, 2008b).[7]

We implemented color cuts applying the following criteria: $-0.3 < (g-r)_0 < 0.1$ and $0.8 < (u-g)_0 < 1.4$, similar to those used in previous works (Sirko et al. 2004; D11). All the magnitudes were corrected using the extinction coefficients (A_g, A_r, A_u) provided by the SDSS catalog itself and we removed stars with relative errors in the g-band magnitude greater than 1%.

The photometric selection is able to exclude several undesired objects, such as white dwarfs, quasars, and cooler spectral types (Yanny et al. 2009; Vickers et al. 2012), but the major source of contamination, blue-straggler stars (BSSs), remains. The distinction between evolved and main-sequence stars/BSSs is commonly made by investigating spectral features, especially Balmer lines, whose depths are affected by effective temperature ($T_{\rm eff}$) and widths by surface gravity ($\log g$). With the output of SSPP, we can directly inspect these stellar atmospheric parameters. Therefore, we crossmatched the filtered sample with the SSPP catalog using 5″ radius. In addition to color filters, we restricted our sample to stars with moderate signal-to-noise ratio (S/N > 10) and 7500 K < $T_{\rm eff}$ < 10,000 K (Deason et al. 2012; Santucci et al. 2015b), where $T_{\rm eff}$ is the estimate adopted by the pipeline. Duplicated stars with the smallest S/N were removed, which resulted in 16,463 stars. The restrictions above remove poor-quality data and cooler stars that could remain after the color cut, which assures that contamination from other non-BSS stars is minimal.

2.2. BHB Stars

One of the techniques used to disentangle BSSs and BHB stars is the f_m versus $D_{0.2}$ method (Pier 1983), where f_m is the minimum flux relative to the continuum level and $D_{0.2}$ is a measurement of the line width of Balmer lines, so it provides indirect information regarding both $T_{\rm eff}$ and $\log g$. Later, a different approach was proposed by Clewley et al. (2002) based on the parameters of the Sérsic profile (Sersic 1968), which describes the shape of the lines.

[7] Last run on DR9 (Ahn et al. 2012; Rockosi et al. 2022).

BSSs present a stronger $\log g$ than those located in the horizontal branch. Santucci et al. (2015b) showed that these stellar types are clearly distinguishable for magnitudes $g_0 < 18$ with SEGUE/SDSS DR8 data, being possible to classify them by fitting a combination of two Gaussian functions to their $\log g$ distributions. This method was proved to be in good concordance with spectral analysis, with more than 90% agreement. When replicating this procedure with current SEGUE data, we noticed a change in the peaks of both groups and a greater overlap in the $\log g$ distribution as presented in the left column of Figure 1. This is observed even for relatively bright A-type stars ($g_0 < 18$, top panel), which makes it more difficult to separate these objects with that simple approach (dashed lines indicate the Gaussian fits from Santucci et al. 2015b). The differences are probably due to changes between the releases of SDSS on the $\log g$ estimates considered to obtain the final adopted parameter.[8]

2.3. Classification

Given the two-Gaussian-like morphology of the $\log g$ distributions observed for our sample of A-type stars (Figure 1), we used a Gaussian mixture model (GMM) unsupervised approach in order to distinguish BHBs and BSSs. For this task, we utilize the `scikit-learn` (Pedregosa et al. 2011) `GaussianMixture`[9] package. In this GMM implementation, the expectation-maximization algorithm (Dempster et al. 1977) is employed in the search for the best-fit model.

The GMM technique fits the data as a finite combination of K Gaussian distributions. As made previously by Santucci et al. (2015b), K was defined based on visual inspection of $\log g$ estimates presented in Figure 1 and the assumption that the contamination is predominantly of main-sequence stars/BSSs. Therefore, $K = 2$ is an adequate value for the sample. Moreover, GMM can be readily applied to data of arbitrary dimensionality. Therefore, we take advantage of such flexibility and explore a suitable combination of $\log g$ estimates provided by SSPP (we refer the reader to Lee et al. 2008a for details about different approaches to determine $\log g$ from SEGUE spectra). We noticed that the distributions of both $\log g_{\rm ANNRR}$ and $\log g_{\rm SPEC}$ exhibit clearly two peaks, as expected for the BHBs/BSSs dichotomy, while it is not possible to observe this feature in others. These two distributions are shown in the right column of Figure 1. Hence, we proceeded with the GMM separation within the two-dimensional space defined by these $\log g$ estimates. The final $\log g$ adopted by the pipeline was not considered an extra dimension as it consists of a weighted mean of the valid estimates.

In order to guarantee the robustness of our method against uncertainties reported by the SSPP, we constructed a set of 10^4 realizations of each star's $\log g$ estimates in a Monte Carlo framework. Then, we performed the GMM classification for all iterations. Finally, the fraction of instances that a star is attributed to a certain class (either BHB or BSS) is taken as its membership probability for that given group. For this procedure, stars without valid estimates of both $\log g_{\rm ANNRR}$ and $\log g_{\rm SPEC}$ are removed. With this strategy, we achieved a

[8] See https://www.sdss.org/dr16/spectro/sspp_changes/ for detailed information.

[9] https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html#sklearn.mixture.GaussianMixture

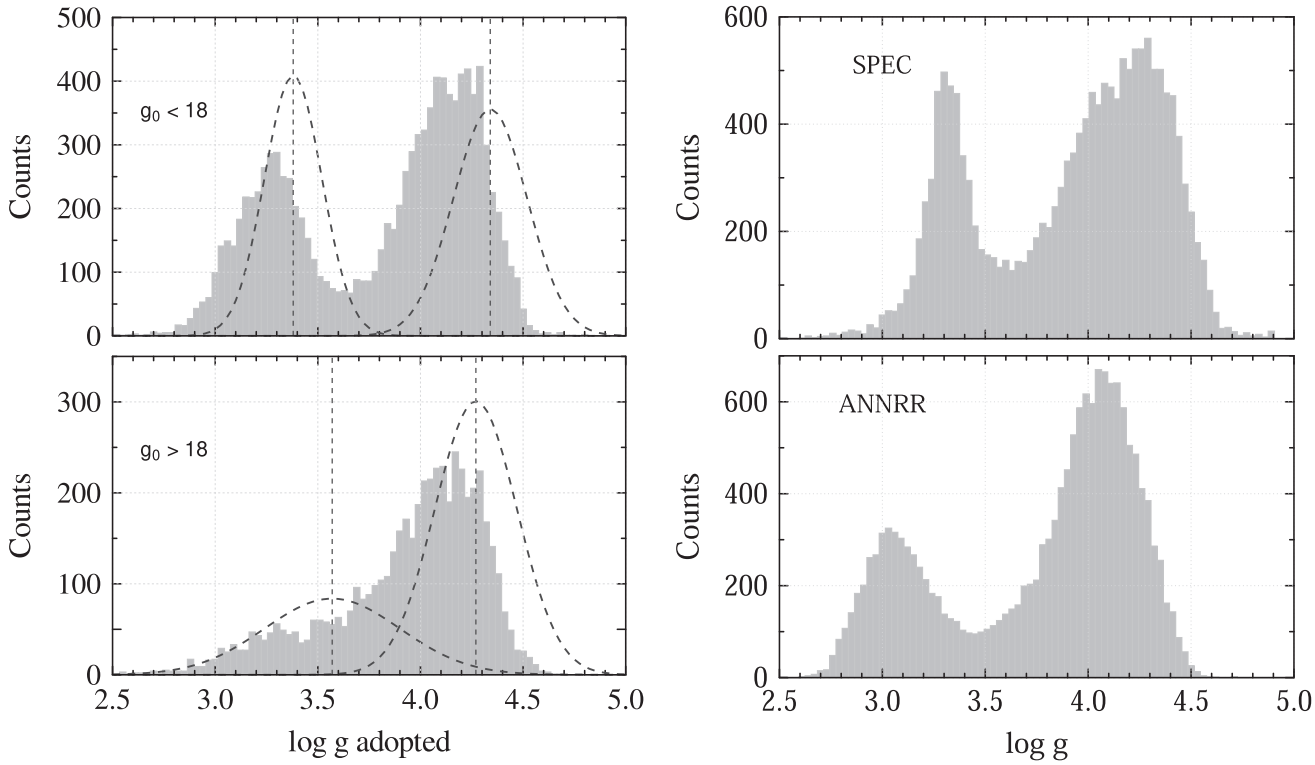

Figure 1. Histograms of log g. Left column: log g adopted by the pipeline for stars with $g_0 < 18$ (top) and $g_0 > 18$ (bottom). Dashed lines are the Gaussian distributions defined by Santucci et al. (2015b). Right column: log g estimates provided by SSPP spectroscopically determined (top) and from ANNRR method (bottom) for all stars.

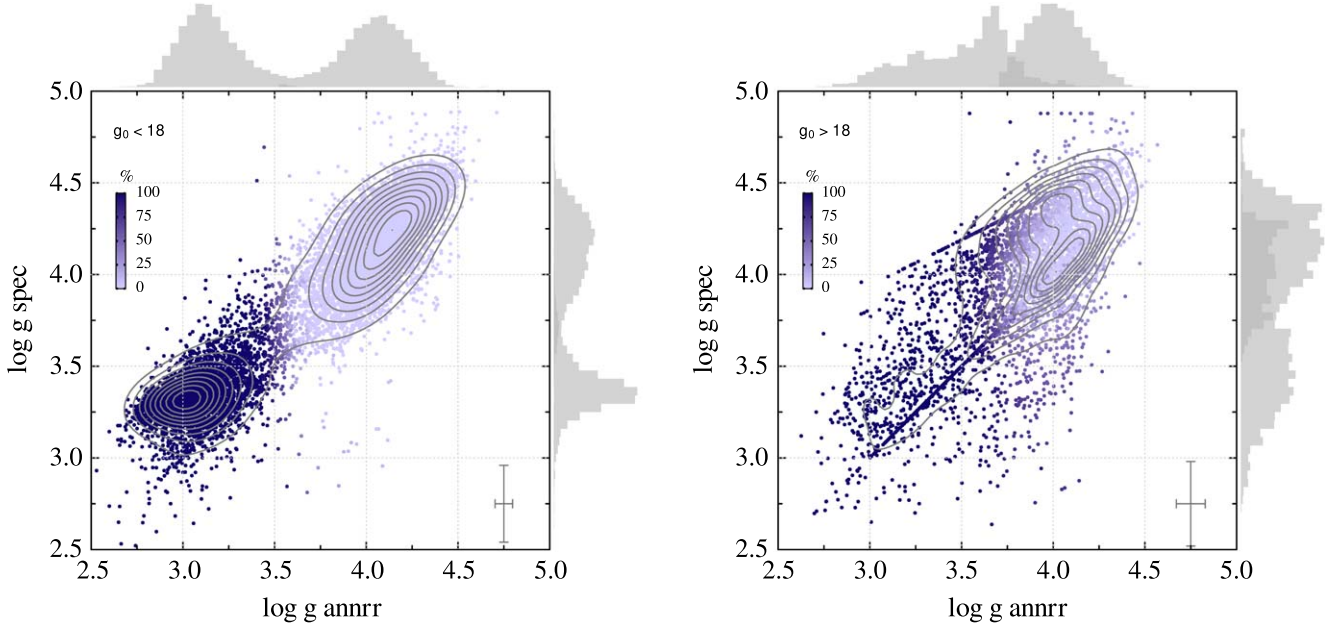

Figure 2. Distribution of classified stars with $g_0 < 18$ (left) and $g_0 > 18$ (right) in the surface gravity space. Median errors are indicated in the bottom right corner. Histograms show the distribution of log g from the respective axis for stars classified as BHB and BSS. Colors indicate the probability of being a BHB star.

sample of 5699/4590 stars classified as BHBs above 50%/99% probability.[10]

The final classification obtained is shown in Figure 2. The difference in the uncertainties of the estimates greatly influences the classification, as the log $g_{\rm ANNRR}$ presents more precise values (∼0.06) than log $g_{\rm SPEC}$ (∼0.21). We crossmatched our sample with the one from Santucci et al. (2015b) to evaluate the fraction of BSS contamination. 10% of our BHB set was classified previously as BSSs, and, among those with a probability greater than 99% of being BHB following the method implemented here, 2% are possibly incorrectly assigned.

[10] The full sample is available at https://github.com/guilhermelimberg/bhb_dist.

Table 1
Heliocentric Distances Provided by Vasiliev & Baumgardt (2021) for Each Globular Cluster, Uncertainties, and Their Respective Distance Moduli

Cluster	D (kpc)	σ_D (kpc)	(m-M)_0 (mag)
NGC 2419	83.0	1.5	19.59
NGC 4147	18.65	0.16	16.35
NGC 5024, M53	18.59	0.15	16.35
NGC 5053	17.30	0.14	16.19
NGC 5272, M3	10.20	0.06	15.04
NGC 5466	16.32	0.13	16.06
NGC 5904, M5	7.49	0.05	14.37
NGC 6205, M13	7.53	0.06	14.38
NGC 6341, M92	8.60	0.05	14.67
NGC 7078, M15	10.73	0.14	15.15
NGC 7089, M2	11.62	0.13	15.33

3. Absolute Magnitude Calibration

3.1. Globular Clusters Stars

The procedure to construct an absolute magnitude relation follows previous works (see Section 1), starting with the selection of BHB stars in globular clusters. We used the photometric catalog from An et al. (2008), which provides magnitudes for crowded fields observed by SDSS. The clusters presenting a well-defined horizontal branch were selected and their magnitudes were corrected using the standard extinction ($E(B-V)$) from Schlegel et al. (1998) along with the relative extinctions from Wang & Chen (2019) for the g, u, and r band. Vasiliev & Baumgardt (2021) attributed a membership probability for stars in globular clusters based on proper motions and parallaxes from Gaia EDR3. We selected stars from several globular clusters that were more likely than 0.99 to belong to those clusters and we obtained their absolute magnitude in the SDSS g band (M_g) with the estimated distance for each cluster given by these authors. The list of clusters, their heliocentric distances, and distance moduli are presented in Table 1.

To create the sample used to implement the calibration, we applied the limits for colors as defined for the initial selection (see Section 2.1). Then, the stars were selected in a single combined color-magnitude diagram (CMD), limiting the M_g between −0.15 and 1.15. After this exercise, the remaining globular cluster members were checked individually at the SIMBAD database (Wenger et al. 2000), and those classified as variables, blue stragglers, and other undesirable types were removed. We also excluded stars with flags in the magnitudes used, leaving us with 744 stars to derive the calibrations from.

3.2. Fitting the Horizontal Branch

Finding the best mathematical relationships to fit observable data is not an easy task. In previous works, the absolute magnitudes for BHB stars have been described as a high-degree polynomial (Preston et al. 1991; Deason et al. 2011; Belokurov & Koposov 2016). Instead of arbitrarily assuming that this function is the best representation of the data, we explore the possible combinations between colors and absolute magnitudes. For this task, we employed the `TuringBot` software (Ashok et al. 2021), a code that performs symbolic regression using a simulated annealing algorithm (Chira & Plionis 2019; Delahaye et al. 2019) in order to search for the best set of parameters and mathematical operations to describe the data.

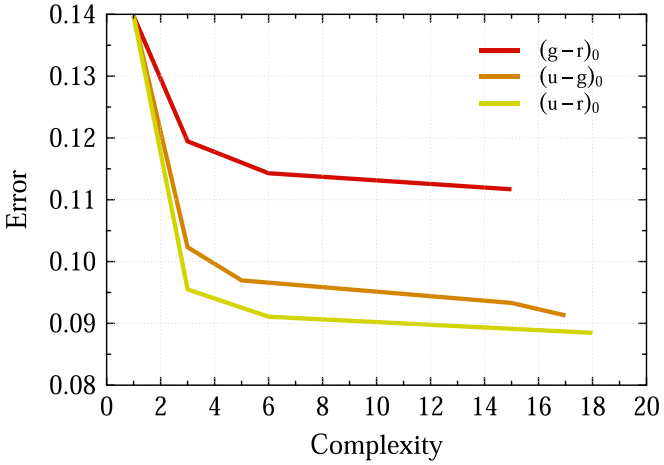

Figure 3. Error comparison between fits for colors $(g-r)_0$, $(u-g)_0$, and $(u-r)_0$ from `TuringBot`.

`TuringBot` is particularly interesting in this case because it allows the visualization of the estimated mathematical laws, allowing the user to choose the most appropriate equations for their needs. Furthermore, the user is free to choose the mathematical operations involved in the fitted functions and the error metric for convergence, as well as the input variables. The best fits are presented in a summarized box, combining the error and the complexity of the equations. The complexity is defined by the sum of the "size" of the mathematical operations, constants, and variables present in the solutions. The program assumes that an input variable, constant, sum, subtraction, and multiplication have size 1 each, division has size 2, and more complex operations have higher sizes.[11] We verified that the use of very complex mathematical operations is unnecessary and does not improve the average error of the equations presented by the software. Hence, we adopted only the basic mathematical operations (sum, subtraction, multiplication, and division) as input for the search for absolute magnitude calibrations. The mean absolute error was used as a criterion for convergence and we tested the dependence of all the most common available observable variables found in the literature for estimates of this type, such as magnitudes, color indices, and metallicity.

After evaluating all the combinations of input variables (u_0, g_0, r_0, $(u-g)_0$, $(u-r)_0$, $(g-r)_0$, and [Fe/H]) presented in Appendix A, we found that there is no significant dependence on metallicity in the calibrations provided by the code, regardless of the mathematical operations adopted and the algorithm convergence time. The colors $(u-g)_0$ and $(u-r)_0$ provided calibrations with smaller errors than the color $(g-r)_0$, traditionally used in the absolute magnitude calibration of BHB stars (D11; Belokurov & Koposov 2016), and also smaller than $(g-r)_0$ with [Fe/H], which means we can achieve more accurate results that do not require metallicity information. The observed improvement with $(u-g)_0$ and $(u-r)_0$ color might be associated with the u filter, whose transmission curve is mostly between 3000 and 4000 Å, i.e., it is positioned in a region of the spectrum where the Balmer discontinuity (∼3645 Å) is located, as well as several hydrogen lines from

[11] More `TuringBot` details can be found in the program documentation, available at https://turingbotsoftware.com/documentation.html.

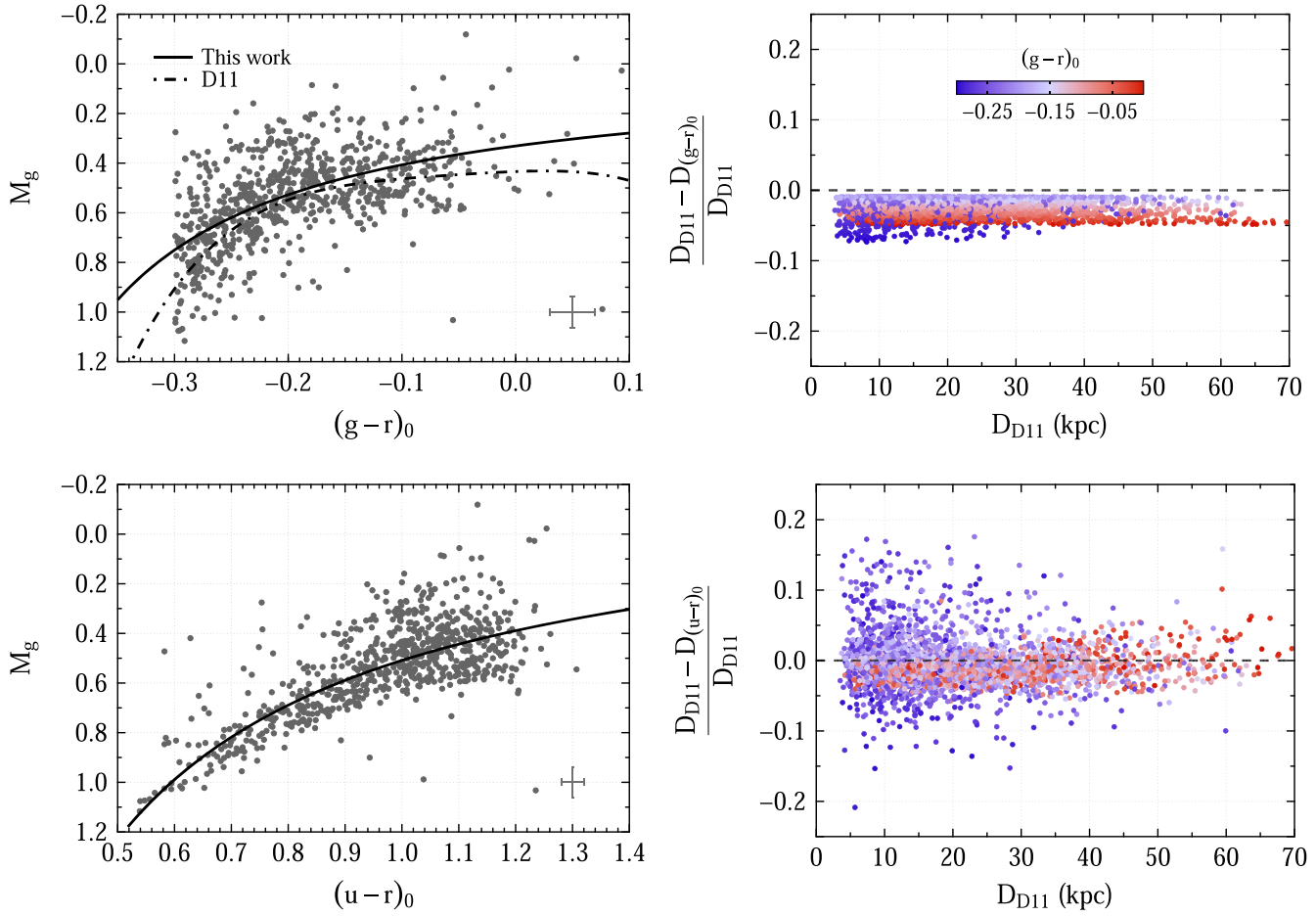

Figure 4. Left panels: color–magnitude diagrams showing BHBs used to define the calibrations. Dashed–dotted line represents the polynomial fit defined in D11. Solid lines represent the calibration for $(g-r)_0$ and $(u-r)_0$ color presented in this work (Equations (1) and (2), respectively, from top to bottom). Median errors of the data are indicated in the bottom right corner of each panel. Right panels: difference between distances calculated with D11's calibration and those presented in this work, respectively, $(g-r)_0$ (top) and $(u-r)_0$ (bottom).

the Balmer series, which makes it a useful indirect indicator of $T_{\rm eff}$ and $\log g$ of the BHBs, atmospheric parameters that are directly linked to the mass of the stars in the horizontal branch (Valcarce & Catelan 2008).

Figure 3 shows the associated errors for each fit in the final BHB sample. Clearly, $(u-r)_0$ presents a better performance than $(g-r)_0$ and $(u-g)_0$. Using the same tool, we find that the relation proposed in D11 is a function of complexity 33 and error = 0.12, while equations of a lower order present a much lower complexity with errors of ∼0.10.

We chose the first functions from which there is no significant decrease in error, i.e., functions of complexity 6 in Figure 3, as those that best describe the data. There exists a singularity in the calibrations; however, it is outside of our color range. Hence, it does not imply an obstacle to their usage in the context of this work:

$$M_g = \frac{0.178}{0.537 + (g-r)_0} \quad (1)$$

$$M_g = \frac{0.721}{(u-r)_0} - 0.212. \quad (2)$$

3.3. Distances Analysis

The left panels in Figure 4 show the distribution of BHB stars in the CMD with colors $(g-r)_0$ (top) and $(u-r)_0$ (bottom). In the top left panel, we can observe how the calibration proposed here (Equation (1)) provides magnitudes lower than D11's, which results in larger distances. The difference is minimal at $(g-r)_0 \sim -0.20$, where both equations come closer, and the smaller values are a consequence of the inclusion of the cluster NGC 7078, whose stars are brighter and were not included in D11. On the other hand, the distribution using $(u-r)_0$ has a lower dispersion (bottom left panel).

In the right panels (Figure 4), we also show the comparison between distances estimated with the relation from D11 and each calibration defined in the present work. For consistence with D11's relation, only stars bluer than $(g-r)_0 = 0$ were considered, and we rejected stars with BHB probabilities of less than 99% to reduce the number of misclassified stars. The new calibration using color $(g-r)_0$ provides distances about 5% larger than D11's for the reddest stars, while the other end of the color window attains a relative difference of up to 9% (top right panel). For the color $(u-r)_0$, the scatter is more uniform and much larger for the bluest stars (bottom right panel).

When comparing with purely astrometric heliocentric distances, there is a considerable scatter, even for stars closer than 5 kpc. For this comparison, we selected stars with relative parallax uncertainty from Gaia EDR3[12] in the interval $0 < \sigma_\varpi/\varpi < 0.2$, renormalized unit weight errors within the

[12] Gaia Collaboration (2020).

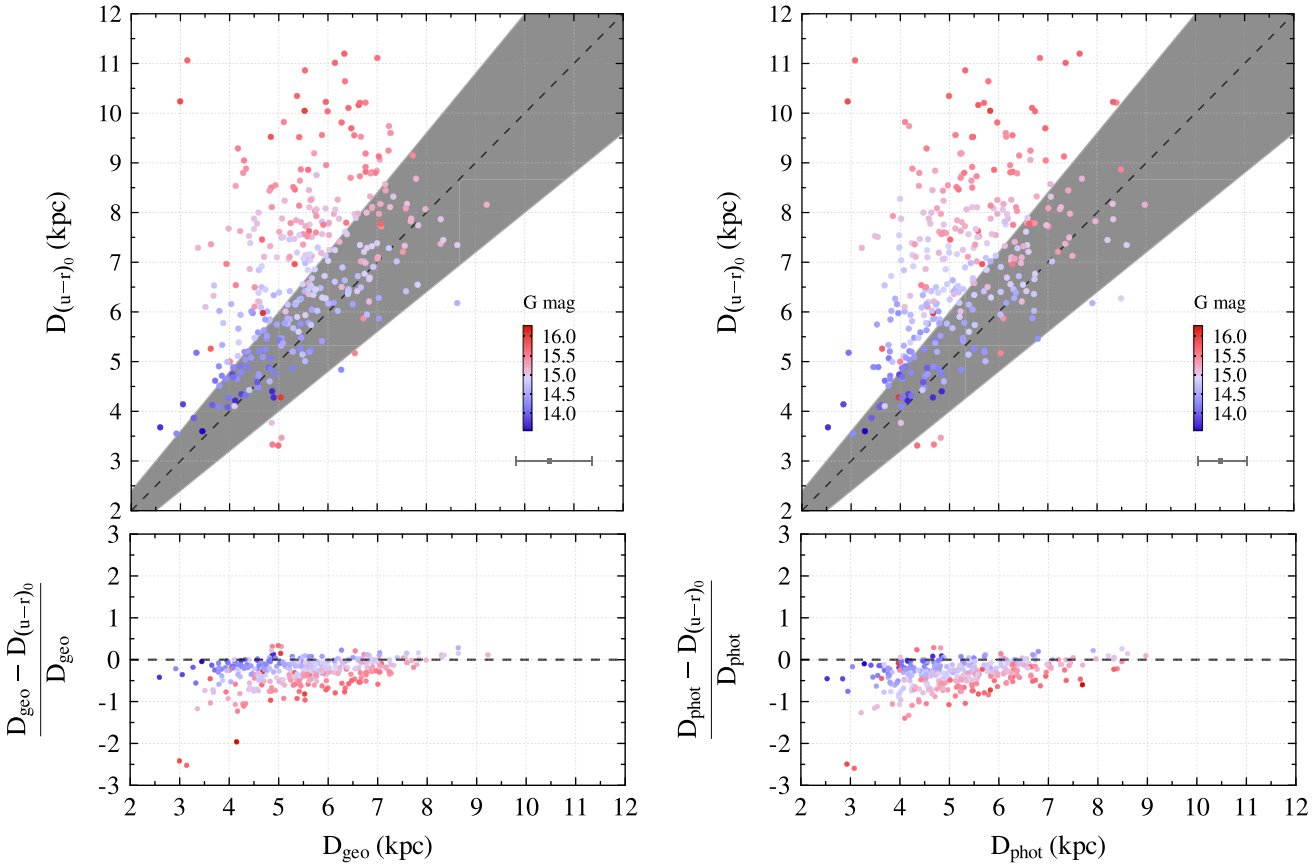

Figure 5. Comparison between the distances estimated by Equation (2) ($D_{(u-r)_0}$), geometric (D_{geo}, left) and photogeometric (D_{phot}, right) distances from Bailer-Jones et al. (2021). Colors indicate Gaia G magnitude and gray area indicates the region of 1σ. Median errors are indicated in the bottom right corner of the top panels.

recommended range (RUWE < 1.4; Lindegren 2018) and also a BHB probability greater than 99% (>300 stars). In Figure 5, we show the comparison between geometric (left) and photogeometric (right) distances provided by Bailer-Jones et al. (2021) and our calibration using $(u-r)_0$ color. For fainter stars, both Bailer-Jones et al.'s (2021) distances are frequently underestimated. The gray region indicates the interval within 20% of distance in the respective horizontal axes, where we find 65% of the stars when using photogeometric estimates and 54% with the purely geometric outside of it. Gaia's parallax measurements potentially are not accurate enough for these BHBs and so the final results are not representative of the sample (since the distances inferred from Bayesian methods are strongly dependent on the measured parallax). We also point out that this effect is unlikely to be the consequence of an inappropriate classification, since 94% of the stars possess $\log g_{ADOP} < 3.6$ and would receive the same label by Santucci et al.'s (2015b) method. Finally, we cannot endorse the compatibility between D11's distances for BHB stars found in the Pristine survey (Starkenburg et al. 2019) and Gaia DR2's parallaxes. As the Pristine data are not publicly available, it is not possible to evaluate whether the difference is due to the BHB sample used.

Similar inconsistencies between photometric and astrometric-inferred distances were also observed by previous works. Using OB stars, Shull & Danforth (2019) noted an increase in the discrepancies at $d > 1.5$ kpc, with B-type stars showing smaller values of distances when considering parallaxes alone. Our A-type stars sample seems to follow this same trend.

4. Discussion and Summary

Since the last absolute magnitude calibration published for BHB stars we had the advent of Gaia data, which allowed us to review the previous relationship thanks to the better characterization of globular clusters (Vasiliev & Baumgardt 2021).

Using data from the SSPP catalog, we obtained a sample of ~5700 BHB stars implementing the GMM algorithm. This new approach is an alternative to the previous individual Gaussian fits, as it is clear that they do not distinguish the latest SSPP $\log g$ distribution properly (Figure 1).

To find which kind of function better describes the distribution of BHBs in the CMD, we used a software for implementing symbolic regression. We suggest two new color–magnitude calibrations based on photometry, including a relation with $(u-r)_0$ color that has not been used before. This calibration provides more accurate estimates than $(g-r)_0$ color in most cases. For the bluest stars, the differences can exceed 10% of nominal values (Equation (2)). However, this difference decreases for redder and more-distant stars. Here, we show that the calibrations can be simpler and achieve an acceptable result, very similar to those from by D11's relation.

We noted substantial differences between photometric and geometric/photogeometric distances. A possibility would be inaccurate estimates of $\log g$ provided by SEGUE, as it was

mainly designed for cool stars ($T_{\rm eff} < 7500$ K). However, Santucci et al. (2015b) showed that this is unlikely to be the case. It could also be due to an incorrect value of extinction for the SDSS photometry, yet it also does not seem to explain the disparity. Most stars were observed in regions of low extinction and we could not find any relation between the extinction values and the inconsistency observed. The observed differences also lead us to believe that measured parallaxes for these stars could be unreliable, as there is no agreement between the distance estimates even for the closest stars. It would be interesting to investigate whether the same discrepancy can be observed with other halo tracers.

The new sample made available here can help to improve results already known about the structures of the Milky Way stellar halo. For example, these BHBs can be used to revisit the duality of the stellar halo or reevaluate the Galaxy's mass estimate.

We thank the anonymous referee for the careful review and all the suggestions, which greatly improved our work. This research was financed with public funds, without which it would not have been possible. F.O.B. acknowledges CAPES (PROEX; Proc. 88887.604787/2021-00). R.M.S. acknowledges CNPq (Proc. 306667/2020-7). S.R. acknowledges partial financial support from FAPESP (Proc. 2015/50374-0 and 2014/18100-4), CAPES, and CNPq. G.L. acknowledges FAPESP (Proc. 2021/10429-0). A.P.-V. acknowledges the DGAPA-PAPIIT grant IA103122. H.D.P. thanks FAPESP (Proc. 2018/21250-9 and 2022/04079-0).

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). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This research has also made use of RStudio (RStudio Team 2022) and TOPCAT (http://www.starlink.ac.uk/topcat/, Taylor 2005).

Appendix A
Fitting the Horizontal Branch with [Fe/H]

For the analysis of dependence of metallicity, we have a smaller sample of stars than the one used for the calibrations; about 40 stars from the original sample were available in the SSPP data. The same procedure was done with both the pure photometric sample and this one crossmatched with SSPP.

Each solution evaluation (considering different input variables) was taken in a period of approximately 10 minutes, which is enough for the convergence of several solutions, as `TuringBot` needs less time to converge than other similar software (Ashok et al. 2021). The hardware involved in this process is highly important for the convergence time. In our case, the program was executed in a computer with an Advanced Micro Devices Ryzen 7 2700 processor, with 16 threads, all used at once. Tests were also made by running the program longer and no significant improvement was observed.

Figure 6 shows how the error decreases with the increase of the complexity of the functions (see Section 3.2 for the definition of "complexity"). Some fits coincide since the program can create colors from magnitudes, if it is better than the magnitudes alone, and not use all the variables provided. This is the reason why the line for all the parameters is not visible. We can see that, for equations up to complexity 25, there is no advantage of adding the metallicity information as we can achieve similar errors using only magnitudes.

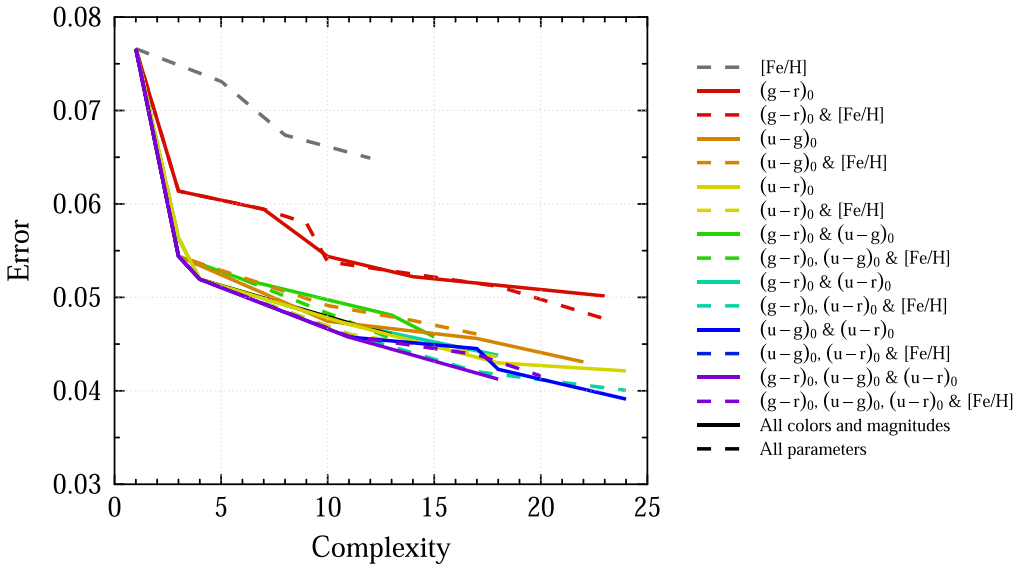

Figure 6. Error comparison between fits for some combinations of magnitudes, colors, and [Fe/H] from `TuringBot`.

Appendix B
Globular Cluster Distances

Figure 7 presents a boxplot for the distances obtained with both calibrations and the one from D11 compared to those provided by Vasiliev & Baumgardt (2021). The size of each box represent the spread (25th and 75th quartiles) and the center line indicates the median value. The number of stars in each cluster is displayed above their identifier from the New General Catalogue (NGC).

The overall results indicate that Equation (2) is in general more accurate than D11's relation. This conclusion is supported by better distance predictions where 10 of the 11 globular clusters were better constrained with Equation (2). The calibration with color $(u-r)_0$ revealed to be more accurate than using the relation provided in this work for color $(g-r)_0$, as nine clusters present lower dispersion and eight clusters show medians close to zero when using the former color. NGC 2416 was the only cluster where the color $(g-r)_0$ presented a better performance; however, this cluster was represented by only nine members and therefore this result may be due to subsampling. The sample considered has a bias to closer cluster, which is apparent by huge distance gap between NGC 2416, the farthest cluster, and NGC 5466, the penultimate. Therefore, we cannot ascertain whether or not the accuracy of distances estimated using color $(u-r)_0$ varies with the distance.

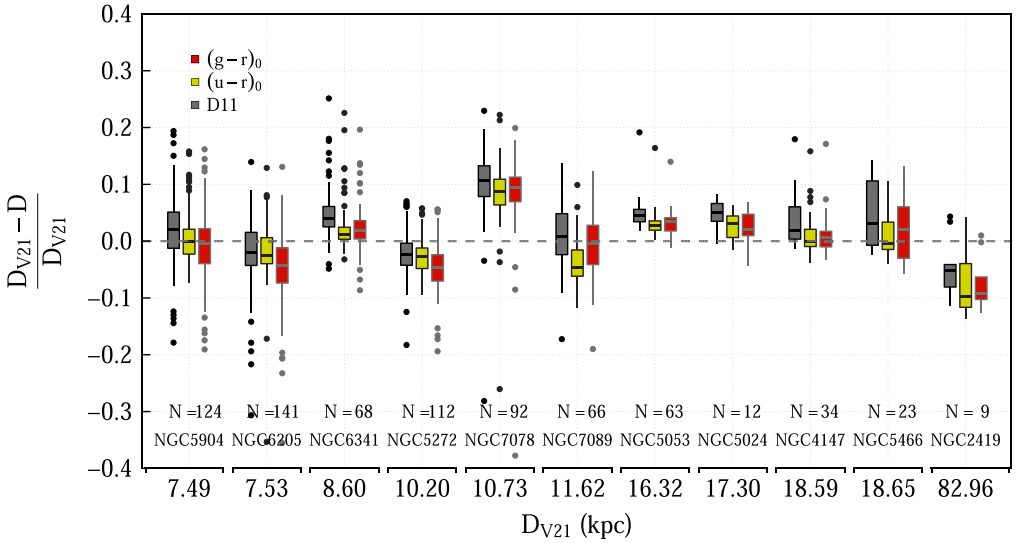

Figure 7. Boxplot for relative distance difference using each calibration presented in this work (Equations (1) and (2)) and that from D11. D_{V21} is the distance provided by Vasiliev & Baumgardt (2021), inferred from Gaia EDR3 data. The number of data points in each cluster is displayed at the bottom of the panel.

ORCID iDs

Fabrícia O. Barbosa https://orcid.org/0000-0002-8262-2246
Rafael M. Santucci https://orcid.org/0000-0002-7529-1442
Silvia Rossi https://orcid.org/0000-0001-7479-5756
Guilherme Limberg https://orcid.org/0000-0002-9269-8287
Angeles Pérez-Villegas https://orcid.org/0000-0002-5974-3998
Hélio D. Perottoni https://orcid.org/0000-0002-0537-4146

References

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29
An, D., Johnson, J. A., Clem, J. L., et al. 2008, ApJS, 179, 326
Ashok, D., Scott, J., Wetzel, S., Panju, M., & Ganesh, V. 2021, in Proc. AAAI Conf. on Artificial Intelligence, 35, 15753
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147
Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611
Belokurov, V., & Koposov, S. E. 2016, MNRAS, 456, 602
Belokurov, V., Koposov, S. E., Evans, N. W., et al. 2014, MNRAS, 437, 116
Bird, S. A., Xue, X.-X., Liu, C., et al. 2022, MNRAS, 516, 731
Brown, A. G. A. 2021, ARA&A, 59, 59
Cacciari, C. 1999, in ASP Conf. Ser. 167, Harmonizing Cosmic Distance Scales in a Post-HIPPARCOS Era, ed. D. Egret & A. Heck (San Francisco, CA: ASP), 140
Chira, M., & Plionis, M. 2019, MNRAS, 490, 5904
Clewley, L., Warren, S. J., Hewett, P. C., et al. 2002, MNRAS, 337, 87
Deason, A. J., Belokurov, V., & Evans, N. W. 2011, MNRAS, 416, 2903
Deason, A. J., Belokurov, V., Evans, N. W., et al. 2012, MNRAS, 425, 2840
Deason, A. J., Belokurov, V., Koposov, S. E., & Lancaster, L. 2018, ApJL, 862, L1
Deason, A. J., Belokurov, V., Koposov, S. E., et al. 2017, MNRAS, 470, 1259
Deason, A. J., Erkal, D., Belokurov, V., et al. 2021, MNRAS, 501, 5964
Delahaye, D., Chaimatanan, S., & Mongeau, M. 2019, in Handbook of Metaheuristics, ed. M. Gendreau & J.-Y. Potvin (Berlin: Springer), 1
Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, J. R. Stat. Soc. B, 39, 1
Donlon, T. I., Newberg, H. J., Sanderson, R., & Widrow, L. M. 2020, ApJ, 902, 119
Erkal, D., Deason, A. J., Belokurov, V., et al. 2021, MNRAS, 506, 2677
Fermani, F., & Schönrich, R. 2013, MNRAS, 430, 1294
Gaia Collaboration 2020, Gaia Source Catalogue EDR3, IPAC, doi:10.26131/IRSA541
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Hayes, D. S., & Philip, A. G. D. 1979, PASP, 91, 71
Helmi, A. 2020, ARA&A, 58, 205
Koppelman, H., Helmi, A., & Veljanoski, J. 2018, ApJL, 860, L11
Lancaster, L., Koposov, S. E., Belokurov, V., Evans, N. W., & Deason, A. J. 2019, MNRAS, 486, 378
Leavitt, H. S., & Pickering, E. C. 1912, HarCi, 173, 1
Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008a, AJ, 136, 2022
Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008b, AJ, 136, 2050
Li, T. S., Ji, A. P., Pace, A. B., et al. 2022, ApJ, 928, 30
Lindegren, L. 2018, http://www.rssd.esa.int/doc_fetch.php?id=3757412
Malhan, K., Ibata, R. A., & Martin, N. F. 2018, MNRAS, 481, 3442
Martin, N. F., Venn, K. A., Aguado, D. S., et al. 2022, Natur, 601, 45
Myeong, G. C., Evans, N. W., Belokurov, V., Sanders, J. L., & Koposov, S. E. 2018, ApJL, 856, L26
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Peñarrubia, J., & Petersen, M. S. 2021, MNRAS, 508, L26
Petersen, M. S., & Peñarrubia, J. 2021, NatAs, 5, 251
Pier, J. R. 1983, ApJS, 53, 791
Preston, G. W., Shectman, S. A., & Beers, T. C. 1991, ApJ, 375, 121
Rockosi, C. M., Sun Lee, Y., Morrison, H. L., et al. 2022, ApJS, 259, 60
RStudio Team 2022, RStudio: Integrated Development Environment for R, RStudio, PBC, Boston, MA, http://www.rstudio.com/
Santucci, R. M., Beers, T. C., Placco, V. M., et al. 2015a, ApJL, 813, L16
Santucci, R. M., Placco, V. M., Rossi, S., et al. 2015b, ApJ, 801, 116
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Sersic, J. L. 1968, Atlas de Galaxias Australes (Observatorio Astronomico: Universidad Nacional de Cordoba)
Shapley, H. 1916, ApJ, 43, 217
Shull, J. M., & Danforth, C. W. 2019, ApJ, 882, 180
Sirko, E., Goodman, J., Knapp, G. R., et al. 2004, AJ, 127, 899
Starkenburg, E., Youakim, K., Martin, N., et al. 2019, MNRAS, 490, 5757
Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
Thomas, G. F., McConnachie, A. W., Ibata, R. A., et al. 2018, MNRAS, 481, 5223
Utkin, N. D., & Dambis, A. K. 2020, MNRAS, 499, 1058
Valcarce, A. A. R., & Catelan, M. 2008, A&A, 487, 185
Vasiliev, E., & Baumgardt, H. 2021, MNRAS, 505, 5978
Vickers, J. J., Grebel, E. K., & Huxor, A. P. 2012, AJ, 143, 86
Wang, S., & Chen, X. 2019, ApJ, 877, 116
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
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. 2022, ApJ, 924, 23
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
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
York, D. G., Adelman, J., Anderson, J. E. J., et al. 2000, AJ, 120, 1579
Yuan, Z., Chang, J., Beers, T. C., & Huang, Y. 2020, ApJL, 898, L37
Yuan, Z., Malhan, K., Sestito, F., et al. 2022, ApJ, 930, 103
Yuan, Z., Smith, M. C., Xue, X.-X., et al. 2019, ApJ, 881, 164