BSEC Method for Unveiling Open Clusters and its Application to Gaia DR3: 83 New Clusters (Postprint)

Authors: Zhong-Mu Li and Cai-Yan Mao

Date: 2024-05-24T00:00:00+00:00

Abstract

Open clusters (OCs) are common in the Milky Way, but most of them remain undiscovered. There are numerous techniques, including some machine-learning algorithms, available for the exploration of OCs. However, each method has its limitations and therefore, different approaches to discovering OCs hold significant values. We develop a comprehensive approach method to automatically explore the data space and identify potential OC candidates with relatively reliable membership determination. This approach combines the techniques of Hierarchical Density-Based Spatial Clustering of Applications with Noise, Gaussian mixture model, and a novel cluster member identification technique, color excess constraint. The new method exhibits efficiency in detecting OCs while ensuring precise determination of cluster memberships. Because the main feature of this technique is to add an extra constraint (EC) for the members of cluster candidates using the homogeneity of color excess, compared to typical blind search codes, it is called Blind Search-Extra Constraint (BSEC) method. It is successfully applied to the Gaia Data Release 3, and 83 new OCs are found, whose color–magnitude diagrams (CMDs) are fitted well to the isochrones. In addition, this study reports 621 new OC candidates with discernible main sequence or red giant branch. It is shown that BSEC technique can discard some false negatives of previous works, which takes about three percentage of known clusters. It shows that as an EC, the color excess (or two-color) constraint is useful for removing fake cluster member stars from the clusters that are identified from the positions and proper motions of stars, and getting more precise CMDs, when differential reddening of member stars of a cluster is not large (e.g., ΔE(GBP $-$ GRP) < 0.5 mag). It makes the CMDs of 15% clusters clearer (in particular for the region near turnoff) and therefore is helpful for CMD and stellar population studies. Our result suggests that the color excess constraint is more appropriate for clusters with small differential reddening, such as globular

clusters or older OCs, and clusters that the distances of member stars cannot be determined accurately.

Full Text

Preamble

ChinaXiv Research in Astronomy and Astrophysics, 24:055014 (16pp), 2024 May

https://doi.org/10.1088/1674-4527/ad3a2b

BSEC Method for Unveiling Open Clusters and its Application to Gaia DR3: 83 New Clusters

Zhong-Mu Li and Cai-Yan Mao
Institute of Astronomy and Information, Dali University, Dali 671003, China; zhongmuli@126.com

Received 2023 December 30; revised 2024 March 25; accepted 2024 March 28; published 2024 May 13

Abstract

Open clusters (OCs) are common in the Milky Way, but most of them remain undiscovered. Numerous techniques, including some machine-learning algorithms, are available for exploring OCs, yet each method has its limitations. Consequently, different approaches to discovering OCs hold significant value. We develop a comprehensive method to automatically explore the data space and identify potential OC candidates with relatively reliable membership determination. This approach combines Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), Gaussian mixture model (GMM), and a novel cluster member identification technique called color excess constraint. The new method exhibits efficiency in detecting OCs while ensuring precise determination of cluster memberships. The main feature of this technique is to add an extra constraint (EC) for cluster candidate members based on the homogeneity of color excess. Compared to typical blind search codes, when successfully applied to Gaia Data Release 3, the method finds 83 new OCs whose color–magnitude diagrams (CMDs) fit well to isochrones. In addition, this study reports 621 new OC candidates with discernible main sequence or red giant branch. The BSEC technique can discard some false negatives from previous works, accounting for about three percent of known clusters. As an EC, the color excess (or two-color) constraint is useful for removing fake cluster member stars identified from positions and proper motions, yielding more precise CMDs when differential reddening of member stars is not large (e.g., $\Delta E(GBP - GRP) < 0.5$ mag). This makes the CMDs of 15% of clusters clearer (particularly near the turnoff region) and is therefore helpful for CMD and stellar population studies. Our results suggest that the color excess constraint is most appropriate for clusters with small differential reddening, such as globular clusters or older OCs, and for clusters where member star distances cannot be determined accurately.

Key words: Galaxy: stellar content – (Galaxy:) open clusters and associations: general – stars: fundamental parameters

1. Introduction

Open clusters (OCs) are widely recognized as valuable laboratories for studying stellar evolution and as tracers of Milky Way structure. They serve as excellent subjects for investigating stellar evolution because stars with the same metallicity and age occupy different evolutionary stages along an isochrone. Additionally, they provide valuable insights into Galactic structure through the observation and measurement of numerous distant OCs and their respective stellar properties. Consequently, extensive efforts have been dedicated to searching for OCs using diverse datasets, resulting in a significant increase in the number of identified OCs following the release of Gaia data (Gaia Collaboration et al. 2016, 2018, 2021, 2023) in recent years. The number of well-known OCs now exceeds 3000 (see, e.g., Castro-Ginard et al. 2022; Hunt & Reffert 2023). However, this is still significantly lower than the predicted value (>10,000; Minniti 2023). Besides limitations in observational data, the technique for searching for OCs plays an important role.

Notably, following the substantial increase in OC sample size achieved by Mermilliod (1995) (1200 OCs), the largest increase in the number of Galactic OCs came with the application of new search methods. For example, the rise in OC numbers in the catalog of Kharchenko et al. (2013) (hereafter MWSC) can be attributed to both Gaia data releases and advancements in algorithms designed for identifying clusters. Many studies (Castro-Ginard et al. 2018, 2019, 2020, 2022; Liu & Pang 2019; Sim et al. 2019; Cantat-Gaudin et al. 2021; He et al. 2022; Hao et al. 2019; Li et al. 2021, 2022a, 2022b; Hunt & Reffert 2023; Qin et al. 2023) have reported the discovery of over a few thousand candidate OCs based on Gaia data. Some famous unsupervised machine learning algorithms—such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al. 1996), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (Campello et al. 2013), Gaussian mixture model (GMM) (Dempster et al. 1977), Unsupervised Photometric Membership Assignment in Stellar Clusters (UPMASK) (Krone-Martins & Moitinho 2014; Pera et al. 2021), and friend-of-friend (FoF) (Yang et al. 2005)—have been employed for blind searches in the Gaia dataset to effectively identify OCs. However, none of these algorithms is flawless in terms of both efficiency and precision, as highlighted by previous studies (see, e.g., Hunt & Reffert 2021). During testing, some known clusters were not detected while fake clusters were occasionally reported. Each method has its own advantages and disadvantages. According to some tests, GMM is good for determining cluster membership in small regions, whereas HDBSCAN is more suitable for large-scale OC searches (Hunt & Reffert 2021). DBSCAN is also effective for identifying OCs with similar density (Castro-Ginard et al. 2019). FoF seems suitable for large-scale blind search, but it often reports false positives, as does HDBSCAN. If membership probabilities are required, GMM and UPMASK can be utilized accordingly.

To develop an efficient and relatively precise method for searching OCs, we conducted this study. The work aims to build a new star cluster search method, Blind Search-Extra Constraint (BSEC), which leverages the advantages of two kinds of existing methods and introduces a new cluster member constraint.

The structure of this paper is as follows. In Section 2 we introduce the design of the new method. In Section 3 we apply the new method to Gaia DR3 data (Gaia Collaboration et al. 2021). In Section 4 we show the crossmatch results and color–magnitude diagrams (CMDs) of new OCs. Finally, in Section 5 we conclude this work.

2. Design of BSEC Method

When designing the BSEC method, we combined the advantages of several different approaches. This is achieved through the following steps. First, an effective algorithm is employed to find as many star cluster candidates as possible. HDBSCAN and FoF are ideal for this purpose. Second, a re-identification algorithm is used to re-identify the OC candidates and remove fake member stars (i.e., field stars). This step makes the cluster catalog and member stars more reliable, which is important for other works. Finally, an Extra Constraint (EC) algorithm is applied to constrain the member stars again based on well-known properties of star clusters. In this work, we used the knowledge that all member stars of a cluster have similar color excess, in addition to similar spatial coordinates and proper motions. When stars have the same color excess and are not too red, regardless of metallicity and age, they will distribute along a fixed curve in a color–color diagram (CCD) and along a curve in a CMD. Figure 1 shows the CCD of all stars in the PARSEC-COLIBRI isochrone database (Marigo et al. 2017). These stars cover large ranges in metallicity ($0.0002 \leq Z \leq 0.05$) and age (3.98 Myr $\leq$ age ≤ 13.2 Gyr). We see clearly that when stellar colors are not too red (e.g., $\mathrm{GBP} - \mathrm{G} \leq 3.0$ or $\mathrm{GBP} - \mathrm{GRP} \leq 2.0$), the color–color relation (CCR) can be described well using a polynomial function with an uncertainty of about 0.05 mag. For most redder stars, their CCR can also be described by the same fitting function. As we see, the fitted CCR is almost independent of stellar mass, age, and metallicity. There is a fixed fitting function for a group of stars if they have the same color excess, although the fitting formulae differ for various color excesses. This is very much the case for a star cluster, because member stars have the same color excess for any fixed color. We can therefore apply the CCR to provide a new constraint on cluster member stars. Stars that distribute near the CCR of a cluster are possible members, while those far from the CCR are not. Any two colors can be used to build the CCR of a cluster. In this work, we take (GBP − G) and (GBP − GRP) to build the CCRs. For a fixed cluster observed by the Gaia satellite, the CCR is similar to those shown in Figure 2. Note that the observed CCR differs from that of PARSEC-COLIBRI stellar isochrones because of color excesses in the two colors. We can eliminate stars that are obviously distant from the CCR curve, as they cannot be cluster members. This technique is called color excess constraint or two-color constraint, and it constitutes the first part of the EC for star clusters. Our experiments show that the two-color constraint eliminates many fake clusters or stars with large photometric uncertainties.

However, the CCR is affected by differential reddening, which may result in significant errors. To test how differential reddening affects the CCR, we calculated CCRs with maximum differential reddening values ΔE(GBP − GRP) from 0.1 to 1.5 mag. Figure 3 shows two examples with maximum differential reddening values of 0.5 and 1.0 mag. In the figure, purple squares are stars with random differential reddening ΔE(GBP − GRP) lower than 0.5 mag. The orange line is the fitting relation for these stars, while the blue dashed lines denote an error of 0.05 mag around the fitting curve. We can see that almost all purple squares lie within the range bounded by the blue lines, indicating that the CCR is only slightly affected by differential reddening. Gray dots represent stars with differential reddening smaller than 1.0 mag, showing that the color dispersion is significantly larger than in the case of 0.5 mag differential reddening. Table 1 lists the errors of the best-fitting CCRs for different values of maximum differential reddening. We observe that the fitting error increases with differential reddening. It can be concluded that CCR is an effective method for constraining member stars when differential reddening is not large (e.g., ΔE(GBP − GRP) $<$ 0.5 mag). Therefore, it is more suitable for studying globular clusters and OCs older than 0.3 Gyr because of the scarcity of gas and dust. In other words, the color excess constraint may not be appropriate for star clusters with large differential reddening, where there is an obvious difference between the reddening of cluster member stars and other stars.

According to the CMDs of the 83 newly discovered OCs in this work, the color spread of main sequence stars near the turnoff is not larger than about 0.6 mag (including a spread of 0.1 mag caused by binaries), and that of red giant stars is lower than 0.5 mag. This suggests that the differential reddening of these OCs is likely smaller than 0.5 mag, so the CCR can be used to provide an EC on the member stars of these clusters.

3. Application to Gaia DR3

3.1. Blind Search of OC Candidates

3.1.1. Data This work uses data from the latest Gaia release, Gaia DR3 (Gaia Collaboration et al. 2023), which contains astrometry and broadband photometry already published as part of Gaia EDR3 (Gaia Collaboration et al. 2021). This release provides more stars and more accurate astrometric and photometric data compared to the first and second releases (DR1 and DR2) (Gaia Collaboration et al. 2016, 2018). No cuts are applied to the observational data because we need to find as many OC candidates as possible. In total, the catalog contains relatively accurate astrometry and photometry for more than 1.5 billion sources (Damljanović 2021; Forveille & Kotak 2021; Gaia Collaboration et al. 2021, 2023). As is well known, Gaia data have been widely used in star cluster studies.

3.1.2. Algorithm for Blind Search The well-known cluster search algorithm HDBSCAN is used for blind search of OC candidates. HDBSCAN is a clustering algorithm that can find cluster candidates across various densities, making it possible to identify small clusters containing tens of member stars, which is suitable for this work. A minimum cluster size mclSize is set to 25, rather than the suggested value of 10, because mclSize = 10 results in too many cluster candidates—far more than previous findings. The leaf mode rather than the default mode (Excess of Mass, i.e., EoM) is used because this mode is more effective. In addition, two candidates within 0°.5 are combined in the cluster search because the largest OCs can reach sizes of about 0°.5. However, this method may incorrectly combine two clusters at different distances, so we check the results with parallax distributions later. Because this is a test, we take each original data file of Gaia DR3 as a grid for blind search. Although this is somewhat rough, it does not obviously affect the final result, as each file covers a sufficiently large coordinate space.

3.1.3. Result The blind search with HDBSCAN reports 6908 cluster candidates. This number is similar to that in a recent work by Hunt & Reffert (2023). However, as pointed out in some papers (e.g., Hunt & Reffert 2023), there are possibly many false positive identifications in HDBSCAN results. In addition, the CMDs of some candidates appear strange or contain too many stars (more than 1 million) because no cuts were applied to the observational data. This makes the results insufficiently reliable and impossible to use directly for other research. Therefore, a re-identification based on the HDBSCAN results must be carried out.

3.2. Re-identification of Member Stars

3.2.1. GMM Re-identification The GMM algorithm is used to re-identify cluster candidates because it is powerful for detecting cluster members and can provide membership probabilities. Here, some cuts are applied to the HDBSCAN results for each cluster candidate. Specifically, stars are constrained to be brighter than 21 mag, with parallax between 0 and 7 mas and proper motions (_α^* and _δ) between -50 and 50 mas yr^{-1}. To make the results more reliable, only stars around the mean values of proper motions, parallax, and coordinates are selected for GMM re-identification. This eliminates most field stars and makes the re-identification much faster and more effective. After re-identification, we can select member stars of a cluster according to their probabilities. As examples, Figures 4 and 5 show the distributions of member stars after HDBSCAN and GMM processes in coordinate and CMD spaces. We see that CMDs become closer to stellar population isochrones when using larger probability cuts (e.g., 0.5 and 0.9; examples 2 and 3 in Figure 5). In this work, we take a probability cut of 0.9 to ensure reliable member stars. Cluster candidates with preferable CMDs—those that include at least a main sequence or red giant branch—are studied here because their CMDs can be compared to stellar isochrones (e.g., Marigo et al. 2017), allowing the age and/or metallicity of such candidates to be determined. These candidates are more likely to be real star clusters. As a result, the GMM re-identification yields 5411 cluster candidates with more than 20 member stars, of which 1166 have preferable CMDs.

3.2.2. Extra Constraint Because the main structure of some cluster CMDs (e.g., main sequence turnoff and red giant branch) remains unclear and the astrometric parameters of member stars of some candidates are dispersed, an EC constraint is applied to the GMM re-identification results. This process is based on member stars with GMM probabilities larger than 0.9. As the first part of EC, member stars are checked using color excess information via the CCR (two-color constraint). Colors (GBP − GRP) and (GBP − G) are used for this work. As shown in Figure 2, only stars within the possible range are taken as cluster members. An uncertainty of 0.05 mag is adopted because fitting the CCR of stars in the PARSEC-COLIBRI isochrone database shows that almost all stars distribute within this range if they are not too red ($GBP - G \leq 3.0$; see Figure 1). Note that the four examples shown in Figure 2 are chosen randomly and therefore are not named. This constraint eliminates some stars that are possibly not cluster members but does not change the spatial distribution of stars (Figure 4). CMDs become clearer after the two-color constraint, as seen in Figure 5, which makes CMD fitting of clusters more reliable. We see that some faint stars are removed from cluster members by the two-color constraint, likely due to their relatively large photometric uncertainties. In particular, some fake blue stragglers (in the upper left of the turnoff) are eliminated by the two-color constraint (examples 2 and 3). This is helpful for both CMD fitting studies and blue straggler studies of star clusters. Note that there is no clear understanding of how the CCR of blue stragglers differs from normal stars, but according to a test based on data from Cummings & Kalirai (2018), blue stragglers obey the same fitting CCR (U − B versus B − V) as all stars when photometric uncertainties and differential reddening are taken into account.

This constraint is also helpful for obtaining the real red giant branch (see the upper right of example 2). Because the position of the red giant branch in the CMD is sensitive to metallicity, an accurate red giant branch is important for determining cluster metallicities. Thereby, the color excess constraint is useful for obtaining reliable metallicities, as it helps to obtain a more reliable red giant branch. The CMDs of 15% of clusters in our work become clearer after the two-color constraint. However, the CMDs of many cluster candidates are still not similar to stellar population isochrones. Such candidates may not be real clusters. We therefore check the CMDs of all candidates as the second part of EC. The G versus (GBP − GRP) CMD is used here. A visual inspection is finally applied to the CMDs of cluster candidates to identify those with clear CMDs. Additionally, the final clusters are checked via proper motion dispersions using the methods of Cantat-Gaudin & Anders (2020) and Hao et al. (2022). Equation (1) is used to judge whether a candidate is a real cluster. This equation is adopted because newly found clusters have parallax less than about 1 mas and proper motion less than about 0.5 mas yr^{-1}.

4. Crossmatch to Known Clusters and Newly Found OCs

When we crossmatch the OC candidates to many catalogs of known OCs or candidates (e.g., Bica et al. 2001, 2003; Dias et al. 2002, 2021; Reylé & Robin 2002; Chen et al. 2003; Porras et al. 2003; Frinchaboy et al. 2004; Glushkova 2010; Kharchenko et al. 2013; Schmeja et al. 2014; Camargo et al. 2015, 2016a, 2016b; Castro-Ginard et al. 2018, 2019, 2020, 2022; Scholz et al. 2015; Kos et al. 2018; Ryu & Lee 2018; Ferreira et al. 2019, 2021; Kounkel & Covey 2019; Cantat-Gaudin et al. 2020; Hao et al. 2021, 2022; He et al. 2021, 2022a, 2022b; Hunt & Reffert 2021; Li et al. 2022; Chi et al. 2023b; Li & Mao 2023; Qin et al. 2023) and globular clusters, 2941 known clusters or candidates (e.g., NGC 5904, NGC 6005, NGC 2420, NGC 6830, NGC 4590, NGC 6584, NGC 6819, NGC 2627, Alessi 6, UBC 285, King 6, and Gulliver 59) are matched within the 5411 cluster candidates of this work. Among the 1166 candidates with preferable CMDs, 621 with rough CMDs (which include at least a main sequence or red giant branch and have somewhat large dispersion) are found to be new. In particular, 83 with good CMDs (which have small dispersion and are similar to isochrones in the PARSEC-COLIBRI database) are found to be new.

Because the distributions of position, proper motion, color excess, and membership probability of member stars of each cluster candidate have been constrained in previous steps, the CMD can be employed to impose an additional constraint to identify real clusters. Thus, the 83 candidates with good CMDs are more likely to be real clusters, and their stellar populations can possibly be determined from CMDs. We take them as new OCs. The new clusters are judged not only by their good CMDs but also by the distributions of position, proper motion, color excess, and membership probability. Note that if the distance between the centers of a candidate and a known cluster is less than 0.5 deg, the candidate is considered known (or matched). Tables 2 and 3 list the numbers of candidates from this work that are matched to known star clusters or candidates in previous catalogs, for all candidates and those with preferable CMDs, respectively. Figure 6 shows the distribution of matched and unmatched known clusters and candidates, together with all newly found candidates with preferable CMDs and clusters with good CMDs in the Galactic coordinate system. We see clearly that the newly found clusters and known clusters do not overlap. Table 4 shows known clusters that are not recovered by this work, while Tables 5 and 6 list astrometric parameters of newly found OCs with good CMDs.

Meanwhile, Figure 7 shows the distributions of Galactic longitudes and latitudes of matched and unmatched known clusters. We observe that most matched clusters distribute around l = 0 ± 75 deg and b = 0 ± 10 deg. The number of clusters is sensitive to the input parameter of HDBSCAN, mclSize. If a smaller value such as 20 or 10 is used for searching cluster candidates, many more clusters will be reported, and more known clusters will be matched.

As mentioned above, clusters with similar R.A. and decl. but different may be incorrectly combined into one cluster, so we check our results via the parallax of member stars in each cluster. The parallaxes of stars are divided into 5 or 10 bins according to the number of cluster members. We find that none of the parallax distributions of the 83 new clusters exhibits a clearly separate bimodal distribution. This can also be checked via the distribution of member stars in the parallax versus proper motion space (Figure 8). This suggests that the member stars of each cluster distribute in a concentrated area. Therefore, none of the 83 new clusters includes two physically separated clusters.

To understand the properties of the newly discovered OCs, we present their distribution in different spaces. Figure 8 shows the distribution of some example newly found OCs in coordinate, proper motion, proper motion versus parallax, and CMD spaces. Here we plot stars in the parallax versus proper motion space to check whether the stars can be divided into two groups via parallax and to examine the relation between the two parameters. As we see, most OCs with clear CMDs show clustering features in these spaces. Figures 9–11 present the proper motion and parallax distributions of newly found OCs (two samples for those with rough CMDs and good CMDs), together with those of the CG sample (Castro-Ginard et al. 2018, 2019, 2020, 2022). We observe that the distributions of proper motion $\_\alpha^*$ and parallax differ among the three samples, indicating that the fraction of distant OCs is larger in the sample of newly found OCs compared to the CG sample. Thus, BSEC is able to find some distant OCs.

4.1. CMDs of OCs

CMDs are important for studying OCs. Many fundamental cluster parameters such as color excess, metallicity, age, and binary fraction can be determined from CMDs. Although the CMDs of the 83 newly found OCs are relatively clear, only some include both main sequence and red giant branch. Some clusters have only a main sequence or red giant branch. Although their parallaxes suggest they are Milky Way clusters, there are actually obvious uncertainties in the parallaxes. Figures 12–14 present CMDs of the newly found OCs. Fundamental parameters such as distance modulus, metallicity, and age can be determined by comparing the CMDs to theoretical stellar population isochrones. Detailed CMD fitting of these clusters will be carried out and reported in another paper using the stellar population model and CMD fitting code of Li et al. (2017, 2021).

5. Conclusion

This paper presents a new composite method, BSEC, for hunting star clusters. The main feature of BSEC is that cluster members are constrained not only by proper motions and spatial coordinates but also by color excess and CMD shape. The fitted curve of the CCR is used to constrain member stars of a cluster when the maximum differential reddening is not too large. This color excess constraint technique significantly improves cluster membership determination and helps obtain clearer CMDs for about 15% of clusters. However, the constraint cannot be used independently for cluster identification; it is better employed as an EC for cluster members.

Moreover, the EC technique is probably useful for star clusters where member star distances cannot be determined accurately. It can be used to identify background and foreground stars, as their reddening is usually different from cluster members. The BSEC method is then applied to Gaia DR3 data, finally discovering 83 new OCs with CMDs similar to stellar population isochrones and 621 new candidates with rough CMDs. The spatial distributions in the R.A. versus decl. space of a few cluster candidates are not circular because of limitations in the observed files used for the study, but this does not affect the conclusion because this work aims to introduce the new BSEC method and color excess constraint. Moreover, taking the number of known real clusters as 3000, the BSEC method discards some false negatives from previous works, accounting for about 3% of known clusters. The membership and CMDs of these clusters are obtained, and their distributions are studied.

Our results show a larger fraction of distant clusters compared to previous work. The results can be used in many future studies, particularly CMD studies. We conclude that BSEC is a useful method for cluster identification that helps obtain more precise CMDs, especially for clusters with small differential reddening. Although HDBSCAN and GMM blind search algorithms are used in this work, they can be replaced by other algorithms. In fact, the results are somewhat sensitive to some adjustable parameters, which contributes to some clusters not being recovered by this work. The method of EC can also be developed further.

Acknowledgments

We appreciate the constructive comments of the referee and thank Mr. Tao Xia and Ms. Yangyang Deng for their help. This work has been supported by Yunnan Academician Workstation of Wang Jingxiu (202005AF150025), China Manned Space Project (NO. CMS-CSST-2021-A08), and Guanghe Project (gh-fund202302019167).

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.

Data Availability: The data underlying this article can be downloaded from Zenodo (DOI:10.5281/zenodo.10154725).

References

Bica, E., Dutra, C. M., & Barbuy, B. 2003, A&A, 397, 177

Bica, E., Santiago, B. X., Dutra, C. M., et al. 2001, A&A, 366, 827

Camargo, D., Bica, E., & Bonatto, C. 2016a, MNRAS, 455, 3126

Camargo, D., Bica, E., & Bonatto, C. 2016b, A&A, 593, A95

Camargo, D., Bica, E., Bonatto, C., et al. 2015, MNRAS, 448, 1930

Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, Adv. Knowl. Discovery Data Mining, 7819, 160

Cantat-Gaudin, T., & Anders, F. 2020, A&A, 633, A99

Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1

Cantat-Gaudin, T., Krone-Martins, A., Sedaghat, N., et al. 2019, A&A, 624, A126

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2018, A&A, 618, A59

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2019, A&A, 627, A35

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2020, A&A, 635, A45

Castro-Ginard, A., Jordi, C., Luri, X., et al. 2022, A&A, 661, A118

Chen, L., Hou, J. L., & Wang, J. J. 2003, AJ, 125, 1397

Chi, H., Wang, F., Wang, W., et al. 2023a, ApJS, 266, 36

Chi, H., Wei, S., Wang, F., et al. 2023b, ApJS, 265, 20

Cummings, J. D., & Kalirai, J. S. 2018, AJ, 156, 165

Damljanović, G. 2021, POBeo, 100, 75

Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, J. Roy. Stat. Soc.: Ser. B (Methodological), 39, 1

Dias, W. S., Alessi, B. S., Moitinho, A., et al. 2002, A&A, 389, 871

Dias, W. S., Monteiro, H., Moitinho, A., et al. 2021, MNRAS, 504, 356

Ester, M., Kriegel, H. P., & Xu, X. 1996, KDD-96 Proc. 6th Int. Conf. Knowledge Discovery and Data Mining (Portland, OR: AAAI Press)

Ferreira, F. A., Corradi, W. J. B., Maia, F. F. S., et al. 2021, MNRAS, 502, L90

Ferreira, F. A., Santos, J. F. C., Corradi, W. J. B., et al. 2019, MNRAS, 483, 5508

Forveille, T., & Kotak, R. 2021, A&A, 649, E1

Frinchaboy, P. M., Majewski, S. R., Crane, J. D., et al. 2004, ApJL, 602, L21

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

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

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1

Glushkova, E. V. 2010, PAZh, 36, 83

Hao, C. J., Xu, Y., Hou, L. G., et al. 2021, A&A, 652, A102

Hao, C. J., Xu, Y., Wu, Z. Y., et al. 2022, A&A, 660, A4

He, Z., Li, C., Zhong, J., et al. 2022a, ApJS, 260, 8

He, Z., Wang, K., Luo, Y., et al. 2022b, ApJS, 262, 7

He, Z.-H., Xu, Y., Hao, C.-J., et al. 2021, RAA, 21, 093

Hunt, E. L., & Reffert, S. 2021, A&A, 646, A104
Hunt, E. L., & Reffert, S. 2023, A&A, 673, A114
Kharchenko, N. V., Piskunov, A. E., Schilbach, E., et al. 2013, A&A, 558, A53
Kos, J., de Silva, G., Buder, S., et al. 2018, MNRAS, 480, 5242
Kounkel, M., & Covey, K. 2019, AJ, 158, 122
Krone-Martins, A., & Moitinho, A. 2014, A&A, 561, A57
Li, Z., Deng, Y., & Chen, J. 2021, ApJS, 253, 38
Li, Z., Deng, Y., Chi, H., et al. 2022, ApJS, 259, 19
Li, Z., & Mao, C. 2023, ApJS, 265, 3
Li, Z.-M., Mao, C.-Y., Luo, Q.-P., et al. 2017, RAA, 17, 071
Liu, L., & Pang, X. 2019, ApJS, 245, 32
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
Mermilliod, J.-C. 1995, Information & On-Line Data in Astronomy, 203
Minniti, D. 2023, arXiv:2306.12894
Pera, M. S., Perren, G. I., Moitinho, A., et al. 2021, A&A, 650, A109
Porras, A., Christopher, M., Allen, L., et al. 2003, AJ, 126, 1916
Qin, S., Zhong, J., Tang, T., et al. 2023, ApJS, 265, 12
Reylé, C., & Robin, A. C. 2002, A&A, 384, 403
Ryu, J., & Lee, M. G. 2018, ApJ, 856, 152
Schmeja, S., Kharchenko, N. V., Piskunov, A. E., et al. 2014, A&A, 568, A51
Scholz, R.-D., Kharchenko, N. V., Piskunov, A. E., et al. 2015, A&A, 581, A39
Sim, G., Lee, S. H., Ann, H. B., et al. 2019, JKAS, 52, 145
Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2005, MNRAS, 356, 1293

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.