The Astrophysical Journal Supplement Series, 265:12 (13pp), 2023 March https://doi.org/10.3847/1538-4365/acadd6

Hunting for Neighboring Open Clusters with Gaia DR3: 101 New Open Clusters within 500 pc

Songmei Qin (秦松梅)[1,2], Jing Zhong (钟靖)[1], Tong Tang (唐通)[1,2], and Li Chen (陈力)[1,2]
[1] Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, People's Republic of China; qinsongmei@shao.ac.cn, chenli@shao.ac.cn
[2] School of Astronomy and Space Science, University of Chinese Academy of Sciences, No. 19A, Yuquan Road, Beijing 100049, People's Republic of China

Received 2022 October 18; revised 2022 December 13; accepted 2022 December 15; published 2023 February 22

Abstract

We systematically searched for open clusters in the solar neighborhood within 500 pc using the pyUPMASK and HDBSCAN clustering algorithms based on Gaia DR3. Taking into consideration that the physical size for most open clusters is less than 50 pc, we adopted a slicing approach for different distance shells and identified 324 neighboring open clusters, including 223 reported open clusters and 101 newly discovered open clusters, named here as the Open Cluster of Solar Neighborhood. Our discovery has increased the number of open clusters in the solar neighborhood by about 45%. In this work, larger spatial extents and more member stars were attained for our cluster sample. We provided the member stars and the membership probabilities through the pyUPMASK algorithm for each cluster and derived their astrophysical, age, and structural parameters.

Unified Astronomy Thesaurus concepts: Star clusters (1567); Open star clusters (1160)

Supporting material: figure set, machine-readable tables

1. Introduction

Open clusters (OCs) in the Milky Way are a collection of stars that are formed from the same molecular cloud and are gravitationally bound together (Lada & Lada 2003; Portegies Zwart et al. 2010), thus sharing similar specific characteristics (e.g., age, distance, reddening, metal abundance, etc.). OCs provide an ideal laboratory for studying star formation and evolution (Evans et al. 2009). Meanwhile, with its large sample, OCs are powerful tracers of the Galactic disk and can be used to constrain the Galactic structure and evolution history (Janes & Adler 1982; Dias & Lépine 2005).

As most Galactic OCs are located in the thin disk (Kharchenko et al. 2013), observations of OCs are often hampered by contamination from dominant background/foreground field stars, leading to more uncertainties in the characterization of cluster properties. Before the Gaia era, reliable member star selection was fairly difficult due to the limited astrometric precision, which brought about inconsistencies in the determination of basic parameters like the distances, kinematics, and ages of OCs (Netopil et al. 2015).

As one of the most successful and ambitious projects, Gaia enabled a deluge of scientific work including improving the quality of the cluster census. The Gaia DR2 catalog (Gaia Collaboration et al. 2018) presents more than 1.3 billion stars with unprecedented high-precision astrometric and photometric data, greatly improving the reliability of stellar membership determination and characterization of a large sample of OCs. The study of Galactic OCs has ushered in a new era. A more solid assessment of cluster membership using the precise Gaia data greatly promoted subsequent discoveries of new OCs. In the meantime, with the popularity of machine learning, cluster census efficiency has been tremendously raised. Many different methods have been used to search for new or re-detect existing OCs with Gaia data, especially various clustering algorithms such as unsupervised photometric membership assignment in stellar clusters (UPMASK; Krone-Martins & Moitinho 2014), 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; McInnes et al. 2017), Gaussian mixture models (GMMs, Pearson 1894), the Friends of Friends algorithm (FOF; Huchra & Geller 1982), etc. Hunt & Reffert (2021) have made detailed comparisons for three main algorithms (DBSCAN, HDBSCAN, and GMMs) side-by-side, exploring their effectiveness in blind-searching of OCs with the large-scale Gaia data set.

Based on Gaia DR2, Cantat-Gaudin et al. (2018) applied the UPMASK algorithm to select cluster members and provided an updated catalog of 1229 OCs including previously reported clusters and 60 newly discovered clusters. Liu & Pang (2019) reported 76 new OCs by employing an FOF-based cluster finder method to hunt out overdensities in (l, b, ϖ, μ_{α}^{*},[3] μ_{δ}) space. In parallel, using the DBSCAN algorithm, Castro-Ginard et al. (2020) found 582 new cluster candidates located in the low-Galactic-latitude area by investigating the aggregation of stars in the five-dimensional parameter phase space, with part of the candidates overlapped with the result of Liu & Pang (2019). Later on, Cantat-Gaudin et al. (2020) compiled a comprehensive list of 1867 bona fide OCs reanalyzed with UPMASK and Gaia data, providing a large and homogeneous catalog of OC properties and the corresponding member stars.

As the early stage of the Gaia third data release, Gaia EDR3 (Gaia Collaboration et al. 2021) provides astrometric and photometric parameters of 1.5 billion sources with even higher accuracy than Gaia DR2, averagely increasing the precision of proper motion by 2–3 times and parallax by about 20%. Solely based on the Gaia EDR3 database, Castro-Ginard et al. (2022)

---

[3] $\mu_{\alpha}^{*} = \mu_{\alpha} \cos\delta$.

have proposed the OCfinder method, which employs the DBSCAN clustering algorithm for selecting overdensities in the five-dimensional astrometric space and incorporates a deep artificial neural network for distinguishing bona fide OCs in color–magnitude diagrams (CMDs). In their work, 628 new OCs were picked out, mostly located 1 kpc away from the Sun. In our recent work (He et al. 2022b), we carried out a blind search for new OCs using Gaia DR2/EDR3 by dividing low-Galactic-latitude regions into $2° \times 2°$ ($|b| < 5°$) or $3° \times 3°$($|b| > 5°$) grids, sequentially performing the DBSCAN and pyUPMASK clustering algorithms in five-dimensional phase space (d_{l^*}, d_b, v_{α^*}, v_δ, ϖ). Eventually, 541 star clusters unrecorded in the literature were found, with a majority of them located beyond 1 kpc from the Sun, while only about 6% are located within 500 pc to the Sun. Using a similar approach and with Gaia EDR3, He et al. (2022a) expanded their search grid to $12° \times 12°$ or $18° \times 18°$ to increase the number of clusters in the solar vicinity and cataloged 270 newly found cluster candidates within 1.2 kpc of the Sun, of which 179 new OC candidates are within 500 pc.

In most cases of the above systematic searches for new OCs, Gaia's high-precision position (l, b), proper motion (μ_α^*, μ_δ), and parallax (ϖ) data were used to detect the aggregation characteristics of cluster members in multidimensional phase space, photometric data were assisted in confirming the reality of the OCs in CMDs and determining their basic properties by subsequent further analysis. On the other hand, these works have different features in their own methodology and procedure. The newly found objects greatly enrich our understanding of the Galactic OCs population, and they also indicate that the present OC sample is far from complete. It is anticipated that many new OCs still can be detected through the careful analysis of observational data.

The main difficulty in detecting OCs lies in the serious contamination of field stars, especially for nearby OCs, in which member stars cover a wide projected field of view and the proportion of background field stars could be overwhelming. This study aims to find new OCs in the solar neighborhood, with an effective way to decontaminate the background/foreground objects by astrometric data. For this purpose, we proposed a new approach of data "slicing" along the line-of-sight distance. In a systematic search of nearby OCs, we divided all-sky data into sub-blocks with different sizes in every 100 pc interval for reducing contamination of field stars as much as possible and highlighting the aggregation signal of cluster members in the multidimensional phase space.

This paper is the first of our serial work, focused on a comprehensive search for undetected clusters within 500 pc from the Sun, solely based on the most recent precise Gaia Data Release 3 (DR3) data set. With our novel slicing approach, we greatly improved the efficiency of detecting the nearby new OCs. Altogether we updated/derived the physical properties of 324 OCs, with 223 previously reported clusters and 101 newly detected clusters. The data we used in this work are introduced in Section 2. In Section 3, we describe the searching process with the new approach of data slicing along the line-of-sight distance. In Section 4, we describe the catalogs of OCs and their members. A discussion about cluster crossmatching and a property analysis is given in Section 5. Finally, we summarize in Section 6.

2. Data

This study aims to find new OCs in the solar neighborhood, with an efficient way of decontaminating background influences. For this purpose, we restricted our survey to stars within 500 pc, making full use of the high precision of the Gaia DR3 data.

Gaia DR3 (Gaia Collaboration et al. 2022) provides astrometric information with nearly the same high precision as Gaia EDR3 for about 1.8 billion sources over the sky, as well as near-millimagnitude-precision photometric data in three bands (G, $G_{\rm BP}$, and $G_{\rm RP}$). Moreover, Gaia DR3 introduces an impressive wealth of new data products, including the accurate radial velocity (RV) parameters for more than 33 million objects (Katz et al. 2022) and atmospheric parameters ($T_{\rm eff}$, $\log g$, and [M/H]) for about 470 million sources (Fouesneau et al. 2022). For Gaia DR3, the typical proper motion uncertainty respectively goes from 0.07 mas yr^{-1} for $G \approx 17$ mag up to 0.5 mas yr^{-1} for $G = 20$ mag; the parallax uncertainty goes from 0.07 mas at $G \approx 17$ mag up to 0.5 mas at $G = 20$ mag; and the mean G-band photometry uncertainty goes from 1 mmag at $G \approx 17$ mag up to 6 mmag at $G = 20$ mag. It is noteworthy that the newly determined median RVs (Katz et al. 2022) have risen greatly in number compared to Gaia DR2 and its median precision goes from 1.3 km s^{-1} at $G_{\rm RVS} \approx 12$ mag up to 6.4 km s^{-1} at $G_{\rm RVS} = 14$ mag, which greatly help us to study the dynamical evolution of the OCs in the Milky Way.

In this work, we first select stars from Gaia DR3 in the region with $|b| < 30°$ since the majority of identified OCs are located in $|b| < 20°$ (Dias et al. 2002; Kharchenko et al. 2013; Castro-Ginard et al. 2020). Then, we used the astroquery.gaia python package[4] (Ginsburg et al. 2019) to obtain the positions (l, b), proper motions (μ_α^*, μ_δ), parallaxes (ϖ), magnitudes in three photometric filters (G, $G_{\rm BP}$, and $G_{\rm RP}$), RV (rv), and their associated uncertainties from the Gaia Archive.[5] To investigate the OCs in the solar vicinity, we only retained stars with a parallax greater than 2, approximately corresponding to a distance of 500 pc. To reduce the amount of faint background stars, we filtered out stars with $G > 18$ mag. And we applied a cut to the renormalized unit weight error (RUWE) of < 1.4 (Lindegren 2018) to exclude unreliable astrometric and photometric data. Finally, we screened out about 8 million stars as the initial sample.

3. Method

Our search process is divided into four steps:

1. Partitioning the entire sample into multi-blocks to reduce the contamination of field stars; see the details in Section 3.1.
2. Employing an unsupervised clustering method pyUPMASK to five-dimensional data (l, b, ϖ, μ_α^*, μ_δ) to assign probabilities to the stars and remove stars with probabilities <0.1 in each block. Then we pick out those blocks that may contain OCs from proper motion distribution features; see the details in Section 3.2.
3. Separating the stars in each block into members of individual cluster candidates with HDBSCAN; see the details in Section 3.3.

---

[4] https://astroquery.readthedocs.io/en/latest/gaia/gaia.html

[5] https://gea.esac.esa.int/archive/

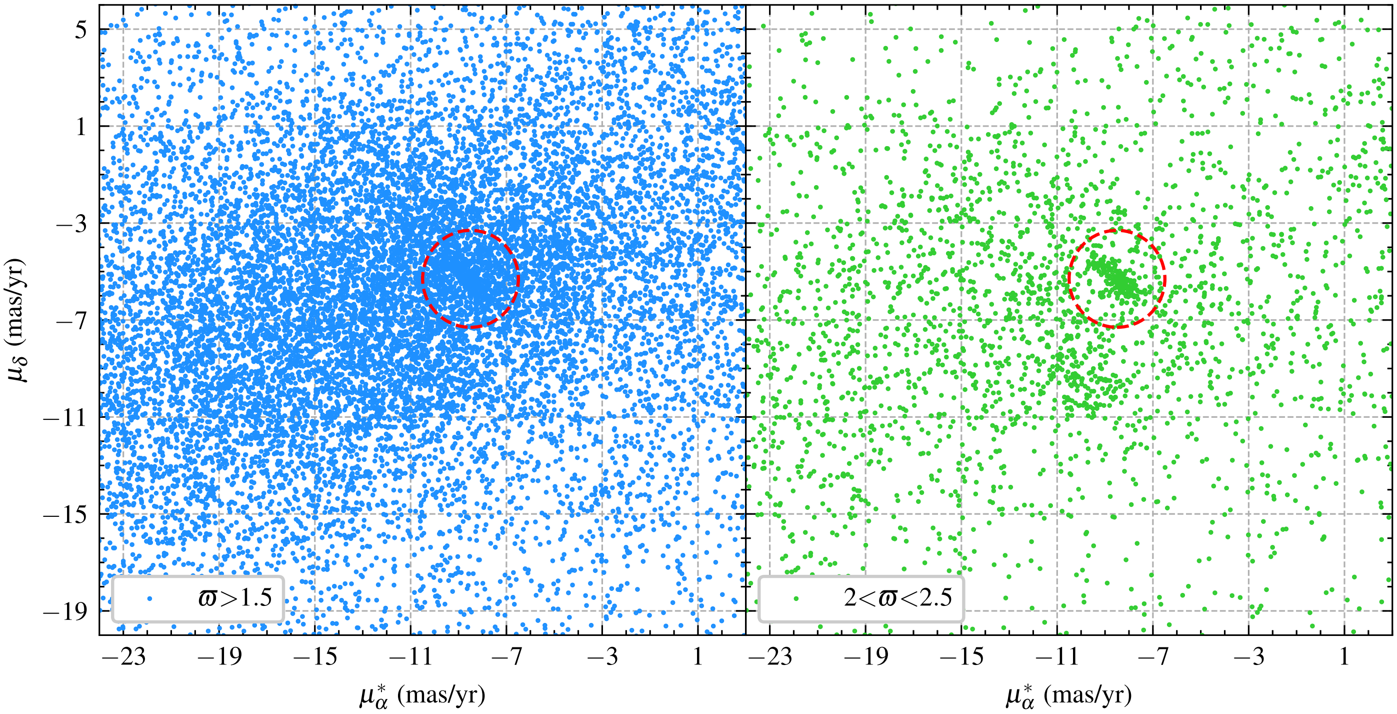

Figure 1. The proper motion distribution of two samples. The blue and green dots refer to stars with different distant slicing. The red dashed circles represent an OC where the members assembled.

4. By applying a visual inspection of the distribution of position (l, b), proper motion (μ_α^*, μ_δ), magnitude–parallax (G, ϖ), and CMD ($G_{\rm BP} - G_{\rm RP}$, G) of each cluster candidate, we confirm the true existence of the star clusters; see the details in Section 3.4.

According to the parameter ranges obtained in the above identification process, we retrieved the astrometric and photometric data for stars in the cluster regions. Then, we obtained the membership probabilities of the stars in each cluster through the pyUPMASK algorithm and regarded stars with probabilities greater than 0.5 as cluster members, which are cataloged in Table 3.

3.1. Data Slicing

Members of an OC that formed in the same molecular cloud generally are severely immersed in a dense star field, and it is hard to distinguish actual star cluster members from the background, especially for those clusters mixed with the vast majority of field stars. The most efficient way of hunting out new clusters is to look for the clustering of stars in velocity space (i.e., the vector-point-diagram) since OC members share a distinctive movement as compared to the field stars. For highlighting the cluster members in the proper motion distribution diagram, we cut the entire sample into various subsamples according to their 3D spatial coordinates (l, b, and ϖ). For example, as we can see in Figure 1, the two panels are the proper motion distributions (μ_α^*, μ_δ) of two subsamples with the same 2D spatial distribution range: $306° \leqslant l \leqslant 312°$ and $-6° \leqslant b \leqslant 0°$. The ϖ of the left sample is greater than 1.5, and in the red dashed circle we can see a vaguely condensed area blurred by the overwhelming distribution of field stars. In the right panel of Figure 1, we limited this sample to $2 < \varpi < 2.5$, thus removing many foregrounds and background stars and the corresponding distribution shows an obvious concentration of cluster member stars around the same region. Hence, ingenious data slicing is critical for hunting out new OCs in our cluster searching volume.

Table 1
Size of the Search Area

dist (pc) ϖ (mas)	<100 >10	100–200 5–10	200–300 3.33–5	300–400 2.5–3.33	400–500 2–2.5
size (deg)	30	20	12	10	6

To reveal OCs hidden in the field stars more fully and efficiently, we divided the entire remaining stellar sample within the Galactic latitude range of $|b| < 30°$ into 1044 blocks according to their 3D spatial coordinates (l, b, and ϖ). First, we separated all the stars into several searching shell regions along ϖ (>10 mas, 5–10 mas, 3.33–5 mas, 2.5–3,33 mas, 2–2.5 mas), approximately corresponding to the distance (<100 pc, 100–200 pc, 200–300 pc, 300–400 pc, 400–500 pc), which is large enough to cover at least one typical OC. In recent years, many investigations based on Gaia data revealed that besides the high-density inner parts, there might be extended low-density outer halos (Zhong et al. 2019; Meingast et al. 2021; Zhong et al. 2022) or elongated tidal tails (Carrera et al. 2019; Zhang et al. 2020; Boffin et al. 2022; Bai et al. 2022) in the outskirt of OCs. To contain the outer halo structure of OCs adequately, we adopted 50 pc as the typical extended spatial scale of OCs (Tarricq et al. 2022; Zhong et al. 2022). As the projected angular size varies with distance, we adopted searching grids with different angular sizes, ranging from 30° at a close distance of 100 pc to about 6° at 500 pc, as listed in Table 1.

3.2. Initial Screening with pyUPMASK

After getting the 1044 slicing blocks, we applied the pyUPMASK algorithm (Pera et al. 2021) to gain the

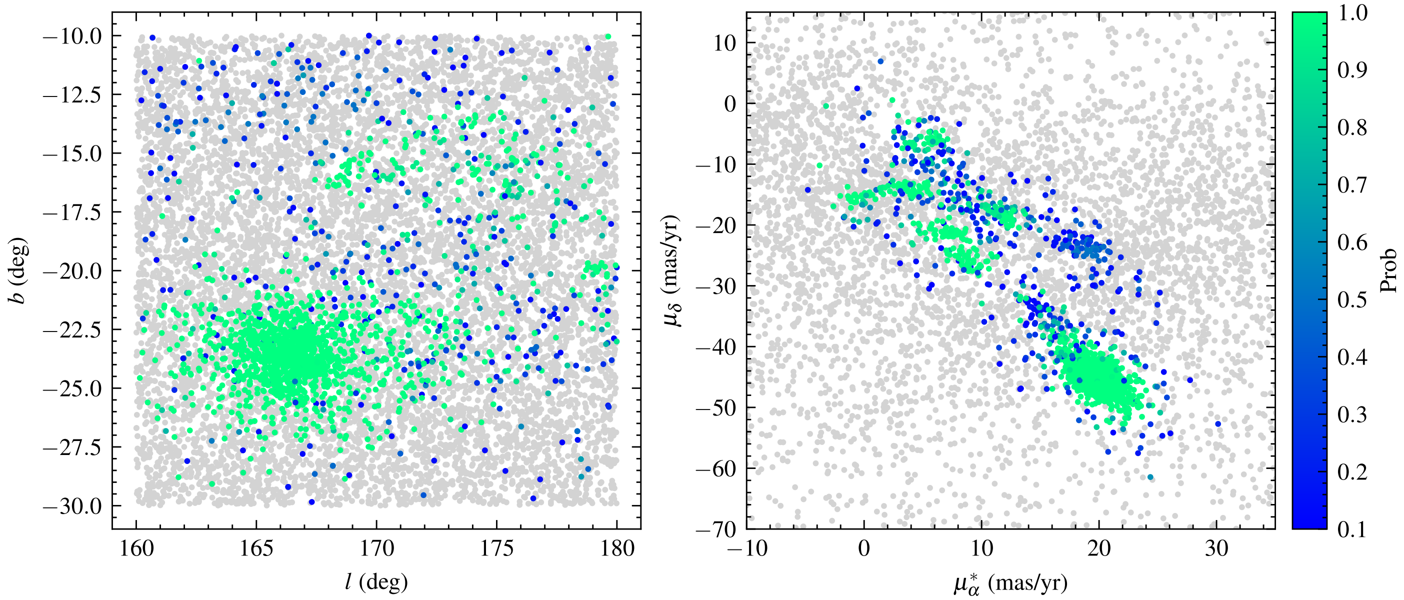

Figure 2. Position and proper-motion distributions of one block sample. The gray dots refer to stars with membership probabilities less than 0.1, while the colored dots are stars with membership probabilities greater than 0.1. The color bar represents the probabilities of the member stars.

membership probability of each star, which is based on the clustering of members compared to field stars in (l, b, ϖ, μ_α^*, μ_δ) phase space. pyUPMASK is an open-source software package compiled in the Python language following the development principle of UPMASK (Krone-Martins & Moitinho 2014), which is a member star determination method developed to process photometric data originally, though later widely used in the determination of member stars based on astrometric data (Cantat-Gaudin et al. 2018, 2020).

This enhanced clustering method contains several major procedures as follows: (i) input data reduction using a principal component analysis; (ii) choose one of the clustering algorithms such as K-means (Chaturvedi et al. 2001), mini-batch k-means (Sculley 2010), GMM (Pearson 1894), agglomerative clustering (Zepeda-Mendoza & Resendis-Antonio 2013), the k-nearest neighbors density method (Rodriguez & Laio 2014), or the Voronoi method (Voronoi 1908) supported by pyUPMASK to process the reduced data; (iii) employ Ripley's K function (Ripley 1976) to assess the authenticity of the clusters (or reject fake clusters with a random uniform distribution); (iv) apply the Gaussian uniform mixture model to level down the field contamination. We skip this step to reserve any poor or non-Gaussian distribution clusters; and (v) evaluate the cluster membership probabilities through the kernel density estimator (KDE).

The continuous KDE probabilities between 0 and 1 are assigned to all the stars defined as $P_{\rm cl} = {\rm KDE}_m/({\rm KDE}_m + {\rm KDE}_{nm})$, where ${\rm KDE}_m$ and ${\rm KDE}_{nm}$ refer to the KDE likelihoods for members and nonmembers, respectively, which reflects the proportion of true cluster stars and field stars. Concerning the imbalance of members and field stars that are mixed within a stellar cluster field, we set the lower limit of member probabilities to 0.1 to preserve cluster members as many as possible. As an example, Figure 2 shows the position and proper motion distributions of a sample with the range of $160° \leqslant l \leqslant 180°$, $-30° \leqslant b \leqslant -10°$, and $5\ {\rm mas} < \varpi < 10\ {\rm mas}$. In Figure 2, the gray dots represent stars with probabilities less than 0.1 (about 90%), while the rest of the colored dots represent stars with probabilities greater than 0.1 (about 10%). We can see that after rejecting a large proportion of field stars, some clumps are exposed in the position and proper motion distributions. Nevertheless, some clusters such as Cand1 in Figure 3 show a relatively sparse spatial distribution that leads to lower probabilities. To avoid losing such sparse OCs, we applied the cut at a probability of >0.1 to exclude most field stars.

After excluding field stars with low probability ($p < 0.1$) in all block samples, we visually inspected the proper motion distributions of the member stars. As shown in Figure 2, the block sample exhibits several distinct clumps both in position and proper motion distributions, which would be reserved in the following analysis in Section 3.3. It is noticed that, the concentrated structure around ($20\ {\rm mas\ yr}^{-1}$, $-45\ {\rm mas\ yr}^{-1}$) is the famous OC Pleiades (M 45) (Cantat-Gaudin et al. 2020; Heyl et al. 2022). Eventually, 394 block samples that might contain OCs were retained.

3.3. Separating OCs with HDBSCAN

For the 394 samples, we normalized the data in five dimensions (l, b, ϖ, μ_α^*, μ_δ) and then used HDBSCAN (McInnes et al. 2017) to separate individual cluster candidate members from field stars. HDBSCAN is primarily proposed by Campello et al. (2013), which combines the density-based approach of DBSCAN (Ester et al. 1996) with hierarchical clustering to deal with data sets of varying densities. The key parameter that affecs the resulting clustering is *min_cluster_size* ($m_{\rm clSize}$), which refers to the minimum possible size of a cluster. A smaller value of $m_{\rm clSize}$ might cause a large cluster to be divided into small clumps, which may generate some fake clusters. However, with a larger value of $m_{\rm clSize}$, small adjacent clusters in phase space would be combined as a big one.

Hunt & Reffert (2021) compared three clustering algorithms, DBSCAN, HDBSCAN, and GMMs, in terms of computational speed and availability, and concluded that HDBSCAN is the most sensitive and effective method for revealing OCs in Gaia data. In their work, the performance of the HDBSCAN algorithm for 100 OCs showed that the value of

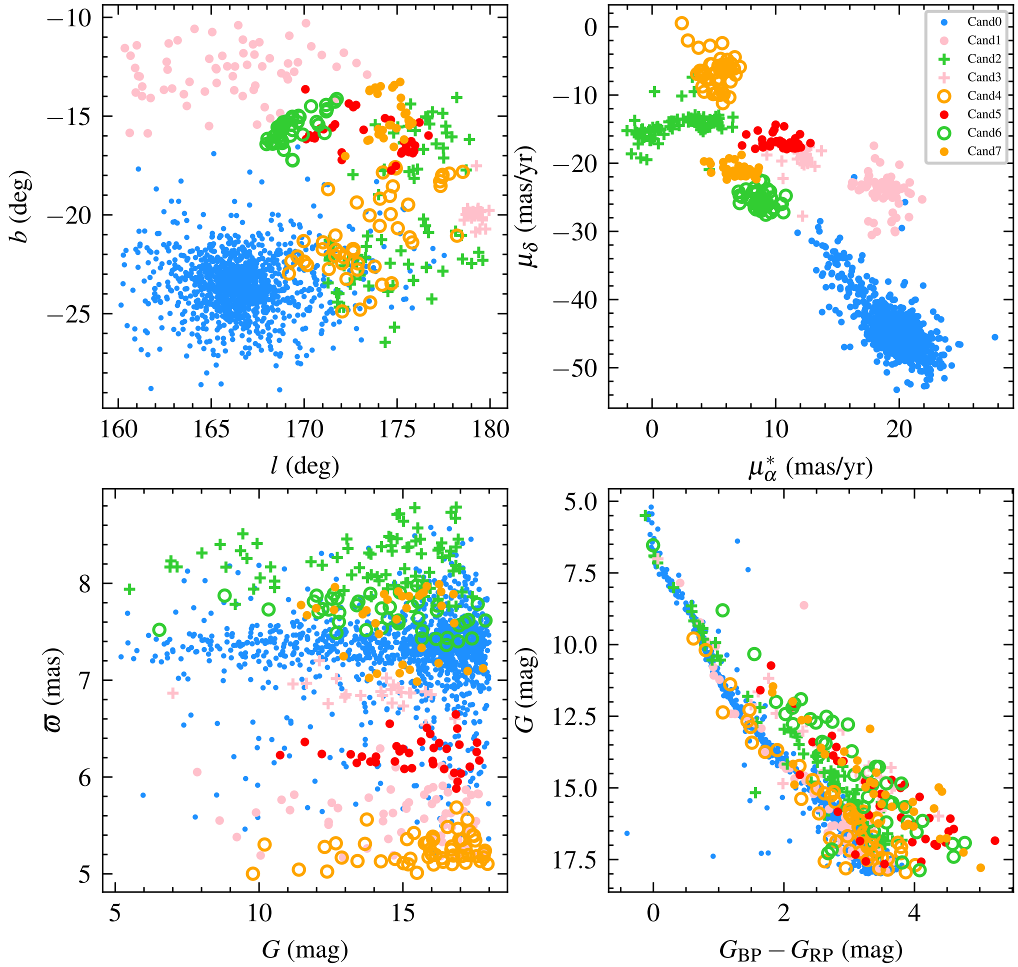

Figure 3. Spatial (l, b), proper motion (μ_α^*, μ_δ), magnitude–parallax (G, ϖ), and CMDs ($G_{\rm BP} - G_{\rm RP}$, G) of the eight separated cluster candidates using HDBSCAN. The colors and symbols represent member stars of different cluster candidates.

"Sensitivity" = TP/(TP + FN) was the largest when adopting $m_{\rm clSize}$ as 10, which corresponds to the strongest ability of HDBSCAN to detect real OCs. Campello et al. (2013) also recommended setting $m_{\rm clSize} = m_{\rm Pts} = 10$ for best sensitivity and speed when running the algorithm. Thus, we assigned $m_{\rm clSize} = 10$. At the same time, for better detecting some sparse OCs, we selected the "leaf" cluster selection method (McInnes et al. 2017). After applying HDBSCAN to separate out cluster groups in the five-dimensional data, we obtained 800 OC candidates. For example, in Figure 3, the same sample of stars as in Figure 2 were separated into eight cluster candidates with HDBSCAN. Although some fake clusters may arise because of the parameter selection in HDBSCAN, we will check every cluster candidate in the five-dimensional data by eyes; see the details in Section 3.4.

3.4. Visual Inspection

The member stars in an OC are comoving and sharing similar parallaxes/distances. In the meantime, the CMD of the members is expected to present a clear main-sequence feature. We visually screen each cluster candidate in terms of position (l, b), proper motion (μ_α^*, μ_δ), magnitude–parallax (G, ϖ) distributions, and CMD ($G_{\rm BP} - G_{\rm RP}$, G) to reject "false positive" clusters.

It is noticed that a given cluster candidate which may happen to be identified within one block may sometimes fall on the borders, or can be detected in more than one block (with different membership results). To merge the split cluster candidates, we adopted the following process. We designated a nearby cluster with the smallest angular distance as a reference cluster for each target cluster identified by eye inspection. And we inspected the position (l, b), proper motion (μ_α^*, μ_δ), magnitude–parallax (G, ϖ) distribution, and CMD ($G_{\rm BP} - G_{\rm RP}$, G) of the target and reference clusters together to assess whether they are the same cluster or not. A few specific cases that arose in our cluster samples are listed as follows:

1. Only one peak appears in both the spatial and proper motion distributions. If the parallaxes of their members are consistent, the two clusters are thought to be combined together; if not, the two clusters are considered to be different individuals.
2. More than one peak arises in both spatial and proper motion distributions. In this case, the two clusters are regarded as different individuals.
3. Only one peak in the spatial distribution but two separated peaks in the parallax histogram. They are identified as distinguished clusters.

Figure 4 shows five-dimensional distributions of three corresponding examples. An example of case (a) is presented

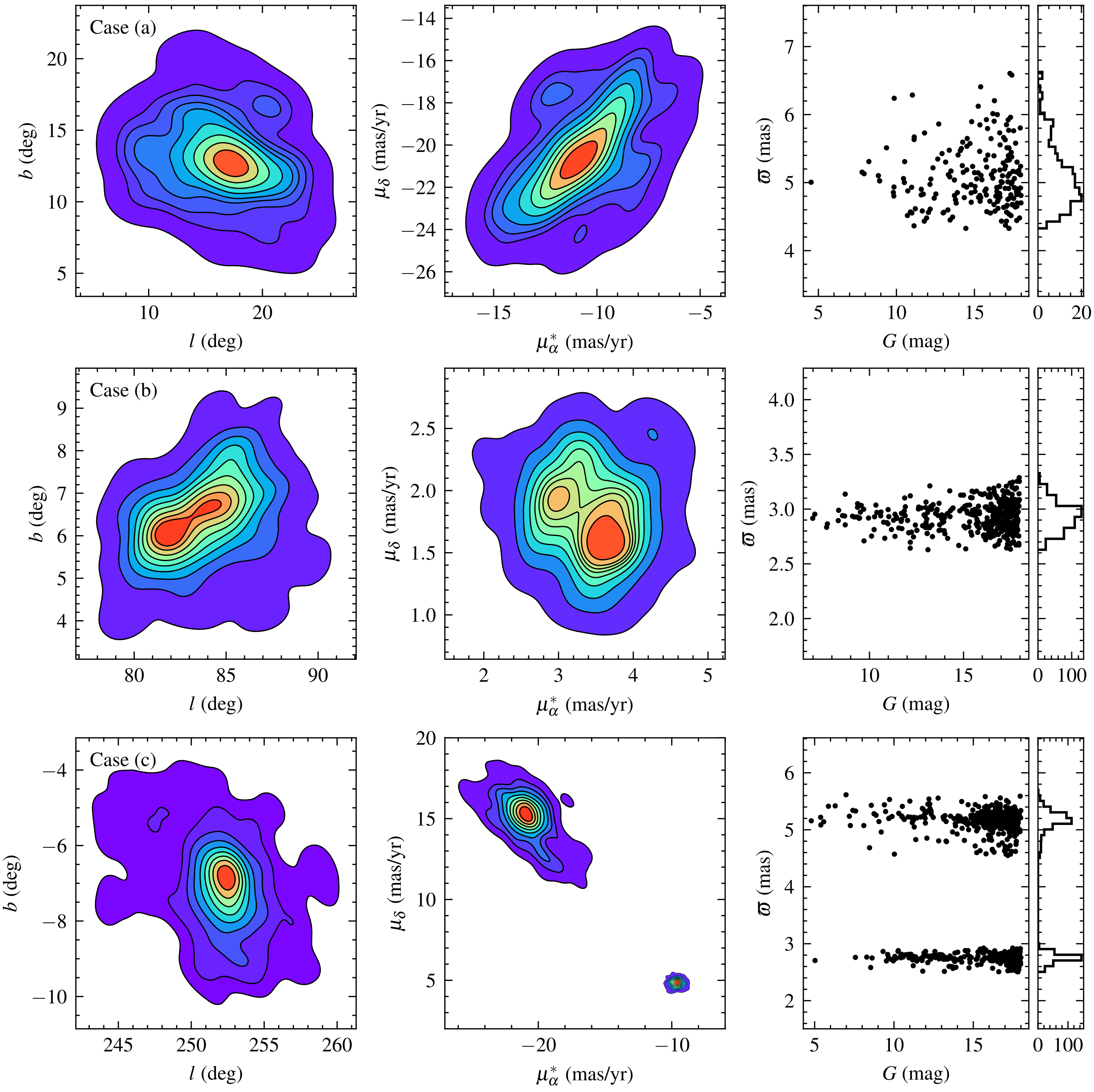

Figure 4. Three examples of target and reference clusters. The left panels present the spatial distributions of the members' isodensity contours. The middle panels present the proper motion distribution of the members' isodensity contours. The right panels present the G-parallax distribution with parallax histograms on the right edge.

in the top panel: only one spatial over-density distribution can be detected and one peak in the proper motion distribution. In addition, there is one peak around 5 mas in the parallax histogram. These properties demonstrate that all members belong to the same cluster. The middle panels display a case (b) example of two clusters, which hold close but notably different peaks. And their mean parallaxes are approximately equal to 2.9 mas. They could be binary clusters at nearly the same distance. To illustrate case (c), we show two distinct clusters which have the same projected position in the bottom panel. Although these clusters overlap in the projected position, which means one is in the front and the other is in the back, they can be well distinguished through the distribution of different proper motions and parallaxes. During the above visual checking process, we adjusted the parameter ranges of each cluster. Finally, we obtained 324 genuine stellar aggregations, most of which are certainly OCs.

4. Cluster and Member Catalog

We regarded stars with membership probabilities >0.5 as cluster members. Based on those members, we attained the

Table 2
Description of the Catalog of OC Properties

Column	Format	Unit	Description
Name	string	⋯	Cluster name in this work
OC_flag	string	⋯	Cluster crossmatch cases
glon	float	deg	Mean Galactic longitude of members
glat	float	deg	Mean Galactic latitude of members
R.A.	float	deg	Mean R.A. of members
decl.	float	deg	Mean decl. of members
pmra	float	mas yr^{-1}	Mean proper motion in R.A. of members
e_pmra	float	mas yr^{-1}	Standard deviation of proper motion in R.A.
pmdec	float	mas yr^{-1}	Mean proper motion in decl. of members
e_pmdec	float	mas yr^{-1}	Standard deviation of proper motion in decl.
plx	float	mas	Mean parallax of members
e_plx	float	mas	Standard deviation of parallax
RV	float	km s^{-1}	Mean RV of members
e_RV	float	km s^{-1}	Standard deviation of the RV
RV_Flag	string	⋯	Label of RV
N_RV	int	⋯	Number of RV members
N	int	⋯	Number of members with membership probabilities higher than 0.5
m–M	float	mag	Cluster distance modulus determined by the isochrone fit
log t	float	⋯	Cluster age determined by the isochrone fit
$E(B–V)$	float	mag	Cluster reddening determined by the isochrone fit
$r_{\rm h}$	float	deg	Angular size of half-number radius
$r_{\rm c}$	float	deg	Cluster core radius
$e\_r_{\rm c}$	float	deg	Uncertainty of cluster core radius
$r_{\rm t}$	float	deg	Cluster tidal radius
$e\_r_{\rm t}$	float	deg	Uncertainty of cluster tidal radius
$r_{\rm o}$	float	deg	Mean radius of cluster outer region
$e\_r_{\rm o}$	float	deg	Uncertainty of the mean radius of cluster outer region
$r_{\rm e}$	float	deg	Cluster boundary radius
$e\_r_{\rm e}$	float	deg	Uncertainty of cluster boundary radius
Ref_Name	string	⋯	Cluster names in the literature works
Ref	string	⋯	References corresponding to cluster names (see Section 5.1)

(This table is available in its entirety in machine-readable form.)

cluster properties. We provide two catalogs in this paper: one for the properties of 324 OCs and the other for parameters of 59,304 member stars.

Table 2 describes the catalog of our OC properties. Crossmatching with the published catalogs, we verified the reported and new OCs and provided their Open Cluster of Solar Neighborhood (OCSN) "Names" (Col. 1). The "OC_flag" (Col. 2) was given according to the crossmatching cases; see the details in Section 5.1. The central coordinates of the clusters (Cols. 3–6) were obtained through a two-dimensional Gaussian KDE and the bandwidth of the kernel was calculated via the well-known Scott's rule (Scott 1992, 2015). We derived the mean values of the proper motion and parallax of each cluster and their corresponding standard deviations (Cols. 7–12). Meanwhile, we fitted the mean RV for each cluster with a Gaussian profile (Cols. 13–16). By visually inspecting the match of the isochrones to the observed cluster CMDs, we further obtained the ages, distance moduli, and reddening parameters (Cols. 18–20). In order to reveal the structural characteristics of the nearby star clusters better, we provided the radii parameters (r_c, r_t, r_o, r_e) of OCSN clusters (Cols. 22–29) according to the two-component model proposed by Zhong et al. (2022; hereafter Zhong2022). We listed the reported cluster names in the literature as well as the corresponding reference work (Cols. 30–31).

Table 3
Description of the Catalog of Cluster Members

Column	Format	Unit	Description
Source_id	long	⋯	Unique source identifier
l	double	deg	Galactic longitude
b	double	deg	Galactic latitude
R.A.	double	deg	Right ascension (ICRS at Epoch = 2016)
decl.	double	deg	decl. (ICRS at Epoch = 2016)
plx	double	mas	Parallax
plx_err	float	mas	Standard error of parallax
pmra	double	mas yr^{-1}	Proper motion in R.A.
pmra_err	float	mas yr^{-1}	Standard error of proper motion in R.A.
pmdec	double	mas yr^{-1}	Proper motion in decl.
pmdec_err	float	mas yr^{-1}	Standard error of proper motion in decl.
rv	float	km s^{-1}	RV
rv_err	float	km s^{-1}	RV error
rv_Flag	string	⋯	Label of RV members
Gmag	float	mag	G magnitude
Gmag_err	float	mag	G magnitude error
BPmag	float	mag	BP magnitude
BPmag_err	float	mag	BP magnitude error
RPmag	float	mag	RP magnitude
RPmag_err	float	mag	RP magnitude error
probs	double	⋯	Membership probability obtained from pyUPMASK
Name	string	⋯	Corresponding cluster name in this work

(This table is available in its entirety in machine-readable form.)

Table 3 describes the catalog of cluster members, including the astrometric and photometric parameters from Gaia DR3 (Cols. 1–20), the derived membership probabilities through pyUPMASK (Col. 21), and the corresponding cluster names in this work (Col. 22).

4.1. RVs of the OCs

In our cluster member sample, there are a total of 22,067 stars with RVs from Gaia DR3 (Katz et al. 2022), which can be used to estimate the mean RVs of the 324 clusters. At first, we found that the RVs of the initial cluster members deviate greatly from the median value, especially at $G > 14.5$ mag. It is noticed that a similar situation was also reported by Ye et al. (2022). To derive the reliable mean RV of each cluster, we removed outliers beyond 3σ and used the high-quality members for calculation. The RV outliers, high-quality members, and non-RV members are labeled as "rv_Flag" = 0, 1, 2, respectively in the member catalog (see Table 3). In our sample, the average RVs were derived by the Gaussian fitting for clusters with enough RVs members or simply provided by median RVs for clusters with few RV members, while the

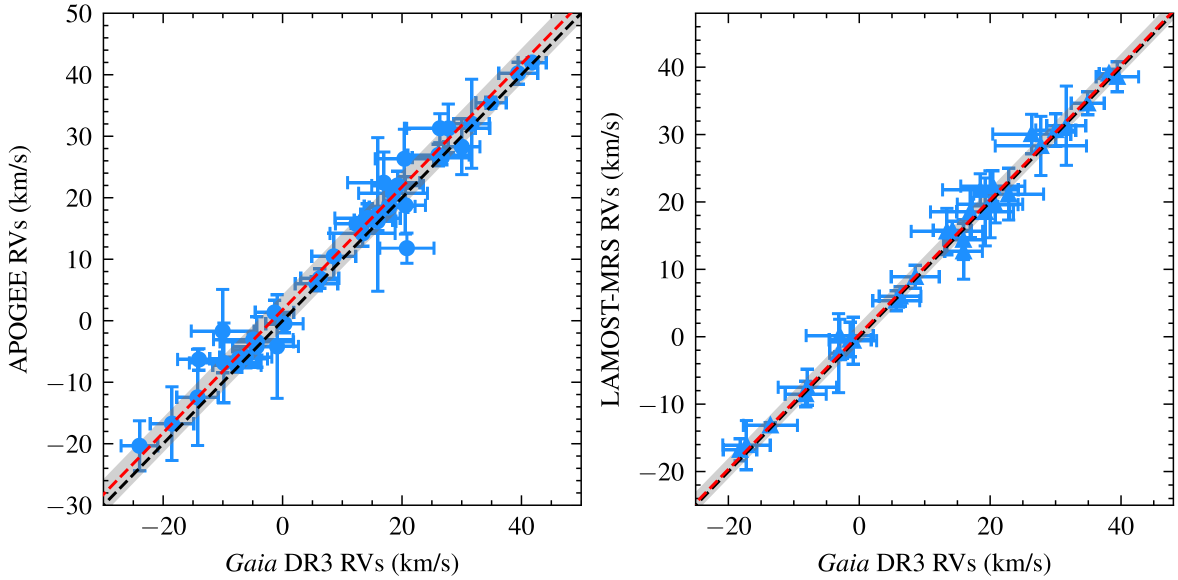

Figure 5. Comparison of the mean OC RVs between Gaia DR3 and APOGEE/LAMOST-MRS. The error bars of the blue points/triangles are the 1σ of the Gaussian fitting or standard deviations for OC RVs. The black dashed lines are 1:1 scale lines. The red dashed lines are the mean values of the differences between Gaia DR3 RVs and APOGEE RVs or LAMOST-MRS RVs, and the gray-filled regions are the corresponding 1σ regions.

corresponding uncertainties were the 1σ of the Gaussian functions or standard deviations, respectively. We provided the corresponding "RV_Flag" = 1 for the mean RVs from Gaussian fitting and 2 for median RVs. Eventually, there are 324 OCs with an observed mean or median RV (hereafter Gaia DR3 RVs) which are listed in Table 2.

At the same time, we calculated the mean or median RVs of our cluster samples with APOGEE DR17 (Abdurro'uf et al. 2022) and LAMOST DR9 of the Medium Resolution Survey[6] (hereafter APOGEE RVs and LAMOST-MRS RVs, respectively) through the same approach. Comparisons of the RVs for 55 common OCs between Gaia DR3 and APOGEE and 42 common OCs between Gaia DR3 and LAMOST-MRS are shown in Figure 5. The mean and standard deviation values of the differences between Gaia DR3 RVs and APOGEE RVs are $-1.42\ \mathrm{km\ s^{-1}}$ and $2.83\ \mathrm{km\ s^{-1}}$, respectively, and the corresponding values of the differences between Gaia DR3 RVs and LAMOST-MRS RVs are $-0.48\ \mathrm{km\ s^{-1}}$ and $1.50\ \mathrm{km\ s^{-1}}$, respectively, demonstrated as the red dashed lines and the gray-filled regions in Figure 5. It is evident that the Gaia DR3 RVs, APOGEE RVs, and LAMOST-MRS RVs are mostly consistent.

4.2. Metallicity and Isochrone Fitting

Before getting the age parameters for our OC samples, we also collected the available [Fe/H] metallicity for clusters from literature spectroscopic work. The Open Cluster Chemical Abundances and Mapping survey (Donor et al. 2020; hereafter Donor2020) provided [Fe/H] abundances for a sample of 128 OCs from APOGEE DR16. After crossmatching our cluster samples with Donor2020, 14 common clusters were found. We also acquired seven common clusters within 500 pc from Netopil et al. (2022) (hereafter Netopil2022). In addition, we gathered the LAMOST spectroscopic parameters of the other 34 OCs from Zhong et al. (2020) (hereafter Zhong2020). To ensure the reliability of [Fe/H] for OCs, we select 31 OCs with more than five [Fe/H] members, as shown in Table 4, which would be used in the following isochrone-fitting process.

To determine the age parameters of the OCs found in the solar neighborhood, we use a set of Padova isochrones (Marigo et al. 2017) to perform the CMD fitting. The grid of logarithm ages in isochrones is from 6.0 to 10.10 with an interval of 0.05, and the photometric system is the Gaia photometric system (Riello et al. 2021) from CMD 3.6[7] For clusters whose metallicity is reported by literature, we adopted the abundances in Table 4 as an input parameter to derive the Padova isochrone, while the isochrones of the other clusters were adopted with a solar metallicity of $Z_\odot = 0.0152$ (Caffau et al. 2009, 2011). We carefully inspected the match of the isochrones to the significant characteristic regions, such as the upper main sequence, the turn-off point, and the red giant or red clump features in the CMDs. By adjusting the isochrones to achieve the best fitting of cluster members in the CMDs, we obtained the age, distance modulus, as well as reddening of each cluster. Then, we use the formulas $A_G = 2.74 \times E(B-V)$, $E(BP-RP) = 1.339 \times E(B-V)$, and $E(G-RP) = 0.705 \times E(B-V)$ (Casagrande & VandenBerg 2018; Zhong et al. 2019) to calculate the $E(B-V)$ values. Figure 6 shows the isochrone-fitting examples for four clusters and the final fitting results (age: $\log t$; distance modulus: $m-M$; reddening value: $E(B-V)$) are shown in Table 2.

4.3. Structural Parameters

It is evident that the discoverable spatial scale of OCs is greatly expanded in the Gaia era. More and more members located in the extended region were identified through their kinematic properties (Meingast & Alves 2019; Meingast et al. 2021). Our investigation of nearby clusters also shows that many clusters have an extended outer structure. Many reported clusters are just tight core components in our OCSN catalog (See Section 5.1).

To describe better the radial density profile (RDP) of clusters with extended outer regions, it is proposed by Zhong2022 to

---

[6] http://www.lamost.org/dr9/

[7] http://stev.oapd.inaf.it/cgi-bin/cmd

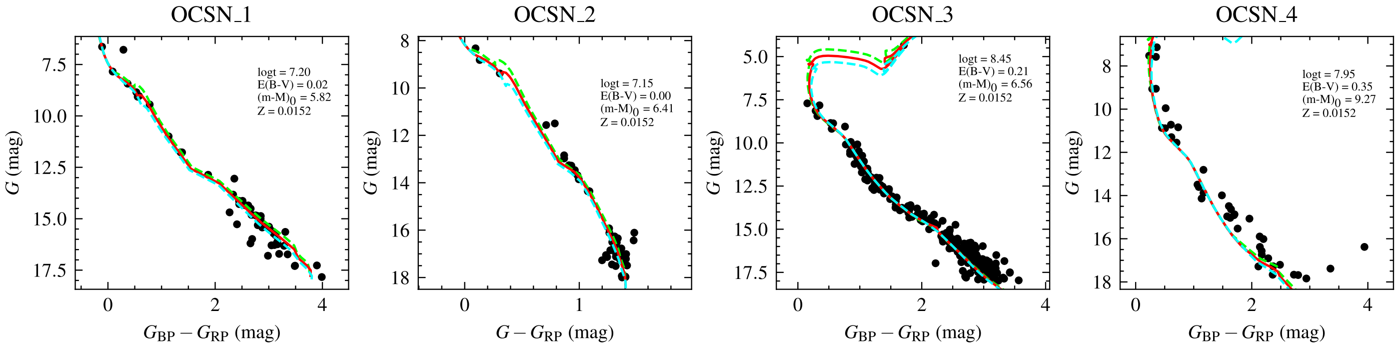

Figure 6. Example of the isochrone-fitting results for four clusters: OCSN_1, OCSN_2, OCSN_3, and OCSN_4. The black dots refer to the members identified in this work. The red solid lines indicate the best-fitting isochrones while the lime or cyan dashed lines denote the ischrones whose ages (log *t*) are 0.5 larger or smaller than the best-fitting isochrones. All the isochrones applied at solar metallicity ($Z_\odot = 0.0152$). The best-fitting parameters are shown in the panels for each cluster.

(The complete figure set (324 images) is available.)

use a two-component model instead of only using a King model (King 1962, 1966). After deriving the RDP of each cluster through a two-dimensional Gaussian KDE on the two-dimensional position space, we further attempted the two-component model to fit the RDP:

$$F(r) = f(r) + g(r), \tag{1}$$

where $f(r)$ is the King model that mainly describes the RDP of core members and $g(r)$ is a logarithmic Gaussian function that described the RDP of corona members (Zhong2022).

In the fitting procedure, the two-component model performs a more reliable approximation of the RDP of most OCSN clusters. In particular, we noted that there are about 33% of OCSN clusters whose RDPs can be well approximated by the single King model. The fraction of clusters that well follow a single King profile is larger than the fraction (about 10%) in Zhong2022. We speculate that it is because the outer extended structure of some nearby clusters can be extended to dozens of degrees, and the OCSN clusters we identified may still be core components. However, there is still a fraction of star clusters (43 of 324) that cannot be decently fitted by the single-component or two-component method, possibly due to their sparse distribution, extended tail-like structures, or even multiple cores. And it is quite obvious that not all of the OCSN clusters have a clear core.

5. Discussion

5.1. Comparison with Reported Clusters

Based on Gaia data, several OC-hunting studies have published more than 3000 OCs by applying multifarious clustering algorithms or manual-searching approaches. We collected the published OCs within 500 pc from previous works (Liu & Pang 2019; Sim et al. 2019; Cantat-Gaudin et al. 2020; Hunt & Reffert 2021; He et al. 2022a, 2022b; Li et al. 2022, hereafter LP19, Sim19, CG20, HR21, He22a, He22b, Li22, respectively), including about 10% of the OCs within 500 pc of the solar neighborhood. Many nearby young associations and moving groups within 500 pc to the Sun also have been investigated with Gaia data, such as the Orion complex with a strong sign of radial expansion attributed to a supernova expansion (Kounkel et al. 2018, hereafter K18), Vela OB2, which hosts complex spatial filamentary substructures (Cantat-Gaudin et al. 2019; Beccari et al. 2020; Pang et al. 2021, hereafter CG19, B20, Pang21, respectively), Taurus region consists of 22 groups (Liu et al. 2021; hereafter Liu21), Chamaeleon I with two sub-clusters (Roccatagliata et al. 2018; hereafter R18), Corona Australis with "off-cloud" and "on-cloud" populations (Galli et al. 2020a, hereafter G20a), ϵ Cha Association (Dickson-Vandervelde et al. 2021; hereafter DV21), ρ Ophiuchi with two young populations (Grasser et al. 2021; hereafter G21), Perseus with five clustered groups Autochthe, Alcaeus, Mestor, Electryon, and Heleus (Pavlidou et al. 2021;

Table 4
Metallicity of Some Reported Clusters

Name	Ref_Name	[Fe/H] (dex)	e_[Fe/H] (dex)	N_[Fe/H]	Ref
OCSN_127	ASCC_16	−0.06	0.06	33	Donor2020
OCSN_128	ASCC_19	−0.07	0.05	19	Donor2020
OCSN_129	ASCC_21	−0.13	0.04	10	Donor2020
OCSN_130	ASCC_41	−0.11	0.07	8	Zhong2020
OCSN_141	Alessi_20	+0.14	0.04	11	Zhong2020
OCSN_189	Collinder_69	−0.10	0.05	55	Donor2020
OCSN_192	Collinder_350	−0.10	0.11	31	Zhong2020
OCSN_194	Gulliver_6	−0.13	0.20	18	Zhong2020
OCSN_203	IC_348	−0.17	0.14	11	Zhong2020
OCSN_204	IC_2391	−0.03	0.04	11	Netopil2022
OCSN_205	IC_2602	−0.02	0.02	7	Netopil2022
OCSN_206	IC_4665	−0.01	0.02	11	Netopil2022
OCSN_207	IC_4756	+0.02	0.04	13	Netopil2022
OCSN_213	L_1641S	−0.09	0.07	22	Donor2020
OCSN_218	Melotte_20	+0.01	0.05	64	Donor2020
OCSN_219	Melotte_22	0.00	0.05	83	Donor2020
OCSN_220	Melotte_25	+0.12	0.04	48	Netopil2022
OCSN_221	NGC_752	−0.08	0.07	49	Zhong2020
OCSN_222	NGC_1039	+0.02	0.06	7	Netopil2022
OCSN_224	NGC_1662	−0.19	0.09	35	Zhong2020
OCSN_226	NGC_1980	−0.08	0.04	9	Donor2020
OCSN_227	NGC_2232	−0.09	0.09	6	Zhong2020
OCSN_228	NGC_2281	−0.08	0.11	73	Zhong2020
OCSN_234	NGC_2632	+0.16	0.07	21	Netopil2022
OCSN_241	NGC_6633	−0.10	0.05	8	Zhong2020
OCSN_255	RSG_1	−0.01	0.09	20	Zhong2020
OCSN_256	RSG_5	+0.07	0.09	13	Zhong2020
OCSN_259	Roslund_6	−0.01	0.10	32	Zhong2020
OCSN_261	Ruprecht_147	+0.12	0.03	33	Donor2020
OCSN_265	Stock_2	−0.11	0.07	19	Zhong2020
OCSN_266	Stock_10	−0.13	0.09	23	Zhong2020

(This table is available in machine-readable form.)

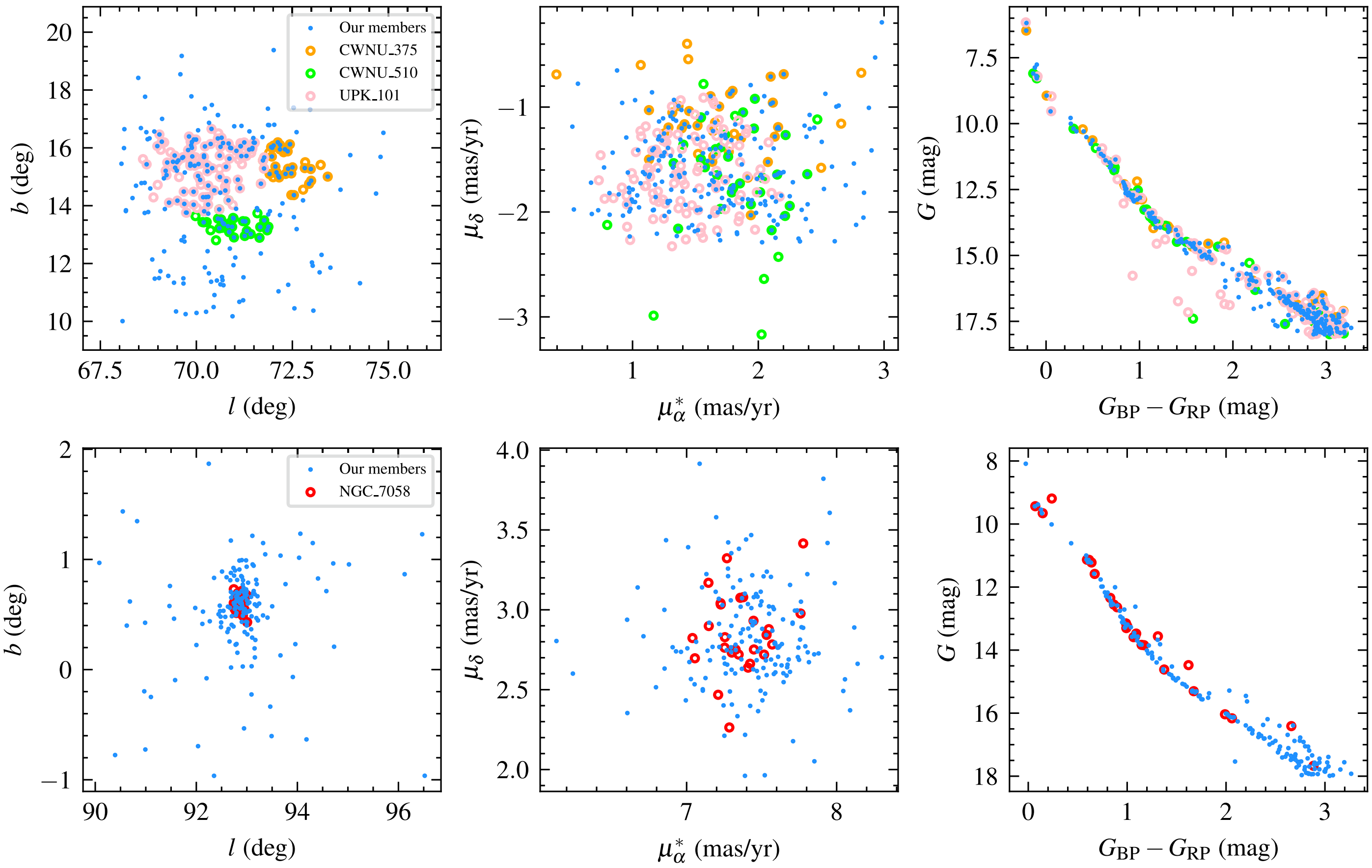

Figure 7. Example of the results of our searching procedure for two clusters. For each cluster, the three scatter plots (from left to right) represent the Galactic coordinate distribution, proper motion distribution, and CMD. The blue dots refer to the members identified in this work. The orange, green, pink, and red circles represent the members of CWNU_375 and CWNU_510 from He22a, UPK_101 from Sim19, and NGC_7058 from CG20, respectively. CWNU_375 and UPK_101 have four common members and CWNU_510 and UPK_101 have three common members in previous results.

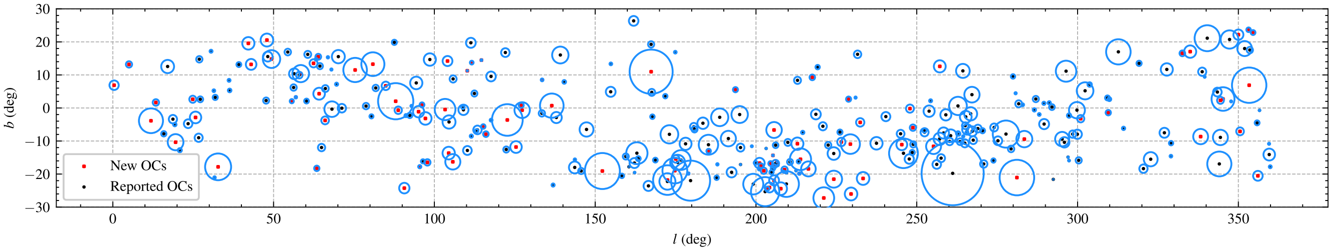

Figure 8. Distribution of the OC population in Galactic coordinates. The black points represent the reported OCs found in this study, and the red squares represent the OCs newly identified in this work using Gaia DR3. The blue circles refer to the r_h values given in this work.

hereafter P21), stellar "snake" (Tian 2020; Wang et al. 2022; hereafter T20 and W22, respectively), Lupus association (Galli et al. 2020b; hereafter G20b), etc. We combined these published OCs as well as young associations and moving groups related to giant molecular clouds to create a reference cluster catalog (hereafter ref_OCs).

In order to analyse carefully the differences between the OCSN catalog and the ref_OCs catalog, we created a common sample for comparison. However, it is not appropriate to crossmatch with two cluster catalogs through their center celestial coordinates. This is because the search area in our work (OCSN catalog) is systematically larger than many previous works (ref_OCs catalog), and the deviation of the cluster center coordinates derived by the different methods in the different works may be very large. Hence, we adopted crossmatching with all cluster members rather than the cluster itself. Finally, with a matching radius of 1″, a common catalog of about 332,000 members was obtained, which is referred to as the common_memb catalog.

For each cluster, we calculated the fraction (f) between the common member stars to members in the OCSN catalog and then assigned a flag according to this fraction. Meanwhile, we checked the position, proper motion, and parallax distributions of those common members through visual inspection to assess whether they are the same clusters. In our catalog, we provided the corresponding OC_flag of the cluster as well as its literature name in Table 2, which mainly have three cases below:

1. OC_flag = 1: $f = 0\%$, the new OCs in the OCSN catalog.
2. OC_flag = 2: $f < 50\%$ and most of the common members are located on the outer part of the cluster in the OCSN catalog.

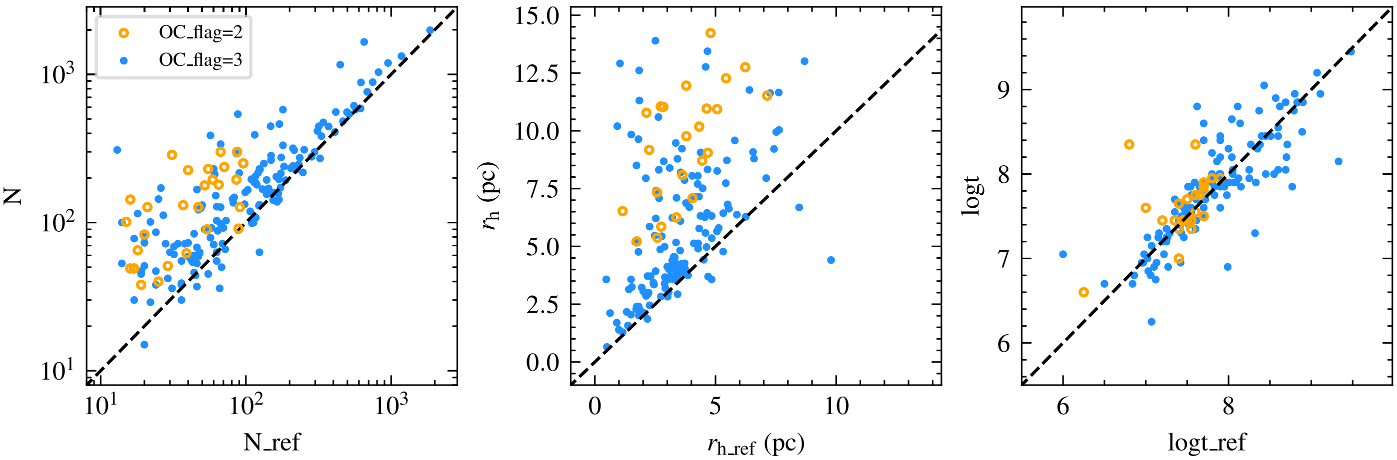

Figure 9. A comparison of the distributions of member numbers, r_h, and age parameters between the reported clusters from ref_OCs and the OCs obtained in this work. The blue dots refer to the OCs with "OC_flag" = 3 and the orange circles refer to the OCs with "OC_flag" = 2. The black dashed lines both show the relation of $y = x$.

Table 5
Binary Clusters in Our OC Sample

Group	Cluster1	Cluster2	Δpos (pc)	ΔV (km s^{-1})
1	OCSN_16	OCSN_18	15.9	1.4
2	OCSN_29	OCSN_286	16.8	2.8
3	OCSN_40	OCSN_41	17.1	1.0
3	OCSN_40	OCSN_158	17.5	3.5
4	OCSN_50	OCSN_51	12.3	3.4
5	OCSN_91	OCSN_92	6.5	3.6
5	OCSN_91	OCSN_237	19.9	1.6
5	OCSN_92	OCSN_237	17.1	2.3
6	OCSN_98	OCSN_100	13.1	1.3
7	OCSN_118	OCSN_271	16.5	1.0
8	OCSN_124	OCSN_184	16.8	3.8
9	OCSN_127	OCSN_129	13.2	1.7
10	OCSN_128	OCSN_283	18.1	2.4
11	OCSN_176	OCSN_178	18.9	0.8
11	OCSN_177	OCSN_178	12.7	2.1
12	OCSN_187	OCSN_188	5.6	4.3
13	OCSN_197	OCSN_198	18.2	1.4
14	OCSN_245	OCSN_246	10.0	3.5
15	OCSN_322	OCSN_324	10.9	2.4

(This table is available in machine-readable form.)

3. OC_flag = 3: $f > 50\%$ or most of the common members are located in the center part of the cluster in the OCSN catalog.

For clusters whose $f < 50\%$, we further inspected the five-dimensional distributions of their members. For instance, the top panels in Figure 7 show the spatial, proper motion, and color–magnitude distributions of members of the three reported clusters. The reported clusters CWNU_375 and CWNU_510 are both published by He22a, and UPK_101 is published by Sim19 . The space distribution clearly shows that the cluster identified in the OCSN catalog incorporates the three reported clusters. In this case, the three reported clusters are regarded as reference clusters that combined as one cluster with OC_flag = 2 in the OCSN catalog (see Table 2). Similarly, the bottom panels in Figure 7 show another cluster case with OC_flag = 3. For this cluster, our work identified more members in a wider range, while the previously reported members of the cluster (NGC_7058) in CG20 are only a core component of this cluster. Finally, 25 OCs with OC_Flag = 2 and 198 OCs with OC_Flag = 3 were reported in the previous catalogs, and 101 new OCs with OC_Flag = 1 were not presented in any literature studies. We noticed that about 10% of our cataloged clusters coincided with known associations and moving groups, which clearly indicates that not all of the clusters are bound OCs.

Figure 8 shows the distribution of 324 OCSN clusters in the Galactic coordinates, while the blue circles represent the half-number radius (r_h) of each cluster. Furthermore, we use the red squares and black dots to present the new OCs (OC_flag = 1) and reported OCs (OC_flag = 2, 3), respectively. It is worth noting that many nearby clusters with large spatial scales (r_h) were discovered for the first time in our work. This also shows that our slicing approach is very effective for searching nearby OCs with a large spatial distribution.

For OCSN clusters that have been previously reported, we also compared their number of members, the half-number radius in physical sizes, and the age with the literature results. Figure 9 shows the comparison results, while the orange circles and blue dots present clusters with OC_flag = 2 and OC_flag = 3, respectively. It can be seen that our cluster sample (especially clusters with OC_flag = 2) contains more member stars than the literature results. Moreover, because the r_h of most clusters is larger than previous results, the updated members in the cluster present a more extended spatial distribution. Our OCSN catalog expands the physical size of many nearby star clusters. On the other hand, because we only increased the number of member stars, the ages of the isochrone-fitting results of the star clusters are still consistent with the literature results.

5.2. Binary Clusters

OCs are born in giant molecular clouds and in some cases also form in groups (Camargo et al. 2016). A number of OCs are found in pairs or higher-order systems (Rozhavskii et al. 1976; Subramaniam et al. 1995; Soubiran et al. 2019). As the most famous double star clusters, h and χ Persei have been extensively studied and many interesting results have been obtained (Slesnick et al. 2002; Zhong et al. 2019; Li et al.

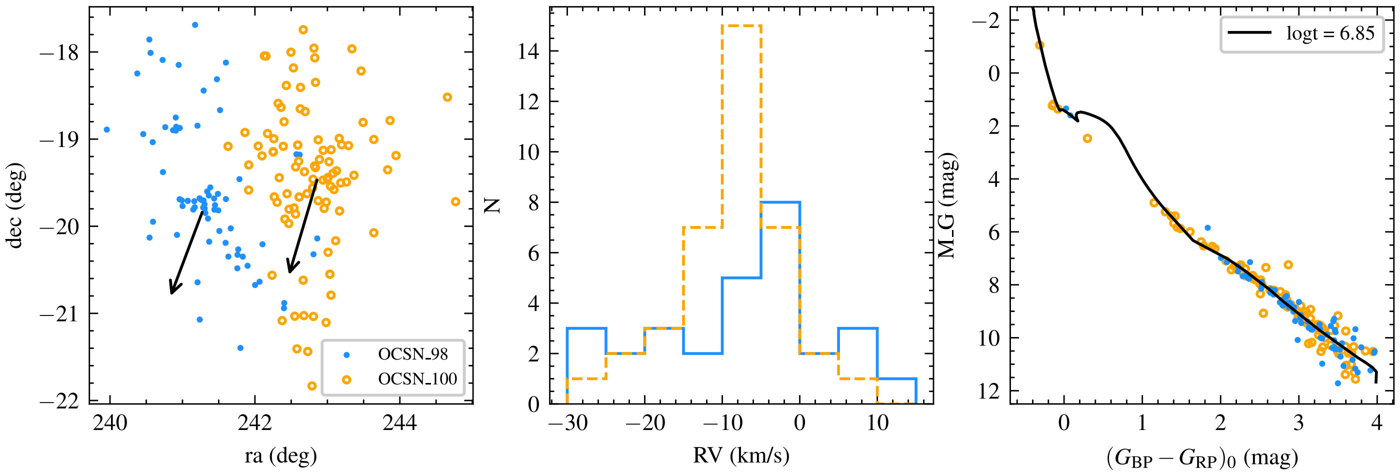

Figure 10. An example of binary cluster in our samples, OCSN_98 in blue and OCSN_100 in orange. Left panel: the spatial distributions of the two OCs, and the black arrows refer to their projected moving direction. Middle panel: histograms of the RVs of their members. Right panel: absolute CMDs for the two OCs, and the black line refers to an isochrone with a solar metallicity of $Z_\odot = 0.0152$ (Caffau et al. 2009, 2011) and a logarithm of age of log $t = 6.85$.

2019). Based on the high-precision kinematic information provided by Gaia data, more and more binary clusters have been confirmed (Soubiran et al. 2019; Casado 2021; Bisht et al. 2021; Angelo et al. 2022), which are important for studying the formation and dynamical evolution of OCs (de La Fuente Marcos & de La Fuente Marcos 2009; de la Fuente Marcos & de la Fuente Marcos 2010; Arnold et al. 2017). We performed a preliminary screening in our cluster samples with spatial separations of Δpos < 20 pc (Subramaniam et al. 1995) and velocity differences of $\Delta V < 5$ km s^{-1} (Soubiran et al. 2019). As a result, we got 15 groups of OCs including binary and triple cluster systems. The results are listed in Table 5.

In our catalog, there are 19 OC pairs with a common origin, whose age differences are less than 30 Myr. We show a pair of two new OCs in Figure 10: OCSN_98 and OCSN_100. It is clear that the mean central position of the two clusters is very close (Δpos = 13 pc), while the tangential velocities (black arrows in the left panel) and RVs (histogram in the middle panel) are also similar. The total velocity difference between the two clusters is $\Delta V = 1.3$ km s^{-1}. The absolute CMDs of the two clusters are presented in the right panel, which presents the same visual fitting age with log $t = 6.85$. The similarity of the two clusters suggests that they may have a common origin.

Furthermore, along with more new star clusters being added, we found three groups (Group 3, Group 5, and Group 11 in Table 5) containing triple OCs, whose age differences are less than 10 Myr. For example, Group 3 comprises two new OCs and one already reported OC: OCSN_40, OCSN_41, and OCSN_158. As shown in Table 5, the two pairs of OCSN_40, OCDN_41 and OCSN_40, OCSN_158 have similar positions (Δ pos ≈ 17 pc) and similar velocities ($\Delta V < 3.5$ km s^{-1}). At the same time, since these three clusters have almost the same age (log $t = 7.25$, 7.20, and 7.20), it can be inferred that the triple clusters also formed together from the same molecular cloud.

6. Summary

In this paper, we performed a systematic blind search for OCs at Galactic latitudes $|b| \leqslant 30^\circ$ within 500 pc of the solar neighborhood by choosing different slicing box sizes in different distance grids with Gaia DR3 data. By utilizing the clustering algorithms pyUPMASK and HDSBSCAN, we identified a total of 324 OCs. Our results include 101 new clusters that were never reported before, increasing the OC census within 500 pc by about 50%. Meanwhile, 223 reported clusters and their members were updated by carefully comparing their spatial distributions and other properties with a previous cluster catalog (ref_OCs). In the OCSN catalog, we provided the membership probabilities of member stars and further estimated the mean positions, proper motions, parallaxes, and structural parameters for each cluster. We also derived the mean RVs of the OCs through Gaussian fitting based on Gaia DR3. Subsequently, we performed visual isochrone fitting to obtain the ages, distance moduli, and reddening values for the clusters according to the distribution of member stars on the CMDs.

In particular, we compared the star clusters in the literature with our star clusters and use an OC_flag to classify the OCSN clusters into three samples. Our classification based on manual inspection not only marks new clusters but also combines some duplicate or partially reported clusters. Additionally, 19 pairs of clusters were identified as binary clusters in the solar neighborhood, and three groups of OCs were confirmed as triple cluster systems, with a spatial separation less than 20 pc, a velocity difference less than 5 km s^{-1}, and similar ages.

For our hunted OC samples within 500 pc in the solar neighborhood, more detailed analyses are needed to investigate further their properties, such as the mass function and the dynamical states. Especially more spectroscopic data for the member stars will be of prime importance to determine the dynamical and chemical evolution of these clusters.

This work is supported by the National Key R&D Program of China No. 2019YFA0405501. L.C. acknowledges support from the National Natural Science Foundation of China (NSFC) through grants 12090040 and 12090042. J.Z. would like to acknowledge the NSFC under grant 12073060, and the Youth Innovation Promotion Association CAS. We acknowledge the science research grants from the China Manned Space Project with NO. CMS-CSST-2021-A08.

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.

ORCID iDs

Songmei Qin (秦松梅) https://orcid.org/0000-0003-3713-2640
Jing Zhong (钟靖) https://orcid.org/0000-0001-5245-0335
Tong Tang (唐通) https://orcid.org/0000-0003-1864-8721
Li Chen (陈力) https://orcid.org/0000-0002-4907-9720

References

Abdurro'uf, K., Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35
Angelo, M. S., Santos, J. F. C., Maia, F. F. S., & Corradi, W. J. B. 2022, MNRAS, 510, 5695
Arnold, B., Goodwin, S. P., Griffiths, D. W., & Parker, R. 2017, MNRAS, 471, 2498
Bai, L., Zhong, J., Chen, L., Li, J., & Hou, J. 2022, RAA, 22, 055022
Beccari, G., Boffin, H. M. J., & Jerabkova, T. 2020, MNRAS, 491, 2205
Bisht, D., Zhu, Q., Yadav, R. K. S., et al. 2021, MNRAS, 503, 5929
Boffin, H. M. J., Jerabkova, T., Beccari, G., & Wang, L. 2022, MNRAS, 514, 3579
Caffau, E., Ludwig, H.-G., Steffen, M., Freytag, B., & Bonifacio, P. 2011, SoPh, 268, 255
Caffau, E., Maiorca, E., Bonifacio, P., et al. 2009, A&A, 498, 877
Camargo, D., Bica, E., & Bonatto, C. 2016, MNRAS, 455, 3126
Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, PAKDD 2013: Advances in Knowledge Discovery and Data Mining (Berlin: Springer), 160
Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93
Cantat-Gaudin, T., Jordi, C., Wright, N. J., et al. 2019, A&A, 626, A17
Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1
Carrera, R., Pasquato, M., Vallenari, A., et al. 2019, A&A, 627, A119
Casado, J. 2021, ARep, 65, 755
Casagrande, L., & VandenBerg, D. A. 2018, MNRAS, 479, L102
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
Chaturvedi, A., Green, P. E., & Caroll, J. D. 2001, J. Classif., 18, 35
de La Fuente Marcos, R., & de La Fuente Marcos, C. 2009, A&A, 500, L13
de la Fuente Marcos, R., & de la Fuente Marcos, C. 2010, ApJ, 719, 104
Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871
Dias, W. S., & Lépine, J. R. D. 2005, ApJ, 629, 825
Dickson-Vandervelde, D. A., Wilson, E. C., & Kastner, J. H. 2021, AJ, 161, 87
Donor, J., Frinchaboy, P. M., Cunha, K., et al. 2020, AJ, 159, 199
Ester, M., Kriegel, H. P., Sander, J., & Xu, X. 1996, in KDD-96 Proc. of the Second Int. Conf. on Knowledge Discovery andData Mining (Palo Alto, CA: AAAI Press), 226
Evans, N. J. I., Dunham, M. M., Jørgensen, J. K., et al. 2009, ApJS, 181, 321
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2022, arXiv:2206.05992
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, Vallenari, A., Brown, A. G. A., et al. 2022, arXiv:2208.00211
Galli, P. A. B., Bouy, H., Olivares, J., et al. 2020a, A&A, 634, A98
Galli, P. A. B., Bouy, H., Olivares, J., et al. 2020b, A&A, 643, A148
Ginsburg, A., Sipocz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98
Grasser, N., Ratzenböck, S., Alves, J., et al. 2021, A&A, 652, A2
He, Z., Wang, K., Luo, Y., et al. 2022a, ApJS, 262, 7
He, Z., Li, C., Zhong, J., et al. 2022b, ApJS, 260, 8
Heyl, J., Caiazzo, I., & Richer, H. B. 2022, ApJ, 926, 132
Huchra, J. P., & Geller, M. J. 1982, ApJ, 257, 423
Hunt, E. L., & Reffert, S. 2021, A&A, 646, A104
Janes, K., & Adler, D. 1982, ApJS, 49, 425
Katz, D., Sartoretti, P., Guerrier, A., et al. 2022, arXiv:2206.05902
Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Roser, S., & Röser, R.-D. 2013, A&A, 558, A53
King, I. 1962, AJ, 67, 471
King, I. R. 1966, AJ, 71, 276
Kounkel, M., Covey, K., Suárez, G., et al. 2018, AJ, 156, 84
Krone-Martins, A., & Moitinho, A. 2014, A&A, 561, A57
Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57
Li, C., Sun, W., Grijs, R. d., et al. 2019, ApJ, 876, 65
Li, Z., Deng, Y., Chi, H., et al. 2022, ApJS, 259, 19
Lindegren, Z. 2018, Renormalising the Astrometric Chi-square in Gaia DR2GAIA-C3-TN-LU-LL-124-01Lund Observatory, http://www.rssd.esa.int/doc_fetch.php?id=3757412
Liu, J., Fang, M., Tian, H., et al. 2021, ApJS, 254, 20
Liu, L., & Pang, X. 2019, ApJS, 245, 32
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
McInnes, L., Healy, J., & Astels, S. 2017, JOSS, 2, 205
Meingast, S., & Alves, J. 2019, A&A, 621, L3
Meingast, S., Alves, J., & Rottensteiner, A. 2021, A&A, 645, A84
Netopil, M., Oralhan, İ A., Çakmak, H., Michel, R., & Karataş, Y. 2022, MNRAS, 509, 421
Netopil, M., Paunzen, E., & Carraro, G. 2015, A&A, 582, A19
Pang, X., Yu, Z., Tang, S.-Y., et al. 2021, ApJ, 923, 20
Pavlidou, T., Scholz, A., & Teixeira, P. S. 2021, MNRAS, 503, 3232
Pearson, K. 1894, RSPTA, 185, 71
Pera, M. S., Perren, G. I., Moitinho, A., Navone, H. D., & Vazquez, R. A. 2021, A&A, 650, A109
Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, ARA&A, 48, 431
Riello, M., de Angeli, F., Evans, D. W., et al. 2021, yCat, J/A+A/649/A3
Ripley, B. D. 1976, J. Appl. Probab., 13, 255
Roccatagliata, V., Sacco, G. G., Franciosini, E., & Randich, S. 2018, A&A, 617, L4
Rodriguez, A., & Laio, A. 2014, Sci, 344, 1492
Rozhavskii, F. G., Kuz'mina, V. A., & Vasilevskii, A. E. 1976, Ap, 12, 204
Scott, D. W. 1992, Multivariate Density Estimation: Theory, Practice, and Visualization (New York: Wiley)
Scott, D. W. 2015, Multivariate Density Estimation: Theory, Practice, and Visualization (2nd ed.; New York: Wiley), doi:10.1002/9780470316849
Sculley, D. 2010, in WWW '10: Proc. of the 19th Int. Conf. on the World Wide Web (New York: Association for Computing Machinery), 1177
Sim, G., Lee, S. H., Ann, H. B., & Kim, S. 2019, JKAS, 52, 145
Slesnick, C. L., Hillenbrand, L. A., & Massey, P. 2002, ApJ, 576, 880
Soubiran, C., Cantat-Gaudin, T., Romero-Gómez, M., et al. 2019, A&A, 623, C2
Subramaniam, A., Gorti, U., Sagar, R., & Bhatt, H. C. 1995, A&A, 302, 86
Tarricq, Y., Soubiran, C., Casamiquela, L., et al. 2022, A&A, 659, A59
Tian, H.-J. 2020, ApJ, 904, 196
Voronoi, M. G. 1908, Journal fur die Reine und Angewandte Mathematik, 133, 97
Wang, F., Tian, H., Qiu, D., et al. 2022, MNRAS, 513, 503
Ye, X., Zhao, J., Oswalt, T. D., Yang, Y., & Zhao, G. 2022, AJ, 164, 132
Zepeda-Mendoza, M. L., & Resendis-Antonio, O. 2013, Encyclopedia of Systems Biology (New York: Springer), 886
Zhang, Y., Tang, S.-Y., Chen, W. P., Pang, X., & Liu, J. Z. 2020, ApJ, 889, 99
Zhong, J., Chen, L., Jiang, Y., Qin, S., & Hou, J. 2022, AJ, 164, 54
Zhong, J., Chen, L., Kouwenhoven, M. B. N., et al. 2019, A&A, 624, A34
Zhong, J., Chen, L., Wu, D., et al. 2020, A&A, 640, A127