A&A 624, A126 (2019)
https://doi.org/10.1051/0004-6361/201834453

Astronomy & Astrophysics

Gaia DR2 unravels incompleteness of nearby cluster population: new open clusters in the direction of Perseus⋆

T. Cantat-Gaudin[1], A. Krone-Martins[2], N. Sedaghat[3], A. Farahi[4], R. S. de Souza[5], R. Skalidis[6,7], A. I. Malz[8,9], S. Macêdo[10], B. Moews[11], C. Jordi[1], A. Moitinho[2], A. Castro-Ginard[1], E. E. O. Ishida[12], C. Heneka[13], A. Boucaud[14,15], and A. M. M. Trindade[16] for the COIN Collaboration

[1] Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
e-mail: tcantat@fqa.ub.edu
[2] CENTRA, Faculdade de Ciências, Universidade de Lisboa, Ed. C8, Campo Grande, 1749-016 Lisboa, Portugal
[3] Department of Computer Science, University of Freiburg, Georges-Koehler-Allee 052, 79110 Freiburg, Germany
[4] Department of Physics, McWilliams Center for Cosmology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA
[5] Department of Physics & Astronomy, University of North Carolina at Chapel Hill, NC 27599-3255, USA
[6] Department of Physics and Institute for Theoretical and Computational Physics, University of Crete, 71003 Heraklion, Greece
[7] Foundation for Research and Technology – Hellas, IESL, Voutes, 7110 Heraklion, Greece
[8] Center for Cosmology and Particle Physics, New York University, 726 Broadway, New York, NY 10004, USA
[9] Department of Physics, New York University, 726 Broadway, New York, NY 10004, USA
[10] Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco, IFPE, 50670-430, Recife-PE, Brazil
[11] Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK
[12] Université Clermont Auvergne, CNRS/IN2P3, LPC, 63000 Clermont-Ferrand, France
[13] Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
[14] LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France
[15] Paris-Saclay Center for Data Science, Université Paris-Saclay, Orsay, France
[16] Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal

Received 18 October 2018 / Accepted 14 March 2019

ABSTRACT

Context. Open clusters (OCs) are popular tracers of the structure and evolutionary history of the Galactic disc. The OC population is often considered to be complete within 1.8 kpc of the Sun. The recent *Gaia* Data Release 2 (DR2) allows the latter claim to be challenged.

Aims. We perform a systematic search for new OCs in the direction of Perseus using precise and accurate astrometry from *Gaia* DR2.

Methods. We implemented a coarse-to-fine search method. First, we exploited spatial proximity using a fast density-aware partitioning of the sky via a k-d tree in the spatial domain of Galactic coordinates, (l, b). Secondly, we employed a Gaussian mixture model in the proper motion space to tag fields quickly around OC candidates. Thirdly, we applied an unsupervised membership assignment method, UPMASK, to scrutinise the candidates. We visually inspected colour-magnitude diagrams to validate the detected objects. Finally, we performed a diagnostic to quantify the significance of each identified over-density in proper motion and in parallax space.

Results. We report the discovery of 41 new stellar clusters. This represents an increment of at least 20% of the previously known OC population in this volume of the Milky Way. We also report on the clear identification of NGC 886, an object previously considered an asterism. This study challenges the previous claim of a near-complete sample of OCs up to 1.8 kpc. Our results reveal that this claim requires revision, and a complete census of nearby OCs is yet to be found.

Key words. open clusters and associations: general – methods: numerical

1. Introduction

Galactic stellar clusters, traditionally called open clusters (OCs), are tracers of the structure and evolution of the Milky Way (e.g. Janes & Adler 1982; Dias & Lépine 2005; Piskunov et al. 2006; Moitinho 2010), playing a fundamental role in studies of star formation environment and evolution (e.g. Lada 2010). These clusters are found primarily in the Galactic plane and are composed of dozens to several thousands of stars of similar age, metallicity, kinematics, and distance. As such, they provide a testbed for stellar evolution models.

It has been often stated in the literature that the census of the OC population in the solar neighbourhood is complete. Piskunov et al. (2006) put the completeness radius at 850 pc. This was contested by Moitinho (2010), who showed that, based on the photometry used by Piskunov et al. (2006), many sparse and/or old OCs are expected to be missed. Kharchenko et al. (2013), working with Two Micron All-Sky Survey photometry (2MASS; Skrutskie et al. 2006), made the claim that the OC sample is almost complete within 1.8 kpc from the Sun. Few discoveries of nearby OCs had been made since the publications of Alessi et al. (2003) and Kharchenko et al. (2005; with the notable

⋆ The list of members are only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/624/A126 exception of Röser et al. 2016 who identified nine new objects within 500 pc), making the completeness claim appear plausible. While the astrometric and photometric analysis followed in the aforementioned works provide a well-established methodology for identifying members and deriving properties of clusters, a significant part of the problem seems to arise from the data used in those studies. The catalogues they are based on, such as HIPPARCOS (ESA 1997), 2MASS (Skrutskie et al. 2006), PPMXL (Roeser et al. 2010), or the United States Naval Observatory CCD Astrograph Catalog (UCAC; e.g. Zacharias et al. 2013) have brought enormous contributions to astronomy, but as any data set they have sensitivity limitations and should not be overinterpreted. As noted in Moitinho (2010), many objects fall on the borderline of the data sensitivity limits, which causes incompleteness and creates false positives in the sample.

The improvement in quality and the depth of the recently published *Gaia* Data Release 2 (DR2; Gaia Collaboration 2018) allow us to identify new objects and address the question of the reality of previously identified objects. In a recent publication, Castro-Ginard et al. (2018) reported on the discovery of 11 OCs closer than 500 pc in the *Gaia* DR1 data (Gaia Collaboration 2016). Cantat-Gaudin et al. (2018a) have identified 34 new OCs within 2 kpc in the *Gaia* DR2 data and have shown that the samples on which the completeness analyses are based are also highly contaminated by false positives.

In this study we search for known and unknown stellar clusters and make use of stars in the *Gaia* DR2 catalogue up to magnitude $G = 18.0$, using a coarse-to-fine search methodology combining machine learning and statistical techniques. Recent data releases of *Gaia* have changed the state of the field by providing accurate and precise measurements of stellar population kinematics and parallaxes. We use the spatial, kinematic, and parallax information in a hierarchical and automated fashion to scan the sky, flag the candidates, and validate them. To validate or reject the flagged candidates, we employ the UPMASK method (Krone-Martins & Moitinho 2014), which uses all of the above information simultaneously. We neither make use of photometric data to detect clusters nor to perform stellar membership. The photometric information is employed solely for an independent validation of each OC candidate by visual inspection of their colour-magnitude diagrams.

In Sect. 2, we briefly present the *Gaia* DR2 data and the source selection criteria adopted in this study. The methods developed and employed to search for OCs are detailed in Sect. 3. We present the results and comment on some specific objects in Sect. 4. Finally, we summarise our findings in Sect. 5. The mean astrometric parameters of all discovered objects are listed in Table A.1, and the list of their members is available at the CDS.

2. Data

The ESA *Gaia* space mission (Gaia Collaboration 2016) is producing an unprecedented all-sky survey in terms of its sheer size, dimensionality, and history-changing astrometric precision and accuracy. In particular, the most recent data release, *Gaia* DR2, contains more than 1.6 billion sources as faint as $G \sim 21$, providing five-parameter astrometric solutions accurate to hundreds of micro-arcseconds, as well as magnitudes in three photometric bands, for more than 1.3 billion sources. (Gaia Collaboration 2018). This unprecedented precision facilitates the identification of stellar clusters by increasing the contrast between the cluster members and field objects in the proper motion and parallax space.

The *Gaia* DR2 data were retrieved in two different ways. First, we used the *Gaia* Archive bulk retrieval data facility[1] to obtain the entire region defined by Galactic coordinates $l \in [120°; 200°]$ and $b \in [-10°; 10°]$ and magnitudes $G \leq 18$. This magnitude cut corresponds to typical astrometric uncertainties better than 0.3 mas yr^{-1} in proper motion and 0.15 mas in parallax. From these data, we selected the information for spatial tiling and candidate flagging (positions l, b, proper motions $\mu_{\alpha*}, \mu_\delta$, and parallax, ϖ; see Sects. 3.1 and 3.2). Secondly (described in Sect. 3.3), we validated the candidates querying data through the *Gaia* archive facility at the European Space Agency Centre (Salgado et al. 2017) with `pygacs`[2] to extract the positions (α, δ, l, b), parallaxes (ϖ), proper motions $(\mu_{\alpha*}, \mu_\delta)$, fluxes in the G, G_{BP} and G_{RP} passbands, and their associated uncertainties and covariances, from the table `gaiadr2.gaia_source` via Astronomical Data Query Language (ADQL; Ortiz et al. 2008) queries launched using the Table Access Protocol (TAP; Dowler et al. 2018).

3. Method

To perform a systematic search for OCs, we designed and employed a coarse-to-fine search algorithm tailored to be used on astrometric catalogues. Our algorithm consists of three steps: (i) fast density-aware sky tiling to exploit the proximity of cluster members on the plane of the sky; (ii) OC candidate region flagging based on proper motion similarity via a fast recommendation system; and (iii) cluster membership assignment and independent validation. We detail these steps in the following subsections.

3.1. Density-aware spatial tiling

As the very first step, we partitioned the sample into computationally tractable subsets, i.e. tiles. Partitioning data sets of *Gaia* size requires an extremely fast and scalable method. At the same time, we would like to avoid loss of spatial information, or, informally speaking, splitting a potential cluster across many tiles, which lowers the signal-to-noise ratio (S/N). Hence, the adopted algorithm should be aware of the spatial spread of data points. To this end, we employed k-d trees for spatial partitioning (Bentley 1975).

Traditionally used for fast nearest neighbour searches, k-d trees are data structures based on binary trees (Garnier & Taylor 2009). In this work, however, we used the accompanying fast partitioning algorithm to create a rectangular tessellation of the target sky region, which preserves a predefined minimum number of stars in each tile. These k-d trees are known to perform best when the number of dimensions (k) is not large (Friedman et al. 1977; Sproull 1991). We used the two Galactic spatial coordinates l and b, hence $k = 2$ to partition the region of interest with ~15 million sources. The algorithm outputs 2048 tiles, each containing ~7400 stars on average[3].

3.2. Candidate flagging

To quickly search over 2048 tiles and flag OC candidates, we employed a Gaussian mixture model (GMM; e.g. Pearson 1894; Melchior & Goulding 2018) in the proper motion space $(\mu_{\alpha*}, \mu_\delta)$.

[1] http://cdn.gea.esac.esa.int/Gaia/gdr2/gaia_source/csv/

[2] https://github.com/Johannes-Sahlmann/pygacs

[3] We make use of the *SciPy* (Jones et al. 2001) implementation of the k-d tree (cKDTree algorithm from Maneewongvatana & Mount 1999).

In this case, the GMM is used to decrease the clutter and information overload, rather than as an optimal model of the proper motion distribution. The Gaussian covariances are used as the score metric to decide on the follow-up analysis of the field.

A GMM is a parametric model that consists of a weighted sum of Gaussian components. The GMM allows us to separate a candidate region from the background and foreground sources because the measured stellar proper motions of OC members can be roughly approximated by a Gaussian distribution, where the means equal the bulk proper motion of the OC and a variance smaller than that of the field stars. Therefore, GMMs are a natural choice to flag candidate regions in which an OC may exist. This technique has been traditionally used to study and search OCs (Vasilevskis et al. 1958; Sanders 1971) and has been continuously applied in astronomy (e.g. Krone-Martins et al. 2010; Ducourant et al. 2017; de Souza et al. 2017).

We adopted GMMs with ten multivariate Gaussian components[4] to perform one global search in the entire data within each tile, and multiple targeted searches in parallax bins of 0.2 mas (increased by steps of 0.2 mas to contain at least 1000 objects per bin). The targeted searches mitigate false negatives caused by the relative S/N of the candidate regions when compared to the field population. Since we expect members of the same stellar cluster to share similar parallaxes, this parallax screening increases the search efficiency. Finally, the candidate region is kept for further scrutiny if its Gaussian component exhibits a variance smaller than 0.1 mas^2 yr^{-2} in both $\mu_{\alpha}*$ and μ_{δ} (corresponding to a standard deviation of ~0.3 mas yr^{-1}), and if it has at least ten members falling within its interquartile range.

3.3. Candidate analysis

To further analyse each candidate region, we employed an unsupervised method, called UPMASK, which relies on minimal physical assumptions about stellar clusters (Krone-Martins & Moitinho 2014). The key assumptions are that the cluster member stars must share some common properties, thereby being clustered in some parameter spaces (here proper motion and parallax), and that the spatial distribution of the member stars should be incompatible with a uniform spatial distribution at the same time. This method was already successfully applied to astrometric data (i.e. Cantat-Gaudin et al. 2018b,a), where it serendipitously revealed 60 new Milky Way stellar clusters while analysing previously known clusters.

After querying the *Gaia* archive data around the candidates, we applied UPMASK, the workflow of which is outlined below.

1. Sample a new data set drawn from the probability distribution functions defined by the original measurements and their reported covariance matrices.
2. Create small groups in the parameter space ($\mu_{\alpha*}, \mu_{\delta}, \varpi$) through a k-means clustering algorithm (Forgy 1965; Lloyd 1982) that has a large k with respect to the data set size, guaranteeing ~10–15 objects per group.
3. Test each small group for compatibility with a uniform spatial distribution in (l, b) based on the branch lengths of minimum spanning trees (e.g. Graham & Hell 1985), after which those that are compatible are discarded as field stars.
4. Repeat steps (2–3) until no star is discarded, and either the remaining are assigned as stellar cluster members at this iteration or all stars are discarded and no cluster is detected.
5. Repeat steps (1–4) up to the maximum number of iterations (in this study we performed the loop 100 times).
6. Compute a membership score as the frequency with which each star was assigned as a cluster member.

The resulting clustering score is therefore based on the 5D ($\alpha, \delta, \mu_{\alpha*}, \mu_{\delta}, \varpi$) information of each star and associated nominal uncertainties.

[4] The results are insensitive to the exact number of Gaussians, as long as this number is large enough to disentangle the background.

4. Identification of clusters in the Perseus direction

After an automated analysis and human visual inspection of all colour-magnitude diagrams and positional maps of the candidates, we ended up with 133 stellar aggregates that look like potential clusters. Forty-one are hitherto unreported clusters and an additional five are known objects that had not yet been identified in the *Gaia* DR2 data. The location of their members and their colour-magnitude diagrams are shown in Figs. B.1 to D.5. The full membership list is available at the CDS.

4.1. Re-identification of known objects

The majority of the identified aggregates correspond to known clusters. Of these 87 were already identified with *Gaia* DR2 astrometry by Cantat-Gaudin et al. (2018a) in a search that relied on prior information on the location and expected dimension of clusters from the literature, which allowed these authors to identify 227 objects in the area investigated in the present study. Our unsupervised search did not recover all of those 227, but was however able to recover five objects (Czernik 5, Czernik 15, FSR 0494, FSR 0519, and NGC 886) missed by Cantat-Gaudin et al. (2018a). In all five cases, the apparent sizes listed in the catalogues of Dias et al. (2002) and Kharchenko et al. (2013) are too small for the clusters to appear as a clear over-density in the field of view.

The nearest of these five mischaracterised clusters is NGC 886, an OC in Cassiopeia. This cluster was first observed by J. Herschel in 1829 (No. 214 in Herschel 1833) and is listed under its current name in the original New General Catalogue of Nebulæ and Clusters of Stars (Dreyer 1888). Rediscovered as Stock 6 in the 1950s, it was later flagged as “non-existent” in the Revised New General Catalogue of Nonstellar Astronomical Objects (RNGC; Sulentic et al. 1973). The catalogue of Dias et al. (2002) lists Stock 6, which has an apparent radius of 0.12°, but flags it as a dubious grouping, while Kharchenko et al. (2013) lists NGC 886, which has a total radius of 0.17°. In this study, we find a clear centrally concentrated distribution of stars within 0.5° of the reported position of NGC 886/Stock 6 (see Fig. D.4).

4.2. Newly discovered clusters

The remaining 41 groups are not (to the best of our knowledge) listed in the literature (Dias et al. 2002; Kharchenko et al. 2013; Schmeja et al. 2014; Scholz et al. 2015; Röser et al. 2016; Castro-Ginard et al. 2018; Ferreira et al. 2019). We remark that the colour-magnitude diagrams of most of the newly reported clusters present a broadened sequence (see Figs. B.1 to C.13) and red turn-off point (G_{BP}-G_{RP}~0.6 to more than 1), which suggests that they are affected by differential reddening. All these clusters are however clearly visible in astrometric space. Their positions and mean proper motions and parallaxes are given in Table A.1, along with the radius r_{50} containing half the cluster members we identified. We also provide a rough distance from the mean parallax, after correcting for a zero point offset of 0.029 mas (Lindegren et al. 2018; Arenou et al. 2018).

We performed an additional statistical diagnostic to assess the significance of the signature of the clusters in astrometric space. Working with all stars in a field of view of radius $2 \times r_{50}$,

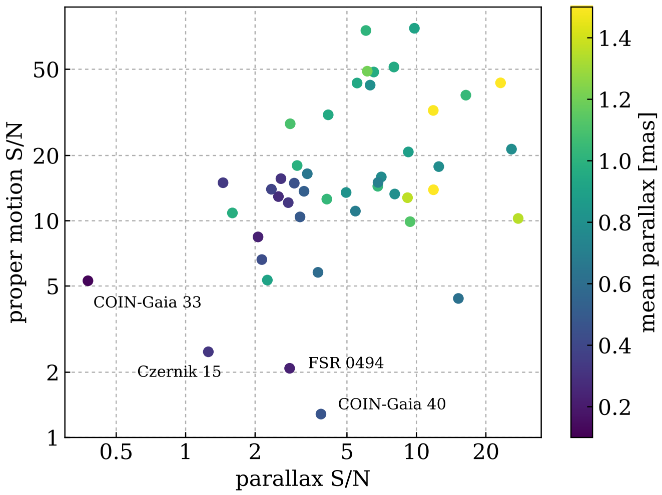

Fig. 1. Signal in parallax and proper motion space (see Sect. 4.2) for the 41 newly discovered clusters and five clusters re-identified in this study.

we binned stars in square cells of 0.5 mas yr^{-1} in proper motion space and computed the density distribution of parallaxes (using a Gaussian kernel of sigma 0.1 mas) in each cell[5]. We show in the middle right panels of Figs. B.1 to D.5 that the parallax distribution in the proper motion bin containing the cluster is always more peaked (denser relative density) than the mean of the other bins. To quantify how the proper motion selection reveals a peaked parallax distribution we calculated the difference between the density in the cluster bin and the mean density of the field bins, divided by the standard deviation of the field bins. This number can be interpreted as a S/N of the signature of the cluster in parallax space.

Reciprocally, we quantified how the density in proper motion space is enhanced when selecting stars in a 0.2 mas range centred on the cluster parallax, compared to the mean density in bins not containing the cluster. The proper motions of stars (highlighted from their parallax only) are shown in the right panels of Figs. B.1 to C.13.

In Fig. 1 we show the S/N in parallax and proper motion space for the 41 clusters discovered in this study and the five clusters we re-identified. Unsurprisingly, the signal in parallax space is stronger for the most nearby clusters. The most distant new cluster (COIN-*Gaia* 33, ϖ~0.12 mas) barely stands out as an over-density in parallax space, but its signature is clearly visible in proper motion space.

Based on their aspect in positional space and the aspect of their colour-magnitude diagram, we divide the 41 new clusters into 28 grade A (most certain candidates) and 13 grade B clusters. Their main parameters (location, apparent size, proper motions, and parallax) are listed in Table A.1. The grade B clusters tend to be more distant, more reddened, and their colour-magnitude diagrams are not well-defined, likely because of differential extinction. We remark that the four known objects Czernik 5, Czernik 15, FSR 0494, and FSR 0519 present a similarly blurred colour-magnitude diagram and weak signal in proper motion and parallax space.

4.3. Consequences for the cluster census

The unprecedented quality of the astrometry provided by the *Gaia* mission allows for new discoveries of groups of stars sharing a common proper motion and parallax, which is especially powerful when investigating populations projected against a dense background. We verified that none of the objects characterised in this study are visible as significant over-densities in optical images of the Galactic plane.

Table 1. Number of clusters listed by various authors in the region investigated in the present study.

Distance	DAML $^{(a)}$	MWSC $^{(b)}$	GDR2 $^{(c)}$	COIN-*Gaia* $^{(d)}$
<1 kpc	30	33	23	14
<2 kpc	140	180	69	33
all	363	430	227	41

Notes. $^{(a)}$From Dias et al. (2002); $^{(b)}$Kharchenko et al. (2013); $^{(c)}$known clusters identified from *Gaia* DR2 astrometry in Cantat-Gaudin et al. (2018a) and this study; $^{(d)}$previously unknown clusters first reported in this study.

The majority of new clusters reported in this work are nearby objects, 33 of which are located within 2 kpc of the Sun. Their discovery represents a ~50% increase in the number of clusters identified from *Gaia* DR2 astrometry in this direction and distance range[6]. These 33 new nearby clusters also represent a ~20–25% increase with respect to the catalogues of Dias et al. (2002) and Kharchenko et al. (2013). These two widely used reference catalogues include a significant fraction of objects flagged as putative or dubious, some of which were only very recently found to be asterisms (e.g. Han et al. 2016; Kos et al. 2018). Table 1 contains a summary of the total number of clusters listed by different authors in the region investigated in the present study.

Since our blind search is unable to recover all the known objects within 2 kpc, it also certainly failed to detect a significant number of unknown OCs that other methods might be able to uncover in the *Gaia* DR2 data. In particular, the flagging of candidates described in Sect. 3.2 is based on proper motions only, and is therefore likely biased towards clusters whose proper motions are significantly different from the field stars.

The present study shows that the assumption of completeness often made in OC studies (e.g. Buckner & Froebrich 2014; Lin et al. 2015; Joshi et al. 2016; Matsunaga et al. 2018; Piskunov et al. 2018) needs to be seriously re-evaluated in the *Gaia* era. Cantat-Gaudin et al. (2018a) have also shown that the samples on which the completeness analyses are based are also highly contaminated by false positives. We note that none of the new objects are located in the gap of the Perseus arm (see Figs. 2 and 3), and the region $l \in [140; 160]$ still appears to be almost devoid of OCs in the distance range ~1–2 kpc (as already noted by e.g. Vázquez et al. 2008).

Although the *Gaia*-DR2 catalogue represents an unprecedented improvement in the amount and quality of astrometric data available to astronomers, we point out that the next *Gaia* data releases planned for the upcoming years will incrementally refine the parallax and proper motion measurements, allowing us to better discern stellar clusters and possibly to discard some groupings identified in the *Gaia*-DR2 data as false positives.

[5] We only consider cells containing at least ten stars to mitigate the noise introduced by sparsely populated cells.

[6] Cantat-Gaudin et al. (2018a) have identified 68 clusters in this region within 2 kpc of the Sun. The present study adds NGC 886 to the list of known clusters in this region and distance range. None of the recent discoveries of Castro-Ginard et al. (2018) and Ferreira et al. (2019) fall within this region.

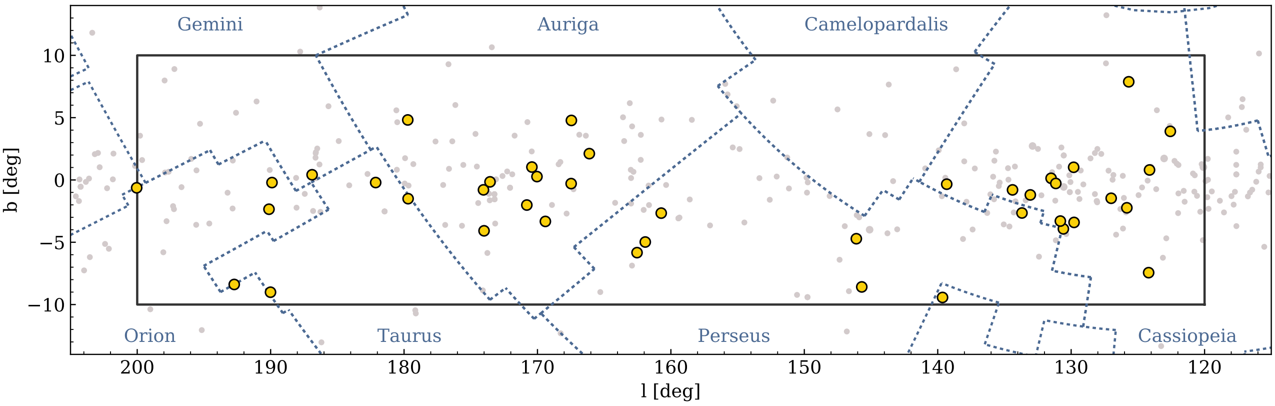

Fig. 2. Projected Galactic map of previously known OCs (grey points) and the new COIN-*Gaia* OCs (yellow dots) detected in this work.

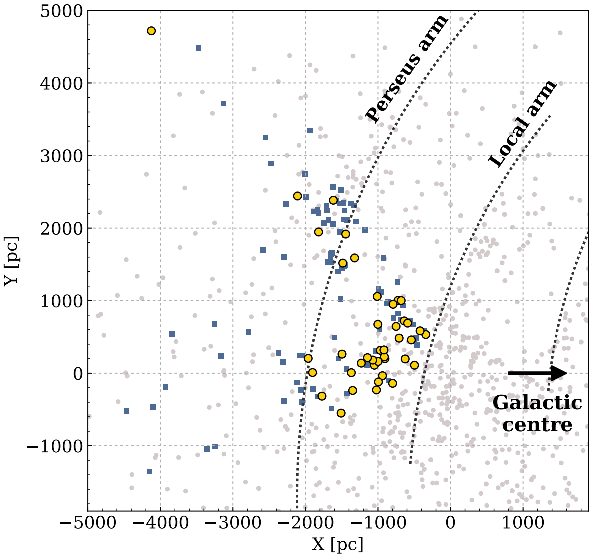

Fig. 3. Locations of OCs on the projected Galactic plane, for both previously known OCs (grey dots) and new COIN-*Gaia* OCs (yellow dots). In addition, blue symbols indicate known OCs that our blind search re-identified. The dotted lines indicate the spiral arm model of Reid et al. (2014). The Sun is located at coordinates (0, 0).

5. Conclusions

This study reports the discovery of 41 new OCs in the direction of Perseus at the Galactic coordinates $l \in [120°; 200°]$ and $b \in [-10°; 10°]$. This work employs a fully automated and scalable coarse-to-fine search algorithm, tailored to astrometric catalogues. The search is composed of three main steps: a fast density-aware sky tiling, a fast recommendation system for flagging OC candidate regions based on proper motion similarity, and cluster membership assignment and independent validation.

The majority of the 41 new clusters reported in this work are nearby objects, 33 of which are located within 2 kpc of the Sun. Their discovery represents a significant increase of at least 20% in the OC population relative to previously known objects in the same region. The sample is divided in 28 high certainty OCs and 13 plausible OCs, for which we provide the location, apparent size, and mean astrometric parameters. This works challenges the previous claim that the cluster census was complete up to 1.8 kpc and suggests that many discoveries are still to be made in our nearby Galactic environment.

Acknowledgements. We thank the anonymous referee for the comments and suggestions that helped improve this manuscript. This work was created during thet 5th COIN Residence Program (CRP#5) held in Chania, Greece in September 2018, with support from CNRS and IAASARS. We thank Vassilis Charmandaris for encouraging the accomplishment of this event. This project is financially supported by CNRS as part of its MOMENTUM programme over the 2018–2020 period. T.C.G. acknowledges support from Juan de la Cierva - formación 2015 grant, MINECO (FEDER/UE). This work was supported by the MINECO (Spanish Ministry of Economy) through grant ESP2016-80079-C2-1-R (MINECO/FEDER, UE) and MDM-2014-0369 of ICCUB (Unidad de Excelencia "María de Maeztu"). A.K.M. acknowledges the support from the Portuguese Fundação para a Ciência e a Tecnologia through grants SFRH/BPD/74697/2010, PTDC/FIS-AST/31546/2017, and from the ESA contract AO/1-7836/14/NL/HB. A.K.M. and A.M. acknowledge the support from the Portuguese Strategic Programme UID/FIS/00099/2013 for CENTRA. R.S.S. acknowledges the support from NASA under the Astrophysics Theory Program grant 14-ATP14-0007. R.S. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 771282 A.F. is supported by a McWilliams Postdoctoral Fellowship. AIM is a New York University Graduate School of Arts and Sciences Ted Keusseff Fellow advised by David W. Hogg. B.M. is a Principal's Career Development Scholar at the University of Edinburgh. We acknowledge feedback from Madhura Killedar, whose attendance at CRP#5 was supported by the Sydney Informatics Hub at The University of Sydney. We acknowledge partial support from the Fundação de Amparo à Pesquisa do Estado de São Paulo grant 2014/13407-4. This work has made use of the computing facilities of the Laboratory of Astroinformatics (IAG/USP, NAT/Unicsul), whose purchase was made possible by the Brazilian agency FAPESP (grant 2009/54006-4) and the INCT-A, and we thank the entire LAi team, especially Ulisses Manzo Castello and Alex Carciofi for the support. This work has made use of results from the ESA space mission *Gaia*, the data from which were processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. The *Gaia* mission website is `http://www.cosmos.esa.int/gaia`. Some of the authors are members of the *Gaia* DPAC. The Cosmostatistics Initiative[7] (COIN) is a non-profit organisation whose aim is to nourish the synergy between astrophysics, cosmology, statistics, and machine learning communities. COIN acknowledges the support from the `Overleaf`[8] collaborative platform.

References

Alessi, B. S., Moitinho, A., & Dias, W. S. 2003, A&A, 410, 565
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17
Bentley, J. L. 1975, Commun. ACM, 18, 509
Buckner, A. S. M., & Froebrich, D. 2014, MNRAS, 444, 290
Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018a, A&A, 618, A93
Cantat-Gaudin, T., Vallenari, A., Sordo, R., et al. 2018b, A&A, 615, A49
Castro-Ginard, A., Jordi, C., Luri, X., et al. 2018, A&A, 618, A59
de Souza, R. S., Dantas, M. L. L., Costa-Duarte, M. V., et al. 2017, MNRAS, 472, 2808
Dias, W. S., & Lépine, J. R. D. 2005, ApJ, 629, 825
Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, A&A, 389, 871
Dowler, P., Rixon, G., Tody, D., & Demleitner, M. 2018, TAP v. 1.1 IVOA Proposed Recommendation 30 August 2018
Dreyer, J. L. E. 1888, MNRAS, 49, 1

[7] `https://cosmostatistics-initiative.org`
[8] `https://www.overleaf.com`

Ducourant, C., Teixeira, R., Krone-Martins, A., et al. 2017, A&A, 597, A90
ESA 1997, The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission (Netherlands: ESA Special Publication), 1200
Ferreira, F. A., Santos, J. F. C., Corradi, W. J. B., Maia, F. F. S., & Angelo, M. S. 2019, MNRAS, 483, 5508
Forgy, E. W. 1965, Biometrics, 21, 768
Friedman, J. H., Bentley, J. L., & Finkel, R. A. 1977, ACM Trans. Math. Softw., 3, 209
Gaia Collaboration (Brown, A. G. A., et al.) 2016, A&A, 595, A2
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Garnier, R., & Taylor, J. 2009, Discrete Mathematics: Proofs, Structures and Applications, 3rd edn (Boca Raton: CRC Press)
Graham, R. L., & Hell, P. 1985, Ann. Hist. Comput., 7, 43
Han, E., Curtis, J. L., & Wright, J. T. 2016, AJ, 152, 7
Herschel, J. 1833, A Treatise on Astronomy, Cabinet cyclopaedia (London: Longman)
Janes, K., & Adler, D. 1982, ApJS, 49, 425
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python
Joshi, Y. C., Dambis, A. K., Pandey, A. K., & Joshi, S. 2016, A&A, 593, A116
Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R.-D. 2005, A&A, 440, 403
Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R. D. 2013, A&A, 558, A53
Kos, J., de Silva, G., Buder, S., et al. 2018, MNRAS, 480, 5242
Krone-Martins, A., & Moitinho, A. 2014, A&A, 561, A57
Krone-Martins, A., Soubiran, C., Ducourant, C., Teixeira, R., & Le Campion, J. F. 2010, A&A, 516, A3
Lada, C. J. 2010, Phil. Trans. R. Soc. London, Ser. A, 368, 713
Lin, C. C., Chen, W. P., Hou, J., Chen, L., & Shao, Z. 2015, Proc. IAU, 12, 163
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
Lloyd, S. 1982, IEEE Trans. Inf. Theory, 28, 129
Maneewongvatana, S., & Mount, D. M. 1999, in Center for Geometric Computing, 4th Annual Workshop on Computational Geometry, 2, 1
Matsunaga, N., Bono, G., Chen, X., et al. 2018, Space Sci. Rev., 214, 74
Melchior, P., & Goulding, A. D. 2018, Astron. Comput., 25, 183
Moitinho, A. 2010, in Star Clusters: Basic Galactic Building Blocks Throughout Time and Space, eds. R. de Grijs & J. R. D. Lépine, IAU Symp., 266, 106
Ortiz, I., Lusted, J., Dowler, P., et al. 2008, ADQL v.2.0 IVOA Recommendation 30 Oct 2008
Pearson, K. 1894, Phil. Trans. R. Soc. London, Ser. A: Math. Phys. Eng. Sci., 185, 71
Piskunov, A. E., Kharchenko, N. V., Röser, S., Schilbach, E., & Scholz, R.-D. 2006, A&A, 445, 545
Piskunov, A. E., Just, A., Kharchenko, N. V., et al. 2018, A&A, 614, A22
Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130
Roeser, S., Demleitner, M., & Schilbach, E. 2010, AJ, 139, 2440
Röser, S., Schilbach, E., & Goldman, B. 2016, A&A, 595, A22
Salgado, J., González-Núñez, J., Gutiérrez-Sánchez, R., et al. 2017, Astron. Comput., 21, 22
Sanders, W. L. 1971, A&A, 14, 226
Schmeja, S., Kharchenko, N. V., Piskunov, A. E., et al. 2014, A&A, 568, A51
Scholz, R.-D., Kharchenko, N. V., Piskunov, A. E., Röser, S., & Schilbach, E. 2015, A&A, 581, A39
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Sproull, R. F. 1991, Algorithmica, 6, 579
Sulentic, J., Tifft, W., & Dreyer, J. 1973, The Revised New General Catalogue of Nonstellar Astronomical Objects (Tucson: University of Arizona Press)
Vasilevskis, S., Klemola, A., & Preston, G. 1958, AJ, 63, 387
Vázquez, R. A., May, J., Carraro, G., et al. 2008, ApJ, 672, 930
Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, AJ, 145, 44

Appendix A: COIN-*Gaia* open clusters

Table A.1. Mean parameters of the OCs characterised in this study.

OC	α (deg)	δ (deg)	l (deg)	b (deg)	r_{50} (deg)	N	$\mu_{\alpha*}$	$\sigma_{\mu_{\alpha*}}$ (mas yr^{-1})	μ_{δ}	$\sigma_{\mu_{\delta}}$	ϖ (mas)	σ_{ϖ} (mas)	d (pc)
Grade A clusters													
COIN-*Gaia* 1	11.933	66.769	122.566	3.9	0.357	88	−5.03	0.12	−3.04	0.11	1.55	0.04	635
COIN-*Gaia* 2	15.06	55.409	124.192	−7.441	0.187	151	−4.46	0.12	−1.96	0.12	0.79	0.06	1226
COIN-*Gaia* 3	18.739	60.505	125.825	−2.235	0.126	98	−2.43	0.1	−1.81	0.12	0.78	0.05	1243
COIN-*Gaia* 4	26.129	58.756	129.782	−3.408	0.107	66	−0.95	0.07	−1.04	0.08	0.45	0.03	2086
COIN-*Gaia* 5	27.408	58.078	130.583	−3.927	0.224	39	−2.77	0.17	−0.53	0.1	1.06	0.08	916
COIN-*Gaia* 6	28.101	58.636	130.809	−3.3	0.077	132	−2.35	0.11	−0.44	0.14	0.28	0.06	3267
COIN-*Gaia* 7	33.738	58.466	133.686	−2.651	0.152	53	−1.04	0.1	−1.56	0.09	0.65	0.05	1477
COIN-*Gaia* 8	39.048	50.013	139.64	−9.43	0.312	69	2.51	0.11	−2.49	0.11	1.36	0.04	720
COIN-*Gaia* 9	47.748	48.023	145.694	−8.585	0.265	41	−1.93	0.11	−2.79	0.12	1.12	0.06	871
COIN-*Gaia* 10	68.385	40.509	161.919	−4.981	0.177	41	2.02	0.09	−3.37	0.11	0.94	0.04	1029
COIN-*Gaia* 11	68.11	39.479	162.538	−5.834	0.336	92	3.57	0.15	−5.63	0.13	1.49	0.05	661
COIN-*Gaia* 12	79.209	41.708	166.113	2.112	0.306	90	2.62	0.13	−4.66	0.09	1.03	0.05	944
COIN-*Gaia* 13	83.186	42.087	167.459	4.776	1.003	165	−3.83	0.18	−1.66	0.17	1.93	0.09	511
COIN-*Gaia* 14	77.696	39.195	167.476	−0.293	0.157	24	1.33	0.09	−6.57	0.08	1.04	0.03	933
COIN-*Gaia* 15	76.09	35.831	169.408	−3.328	0.147	134	0.44	0.15	−3.75	0.1	0.82	0.07	1172
COIN-*Gaia* 16	80.179	37.438	170.038	0.27	0.059	49	1.26	0.14	−3.74	0.14	0.62	0.06	1535
COIN-*Gaia* 17	81.244	37.558	170.418	1.035	0.151	83	0.36	0.16	−4.28	0.12	0.89	0.07	1094
COIN-*Gaia* 18	78.408	35.498	170.797	−2.014	0.239	59	0.77	0.12	−4.92	0.14	0.95	0.07	1018
COIN-*Gaia* 19	82.188	34.29	173.556	−0.153	0.172	88	−1.48	0.13	−4.64	0.09	0.77	0.06	1245
COIN-*Gaia* 20	78.634	31.691	174.004	−4.082	0.258	55	0.55	0.14	−1.45	0.09	0.91	0.05	1063
COIN-*Gaia* 21	84.766	28.402	179.696	−1.504	0.098	42	−0.12	0.09	−3.82	0.08	0.7	0.04	1378
COIN-*Gaia* 22	91.06	31.602	179.721	4.815	0.086	104	−0.71	0.15	−3.28	0.11	0.49	0.07	1927
COIN-*Gaia* 23	87.449	27.008	182.127	−0.209	0.275	108	−0.32	0.16	−0.96	0.13	1.03	0.08	942
COIN-*Gaia* 24	90.693	23.203	186.893	0.416	0.194	70	2.54	0.1	−2.97	0.09	0.96	0.05	1006
COIN-*Gaia* 25	91.691	20.276	189.899	−0.211	0.291	112	−0.5	0.15	−2.65	0.12	1.2	0.07	813
COIN-*Gaia* 26	83.771	15.721	190.008	−9.01	0.107	81	0.27	0.11	−2.39	0.11	0.69	0.06	1395
COIN-*Gaia* 27	85.76	13.743	192.728	−8.387	0.195	78	0.7	0.13	−3.6	0.12	0.91	0.05	1063
COIN-*Gaia* 28	96.333	11.159	200.041	−0.63	0.115	124	−1.12	0.15	−0.94	0.15	0.59	0.07	1616
Grade B clusters													
COIN-*Gaia* 29	15.548	63.648	124.125	0.801	0.177	29	−2.65	0.05	−0.32	0.04	0.31	0.03	2923
COIN-*Gaia* 30	21.08	70.574	125.684	7.878	0.254	92	−6.14	0.11	2.11	0.16	1.35	0.05	727
COIN-*Gaia* 31	21.307	61.135	127.0	−1.469	0.093	37	−1.98	0.07	−0.53	0.05	0.38	0.04	2436
COIN-*Gaia* 32	28.194	63.066	129.813	1.016	0.139	24	0.05	0.07	−2.2	0.05	0.78	0.03	1245
COIN-*Gaia* 33	30.276	61.475	131.15	−0.28	0.05	25	−1.31	0.03	−0.08	0.07	0.12	0.04	6576
COIN-*Gaia* 34	31.231	61.776	131.504	0.133	0.183	47	−2.23	0.16	0.88	0.09	1.01	0.06	966
COIN-*Gaia* 35	33.472	60.039	133.06	−1.201	0.043	31	−0.53	0.07	−0.39	0.08	0.34	0.04	2702
COIN-*Gaia* 36	36.256	59.975	134.394	−0.798	0.105	7	−0.97	0.06	−0.56	0.04	0.43	0.01	2173
COIN-*Gaia* 37	45.377	58.329	139.326	−0.341	0.845	29	0.85	0.09	−2.04	0.05	0.98	0.03	995
COIN-*Gaia* 38	51.472	51.072	146.101	−4.718	0.132	73	2.03	0.14	−6.82	0.11	0.79	0.05	1217
COIN-*Gaia* 39	69.612	42.95	160.724	−2.663	0.244	32	0.24	0.11	−2.46	0.08	0.99	0.04	980
COIN-*Gaia* 40	81.874	33.526	174.048	−0.794	0.073	28	0.39	0.06	−2.73	0.05	0.47	0.04	2003
COIN-*Gaia* 41	89.832	19.028	190.123	−2.349	0.15	85	−0.32	0.1	−3.66	0.1	0.52	0.07	1820
Known clusters													
FSR 0494	6.416	63.754	120.087	1.025	0.056	72	−2.5	0.11	−0.85	0.07	0.23	0.05	3943
FSR 0519	13.092	64.596	123.032	1.724	0.085	16	−2.42	0.03	−0.33	0.05	0.32	0.01	2833
Czernik 5	28.927	61.355	130.557	−0.562	0.036	74	−2.02	0.08	0.45	0.11	0.23	0.07	3882
NGC 886	35.898	63.8	132.893	2.725	0.34	174	3.19	0.1	−4.0	0.11	0.95	0.04	1017
Czernik 15	50.781	52.223	145.105	−3.995	0.111	36	0.33	0.09	−1.07	0.09	0.32	0.04	2853

Notes. r_{50}: radius containing half the members identified in this study. N: number of stars with the most frequent membership probability ≥ 50%. d: mode of the distance likelihood after adding +0.029 mas to the measured parallaxes.

Appendix B: Grade A clusters

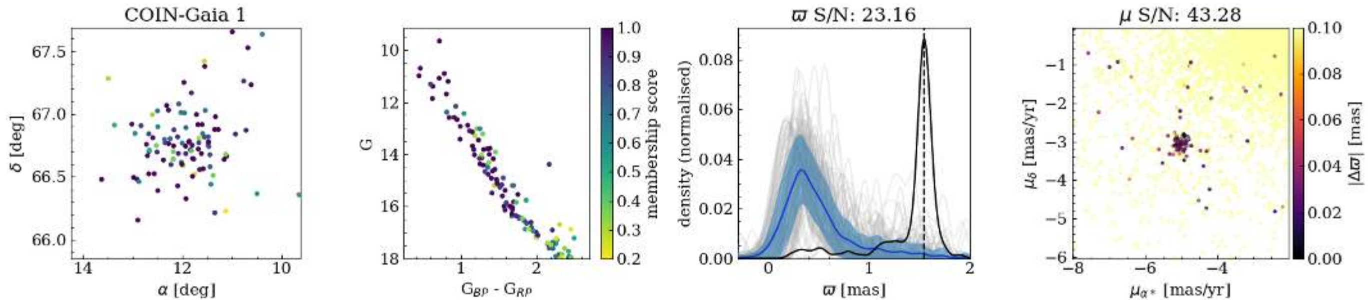

Fig. B.1. *Left*: spatial distribution of the stars of COIN-*Gaia* 1 with membership probabilities $p > 20\%$. *Middle left*: colour-magnitude diagram of the same stars. *Middle right*: parallax distribution density for stars binned by proper motion (see Sect. 4.2). *Right*: proper motion of all stars in the field of view, colour coded by their difference in parallax with the value listed in Table A.1.

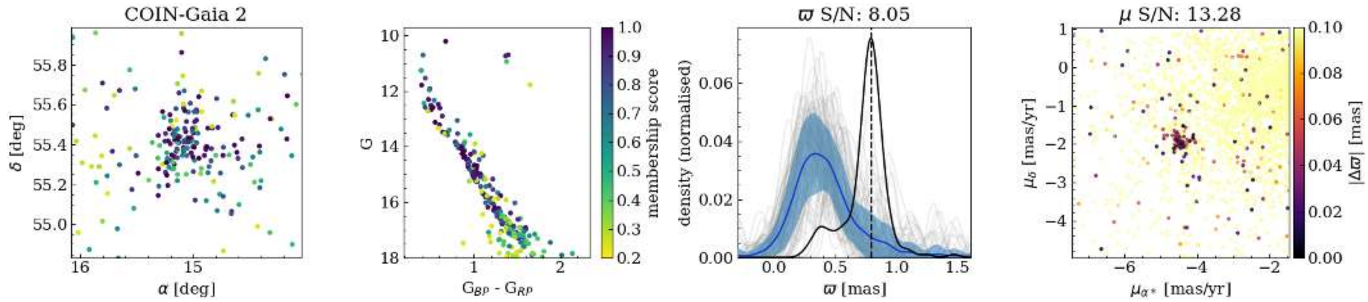

Fig. B.2. Same as Fig. B.1 for COIN-*Gaia* 2.

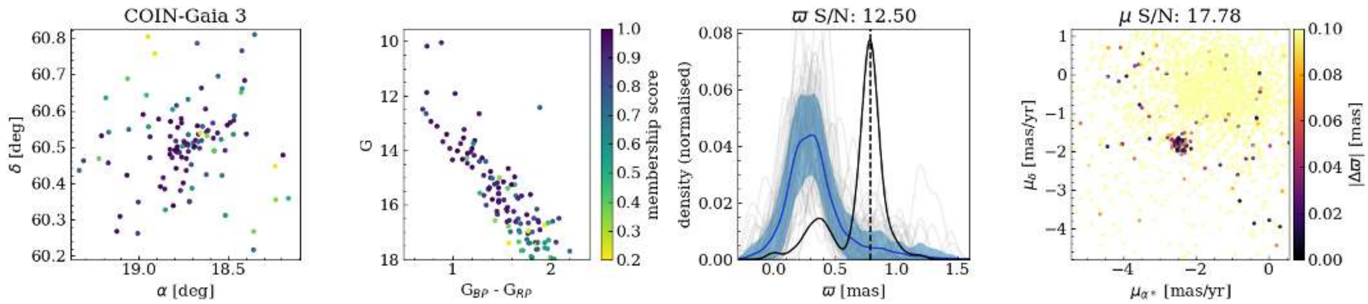

Fig. B.3. Same as Fig. B.1 for COIN-*Gaia* 3.

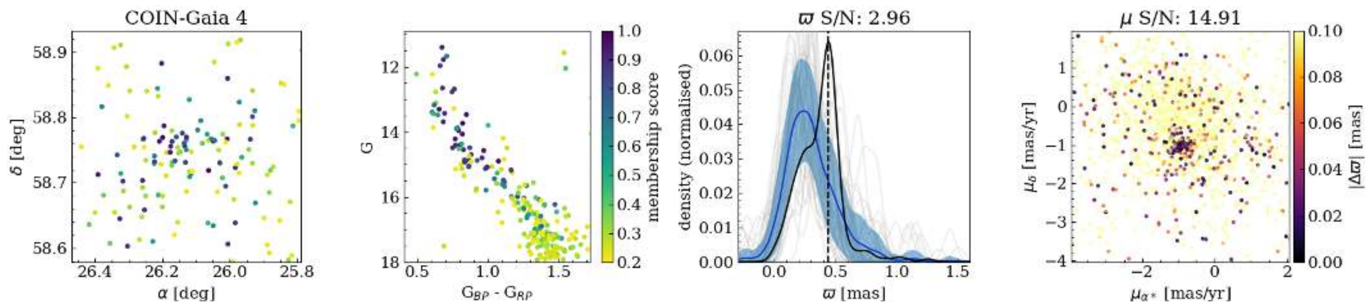

Fig. B.4. Same as Fig. B.1 for COIN-*Gaia* 4.

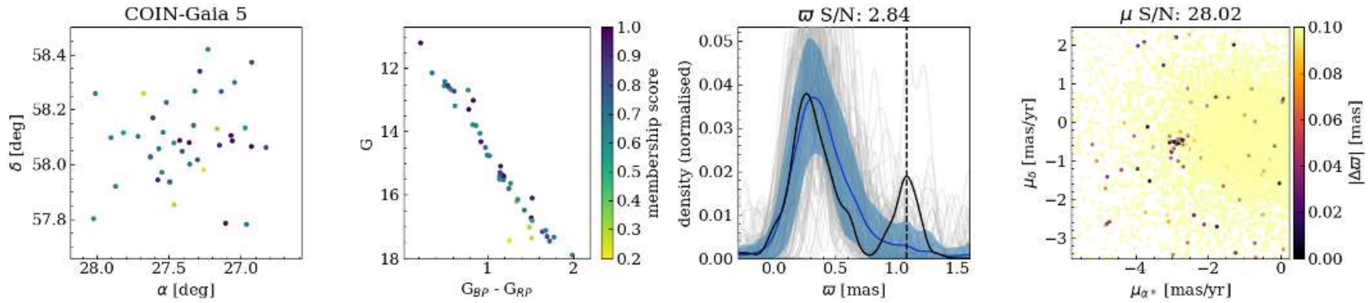

Fig. B.5. Same as Fig. B.1 for COIN-*Gaia* 5.

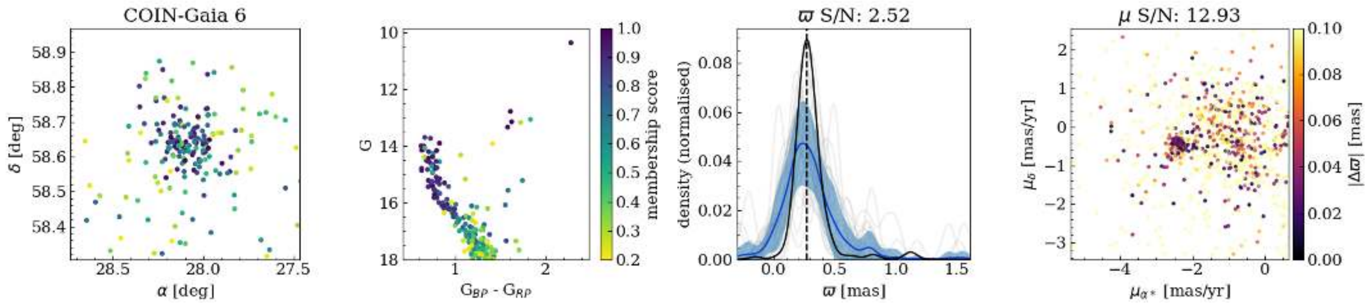

Fig. B.6. Same as Fig. B.1 for COIN-*Gaia* 6.

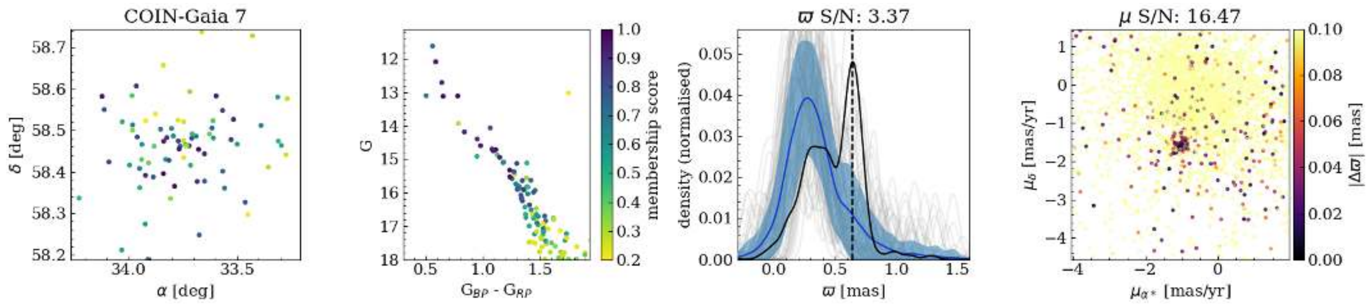

Fig. B.7. Same as Fig. B.1 for COIN-*Gaia* 7.

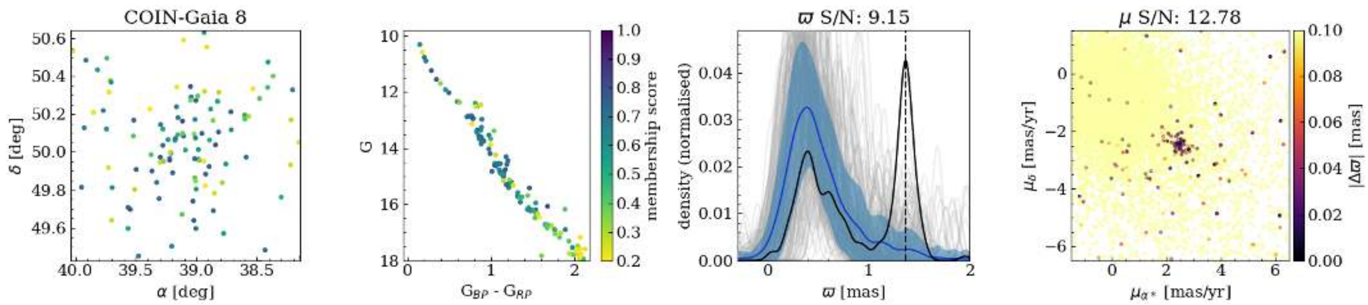

Fig. B.8. Same as Fig. B.1 for COIN-*Gaia* 8.

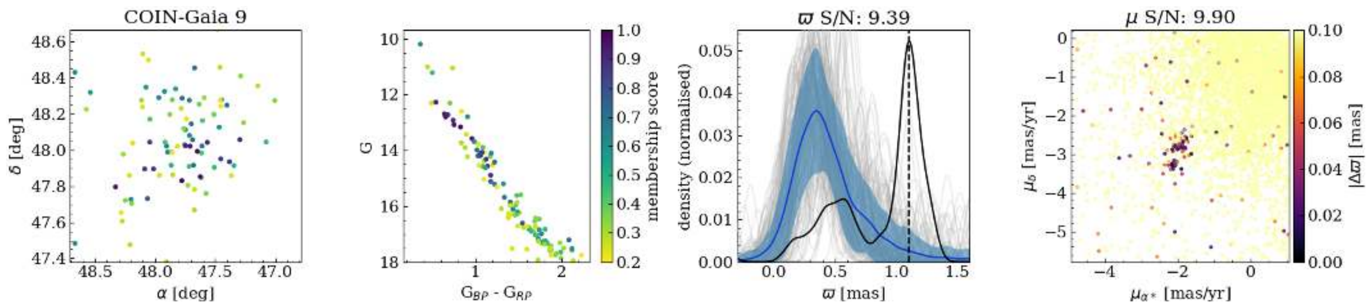

Fig. B.9. Same as Fig. B.1 for COIN-*Gaia* 9.

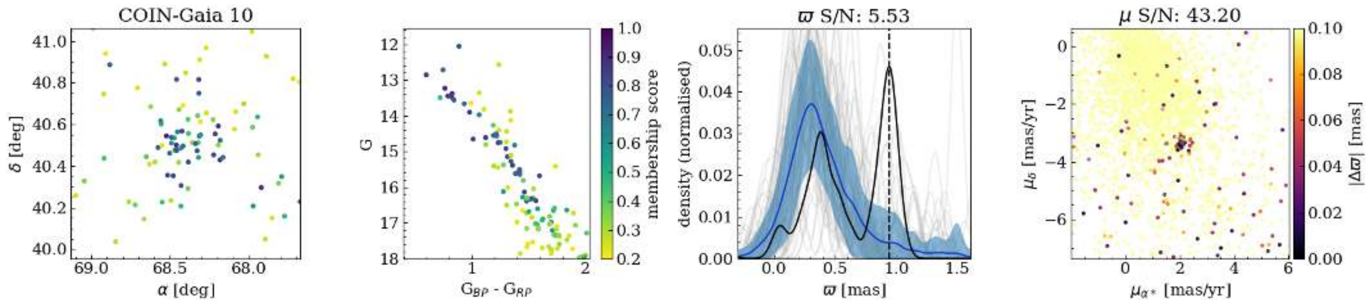

Fig. B.10. Same as Fig. B.1 for COIN-*Gaia* 10.

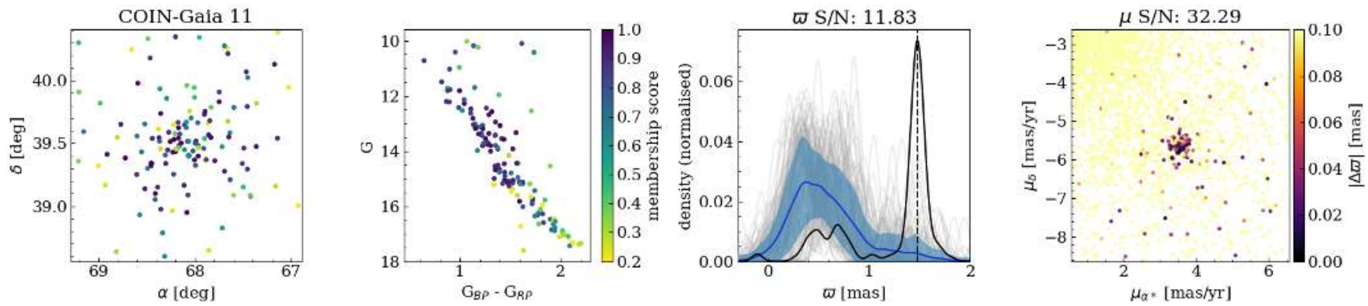

Fig. B.11. Same as Fig. B.1 for COIN-*Gaia* 11.

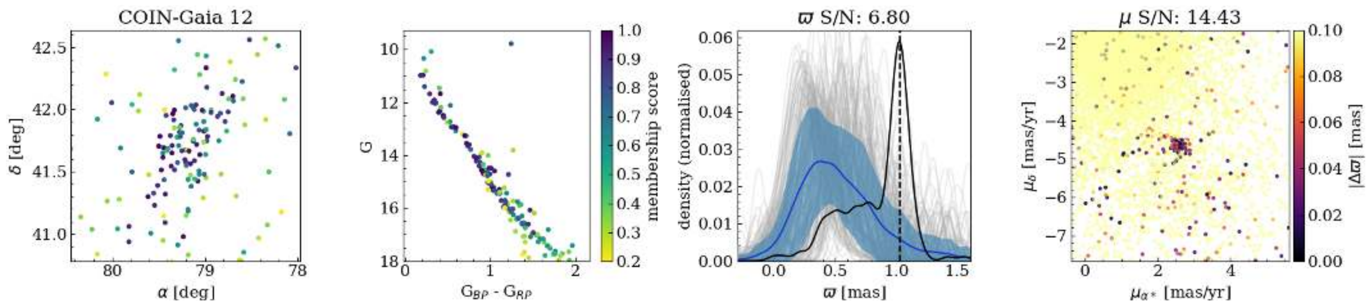

Fig. B.12. Same as Fig. B.1 for COIN-*Gaia* 12.

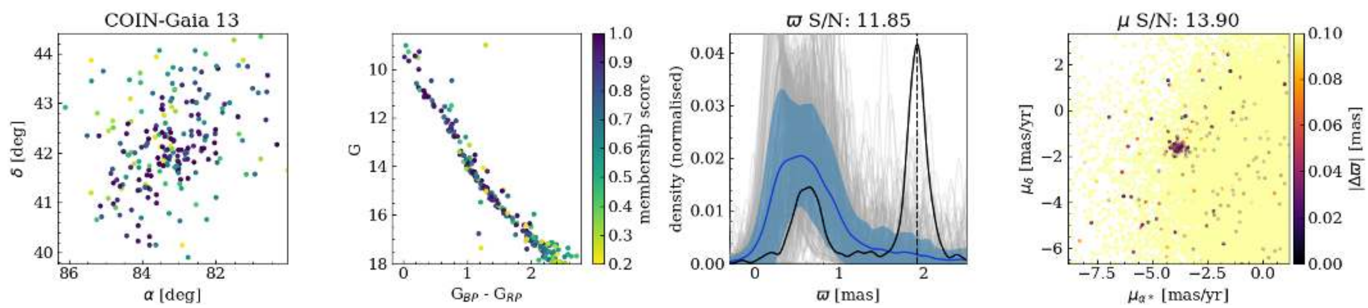

Fig. B.13. Same as Fig. B.1 for COIN-*Gaia* 13.

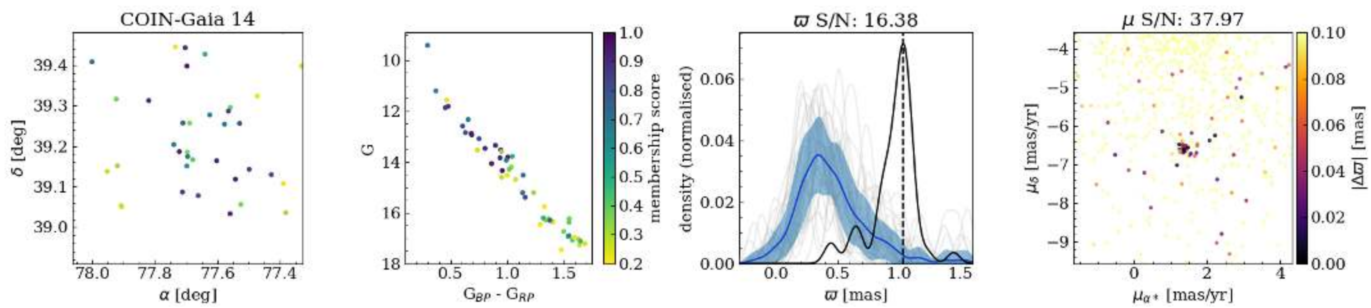

Fig. B.14. Same as Fig. B.1 for COIN-*Gaia* 14.

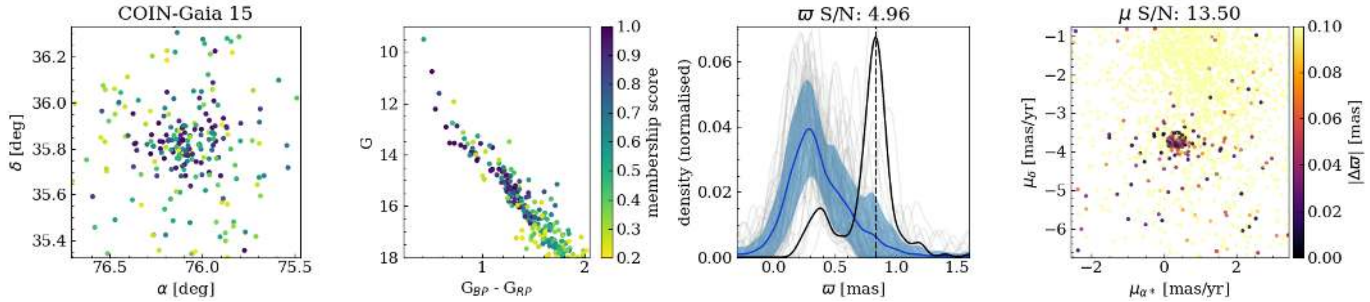

Fig. B.15. Same as Fig. B.1 for COIN-*Gaia* 15.

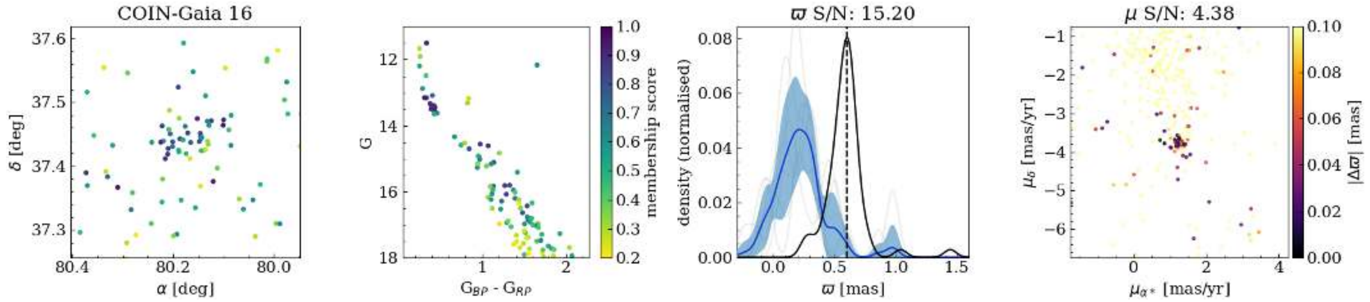

Fig. B.16. Same as Fig. B.1 for COIN-*Gaia* 16.

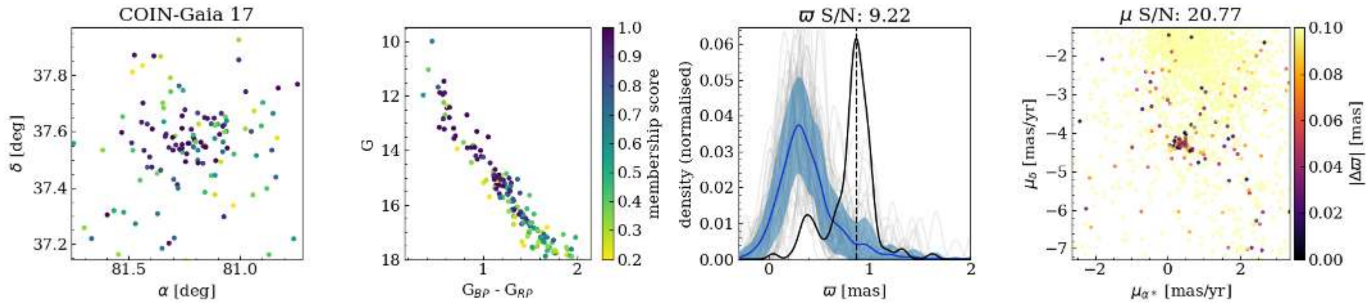

Fig. B.17. Same as Fig. B.1 for COIN-*Gaia* 17.

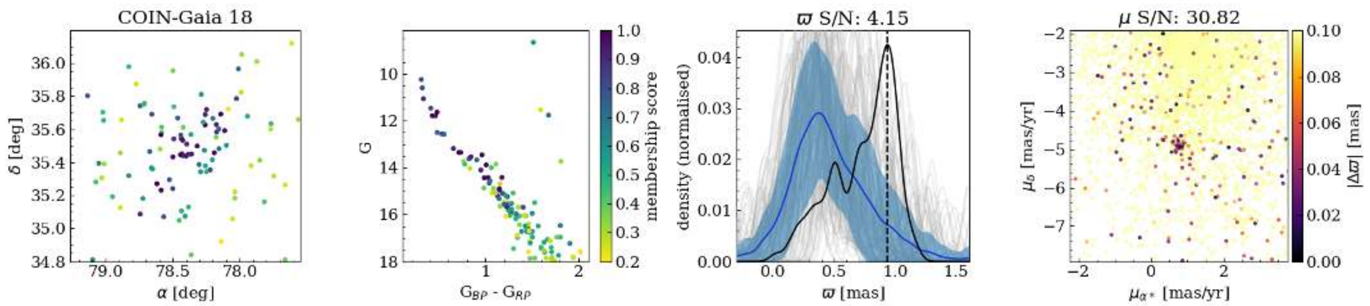

Fig. B.18. Same as Fig. B.1 for COIN-*Gaia* 18.

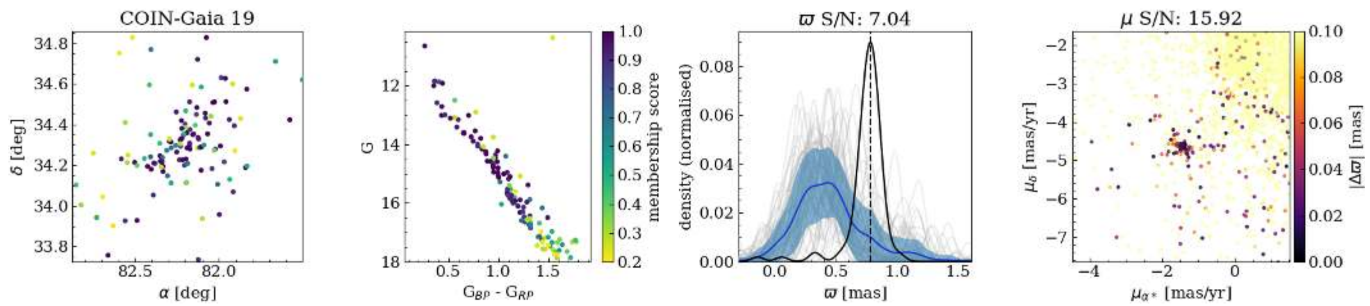

Fig. B.19. Same as Fig. B.1 for COIN-*Gaia* 19.

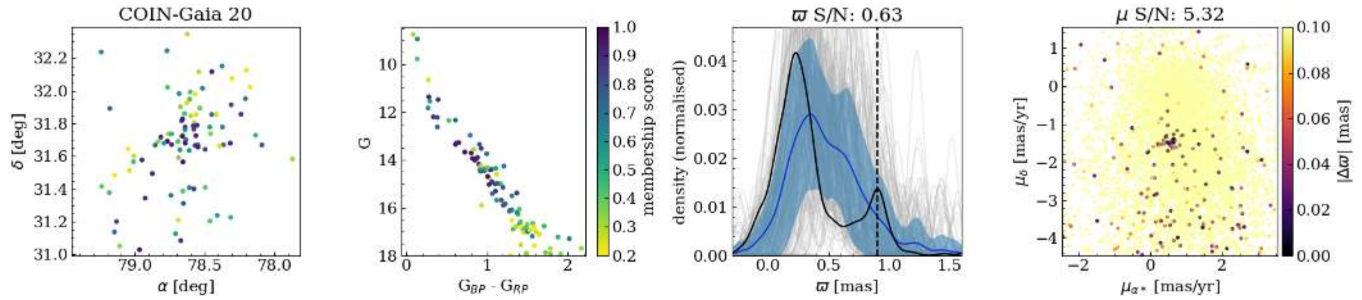

Fig. B.20. Same as Fig. B.1 for COIN-*Gaia* 20.

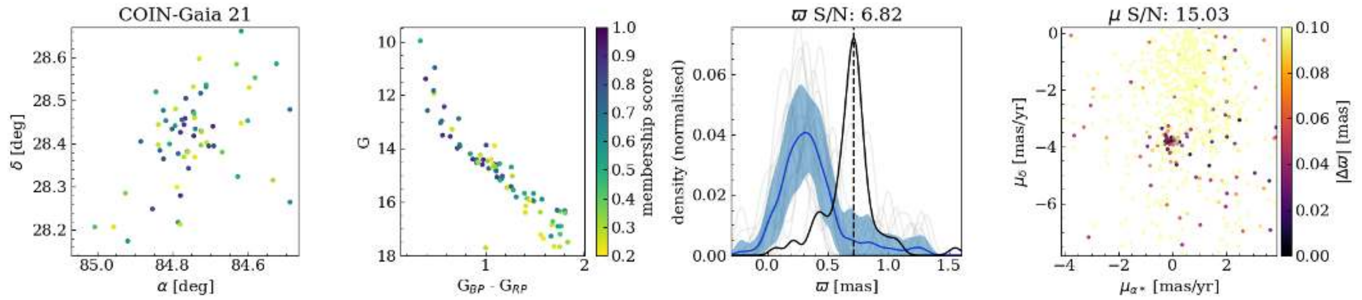

Fig. B.21. Same as Fig. B.1 for COIN-*Gaia* 21.

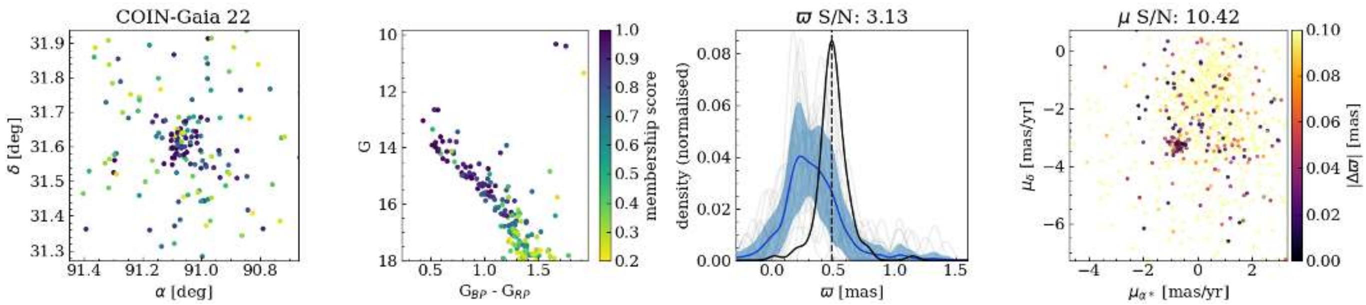

Fig. B.22. Same as Fig. B.1 for COIN-*Gaia* 22.

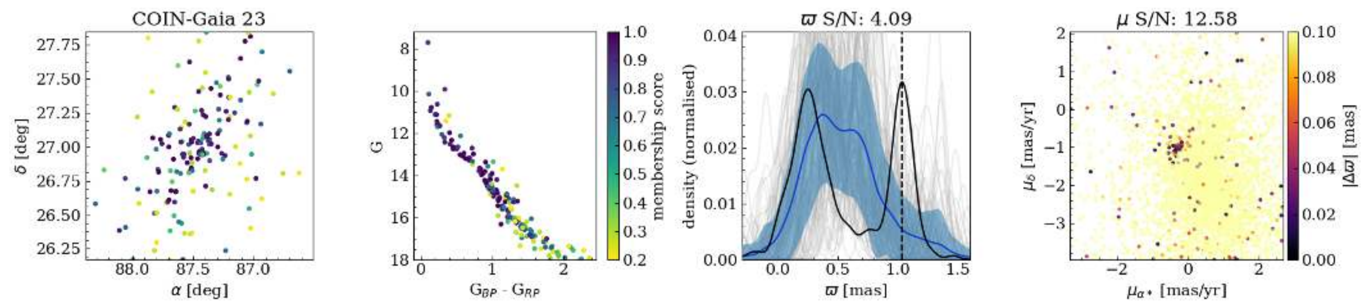

Fig. B.23. Same as Fig. B.1 for COIN-*Gaia* 23.

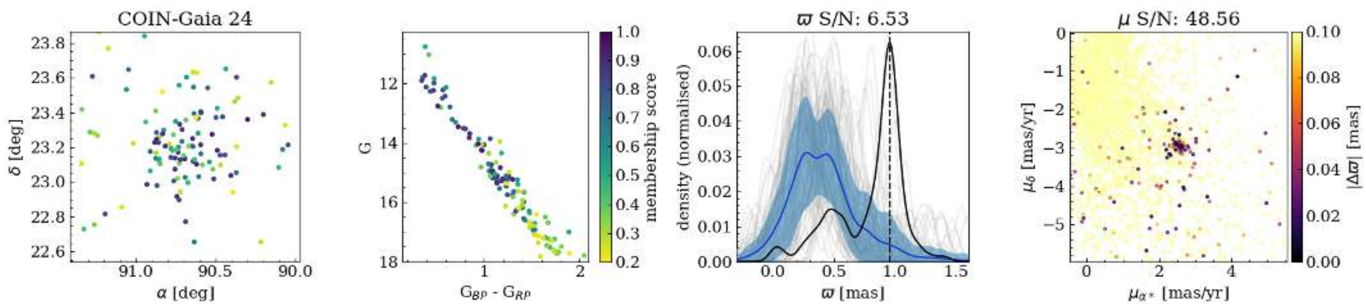

Fig. B.24. Same as Fig. B.1 for COIN-*Gaia* 24.

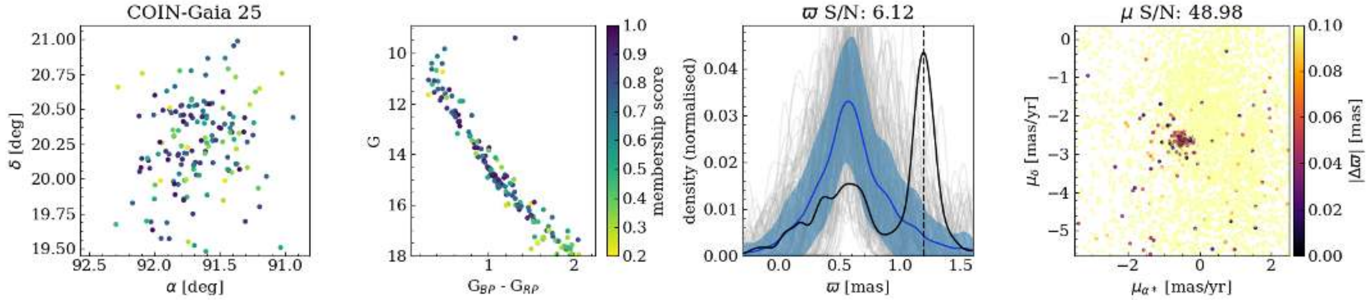

Fig. B.25. Same as Fig. B.1 for COIN-*Gaia* 25.

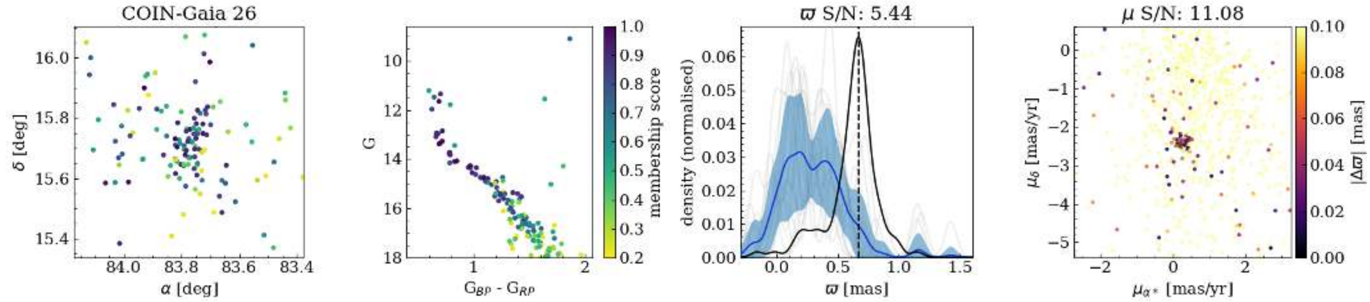

Fig. B.26. Same as Fig. B.1 for COIN-*Gaia* 26.

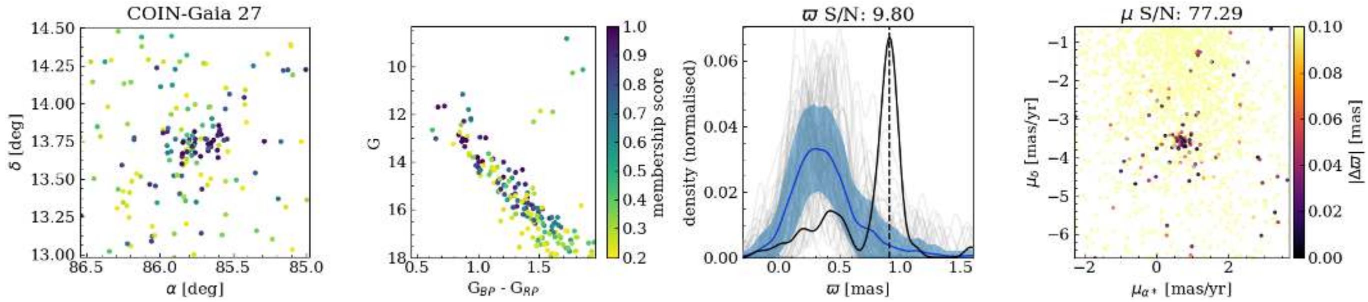

Fig. B.27. Same as Fig. B.1 for COIN-*Gaia* 27.

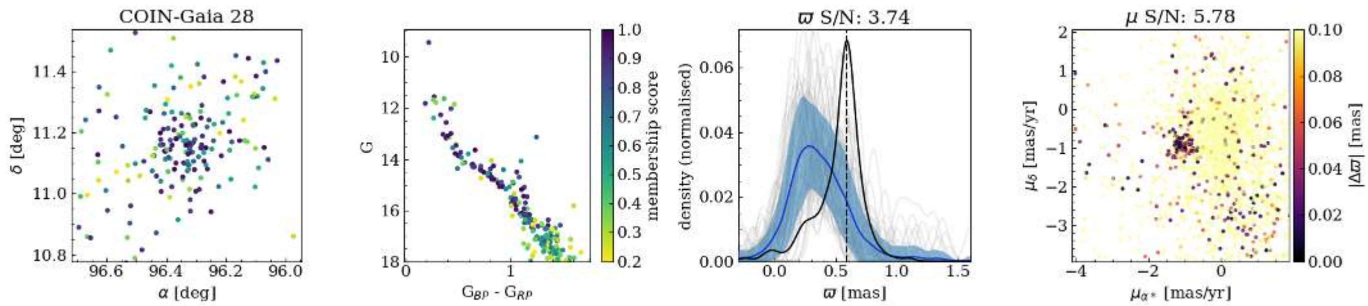

Fig. B.28. Same as Fig. B.1 for COIN-*Gaia* 28.

Appendix C: Grade B clusters

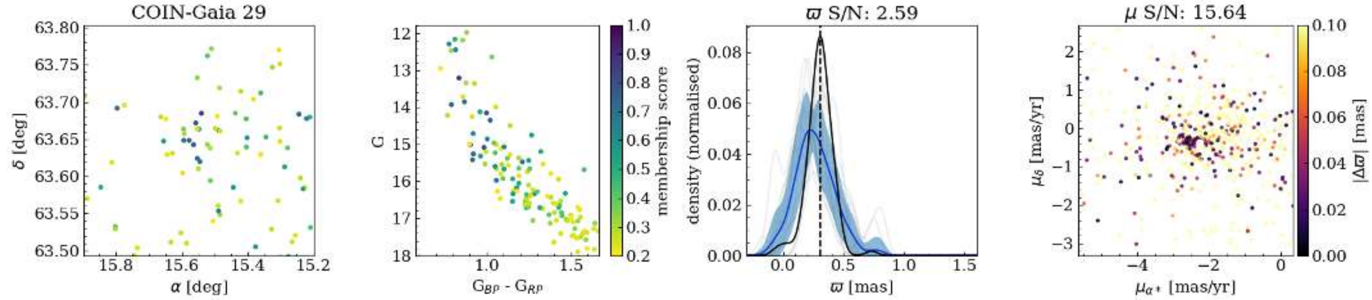

Fig. C.1. Same as Fig. B.1 for COIN-*Gaia* 29.

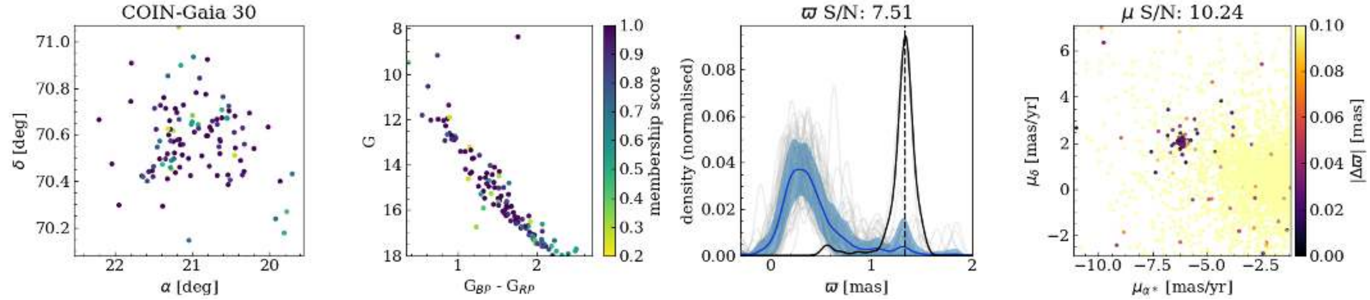

Fig. C.2. Same as Fig. B.1 for COIN-*Gaia* 30.

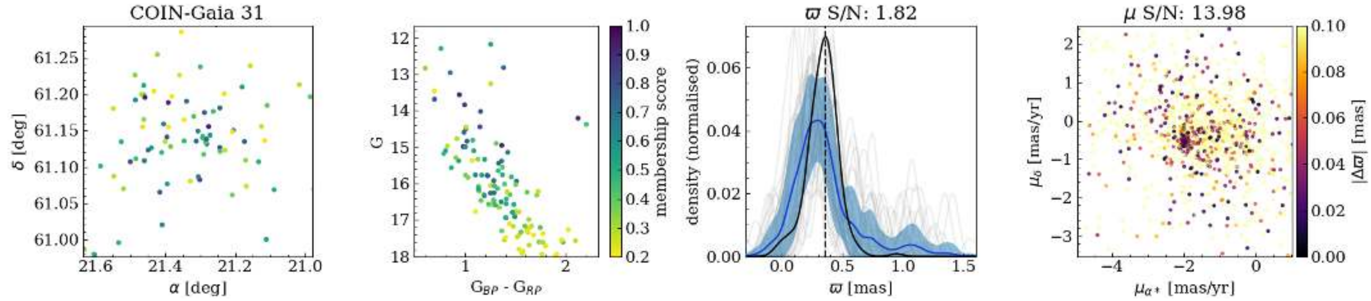

Fig. C.3. Same as Fig. B.1 for COIN-*Gaia* 31.

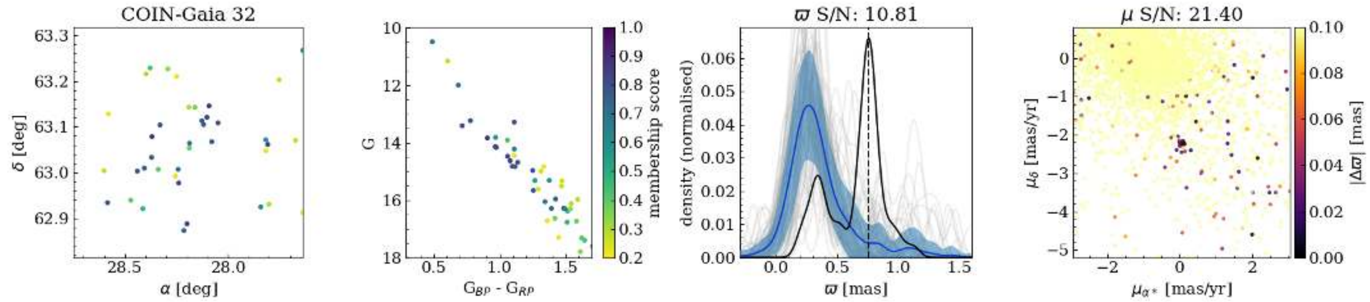

Fig. C.4. Same as Fig. B.1 for COIN-*Gaia* 32.

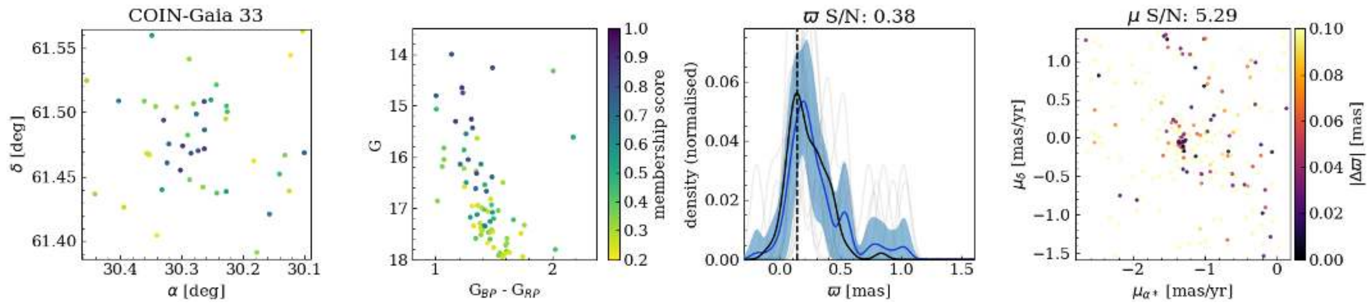

Fig. C.5. Same as Fig. B.1 for COIN-*Gaia* 33.

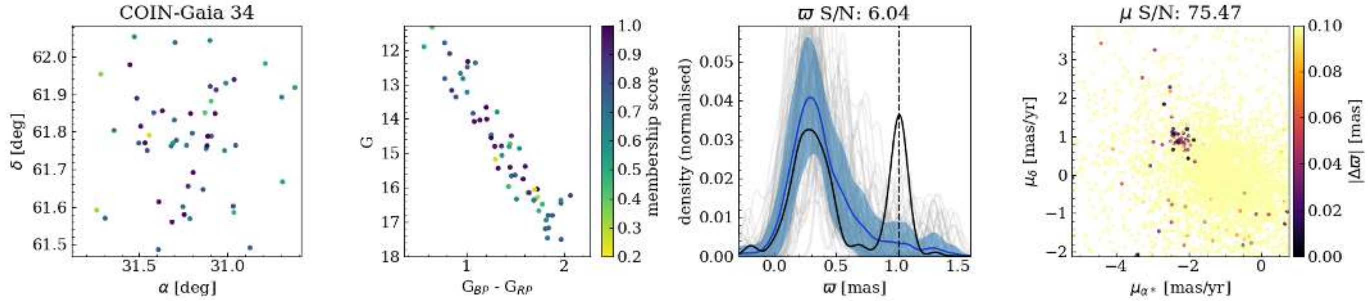

Fig. C.6. Same as Fig. B.1 for COIN-*Gaia* 34.

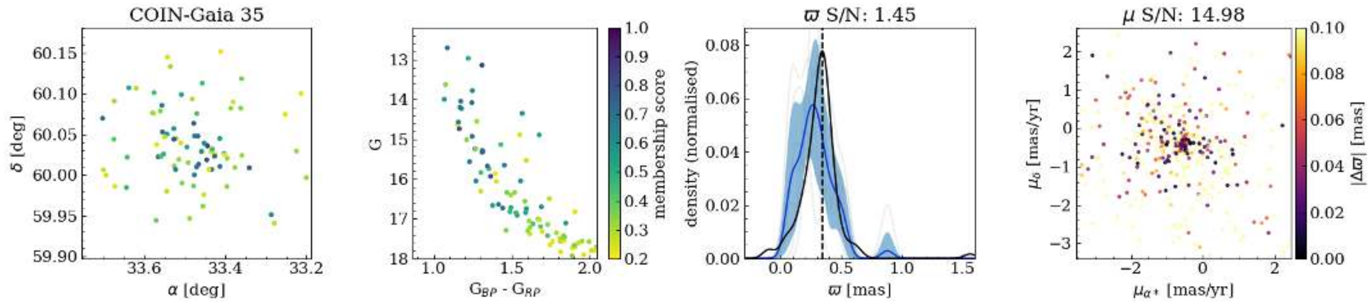

Fig. C.7. Same as Fig. B.1 for COIN-*Gaia* 35.

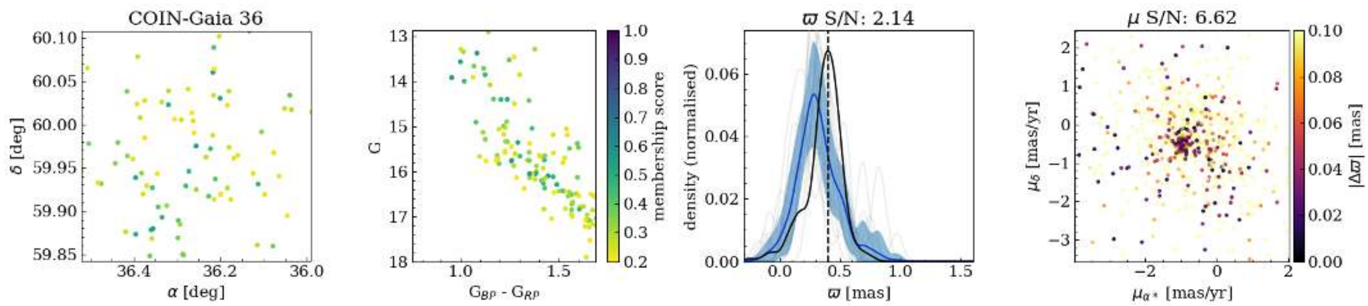

Fig. C.8. Same as Fig. B.1 for COIN-*Gaia* 36.

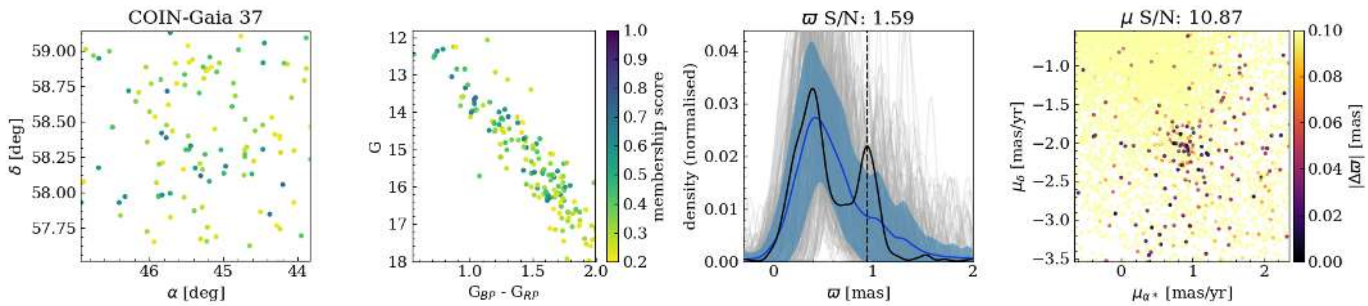

Fig. C.9. Same as Fig. B.1 for COIN-*Gaia* 37.

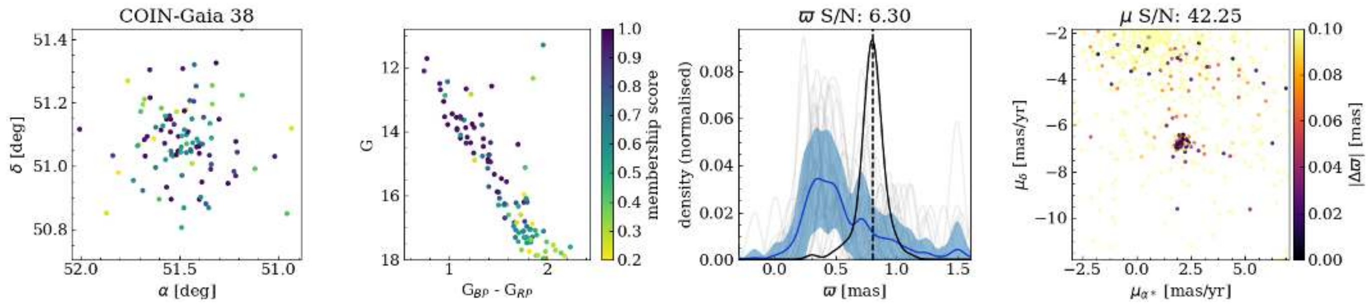

Fig. C.10. Same as Fig. B.1 for COIN-*Gaia* 38.

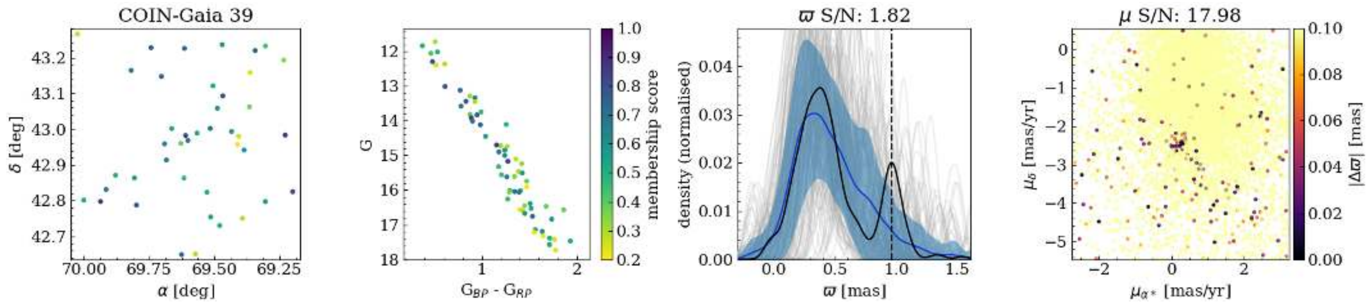

Fig. C.11. Same as Fig. B.1 for COIN-*Gaia* 39.

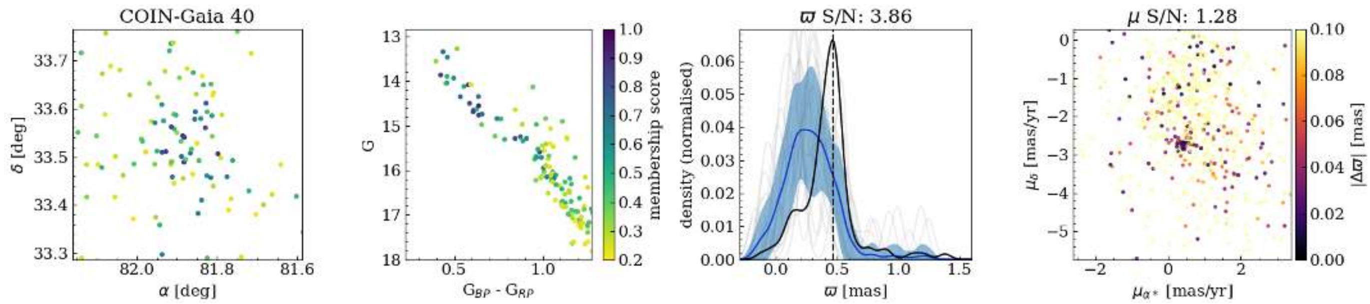

Fig. C.12. Same as Fig. B.1 for COIN-*Gaia* 40.

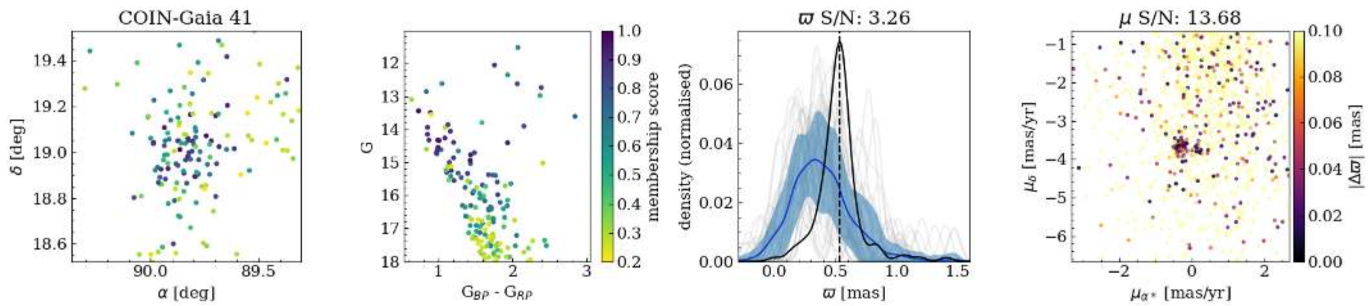

Fig. C.13. Same as Fig. B.1 for COIN-*Gaia* 41.

Appendix D: Known clusters

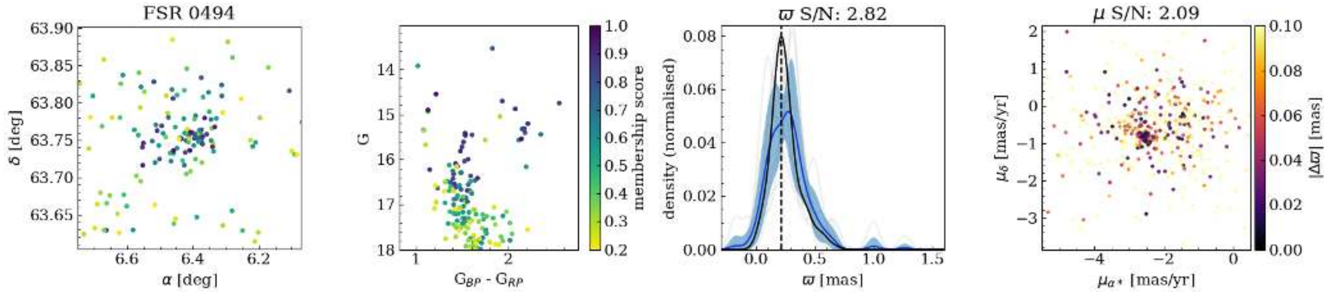

Fig. D.1. Same as Fig. B.1 for FSR 0494.

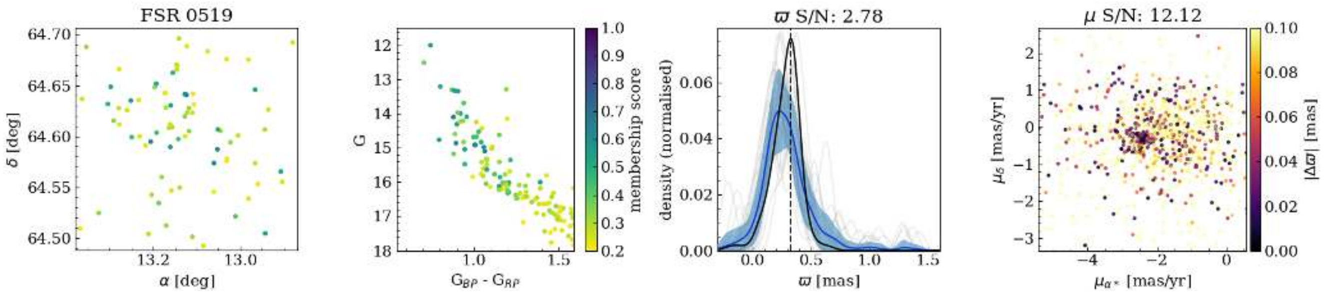

Fig. D.2. Same as Fig. B.1 for FSR 0519.

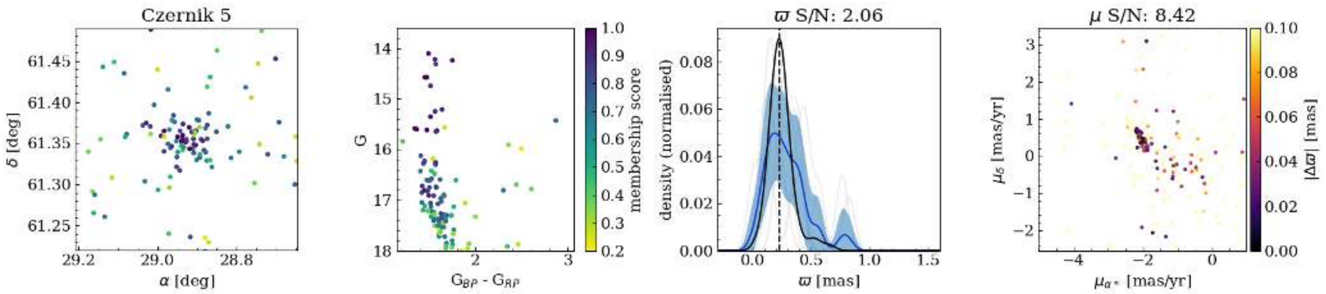

Fig. D.3. Same as Fig. B.1 for Czernik 5.

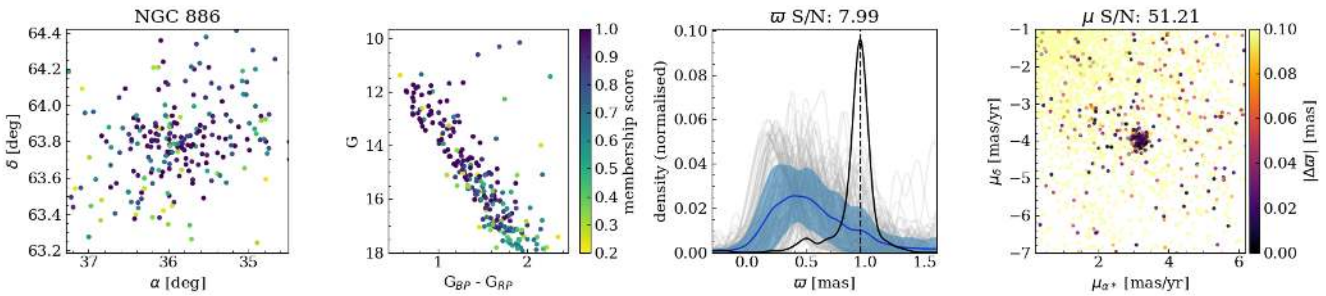

Fig. D.4. Same as Fig. B.1 for NGC 886.

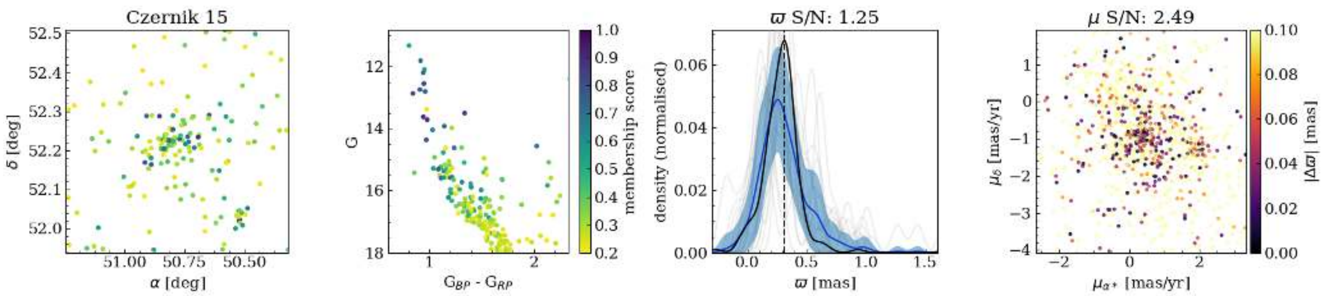

Fig. D.5. Same as Fig. B.1 for Czernik 15.