Gaia keeps on delivering: expanding the open cluster population with EDR3

Castro Ginard, A.

Citation

Castro Ginard, A. (2022). Gaia keeps on delivering: expanding the open cluster population with EDR3. doi:10.5281/zenodo.7104494

Gaia keeps on delivering: expanding the open cluster population with EDR3

Alfred Castro-Ginard

Leiden Observatory, Leiden University. Niels Bohrweg 2, 2333 CA Leiden, The Netherlands.

The improvements in the precision of the astrometric data, particularly of the parallax and proper motion measurements, of *Gaia* EDR3 with respect to *Gaia* DR2 provides the opportunity to search for farther and fainter open clusters (OCs) that have so far been undetected. This will allow us to have a more complete view of the OC population in the Milky Way, enabling studies about the structure and the evolution of our Galaxy.

The `OCfinder` method

In order to blindly search for OCs in *Gaia* data, we have developed the method `OCfinder` [see 2, for a detailed explanation]. The method, devised in [2] and successfully applied to find hundreds of new open clusters in [3, 4, 6], consists in two well-known machine learning algorithms. First, we apply a density-based clustering method, DBSCAN [7], to find statistical stellar overdensities in the five astrometric dimensions (*i.e.* position, parallax and proper motion). Second, we classify these overdensities into random statistical or real open clusters using an artificial neural network [ANN, 8] trained to identify isochrones in a color-magnitude diagram (CMD).

We apply the methodology to $232\,463\,114$ stars in the Galactic disc, defined as $|b| < 20^\circ$, up to magnitude $G = 18$. We enable the analysis of such a large number of stars thanks to the deployment of the methodology in a Big Data environment [4], *i.e.* the MareNostrum Supercomputer located in the Barcelona Supercomputing Centre[**].

New open clusters in EDR3

The application of `OCfinder` to *Gaia* EDR3 data results in the discovery of 628 new open clusters, together with the re-detection of about 80% of the open clusters characterised with *Gaia* DR2 data [1]. As shown in Figure 1, the OCs found in *Gaia* EDR3 are in general more compact in both proper motion and parallax. This is due to i) the improvements on the astrometric precision of *Gaia* EDR3 with respect to *Gaia* DR2, and ii) the fact that we expand the search up to $G = 18$ (compared to our previous searches up to $G = 17$), which allow us to detect farther objects. In fact, out of the 628 new OC, only three objects are closer than 1 kpc, and 75 objects are located within 1 and 2 kpc. This is due to a better completeness of previous surveys in these regions, together with the aforementioned methodological effects.

The Table with the new UBC clusters can be found online at the CDS[††]. It contains, for each cluster, information about the mean astrometric parameters (including radial velocity when available), together with the number of members of each cluster and an estimation of the age, distance and line-of-sight extinction. The clusters are divided into class A, class B and class C, according to their reliability which we assess by inspecting the distributions of the member stars in the astrometric dimensions and the CMD, aided with the radial velocity measurements when available.

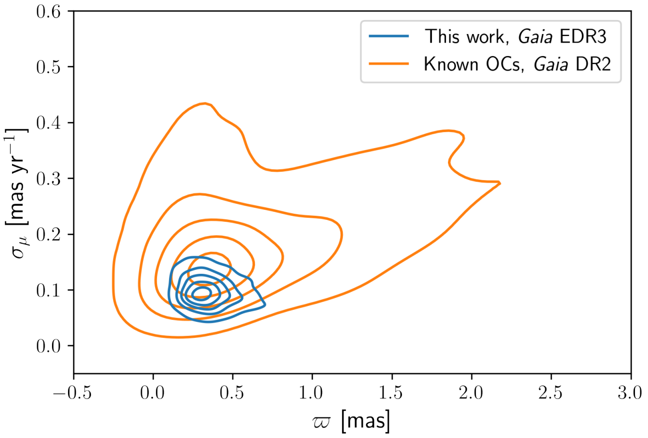

Figure 1: Total proper motion dispersion as a function of parallax for the OC characterised in DR2 (orange) and EDR3 (blue).

Ages, distances and line-of-sight extinctions

For each of the newly detected OCs, we estimate their age, distance and line-of-sight extinction. The estimation is done using an ANN trained on well-characterised open clusters [see 1, for details on the ANN]. The ANN takes the CMD of the OC member stars, together with the OC mean parallax and two additional quantities derived from the CMD to aid the estimation. By comparing their results with the reported values of the well-characterised OC in the training set, [1] report the uncertainties of their estimation of log age to be in the range from 0.15 to 0.25 dex for young clusters (≤ 8.5 dex), and from 0.1 to 0.2 dex for older OCs. The uncertainties for the extinctions range from 0.1 to 0.2 dex in A_v, and in the case of the distance modulus the uncertainties are also within 0.1 and 0.2 dex, which correspond to a 5% to 10% uncertainty in distance.

Figure 2 shows a heliocentric spatial distribution of the known previous to *Gaia* EDR3 and newly detected OCs in different age bins. The increase in the OC population in the different age bins is seen when comparing known OC (triangles) and newly detected in *Gaia* EDR3 (crosses): (i) 703 and 276 known and new OC younger than 100 Myr, (ii) 675 and 248 known and new OCs from 100 to 500 Myr, (iii) 229 and 58 known and new OCs with ages within 500 Myr and 1 Gyr, and (iv) 260 and 46 known

---

[**] `https://www.bsc.es/marenostrum`

[††] `http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/661/A118`

and new OCs older than 1 Gyr. Moreover, in the younger bin, we find clear overdensities that trace the spiral arms [5]. These young overdensities are dispersed in the following older age intervals.

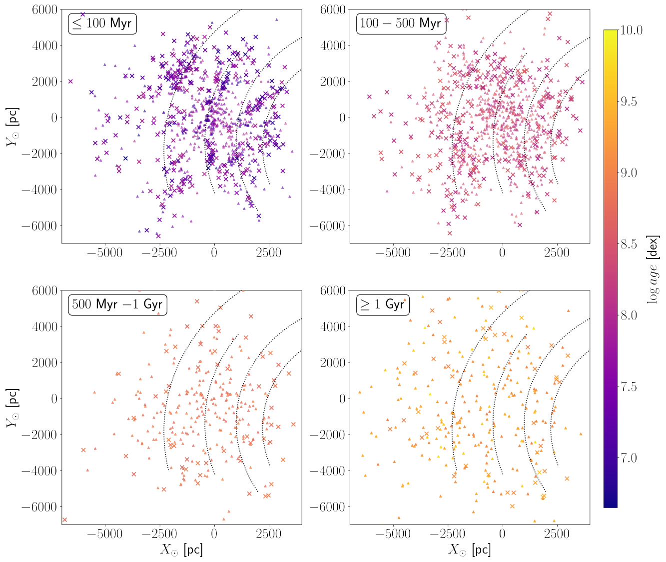

Figure 2: $X_\odot$ *vs* $Y_\odot$ distribution of the newly found OC (crosses) and OC characterised with DR2 data (triangles), for different age bins: younger than 100 Myr *(top left)*, $100 - 500$ Myr *(top right)*, 500 Myr - 1 Gyr *(bottom left)* and older than 1 Gyr *(bottom right)*. The dotted lines show the spiral arms fitted in [5].

OCs as tracers of the Milky Way spiral arms

As already seen in Figure 2 (top left panel), overdensities of young OCs trace the present-day locations of spiral arms. We used Gaussian mixture models to detect these overdensities and relate them to the Perseus, Local, Sagittarius and Scutum spiral arms [see 5, for a description of the method and a summary of the main results]. Once the present-day spiral arms are described, we trace back the OCs associated with the each spiral arm to their birth position [using radial velocities compiled in 9] to find the evolution of these spiral arms for the last 80 Myr [5].

We find that each spiral arm has its own pattern speed, discarding a common pattern speed for all the spiral arms explored. Moreover, the different spiral pattern speeds decrease as their Galactocentric radius increase, closely following the Galactic rotation curve. This behaviour of the Galactic spiral arms allow us to disfavour classical density waves as the main mechanism for the formation of the Milky Way spiral arms, favouring short-lived transient spirals.

Summary and outlook

This contribution presented the `OCfinder` method, devised to automatically search for OC in *Gaia* data, and its application to *Gaia* EDR3. So far, about $1\,300$ OCs have been detected with `OCfinder`, representing almost 50% of the currently known OC population. This shows the importance of using machine-learning aided methodologies to extract knowledge and better understand the high-quality data provided by surveys as *Gaia*.

We also showed how an improved OC census, both in terms of quantity of OCs and quality of their estimated parameters, can improve the science we derive from this population. By tracing the Milky Way spiral arms with OCs, we found that the mechanisms driving the formation of the spiral arms in our Galaxy are in better agreement with simulation-based approaches that tend to favour a transient nature.

The next *Gaia* DR3 release, will provide more complete information about the Solar neighbourhood. In particular, and relevant for `OCfinder`, *Gaia* DR3 will add radial velocity measurements for 33 million stars with $G_{RVS} \leq 14$, adding an extra dimension to not only help the search for new OCs but also to improve the characterisation of the known OC population. The volume of available high-quality data will only increase with future *Gaia* data releases, as well as with future photometric and spectroscopic surveys, therefore the need for improved and more robust data-driven machine-learning methodologies.

References

[1] Cantat-Gaudin, T., et al. 2020, A&A, 640, 1.
[2] Castro-Ginard, A., et al. 2018, A&A, 618, 59.
[3] Castro-Ginard, A., et al. 2019, A&A, 627, 35.
[4] Castro-Ginard, A., et al. 2020, A&A, 635, 45.
[5] Castro-Ginard, A., et al. 2021, A&A, 652, 162.
[6] Castro-Ginard, A., et al. 2022, A&A, 661, 118.
[7] Ester, M., et al. 1996, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining KDD'96 (AAAI Press) 226.
[8] Hinton, G., 1989, Artif. Intell., 40, 185.
[9] Tarricq, Y., et al. 2021, A&A, 647, 19.

Short CV

2016–2017: MSc in Astronomy, University of the Balearic Islands, Spain
2017–2021: PhD in Astronomy, University of Barcelona, Spain
2021–present: Postdoctoral researcher, Leiden Observatory, The Netherlands