Revisiting the memberships, structural parameters and kinematics of stellar clusters in the Milky Way.

Yoann Tarricq

THESIS PRESENTED

FOR THE DEGREE OF

DOCTOR
FROM THE UNIVERSITY OF BORDEAUX

ECOLE DOCTORALE: SCIENCES PHYSIQUES ET DE L'INGÉNIEUR

SPECIALITY: ASTROPHYSICS

By **Yoann TARRICQ**

Revisiting the memberships, structural parameters and kinematics of stellar clusters in the Milky Way.

Under the supervision of: **Caroline SOUBIRAN**
Thesis advisor: **Laia CASAMIQUELA**

02/12/2021

Jury members :

Ms. Estelle MORAUX	Lecturer, IPAG, University Grenoble Alpes	Referee
Ms. Céline REYLÉ	Astronomer, UTINAM, Observatory of Besançon	Referee
Mr. David KATZ	Assistant Astronomer, GEPI, Observatory of Paris	Examiner
Mr. Hervé BOUY	Professor, LAB, University of Bordeaux	Jury president
Ms. Caroline SOUBIRAN	Research Director, LAB, CNRS	Thesis supervisor
Ms. Laia CASAMIQUELA	Post-Doctoral Researcher, LAB, University of Bordeaux	Thesis advisor

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ECOLE DOCTORALE: SCIENCES PHYSIQUES ET DE L'INGÉNIEUR

SPÉCIALITÉ: ASTROPHYSIQUE

Par **Yoann TARRICQ**

Révision des listes de membres, des paramètres structurels et de la cinématique des amas stellaires de la Voie Lactée.

Sous la direction de : **Caroline SOUBIRAN**
Co-encadrante : **Laia CASAMIQUELA**

Membres du jury :

Mme. Estelle MORAUX	Maître de Conférence, IPAG, Université Grenoble Alpes	Rapportrice
Mme. Céline REYLÉ	Astronome, UTINAM, Observatoire de Besançon	Rapportrice
M. David KATZ	Astronome-Adjoint, GEPI, Observatoire de Paris	Examinateur
M. Hervé BOUY	Professeur, LAB, Université de Bordeaux	Président du jury
Mme. Caroline SOUBIRAN	Directrice de Recherche, LAB, CNRS	Directrice de thèse
Mme. Laia CASAMIQUELA	Post-Doctorante, LAB, Université de Bordeaux	Co-Encadrante

"It's easy to know what you are against, but quite another to know what you are for."

Damien, *The Wind that Shakes the Barley* from Ken Loach.

Résumé

Révision des listes de membres, des paramètres structurels et de la cinématique des amas stellaires de la Voie Lactée.

Les amas stellaires sont des groupes d'étoiles gravitationnellement liées. Les astronomes distinguent deux types d'amas stellaires : les amas gobulaires et les amas ouverts (AO). Cette thèse se concentre sur ces derniers. Les étoiles appartenant à un AO donné ont toutes été formées à partir du même nuage de gaz moléculaire lors du même événement de formation stellaire. Des étoiles de masses très différentes se forment lors de tels événements mais elles partagent toutes la même cinématique, la même composition chimique et le même âge. Les membres d'un AO peuvent être dispersés sur de vastes régions car les AO se désintègrent au cours de leur vie, à cause des interactions gravitationnelles avec leur environnement mais la détermination des propriétés des AO est tout de même beaucoup plus précise que pour les étoiles isolées car basée sur plusieurs étoiles. Pour cette raison, les AO constituent des sondes efficaces dans le domaine de l'archéologie galactique, destinées à élucider les propriétés et l'évolution de notre Galaxie.

Le volume et la précision sans précédent des données de la mission *Gaia* ont radicalement changé notre vision de la population d'AO en nous révélant la position en trois dimensions ainsi que la cinématique en deux dimensions de plus d'un milliard d'étoiles. Dans cette thèse, nous visons à identifier les amas en voie de désintégration et leurs membres, ainsi qu'à revisiter les propriétés des AO à la lumière des données *Gaia*.

Nous avons utilisé de nouvelles techniques d'apprentissage automatique afin de détecter les étoiles partageant les mêmes cinématiques et parallaxes à proximité des amas. Si les noyaux des amas ont déjà fait l'objet de nombreuses études, notre méthodologie de regroupement appliquée à plusieurs centaines d'amas nous a permis de révéler qu'ils sont plus étendus que prévu, et ce indépendamment de leur âge. Nous publions de nouvelles listes de membres qui nous ont permis d'identifier 70 amas présentant de remarquables queues de marée, un signe de désintégration. Nous multiplions par plus de quatre le nombre de ces structures identifiées et mettons en évidence les différents processus d'évaporation des amas et les courtes échelles de temps dont ils ont besoin pour les affecter. Nous déterminons, de manière homogène et automatique, plusieurs paramètres morphologiques comme la taille des amas, les profils de densité radiale et les niveaux de ségrégation de masse.

Nous avons également combiné les mesures de vitesse radiale (VR) de la mission *Gaia* et de plusieurs relevés au sol pour mesurer la VR des AO et étudier leur cinématique en trois dimensions. Nous fournissons le plus grand catalogue de VR disponible pour les

AO, compilé et homogénéisé à partir de tous les catalogues disponibles de spectroscopie haute résolution complémentaires à *Gaia*. Cela nous permet de caractériser les propriétés cinématiques de la population d'AO et de les comparer à celles des étoiles de champ. Notre comparaison des différentes composantes galactiques de la vitesse montre que les nuages moléculaires géants sont moins efficaces pour disperser les AO que les étoiles de champ. Enfin, une analyse des orbites des amas montre que les AO naissent sur des orbites circulaires, et que plus ils vieillissent, plus leurs orbites sont enclines à subir des perturbations.

Mots-clefs: Galaxie: cinématique et dynamique – Galaxie: structure – Amas ouverts et associations: general– Méthodes: statistique – Galaxie: amas et associations – Galaxie: disque – étoiles: cinématique et dynamique

Abstract

Revisiting the memberships, structural parameters and kinematics of stellar clusters in the Milky Way.

Stellar clusters are gravitationally bound groups of stars. Astronomers distinguish two types of stellar clusters: Gobular and Open Cluster (OC). This thesis is focused on the latter. The stars belonging to a given OC have all been formed from the same cloud of molecular gas during the same event of star formation. Stars of very different mass are formed during such events but they all share the same kinematics, chemical composition and age. The members of an OC can be spead on vast regions because OCs disrupt during their lifetimes due to their gravitational interactions with their surroundings. The determination of the OCs properties is still much more accurate than for isolated stars as it is based on several étoiles. This makes OCs efficient probes for Galactic archeology studies, dedicated to unravel the properties and the evolution of our Galaxy.

The unprecedented volume and accuracy of the *Gaia* mission data dramatically changed our vision of the OC population by giving us access to the three dimensional positions and two dimensional kinematics of more than 1 billion stars. In this thesis, we aim at identifying disrupting clusters and their members and at revisiting the properties of OCs in the light of *Gaia* data.

We used novel machine learning techniques in order to detect the stars sharing the same kinematics and parallaxes in the surroundings of clusters. Our clustering methodology applied on several hundreds of clusters allowed us to reveal that clusters are not limited to their well studied cores but are more extended than previously expected, regardless of their age. We provide new lists of members which allowed us to identify 70 clusters with remarkable tidal structures, a sign of ongoing disruption. We multiply by more than four the number of such identified structures and highlight the different cluster evaporation processes and the short timescales they need to affect the clusters. We determine several morphological parameters like cluster sizes, radial density profiles and mass segregation levels in a homogeneous and automatic way.

We also combined Radial Velocity (RV) measurements from the *Gaia* mission and from several ground based surveys to measure the RV of OCs and to study their kinematics in three dimensions. We provide the largest RV catalogue available for OCs, compiled and homogeneised from all available catalogues of high resolution spectroscopy complementary to Gaia. This allows us to characterize the kinematical properties of the

OC population and to compare them with that of field stars. Our comparison of the different Galactic components of the velocity show that Giant Molecular Clouds are less efficient to scatter OCs than field stars. Finally, an analysis of the orbits of the clusters show that OCs are born in circular orbits, and as age increases they are more prone to suffer perturbations of their orbits.

Keywords: Galaxy: kinematics and dynamics – Galaxy: structure – open clusters and associations: general– methods: statistical – Galaxy: clusters and associations – Galaxy: disc – stars: kinematics and dynamics

Remerciements

Je remercie en premier lieu ma directrice de thèse, Caroline Soubiran. Merci tout d'abord pour avoir accepté ma candidature en tant que doctorant. Je te remercie surtout pour m'avoir fait confiance et pour m'avoir laissé la liberté de m'égarer dans la masse de données Gaia. Merci également pour avoir su me recadrer dans les moments où j'étais un peu perdu et pour m'avoir parfois mis la pression quand je n'avançais pas assez. Un énorme merci à Laia Casamiquela qui m'a co-encadré avec brio durant ses trois années. Merci pour tes encouragements, ta patience et ta gentillesse. Merci à toutes les deux pour m'avoir appris tant de choses et pour m'avoir inlassablement relu, aidé et corrigé et avoir ainsi considérablement amélioré la qualité de cette thèse.

Comme tout travail de recherche, cette thèse est le résultat d'un travail collaboratif impliquant de nombreuses personnes. J'aimerais donc remercier l'ensemble des collaborateurs avec qui j'ai travaillé et grâce à qui cette thèse a été possible.

En particulier, je tiens à remercier Laia, Javier, Alfred, Nuria ainsi que Phillip pour leurs innombrables conseils et suggestions qui m'ont considérablement aidé dans la compréhension et l'utilisation d'HDBSCAN. Un grand merci supplémentaire à Javier pour ses ultimes relectures et corrections !

Je tiens aussi à remercier celles qui font tourner le laboratoire au quotidien et qui nous aident énormément dans toutes les démarches administratives : merci à Cécile, Nathalie, Annick et Leïla

Merci aussi à tous les jeunes chercheurs du LAB : à Sacha évidemment, mon fidèle compagnon de bureau toujours partant pour aller manger au CNRS et boire des bières. Merci aussi pour être arrivé au labo pratiquement systématiquement après moi, un tel acharnement dans la flemme mérite d'être salué ! Un grand merci également à William pour les doodles destinés à aller boire des coups toustes ensembles. Merci à Yassin, Mélisse, Baptiste, Nuria, Laia, Maxime, Lars, Tiziano, Aurélien, Jean-Paul, Benjamin et Angèle pour avoir été là et avoir rendu la présence au labo plus sympa que le télétravail !

Un grand merci à François, Clémence, Anne-So, Tim, Antoine, Axel, Jérôme, Hélène et Théo pour être à mes côtés depuis des années. Merci en particulier à François de répondre toujours présent et à vous tous pour tous les bons souvenirs qu'on partage et partagera encore ensemble. Je ne pourrai pas rêver avoir de meilleurs amis !

Je remercie également Victor et Geoffrey pour nos conversations sans fin, nos coups de gueule et notre lassitude que l'on a partagé tout au long de cette thèse mais que l'on partage depuis bien plus longtemps. Merci de m'avoir remotivé quand j'étais déprimé et de m'avoir aidé contre un saucisson. Tout ça c'est pour le torrent ! Merci également à Geoffrey et Aude de m'avoir proposé de cuisiner pour le pot de thèse et m'avoir ainsi enlevé son organisation des mains. Merci aussi à ceux qui m'accompagnent depuis mes années de licence: Thibaut, Olivier, Julio et Charles.

Merci à tous les gens que j'ai rencontré et cotoyé lors des mobilisations, au sein des Preks et de l'Association Continuité Alimentaire Bordeaux. Merci pour avoir donné du sens à ces longs mois de confinement. Merci surtout à Yannis, Laure, Ludovic, Charles,

Natacha et Marcu-Antone qui sont (re)devenus des amis formidables.

Je remercie particulièrement mon père. Merci d'avoir essayé de satisfaire ma curiosité, d'avoir supporté et essayé de répondre à mes questions incessantes quand j'étais petit. Merci surtout de m'avoir toujours soutenu et encouragé. Un grand merci à ma sœur, Coralie, pour ne pas m'avoir fait taire même à la Cité de l'Espace. Merci pour m'avoir fait découvrir tant de bons groupes et de bons livres qui m'ont permis de m'échapper un peu pendant ces deux dernières années.

Enfin, merci à Judith pour ta gentillesse et tes encouragements. Merci pour tes conseils et pour ton aide. Merci de me faire rire et d'avoir rendu ces derniers mois beaucoup plus agréables. Cette période n'aurait pas été la même sans toi.

Résumé Substantiel

Introduction

Les amas stellaires sont des groupes d'étoiles liées gravitationnellement. Deux types d'amas stellaires sont observés au sein de la Voie Lactée : les Amas Globulaires que l'on trouve principalement dans le bulbe, le halo et le disque épais et les Amas Ouverts (AO) qui sont ceux qui nous intéressent dans cette thèse. Les AO sont constitués d'étoiles qui se sont toutes formées lors du même évènement de formation stellaire, lorsque le cœur d'un Nuage Moléculaire Géant s'est effondré gravitationnellement. Ces étoiles couvrent une large gamme de masses mais partagent le reste de leurs propriétés telles que leur âge, composition chimique, position et cinématique. Peu d'AO survivent plusieurs centaines de millions d'années et la plupart se désintègrent au cours de leur vie, leurs étoiles se dispersant sur de vastes régions. De nombreux processus sont en cause dans la dislocation des AO tels que les interactions avec des Nuage Moléculaire Géants, l'évolution stellaire ou le passage d'un bras spiral. Malgré cela, les AO sont d'une importance cruciale dans le domaine de l'archéologie galactique qui vise à étendre notre compréhension de l'évolution et de la structure de la Voie Lactée. En effet, la détermination de leurs paramètres est bien plus précise que pour les étoiles isolées car elle est basée sur plusieurs étoiles. Sachant que la formation d'étoiles a lieu dans les bras spiraux, les amas jeunes peuvent être utilisés pour retracer leurs position et comprendre leur nature. En outre, abritant des étoiles couvrant une large gamme de masse et de stades d'évolution stellaire, les AO constituent des laboratoires permettant de tester les théories de formation et d'évolution stellaire.

La mission *Gaia* et en particulier sa DR2 a révolutionné notre connaissance des AO. La position en trois dimensions ainsi que la cinématique en deux dimensions de plus d'un milliard d'étoiles nous est maintenant accessible avec une précision inégalée. La précision remarquable des mesures astrométriques de *Gaia* ainsi que sa couverture de tout le ciel a permis d'établir de nouvelles listes de membres pour les AO déjà connus, d'identifier beaucoup plus de membres pour ces amas et d'écarter certains astérismes, précédemment considérés comme des AO. Elle a également permis la détection de structures proéminentes autour de certains AO par plusieurs équipes. Les halos et les queues de distribution d'un certain nombre d'amas proches ont été détectées et caractérisées, notamment par Meingast et al. (2021) et Pang et al. (2021).

La DR2 de *Gaia* a également révélé la complexité de la distribution locale des vitesses avec une résolution sans précédent. En particulier, le mouvement des AO peut être utilisé pour comprendre le potentiel gravitationnel de la Voie lactée et les différentes perturbations qui agissent sur la structure et la dynamique du disque galactique.

Dans cette thèse, nous avons revisité les listes de membres des AO à la lumière des données *Gaia* en nous concentrant sur l'identification des étoiles appartenant aux amas et se situant à grande distance de leur centre. Nous avons étudié les morphologies en deux dimensions de ces amas en voie de désintégration et redéterminé leurs paramètres

structurels. Enfin, en tirant parti des dernières listes de membres et de multiples catalogues de VR, nous avons étudié la cinématique en trois dimensions des AO. Nous avons porté un intérêt particulier à l'étude de l'évolution des propriétés des amas en fonction de leur âge.

Partitionnement: comment identifier les membres d'un Amas Ouvert ?

De nombreuses méthodes statistiques dédiées à la détection des queues de marée des amas ont été utilisées depuis la publication de *Gaia*-DR2. Dans une première approche, nous avons essayé d'identifier les membres des amas en nous basant sur les 3 composantes de leurs vitesses héliocentrique. Malgré les défauts inhérents à cette approche tels que la limite en magnitude et l'incomplétude du *Gaia*-RVS, cette approche nous a tout de même permis d'identifier la queue de marée de Ruprecht 147, une détection confirmée par la suite par Yeh et al. (2019). Pour améliorer l'étude des zones externes des amas, nous avons développé une méthodologie capable d'identifier les membres dans la zone périphérique des AO, jusqu'à 50 pc de leur centre. La méthode est basée sur l'algorithme de regroupement non supervisé HDBSCAN qui peut détecter des surdensités dans l'espace astrométrique $(\mu_{\alpha^*}, \mu_\delta, \varpi)$, et ce même dans des ensembles de données de densité variable. Nous avons appliqué cette méthode aux 467 AO de Cantat-Gaudin et al. (2020) situés à moins d'1,5 kpc et âgés de plus de 50 millions d'années. Nous rapportons des listes de membres pour 389 d'entre eux, les 78 restants étant trop proches de leurs voisins dans l'espace astrométrique utilisé pour que notre méthode puisse les séparer correctement. Pour la grande majorité des amas, nous identifions beaucoup plus de membres que ce qui était connu auparavant.

Nous identifions de vastes halos d'étoiles autour de presque tous les amas, atteignant dans la plupart des cas le rayon maximal de la zone étudiée. Nous identifions également des queues de marée pour 71 AO. Les précédentes détections de couronnes ou de queues de marée d'AO étaient concentrées sur des échantillons plus petits d'amas proches. Comme nous avons travaillé sur de grands ensembles de données qui incluent la sphère de 50 pc entourant chaque amas, nous avons pu établir des listes de membres à grand rayon, même pour les AO situés à de grandes distances héliocentriques. Nous avons ainsi multiplié par plus de 4 le nombre d'amas identifiés avec une queue de marée.

Paramètres structurels des Amas Ouverts

Dans cette partie, notre principal objectif était de déterminer les paramètres structurels d'amas identifiés depuis la publication de *Gaia*-DR2. Nous avons d'abord essayé de mesurer la forme tridimensionnelle des AO mais le nombre d'amas suffisamment proches pour être étudiés avec précision en trois dimensions est limité aux $\sim$ 17 AO situés à moins de 250 pc. Nous nous sommes donc concentrés sur la morphologie bidimensionnelle des amas pour lesquels nous avons obtenu de nouveaux membres à grand rayon. Pour ce faire, nous avons ajusté les profils de densité radiale de chaque amas comptant plus de 100 membres avec une fonction de King afin d'étudier leurs rayons de cœur et de marée.

Nous trouvons des rayons de cœur similaires à ceux publiés dans les études précédentes, mais comme nous trouvons des membres à un rayon beaucoup plus grand que dans les études précédentes, nos rayons de marée sont également beaucoup plus grands. Notre distribution de rayons de cœur montre une concentration entre 1 et 2.5 pc indépendamment de l'âge ou du nombre de membres de l'amas. Nous constatons également que les amas les plus anciens ont tendance à avoir des R_c plus petits comparés aux jeunes amas avec des valeurs convergeant vers 1.85 pc à l'âge d'1 milliard d'années. Nos rayons de marée culminent autour de 30 pc, mais plus important encore, ils semblent augmenter avec l'âge, ce qui pourrait être un signe de ségrégation de masse, de dissolution ou d'une combinaison des deux processus. La distribution des rayons de marée pourrait être biaisée en raison de la limite de 50 pc que nous avons utilisée pour interroger le catalogue *Gaia*-eDR3. Nous avons également ajusté des profils EFF et de Plummer sur le profil de densité radiale de notre échantillons d'amas. Le rayon caractéristique du profil de Plummer a confirmé la diminution de la taille des noyaux avec l'âge, tandis que le paramètre γ du profil EFF a montré que, quelle que soit leur taille, les profils de densité radiale des anciens amas ont tendance à diminuer au même rythme à grand rayon.

Nous avons également ajusté des modèles de mélange gaussien sur la distribution spatiale des membres projetés sur un plan perpendiculaire à la sphère céleste. Ceci est particulièrement adapté aux amas pour lesquels nous avons détecté une queue de marée allongée car les profils de densité radiale décrits précédemment supposent une distribution sphérique des membres. Nous avons utilisé une fonction à trois composantes gaussiennes sur les 71 amas présentant une queue de marée: une composante représentant le noyau de l'amas, une la queue de marée et une la couronne. Pour les autres amas, nous avons adopté un modèle de mélange gaussien à deux composantes avec un noyau et un halo. Nous avons cherché des corrélations entre les paramètres des gaussiennes ajustées et les caractéristiques des amas (c'est-à-dire leur âge, leur emplacement, le nombre de membres, etc). Pour les 71 amas de notre échantillon présentant une queue de marée, nous observons que les vieux amas ont tendance à avoir de plus petits noyaux que les amas jeunes. Nous notons également que le poids relatif du halo des vieux amas est en moyenne plus faible que celui des jeunes. Ceci implique que lorsque l'âge des amas augmente, la proportion d'étoiles dans leurs halos diminue, soit parce que leurs étoiles se déplacent vers leurs centres, soit parce que leurs étoiles externes sont éjectées.

Nous avons appliqué la méthode proposée par Allison et al. (2009b) pour mesurer le degré de ségrégation de masse de notre échantillon d'AO. Nous ne remarquons aucune tendance particulière du taux de ségrégation de masse, mesuré à travers les 10 étoiles les plus massives, avec l'âge, la distance galactocentrique ou avec l'altitude de l'amas au-dessus du plan galactique. Cependant, les amas ayant un nombre significatif d'étoiles avec un taux de ségrégation de masse élevé sont en moyenne plus âgés que les amas avec peu d'étoiles fortement ségréguées en masse. Couplé à une plus faible proportion d'étoiles peuplant les halos d'amas, cela met en évidence le fait que les différents processus physiques en jeu dans la dislocation des amas agissent sur des échelles de temps plus courtes que la ségrégation de masse.

Propriétés cinématiques des Amas Ouverts

En combinant *Gaia* ainsi que des relevés et catalogues au sol avec les nouvelles listes de membres et déterminations de l'âge des AO de Cantat-Gaudin et al. (2020), nous avons assemblé le plus grand catalogue de VR pour AO destiné à l'étude de leur cinématique. Comme sous-produit de notre étude, qui inclut le *Gaia* DR2-RVS, les relevés *Gaia*-ESO, APOGEE, RAVE, GALAH, et des catalogues plus petits, nous avons comparé les VR des différents relevés entre eux afin d'évaluer leur précision et leur point zéro. Nous avons constaté que les points zéro de chque catalogue sont consistants à moins d'1 km s^{-1}. La dispersion des comparaisons indique que la précision réelle de chaque catalogue est compatible avec les incertitudes individuelles qu'il contient. Toutes les mesures de VR non *Gaia* ont été corrigées pour les aligner sur le point zéro de *Gaia*-RVS. La moyenne pondérée des VR de chaque étoile et de chaque amas nous a permis d'obtenir 1 382 AO avec une VR dont 38% ont une VR extrêmement précise basé sur plus de trois étoiles et avec une incertitude inférieure à 3 km s^{-1}.

Nous avons calculé les vitesses cartésiennes et cylindriques héliocentriques et galactocentriques pour cet échantillon d'AO et défini un échantillon de haute qualité composé des 418 AO ayant les vitesses les plus fiables. Parmi ces 418 amas, 411 d'entre eux ont un âge déterminé. Nous avons constaté que la plupart des AO tombent dans une bande située entre les deux arcs principaux dessinés par les étoiles de champ dans le plan $V_r - V_\phi$, tandis qu'ils semblent suivre les surdensités décrites par les crêtes diagonales dans le plan $R_{GC} - V_\phi$. La courbe de rotation tracée par nos AO montre deux creux significatifs : à $R_{GC} \sim 7$ kpc, et un autre plus important autour de $R_{GC} \sim 9{,}7$ kpc. Les emplacements et les profondeurs de ces creux sont en accord avec les perturbations que l'on pourrait attendre des composantes non-axisymétriques du disque, qui dessinent également les crêtes observées dans le plan $R_{GC} - V_\phi$.

Avec les âges de presque tous les amas de notre échantillon, nous avons étudié en détail la relation âge-vitesse des AO. Elle montre une nette anisotropie des trois composantes de la vitesse. Comparé aux études sur les étoiles de champ, le paramètre de chauffage β des AO s'est avéré être similaire dans les directions radiale et azimutale, mais significativement plus faible dans la direction verticale. Ce faible taux de chauffage dans la coordonnée Z peut être dû à la dislocation des vieux amas, qui sont les plus susceptibles d'atteindre de hautes altitudes au-dessus du disque, ou à un chauffage cinématique moins efficace des AO par les nuages moléculaires géants. Nous sommes cependant conscients que les coupures de qualité que nous avons appliquées à notre échantillon ont écarté des amas distants, ce qui a entraîné un biais dans notre échantillon.

Nous avons utilisé les informations 6D ainsi que l'âge de notre échantillon d'AO pour calculer et étudier les orbites et les variables d'action. Nous avons analysé les dépendances des paramètres orbitaux récupérés en fonction de l'âge. La plupart des amas atteignent une altitude maximale au-dessus du plan galactique qui est inférieure à 400 pc, et seuls ceux âgés de plus d'1 milliard d'années sont capables de s'éloigner davantage du plan médian. Ils s'eloignent toutefois rarement de plus d'1 kpc. Les amas âgés de moins de 30 millions d'années montrent une excentricité très faible (~ 0.018), et pour les amas plus

anciens, en particulier ceux de plus de 100 millions d'années, l'excentricité montre une relation croissante avec l'âge. Ces résultats montrent que les AO sont nés sur des orbites circulaires, et que plus l'âge augmente, plus ils sont enclins à subir des perturbations de leurs orbites. Cela se voit également après le calcul des variables d'action, où lorsque l'âge augmente, la distribution dans le plan $(L_Z, \sqrt{J_R})$ tend à s'étendre au-delà de $\sim$ (1,0). Nous avons mis en relation la distribution de nos groupes dans cet espace d'action avec la localisation des groupes mobiles connus en fonction de l'âge. Nous concluons que les groupes mobiles des Pléiades, des Hyades et de Coma semblent être peuplés de jeunes amas, tandis que la région de Sirius semble abriter un certain nombre d'amas d'âge $\gtrsim 300$ Myr.

Conclusion

Tout au long de cette thèse, nous avons essayé de mettre en évidence les processus physiques à l'œuvre dans notre galaxie à travers l'étude des "pierres angulaires" de notre Voie Lactée : les Amas Ouverts. L'étude de leur morphologie, cinématique et des échelles de temps nécessaires pour influencer ces paramètres nous renseignent sur les caractéristiques de la Voie Lactée. Nous avons établi de nouvelles listes de membres étendues pour des AO situés à moins d'1.5 kpc et agés de plus de 50 millions d'années et mis en évidence la présence de vastes halos stellaires environnant une large majorité d'entre eux. L'étude de la morphologie de ces AO et de leur ségrégation de masse nous a permis d'oberver un dépeuplement des halos des amas les plus âgés et de postuler que les différents processus physiques en jeu dans la dissolution des amas tels que les rencontres avec les Nuages Moléculaires Géants ou les passages des bras spiraux agissent sur des échelles de temps bien plus courtes que la ségrégation de masse.

En croisant les dernières listes de membres obtenues grâce aux données *Gaia* avec de nombreux relevés de VR, nous avons également construit le plus large catalogue de VR pour AO à ce jour. Ce catalogue nous a permis d'étudier avec précision la cinématique en trois dimensions des AO et d'observer des différences notables entre la cinématique des amas et des étoiles de champ. En plus de ne pas être alignés avec les structures de l'espace des phases dessinées par les étoiles de champ dans le voisinage solaire, les AO souffrent également de moins de chauffage cinématique que ces étoiles dans la direction perpendiculaire au plan Galactique. Les rencontres avec les Nuages Moléculaires Géants, source principale de chauffage cinématique des étoiles de champ sont ainsi potentiellement plus efficace pour disloquer les amas que pour les disperser dans la direction verticale.

Contents

List of Figures

List of Tables

List of Acronyms

2MASS Two Micron All-Sky Survey

4MOST 4-metre Multi-Object Spectroscopic Telescope

APOGEE Apache Point Observatory Galactic Evolution Experiment

AVR Age-Velocity Relation

CMD Color-Magnitude Diagram

CG+18 Cantat-Gaudin et al. (2018a)

CG+20 Cantat-Gaudin et al. (2020)

DR Data Release

DBSCAN Density-Based Spatial Clustering of Applications with Noise

eDR3 early Data Release 3

EDSD Exponentially Decreasing Space Density

EFF Elson, Fall & Freeman

EM Expectation-Maximization

ESA European Space Agency

GALAH GALac-tic Archaeology with HERMES

GC Globular Cluster

GMC Giant Molecular Cloud

GMM Gaussian Mixture Model

HDBSCAN Hierarchical Density-Based Spatial Clustering of Applications with Noise

HDI Highest Density Interval

HQS High-Quality Sample

HR Hertzsprung-Russell

LAMOST Large Sky Area Multi-Object Fibre Spectroscopic Telescope

LSR Local Standard of Rest

MAD Median Absolute Deviation

ML Maximum Likelihood

MST Minimum Spanning Tree

MSR Mass Segregation Ratio

OC Open Cluster

OCCASO Open Cluster Chemical Abundances from Spanish Observatories

PDF Probability Density Function

RAVE Radial Velocity Experiment

RV Radial Velocity

RDP Radial Density Profile

RVS Radial Velocity Spectrometer

SFR Star Forming Regions

S/N Signal-to-Noise ratio

UBC University of Barcelona Clusters

UCAC US Naval Observatory CCD Astrograph Catalog

UPMASK Unsupervised Photometric Membership Assignment in Stellar clusters

WEAVE WHT Enhanced Area Velocity Explorer

YSO Young Stellar Objects

Chapter 1

Introduction

Contents

1.1 Historical context

Ever since antiquity, star clusters have been examined by astronomers. The appearance of the Pleiades in the night sky was scrutinized every year as its emergence marked the beginning of the harvest in the northern hemisphere. Later the roman astronomer Ptolemy mentions in his work several clusters such as Praesepe, the Double Cluster or NGC 6475 (also known as the Ptolemy cluster, see Fig. 1.1) and classifies them as nebulae (Moore and Rees, 2011). The invention of the refracting telescope at the beginning of the 17$^{\text{th}}$ century allowed its first users to resolve these nebulae into their members. In particular, Galileo turned his telescope towards some of the nebulae identified by Ptolemy and reported 50 stars belonging to the Praesepe cluster (Maran and Marschall, 2009). At the end of the 18$^{\text{th}}$ century, the astronomers William and Caroline Herschel started a systematic study of objects having a nebulous appearance: the *Catalogue of Nebulae and Clusters of Stars* (Hoskin, 1987). They realised that many of these objects were composed of individual stars. One century later, hundreds of star clusters were identified and classified in the *New General Catalogue* and in the *Index Catalog* compiled by Dreyer (1888, 1895, 1910). The advent of telescopes revealed that star clusters could be divided into two categories: the so-called Globular and Open Clusters. While the first ones are composed of thousands of stars, are spherically shaped and can be observed all across the sky, the latter are more sparsely populated, have a larger variety of shapes and were generally identified near the Galactic plane. At the beginning of the 20$^{\text{th}}$ century, Ejnar Hertzsprung draws the first Color-Magnitude Diagram (CMD) of OCs for the Pleiades and the Hyades and noticed that these two clusters had different stellar populations (Strand, 1977). This was later explained by the age difference between those two clusters.

Figure 1.1: Images of six OCs, showed for descriptive purposes. From left to right and top to bottom are shown respectively the Pleiades, NGC 6475, NGC 6705, NGC 3766, NGC 2547 and NGC 2422. Crédit: Davide de Martin & ESO.

1.2 Open clusters

1.2.1 Definition

OCs are groups of stars formed from the same event of gravitational collapse of a Giant Molecular Cloud (GMC). Their stellar populations are gravitationally-bound and as they formed from the same cloud of gas and dust, they share the same age, distance, motion and chemical composition.

Having said that, a clear and universal definition of OCs does not exist. In their review on young massive star clusters, Portegies Zwart et al. (2010) simply defined a star cluster as a gravitationally bound group of stars. Focusing on young embedded clusters, Lada and Lada (2003) proposed that a cluster is a set of physically related stars having a stellar mass volume density large enough ($\geq 1.0 M_{\odot}\mathrm{pc}^{-3}$) to resist tidal disruption and populated enough to survive for at least 100 Myr. Because they focused on star clusters of all ages and masses, Krumholz et al. (2019) could not adopt the same definition. Young clusters have complex spatial distributions, usually made of several clumps. In order not to exclude from their study young clusters such as the Orion Nebula Cluster, they defined a cluster as a group of at least 12 stars with a mean density at least a few times greater than the background density and much greater than the local density of dark matter. Even if these definitions do not perfectly overlap, they still allow to discriminate star clusters and associations. The latter are gravitationnally unbound and have stellar density of $\sim$0.01 stars/pc^3 at most, with stars spreading over areas that can reach a few hundred parsecs. Consequently they only survive for a few tens of Myr because they get disrupted easily by the dissipative nature of the Galactic disc (Moraux, 2016). On the contrary, OCs are bound, much denser than associations, have a stellar density $\sim$10 times larger and can survive for several Gyr.

All of these definitions fail to make a clear distinction between OC and Globular Cluster (GC) the other type of clusters found in the Milky Way. Several aspects allow to discriminate between those two types of clusters. First, according to Gratton et al. (2019), GC are only found in the bulge, halo and thick disc of the Galaxy while OCs are mainly located in the thin disc (see Fig. 1.2 for a schematic view of the Milky Way). This implies that GCs are old ($\sim$ 10 Gyr) while most OCs are younger than 1 Gyr and are still forming in the spiral arms of the Galaxy. Because OCs are much younger than GCs, they have been formed from an environment already enriched in metals by the atomic nucleosynthesis of population II stars. GCs are therefore metal poor compared with OCs. Finally, OCs are composed of the same population of stars, and are chemically homogeneous (Moraux, 2016) while GCs can contain multiple stellar populations (Piotto et al., 2005) and can be chemically inhomogeneous (Gratton et al., 2012). In this thesis, we are focusing on OCs from the Milky Way.

1.2.2 Properties of Open Clusters

OCs were long thought to be the only environment in which stars are formed. However, observations with the *Spitzer Space Telescope* revealed the filamentary structure of Young

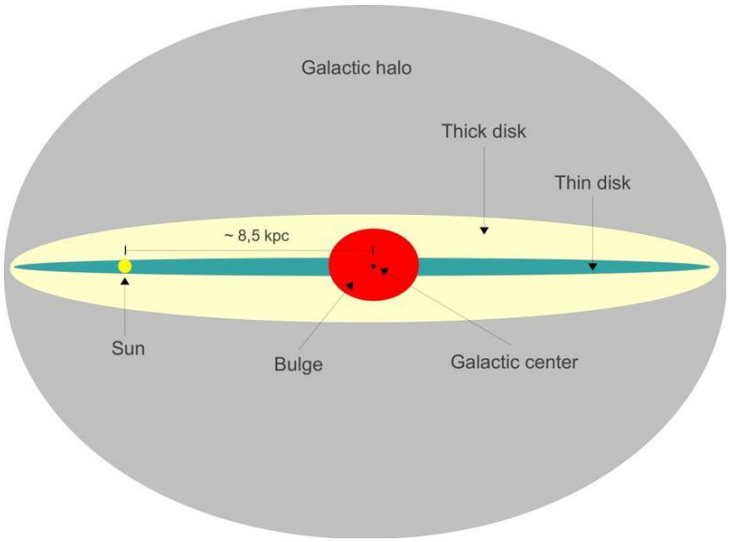

Figure 1.2: Schematic edge on view of the Milky Way with its main structures indicated. Credit: Wikipedia

Stellar Objects (YSO) in the Star Forming Regions (SFR) NGC 1333 and highlighted the fact that they follow the gas distribution (Gutermuth et al., 2008). This illustrates that the importance of clusters in star formation process has possibly been overestimated in the past as stars can also formed hierarchically in environments of varying densities from loose associations to dense clusters (Bastian, 2011; Gieles and Portegies Zwart, 2011). Nowadays it would be more accurate to state that stars form in groups rather than in clusters (Moraux, 2016). More precisely, bound clusters are formed in the dense cores of GMC and associations in their low density regions.

Star clusters are key objects to a better understanding of the star formation process. They are very useful probes to test theories of star evolution: stellar populations in OCs usually cover a wide range of masses and stellar evolutionary phases, making their Hertzsprung-Russell (HR) diagrams a snapshot of stellar evolution. Moreover, their properties (mainly distance and age) are determined with much higher accuracy than for field stars. That is why they are ideal laboratories to test theories of stellar evolution. Because star formation occurs in the spiral arms, young OCs can be used to trace their position and understand the mechanisms at play in their formation (Dias and Lépine, 2005; Castro-Ginard et al., 2021). Old clusters hosting more evolved stars, they constitute excellent spectroscopic targets which can be used to determine with accuracy their RV or their chemical composition (Tarricq et al., 2021a; Casamiquela et al., 2020). Besides, as their ages cover the entire lifespan of the Galactic disc, tracing the young to old thin disc components, they are also widely used to trace the history of the Galactic disc (Friel, 1995).

Repeated efforts have been made in the past to build large catalogs of OCs and to provide an estimation of their mean positions, proper motions, distance, etc. The first major step in this direction comes with the Lund catalogue (Lynga, 1982) which gathers around 1200 entries. In the 1990s, the Database for Galactic Open Clusters

(Mermilliod, 1995) covers data for around 100 000 stars from 500 clusters, mostly part of the Lund catalog. Dias et al. (2002) updated the Lund catalogue of OCs by gathering new discoveries using Two Micron All-Sky Survey (2MASS) data (Skrutskie et al., 1997). They provided -when enough data was available- the age, RV and proper motion for more than 1 500 clusters in a systematic way. Later, Kharchenko et al. (2013) published a catalogue ∼3 000 clusters with estimations of their radii, age, position and proper motions by exploiting both 2MASS and the PPMXL survey (Roeser et al., 2010).

1.2.2.1 Clusters formation & evolution

Since the *Herschel* space mission, GMCs are known to be composed of a network of filaments and clumps (Molinari et al., 2014). Within these clouds, star formation is a multi-scale process which involves at each scale different physical process. Star formation occurs when the mass of a clump exceeds the Jeans mass (Jeans, 1902) triggering the gravitational collapse of the cloud. Such gravitational collapse can also be triggered by the passage of a spiral arm (among other processes) (Shu et al., 1972). Molecular clouds are in addition subject to internal turbulence. At the scale of the clouds, turbulence generally acts against gravitational collapse but at smaller scales, it compresses the gas and fragments the cloud into clumps and filaments (McKee and Ostriker, 2007). Massive clumps can suffer from multiple events of such gravoturbulent fragmentation, leading to the formation of multiple stars. A more detailed description of the molecular clouds or the physical process triggering the onset of star formation are out of the scope of this thesis.

OCs form in the massive and dense clumps located at the intersection of the long filamentary substructures of GMCs (Lada, 2010). As the star-formation efficiency is relatively low (Lada and Lada, 2003; Clark et al., 2004), only a fraction of the gas from the GMC is actually consumed to form stars. A recently born cluster is therefore composed of both stars and gas in virial equilibrium (Kroupa, 2005). Due to stellar winds or supernova explosion, the gas is then expelled from the cluster, lowering its gravitational potential. Having lost its gas, the cluster is no longer in virial equilibrium and has to dynamically relax to reach another state of equilibrium. During this process, a significant number of the stars from a cluster are expelled from it and most galactic star clusters are destroyed. The surviving ones are thought to suffer from a dramatic star loss (Baumgardt and Kroupa, 2007). Lada and Lada (2003) postulated that $\sim 90\%$ of the clusters suffered from such a destruction and refers to this process as infant mortality. In the surviving clusters, this star loss can also lead to the formation of tail like structures (Hu et al., 2021; Meingast et al., 2021).

Embedded clusters still being surrounded by gas and dust are obscured in the optical photometric bands and only infrared observations can reveal their structures. Lada et al. (1993) and Lada et al. (1996) respectively conducted pioneering infrared observations of NGC 2264 and NGC 1333 in the J, H and K bands. They showed that some embedded clusters are not necessarily centrally concentrated systems but exhibit many clumps and many structures showing they kept in their spatial distribution the imprint of their GMC progenitor (Gutermuth et al., 2008). The surface density of these hierarchically formed

clusters do not have smooth radial profiles and are poorly described by simple power law functions. On the contrary, centrally concentrated systems which formed at the intersection of gas filaments are well described by functions like the Elson, Fall & Freeman (EFF) (Elson et al., 1987) or the King profile (King, 1962). The EFF profile does not assume an energetic equilibrium of the cluster so it is most suited for young clusters, which have not completely relaxed yet. After a relaxation time (i.e. the time needed for a system to reach an energetic equilibrium, typically a few tens of Myrs), clusters have lost their memory of their initial conditions, they are fully relaxed and their radial profile becomes well fitted by the King function. The relaxation time is defined as:

$$t_{relax} \simeq \frac{0.1N}{\ln N} t_{cr}, \tag{1.1}$$

(Binney and Tremaine, 1987; Lada and Lada, 2003) with N the number of cluster members and t_{cr} the dynamical crossing time which represents the time needed for a typical star to orbit around the system.

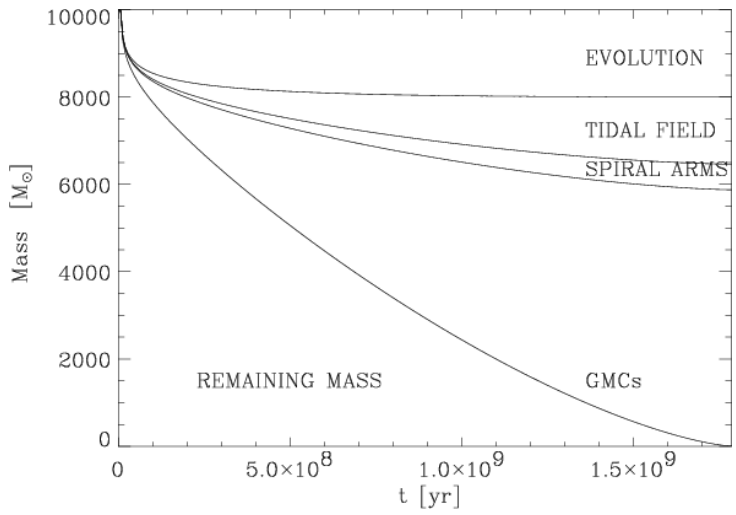

Figure 1.3: Mass evolution of a cluster from the solar neighborhood with an initial mass of 10^4 $M_\odot$. The solid line represents the contribution to this mass loss from respectively (frop top to bottom) stellar evolution, galactic tidal field harassment, spiral arm shocking and molecular clouds encounters. From Lamers and Gieles (2006)

Internal cohesion of OCs continuously undergo perturbations due to both internal and external physical processes. Stellar evolution causes the clusters to loose some of its mass through supernova explosion and stellar winds. Gravitational interactions with GMCs are also known to greatly perturb the clusters equilibrium (Gieles et al., 2006). Spiral arm shocking also contribute to the evaporation of stars from stellar clusters (Gieles et al., 2007). Finally, Galactic harassment i.e. simple interactions with the Galactic field can also progressively contribute to the disruption of clusters (Baumgardt and Makino, 2003). Lamers and Gieles (2006) used an approximative analytical model to estimate the mass loss experimented by OCs in the solar neighborhood, taking all these processes into account. Figure 1.3 is adopted from their work where we can see that the dissolution of

a cluster with an initial mass of 10^4 $M_\odot$ is dominated by the encounters with GMCs.

These events are complemented by another dynamical process: mass segregation (Spitzer, 1969). Because of equipartition of kinetic energy in gravitational encounters between clusters members, massive stars move around the clusters centers slower than less massive members. Massive stars therefore sink towards the clusters centers on lower orbits than less massive stars. Mass segregation has been highly studied in OCs and is believed to contribute to cluster evaporation (Mathieu, 1984; Kroupa, 1995; de La Fuente Marcos, 1996). Low mass stars on outer orbits are indeed more likely to be stripped from the clusters gravitational potential by the tidal shocking events previously discussed. The multitude of physical processes at play in the disruption of a cluster implies that the age distribution of OC is highly biased: very few cluster reach ages of more than 1 Gyr compared to the number of cluster which are formed.

1.2.2.2 Clusters spatial distribution

The study of the structure of the Milky Way is not an easy task given the location of the Sun inside the Galactic disc. Thus, a detailed view of the spiral arm features is still missing. The origin, location and even the number of the spiral arms still remain uncertain. As star formation is believed to happen in the spiral arms, their position can be traced by young populations, such as young OCs (Roberts, 1969). Dias and Lépine (2005) used the young clusters from their sample published in Dias et al. (2002) to investigate the nature of the spiral arms. Figure 1.4 shows the positions of their clusters in three different age bins. It is clear that the very young clusters in their sample constrain efficiently the position of the spiral arms. We also note that they quickly fill the inter arm region: the spiral arms are still distinguishable in the middle panel even though clusters already started to drift. No structures are visible in the bottom panel anymore, illustrating the heating of the clusters by molecular clouds and by the non-axisymmetric features of the disc. More thorough studies have tried to map the spiral arms in the surroundings of the Sun. For example, (Vázquez et al., 2008) noticed a gap in the distribution of clusters in the Perseus arm. This gap also seen in the high-mass star-forming regions from Reid et al. (2014) and in the distribution of HII regions from Hou and Han (2014) could change our view of the Perseus and local arms but no conclusion has been reached yet.

The scattering of the OCs from their birthplaces can also be witnessed in the vertical direction by studying the scaleheight of the disc. OCs are mostly located in the thin disc even though old ones which have been scatterred by their encounter with GMCs and by the spiral arms (Spitzer and Schwarzschild, 1951; Lacey, 1984; Jenkins and Binney, 1990) can also venture in the thick disc. Old clusters therefore tend to be located at higher Galactic altitudes than young ones (Friel, 1995; Cantat-Gaudin et al., 2018a). In addition, clusters which remain for a long time in the Galactic mid-plane are more likely to be disrupted by the different processes discussed in Sect. 1.2.2.1. On the contrary, clusters which move away from the mid-plane survive longer as they suffer less from these effects, in particular from GMC encounters. The lack of old clusters at low altitudes is also a product of this selection effect.

Bonatto et al. (2006) examined the vertical distribution of a sample of more than 600

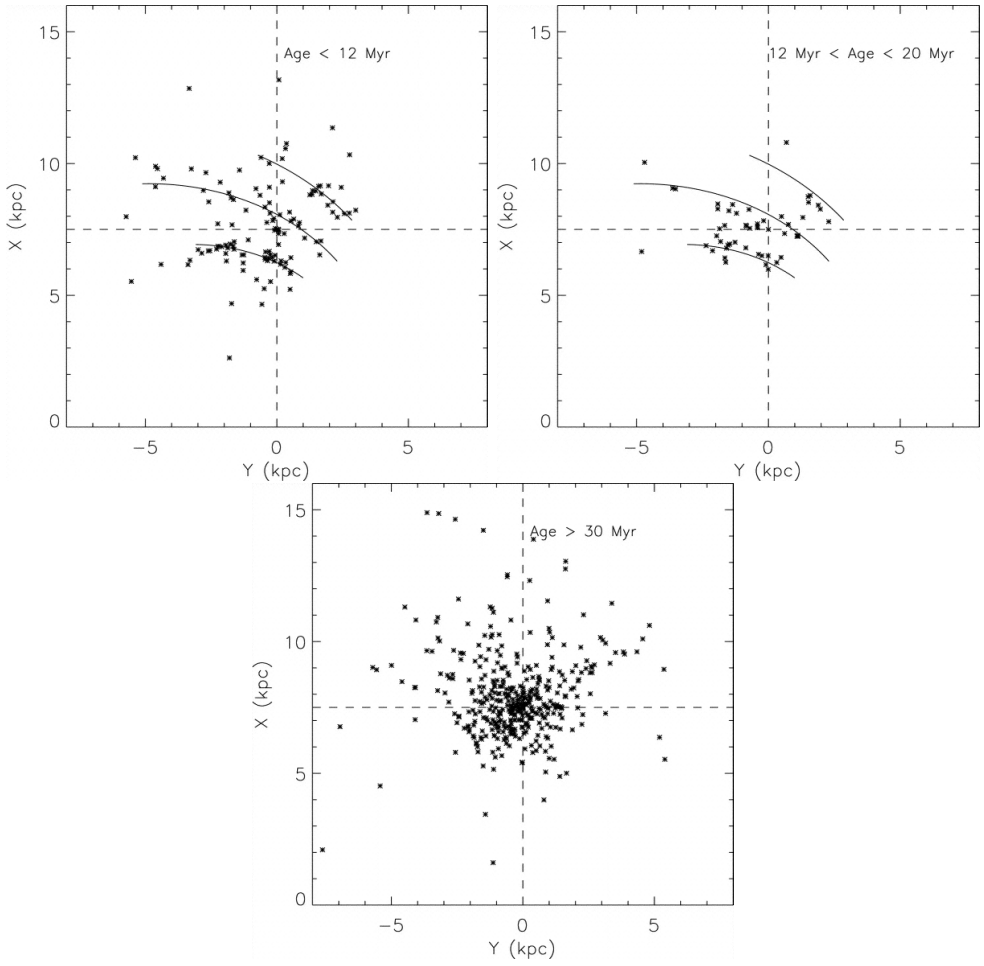

Figure 1.4: Distribution in the $(X - Y)$ plane of the clusters from Dias and Lépine (2005). In the top left panel are shown the clusters younger than 12 Myrs and in the top right panel are shown the clusters with an age comprised between 12 and 20 Myr. The solid line represent the present-day position of the spiral arms. The bottom panel shows the distribution of clusters older than 30 Myr. The Sun is located at the coordinates (0, 7.5) in all panels.

OCs. Assuming their distribution could be described by a simple exponentially decay profile and that young OCs were symmetrically distributed above and below the plane, they calculated the scale height of the clusters in different age bins and the vertical position of the Sun. They found the Sun to be at an altitude of 14 pc above the Galactic mid-plane. As for the scale-height of OCs, they reported an increase by a factor of three between clusters younger than 200 Myr and clusters with intermediate age (comprised between 200 Myr and 1 Gyr) but a uniform distribution of clusters older than 1 Gyr. They also report a clear increase of the scale height of clusters located in the outer Galaxy compared with inner clusters. Like before, this might just be a selection effect due to a higher survival rate of OCs at large Galactocentric radius.

1.2.2.3 Clusters kinematics

The first studies of the kinematics of OCs were usually dedicated to investigate the local kinematical field. Hron (1987) conducted one of these studies by compiling RVs for 105 young clusters. He used these data to measure the solar motion and one of the Oorts constants. He also used his sample of clusters to trace the Galactic rotation curve and found that their distribution of azimuthal velocities agrees at the solar radius with the ones from HII-regions and molecular clouds derived respectively by Georgelin and Georgelin (1976) and Clemens (1985). Later, Scott et al. (1995) investigated the kinematics of clusters older than 1 Gyr and found them to be consistent with the ones of Hron (1987) but with a much larger dispersion of 29 kms^{-1} instead of 10 kms^{-1} for young OCs. Without any information on the proper motions of their sample of clusters, Scott et al. (1995) already witnessed the increase of Age-Velocity Relation (AVR) with OCs.

In the following years, the publication of the *Hipparcos*, Tycho-2, US Naval Observatory CCD Astrograph Catalog (UCAC) and PPMXL catalogues added 2 more dimensions to the investigation of the kinematical propreties of clusters. We saw in Sect. 1.2.2.2 that clusters could be used to investigate the position of the spiral arms. By adding the full 3D kinematical information, Dias and Lépine (2005) determined the rotation speed of the spiral arms. By integrating the orbits of their young clusters, they found a unique angular velocity for all the explored spiral arms, indicating that spiral arms rotate like rigid bodies.

Proper motion information combined with RVs on large amounts of OCs made a three dimensional investigation of the AVR with star clusters possible. By dividing their sample of 488 clusters into different age bins, Wu et al. (2009) found that the dispersion increased with age for each component of the velocity. This increase of the velocity dispersion with age, already observed with field stars is a probe of the heating of the clusters in the Galactic disc (Wielen, 1977; Binney and Tremaine, 2008). Later, Soubiran et al. (2018a) provided the 6D phase-space information for 861 clusters and also looked at the AVR. They found a significant increase of the vertical velocity dispersion for their oldest population with a velocity dispersion representative of the thin disc.

1.2.2.4 Clusters chemical composition

As explained in Sect. 1.2.1, OCs are chemically homogenous groups of stars. The determination of the chemical composition of a cluster therefore benefits from the same advantages than with other clusters' parameters: as it can be based on multiple stars, the chemical composition of a cluster is much more reliable than for a single star. The chemical signatures of OCs is reminiscent of the chemical abundances of the GMCs from which they formed. As clusters are found at all ages and locations in the Galactic disc, they are key objects to trace its chemical composition and in particular its enrichment in heavy elements through nucleosynthesis and chemical evolution.

In the catalog of clusters from Dias et al. (2002), the OCs metallicity spans the range $-1.0 < [\mathrm{Fe/H}] < 0.5$ dex with $\sim 75\%$ of the clusters having a metallicity comprised in

the range $-0.4 < [Fe/H] < 0.08$ dex. The first study of the disc chemical signature using OCs was conducted by Janes (1979) who found a decrease of the metallicity distribution with an increasing Galactocentric radius. These findings were confirmed by numerous other studies like the review from Friel (1995). Twarog et al. (1997) proposed that the metallicity gradient of OC was not a simple linear function of the Galactocentric radius but was better described by two linear functions with different slopes and a transition near the Solar radius. This seems to be confirmed by the most recent data of large scale spectroscopic surveys like Apache Point Observatory Galactic Evolution Experiment (APOGEE) (Donor et al., 2020).

Precise determinations of OCs abundances also allow to assess the level of chemical homogeneity within the different stars of a cluster (Liu et al., 2016). With the differential analysis of high resolution and high signal to noise spectra of twin stars, Casamiquela et al. (2020) investigated in this direction. They detected a slight level of inhomogeneity (0.02-0.03 dex) in the three investigated clusters: Ruprecht 147, the Hyades and NGC 2632. Although the level of inhomogeneity was too low to draw any conclusion on star formation, some of the tidal tails stars identified by Röser et al. (2019) were discarded based on their chemical signature.

1.2.3 Membership determination

Identifying cluster members requires large scale observational data and the combination of several types of observations. As seen in Sect. 1.2.1, the stars belonging to one cluster are born at the same moment and share the same age and metallicity. Photometric data can therefore be used to draw the HR diagram of the cluster candidates and check if an isochrone can fit the distribution of the stars. However, stars can follow the main sequence of a cluster just by chance alignment. Follow-up spectroscopic observations of the stars can then be performed in order to compare the metallicity or the RV of the stars.

Other properties shared by cluster members can also be used to establish OCs membership lists: the distance and the Galactic velocities of the stars from a same cluster are very similar. With the advent of the large scale astrometric mission *Hipparcos* and the successive publications of the Hipparcos and Tycho-2 catalogues (Perryman et al., 1997; Høg et al., 2000), several clusters were identified through systematic searches designed to track overdensities of stars in the space of position or proper motions complemented with photometric criteria (Platais et al., 1998; Alessi et al., 2003). When full sky proper motion catalogues such as the PPMXL (Roeser et al., 2010) and the fourth UCAC (Zacharias et al., 2013) surveys were released, astrometric data became systematically used in the membership analysis of OCs (Kharchenko et al., 2013; Dias et al., 2014).

Powerful statistical methods able to combine both astrometric and photometric information have also been developped (Sarro et al., 2014; Olivares et al., 2018b). Using the stars magnitudes, colours as well as their proper motions and positions allows these methods to be less biased by the high proper motions uncertainties of faint stars. They also allow a more complete and comprehensive selection of members but are less easy to

adapt to a large number of clusters. With the *Gaia* mission, the hunt for OCs and their members has been brought to a whole new scale. The following of this thesis will focus and bring more details to the *Gaia* related discoveries.

1.3 Open Clusters in the *Gaia* era

1.3.1 *Gaia*

The *Gaia* mission is an European Space Agency (ESA) project designed to map the Milky Way in order to bring light to its structure, formation and evolution (Gaia Collaboration et al., 2016). These being its main scientific objectives, *Gaia* also provides crucial information on star formation and evolution or allow to detect exoplanets, supernovae, brown and white dwarfs. To do so, it measures the astrometry (i.e. the position, parallaxes and proper motions) of more than 1 billion stars with unprecedented accuracies and it reports their optical spectrophotometry as well as the RV for several million stars. This represents astrophysical meaningful information for approximately 1% of the Milky Way stars. While *Hipparcos*, the previous ESA astrometric mission provided a catalogue of ∼100 000 objects with a full astometric solution, *Gaia* multiplied by more than 10 000 this number and improved the astrometric accuracy by a factor of ∼ 100.

The *Gaia* consortium publishes periodically a catalogue gathering all the available data. The second Data Release (DR), published in April 2018 was based on 22 months of observations (Gaia Collaboration et al., 2018b). An overview of its contents is shown in the top panel of Fig. 1.5. A full astometric solution is provided for more than 1 billion stars with a magnitude range of $3 < G < 21$ mag. The uncertainties strongly depend on the magnitude of the stars but are typically smaller than 1 mas in parallax and 1 mas/yr in proper motions even for fainter targets. The RVs provided by *Gaia* (Katz et al., 2019) for more than 7 million stars with a magnitude range of $4 < G < 13$ mag already constitute the biggest RV catalog. The next DR has been splitted into two releases: the early Data Release 3 (eDR3) published in December 2020 and the DR3 expected for the first semester of 2022. As seen in the bottom panel of Fig. 1.5, the number of stars with a full astrometric solution is ∼ 10% higher in the eDR3 compared with the DR2. Being based on 34 months of observations, the accuracies of the measurements of the positions and parallaxes have improved by a factor of 0.8 while the uncertainties on the proper motions have improved by a factor of 0.5 (Lindegren et al., 2021b). The number of sources with G-band, G_{BP} and G_{RP} photometry have also increased by roughly ∼ 10%. In the full DR3, information on variable sources, effective temperatures, extinction, reddening, radius and RV are expected for much more stars than for the DR2. The successive *Gaia* catalogs constitute the biggest and most homogeneous sample up to date in the field of stellar and galactic physics.

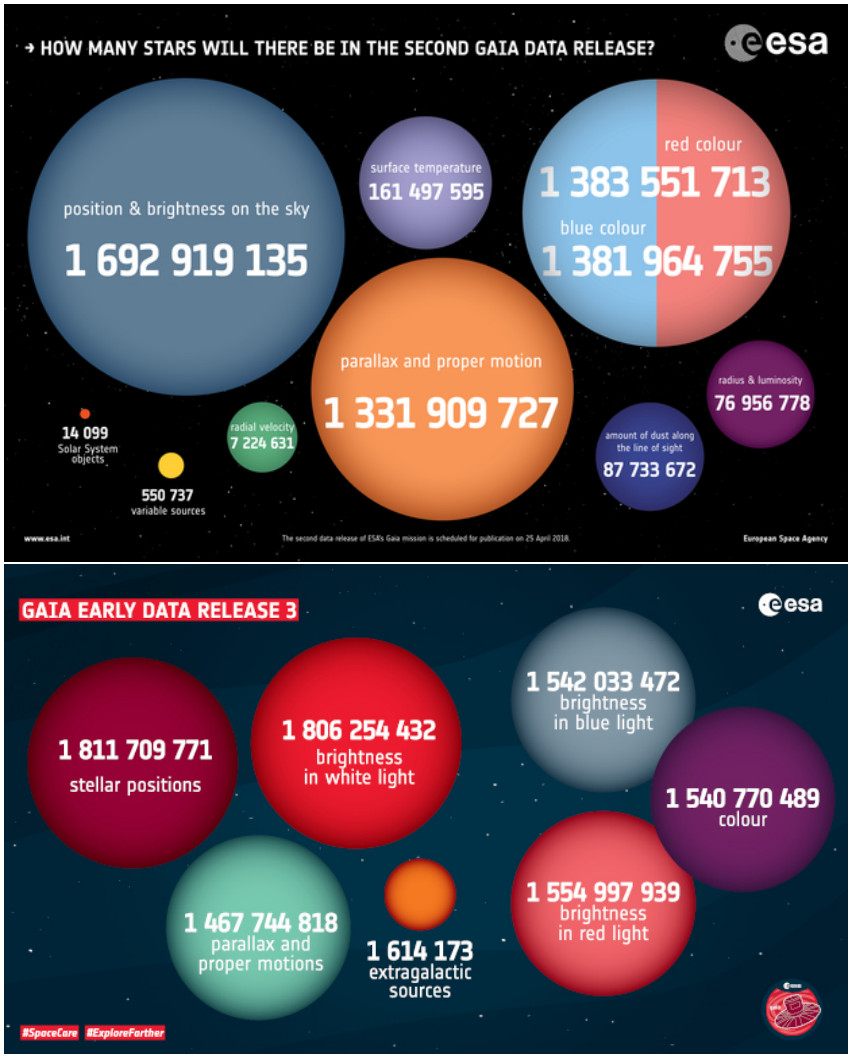

Figure 1.5: Overview of the contents of *Gaia* DR2 (top) and eDR3 (bottom). Credit: ESA

1.3.2 Some results from *Gaia*

Since the release of the DR2, *Gaia* has led to the understanding of the origin of some stars in the Galaxy. For example, Helmi et al. (2018), by looking at the kinematical signatures of stars in a 2.5 kpc bubble around the Sun detected that a significant fraction of them had a retrograde motion. Both photometric and chemical analysis of these stars confirmed that these stars located in the nearby halo could not have been formed in the thick disk but came from a different system which was accreted to the Milky Way. They called this system which merged with our Galaxy some 10 Gyr ago the *Gaia*-Enceladus system. The unprecedented precision and all-sky coverage of *Gaia* data also unveiled previously unseen structures. Prominent thin diagonal ridges appear in the space of azimuthal velocity versus Galactocentric radius (see Fig. 1.6). These diagonal ridges could have multiple origins: they could be the signs of an on-going phase mixing in the Galactic disc or they could be induced by the bar and the spiral arms. Indeed, the bar and the spiral arms create resonant orbital structures in the disc i.e. regions of stable

and unstable orbits. Regions of unstable orbits translate into gaps in the distribution of stars while regions of stable orbits translate into overdensities of stars identifiable as the ridges in Fig. 1.6. Alternatively, these diagonal ridges could be created by a combination of both phase mixing and resonant orbital structures.

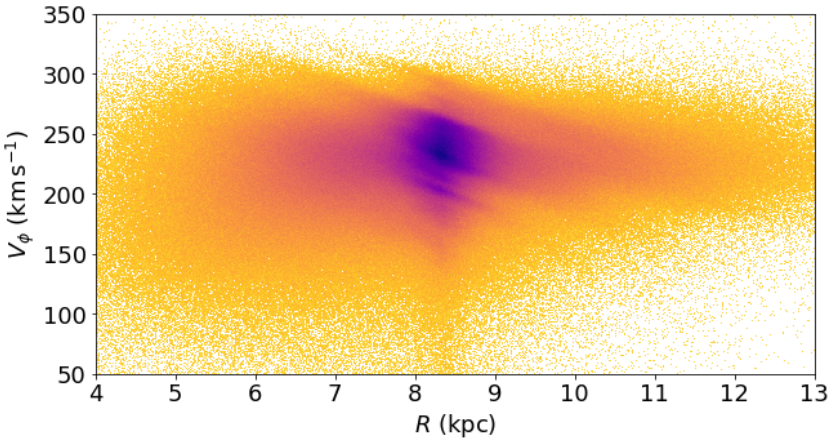

Figure 1.6: Two dimensional histogram of the azimuthal velocities versus Galactocentric radius showed with bins of 1 km s^{-1} in V_ϕ and 0.01 kpc in R. From (Antoja et al., 2018).

In parallel, the view on OCs has been drastically improved since the publication of the *Gaia* DR2. Previous OC catalogs like the ones from Kharchenko et al. (2013) and Dias et al. (2002) were heterogeneous collection of different surveys, each with their own characteristics, bias and uncertainties, making a uniform characterisation of clusters very challenging. The publication of the *Gaia*-DR2 catalogue allowed to revisit the OC census in the Milky Way and to determine their mean parameters in a homogeneous way. Based on the list of more than 3 000 known clusters and candidates, mostly from the two aforementionned catalogues, Cantat-Gaudin et al. (2018a) (CG+18) obtained a list of members and mean parameters for only 1 229 of them by using the clustering algorithm Unsupervised Photometric Membership Assignment in Stellar clusters (UPMASK) (Krone-Martins and Moitinho, 2014). Even though some of the clusters reported pre-*Gaia* were either too embedded in their field or observable only at infrared wavelengths, a large number of objects known as clusters turned out not to be real groups of stars. This highlights, according to the authors, the need to re-examine such objects in the light of the recent *Gaia* catalogues. In addition, they serendipitously discovered 60 OCs in the close vicinity of known ones, showing that the OC census, previously thought to be complete up to 1.8 kpc was far from completion.

Hundreds of new OCs and their members were also identified by Castro-Ginard et al. (2018, 2019, 2020) who developed a machine learning approach to spot over-densities in the five dimensional parameter space of positions, parallaxes and proper motions. These works released a catalogue of more than 600 new open clusters. Cantat-Gaudin et al. (2019b) also looked for overdensities in the direction of the Perseus arm and detected 41 new clusters. Sim et al. (2019) visually inspected stellar distributions in the galactic

coordinates and proper motion space and identified 207 new cluster candidates. Also Liu and Pang (2019) used a friend-of-friend method already widely used in the galaxy cluster community to identify 76 unreported clusters. Kounkel and Covey (2019) applied the unsupervised machine learning algorithm HDBSCAN to identify not only clusters but also moving groups and associations within 1 kpc. More recently Cantat-Gaudin et al. (2020) (CG+20), published a catalogue of 2 017 OCs previously identified by the aforementioned authors and determined in a homogeneous way their memberships, distances and ages for most of them. This constitutes the most complete catalogue of OCs.

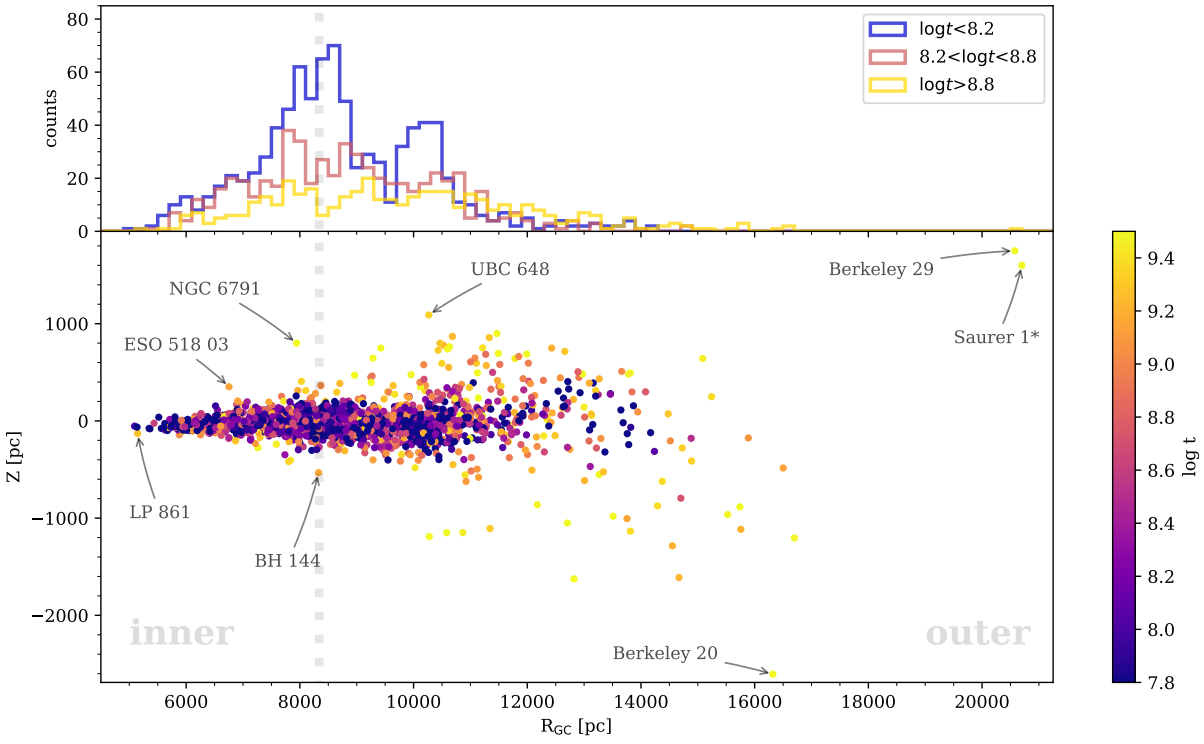

Figure 1.7: Top: galactocentric distribution of the cluster sample from CG+20 in three age bins. Bottom: altitude above the Galactic mid-plane of these clusters as a function of their Galactocentric radius and coloured by age. In both panels the vertical dotted line represents the Galactocentric radius of the Sun. From CG+20.

This homogeneous catalogue allowed the authors to study, among other things, the scale height of the Galactic disc and revisit the results of Bonatto et al. (2006) discussed in Sect. 1.2.2.2 but with an unprecedented amount of clusters with reliable age determinations. Figure 1.7 shows the increase of the altitude of clusters with an increasing age as well as with an increasing Galactocentric radius. They found an altitude of the Sun of 23±3 pc. While Bonatto et al. (2006) could not see an increase of the scale height for very old clusters, CG+20 report that it reaches several hundreds of parsecs.

Castro-Ginard et al. (2021) also revisited the work done by Dias and Lépine (2005) on the spiral arms nature with *Gaia* data. Using the largest sample of clusters with full kinematic information and age determinations (proper motions and ages coming from CG+20 and RV from Tarricq et al., 2021a), they derived different pattern speeds for each

investigated spiral arms. Their findings are in better agreement with both simulations and observational studies based on field stars kinematics as they do not do not find the spiral arms to rotate as rigid bodies. This would imply that rather than being long-lived structures according to density wave origin (Lin and Shu, 1964), spiral arms would be short-lived with a transient behaviour (Toomre, 1964).

1.4 Overview of the thesis

The goal of this thesis is to uncover some of the physical processes at work in our Galaxy, the Milky Way by studying in particular OCs, sometimes called the "building blocks" of galaxies. Many things remain poorly understood in the Milky Way: from the number and nature of the spiral arms to its dynamical history. To unveil some of its properties, theoretical models have to be enriched and better constrained in order for them to be able to fully reproduce the complexity and extremely large variety of effects at play. The observation of OCs is one of the best ways to achieve this, as the determination of the clusters parameters is much easier and accurate than any other object in the Galaxy. The *Gaia* mission with its unprecedent volume of astrometric data is helping to constrain the theoretical models by providing us kinematical information for more than 1 billion stars. In the field of stellar clusters, it allowed to revisit the OC census and to provide new high confidence memberships.

We focused mainly on the investigation of the relation between the morphology and kinematics of OCs with their properties such as their age or their position in the Galaxy. A particular attention was brought on the external regions of the clusters in order to unveil the different mechanisms at play in their disruption.

This manuscript is organized as follows. In Chapter 2, we present the methodology used in order to detect new members at a large distance from the clusters centers. We present the inevitable bias of each method and show the results of this clustering approach. The Chapter 3 is dedicated to the study of the morphology of OCs. We show here a 3-dimensional study of nearby clusters, the limitations to this study due to the parallax uncertainties of *Gaia*. In the last part of this chapter (Sect. 3.5), we present the morphological analysis of the clusters using the lists of members obtained in Chapter 2 and evaluate how their properties correlate with the age and the galactic position of the clusters. In Chapter 4, we explain the different steps in the constitution of our RV catalogue and we use these RV compute 6D galactocentric coordinates for the largest sample of OCs up to date. We then investigate the kinematical properties of our sample of clusters as well as their orbital parameters. Finally, in Chapter 5, we present the main results of this thesis along with its limits and perspectives.

Chapter 2

Clustering: How to identify the members of an Open Cluster?

Contents

2.1 Context: a large variety of clustering methods

As seen in Sect. 1.2.3, the most commonly used methods to identify an OC and its members are aimed at identifying stars sharing similar kinematics and positions. Stars from the same OCs also share the same ages and chemical composition but the determination of these parameters for individual stars requires spectroscopic follow-up on each of the targeted stars and is therefore very time consuming. Even before *Gaia*, the search and identifications of OCs was therefore primarily conducted by looking for overdensities of stars in proper motion or position in the celestial sphere. The two largest catalogues of OCs prior to *Gaia* were respectively compiled this way (Dias et al., 2002; Kharchenko et al., 2013), but with data coming from different heterogeneous surveys (mainly 2MASS, UCAC and PPMXL: Skrutskie et al., 1997; Zacharias et al., 2013; Roeser et al., 2010).

The successive publication of the second *Gaia* data release (DR2, Gaia Collaboration et al., 2018b) and of the early data release 3 (eDR3, Gaia Collaboration et al., 2021), led to what could be called a revolution in the study of OCs. With almost 1.5 billion sources with a full astrometric solution (position, proper motions and parallaxes), the census of OCs as well as their characterization have been drastically improved.

The combination of the striking precision of the *Gaia* astrometric measurements and its all sky coverage allowed to compute new memberships for already known OCs, to identify many more cluster members and to discard some asterisms (see Sect. 1.3.2 for more details). It also allowed the detection of prominent structures around some OCs by several groups. Röser et al. (2019), Meingast and Alves (2019) and Jerabkova et al. (2021) characterized the tidal tails of the Hyades at large spatial scale with different methods. The tidal tails of Coma Berenices and Blanco 1 was discovered respectively by Tang et al. (2019) and Zhang et al. (2020) using the unsupervised clustering algorithm `StarGO` (Yuan et al., 2018). Yeh et al. (2019) discovered the tail of Ruprecht 147 by using a simple selection based on proper motions. The tidal tails of Praesepe were identified using the convergent point method by Röser and Schilbach (2019). To characterize the tail of NGC 2506, Gao (2020) used a principal-component analysis and a Gaussian Mixture Model (GMM). Finally, Bhattacharya et al. (2021) used the membership determination algorithm `ML-MOC` (Agarwal et al., 2021) to identify the tails of NGC 752. Meingast et al. (2021) studied ten nearby (located closer than 500 pc), prominent and young OCs and identified around almost all of them an extended population of stars, referred as a corona. To identify cluster members, they used a two steps method which first filters out stars with discrepant tangential velocities compared to the cluster bulk velocities. They then use the unsupervised clustering algorithm Density-Based Spatial Clustering of Applications with Noise (DBSCAN) on the heliocentric coordinates (X, Y, Z) of the remaining stars to produce their final membership lists.

Theoretically, the shape of OC can be described with a dense core and an outer halo (or corona) having a low density of stars (Artyukhina and Kholopov, 1964). As pointed out by Nilakshi et al. (2002) and more recently by Meingast et al. (2021), halos are much more extended than the cores and they are suspected to comprise a large number of cluster members. While the detections of tidal tails and halos surrounding OCs are

focused on small samples of clusters, we aim at developping a methodology able to detect such structures on a large sample of clusters. In the following, we tested two methods dedicated to this purpose.

2.2 Selection based on the velocities

The members of a given cluster will share similar proper motion and RV. Consequently, the cluster will stand out as an overdensity in the Galactic space velocity components (Johnson and Soderblom, 1987). We used as a reference frame the Cartesian heliocentric velocities (U, V, W) with each axes respectively oriented towards the Galactic centre, the direction of Galactic rotation, and the north Galactic pole. We took advantage of the mean cluster parameters of CG+18 and of the RV and mean Cartesian heliocentric velocities catalogue of clusters computed by Soubiran et al. (2018a). In a first attempt to detect members at large distance from the clusters centers, we tried to identify them by comparing the stars Cartesian heliocentric velocities to the mean values of the clusters, taking advantage of the $\sim$ 7 million RV measurements from *Gaia*-Radial Velocity Spectrometer (RVS) (Katz et al., 2019).

2.2.1 Method

We selected all the 80 clusters from CG+18 closer than 550 pc and with a RV computed in Soubiran et al. (2018a). We used the mean cluster positions they computed to query a cone search of 100 pc radius around the clusters centers in the *Gaia*-DR2 archive. In addition, for being close to the Sun and occupy a wide area in the sky, we performed an additional cut in parallax to constrain the number of stars per request. We only considered the stars having a parallax $D_{cluster} - 200$ pc $< 1/\varpi < D_{cluster} + 200$ pc with $D_{cluster}$ the distance of the cluster and ϖ the parallax of a star. We chose very wides cuts of 200 pc around the mean cluster distance and 100 pc radius around the mean cluster position for the purpose of detecting extended structures around OCs. We also discarded all the stars which did not have a RV available in *Gaia*-DR2 for which we can not calculate the Cartesian heliocentric velocities.

We now need to define a criteria based on the (U, V, W) velocities of stars allowing us to identify cluster members candidates for each cluster. Based on the typical velocity dispersion of a cluster, Meingast and Alves (2019) selected as cluster candidates of the Hyades all sources within a radius of 2.5 $\mathrm{km\,s^{-1}}$ around the cluster mean velocity. We used the same approach and considered as candidate members for each cluster all the stars with heliocentric velocities along the U, V and W axis comprised in a 5 $\mathrm{km\,s^{-1}}$ interval centered on the cluster mean heliocentric velocity computed by Soubiran et al. (2018a). In Tarricq et al. (2021a) where we studied the kinematics of OCs (see Chapter 4 for more details), the median velocity uncertainty of clusters' members measured with *Gaia*-RVS was 1.33 $\mathrm{km\,s^{-1}}$. This value is of the same order of magnitude than the treshold of 5 $\mathrm{km\,s^{-1}}$ chosen to select our candidate members.

2.2.2 Results

For our sample of clusters which contain according to CG+18 30 079 members, we obtained an additional list of 6 434 candidates located at less than 100 pc from the cluster centers and compatible in terms of heliocentric velocities. We show the results of this procedure on four clusters on Fig. 2.1. For all these clusters, we can see that the list of candidates recovered by our selection based on heliocentric velocities (in orange and in green) is very limited in terms of magnitude by the limits of the RVS. The cores of the clusters previously identified by CG+18 are also identified by our selection as we can see an orange concentration of stars in all of the panels showing the astrometric parameters of members. New candidates are also detected at large distance from the clusters centers.

In the case of Ruprecht 147, the new candidates (in green) seems to trace what could be a tidal tail visible in the (α-δ) plane and especially in the ($X-Y$) plane. In the ($X-Y$) plane, this structure is slighlty tilted compared to the the line of sight traced by the members in blue. This elongation is therefore not due to the radial elongation of clusters due to the parallax uncertainties (this elongation will be studied in more details in Sect. 3.2). This structure matches the orientation of the tidal tail identified by Yeh et al. (2019).

For NGC 2632, more than a hundred of candidates are identified. In the equatorial coordinates and in the ($X-Y$) plane, these candidates seem to form a halo around the core of the cluster. On the CMD, we can however see that some contaminants are considered as candidates by our selection as some stars lie below the main sequence. We can see the same trend with the Pleiades with even more contaminants lying below the main sequence and above the binary sequence in the CMD. There is also a halo-like structure in the (α-δ) and in the ($X-Y$) plane. For these two clusters, the field is too dense in the (U,V,W) space to properly discriminate cluster members and field stars.

As for Collinder 135, a larger number of candidates is identified compared to the number of members published by CG+18. The CMD also shows a thick sequence of candidates between $G \sim 14$ mag and $G \sim 10$ mag. Most of these candidates are therefore likely contaminants. On the equatorial coordinates plane, two clumped substructures are easily identifiable around (α-δ) $\sim$ (122, -49) and (115, -38) degrees. On the ($X-Y$) plane, two substructures are again identifiable around $(X,Y) \sim$ (-110, -345) and (-100, -250) pc. We cross-matched the list of members of these substructures with the list of members from CG+18 and Castro-Ginard et al. (2018) to identify these groups and we highlight them on Fig. 2.2. The clump in magenta, located at (115, -38) degrees in the equatorial coordinates plane and at (-110, -345) pc on the ($X-Y$) plane, turns out to be composed of members of NGC 2451B. The red substructure, identified at (122, -49) degrees on the (α-δ) diagram corresponds to the cluster NGC 2547. Finally, the clump located at (-100, -250) pc corresponds to the OC UBC 7, identified by Castro-Ginard et al. (2018). NGC 2451B, NGC 2547 and UBC 7 are young OCs located close to Collinder 135 and with roughly the same age.

Collinder 135, NGC 2451B, NGC 2547 and UBC 7 form a complex structure, embedded in the Vela OB2 region. Beccari et al. (2020) and Cantat-Gaudin et al. (2019a) studied this complex region composed of many young clusters. Beccari et al. (2020)

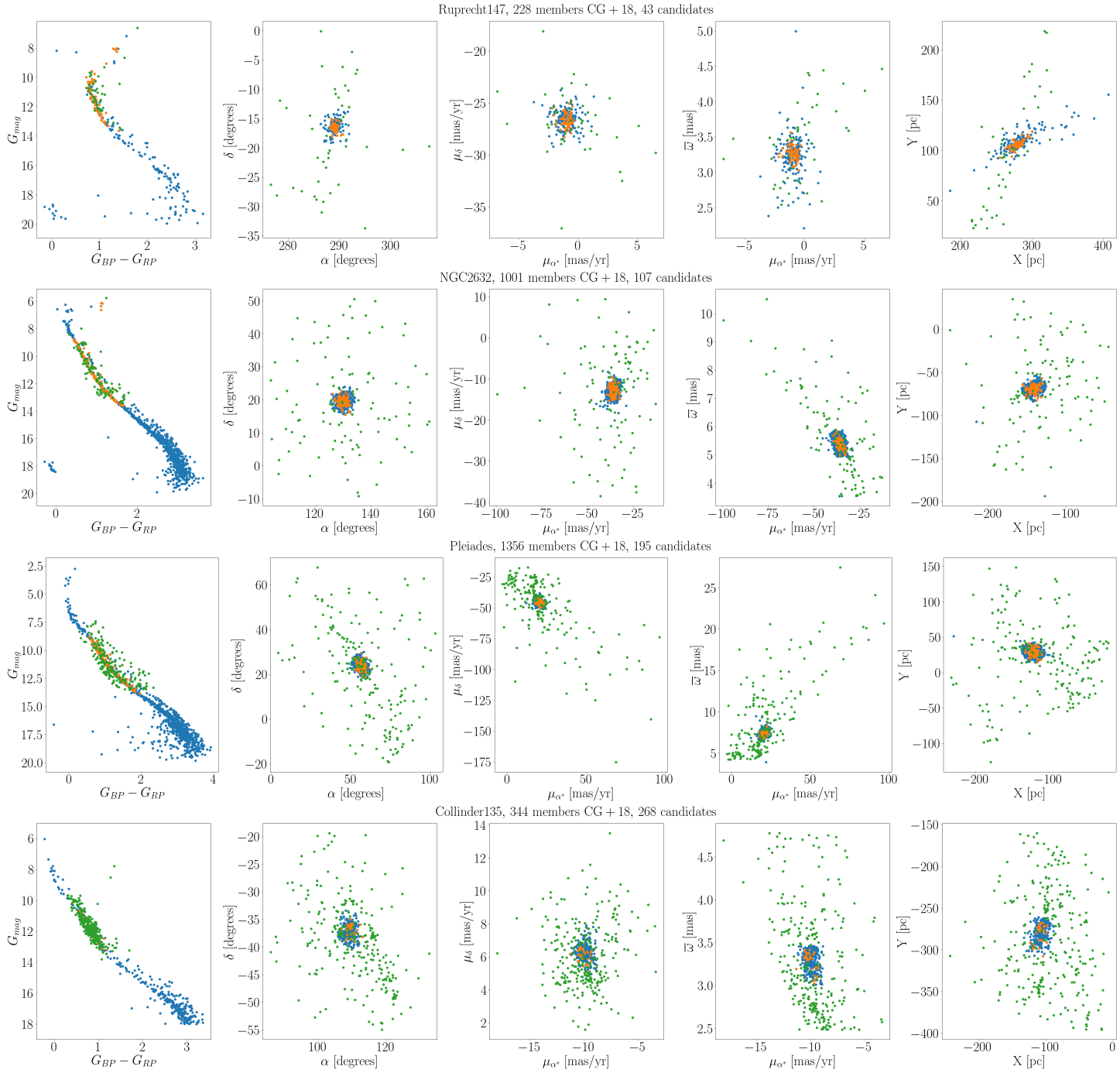

Figure 2.1: Example of the results of our clustering procedure on Ruprecht 147, NGC 2632, the Pleiades and Collinder 135. For each cluster, we show the members recovered by CG+18 (blue), the candidate members recovered by our selection on the heliocentric velocities (green) and the members we have in common in orange. Each panel line stands for one cluster where we show its CMD, the distribution of members in the equatorial coordinate plane, in the proper motion plane, in the proper motion-parallax plane and in the X-Y plane. For each cluster, we indicate the number of members from CG+18 and the number of candidates showed in green.

showed that Collinder 135, NGC 2451B and NGC 2547 belonged to a coeval family of clusters. They identified major filamentary stellar structures connecting the clusters like a bridge that they interpret as imprints of the star formation process. Also Kovaleva et al. (2020) investigated the hypothesis that Collinder 135 and UBC 7 form a physical pair of clusters. By integrating their orbits, they showed that they were much closer in the past than their current separation of 24 pc.

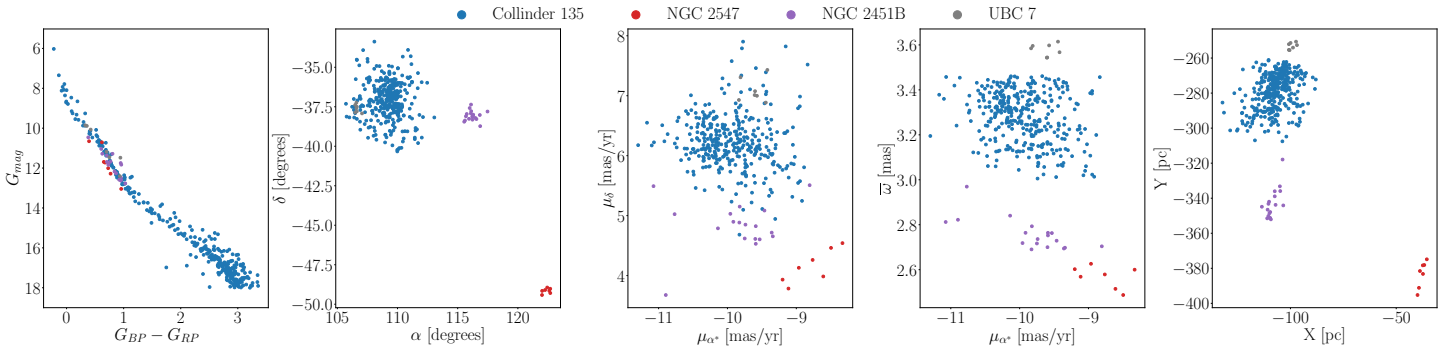

Figure 2.2: Same as Fig. 2.1 for Collinder 135 but with only the members of Collinder 135 from CG+18 and the identified substructures highlighted in colors.

2.2.3 Discussion

The assessment of members by a simple selection on the three dimensional velocities of stars compared to the mean velocity of the clusters is a perfectible method. The 2.5 $\mathrm{km\,s^{-1}}$ treshold has been chosen by Meingast and Alves (2019) and later used by Casamiquela et al. (2020) based on the typical velocity dispersion of young OB associations which is $\sim$ 1.5 $\mathrm{km\,s^{-1}}$ (Preibisch and Mamajek, 2008). As we are studying more evolved structures than OB associations, increasing this 1.5 $\mathrm{km\,s^{-1}}$ treshold seemed relevant but we obtained a satisfying list of candidates only for Ruprecht 147, the oldest cluster in our sample. Ruprecht 147 has a peculiar Cartesian heliocentric velocity which stands out from field stars, making the identification of its members easier. For NGC 2632, the Pleiades and Collinder 135, this method is unfit to separate field stars and clusters members. Many contaminants are considered as candidates for these three clusters and as our procedure does not provide a membership probability, there is no way to distinguish between the contaminants and the actual members. For young clusters like Collinder 135 which are still embedded in their parental structure, such a selection is unreliable and is unable to separate the different clusters located in its neighborhood. This shows that to select members based on their velocities, the value of the treshold around the mean Cartesian heliocentric velocity of a cluster depends of the age and environment of the cluster. It is therefore hard to generalize this method to a large amount of OCs. Besides, this selection is limited on the $\sim$ 7 million stars part of the RVS out of the 1.3 billion stars with proper motions and parallaxes in the *Gaia* DR2.

In some cases, the membership of some stars in a cluster can also either be confirmed or denied by comparing their chemical composition to the chemical signature of the cluster. We conducted such analysis and explain our strategy in details in Appendix A.

2.3 A machine learning approach with HDBSCAN[1]

As seen in the previous section, selecting cluster members only based on a comparison between the stars and the cluster three dimensional Cartesian heliocentric velocities is not satisfying for several reasons. The detection of members at large distances from cluster centers requires more sophisticated methods. To compute membership lists for local OCs at large distances from the center, we developped a method based on the unsupervised density based clustering algorithm: HDBSCAN.

2.3.1 Data

We selected all the OCs from CG+20 closer than 1.5 kpc from the Sun and older than 50 Myr. We found these cuts to be the best compromise in order to discard young clusters still embedded in star forming regions (like Collinder 135) with complex spatial structure and distant clusters with less reliable astrometric parameters. These cuts give optimal results of our clustering procedure (see Sect. 2.3.2). This leaves us with 467 clusters. The CG+20 catalogue includes the most recent improvements of the OC census based on *Gaia* DR2, previously reported in Cantat-Gaudin et al. (2020, 2018a) and Castro-Ginard et al. (2018, 2019, 2020). Here we take advantage of the exquisite astrometric precision of *Gaia* EDR3 to re-visit the memberships of the selected clusters on a wide area around their center. We used the mean proper motions, parallaxes and positions calculated by CG+20 to query the *Gaia* archive for each OC as follows:

- we queried cone of 50 pc radius around the center of each cluster,
- we used the cluster dispersion in proper motion from CG+20 to perform cuts at 10 σ in proper motion to discard very discrepant stars and help the clustering algorithm,
- we considered only stars with $G < 18$ mag and with a Renormalized Unit Weight Error (RUWE) lower than 1.4, following the recommendation of Fabricius et al. (2021).
- for clusters closer than 500 pc, which span a very wide area on the sky, we applied an additional cut in parallax in order to limit the number of stars in the query. Based on $\varpi_{cluster}$, the mean parallax of CG+20, we left a margin of 200 pc so that all the stars with parallaxes ϖ verifying the following relation were selected: $1/\varpi_{cluster} - 200pc < 1/\varpi < 1/\varpi_{cluster} + 200pc$.

For each cluster, these cuts allowed us to discard a significant amount of stars having astrometric measurements inconsistent with the cluster's mean astrometric parameters.

[1]This section is based on Tarricq et al. (2021b). The work detailed in this section has been conducted in collaboration with several co-authors

2.3.2 Clustering

We used the clustering algorithm Hierarchical Density-Based Spatial Clustering of Applications with Noise (Campello et al., 2013) using its recent python implementation (McInnes et al., 2017) to perform our membership study. It aims at improving the performance of the widely used density-based algorithm DBSCAN (Ester et al., 1996), which was successfully applied to the search of OCs by Castro-Ginard et al. (2018, 2019, 2020). One of the main advantages of HDBSCAN over DBSCAN is that it is able to detect overdensities of varying density in a dataset. To do so, HDBSCAN adds a hierarchical approach to DBSCAN. To detect a cluster, DBSCAN draws hyperspheres of radii ε around each star and considers as a cluster the points inside a hypersphere containing more than `minPts`. In other words, a cluster for DBSCAN is defined as the points within an overdensity which is more populated than the chosen parameter `minPts`. We refer the reader to Castro-Ginard et al. (2018) for a detailed explanation of DBSCAN. HDBSCAN does not depend on the radius ε of a hypersphere as it scans all values of ε and uses them to build a hierarchical tree merging these different results. Clusters are defined by two parameters, the parameter `min_cluster_size` which is equivalent to the parameter `minPts` of DBSCAN and the parameter `min_samples` which sets how conservative the algorithm is. Higher values of `min_samples` will discard the clusters with the lowest contrast with respect to the background and consider them as noise even if their number of members is higher than `min_cluster_size`. All clusters containing less stars than `min_cluster_size` will automatically be classified as noise. Once the hierarchical tree of the dataset has been built, HDBSCAN offers two options to select the clusters: the Excess of Mass (EoM) clustering and the leaf clustering. We selected the clusters according to the leaf method which chooses the clusters located at the lowest level of the tree. As noted by Hunt and Reffert (2021), the leaf clustering method almost always performed better in the identification of OCs and we therefore adopted this method.

We ran HDBSCAN on each dataset resulting from the *Gaia* EDR3 query described in Sect. 2.3.1. Following Kounkel and Covey (2019) we chose as initial parameters `min_cluster_size`=40 and `min_samples`=25. We tested this analysis with different input parameters on a subset of clusters representative of our sample and found this choice as the best compromise. If HDBSCAN did not identify a cluster, we lowered iteratively these parameters and try again the clustering. We ran HDBSCAN on the three dimension space of the parallax and proper motions, (μ_{α^*}, μ_δ, ϖ), but not on the sky coordinates to avoid penalizing the stars in the cluster outskirts. For every cluster, we ran HDBSCAN 100 times, each time with a new sample of individual (μ_{α^*}, μ_δ, ϖ) randomly generated from their uncertainties taking into account the correlations between them, as done by Cantat-Gaudin et al. (2018a). This process allowed us to compute membership probabilities: the membership probability of a star corresponds to the frequency with which it was considered as a member by HDBSCAN.

In some cases HDBSCAN identified several clusters, either statistical clusters, asterisms or other physical groups located in the same field. In this case, we took advantage of the previous information that we have for that particular cluster, and we only considered the group identified by HDBSCAN with mean proper motions and parallax the closest

to the value computed by CG+20. This is for instance the case of the cluster NGC 7063 which has two close neighbors: ASCC 113 and UPK 113. With this additional filter we could identify the three clusters.

Visual inspection of the results obtained for the 467 clusters showed a successful membership list for 389 OCs. It performed poorly on some particular cases that we discarded: (1) the clusters with too few stars (less than 30), (2) the clusters with neighbours which overlap in the same field of the query and are also close in the parallax-proper motion space. The latter situation happened in particular around star forming regions and results from the large radius of our query. Even though we discarded most of the star forming regions by considering only clusters older than 50 Myr old, some of the youngest clusters of our sample are still close from their birth location.

2.3.3 Choice of the probability cutoff

We also noticed during this visual inspection that the probability distribution of the members depends on the Galactic coordinates of the cluster and on the density of field stars surrounding the cluster. As a first approach, we represented the histograms of the membership probabilities for each cluster and we chose to consider as the membership probability threshold above which stars are considered as cluster members the probability for which the histogram would reach its minimum. A visual inspection of the members showed that for some clusters a higher cut would be more appropriate while for other ones a lower cut would be better (see Fig B.1 to B.20). For the sake of clarity, we decided to use 0.5 as a single probability cutoff for the membership list which we considered as the best compromise for the large variety of clusters in our sample.

2.3.4 New memberships

For the majority of the 389 OCs, we find a number of members significantly larger than CG+20 as seen in Fig. 2.3 (we recall that CG+20 published the list of members with a membership probability higher than 0.7). A striking case is UBC 480 with 13 members in CG+20 and 470 in our analysis. Our study increases the number of cluster members by a factor of 36. The number of members of NGC 6716 has also been increased by more then 850%, rising from 70 to 568. NGC 6716 was identified by Grice and Dawson (1990) and CG+20 as a sparsely populated cluster but we found it as a quite populated cluster with a dense core and a large halo (or corona).

Figure 2.4 shows an example of the results of our clustering procedure on two well known clusters: Blanco 1 and NGC 2682. Unlike on Fig. 2.1, the members showed here match very well the cluster sequence in the CMD[2]. We can clearly see that in both cases, we recover almost all members identified by CG+20 and that we extend the memberships way further than the cores, reaching the limits of our search radius. For NGC 2682 we identify many halo stars, confirming previous findings of Carrera et al. (2019a). In the case of Blanco 1, a tidal tail already reported by Zhang et al. (2020) is also detected

[2]the results of the clustering procedure are shown for Ruprecht 147 and the Pleiades in Fig. B.15 and B.2 respectively

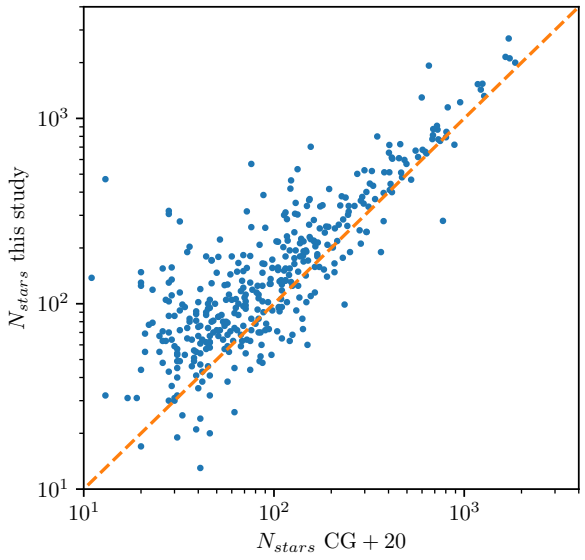

Figure 2.3: Comparison between the number of stars in CG+20 and in this study. The dashed line shows the identity relationship.

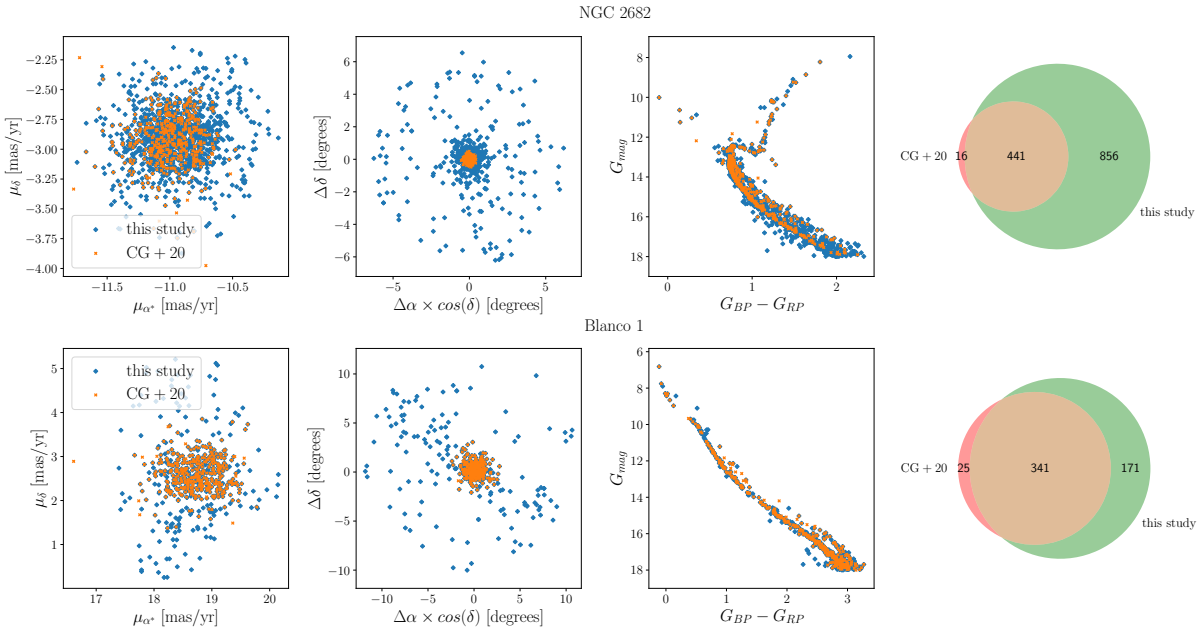

Figure 2.4: Example of the results of our clustering procedure on NGC 2682 (upper panels) and Blanco 1 (lower panels). For each cluster, the three scatter plots represent (from left to right) a comparison between the members from CG+20 (in orange) and ours (in blue) in the proper motion space, in the equatorial coordinate space, and in the color magnitude diagram. The rightmost panel shows for both clusters a Venn diagram comparing the number of members in both studies with members from CG+20 in red, our members in green and the overlap in orange.

by our method in addition to halo stars. We detect vast coronae around a significant number of clusters, similarly to Meingast et al. (2021) who performed a 3D analysis on ten prominent and nearby clusters. We detect similar structures in the five clusters we

have in common. These coronae extend until the edge of our search radius, suggesting that they are even larger. We detected these coronae even for distant clusters such as NGC 2477 located at 1415 pc. We show more complete figures for several clusters in Appendix B.

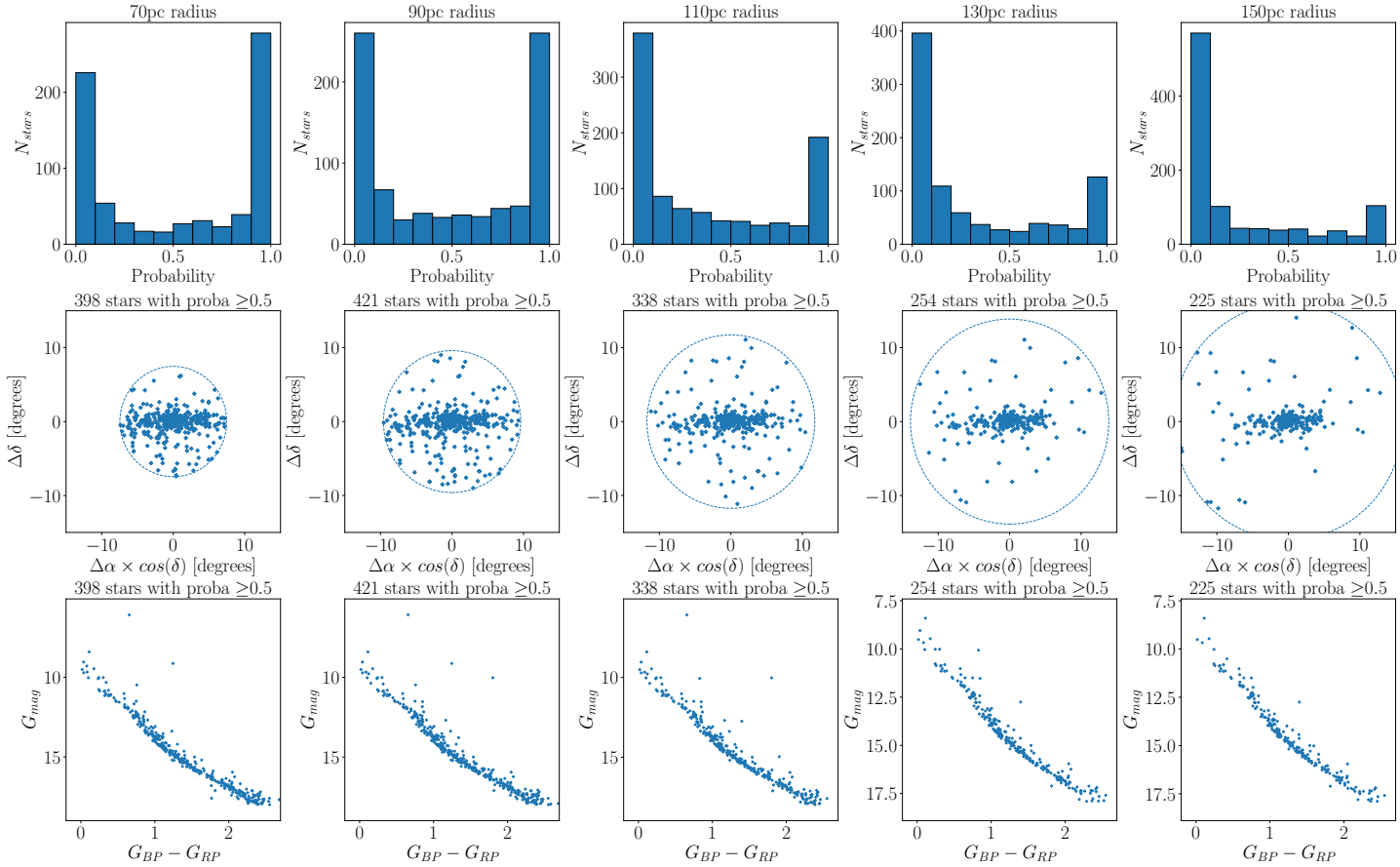

Figure 2.5: Probability distribution (top row), distribution of the recovered members of COIN-Gaia 13 on the celestial sphere (middle row) and color magnitude diagram of the recovered members (bottom row) for concentric search radius going from 70 to 150 pc by steps of 20 pc. On each of the middle panel we show the edge of the search radius in degrees with the blue dotted circle and we indicated the number of stars passing our probability cutoff.

In order to test the hypothesis that the coronae can extend to very large distances from the core and that recovered members are limited by the search radii, we inspected the results of our method on the recently discovered OC COIN-Gaia 13 for different search radii. We queried the *Gaia* archive the same way as described in Sect. 2.3.1 but we did 10 concentric cone searches of increasing radius with steps of 10 pc. We show the resulting probability distributions and members recovered up to a radius of 150 pc on Fig. 2.5. We can see that we keep identifying members until the edge of the cone search even at a radius of 150 pc. However, the number of members stopped increasing after reaching a radius of 80 pc which we interpret as the limit of the corona. This also shows that for very extended fields, a probability cutoff of 0.5 might not be the best compromise anymore.

2.3.5 Center determination

We computed the mean position and parallax of each OC by finding the maximum density point through a kernel density estimation. The mean proper motions are computed differently than the mean positions. As shown on Fig. 2.6, the proper motion distributions are too flat to properly assume the maximum density point. Therefore we used the same method as CG+20: we calculated the median value after removing outliers away from the median by more than three median absolute deviations (MAD).

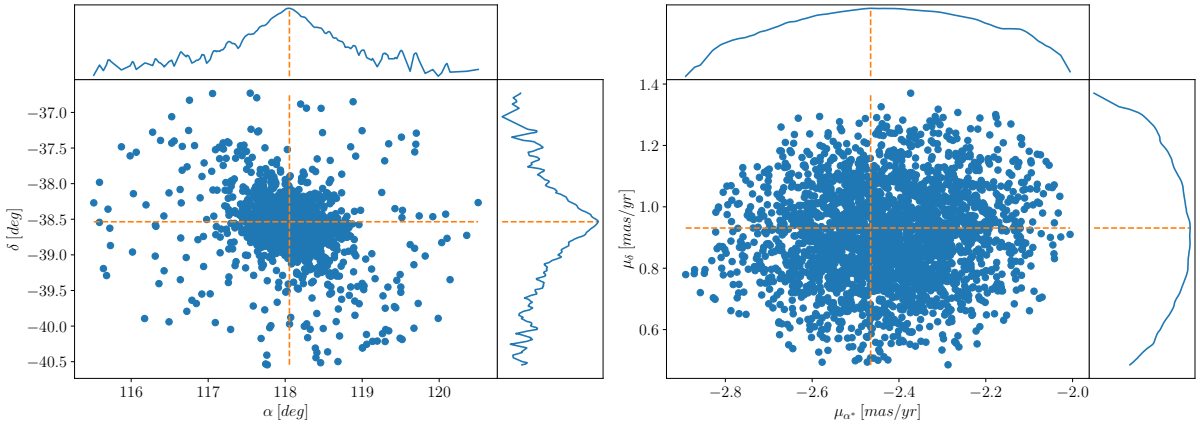

Figure 2.6: Distribution of the recovered members of NGC 2477 in the equatorial coordinate space (left) and in the proper motion space(right). The one dimensional kernel density estimation of α and μ_{α^*} are shown on the top part of each panel and on the right part of each panel is shown the one dimensional kernel density estimation of δ and μ_δ. The maximum density point is shown with the orange dashed line.

The mean astrometric parameters of our OCs sample and the comparison to those of CG+20 are presented in Fig. 2.7. The mean of the residuals of the comparison to CG+20 is well centered on zero for both the positions and proper motions. Inevitably, for some clusters the members of CG+20 and ours are different: either some members were not retrieved by HDBSCAN, or we have many more members now (which represent the majority of cases). This creates significant differences in the mean centers and proper motions of some clusters compared to CG+20, especially for the low populated ones. However for parallaxes, the distribution of the residuals shows a negligible offset of -0.008 mas and a MAD of 0.015 mas showing the agreement between our values and those of CG+20.

2.4 Conclusion

Many statistical methods dedicated to the detection of clusters tidal tails were applied since the publication of *Gaia* DR2. As a first approach, we tried to identify cluster members based on 3 components of the galactic space velocity with respect to the Sun. Despite the inherent flaws of this approach such as the limit in magnitude and the incompleteness of the RVS, this approach still allowed us to identify the tidal tail of Ruprecht 147, a

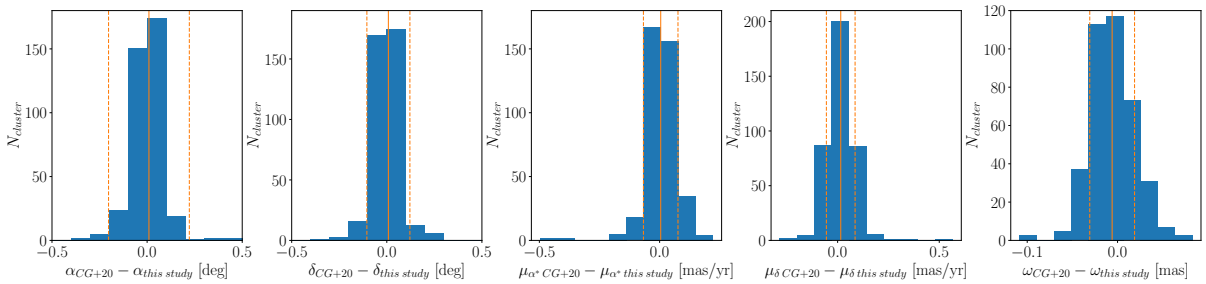

Figure 2.7: Distribution of the residuals of the mean cluster parameters of CG+20 and the ones calculated in this study. The orange solid line represents the mean of the distribution and the orange dashed lines show the 1 σ standard deviation to the mean. For clarity, the offset between the mean positions calculated here and the previous known mean positions are only shown in the range (-0.5, 0.5) degrees, even if nine and five OCs lie beyond this limit for respectively the right ascension and the declination.

detection confirmed by Yeh et al. (2019). To improve the detection of members in the outskirts of clusters, we developed a methodology able to identify members in the peripheral area of OCs, up to 50 pc from their center. The method is based on the unsupervised clustering algorithm HDBSCAN which can detect overdensities in the astrometric space $(\mu_{\alpha^*}, \mu_\delta, \varpi)$ even in datasets of varying density. We applied this method on 467 OCs from CG+20 closer than 1.5 kpc and older than 50 Myr. We report memberships for 389 OCs, the 78 remaining ones being too embedded in their field for our method to properly disentangle them from their neighbors. For the vast majority of clusters, we identify many more members than previously known.

We identify vast coronae around almost all the clusters, reaching in most cases the maximum radius of the investigated area. We also identify tidal tails for $\sim$ 71 OCs. Previous detections of corona or tidal tails of OCs were focused on smaller samples of nearby clusters. Since we worked on large datasets which include the 50 pc sphere surrounding each cluster, we were able to perform a membership analysis at large radius, even for OCs at large heliocentric distances and to multiply by more than 4 the number of identified clusters with a tidal tail. The scientific exploitation of this catalogue of members is explained in Sect. 3.5.

Chapter 3

Structural parameters of Open Clusters

Contents

3.1 Context

OCs are essential objects to better understand the evolution of the stellar disc of the Milky Way. Most stars of the disc are believed to be born in OCs (Lada and Lada, 2003) which dissipate into the field due to relaxation-driven mass loss or tidal perturbations, as recently reviewed by Krumholz et al. (2019). The morphology of OCs is directly related to these processes. OCs have first to survive an initial gas expulsion (Baumgardt and Kroupa, 2007) following the formation of their first stars. Then, they experience a violent phase of relaxation during which stars can be expelled and form tail-like structures depending on star formation efficiency (Dinnbier and Kroupa, 2020b) and the timescales of gas expulsion (Dinnbier and Kroupa, 2020a), among other processes. In addition, young OCs may keep in their morphology the imprint of substructures from their parent molecular clouds (Alves et al., 2020). Recently, the hierarchical formation scenario has been proposed (McMillan et al., 2007) in order to explain the evidence of mass segregation in young clusters such as the Orion Nebula Cluster (Hillenbrand and Hartmann, 1998). Standard dynamical evolution being unable to explain such levels of mass segregation in young clusters (Bonnell and Davies, 1998), this scenario postulates that stars form in small clumps which later merge to form larger mass segregated systems. On the other hand, older clusters are believed to be governed by both internal and external effects. Equipartition of kinetic energy via two-body relaxation has a direct consequence on the distribution of stars within a cluster. Massive stars within a cluster move towards its center whereas lighter stars move towards its outskirts in a mass segregation process (Mathieu, 1984; Kroupa, 1995; de La Fuente Marcos, 1996). However, no direct relation between mass segregation and the ages of cluster has been found so far. At the same time, gravitational perturbations by giant molecular clouds, tidal stripping due to the galactic potential or spiral arm shocks are believed to perturb the cohesion of star clusters and to shape escaping stars into "S-shaped" tidal structures (Küpper et al., 2008; Dinnbier and Kroupa, 2020a). Eventually, at their final stages, clusters may disintegrate (Lamers and Gieles, 2006) while their members mix into the galactic field. Studying in detail the spatial distribution of members in OCs of various ages and in different environments can shed light on all these processes. In particular structural parameters such as the size of the core, the presence of a halo or a tidal tail in the peripheral region and the degree of mass segregation can bring new constraints to theoretical models.

The most complete and enlightening studies of the morphology of OCs are those conducted in 3D. However, their major drawback is that a 3D study requires to convert parallaxes into distances which is not a trivial transformation as we need to take the uncertainties into account. As showed (among others authors) by Bailer-Jones (2015), it requires the use of Bayesian inference and the choice of a prior, prior which depends on the aim of the study. Indeed, assuming that the Probability Density Function (PDF) of a parallax measurement is normally distributed, its inversion will result in an asymmetric distribution in the space of distances, with a longer tail towards the larger distances. A simple inversion of the parallax of the members of a cluster therefore translates into an elongated shape of the clusters along the line of sight. This radial distortion is larger for

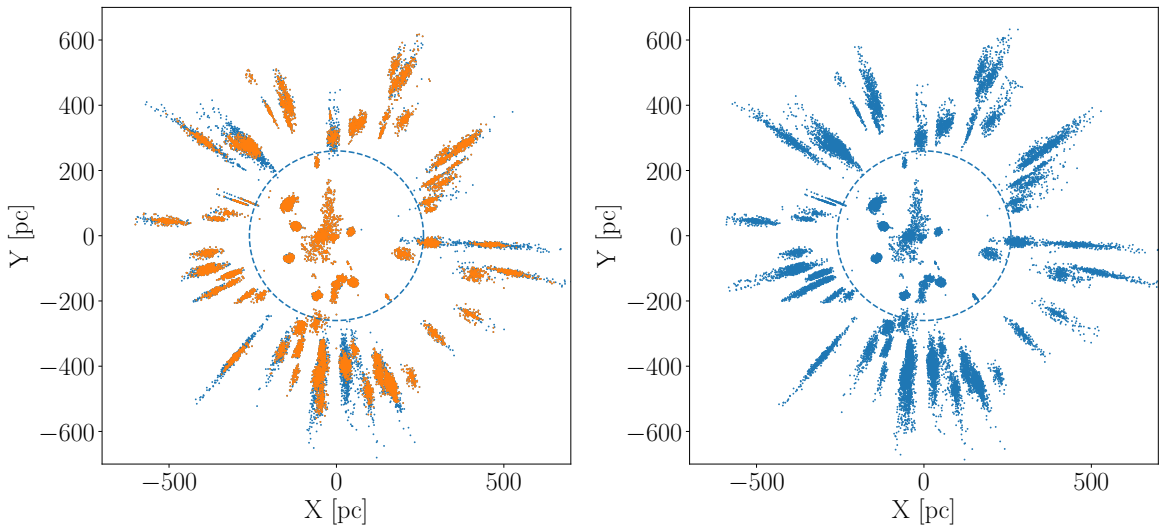

Figure 3.1: Distribution of the members of OCs closer than 550 pc in the X-Y plane, centered on the Sun. On the left panel, the positions in the X-Y plane have been calculated with a simple inversion of the parallax from *Gaia*-DR2. In blue, we show all the stars and in orange, we performed a selection on the quality of the parallax keeping only the stars with $\varpi/\sigma_{\varpi} > 5$. On the right panel, we show the same distributions but the positions have been calculated with the distances from Bailer-Jones et al. (2018) (see Sect. 3.3.1). In both panels, the blue dotted circle represents a limit of 250 pc around the Sun.

sources with large relative errors σ_{ϖ}/ϖ and is unnoticeable when the parallax relative errors become negligible. Even if the uncertainties in *Gaia*-EDR3 have significantly improved compared to *Gaia*-DR2 (Lindegren et al., 2021a), they still translate into an elongated shape of the clusters along the line of sight. Consequently 3D studies are limited to very nearby (<500pc) OCs according to Piecka and Paunzen (2021). In order to study systematically the morphology of clusters further than 500 pc from the Sun it is more efficient to work in 2D.

3.2 Distances of stars with *Gaia*

Having started this PhD a few months after the release of *Gaia* DR2 (Gaia Collaboration et al., 2018b) our first goal was to take advantage on both the extraordinary precisions and volume of the astrometric measurements of *Gaia* DR2. The new memberships determination by CG+18, Gaia Collaboration et al. (2018a) and Röser et al. (2019) using these data allowed us to study the three dimensional morphology of OCs of the solar neighborhood. Even if the OC census of the solar neighborhood could not be considered as complete, as pointed out by Piskunov et al. (2018), *Gaia* DR2 provided us the best overview of the OC population.

In order to measure the shape of OCs, we aimed at fitting a Gaussian on the three dimensional spatial distribution of the OCs members. We aimed at doing that on each OC closer than 550 pc. We chose 550 pc as a limit for two reasons: there are 84 clusters

spanning a large range of properties lying below this limit. This allows us to conduct a statistical study of the OCs shapes depending on their age and environment. In addition, we thought the parallaxes below this limit would be precise enough to be able to conduct a tree dimensionnal study[1].

We represented in the left panel of Fig. 3.1, the distribution in the $(X - Y)$ plane of OCs members identified at this moment by CG+18, Gaia Collaboration et al. (2018a) and Röser et al. (2019). For the members from CG+18, we represented only the ones with a membership probability higher than 0.5. We can easily see a strong radial extension of the OCs below the limit of 250 pc highlighted by the blue dotted circle. This radial extension comes from the large relative errors of distant sources and persists even when we only select the stars with a relative parallax error below 20% (as shown with the orange points in the left panel of Fig. 3.1). We can note several things from this figure:

- Discarding the stars with large relative parallax error does not affect the shapes of the clusters located closer than 250 pc.
- Discarding these stars mainly affects the most distant clusters, reducing their distortion along the line of sight,
- The clusters located further than 250 pc show an unrealistic elongation compared to the ones located closer than 250 pc even when considering only the members with the best parallaxes.

Figure 3.1 illustrates the problem raised in Sect. 3.1 regarding the calculation of the distances. A simple inversion of the parallax produces a strong distorsion of the radial distribution of clusters members which is biasing a 3D analysis of these clusters. In the following, when the distribution of members is unrealistically extended in the direction of the line of sight, we will describe it as distorded, in opposition with the natural elongation of a cluster like the Hyades, for which the tidal tail is identified by Röser et al. (2019).

3.3 Distance inference through bayesian statistics

As seen in Fig. 3.1, calculating a distance from a parallax is not as trivial as it seems. In most of the cases, and especially when the fractional parallax error is larger than 20%, inverting the parallax fails to recover the true distance of a star. It also fails for negative parallaxes and it is unable to give a satisfying value of the uncertainty on the distance (Bailer-Jones, 2015; Luri et al., 2018). Knowing the parallax of a star, the most physically meaningful way to infer its distance and its distance uncertainty requires to do a probabilistic analysis by specifying a likelihood and a prior.

To properly infer the distance r of a star given a parallax ϖ and its uncertainty σ_ϖ

[1]This work was conducted at the very beginning of this thesis, before the publication of the work by Piecka and Paunzen (2021). This explains the slight discrepancy between the 550 pc limit chosen here and the 500 pc limit they recommend to conduct a three dimensional study.

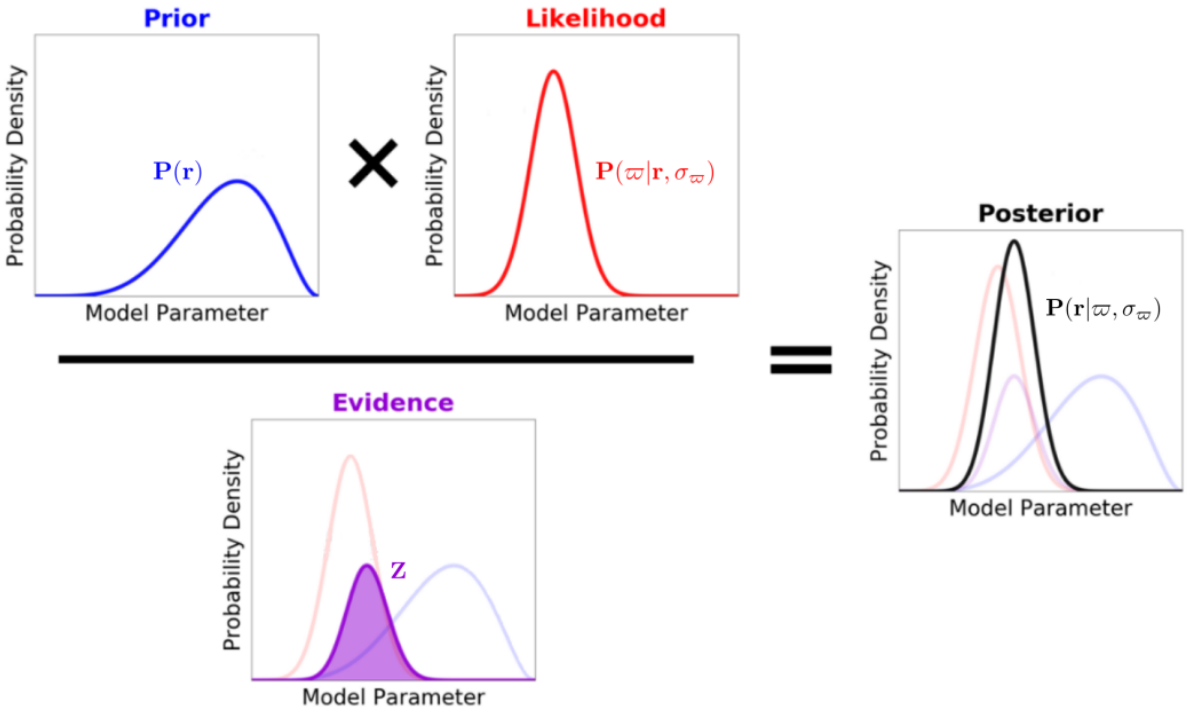

Figure 3.2: Illustration of Bayes' theorem. The posterior distribution (in black) $P(r|\varpi, \sigma_\varpi)$ is defined as the product of the prior distribution (in blue) $P(r)$ with the likelihood of the model (in red) $P(\varpi|r, \sigma_\varpi)$ normalized by the evidence (in purple) Z. From Speagle (2019).

we can then use the Bayes theorem wich is defined in our case as:

$$P(r|\varpi, \sigma_\varpi) = \frac{1}{Z} P(\varpi|r, \sigma_\varpi) P(r), \tag{3.1}$$

with Z the normalization factor

$$Z = \int_0^\infty P(\varpi|r, \sigma_\varpi) P(r)\, dr. \tag{3.2}$$

$P(r)$ is refered to as the prior. It is built by incorporating any a priori knowledege on the properties of the expected distance distribution of the stars. The most obvious property to include in the prior is the positivity of the distance distributions (whereas a parallax can be negative and still contains information). The term $P(\varpi|r, \sigma_\varpi)$ is the likelihood (or the measurement model). It gives the PDF of a star to have a parallax ϖ knowing its distance r and its uncertainty σ_ϖ. Finally, the term $P(r|\varpi, \sigma_\varpi)$ is the posterior distribution. The Bayes theorem is illustrated on Fig. 3.2. Basically, the Bayes' theorem gives us the probability of a star to be at the distance r (which we do not know a priori) given its measured parallax and parallax uncertainty by transforming the expression of the likelihood which represents the probability of a star to have a measured parallax given its distance.

However the choice of a prior needs to be done with care, as it can highly influence the results of any bayesian approach. For example, if the prior is more peaked than the

likelihood, we will end up with a posterior distribution looking like the prior.

3.3.1 With an exponentially decreasing space density prior

Shortly after the publication of the *Gaia* DR2, Bailer-Jones et al. (2018) published a catalog where they recalculated the distances of all 1.33 billion stars with parallaxes published in the *Gaia* second data release. To do so, they used the work of Lindegren et al. (2018) who checked that the standardized astrometric measurements from *Gaia* were normally distributed.

The authors could then define the likelihood as based on the fact that the measured parallax ϖ of a star was normally distributed with a standard deviation σ_ϖ and a mean which is the unknown true parallax defined as $1/r$, r being the true distance of the star:

$$P(\varpi|r,\sigma_\varpi) = \frac{1}{\sqrt{2\pi}\sigma_\varpi} exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi - \frac{1}{r}\right)^2\right]. \tag{3.3}$$

Bailer-Jones (2015) and Astraatmadja and Bailer-Jones (2016) investigated several priors and concluded in favor of a simple Exponentially Decreasing Space Density (EDSD) prior as it was giving the best performances among the several priors they tested. It was also less computationnaly expensive that a more complex prior such as their Milky Way prior. The EDSD prior is some kind of a compromise between the Milky Way prior and the isotopic prior which is the simpler prior they used. It is defined as :

$$P(r|L) = \begin{cases} \frac{1}{2L^3} r^2 e^{-r/L} & \text{if r} > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3.4}$$

with L>0 the scale length of the stellar disc of the Milky Way.

We checked if these new and more reliable distance estimations were good enough for us to study the three dimensional spatial distribution of nearby clusters. We show in the right panel of Fig. 3.1 the same distribution as in the left panel but with the distances published by Bailer-Jones et al. (2018). We did not perform any cuts based on the quality of the parallaxes as the bayesian inference of the distances is mainly dedicated to improve the distance estimations of stars with a large relative parallax error. It is obvious from this figure that the OCs located further than 250 pc are still distorded and that the bayesian distances have not been able to solve this problem. This is in agreement with Piecka and Paunzen (2021) who established that for OCs within 2 kpc, the bayesian approach of Bailer-Jones et al. (2018) gave similar results to the inverse-parallax method regarding their structural parameters.

3.3.2 With *Kalkayotl*

The prior by Bailer-Jones et al. (2018) has proven to be useful for field stars and the study of the Galaxy as a whole (e.g. to compute an extinction map or study the metallicity-age relation of the Milky Way disc: Chen et al., 2019; Feuillet et al., 2019). However, as noted by Carrera et al. (2019b), the usage of a cluster oriented prior is much more efficient to

recover the three dimensional spatial distribution of a cluster than the EDSD prior. This is precisely the aim of Olivares et al. (2020)[2] who published a free and easy to use code: *Kalkayotl*. It is designed to infer the distances and sizes of OCs at the same time with the individual distances of OCs' members. To do so, it uses cluster oriented priors assuming that clusters' members share the same distribution of their astrophysical parameters and in particular their distances.

Two types of priors are implemented in *Kalkayotl*. The statistical priors assume the clusters stars follow a uniform, gaussian or cauchy distribution. They can also follow a distribution defined as a mixture of gaussians. On the other hand the astrophysical priors are inspired by the functions used in the litterature to describe the surface density of star clusters either with a King function (King, 1962) or with an EFF profile (Elson et al., 1987).

We tested the performances of *Kalkayotl* on several clusters located between the 250 and the 550 pc limit discussed in Sect. 3.2. For some clusters, there is not always enough information in the data to properly constrain complex models such as the King, EFF or GMM prior. This typically happen for clusters with few members or large parallax uncertainties. In order to adopt a uniform approach for all the clusters in our sample, we used mainy the Uniform and the Gaussian priors.

We show the results for two of them: Ruprecht 147 and NGC 2422 located respectively at 305 pc and 476 pc on Fig. 3.3. We assumed for both clusters a scale length (i.e. the elongation along the line of sight) of 50 pc.

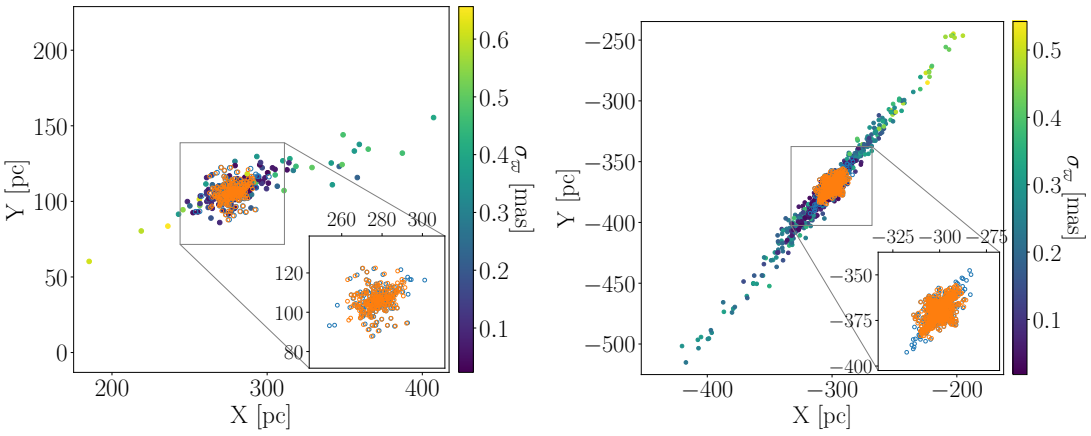

Figure 3.3: X-Y distribution of the stars of Ruprecht 147 (left) and NGC 2422 (right). The dots represents the X-Y positions calculated with a simple parallax inversion with a colour code standing for the parallax uncertainties. The blue and orange circles represents the X-Y distributions obtained with *Kalkayotl* distances using respectively a Gaussian and uniform prior. For clarity, the inner part of the cluster is zoomed in the bottom right pannel where we represented only the X-Y distributions obtained with *Kalkayotl*.

From Fig. 3.3, we can see that *Kalkayotl* allows to get rid of most of the radial distorsion due to the parallax uncertainties. It is particularly striking with Ruprecht 147 on the left panel. There is no sign of radial distortion any more when both Uniform

[2] in collaboration among other authors with Y. Tarricq

and Gaussian priors are used while the cluster originally spread over $\sim$ 260 pc. The stars with large parallax uncertainties (in yellow and in light green) have been dragged towards the cluster center by the use of both priors. We can also see a slight difference of the spatial distribution of the stars between the two priors. On the right panel, the spatial distribution of NGC 2422's members is also drastically modified by the use of *Kalkayotl*. The cluster which originally spread over $\sim$ 350 pc is shrinked to $\sim$ 50 pc along the direction of the line of sight. However with both priors, NGC 2422 is still distorted along the line of sight and we can also note a slight extension of the distribution of members in the direction perpendicular to the line of sight. For both clusters, we can also note that the spatial distribution of the members depends on the choice of the priors.

A three dimensionnal analysis of the cluster spatial distribution using *Kalkayotl* distances would therefore be biased by the choice of the prior and by the distances of the cluster as we can not totally get rid of the radial distortion of clusters, even for clusters located at less than 550 pc. In addition, *Kalkayotl* is computationnaly expensive and the users are recommanded to check the different steps of the computation of these bayesian distances for each prior they used and each cluster analysed. Even if *Kalkayotl* is the best tool available to correct the radial distortion of OCs, a three dimensional study of the clusters morphology would still be biased by this effect so we decided not to use this tool.

3.4 Three dimensional morphology

Because of the parallax uncertainties which translate into an distorted shape of the clusters along the line of sight, the study of the three dimensional morphology of OCs has to be restricted to a subsample of nearby clusters. Recently, Piecka and Paunzen (2021) recommended to conduct this kind of study only for clusters closer than 500 pc. However, we saw in Sect. 3.2 that a three dimensional study was biased for clusters located at more than 250 pc from the Sun. This can be seen on Fig. 3.4 taken from Piecka and Paunzen (2021). They represented here the members from CG+18 and calculated their (X, Y, Z) positions based on the bayesian distances from Bailer-Jones et al. (2018). It is clear from Fig. 3.4 that the three dimensional structure of the clusters become more and more distorted as the distance increases. Even a nearby cluster like NGC 1039 (located at 505 pc) shows a strong distortion along the line of sight compared to NGC 2632 (located at 186 pc). As mentionned in 3.2, we therefore chose to apply a more drastic distance cut of 250 pc for our three dimensional study of OCs.

Starting from the combined catalog of CG+18 and Gaia Collaboration et al. (2018a), we found seventeen OCs closer than 250 pc. Röser et al. (2019) and Meingast and Alves (2019) had recently characterized the tidal tail of the Hyades, and we also included the members identified by Röser et al. (2019). For the members coming from the catalog of CG+18, we kept only the members with a membership probability greater than 0.5. For three clusters, Platais10, Mamajek 1 and Alessi 13, this cut left them with less than 50 members so we discarded them from our sample.

In order to study the three dimensional morphology of the fourteen remaining clusters,

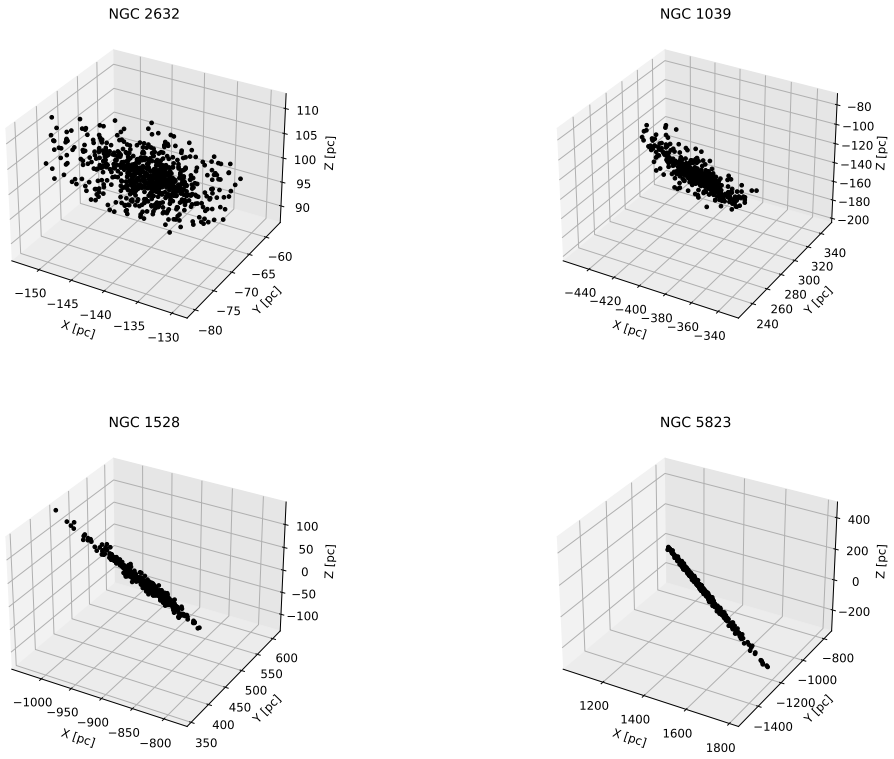

Figure 3.4: Three dimensional spatial distribution of four clusters NGC 2632, NGC 1039, NGC1528 and NGC 5823 (from top left to bottom right) located respectively at 186, 505, 1021 and 1813 pc. From Piecka and Paunzen (2021).

we fitted a Gaussian on the heliocentric coordinates of the clusters members. To do so, we took advantage of the Expectation-Maximization (EM) algorithm [3]. In this algorithm, the parameters of the Gaussian are randomly initialized (the user can also provide a first guess). The parameters of the Gaussian are then tuned in order to maximise the likelihood of the data under this model. Tuning the parameters of the Gaussian over a sufficient number N of iterations always allows to converge to a local maximum of the likelihood. For every cluster, we randomly generated new values of $(\alpha,\ \delta, \varpi)$ from their correlation matrix and we calculated the corresponding heliocentric positions. We run this fitting algorithm a 1000 times. We estimated the errors on the characteristic parameters of the ellipses as the standard deviation of the resulting distributions. The results of the fitting procedure are shown as an example for two clusters on Fig. 3.5.

The fourteen clusters for which we studied the three dimensional morphology are shown in Fig. 3.6. We show here the distribution of their members in the heliocentric coordinates in the three planes (X-Y, X-Z and Y-Z) alongside with the projection on each plane of the ellipse fitted on their spatial distribution. In the X-Y plane, five clusters seem to show an elongation of their fitted ellipse: the Hyades, Platais 9, NGC 2451A, ASCC 123 and NGC 2632. In the case of the Hyades, the only cluster for which the tidal tail was identified in this sample, the elongation of the ellipse follows the well known

[3] Algorithm implemented in the scikit-learn python package (Pedregosa et al., 2011)

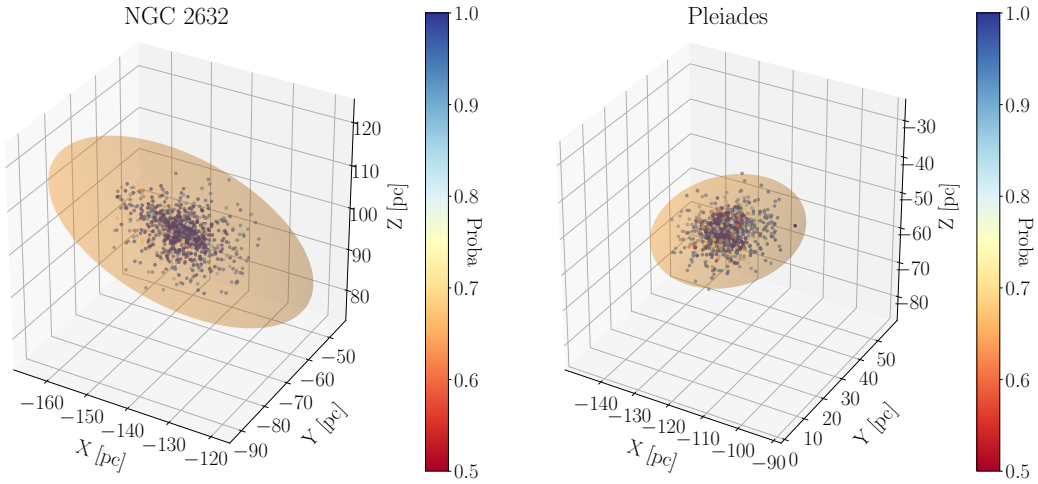

Figure 3.5: Three dimensional spatial distribution of the members of NGC 2632 (left) and the Pleiades (right). The orange ellipsoid represents the 3 σ gaussian fitted on the distribution of the stars and the color bar stands for the membersip probability.

S-shaped structure of a tidal tail. The leading arm, oriented in the direction of the Galactic rotation clearly shows stars falling towards the Galaxy center while the trailing arm shows stars with higher Galactocentric velocities (Kroupa, 2008, p. 20). For the others clusters, the elongation is aligned with the direction of the line of sight. This elongation might therefore not be due to a genuine elongation of the clusters but to the already discussed distortion induced by the parallax uncertainties in the radial direction. In the X-Z plane, besides the Hyades and the clusters distorded along the line of sight like Blanco 1 or NGC 2632, only Platais 3 shows a significant elongation. In the Y-Z plane, only Platais 9 shows an elongation which can not be confused with a distortion along the line of sight. Only three clusters show, in at least one of these three spaces, an elongation of their fitted ellipse which is independant from the radial distortion: the Hyades, Platais 3 and Platais 9. This does not allow us to reach any conclusion regarding a preferential orientation of the three dimensional structure of the clusters in our sample.

We still tried to investigate an eventual correlation between the parameters of the fitted ellipse and the age of the clusters in our sample. We represented in Fig. 3.7 the lengths of the characteristic axis of the ellipse fitted on the three dimensional structure of the clusters as a function of clusters' ages as well as the eccentricities of these ellipses. When available, we took advantage of the catalog of ages published by Bossini et al. (2019). If not, we used the ages of Dias et al. (2002), Gaia Collaboration et al. (2018a) or Kharchenko et al. (2013). The Hyades clearly stands out with the largest values for the three axes of its ellipse. This is not surprising as we consider in our fit the members of its tidal tail identified by Röser et al. (2019). Yet, the Hyades is not much more eccentric than the rest of the sample. All of the ellipses have eccentricities higher than 0.6 regardless of the ages of the clusters. As very young clusters are believed to be in a nearly spherical state (Krumholz et al., 2019, p. 47), we would expect their eccentricity to increase with time but we can not reach such a conclusion based on our sample. This might be due to the fact that clusters get their flattened shape very early in their life. We

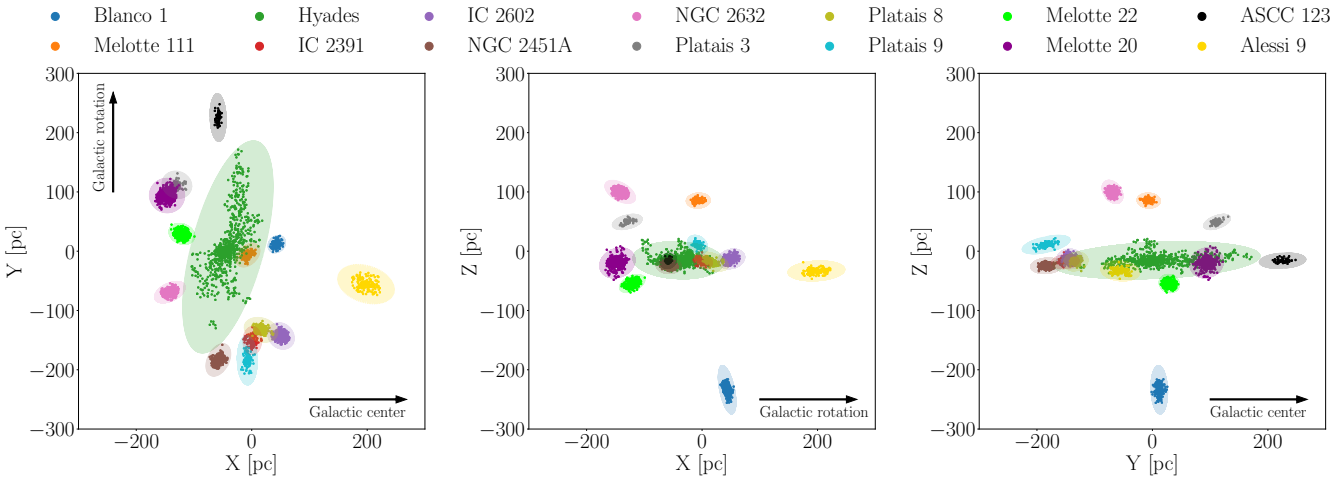

Figure 3.6: Spatial distributions of the clusters members in the heliocentric reference frame. A projection of the fitted 3 σ gaussian is overplotted on the distribution of members for each cluster. The Sun is located at the origin of each panel and the direction of the Galactic center and Galactic rotation are indicated with arrows.

can not verify this hypothesis as the youngest cluster of our sample (IC 2602) is already more than 35 Myrs old. Alternatively, selecting only the clusters located closer than 250 pc from the Sun might not be enough to get rid of the distortion of clusters in the line of sight. We saw in Fig. 3.6 that lots of them were still distorded.

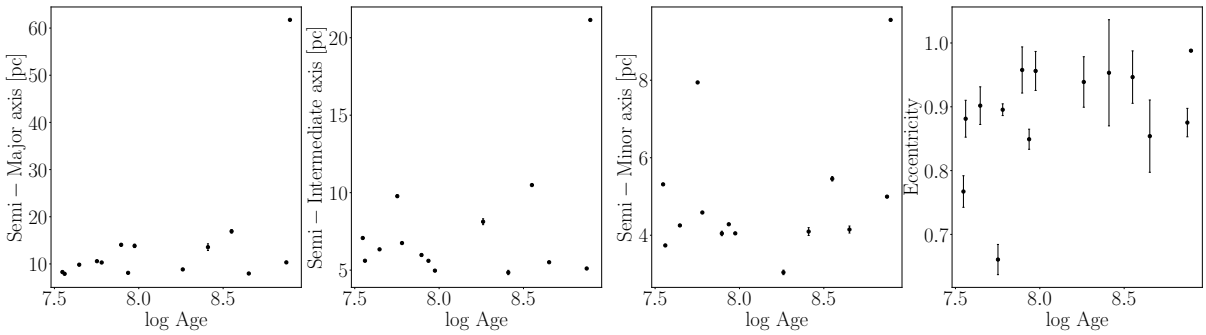

Figure 3.7: Respectively: lengths of the semi-major, semi-intermediate and semi-minor axis and eccentricity of the ellipse fitted on the clusters' three dimensional spatial distribution as a function of the clusters' ages.

Recently, Meingast et al. (2021) investigated the spatial distributions of the members of nearby clusters. By deconvolving the distances of clusters members with a mixture of Gaussians in order to mitigate the distortion of the clusters, they managed to infer the true spatial distribution of the clusters in their sample. Out of their ten clusters sample, they detected vast stellar halos around nine clusters. These stellar halos are similar to the tidal tail of the Hyades shown in Fig. 3.6 both in their shapes and orientations. However, they did not find any correlation between the age of the clusters and their shapes, orientation or elongation.

Only fourteen clusters are located closer than 250 pc from the Sun. In addition, they might still suffer from the radial distortion discussed throughout this section. For these

reasons, we decided to study the morphology of clusters only in two dimensions. This allows our study to not be affected by the radial distorsion and to include much more clusters in our sample.

3.5 Two dimensional morphology[4]

In this section, we only focus on the two dimensional morphology of OCs. We took advantage of our membership analysis explained in Sect. 2.3. This analysis contains new lists of members extended to the outskirts of the clusters for 389 of them. As we detected previously unreported stellar halos around a significant amount of clusters, this selection of members is more suitable than any other (e.g. CG+18, CG+20, Castro-Ginard et al., 2018, 2019, 2020) to study the extent of OCs. In addition, this constitute one of the largest sample dedicated to an homogeneous and systematic determination of OCs' structural parameters.

3.5.1 Radial Density Profiles

The Radial Density Profile (RDP) is a good indicator to study the extension of the spatial distribution of the clusters members. Once obtained, the resulting density profile can be characterized by means of the fit of the widely used King empirical function (King, 1962), the EFF profile (Elson et al., 1987) and the Plummer function (Plummer, 1911).

3.5.1.1 Fitting procedure

The King's profile is widely used to fit the RDP of OCs although it was first introduced to describe the surface density of globular clusters (King, 1962). It is defined as:

$$n(R) = \begin{cases} k \cdot \left(\frac{1}{\sqrt{1+(R/R_c)^2}} - \frac{1}{\sqrt{1+(R_t/R_c)^2}} \right)^2 + c & \text{if } R < R_t \\ c & \text{if } R \geq R_t, \end{cases} \tag{3.5}$$

where k is a scaling constant related to the central density, R_c the core radius, R_t the tidal radius and $n(R)$ is the surface density in stars per squared parsecs. Following Küpper et al. (2010), we added a constant c to the original formula of King (1962), also in stars per squared parsecs. We expect c to be close to zero since we are considering the most reliable cluster members. This constant improved significantly the quality of the fits for many clusters. The core radius is defined as the radius for which the value of the density is equal to half the central density. The tidal radius is the radius where the cluster becomes indistinguishable from the field (King, 1962). In our case the tidal radius is therefore the radius for which the density is equal to c.

[4]This chapter is based on Tarricq et al. (2021b). The work detailed in Sect. 3.5 and 3.6 has been conducted in collaboration with several co-authors who contributed to the improvement of the radial density profile fitting procedure and to the presentation and interpretation of the results.

The EFF profile was proposed by Elson et al. (1987) in order to better describe the outer parts of young OCs in the Large Magellanic Cloud, which were poorly described by the King profile. Its surface density is defined as :

$$n(R) = k' \cdot \left(1 + (R/a)^2\right)^{-\gamma/2} + c, \tag{3.6}$$

where k' is a constant also related to the central density, a the radial scale (i.e. the characteristic radius of the EFF profile) and γ the slope of the profile at a large radius, much larger than R_c. As for the King profile, we added a constant c acting as a field constant density to improve the quality of the fit. The radial scale is defined to match the core radius from the King profile according to the following relation: $r_c = a\sqrt{2^{2/\gamma} - 1}$.

The last template we used in this study was first used by Plummer (1911) in order to fit the brightness surface density of globular clusters. Applied on OCs, its density is equal to :

$$n(R) = k'' \cdot \left(\frac{1}{1 + (R/b)^2}\right)^2 + c, \tag{3.7}$$

with k'' a constant and b the scale value of the Plummer profile which, interestingly is proportional to the half mass radius of the cluster : $r_{hm} \sim 1.3 \cdot b$.

The first step in order to fit a model such as the King, EFF or Plummer profile to a cluster is to determine its RDP. To do so, we first needed to calculate the distance between each cluster's stars and its center. The centers of the clusters are computed as described in Sect 2.3.5. Some of the clusters of our sample such as Ruprecht 98 have high declinations and the distribution of their members in the sky is therefore subject to strong projection effects. Some clusters, especially the most nearby ones, are also sensitive to projection effects due to the curvature of the celestial sphere, especially for nearby clusters. To avoid these biases, we projected each stars coordinates on the plane of the sky, tangential to the celestial sphere at the coordinates of the clusters centers, as suggested by van de Ven et al. (2006) and Olivares et al. (2018a). The projected coordinates are defined for each cluster star as:

$$\begin{aligned} x &= D \cdot \sin(\alpha - \alpha_c) \cdot \cos(\delta) \\ y &= D \cdot \cos(\delta_c) \cdot \sin(\delta) - \sin(\delta_c) \cdot \cos(\delta) \cdot \cos(\alpha - \alpha_c), \end{aligned} \tag{3.8}$$

where D is the heliocentric distance of the cluster computed by CG+20 in pc, and α_c and δ_c are the clusters mean right ascension and declination.

The radial distance R of each star to the center of the cluster is:

$$R = \sqrt{x^2 + y^2}. \tag{3.9}$$

We divided the spatial distribution of the stars on these projected coordinates in concentric rings. We used 10 bins of one parsec width for the inner parts of the clusters and then we progressively increased the width of these rings. We computed the density

defined as the number of stars per square parsec in each ring.

We fitted the aforementionned profiles with a Maximum Likelihood (ML) estimator considering Poissonian uncertainties for each point. We used the MCMC sampler `emcee` (Foreman-Mackey et al., 2013) and initialized eight "walkers" for the King and EFF profiles and six for the Plummer profile (two per parameter to fit). For each walker we assigned 10,000 iterations to converge, and we discarded the first 2,000 iterations to compute the posterior. As recommended, the convergence of the walkers was systematically checked based on the integrated auto-correlation time (i.e. an estimation of the needed length of each walker Goodman and Weare, 2010). As a test we tried to run our fitting procedure on a simulated cluster. The spatial distributions of this synthetic cluster's stars was initialized with the code *mcluster* (Küpper et al., 2011) and the positions in the sky and their error were calculated through the *PyGaia* python package. The density profile of this synthetic cluster was chosen to be an EFF profile with a core radius $R_c = 1$ pc and a power law of $\gamma = 5$. We recovered a core radius of $R_c = 1.03^{+0.21}_{-0.08}$ pc and a power law $\gamma = 4.99^{+0.97}_{-0.39}$. The true parameters of the density profile being recovered by our fitting procedure, we considered it as satisfying. The results of the fitting procedure of a King profile are shown as an example for the cluster NGC 752 in Fig. 3.8 where we found a core radius $R_c = 2.04^{+0.31}_{-0.3}$ pc and a tidal radius of $R_t = 26.45^{+5.15}_{-8.46}$ pc. We applied the aforementioned procedure on the 233 clusters in our sample counting more than 100 members. We chose this lower limit in order to have sufficient number of stars in the circular rings.

We only considered as satisfactory results the fits for which no flag was risen by the integrated auto-correlation time regarding the convergence of the walkers of the fitting procedure. For the King profile, we also discarded the determinations of R_c with errors greater than 2.5 pc and the determinations of R_t with errors greater than 15 pc. This leaves us with estimations of R_c and R_t for respectively 172 and 146 OCs. Both of these quality cuts are mostly useful to discard the cases where the walkers did not converge but went through the integrated auto-correlation check. We performed the same convergence checks for the EFF and the Plummer profile. In addition, we considered only the determinations of the characteristic radii a and b with errors lower than 2.5 pc and the determinations of γ with errors lower than 1.5.

3.5.1.2 Discussion

Among the three investigated profiles, the King's function generally provided a better fit. It has also been much more used than the other two profiles in the literature and it is the only one providing an estimation of the tidal radius which particurlarly interests us as we aim at studying the sizes of OCs. For these reasons, a particular emphasis is put on this profile.

The tidal radius estimated in this study through the fit of a King's profile shall not be confused with the Jacobi radius introduced by Binney and Tremaine (1987, p. 450). The tidal radius is a parameter of the King's RDP while the Jacobi radius R_j is often referred to as the tidal radius but is only a crude estimate of it. Unlike the tidal radius, the Jacobi radius does not involve the density to be equal to zero at $R = R_j$.

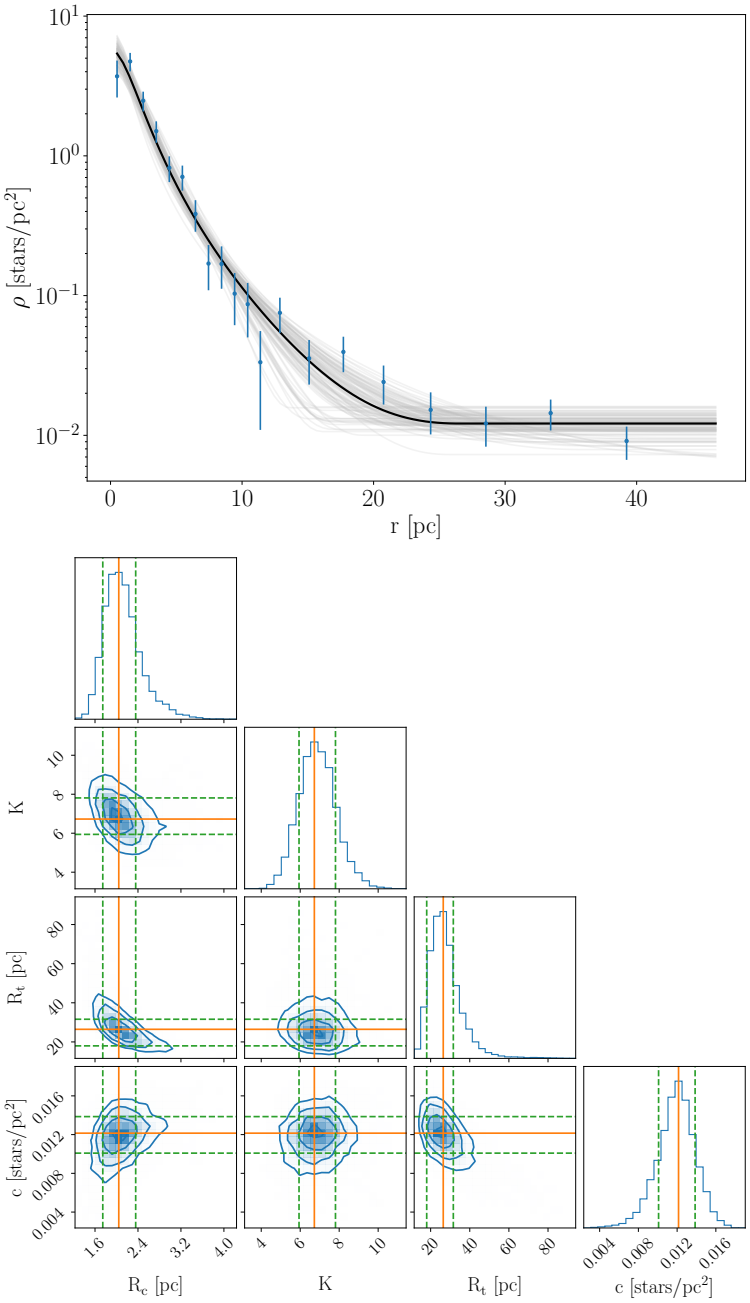

Figure 3.8: Results of the King's profile fit on the cluster NGC 752. In the top panel, the blue dots are shown with Poissonian uncertainties. It also shows the best fit obtained with a ML estimator (black solid line), defined as the mode of the distributions of the parameters obtained though the 64 000 fits performed. The grey lines represent the uncertainties on the fits: we show 100 fits taken from the posterior distribution of the ML. The bottom panel shows the corresponding projection of the parameters posterior distribution. The orange lines show the mode of each distribution and the green dashed line shows the 68% HDI.

The distributions of the fitted cores and tidal radii are shown in Fig. 3.9. The core and tidal radii are computed as the mode of the parameter distributions of our ML procedure's walkers. The uncertainties represent the lower and upper bound of the 68% HDI of the ML walkers. The vast majority of clusters have a core radius between 1 and

2.5 pc, regardless of their age and number of members. The most frequent value of the core radius is $\sim$ 1.85 pc. We can also note that the vast majority of the clusters with less than 250 members (in blue in the figure) have values of R_c slightly smaller than the mode of the distribution while more populated clusters tend to have larger values. Finally, the dispersion of the core radius decreases with the increasing age of the clusters. This indicates that even if young clusters can have very concentrated cores, this feature is more common for old clusters. This is in agreement with the hypothesis discussed by Heggie and Hut (2003) that the evolution of the inner parts of the cluster is dominated by two-body relaxation causing the cluster core to shrink. Two-body relaxation is also known to cause mass segregation in clusters: massive stars concentrate in the cores of the clusters while less massive stars move in their outskirts. It could be connected to the observed decrease of the core radius with age. As more massive stars concentrate towards the cores of the clusters, the cores' gravitational potential increases which causes it to be denser. Mass segregation will be studied in details in Sect. 3.6. We want to point out that the clusters which deviates the most from the mode of the distribution are the ones with the largest errors.

Looking at the right panel of Fig. 3.9, we can see that the distribution of the tidal radius is bimodal. It peaks around 28 pc with a secondary peak at $\sim$ 18 pc. The majority of clusters with less than 250 members (in blue) have values of $R_t \sim$ 18 pc or lower while almost all of the populated clusters (in yellow) have larger values. Additionally, the tidal radius has a mild increase with the age of the clusters. We illustrate this increase by overplotting a linear regression of the tidal radii versus the age of the clusters. To perform the fit, we used a simple least square method taking into account the uncertainties in the tidal radii. We obtained values of 4.64$\pm$1.63 and -17.00$\pm$13.87 respectively for the slope and the y-intercept of the fit. This increase could again be connected to mass segregation or to cluster evaporation: because more stars have moved to the outskirts of the older clusters they are more likely to be torn off from the clusters and the tidal radius increases with age.

Core and tidal radii of OCs have often been determined in the past through the fit of a King profile. The most extensive catalogue of radii was published before the launch of the *Gaia* mission by Kharchenko et al. (2013). Also before *Gaia*, Piskunov et al. (2007) published a catalogue of radii for 236 OCs out of the 650 clusters with reliable memberships from the catalogue ASCC-2.5. More recently, Angelo et al. (2021) studied in details the structural parameters of 38 OCs with *Gaia* DR2 data. The comparison of our determinations of core and tidal radii with these three studies is shown in Fig. 3.10. We note that Angelo et al. (2021) looked for members at a maximum radius of 1° around the clusters centers computed by Dias et al. (2002). This is equivalent for most of their clusters to a radius smaller than 30 pc. We therefore have very different list of members, which makes a close comparison difficult. Nevertheless, we can compare the distributions of R_c and R_t as shown in Fig. 3.10, the similarity between the distribution of R_c in all the studies is striking while the tidal radii computed here are much larger than those computed in the other studies. This is a direct consequence of our choice to search for members at larger distances from the center of each cluster, compared to the previous

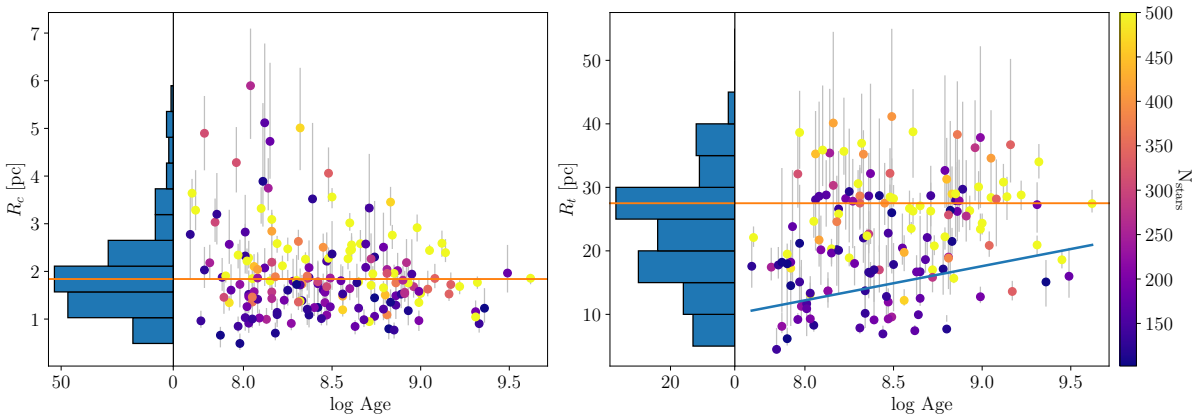

Figure 3.9: Distributions of the fitted core radii R_c (left) and tidal radii R_t (right) from the King's profile as a function of the logarithm of the cluster ages and their corresponding histograms. The color bar stands for the number of cluster members and the mode of the distributions is overplotted with the orange solid line. A linear regression of the tidal radii versus age has been fitted with a least square method (blue line).

studies. On another hand, in the case of M 67/NGC 2682, Carrera et al. (2019a) looked for members up to 150 pc around the center of the cluster. They fitted a King function to the RDP and estimated a tidal radius of 80 pc while we find a value of $31.08^{+1.37}_{-1.48}$ pc. This suggests, as previously reported by Olivares et al. (2018a) that the determination of the tidal radius is highly dependant of the size of the survey. Therefore our distribution of tidal radii is likely truncated due to our queries around each cluster limited to 50 pc.

We estimated the degree of contamination of our sample by looking at the c constant from Eq. 3.5. This constant, in stars per squared parsecs, acts as a field constant density and should be closed to 0 as we consider only member stars in our fitting procedure. Knowing the area occupied by the cluster, it gives an estimated number of the contaminants. The median proportion of contamination in our sample is at $\sim 13\%$. This estimation of the contamination rate is biased for some clusters, leading to high contamination rate. As explained in Sect. 2.3.4, our distribution of members is likely truncated for populated clusters. This leads to an underestimation of the tidal radii of these clusters with typical values of ~ 30 pc. The members detected beyond this estimation of the tidal radii act as a background field density in the fit, leading to an overestimation of the c constant i.e. of the contamination rate. Moreover, the King profile might not be the best way to describe the density of some clusters with extended halos (Küpper et al., 2010), especially if they present an elongation like for Blanco 1: tidal tails stars will also act here as a background field density. That is why we fitted a GMM on the spatial distribution of members (see Sect. 3.5.2).

Schilbach et al. (2006) noted an increase of the size of the clusters with their altitude above the galactic plane and that this increase was especially significant for clusters older than ~ 22 Myr. They also reported that large clusters were found at large Galactocentric distances. They concluded that clusters located within the solar orbit and with a low inclination of their orbit with respect to the Galactic plane are likely to be rapidly dissolved by encounters with giant molecular clouds or by Galactic tidal stripping. On

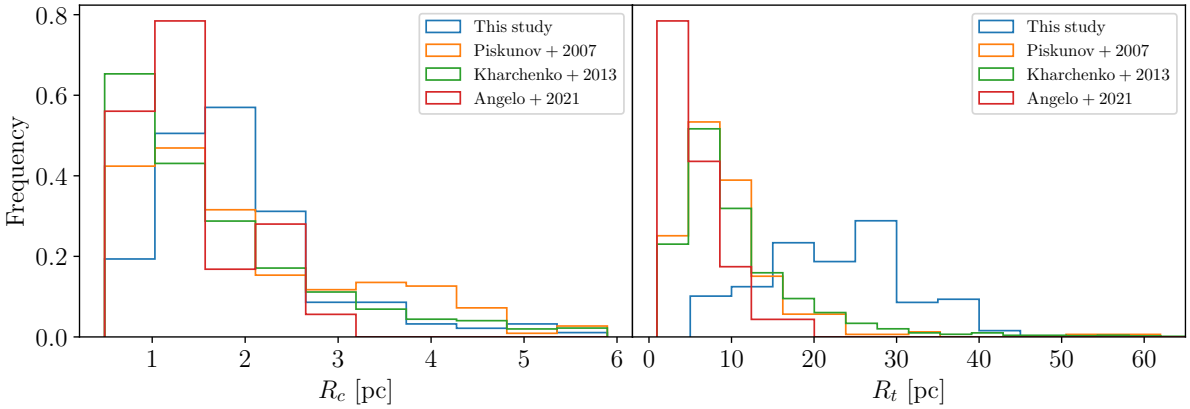

Figure 3.10: Distribution of the core (left) and tidal (right) radii computed in this study and by Piskunov et al. (2007), Kharchenko et al. (2013) and Angelo et al. (2021).

the contrary, clusters with an orbit outside of the solar one and reaching high altitudes are more likely to survive longer and to not be stripped of their members. We do not see any of those correlations, even when we divide our sample of OCs in different age bins so we can not confirm these findings. More recently, Dib et al. (2018) also failed to confirm these findings.

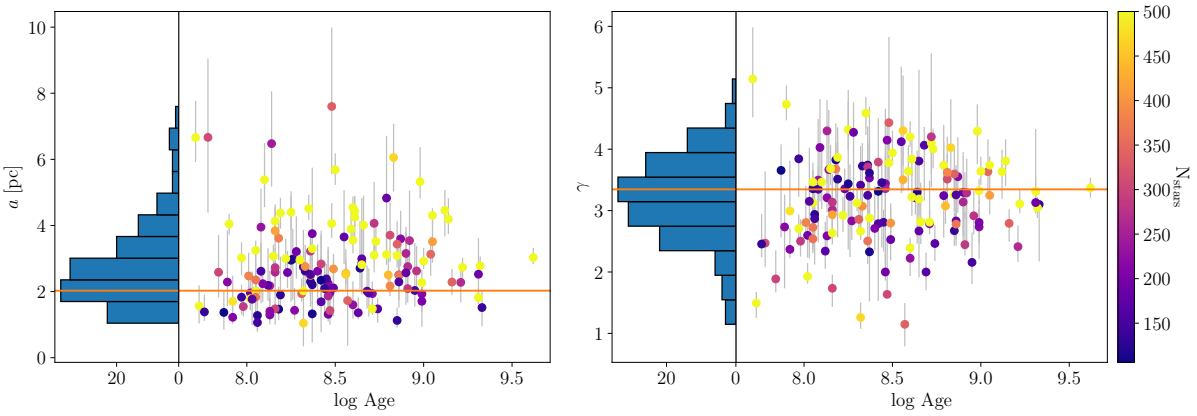

Figure 3.11: Distributions of the fitted scale radii a (left) and γ (right) from the EFF profile as a function of the logarithm of the cluster ages and their corresponding histograms. The color bar stands for the number of cluster members and the mode of the distributions is overplotted with the orange solid line.

The distributions of the EFF and Plummer characteristic parameters are shown in Fig. 3.11 and Fig. 3.12. The parameters and their uncertainties are determined in the same way as for the King's core and tidal radii: by respectively considering the mode and the 68% HDI of the parameters' distribution of our ML procedure's walkers. In Fig. 3.11, we can see that like for the King's core radii, the scale radii of the EFF profile concentrate between 1.5 and 3 pc regardless of the age of the clusters. The mode of the distribution is $\sim$ 2 pc. The large majority of clusters with less than 250 members have a scale radii smaller than $\sim$ 3 pc while more populated clusters have larger values. There

is no apparent trend of the dispersion of the core radii here. The right panel of Fig. 3.11 shows the distribution of the γ parameter with the clusters ages. This parameter seems to follow a normal distribution centered around $\gamma \sim 3.3$. There is no correlation of γ with age or with the number of members per cluster. However, there seems to be a decrease of the dispersion of γ as age increases, with values converging towards the mode of the distribution. This might indicate that the slope of the RDP at large radius of young OCs is randomly distributed and that as clusters grow old and reach equilibrium, their RDP at large radius tend to normalize.

Looking at Fig. 3.12, we can see that the characteristic radius b of the Plummer profile follow the same trend as the King's core radii in Fig. 3.9. The scale radii b concentrate between 3 and 6 pc regardless of the clusters ages. Its dispersion decreases as age increases and the radii converge towards values of ~ 5 pc for old clusters. Figure 3.12 confirms the decrease of the dispersion of the core radii of clusters with age already observed in Fig. 3.9.

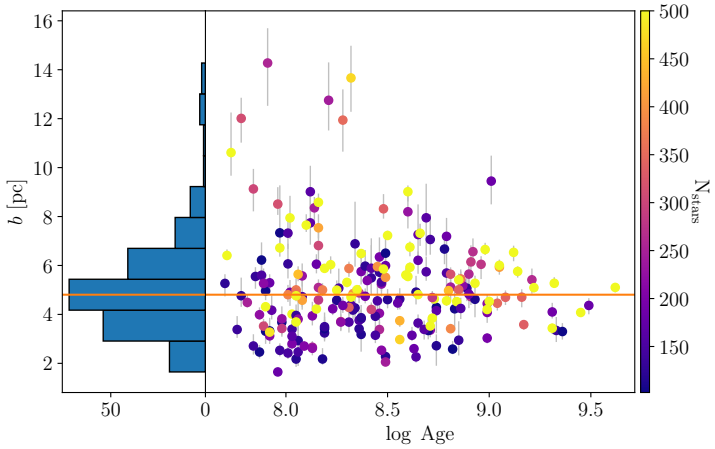

Figure 3.12: Distributions of the characteristic radii b from the Plummer profile as a function of the logarithm of the cluster ages and its corresponding histograms. The color bar stands for the number of cluster members and the mode of the distribution is overplotted with the orange solid line.

3.5.2 Gaussian Mixture Models

The functions describing the RDP used in the previous subsection assumes a circular distribution of the members. In order to study the morphology of the OCs without making this assumption, we fitted a GMM on the spatial distribution of members projected on the sky for each OC. A GMM is a probabilistic model assuming that the data can be described by a combination of a finite number of Gaussian distributions.

3.5.2.1 Fitting procedure

To get rid of projection effets due to clusters located at high galactic latitudes and to members located far from the clusters' centers, we used Eq. 3.8 to project the galactic

latitude and longitude of each star in a cluster on a plane tangential to the celestial sphere. In order to fit a GMM on these coordinates, we used a variational GMM with a Dirichlet process prior[5]. This algorithm is a variant of the classical GMM and allows to infer the effective number of components from the data. Usually, classical GMM fitting take advantage of EM algorithm already mentionned in Sect. 3.4. The variational inference extends the EM approach by adding information through a prior distribution: the Dirichlet process. With the Dirichlet process prior, the number of components set by the user is only used as an upper bound, the algorithm will automatically draw the number of components from the data, activating a component and attributing stars to this component only if necessary: if the maximum number of components is set to 3 but the Dirichlet process only detects 2, it will set the relative weight of one of the component to $\sim$ zero.

As explained in Sect. 2.3.2, we realised by visual inspection that some clusters presented an elongation in their outskirts which could correspond to a tidal tail. A fraction of these OCs has a number of members comprised between 50 and 100. So in order to characterize as many tidal tails as possible, we fitted a GMM on all the clusters with more than 50 members. We also noticed that most of the clusters present a prominent core and an extended halo. Consequently we systematically tried to fit three components to the sky distribution of the members of each cluster. One component would correspond to the core of the cluster, the second one to the eventual tidal tail and the third one to the cluster's halo or coronae. Because not all the clusters show a tidal tail, the Dirichlet process is therefore well suited for our purpose: if only two components are detected, the weight of the component standing for the tidal tail is supposed to be close to zero.

Since there is a stochastic initialisation in the variational inference algorithm, the fit does not always converge towards the same solution. Therefore, in order to have a better estimation of the parameters of the Gaussians, we run the fitting algorithm for each cluster a thousand times. The parameters of the resulting Gaussians were then chosen as the mode of the resulting distributions and their standard errors were computed as: $MAD/\sqrt{N_{iter}}$ with $N_{iter} = 1000$. In addition we noticed the fits were better when the algorithm is forcing the gaussians to be concentric.

For some clusters, the algorithm found a prominent and elongated component of the GMM which could possibly be associated with a tidal tail. However, we could not find a clear cut in weights or in eccentricity to separate clusters with and without elongation, owing to the large variety of clusters and environments we are dealing with. In particular, the most populated clusters (more than 500 members) with a dense core always end up with a second component having a significant weight even if it does not represent a tidal tail but more likely the outskirts of the core. In most of these cases the results with two components were sufficient to fit the distribution . We also tried to separate the clusters with and without a tidal tail using the length of the semi-major axis of the second component, the number of stars attributed to each components or the ratio between the semi minor and the semi major axis of the second component but we did not find an ideal way to separate the two subsamples with accuracy. That is why we visually identified

[5]Algorithm implemented in the scikit-learn python package (Pedregosa et al., 2011)

them. We defined a subsample of 71 OCs with this feature. We show the results for four clusters which present a remarkable elongation in Fig. 3.13. A prominent halo is also noticeable around the cores of the four clusters. The legend in each panel of Fig. 3.13 shows the weights attributed to each component by the algorithm.

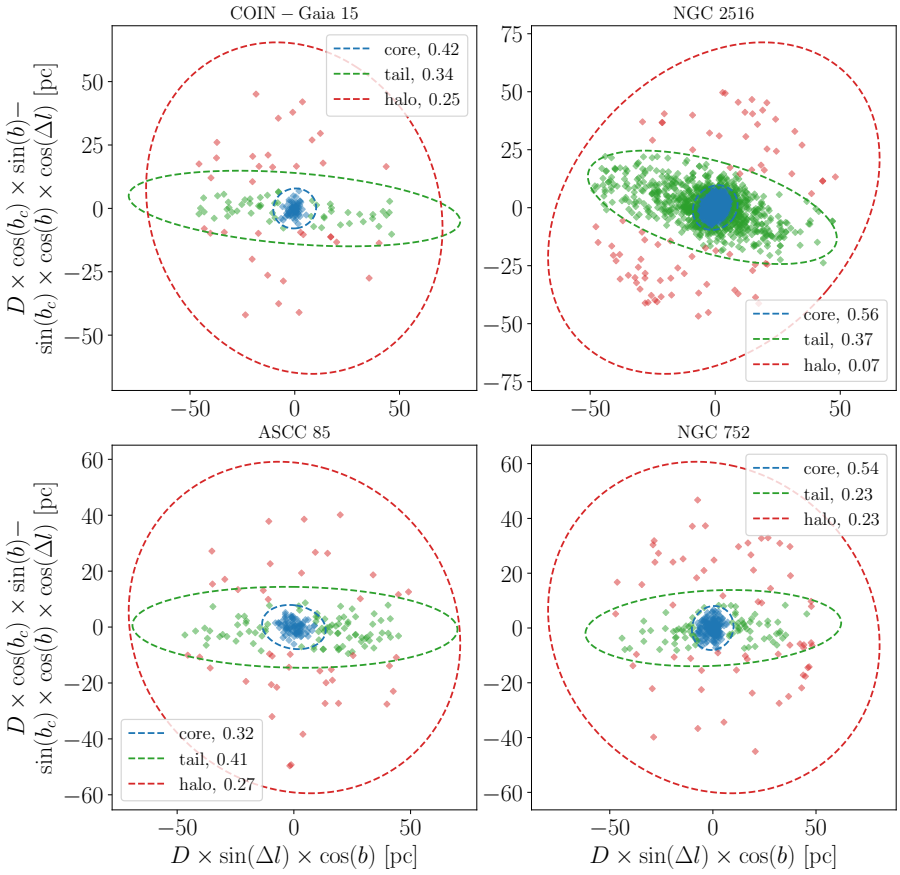

Figure 3.13: Example of four clusters for which we detect a tidal tail. The blue, green and red ellipses represent the 3 σ ellipse fitted on the distribution of the stars standing respectively for the core, the tidal tail and the halo. The stars are colored according to which components they are the most likely to belong to. The relative weights of each component is indicated on each panel.

In the following, for the 71 clusters identified as having a tidal tail, we study the parameters of the 3 components solution. For the rest of the clusters, we adopt the parameters of the 2 components solution.

Recently, Hu et al. (2021) used a similar approach and fitted a two components model using a Least Square Ellipse Fitting on 265 OCs from the membership catalogue by Cantat-Gaudin et al. (2018b). A direct and systematic comparison of their results and ours is not appropriate because they investigate much smaller areas around each cluster compared to our study. However we checked some clusters individually. For instance, we find a tidal tail around NGC 752 (Fig 3.13) with roughly the same orientation and eccentricity as them, according to their Fig. 1. We found an orientation of 6.146 degrees with the Galactic plane and an ellipticity of 0.776±0.002 while they found an angle of

4.697 degrees and an ellipticity of 0.615±0.342.

3.5.2.2 Discussion

The majority of the clusters have a significant corona. Among the clusters well described by a two components solution, there are 254 of them with a weight higher than 0.1 attributed to the corona. In Fig. 3.14 we show the weights of the halo as a function of the age for these clusters. We can see that even if young clusters can have halos with very various weights, we did not find any old clusters with a halo having a weight higher than 0.4. This indicates that as clusters grow old, less stars are part of the corona and that a higher proportion of their stars tend to concentrate in the cores of the clusters. This process might be connected to both the mass segregation and the cluster evaporation. Mass segregation tends to make the most massive stars of a cluster sink into its center and the less massive stars move towards its outskirts. On the other hand, if the cluster progressively evaporates, the outskirt stars are teared out from the corona as time goes by, thus reducing its weight as it is shown in Fig. 3.14.

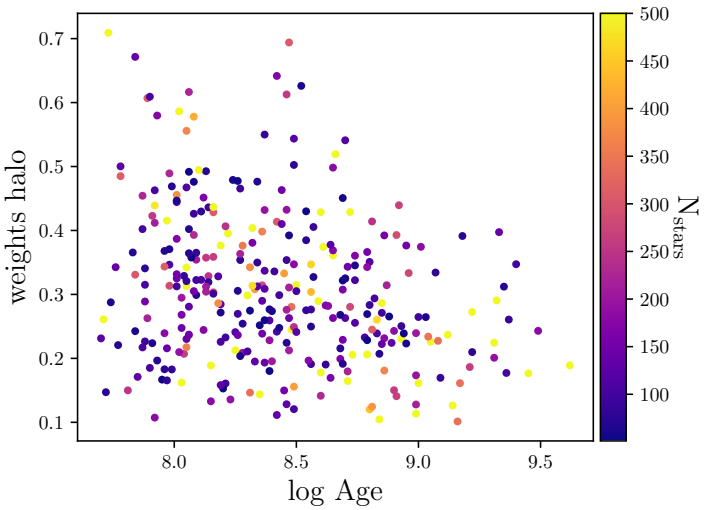

Figure 3.14: Distribution of the weights of the second Gaussian component (associated to a halo structure) as a function of the age for the subsample of clusters which is best fitted with a two components model.

Hu et al. (2021) and Zhai et al. (2017) also fitted ellipses on the distribution of members projected on the plane of the sky. They noticed an increase of the ellipticity of the outer parts with age on a sample of 265 and 154 clusters respectively. Based on 31 OCs, Chen et al. (2004) also noticed an increase of the circularity of the inner parts of OCs with age, especially at high altitudes. They attributed this process to internal dynamical relaxation process at stake in OCs, internal dynamics being able to shape clusters cores after $\sim$ 100 Myr while younger clusters inherit their shape from the cluster formation initial conditions. We considered the fitted parameters of the core and halo ellipses: length of the semi major axis, eccentricity and orientation. We looked for correlations

with the Galactocentric radius, age and number of stars and we do not find any relevant trend.

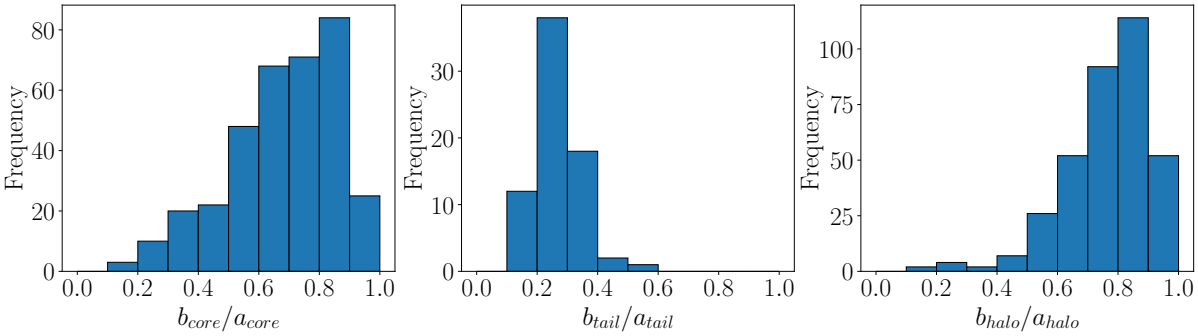

Figure 3.15: Distribution of the axis ratio of each component.

As mentioned in Sect. 3.5.2.1, we detected 71 tidal tails. We show on Fig. 3.15 the distribution of the axis ratio of each fitted component for our whole sample of clusters. In the left hand panel, we show the distribution of the axis ratio of the cores. It shows that the vast majority of the clusters have a nearly circular core as the distribution peaks between 0.8 and 0.9. We checked the clusters with a ratio between the axes lower than 0.5. They do not contain much stars in their center, making their ellipticity less reliable. In the middle panel is shown the distribution of the axis ratio of the tidal tails. They all are very elongated with $\sim 70\%$ of the identified tidal tails having an axis ratio lower than 0.3. Finally, the right hand panel shows the distribution of the axis ratio of the corona. In most of the cases, the corona is almost circular, and in the few cases where the ratio between the axis is lower than 0.5, again the number of stars in the halo is very low.

We can characterize the components of the 71 clusters in our sample showing a tidal tail by looking at the properties of their components. We show on Fig. 3.16 the length of the core semi major axis as a function of clusters ages and of the logarithm of the number of stars belonging to it. The left panel shows that the clusters with tidal tail also follow the decreasing relation of the core radii with age found in Fig. 3.9 using all clusters. In the right panel, it is shown that the length of the axis also has a decreasing dependence with the number of stars belonging to it. This means that even if populated clusters could be thought to have bigger cores, it seems on the contrary that their stronger gravitational binding make them much denser OCs. We looked for correlations between both tidal tails and halos semi-major axis lengths, eccentricities and orientations with the age of the clusters or with their locations in the Galaxy without finding any particular trend.

Sixteen tidal tails have already been characterised in the literature. Eight of them are part of our sample : Coma Berenices, Ruprecht 147, Praesepe, Blanco 1, NGC 752, NGC 7092, NGC 2516 and Platais 9. From these, our study identifies tidal tails for Coma Berenices, Blanco 1, NGC 752, NGC 7092 and NGC 2516 previously found respectively by Tang et al. (2019), Zhang et al. (2020), Bhattacharya et al. (2021) and Meingast et al. (2021). We found the same orientations as these authors. The tidal tails of Ruprecht 147, Praesepe, and Platais 9 were characterised respectively by Yeh et al.

(2019), Röser and Schilbach (2019), Gao (2020) and Meingast et al. (2021) who found them to be roughly aligned with the line of sight. We cannot detect such tidal tails owing to our 2D analysis of the distribution of stars projected on the sky.

Meingast et al. (2021) identified extended stellar populations similar to tidal structure in nine out of 10 OCs. They attributed these structures to the imprint of the parent molecular cloud relic structure for the clusters younger than 50 Myr and to stripped cluster stars for the clusters with age $\gtrsim$ 100 Myr. Pang et al. (2021) also found elongated shapes in 8 of their 13 OC sample in the form of filamentary like substructures, reminiscent of the star formation history of the cluster for the clusters younger than 50 Myr and to tidal stripping for the oldest ones (NGC 2516, Blanco 1, Coma Berenices, NGC 6633 and Ruprecht 147). We find the same structures for the clusters we have in common (with the exception of Ruprecht 147). According to Lada and Lada (2003), Bonnell and Davies (1998) and Bastian et al. (2009), in only a few million years, a cluster would reach a state of equilibrium and get rid of its star distribution inherited from the star formation process. As we do not have any clusters younger than 50 Myr in our sample, the vast majority of the tidal tails in our sample can be attributed to dynamical effects and to tidal stripping.

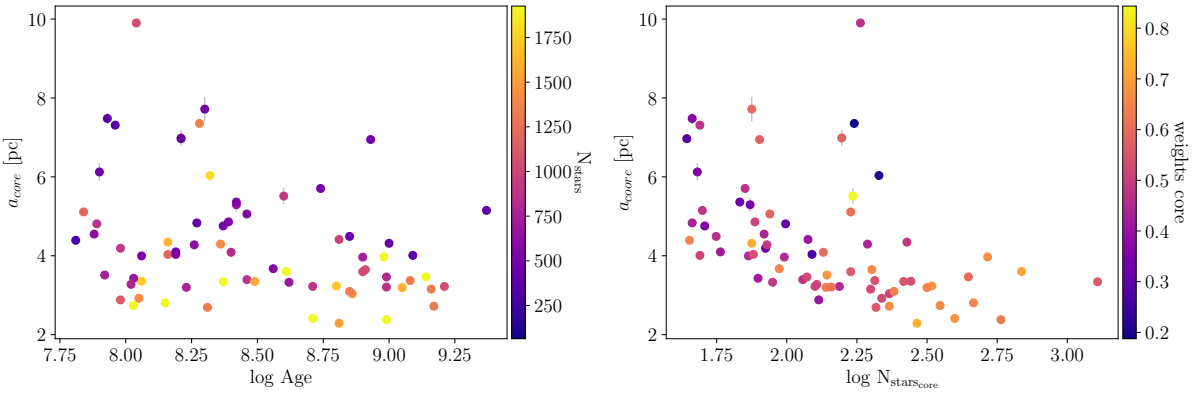

Figure 3.16: Length of the semi major axis of the core of our clusters with age (left) and with the logarithm of the number of stars belonging to it (right) for the subsample of clusters with a tidal tail. On the left panel, the color bar stands for the total number of stars belonging to the cluster and on the right panel, it represents the relative weights of the core.

3.6 Mass segregation

According to the standard view, mass segregation in OCs is believed to increase with age (Kroupa, 1995; Dib et al., 2018). Old clusters are therefore believed to be more frequently mass segregated than young clusters. Thanks to our membership analysis over extended regions of OCs of various age, this is an hypothesis that we can try to verify.

3.6.1 Method

In order to measure the degree of mass segregation we applied the method proposed by Allison et al. (2009b) and widely used to quantify and detect mass segregation in stellar clusters (Nony et al., 2021; Dib et al., 2018; Plunkett et al., 2018; Román-Zúñiga et al., 2019). This method works by comparing the length of the Minimum Spanning Tree (MST) of the most massive stars of a cluster with the length of the MST of a set of the same number of randomly chosen stars. A MST of a set of points is the path connecting all the points, with the shortest pathlength possible and without any closed loops. In a given set of points, only one MST can be drawn. We computed the MST by using the `csgraph` routine implemented in the scipy python module (Virtanen et al., 2020). In all the cases, we drew the minimum spanning tree in the same set of coordinates as defined in Eq. 3.8. The Mass Segregation Ratio (MSR) Λ_{MSR} is then defined as follows:

$$\Lambda_{MSR}(N) = \frac{< l_{random} >}{l_{massive}} \pm \frac{\sigma_{random}}{l_{massive}}, \tag{3.10}$$

with $< l_{random} >$ being the average length of the MST of N randomly chosen stars and $l_{massive}$ the length of the MST of the N most massive stars. The average length $< l_{random} >$ was calculated over 100 iterations where at each iteration we draw a different subsample of random stars allowing us to calculate at the same time σ_{random}, the standard deviation of the length of the MST of these N stars. We used the G magnitude of the stars as a proxy for the mass. The mass segregation ratio Λ_{MSR} was always calculated for a subset of N stars. A result of Λ_{MSR} greater than 1 means that the N most massive stars are more concentrated compared to a random sample and therefore that the cluster shows a sign of mass segregation. We conducted this analysis for all the clusters containing more than 50 stars. In each cluster we calculated the MSR starting at $N = 5$ up to the number of cluster members. We started at $N = 5$ because for lower N the value is not statistically significant. For clusters with more than 100 stars, we stopped at $N = 100$ since the MSR only shows a gradual decrease to reach unity.

Figure 3.17 shows the mass segregation ratio for the cluster Collinder 394 for increasing values of N. There are several degrees of mass segregation. First the 20 most massive stars have a value for $\Lambda_{MSR} \sim 3.2$. There is then a drop of Λ_{MSR} with a plateau for $20 < N < 26$ around a value of ~ 1.7. The mass segregation ratio then drops to 1.4 and progressively decreases. This analysis tells us that in Collinder 394, the 20 most massive stars are 3.2 times closer to each other compared to the typical separation of 20 random stars in the cluster and that the 25 most massive stars of cluster are 70% more concentrated compared to any set of 25 members. After that, the rest of the stars progressively approach $\Lambda_{MSR} \sim 1$.

3.6.2 Discussion

Maschberger et al. (2010) studied the very early stages of clusters through N-body simulations and noted that the 10 most massive stars of a cluster quickly gather into a very

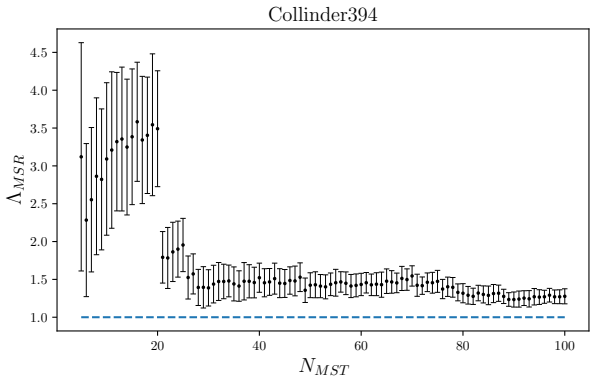

Figure 3.17: Mass segregation ratio $\Lambda_{MSR}(N)$ for the cluster Collinder 394 as a function of the number of stars used to draw the MST. The blue dotted line shows the limit $\Lambda_{MSR} = 1$ after which stars do not show any signs of mass segregation.

concentrated system after the clusters formed. We represent in Fig. 3.18 the distribution of the mass segregation ratio of the 10th most massive stars (Λ_{10}) as a function of their parent cluster ages, Galactocentric radii and altitude above (or below) the galactic plane. We do not notice any particular trend regarding the evolution of mass segregation with these parameters even though we would expect an increase of the mass segregation with age (Dib et al., 2018). For consistency, we also checked if some trend appeared when looking at the MSR of the fifth to the twentieth most massive stars but that was not conclusive.

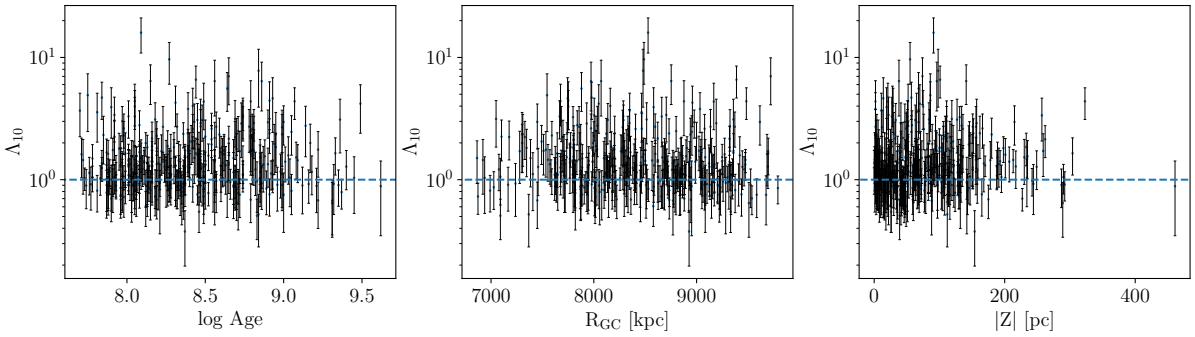

Figure 3.18: Mass segregation ratio Λ_{10} of the 10 most massive stars of each cluster as a function of OCs ages (left), Galactocentric radii (middle) and absolute value of the altitude above the galactic mid-plane (right). The blue dotted line shows the limit $\Lambda_{MSR} = 1$.

The lack of a net relation between Λ_{10} with age might be related to the fact that OCs formed with very different levels of mass segregation (Dib et al., 2018). For instance, the very young cluster Trapezium, the core of the Orion Nebula Cluster, shows evidence of mass segregation even though its age is $\sim$ 1 Myr (Bonnell and Davies, 1998; Allison et al., 2009b). This was referred to as primordial mass segregation (de Grijs et al., 2003). Mass segregation in young clusters was first thought to be caused by the initial conditions of the cluster formation, but recent N-body simulations suggested that mass segregation occurs on time scales of the order of a few Myr. This implies that clusters younger than their dynamical relaxation time can show signs of mass segregation. This could be due

to dynamical interactions through the merging of smaller substructures. In this scenario, clusters are born with a significant amount of clumps. Each of these small clumps can then mass segregate in short time scales through dynamical interactions. The merging of these multiple clumps later gives birth to a cluster that inherited the substructure's segregation (McMillan et al., 2007; Allison et al., 2009a, 2010; Maschberger et al., 2010). This gives a more complex view of what is expected to be observed in our sample of OCs.

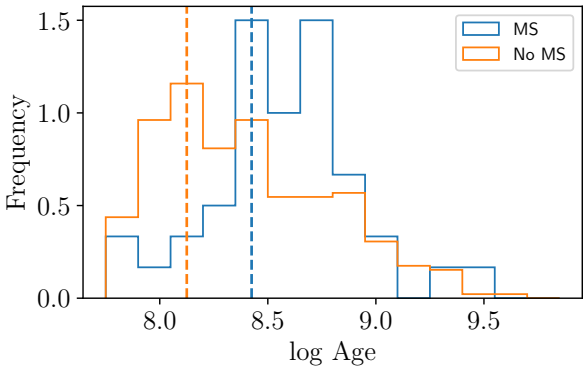

Figure 3.19: Age distribution of two subsamples of OCs: one where more than 10% of the stars have $\Lambda_{MSR} > 2$ (in blue) and one where less than 10% of the stars have $\Lambda_{MSR} > 2$ (in orange). The vertical dotted line shows the mode of each distribution.

In order to investigate in a different way the age dependency of MS, we measured the proportion of stars per cluster having $\Lambda_{MSR} > 2$. We defined two subsamples : one where more than 10% of the stars have $\Lambda_{MSR} > 2$ and the other one being the complementary subsample. For example, we can see in Fig. 3.17 that in Collinder 394, 20 stars have a MSR higher than 2. As it counts 703 members, only 2% of Collinder 394 stars have $\Lambda_{MSR} > 2$ and Collinder 394 is therefore part of the complementary subsample where less than 10% of the clusters members are highly mass segregated. The age distribution of these two subsamples is shown in Fig 3.19. Even if no trend between Λ_{10} and the cluster ages was noticeable in Fig 3.18, it is clear here that OCs which have a large proportion of stars showing a strong mass segregation are -in average- older than the clusters with few stars highly mass segregated.

This could be related with the signs of evaporation observed in Sect 3.5.2. In Fig. 3.14, we noted that old clusters have -in average- less stars in their halo than young ones. As old clusters are proportionally more mass segregated than young ones, more small stars should be pushed to the outskirts of the clusters, increasing the relative weights of the halo. As we observe the opposite, it could indicate that cluster evaporation process is very efficient compared to mass segregation.

3.7 Conclusion

The primary goal of this study was to determine the structural parameters of the clusters identified since the publication of *Gaia* DR2. We first tried to measure the OCs three dimensional shape but the number of clusters close enough to be studied with accuracy

in three dimensions is limited to the $\sim$ 17 OCs closer than 250 pc. So we focused on the two dimensional morphology of clusters for which we obtained new members at large radius. To do so, we fitted the RDP of each cluster with more than 100 members with a King function in order to study their core and tidal radii. We find similar core radii to the ones published in previous studies but as we find members at much larger radius than in previous studies, we also find much larger tidal radii. The distribution of the fitted core radii shows a concentration between 1 and 2.5 pc regardless of the age or of the number of cluster members. We see that older clusters tend to have smaller R_c compared to young clusters with values converging towards 1.85 pc at the age of 1 Gyr. The tidal radii peaks around 30 pc, but more importantly seems to increase with age in what could be a sign of mass segregation, dissolution or a combination of both processes. The tidal radius distribution could be biased due to the limit of 50 pc that we used to query the *Gaia*-eDR3 catalogue. A fraction of the investigated clusters may extend to larger radii than this, as shown in Sect. 2.3.4 with the example of COIN-Gaia 13. We also fitted an EFF and a Plummer function on the RDP of our cluster samples. The characteristic radius of the Plummer profile confirmed the decrease of the cores sizes with age while the γ parameter of the EFF profile showed that regarless of their sizes, the RDP of old clusters tend to decrease at the same rate at large radius.

We fitted GMMs on the spatial distribution of members projected on a plane perpendicular to the celestial sphere. This is particularly suited for the clusters for which we detected an elongated tidal tail as the King function previously used assumes a spherical distribution of members. We used a three components GMM on the 71 clusters with a tidal tail, one component representing the core of the cluster, one the tidal tail and one the corona. For the other clusters, we adopted a two component GMM with a core and a halo. We looked for correlations between the parameters of the fitted Gaussians with the characteristics of the clusters (i.e. their age, location, number of members, etc). For the 71 clusters of our sample with a tidal tail, we observe that old ones are more prone to have small cores than young ones. We also note that the relative weight of the corona of old clusters was in average lower than for young ones. This implies that with an increasing age, the proportion of stars in the cluster halos decreases, either because stars move to the center of the cluster or because outer stars are ejected from the cluster.

We applied the method proposed by Allison et al. (2009b) to measure the degree of mass segregation of our sample of OCs. We do not notice any trend of the mass segregation ratio, measured through the 10 most massive stars, with age, Galactocentric distance or with the altitude of the cluster above the Galactic mid-plane. However, clusters having a significant number of stars with a strong mass segregation ratio, are on average older than clusters with few stars strongly mass segregated. Coupled with a lower proportion of stars populating the clusters halos, this highlights the fact that the various physical processes in play in the disruption of clusters acts on shorter time scales than mass segregation.

Chapter 4

Kinematical properties of Open Clusters

Contents

This chapter is based on Tarricq et al. (2021a). This work has been conducted in collaboration with several co-authors. In this work, my contribution lies mainly in the compilation of data from different surveys to produce a uniform RV catalogue explained in Sect. 4.2.1, in the computation of the velocities used in Sect. 4.3, in the analysis of the age dependance of Galactic Velocities from Sect. 4.3.3, in the fit and analysis the AVR in 4.3.4 and in the integration of the orbits of the OCs and the subsequent discussion detailed in Sect. 4.4.1.

4.1 Context

The study of kinematical properties of the Milky Way OCs has a long tradition. Their motion can be used to understand the Milky Way gravitational potential and the various perturbations that act on the structure and dynamics of the Galactic disc. The solar neighbourhood is known to be clumpy (e.g. Eggen, 1996; Dehnen and Binney, 1998; Antoja et al., 2012), and the relation of observed substructures with some OCs, in particular the Hyades, is established (Eggen, 1958; Chereul et al., 1999). While stellar streams were initially thought to be remnants of star clusters, their origin is now thought to be the result of the disruption of star clusters or of dynamical origin. This dynamical origin can be the result of the resonant trapping by the bar and spiral arms (Famaey et al., 2008) or the passage of the Sagittarius dwarf galaxy within the Galactic plane (Monari et al., 2018; Khanna et al., 2019). Stellar streams could also originate from a combination of these processes. *Gaia* DR2 data (Gaia Collaboration et al., 2018b) revealed the complexity of the local velocity distribution with unprecedented resolution, showing in particular that the velocity distribution of nearby OCs overlaps prominent arched overdensities of moving groups well (see Fig. 4.1 from Gaia Collaboration et al., 2018d).

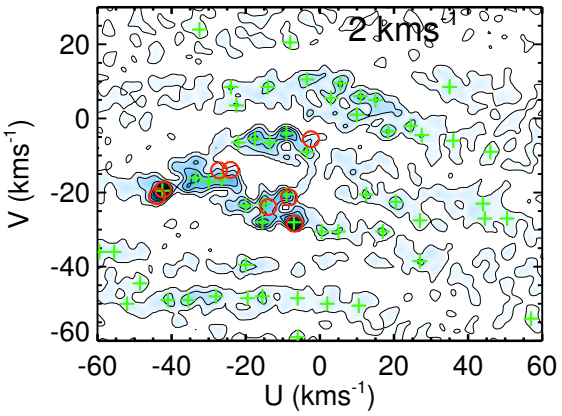

Figure 4.1: Wavelet transformation of the heliocentric velocity plane in the solar neighbourhood showing the substructures at a 2 $\mathrm{km\,s^{-1}}$ scale. The green crosses represents the peaks more significant than 3σ. The red circles stand for OCs closer than 200 pc. From Gaia Collaboration et al. (2018d).

The physical and kinematical properties of OCs have also been revisited based on *Gaia*. Using the memberships from CG+18 and only *Gaia* data, Bossini et al. (2019) determined the age, distance modulus, and extinction of 269 OCs, and Soubiran et al. (2018a) computed the 6D phase-space information of 861 clusters. Carrera et al. (2019a) increased the number of OCs with full 6D phase-space information by 145 by searching for members in the GALac-tic Archaeology with HERMES (GALAH) and APOGEE spectroscopic surveys. The vertical distribution of young clusters was found to be very flat, with a dispersion of vertical velocities of 5 km s^{-1}, while clusters older than 1 Gyr span distances to the Galactic plane of up to 1 kpc with a vertical velocity dispersion of 14 km s^{-1}, typical of the thin disc. Recently, CG+20 derived physical properties of all known OCs identified in the *Gaia*-DR2 data in a homogeneous fashion, with a method based on isochrone fitting and an artificial neural network. They finally assembled 1867 clusters with reliable ages. In the following, we take advantage of these latest age determinations of clusters to study the correlations of the 6D phase-space coordinates and orbital properties with age. We also investigate the phase-space distribution, AVR, and action distribution to compare them to those of field stars.

4.2 Radial Velocity of OCs

We took advantage of the catalogue of OCs by CG+20, which provides a list of probable members for 2017 OCs that the authors used to estimate ages. Most of these OC memberships are taken from the 1481 clusters studied by Cantat-Gaudin and Anders (2020), who used the unsupervised classification scheme UPMASK (Krone-Martins and Moitinho, 2014; Cantat-Gaudin et al., 2018a). These authors also applied UPMASK to the clusters discovered by Liu and Pang (2019) and collected the list of members provided for the recently discovered University of Barcelona Clusters (UBC) by Castro-Ginard et al. (2018, 2019, 2020). For the Hyades and Coma Berenices, they adopted the list of members published by Gaia Collaboration et al. (2018a) because UPMASK cannot recover members for populated clusters that are too extended on the sky.

4.2.1 Gathering measurements from different surveys

Based on this list of members that contains $\sim$ 475 000 stars, we gathered all the RV measurements available for them in various surveys and catalogues. Our main source was *Gaia* DR2, which includes RVs for about seven million stars (Sartoretti et al., 2018; Katz et al., 2019). We also queried several catalogues from large spectroscopic surveys in addition to *Gaia*: the latest public version of the *Gaia*-ESO survey (Randich et al., 2013), APOGEE DR16 (Ahumada et al., 2020), Radial Velocity Experiment (RAVE) DR6 (Steinmetz et al., 2020), and GALAH DR3 (Buder et al., 2021). We did not consider Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST) (Cui et al., 2012; Xiang et al., 2015) because its RV precision and accuracy ($\sim$ 5 km s^{-1}) are not at the same level as the other mentioned catalogs. We included RVs derived in the Open Cluster Chemical Abundances from Spanish Observatories (OCCASO) survey (Carrera et al.,

2021; Casamiquela et al., 2016). We also considered the RV catalogues by Soubiran et al. (2018b); Mermilliod et al. (2009, 2008); Worley et al. (2012); Nordström et al. (2004). Several quality cuts were applied to individual measurements. For RAVE we applied the criteria suggested by Steinmetz et al. (2020) and selected stars that met |correctionRV| <10 $\mathrm{km\,s^{-1}}$, $\sigma_{\mathrm{RV}} < 8\ \mathrm{km\,s^{-1}}$ and correlationCoeff >10. For APOGEE, we rejected the stars flagged as VERY_BRIGHT_NEIGHBOR, BAD_PIXELS or LOW_SNR as recommended in the online documentation of the survey. For *Gaia* RVS we filtered the erroneous RVs found by Boubert et al. (2019). Despite these cuts, some individual measurements still had large uncertainties that were incompatible with the precision we required. This convinced us to filter out all the individual RVs with uncertainties larger than 8 $\mathrm{km\,s^{-1}}$, which is the same cut as for RAVE. The rejected values represent 6% of the full set, and 10% for the RVS set. Then we realised that 43 stars had $|\mathrm{RV}| > 200\ \mathrm{km\,s^{-1}}$ which were mainly OB-type, Wolf-Rayet, or spectroscopic binary stars according to Simbad. These stars with unreliable or variable RVs were rejected.

Table 4.1 gives the number of cluster members retrieved in each of the catalogues after filtering, with the median uncertainty of the corresponding RVs, as quoted in the catalogues. Soubiran et al. (2018b) reported that the uncertainty corresponds to the standard error of the weighted mean for stable stars that were followed-up for exoplanet detection. The high precision of individual measurements and the number of observations explain the very low median uncertainty of that catalogue (0.002 $\mathrm{km\,s^{-1}}$). APOGEE, OCCASO, and the catalogues from Mermilliod et al. (2009, 2008) and Nordström et al. (2004) also contain stars that have been observed several times in order to identify binaries. In that case, the RV uncertainty corresponds to the quadratic sum of the single measurement error with the scatter of the measurements. RAVE also has multiple observations for a small fraction of stars and provides the individual values with their uncertainty corresponding to the error of a single measurement. Finally, the *Gaia*-ESO survey includes RVs of the same star obtained with different setups that give different measurement errors (Jackson et al., 2015).

4.2.2 Zero-points of RV catalogues

This sample allowed us to assess the consistency of RVs in the different catalogues. The comparison of RVs for stars in common in two catalogues gives an idea of potential offsets due to zero-point differences, together with their typical precision. Zero-point differences are a result of the different observing modes, instrumental characteristics, and calibration procedures of each instrument. They have to be taken into account when RVs of different origins are combined. A subset of 3 116 stars has measurements in two or more catalogues. The RV difference between catalogues that have more than 20 stars in common is presented in Fig. 4.3, and the corresponding comparison of catalogues is listed in Table 4.2. The RVs agree well in general, with offsets smaller than 0.5 $\mathrm{km\,s^{-1}}$ and between 0.5 and 1.4 $\mathrm{km\,s^{-1}}$ for comparisons that involve RAVE or GALAH. The dispersions (measured by the median absolute deviation MAD) are typically about 1 $\mathrm{km\,s^{-1}}$ or lower, consistent with the precision of the catalogues listed in Table 4.1 for all

Table 4.1: Number of stars in common in the catalogue of CG+20 and catalogues of RVs, with the corresponding median uncertainty (in $\mathrm{km\,s^{-1}}$), after the quality cuts described in the text. Catalogues are designated by short names as defined below the table.

survey / catalogue	N_{stars}	median RV error
RVS	10741	1.33
GES	9894	0.50
APO	3212	0.21
GAL	1724	0.59
MER	1313	0.28
RAV	386	1.59
OCC	187	0.04
S18	138	0.002
NOR	73	0.60
WOR	10	0.20

RVS: *Gaia* DR2; GES: *Gaia* ESO survey; APO: APOGEE DR16; GAL: GALAH DR3; MER: Mermilliod et al. (2008, 2009); RAV: RAVE DR6; OCC : OCCASO Casamiquela et al. (2016); S18: Soubiran et al. (2018b); NOR: Nordström et al. (2004); WOR : Worley et al. (2012)

combinations of surveys. In these comparisons of the different catalogues, we did not see any significant trend with colour or apparent magnitude.

Most of the RVs available for the cluster members are provided by *Gaia* DR2 or the *Gaia* ESO survey. A large fraction of the GES observing program is dedicated to OCs and nicely increases the number of clusters for which a RV can be computed. *Gaia* and GES have an offset of ~ 0.3 $\mathrm{km\,s^{-1}}$, similar to the offset between GES and APOGEE, and *Gaia* and APOGEE agree at a level better than 0.1 $\mathrm{km\,s^{-1}}$.

The zero-point of RAVE is found to be different from that of the other surveys by $\sim$ 1 $\mathrm{km\,s^{-1}}$ (RAVE underestimates the RVs compared to the others), which is larger than reported in previous studies. Steinmetz et al. (2018, 2020) and Sartoretti et al. (2018) reported an offset of ~ 0.3 $\mathrm{km\,s^{-1}}$ between RAVE and *Gaia* DR2. Even when we applied strong quality cuts on both RAVE and *Gaia* DR2, the median difference remained 1 $\mathrm{km\,s^{-1}}$. RAVE also differs from APOGEE by 1.42 $\mathrm{km\,s^{-1}}$, and from Mermilliod et al. (2008, 2009) by 0.79 $\mathrm{km\,s^{-1}}$. The agreement with S18 is better at a level of 0.2 $\mathrm{km\,s^{-1}}$, however, which might be related to the bright magnitude of the common stars and their colour range. We also note a systematic offset of GALAH DR3 with all the other catalogues, opposite to that of RAVE and of smaller amplitude, but still larger than the value of 0.22 $\mathrm{km\,s^{-1}}$ reported by Buder et al. (2021) for the comparison to *Gaia* DR2. These larger offsets might be related to the fact that our star sample is dominated by dwarfs, with a fraction of early-type stars larger than in RAVE and GALAH in general. For RAVE, we see indeed a trend: the offset is slightly larger for hot stars than for cool stars, as illustrated in Fig. 4.4. In order to set all the RVs on the scale of *Gaia*, we applied a zero-point correction to the individual RVs from all the non-*Gaia* catalogues, according to the offsets listed in Table 4.2.

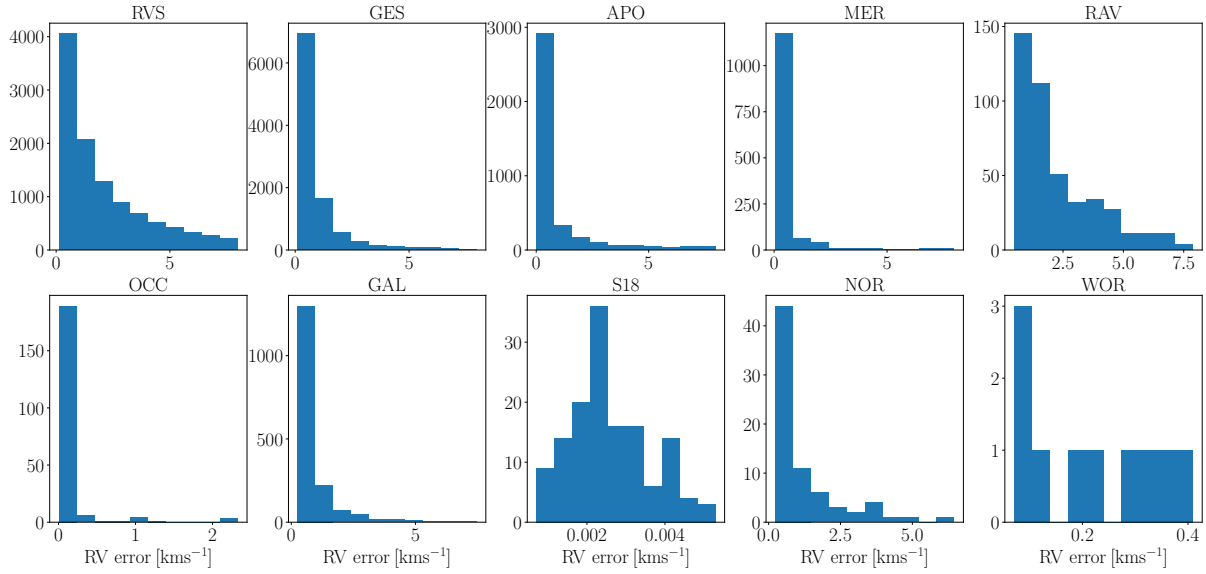

Figure 4.2: Distribution of the RV uncertainties of the OC members in each catalogue, designated by short names as listed in Table 4.1

4.2.3 Mean RVs

In order to compute the mean RV of each OC, we first computed the mean RV of each star, because some of them (12%) have multiple measurements from different catalogues. For both the mean per star and the mean per cluster, we used a weighted procedure based on the errors of individual measurements, following Soubiran et al. (2013, 2018a). For each star or each cluster, the mean RV was computed by attributing to each RV measurement a weight ω_i defined as $\omega_i = 1/\varepsilon_i^2$, where ε_i is the RV error. Outliers were rejected on the basis of a 3σ clipping. A fraction of about 8% of the stars with multiple measurements have an RV uncertainty larger than 3 $\mathrm{km\,s^{-1}}$ and we suspect that they are binaries with variable RVs. These stars were not rejected, but the procedure gives them less weight when the mean RV of their parent cluster is computed. However, in the large majority of cases, we have only one measurement per star so that binaries cannot be identified. Binaries that have large variations in RV may alter the mean RV of the parent OC when there are only a few members.

Our catalogue of stars provides the mean RV in the *Gaia* RVS scale, with its uncertainty and the number of measurements, as well as the membership probability from CG+20. It includes 23 424 unique stars, 97 of which appear twice owing to a non-null probability to belong to two different clusters that are located close to each other on the sky.

In order to compute the mean RV per OC, we considered only the stars with a membership probability higher than 0.4. This is the threshold value found by Soubiran et al. (2018a) to be the best compromise between the largest number of members and the lowest contamination by field stars.

In the end, 1 382 OCs have a mean RV, which represents an improvement by 60% compared to the previous RV catalogue of OCs that was based only on *Gaia*-RVS data

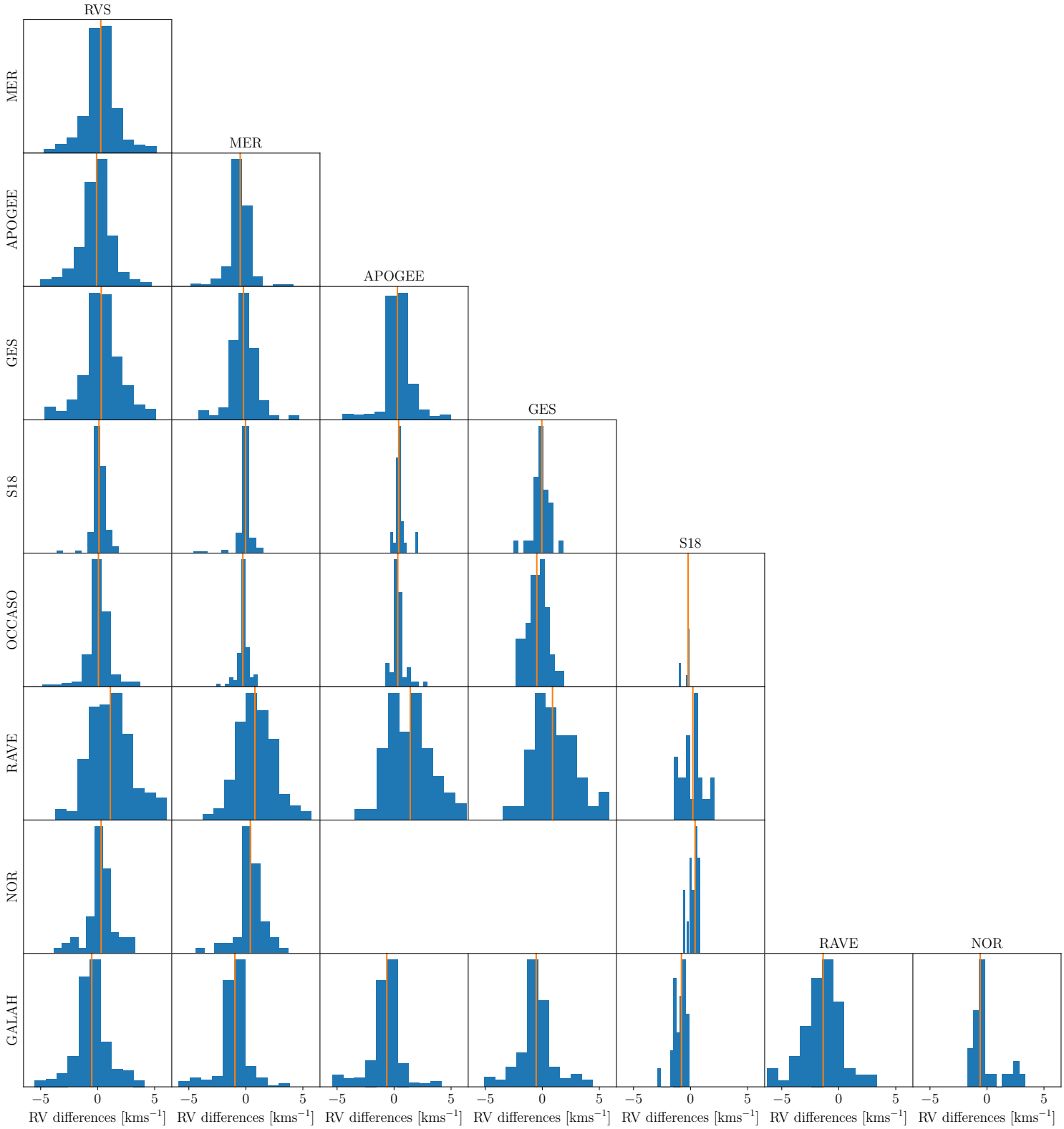

Figure 4.3: Histograms of the RV difference of stars in common in several surveys (zoomed in the range from -5 to 5 km s^{-1}). The solid orange line corresponds to the median value. The statistics of the comparisons are given in Table 4.2.

Table 4.2: Median (MED) and MAD of RV differences for stars observed by two surveys with the number of stars in common. Short names of the catalogues are the same as in Table 4.1.

catalogues	$N_{\rm stars}$	MED	MAD
RVS - MER	1079	0.28	0.68
RVS - APO	784	-0.09	0.86
RVS - GES	486	0.31	1.06
RVS - S18	115	0.12	0.26
RVS - OCC	161	0.07	0.40
RVS - RAV	299	1.12	1.36
RVS - GAL	607	-0.51	0.88
RVS - NOR	65	0.31	0.49
MER - APO	304	-0.50	0.33
MER - GES	152	-0.21	0.72
MER - S18	106	-0.03	0.16
MER - OCC	82	-0.26	0.17
MER - RAV	198	0.79	1.13
MER - GAL	293	-0.96	0.49
MER - NOR	63	0.39	0.47
APO - GES	358	0.29	0.43
APO - S18	30	0.40	0.10
APO - OCC	65	0.33	0.13
APO - RAV	125	1.42	1.60
APO - GAL	534	-0.64	0.59
GES - S18	30	-0.04	0.41
GES - OCC	41	-0.48	0.66
GES - RAV	57	0.89	1.89
GES - GAL	281	-0.54	0.78
S18 - OCC	21	-0.22	0.04
S18 - RAV	24	0.20	0.71
S18 - GAL	29	-0.80	0.36
S18 - NOR	20	0.40	0.23
RAV - GAL	97	-1.37	1.43
GAL - NOR	28	0.60	0.27

(Soubiran et al., 2018a). It also supersedes the catalogue by Carrera et al. (2019a) that was based on *Gaia*, GALAH, and APOGEE. Before *Gaia* DR2, the two largest compilations of cluster RVs were those of Kharchenko et al. (2013) and Dias et al. (2002), with 962 and 703 objects, respectively.

This is the first determination of the mean RV for most of the recently discovered OCs. In particular, half of the UBC clusters (Castro-Ginard et al., 2020) are part of our catalogue, and they represent nearly 20% of the full sample of OCs with a known RV. Of the 75 high-confidence clusters recently discovered by Liu and Pang (2019), 35 are present in the latest catalogue from CG+20, and we provide a RV for most of them. For instance, we provide the mean RV of UBC 274 (also reported as LP 5), a recently discovered disrupting OC (Castro-Ginard et al., 2020; Piatti, 2020), based on 18 members observed with RAVE, GALAH, and *Gaia* RVS (Fig. 4.5). By combining *Gaia* RVS and GES, several OCs now have RVs that are based on hundreds of members. Trumpler 5 and

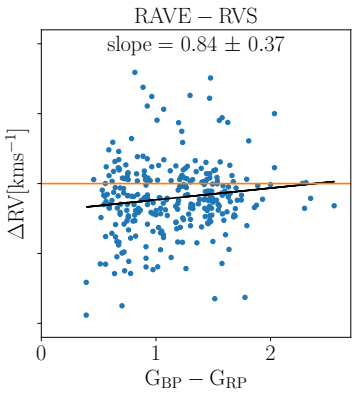

Figure 4.4: RV difference of stars in common between RAVE and RVS as a function of the color of the stars. We overplotted a simple least square linear regression and show the slope of the curve.

NGC 3532 are based on 659 and 664 stars, respectively. For about 18% of the sample, the mean RV is based on more than 10 stars, and for 50% it is based on at least 3 stars. The median uncertainty of the weighted mean RV is 1.13 km s^{-1} when the full sample is considered.

Of the clusters with fewer members, some exhibit a large error that renders their mean RV uncertain. The RVs of 430 OCs are based on only one star, with no information on its potential variability (binarity). This represents 31% of our sample, nearly the same proportion as in Soubiran et al. (2018a). Selecting the most reliable OCs that have a RV uncertainty lower than 3 km s^{-1} based on at least three stars, we obtain 513 clusters. This sub-sample has a median uncertainty of 0.55 km s^{-1} and a median number of nine stars. This means that we have 107 more reliable OCs than in Soubiran et al. (2018a).

The full catalogue includes 21 high-velocity OCs with $|\mathrm{RV}| > 100$ km s^{-1} but 13 are based on a single star and are therefore to be considered with caution. The 8 remaining high-velocity clusters with more members were known before.

The catalogue of OC velocities (released as online data[1]) provides the RV per cluster with its uncertainty and the number of members used on average. It also lists Galactic velocities, orbital parameters, and the actions computed in the next sections.

4.3 Open clusters in phase space

In this section, we combine the mean OC positions, distances, proper motions from CG+20, and our mean RVs to compute heliocentric and Galactocentric Cartesian and cylindrical positions and velocities for the full sample of 1382 OCs. For positions, the same conventions and reference values as Gaia Collaboration et al. (2018d) were used: in Cartesian Galactocentric coordinates, the Sun is located at $X = -8.34$ kpc, $Y = 0$ pc, and $Z = 27$ pc from the centre of the Galaxy. Similarly, we set the azimuthal velocity V_c at the solar radius at 240 km s^{-1}. The velocity of the Sun with respect to the Local

[1] https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/647/A19

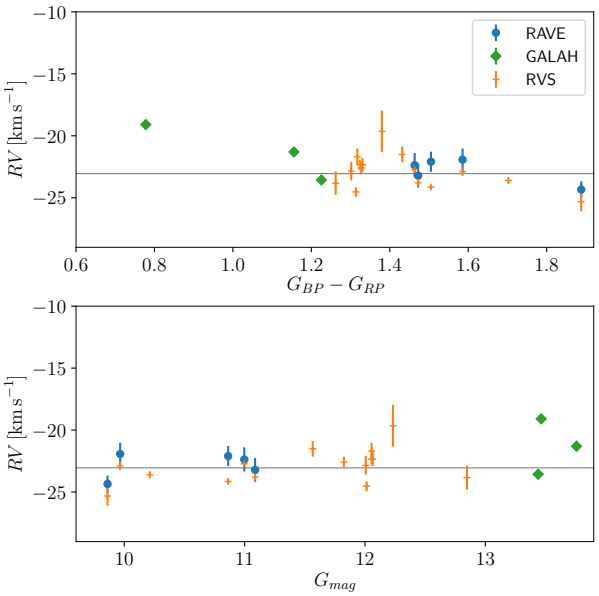

Figure 4.5: Individual RV of UBC 274 members (proba ≥ 0.4) from RAVE in blue, from GALAH in green, both after the zero-point correction indicated in Table 4.2, and from *Gaia*-RVS in orange, as a function of the $G_{BP} - G_{RP}$ colour and G magnitude from *Gaia*. The grey line represents the weighted mean, RV=-23.05±0.25 $\mathrm{km\,s^{-1}}$, based on 18 members.

Standard of Rest (LSR) is set to $(U, V, W) = (11.1, 12.24, 7.25)$ $\mathrm{km\,s^{-1}}$ (Schönrich et al., 2010). The Cartesian heliocentric velocities (U, V, W) are respectively oriented towards the Galactic center, the direction of the Galactic rotation, and the North Galactic pole. In the Galactocentric frame $(R, \phi, Z, V_r, V_\phi, V_z)$, the azimuthal coordinate ϕ is oriented in the direction of the Galactic rotation and its origin is defined as the line drawn by the Sun and the Galactic center. V_ϕ is oriented in the same direction as ϕ and stars having positive V_r move in the direction opposite to the Galactic center. To calculate the uncertainties in the velocity space, we used a Monte Carlo sampling of the astrometric and RV measurements and their uncertainties, which we assumed to be Gaussian. In the following, we consider the standard deviations of the Monte Carlo samples as the uncertainties. We then obtain the full sample of OCs with median standard deviations in the cylindrical velocities of $(\delta V_r, \delta V_\phi, \delta V_z) = (2.8, 3.2, 1.6)$ $\mathrm{km\,s^{-1}}$.

For distant OCs, the astrometric uncertainties translate into large Galactic velocity errors. For 129 OCs we find that one component of the Galactic velocities has a standard deviation higher than 100 $\mathrm{km\,s^{-1}}$. These 129 clusters are all located at distances from the Sun larger than 2.2 kpc. The most extreme cases are NGC 3105 and SAI 109, which have a relative parallax error of 82 and 99%, respectively, and are both located at 6.9 kpc. For these two clusters, the velocity uncertainties reach several thousand $\mathrm{km\,s^{-1}}$ in V_ϕ and V_z. Moreover, as mentioned in the previous section, some OCs do not have a fully reliable mean RV because they only have a few members, lack information about their stability, or because the dispersion among the members is large.

We therefore defined a High-Quality Sample (HQS) composed of the OCs with a reliable mean RV (uncertainty lower than 3 $\mathrm{km\,s^{-1}}$ based on at least three members as defined in the previous section) and with standard deviations on (V_r, V_ϕ, V_z) below 10

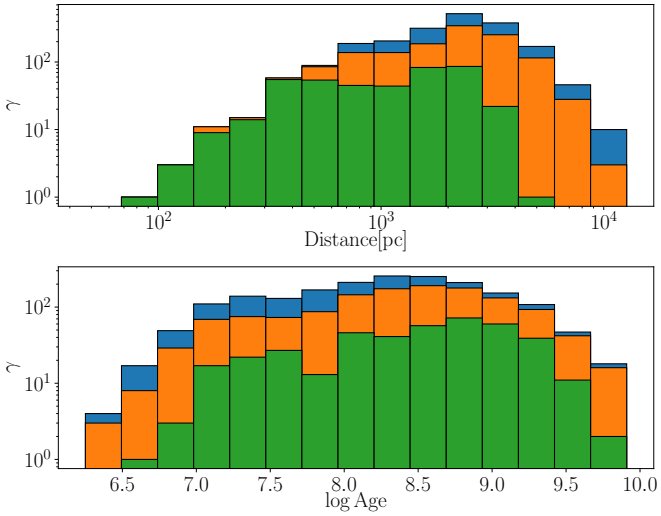

Figure 4.6: Histogram in log scale of the heliocentric distance (top panel) and of the age distribution (bottom panel) for the full sample of OCs from CG+20 in blue, the sub-sample of OCs for which we have a RV measurement in orange, and the HQS in green.

$\mathrm{km\,s^{-1}}$. This HQS is composed of 418 OCs with median uncertainties on (V_r, V_ϕ, V_z) of (1.2, 1.3, 1.0) $\mathrm{km\,s^{-1}}$. This is very similar to the median uncertainty of the field star sample selected by Gaia Collaboration et al. (2018d), who quoted $(\delta V_r, \delta V_\phi, \delta V_z) = (1.4, 1.4, 0.9)$ $\mathrm{km\,s^{-1}}$ for stars with $\varpi/\sigma_\varpi > 5$. For 411 out of these 418 clusters, a homogeneous estimation of their ages is provided by CG+20.

Figure 4.6 shows the histogram of the heliocentric distance and logarithmic ages of our different samples. Our subsample of 1382 OCs with RV measurements is 90% complete with respect to the total sample of 2017 OCs from CG+20 up to 860 pc, and our HQS is 90 % complete with respect with the total sample up to 500pc.

The bottom panel shows that except for the very young clusters, all ages are represented in the HQS, which indicates that our selection does not introduce a noticeable bias in terms of the age distribution for clusters older than 10 Myr. The kinematical selection based on the RV uncertainty removed most of the very young clusters because the RV of young stars is less reliable.

4.3.1 Kinematics of OCs compared with field stars

In this section, we compare the kinematics of the field stars sample from Gaia Collaboration et al. (2018d) with that of the OC sample. In Fig. 4.7 we compare the (V_ϕ, V_r) distribution of the nearby OCs within 500 pc and 200 pc, with the field stars from *Gaia* DR2 closer than 200 pc and having a relative error in parallax < 0.05. In this space, field stars from the solar neighbourhood show clear substructures in form of arches and clumps, which are associated with resonances due to non-axisymmetric features of the Galaxy, as discussed in Gaia Collaboration et al. (2018d) and subsequent works. Large clumps were found to be associated with previously known moving groups (Eggen, 1958; Dehnen and Binney, 1998; Chereul et al., 1999; Nordström et al., 2004), and several over-

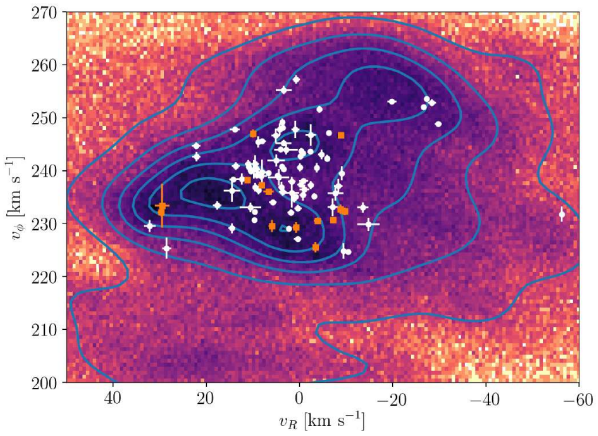

Figure 4.7: (V_ϕ, V_r) distribution of field stars in the solar neighbourhood (closer than 200 pc) compared with clusters closer than 500 pc (white) and 200 pc (orange).

densities of field stars in this space have been linked with OCs before (Gaia Collaboration et al., 2018d). The general trend of our sample of clusters closer than 500 pc is to overlap with the higher density regions of field stars, as previously reported in Soubiran et al. (2018a). The several OCs that we obtained here fall in the central part of the velocity distribution. Although we do not see a clear clumping in the cluster distribution, OCs appear to be most frequent in a band just between the two central arches of the field stars. This shows that OCs do not follow the exact distribution of overdensities drawn by the field stars. This was previously reported based on a smaller number of clusters in Gaia Collaboration et al. (2018d), who asserted that only the Pleiades and Hyades clusters are associated with overdensities, and that other clusters do no show a particular overdensity in this space.

In Fig. 4.8 we plot the distribution of azimuthal velocities as a function of Galactocentric radius of the full sample of OCs and the HQS. Field stars in this figure were retrieved from the *Gaia* archive by selecting stars with the same criteria as Gaia Collaboration et al. (2018d), providing a sample of 6,376,803 stars. In this space, as seen in Sect. 1.3.2, Antoja et al. (2018) showed that the diagonal ridges can be signatures of phase-mixing after a perturbing event, or alternatively, can be due to resonances of a barred potential or of the spiral arms. Kawata et al. (2018) argued that many of the diagonal ridges are likely related with the perturbations from the bar outer Lindblad resonance and spiral arms, which may be explained with the transient spiral arm scenario. Dias et al. (2019) and Barros et al. (2020) claimed that spiral resonances are able to trap stars or OCs orbits inside the corotation radius and associated the corotation with some ridges. In our plot, the clusters seem to qualitatively follow the distribution of diagonal overdensities described by the ridges, at least the most prominent ones. The recent N-body simulations by Khanna et al. (2019) suggest that the ridges are more prominent for stars close to the plane with solar metallicity; these conditions are generally met by a sample of OCs. Probably because of the shape of these ridges, the velocity gradient of V_ϕ decreases towards large radii, which is more clearly visible in the plot that shows the entire cluster sample. This is consistent with the top middle panel of Fig. 4.12, where outer clusters (yellow) have lower velocities (and in general have older ages, see the his-

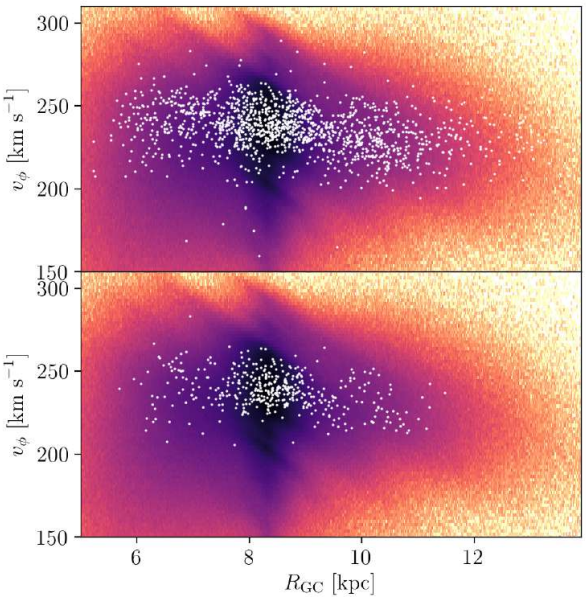

Figure 4.8: (V_ϕ, R_{GC}) distribution of field stars coloured by density, compared with clusters in white. In the top panel we plot the full sample of clusters, and in the bottom panel we show the HQS.

togram). As discussed in more detail in the next subsection, this produces a dip in the rotation curve that is traced by our sample of OCs.

4.3.2 Rotation curve of the Milky Way

We represent in Fig. 4.9 the azimuthal velocities of the full sample and the HQS OCs as a function of the Galactocentric radius, superimposing a theoretical rotation curve extracted from the axisymmetric gravitational potential described in Sect. 4.4. Our sample of OC data describes the rotation curve of the Milky Way for a mostly young population in the range $R_{GC}\sim$ [6.5,10.5] kpc with good precision. In the bottom panels we represent the median value of the rotation velocity in bins of approximately 400 pc. Uncertainties are computed as the standard error of the median: $\sqrt{\pi/2}\cdot\sigma/\sqrt{N}$ (Maindonald and Braun, 2010). We only plot the points computed using more than 20 OCs, which we consider to be reliable estimates and not dominated by outliers.

The HQS and the full sample show a similar behaviour. The samples overlap with what is expected from the axisymmetric curve ("MWPotential2014"). Several bins in both samples depart from the theoretical curve and show a small dip towards the inner Galaxy ($R_{GC}\sim$ 7 kpc), and a significant dip at $R_{GC}\sim$ 9.7 kpc, which departs by more than 2σ from the theoretical curve. In the full sample, the outer dip extends to 11 kpc. Gaia Collaboration et al. (2018d) reported similar dips in the sample of small Z (their fig. 13), but the inner dip is only seen at positive azimuths. In the bottom panels, we overplot the unified rotation curve computed by Sofue (2020), which was obtained from a combination of data from different sources in the literature. Interestingly, these data show a significant drop around 10 kpc, although with a different depth than our values.

A dip in the region $R_{GC}\sim$ 9.5 kpc has been reported variously before (Reid et al.,

2019; Barros et al., 2016; Sofue et al., 2009), but its nature is not fully understood. It has been tentatively explained by a ring density structure observed in the neutral H gas slightly farther out than the solar radius ($R_{\rm GC}\sim$ 10 kpc Nakanishi and Sofue, 2016), which successfully reproduces the dip. Dias et al. (2019) calculated a corotation radius of 8.51 ± 0.64 kpc and found a similar dip in their rotation curve, which can therefore be attributed to the effect of corotation. Clusters located outside the corotation radius have a slower azimuthal velocity than clusters located inside. The depth of the dip is different depending on the data used, the dip we observe here is compatible with the recent study by Reid et al. (2019) from star-forming regions: a small decrease from the model of ~ 5 km s^{-1}. Moreover, the general decreasing slope in the outer Galaxy shown by our data is also seen in the star-forming regions. Other studies from Cepheids (e.g. Mróz et al., 2019) also showed a slight decrease in V_ϕ before and after the solar radius. McGaugh (2019) built a model that reconciled the observed stellar rotation curve with the curve seen by interstellar gas, taking the overdensities of the spiral arms into account. In their Fig. 4, the modelled points (green) show a clear dip inside the solar radius that is compatible with the dip we see in the cluster data in Fig. 4.9.

Compared with Fig. 4.8, these dips are a direct consequence of the shape of the most prominent diagonal ridges, which are clearly shown by field stars in the *Gaia* RVS, but also by our clusters, where a remarkable decreasing velocity gradient is seen towards the outer Galaxy. This effect was analysed in Martinez-Medina et al. (2019) using a non-axisymmetric Galactic model that showed that the two bar-spiral arms produce diagonal-like ridges of stars of constant angular momentum. They showed that these structures tend to clump the stars in the $V_\phi - R_{GC}$ plane, pulling them up at the beginning of the ridge, and down at the end. This translates into wiggles in the rotation curve of the Galaxy. We overplot the rotation curve derived from the model by Martinez-Medina et al. (2019)[2] in Fig. 4.9. Interestingly, the model resembles the two dips we observed in the OC data. We attribute this to the same phenomenon as the dips we observe in the cluster data, thus explained by the effect of non-axisymmetric structures in the angular momentum of the stars.

4.3.3 Age dependence of Galactic velocities

Figure 4.10 shows the distribution of the HQS in the (V_r, V_ϕ), (V_r, V_Z), and (V_ϕ, V_Z) planes coloured by age. Several distant old OCs with extreme velocities ($V_r > 40$ km s^{-1}) have recently been discovered. Those included in the HQS are LP 930, UBC 326, and UBC 324 (highlighted with blue circles in Fig. 4.10), which are all located farther than 1 kpc in heliocentric distance. The most extreme OCs in total velocity ($\sqrt{V_r^2 + V_\phi^2 + V_z^2}$) are Ruprecht 171 and Haffner 5 (highlighted with green circles in Fig. 4.10), which are known as high-velocity objects for the disc.

We show in Fig. 4.11 the distribution of V_Z as a function of age in four bins of $R_{\rm GC}$. We colour the points according to the Z position. In all bins we see a distribution centered

[2] we have scaled the rotation curve of the model to fit the reference frame used in our work, described in Sect. 4.3, using the expression $V_{\rm new} = V_{\rm old} + \frac{R_\odot}{8.5}(\theta_\odot - 214)$, which was also used in Mróz et al. (2019).

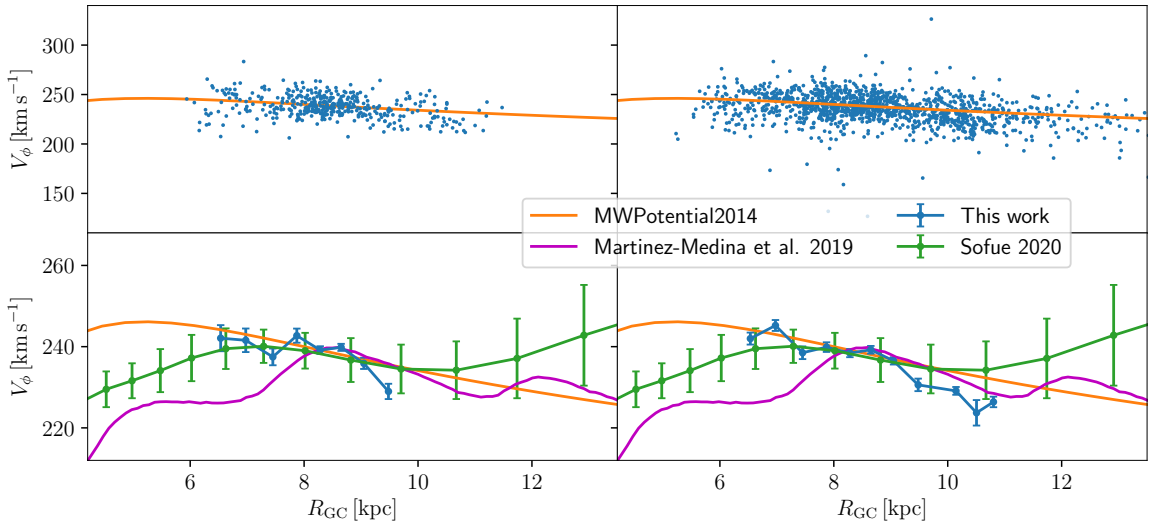

Figure 4.9: Top: Azimuthal velocities of all the OCs in the full sample (right) and the HQS (left) as a function of R_{GC}. The rotation curve of the gravitational potential in Sect. 4.4 is superimposed. Bottom: Median azimuthal velocities in bins of R_{GC}. We also superimpose the unified rotation curve computed by Sofue (2020) and the non-axisymmetric model by Martinez-Medina et al. (2019).

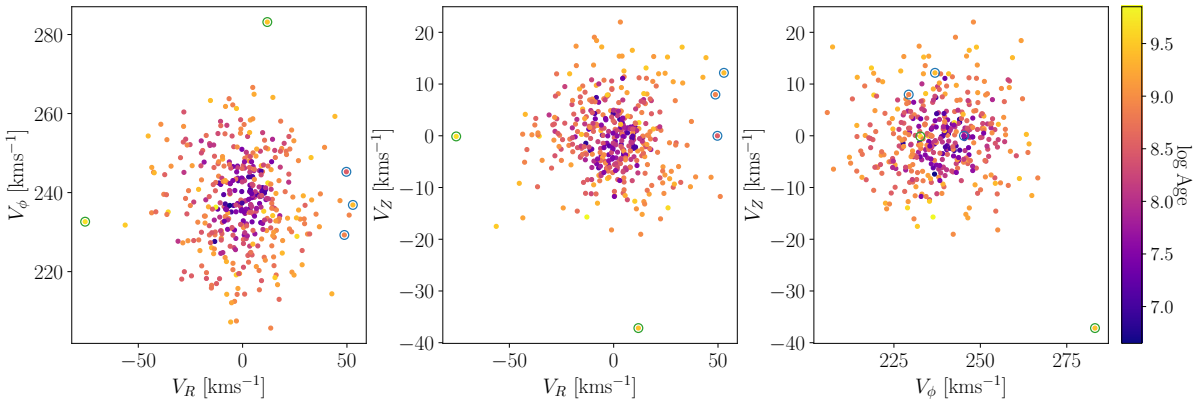

Figure 4.10: Left: Distribution of the HQS in Galactic velocities (V_r, V_ϕ) with colour related to age. Middle: Same with the velocities (V_r, V_Z). Right: Same with the velocities (V_ϕ, V_Z). We highlighted with blue circles the clusters.

on zero, with an increasing dispersion as a function of age, which means that in general, the oldest clusters are those reaching larger V_Z. In the outer disc, the dispersion of Z of the oldest clusters (211 pc) is much larger than the dispersion in the inner disc (86 pc). This implies that we are seeing some of the outer disc clusters in a point of their orbit far from the plane, and others crossing the plane at present time. However, in the inner disc this dispersion is much smaller, and the majority of clusters are confined in the plane, even those with large V_Z. This effect is probably an observational bias and agrees well with what is seen in Fig. 9 from CG+20 where inner clusters are also restricted around the Galactic plane while outer clusters can be located further away from it.

The three components of the velocities of the HQS clusters as a function of age are

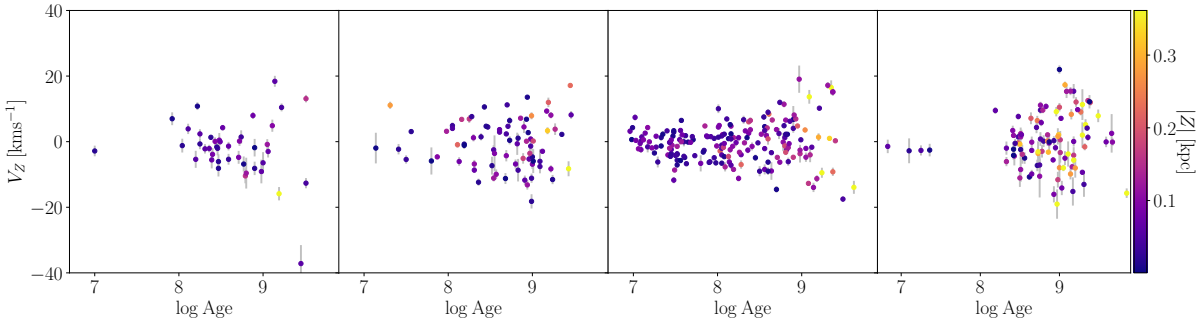

Figure 4.11: Vertical component of the velocity as a function of log Age in four bins of Galactocentric radius for the HQS. The colour corresponds to the Z coordinate.

shown in the top panel of Fig. 4.12. Clearly, they show a significant growing spread as age increases, which is related with the kinematic heating and is discussed in more detail in Sect 4.3.4. The radial and vertical velocities are centred on zero, and the azimuthal velocity is centred on the azimuthal velocity of the LSR. In addition, the middle panel of the top row shows that the azimuthal velocity of the oldest clusters within a given age range decreases when the radius increases. This trend is only noticeable for old clusters. Our kinematical selection based on the RV quality and errors of the Galactic velocities removed most of the young distant clusters of our sample. This bias was previously highlighted by Gaia Collaboration et al. (2018d): RVs for young stars are often less reliable because their temperature is high, because they rotate, or because they are closer to the Galactic plane and cannot be observed spectroscopically because of heavy extinction. Inner clusters, moreover, (in blue) are preferentially old, with a wide range of azimuthal velocities, a dozen of them have $V_\phi < 220$ km s^{-1}. For the vertical velocity shown in the right panel of the top row, the increase in dispersion highlights the fact that the clusters that reach higher V_Z are in general the oldest ones.

The bottom panels of Fig. 4.12 show the histograms of each component of the velocities in several age bins. The youngest clusters form a quite symmetric distribution, centred near zero for the radial and vertical components (median value of $V_r = -0.3$ km s^{-1} and $V_Z = -1.3$ km s^{-1}) and near the azimuthal velocity of the LSR for the azimuthal component (median value of $V_\phi = 239.5$ km s^{-1}). In the radial component, as age increases, the distribution becomes less Gaussian and the peak of the histogram tends to move towards larger V_r, although this is dependent on the binning of the histogram and the age limits that were assumed. Also depending on the binning, there seems to be a sign of bimodality in the four older age bins, although low number statistics does not allow us to reach a significant conclusion. However, regardless of the binning, the oldest bin has a significantly higher maximum value (around ~ 14 km s^{-1}) than the two youngest ones. As age increases, the distribution becomes more and more asymmetric, with a large tail towards negative V_r, and a smaller tail on the positive side becoming more populated at older ages. This effect is also independent of the histogram binning and age limits. This implies that clusters moving inwards do it with a wider range of radial motions than those moving outwards. This trend is noticeable in the whole range of ages represented in our OC population, which is representative of the youngest pop-

ulation in the Galactic disc. The majority of old clusters of our sample have a positive V_r, highlighting that OCs that move outwards have a higher probability of survival than clusters moving inwards. For the azimuthal component, the histograms in bins of ages clearly show the asymmetric drift: in the two youngest bins the distribution is centred nearly on 240 $\mathrm{km\,s^{-1}}$ and the mode of the histogram shifts towards lower values in the older bins. In the three oldest bin, the most probable value is around 230 $\mathrm{km\,s^{-1}}$. In the third and fourth age bins, we can also note a bimodality of the distribution that is likewise seen for the radial component of the velocity, the bimodality being independent of the binning here, however. As age increases, the growing spread of the vertical velocity is clearly visible in the histograms, particularly in the two last age bins, where the distribution becomes asymmetric but remains centred near zero.

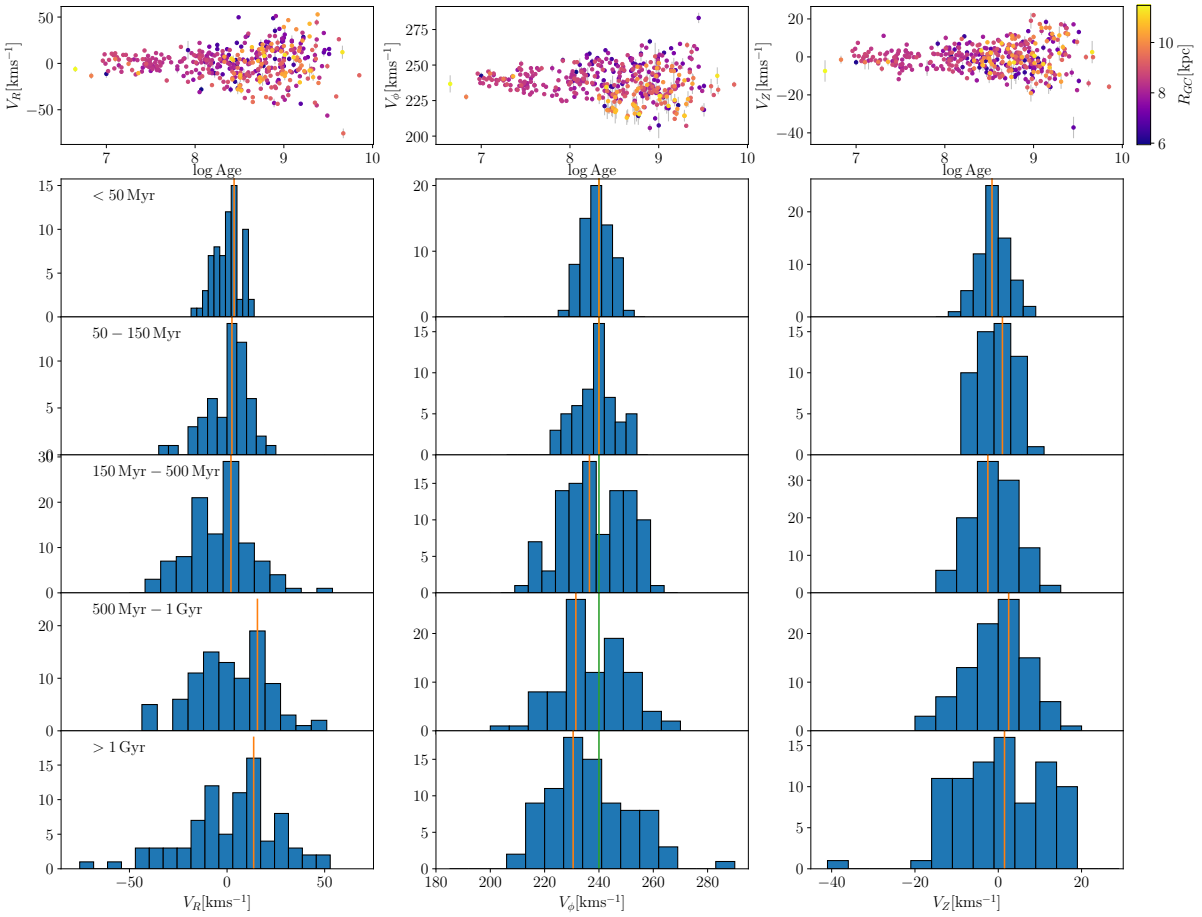

Figure 4.12: Top: Radial, azimuthal, and vertical velocities $V_r, V_\phi,$ and V_Z as a function of age for the HQS with the Galactocentric radius shown with colours as indicated in the colour bar. Bottom: Distribution of V_r, V_ϕ and V_Z of the OCs in different age bins. In each panel, the mode of the distribution is marked with a vertical orange line, and the azimuthal velocity of the LSR is marked with a vertical green line in the middle column.

Figure 4.13 shows the distribution of V_ϕ of the HQS sample in the X-Y plane in four different age bins. The spiral arms modelled by Reid et al. (2014) are represented by the shaded structures, and the updated Cygnus arm from Reid et al. (2019) is represented

by the dashed shaded structure. We note that in the two youngest bins, OCs are mainly located close to the Sun near the local arm, and this spatial distribution is very different from that of the original sample presented in CG+20 (their Fig. 8). This is a consequence of the bias of our sample stated above: we discarded most of the young distant clusters from the HQS based on the quality of their radial and Galactic velocities. The remaining nearby young clusters show a remarkable homogeneity in azimuthal velocities, with a typical velocity dispersion of 5.5 $\mathrm{km\,s^{-1}}$ around solar velocity. In the two oldest age bins, OCs are much more spatially and kinematically dispersed and no longer follow the spiral arms.

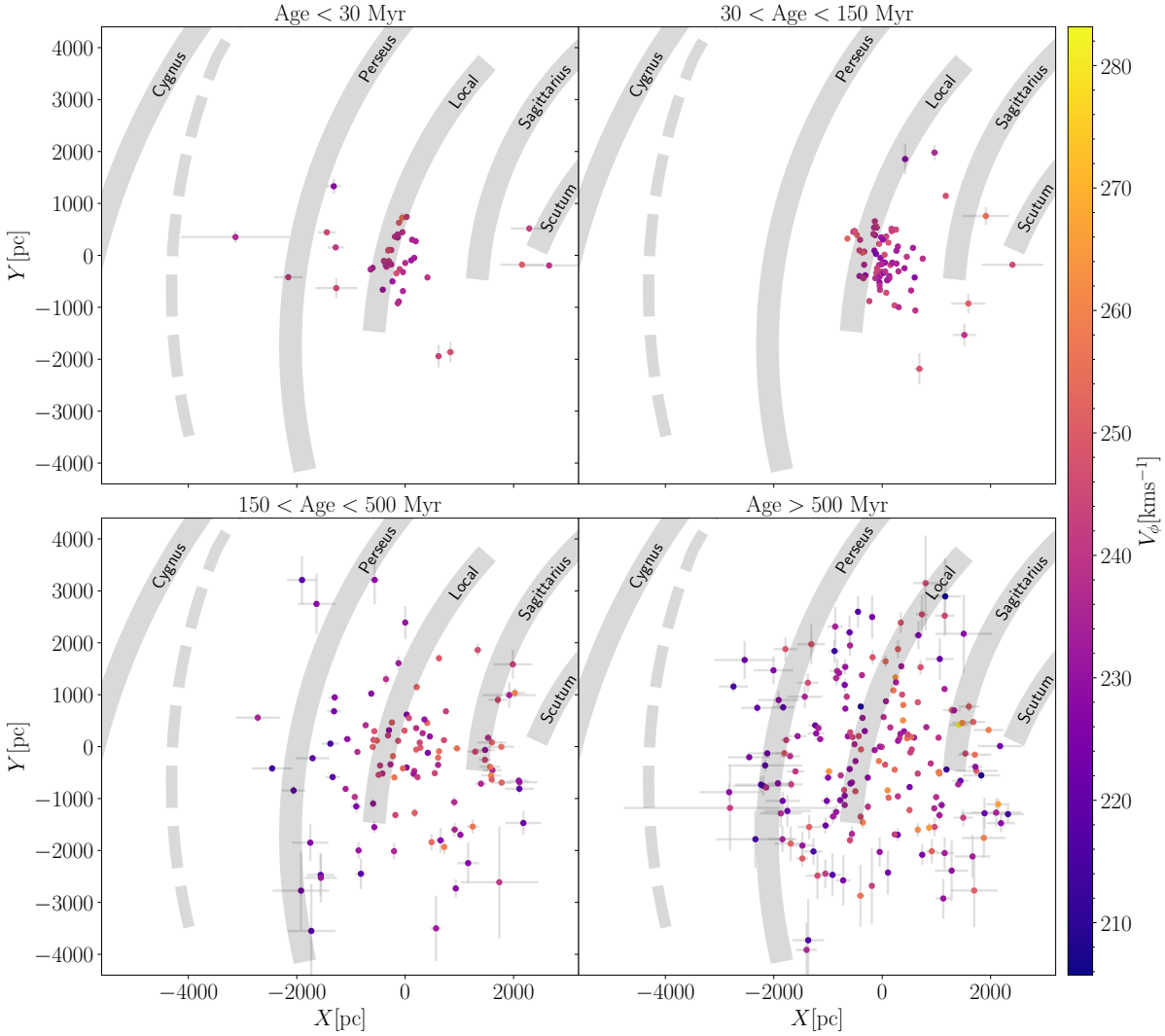

Figure 4.13: Distribution of the HQS OCs in heliocentric Cartesian coordinates. Colours stand for the azimuthal velocity V_ϕ. The shaded structures show the spiral arm modelled by Reid et al. (2014), and the dashed shaded structure represents the updated Cygnus arm from Reid et al. (2019).

4.3.4 Age-velocity relation for open clusters

The increasing trend of the velocity dispersions for field stars with age is known as the AVR and has been discussed for decades (e.g., Wielen, 1977; Freeman, 1987; Nordström et al., 2004). The AVR is reasonably described as a power law with some saturation at the oldest ages (Binney and Tremaine, 2008; Aumer and Binney, 2009; Soubiran et al., 2008), altough the shape of the relation is debated (Martig et al., 2014). The AVR involves different mechanisms, such as the radial migration and heating induced by giant molecular clouds, by the spiral arms (Mackereth et al., 2019), or by the bar (Grand et al., 2016), which affect the velocity components differently. The AVR study is complicated by the mixture of different stellar populations present in the solar neighbourhood, such as the thin and thick discs, which have different scale heights and different star formation and heating histories (Yu and Liu, 2018; Mackereth et al., 2019). Similarly, inner and outer populations are differently affected by the dynamical effect of the bar, the spiral arms, and accretion events. In addition, uncertainties in the determination of stellar ages strongly affect the AVR (Aumer et al., 2016). Ages of OCs are more reliable than ages of individual stars and have been determined homogeneously in CG+20. This allowed us to investigate the AVR of OCs in a wide age range up to ~ 2.5 Gyr (too few clusters are older than this limit) and with a significant number of objects below 1 Gyr. The spatial extension is mostly confined within 200 pc from the Galactic plane, so that the OC population is representative of the thin disc. It is thus interesting to compare the AVR of OCs and field stars in the age range in which they overlap.

Figure 4.14 shows the velocity dispersions measured in seven age bins for the 418 OCs from the HQS with an age determination. In each age bin, we fitted a Gaussian to each velocity component. Each point represented in Fig. 4.14 stands for these Gaussian standard deviations and their uncertainties. As expected, the dispersion of each component increases with age. The values and ratios are given in Table 4.3. The OC kinematics is characterised by a clear anisotropy in the three components, at all ages. The radial dispersion σ_R is always significantly larger than σ_ϕ, which is significantly larger than σ_z, as also seen in local field stars (e.g. Binney and Tremaine, 2008; Kuijken and Tremaine, 1994; Anguiano et al., 2018). The dispersion ratios remain globally stable in the different age bins, with some fluctuations within the error bars. This implies that the velocity ellipsoid maintains the same shape at all ages. Following previous studies of field stars, we fitted the increase in velocity dispersion with age (τ) in the form of a power law $\sigma_V \propto \tau^\beta$, as suggested in Sect. 8.4 of Binney and Tremaine (2008) and by Jenkins (1992). We used a maximum likelihood (ML) estimator on the individual clusters of the HQS (ages younger than 2.5 Gyr), assuming Gaussian velocity errors. The results of the fits are shown in Fig 4.14. The grey lines represent the uncertainties on the fits: we show 100 fits taken from the final sample of the ML. The solid black line shows the best-fit power law obtained with a ML. This best fit is defined as the median value of the 16000 fits. We found $\beta_R = 0.25^{+0.05}_{-0.03}$, $\beta_\phi = 0.23^{+0.03}_{-0.03}$, and $\beta_z = 0.19^{+0.03}_{-0.03}$. These values show that the heating rate of OCs is about the same in all directions, which agrees with the observation that the velocity ellipsoid maintains the same shape at all ages, as pointed out earlier. In the following, we compare this kinematical behaviour with that of field stars.

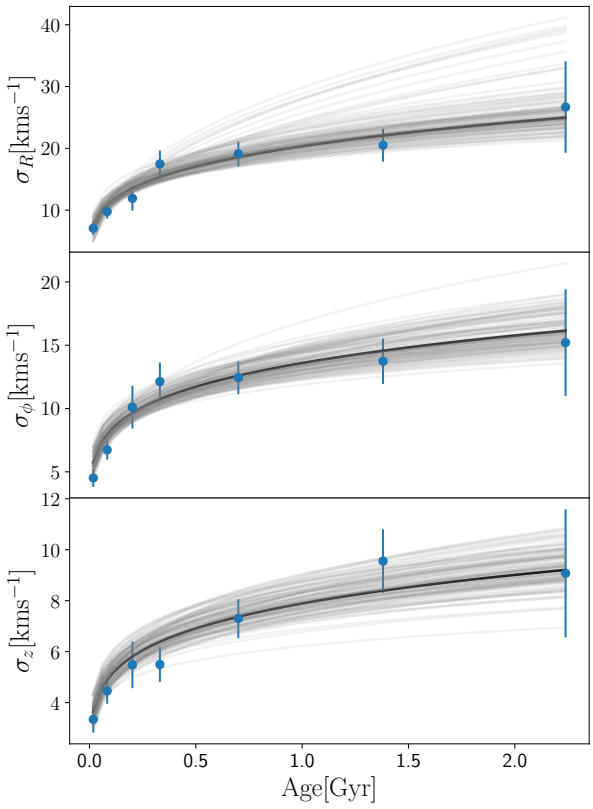

Figure 4.14: Velocity dispersion of the cylindrical Galactocentric components V_r, V_ϕ, and V_z for the HQS. The blue dots and error bars stand for the standard deviations of the velocities and their uncertainties obtained by fitting a Gaussian on each component of the velocities in each of the seven age bins of Table 4.3, the oldest age bin having an upper limit of 2.5 Gyr. The x value of the blue dots is the median value of the corresponding age bin. The black line shows the best-fit power law obtained using a ML on each cluster of the HQS younger than 2.5 Gyr, and the grey lines represent the uncertainties of the fits.

Yu and Liu (2018) measured the AVR of $\sim$ 3 500 local stars, splitting them into different subsamples depending on their Z position and metallicity. Their metal-rich low-Z sample is comparable to our OC sample, which is representative of the local thin disc. For their two youngest age bins, corresponding to 1.4 ± 0.4 and 1.9 ± 0.3 Gyr, respectively, they found dispersions in the different components and dispersion ratios that agree well with what we found for OCs of similar age (see their Table 1). They fitted the heating parameters in the age range 1 to 8 Gyr and determined $\beta_R = 0.28\pm0.08$, $\beta_\phi = 0.30\pm0.09$, and $\beta_z = 0.54 \pm 0.13$, in agreement with our values in R and ϕ, but not in Z.

Mackereth et al. (2019) measured the heating parameters in R and Z with field stars with ages between 1 and 9 Gyr. They found $\beta_z = 0.50$ for stars that were comparable to OCs in metallicity, which is higher than our value. They found that β_R varies from 0.15 to 0.4 depending on the mean orbital radii. Recently, Sharma et al. (2021) assembled several large stellar samples that traced different populations from complementary large surveys and using different age estimators. They found consistent relations for all with $\beta_R = 0.251 \pm 0.006$ and $\beta_z = 0.441 \pm 0.007$, which again agrees with our value for R, but not Z. Mackereth et al. (2019) and Sharma et al. (2021) highlighted a complex relation

Table 4.3: Dispersions of velocity components and ratios in the same bins as in Fig. 4.14

age interval (Gyr)	N	σ_R	σ_ϕ	σ_z	σ_R/σ_ϕ	σ_R/σ_z	σ_ϕ/σ_z
age< 0.03	43	7.05 ± 1.10	4.52 ± 0.71	3.34 ± 0.52	1.56 ± 0.49	2.11 ± 0.66	1.35 ± 0.42
0.03 – 0.15	79	9.75 ± 1.11	6.73 ± 0.77	4.46 ± 0.51	1.45 ± 0.33	2.19 ± 0.50	1.51 ± 0.34
0.15 – 0.25	38	11.90 ± 1.98	10.11 ± 1.68	5.48 ± 0.91	1.18 ± 0.39	2.17 ± 0.72	1.84 ± 0.61
0.25 – 0.50	67	17.48 ± 2.17	12.13 ± 1.50	5.50 ± 0.68	1.44 ± 0.36	3.18 ± 0.79	2.21 ± 0.55
0.50 – 1.00	94	19.11 ± 1.99	12.44 ± 1.30	7.30 ± 0.76	1.54 ± 0.32	2.62 ± 0.55	1.70 ± 0.36
1.00 – 2.00	61	20.51 ± 2.67	13.73 ± 1.79	9.55 ± 1.24	1.49 ± 0.39	2.15 ± 0.56	1.44 ± 0.37
2.00 – 2.50	15	26.68 ± 7.40	15.21 ± 4.22	9.07 ± 2.52	1.75 ± 0.97	2.94 ± 1.63	1.68 ± 0.93

Table 4.4: Comparison of the values of β found here and in previous studies

	β_R	β_ϕ	β_z
This study	$0.25^{+0.05}_{-0.03}$	$0.23^{+0.03}_{-0.03}$	$0.19^{+0.03}_{-0.03}$
Yu and Liu (2018)	0.28 ± 0.08	0.30 ± 0.09	0.54 ± 0.13
Mackereth et al. (2019)	-	-	0.50
Sharma et al. (2021)	0.251 ± 0.006	-	0.441 ± 0.007

between σ_R/σ_z with age, depending on metallicity, angular momentum, and height above the plane.

These comparisons of the heating parameter β (see Table 4.4 for more clarity) tend to show that the OC population has a dynamical evolution similar to the field stars in the radial and azimuthal directions, but not in the vertical direction. OCs seem to have a lower heating rate in Z than field stars. The main difference between our determination of β and that of Yu and Liu (2018); Mackereth et al. (2019) and Sharma et al. (2021) is the age range which extends to young ages for OCs (half of the OCs are younger than 360 million years) with few objects older than 3 Gyr, while the stellar samples typically range between 1 - 10 Gyr. Although their age distributions overlap around 1 - 2 Gyr, the fit of the β parameter is not performed in the same age range for OCs and field stars. It is therefore interesting that the heating rate is found to be similar in the Galactic plane but not perpendicular to the plane. This could mean that clusters do not reach high altitude and older ages because they are disrupted before, which therefore introduces a bias in our sample. The small β_z could also reflect that giant molecular clouds, which are the main cause of the vertical scattering of field stars (Lacey, 1984; Jenkins and Binney, 1990), are not as efficient in scattering OCs, or that the effect of the giant molecular clouds is to disrupt the OCs.

The heating of the Galactic disc and the destruction of clusters have been simulated by Gustafsson et al. (2016), among others. They found that the fraction of massive old OCs, scattered into orbits with $|Z| > 400$ pc, is typically 0.5%. Of the full initial sample of 2017 OCs from CG+20, 4% have $|Z| > 400$ pc. They are mainly old, with a median age of ~ 2 Gyr, and at Galactocentric distances ranging from 7.9 to 20 kpc, as previously reported by CG+20. However, after our quality cuts, only six such clusters remain in the HQS, making the statistics too poor to reach a conclusion on their origin.

4.4 Actions and orbital parameters

In this section we use the full 6D coordinates of the samples of clusters to compute orbits and action variables, and analyze them as a function of the age. We used the python package `galpy` (Bovy, 2015) for Galactic dynamics to integrate the orbits and compute the action-angle variables.

We used the axisymmetric potential MWPotential2014 implemented in `galpy`, which was derived by Bovy (2015) by fitting a simple model to existing dynamical data of the Milky Way. It is composed of a bulge, a Miyamoto-Nagai disc (Miyamoto and Nagai, 1975), and a dark matter halo modelled with a Navarro-Frenk-White (NFW) potential (Navarro et al., 1997). For the sake of comparison, we also used the axisymmetric potential model from McMillan (2017) to confirm our results. This model was also fitted to the mass distribution of the Milky Way and consists of several components representing the cold gas disc close to the Galactic disc, both the thin and thick discs, a bulge, and a dark matter halo. The detailed parameters of each component of these potentials are listed in Bovy (2015) and McMillan (2017).

4.4.1 Orbits

We integrated each cluster orbit with an integration step of 0.01 Myr up to 500 Myr. We did not attempt to integrate more time (up to each cluster age) because the reliability of the results decreases significantly with time due to inaccuracies in the time dependence of the potential, amplification of the uncertainties in distance, and motions, among other effects (Gaia Collaboration et al., 2018c). We computed uncertainties by integrating each orbit 1000 times with a Monte Carlo sampling in the same way as described in Sect. 4.3. Figure 4.15 shows the orbits of two OCs in the (X, Y) plane and the (Z, R_{GC}) plane, as an example of the results we can obtain using `galpy`.

We extracted the orbital parameters of the 1315 OCs for which we have an age estimate to investigate their relation with age. We represent the evolution of the maximum altitude above the Galactic plane (Z_{max}) as a function of the age of the full sample of clusters and the HQS in the left panel of Fig. 4.16. In all the panels, the maximum height of clusters younger than 300 Myr remains constrained close to the Galactic plane. This is shown more clearly in the bottom panel, where the running median of both samples increases only for an age higher than 1 Gyr. For both samples of OCs, the median age of the subsample of clusters that reach an altitude higher than 400 pc is greater than 1.5 Gyr. The left panels of Fig. 4.16 also show the increase in dispersion of the maximum height of the OCs above the plane for ages older than 1 Gyr. This is usually attributed to the vertical heating of the disc: clusters are preferentially formed in the thin disc, and then giant molecular clouds and spiral arms tend to scatter them away from the midplane (Spitzer and Schwarzschild, 1951; Jenkins and Binney, 1990). This effect is consistent with what is shown in more detail in the right panels of Fig. 4.12, as we commented on in Sect. 4.3.

The eccentricity as a function of age in the right panel of Fig. 4.16 shows several things. First, clusters younger than 30 Myr ($\log(\text{Age}) \sim 7.5$) show very low eccentricity.

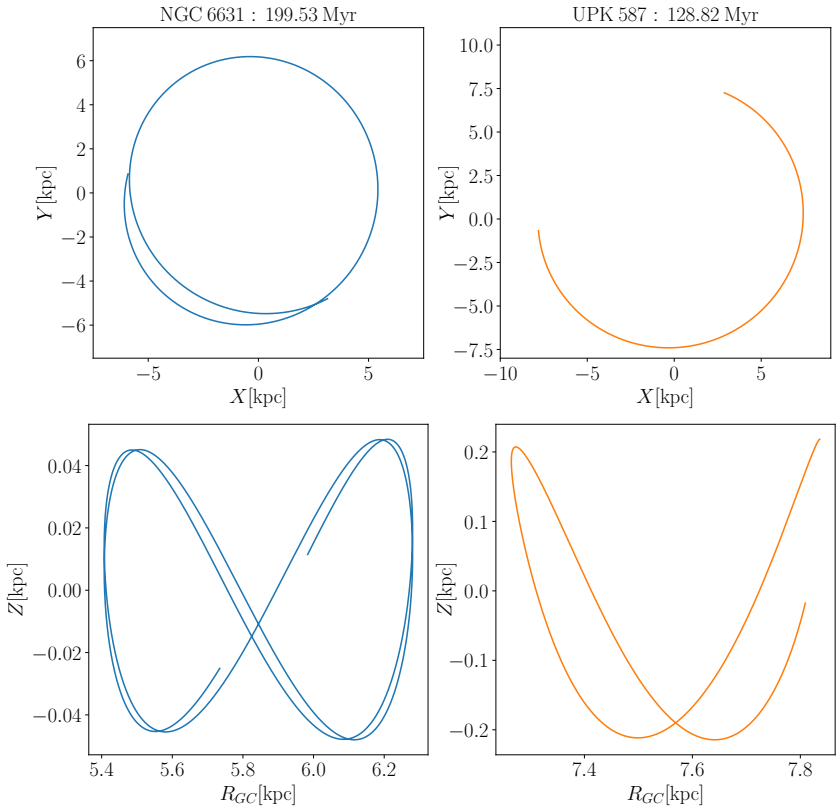

Figure 4.15: Example of an orbit for two OCs NGC 6631 and UPK 587 integrated with `galpy`.

This is best seen in the HQS sample where the maximum value in this range is 0.018, while the full sample contains some outliers with values up to 0.15. Very young clusters therefore have nearly circular orbits. Then, for clusters older than 30 Myr the dispersion of the eccentricities at a given age is large for both samples. The running median of the HQS shows an increase that is quite smooth at least up to 300 Myr ($\log(\text{Age}) \sim 8.5$). The full sample exhibits a similar behaviour as that of the HQS, but with slightly higher mean eccentricity values. And finally, for ages older than 1 Gyr the dispersion in eccentricities starts to be very large for the full sample. For the HQS there seems to be a stabilisation of the eccentricities around a mean value of 0.08.

This shows that OCs are born on nearly circular orbits, and as their age increases, they are more likely to be gravitationally perturbed from non-axisymmetric components. We did not find very young clusters with high eccentricities, but we found old clusters with both high and low eccentricities. We did not see a preferential location in the Galaxy, or differential characteristics between high- and low-eccentricity old clusters.

4.4.2 Action-angle variables

The action-angle variables are a set of canonical coordinates that have been proved to be useful to study the substructure of stars in the 6D phase space. As extensively discussed by several authors (e.g. McMillan and Binney, 2008), orbital actions (J_R, J_ϕ, J_z) are integrals of motion in an axisymmetric potential, but they also provide information of non-

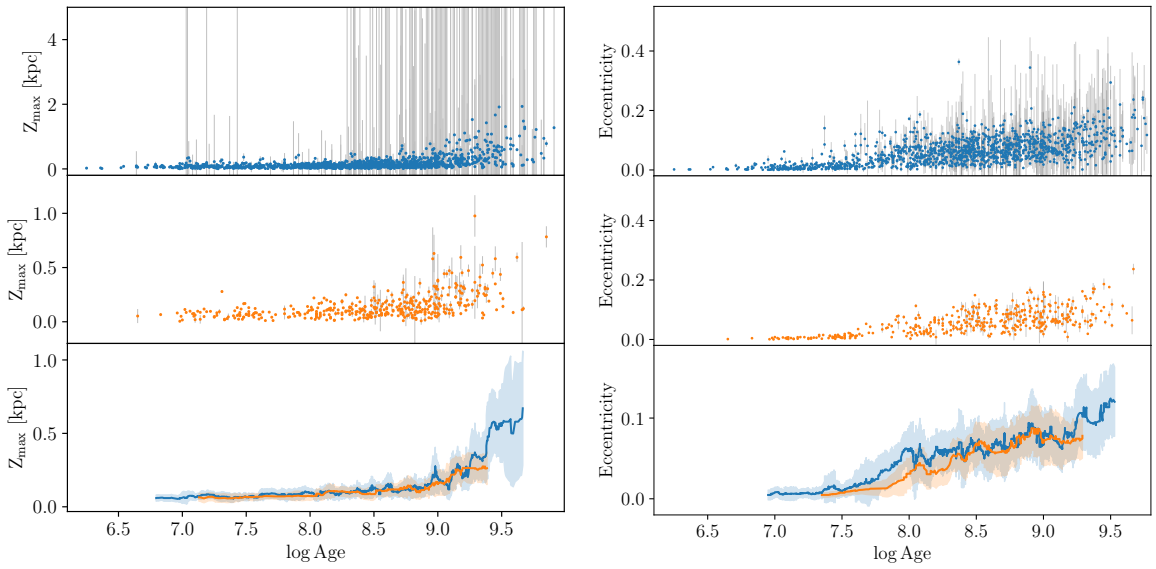

Figure 4.16: Maximum altitude above the Galactic plane (left) and eccentricity (right) of the clusters for which we could integrate their orbits as a function of age. In both panels, we represent in blue (in the top panels) the full sample of OCs and in orange (in the middle panels) the HQS. The bottom panels shows the running median of the two samples, calculated with a window of 30 points. The shaded area corresponds to 1 MAD from the median (line). Note the change in scale between the top left panel and the two bottom left panels and between the two top right panels and the bottom right panel.

axisymmetric perturbations. In addition, they are independent of time and are therefore more reliable to describe the orbital parameters by removing their time dependence. They have been used to describe stellar components in our Galaxy, in particular, the moving groups seen in the solar neighbourhood in the (U, V) plane (Trick et al., 2019).

In an axisymmetric potential, action variables can easily be interpreted as physical quantities. The radial action J_R can be used as a proxy for the orbit eccentricity or as a measure of the oscillations around the guiding-center radius (i.e. the mean orbital radius) of the object. The azimuthal action J_ϕ is equal to the angular momentum in the vertical direction L_Z, which indicates the quantity of rotation of the object around the centre of the Galaxy. Similarly, the last coordinate, the vertical action J_z, can be used as a measure of the oscillations of the object around the plane of the Galaxy and therefore is a proxy for the maximum height of the object along its orbit. These actions are defined as:

$$\begin{cases} J_R = \dfrac{1}{\pi}\displaystyle\int_{R_{min}}^{R_{max}} v_R dR, \\ J_\phi = L_Z = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} v_T \cdot R d\phi = R_g \cdot v_{circ}(R_g), \\ J_z = \dfrac{2}{\pi}\displaystyle\int_0^{z_{max}} v_z dz, \end{cases} \tag{4.1}$$

where R_{min} and R_{max} are the pericenter and apocenter of the OCs orbit, R and v_T are

the OCs radial position and tangential velocity, R_g is its guiding-center radius, $v_{circ}(R_g)$ is the azimuthal velocity at the guiding-center radius and z_{max} is the orbit's maximum height.

All of these quantities are conserved in an axisymmetric potential but are affected by non-axisymmetric structures such as a bar or spiral arms, or by a merger. We refer to Sect. 3 of Binney and Tremaine (2008) for a mathematical and comprehensive description of these variables and their meanings.

We made the computation for the 411 clusters from the HQS with known age. For comparison purposes, we made the same computation with the sample of stars within $d < 200$ pc in *Gaia* DR2 RVS, that is, the same selection as Trick et al. (2019), which counts $\sim$ 350,000 stars.

In Fig. 4.17 we show the distribution of radial action J_R with respect to the vertical component of the angular momentum L_Z of field stars compared with the sample of HQS clusters. In all panels, the sample of field stars is the same because we do not know their ages. We indicate the approximate location of the known moving groups present in the solar neighbourhood (Antoja et al., 2008; Gaia Collaboration et al., 2018d), as analysed in the action space by Trick et al. (2019). In this space, the location of a star is interpreted in relation to its orbital characteristics as follows: (i) the V shape is due to the cut in distance (200 pc) made for the sample selection, (ii) stars with circular orbits are placed at $(L_Z, \sqrt{J_R}) \sim (1, 0)$, while more eccentric orbits appear at larger J_R, and (iii) stars close to their apocenter (pericenter) are placed in the left (right) edge of the V.

We restricted the volume of analysed clusters using their Galactocentric radius instead of their heliocentric distances because of the reduced amount of clusters at $d < 200$ pc. Even though this made the comparison between cluster and field more difficult, the precision in the distances and velocities of the clusters is much better than for individual field stars, so that we expect that the kinematic information is not blurred by this, as usually occurs for individual stars. We made the cuts in $R_{GC} = 8.3 \pm [0.2, 0.3, 0.5]$ pc, shown in each row. Using the information of cluster ages, we were able to add an additional dimension to the figure, therefore we divided the sample into four age bins $[< 30, 30 - 150, 150 - 500, > 500]$ Myr, shown as columns. The differences in the volume selection of the clusters and the field stars are visible in the top row, where some of the clusters stand out of the left and right edges, which would correspond to clusters towards the Galactic centre and anticentre.

Figure 4.17 shows that the bulk of clusters tends to be concentrated towards the position $(1, 0)$, which is expected because most OCs have cold kinematics and tend to have nearly circular orbits. All three distance limits (rows) show that there are few clusters at high eccentricities, but they increase in number at older ages. As a consequence, the density peak moves upwards towards older ages. This is particularly clear in the top row because of the large number of points.

Regarding the relation of the clusters and the moving groups, no clusters populate the two Hercules substreams, which are stars with too eccentric orbits and low angular momentum. Conversely, the location of the Hyades, Pleiades, Coma Berenices, and Sirius moving groups more closely resembles the distribution of the clusters, even at

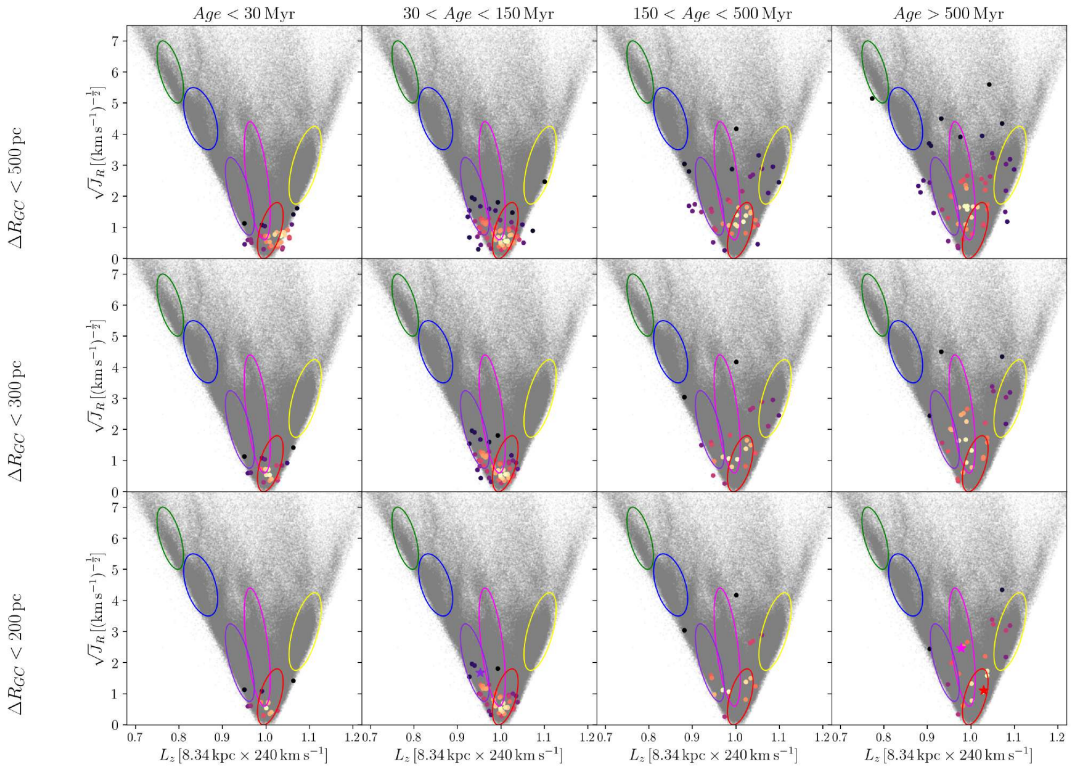

Figure 4.17: Radial action vs. angular momentum (J_R, L_z) distribution of field stars closer than 200 pc in *Gaia*-DR2-RVS (grey), and clusters (coloured dots by the density of points). Each column shows clusters in different age bins, and each row includes clusters selected inside different ranges of Galactocentric radius, instead of heliocentric distances. We indicate the approximate location of the known moving groups analysed by Trick et al. (2019): Hercules (green and blue), Pleiades (dark violet), Hyades (magenta), Coma Berenices (red), and Sirius (yellow). We highlight the values of the three clusters Pleiades, Hyades, and Coma Berenices with a violet (second column in the bottom row), pink, and red star (both in the last column of the bottom row), respectively, in their age-$R_{\rm GC}$ panels.

larger distances (top row). We highlight that the age bin $30 - 150$ Myr (for all distance cuts) shows a significant clump of clusters towards the middle left side (smaller L_z), populating the region where the Pleiades moving group is found. On the other hand, the Sirius moving group is populated mainly by clusters in the two older age bins. These dependences in the population in each moving group depending on age point in the same direction as the findings by Antoja et al. (2008): in their Fig. 14, the Hyades, Pleiades, and Coma gather most of their young stars ($\lesssim 300$ Myr), and at $\gtrsim 300$ Myr, there is a bump of stars related to Sirius. Hercules is associated with stars of ~ 2 Gyr, and the distributions of all moving groups also exhibit a bump at this age. However, we find almost no clusters at this old age so there is no indication of this in our sample.

4.5 Conclusion

By combining *Gaia* ground-based surveys and catalogues, and with new memberships and age determination from CG+20, we assembled the largest catalogue of RV for OCs in order to study their kinematics. As a by-product of our study, which includes the *Gaia* DR2 RVS, *Gaia*-ESO survey, APOGEE, RAVE, GALAH, and smaller catalogues, we compared the RVs from the different sources to each other to assess their typical precision and zero-point. We found RV zero-points to be consistent at a level better than 1 km s^{-1}. The scatter of the comparisons indicates that the real precision of each catalogue is compatible with the individual uncertainties listed in it. All non-*Gaia* RV measurements were corrected to align them on the *Gaia* RVS zero-point. The weighted mean RV of each star and each cluster resulted in 1 382 OCs with an RV, 38% with a highly reliable RV based on more than three stars and with an uncertainty lower than 3 km s^{-1}.

We computed both heliocentric and Galactocentric Cartesian and cylindrical velocities for this sample of OCs and defined a high-quality sample composed of 418 OCs with the most reliable velocities, 411 OCs of which have an age determination. We found that most OCs fall in a band in between the two main arches drawn by field star in the $V_r - V_\phi$ plane, while they seem to follow the overdensities described by the diagonal ridges in the $R_{GC} - V_\phi$ plane. The rotation curve drawn by our OCs shows two significant dips: at $R_{GC} \sim 7$ kpc, and a more prominent one around $R_{GC} \sim 9.7$ kpc. The locations and depths of these dips agree with the perturbations we would expect from the non-axisymmetric components of the disc, which also draw the ridges observed in the $R_{GC} - V_\phi$ plane.

With the ages of almost all the clusters from our sample, we investigated the age velocity relation for OCs in detail. It shows a clear anisotropy of the three velocity components. Compared with field stars studies, the heating parameter β of OCs was found to be similar in the radial and azimuthal directions, but significantly lower in the vertical direction. This low heating rate in the Z coordinate can be due to the disruption of old clusters, which are the most likely to reach high altitudes above the disc, or to a less efficient heating of OCs by giant molecular clouds. We are aware, however, that the quality cuts we applied discarded distant clusters, resulting in a bias in our sample.

We used the 6D + age information of our sample of OCs to compute and investigate orbits and action variables. We analysed the dependences of the recovered orbital parameters as a function of age. Most of the clusters reach a maximum altitude above the plane during their orbits that is lower than 400 pc, and only those older than 1 Gyr are able to move further away from the midplane, but typically less than 1 kpc. Clusters younger than 30 Myr show a very low eccentricity (~ 0.018), and for clusters older than this, especially those older than 100 Myr, the eccentricity shows an increasing relation with age. These results show that OCs are born in circular orbits, and as age increases, they are more prone to suffer perturbations of their orbits. This is also seen after the computation of action variables, where as age increases, the distribution in the $(L_Z, \sqrt{J_R})$ plane tends to spread beyond $\sim$ (1,0). We related our cluster distribution in this action space with the location of the known moving groups as a function of age. We conclude

that the Pleiades, Hyades, and Coma moving groups appear to be more populated by young clusters, while the Sirius region appears to have a clump of clusters of age $\gtrsim 300$ Myr. No clusters populate the two Hercules streams.

Chapter 5

Conclusion, limits and perspectives

5.1 Summary and conclusions

This thesis focuses in the study of the kinematics and morphology of the population of known OCs of the Galaxy. With this purpose we developed a methodology able to detect extended structures around several hundreds of them. We revisited their structural parameters in the light of these new lists of members. We also combined ground based-surveys with *Gaia* and new memberships to study in details the kinematical properties of OCs.

5.1.1 Context

OCs are ideal probes of the Galactic disc history and characteristics. The determination of their physical parameters such as their ages, distances, chemical composition, and kinematics is more precise than for field stars, making them key objects in Galactic archeology studies. Since the publication the DR2 of the *Gaia* mission, the OC census has changed drastically. The combination of the massive amount of data, all sky-coverage and exquisite accuracy allowed to revisit the membership lists of the known clusters from the Milky Way.

In particular CG+18 provided new lists of members for more than 1100 clusters, discarded hundreds of asterisms previously considered as OCs and discovered 60 new clusters, even at low distances while the OC census was thought to be complete up to 1.8 kpc before *Gaia* (Kharchenko et al., 2013). The amount of *Gaia* data to analyse was so huge that discoveries of OCs have followed one another. In addition, while the tidal structure of some GC has been identified before (like for Palomar 5: Odenkirchen et al., 2001), no such structure was known for OCs, even for the closest ones. With *Gaia*-DR2 data, some studies dedicated to small OC samples have revealed structures in the outskirts of very few clusters. The two main ones are the ones by Pang et al. (2021) and Meingast et al. (2021) who revealed such structures for several nearby clusters.

However, at the exception of Castro-Ginard et al. (2020) who identified the tail of UBC 274/LP 5, a cluster located at 1.8 kpc from the Sun, all the studies dedicated to the detection of extended structures around OCs were focused on small samples of clusters. The outskirts of the known OC population studied in an homogeneous and comprehensive way is still lacking while the wealth, accuracy, precision and homogeneity of the data from *Gaia* allows this type of study.

The kinematical study of OCs was also greatly improved by the publication of *Gaia*-DR2. The largest catalogues of RV for OCs increased in sizes twice the year following

the publication of the DR2 with the studies from Soubiran et al. (2018a) and Carrera et al. (2019a) who fully exploited latest membership determinations. The combination of *Gaia*-RVS data with ground based surveys also allows to check for consistency and offsets between the different surveys and to compare the kinematics of OCs with field stars.

5.1.2 Memberships and structure of Open Clusters

We developed a methodology able to recover clusters' members up to 50 pc from the centers of clusters and we provide to the community these lists of members for several hundreds of clusters. To do so, we took advantage of the recently developed machine learning algorithm HDBSCAN. We provide new lists of members for 389 clusters up to 1.5 kpc away from the Sun, representing the largest homogeneous study of the OCs outskirts. These lists of members allow us to identify 71 clusters with a tidal tail. Among these 71 clusters, more than 60 are identified for the first time. In addition, while only a handful of clusters (which had been the subject of specific studies) were known to have extended halos, our methodology highlights that a majority of the OCs in our sample are surrounded by prominent halos which can contain up to several hundreds of stars.

Our new compilation of members allows us to revisit the two dimensional morphology of OCs. With this two dimensional study, we exploit the best measurements of *Gaia*. We find the clusters to be much larger than previously thought, all extending up to our search cone (50 pc). We also find what seems to be a negative correlation between the sizes of the cores of OCs and their age and a positive correlation of their overall sizes with their age. We also investigate the proportion of stars populating the cores and the halos of the clusters and show that the old clusters tend to be in proportion less populated than young ones. If we do not find evidence that the mass segration increases with age, we still find that the clusters showing significant signs of mass segregation are on average older than clusters which do not show these signs.

The increase of the overall sizes of clusters with age as well as the decrease of the proportion of stars populating their halos reflects the two main dynamical processes at play during the lifetimes of clusters: mass segregation and evaporation. We observe at the same time that old clusters halos seem to be sparsely populated and that old clusters seem to be more mass segregated than young ones. This could highlight that the disruption process triggered by encounters with GMCs, Galactic harassment and spiral arms shocking are more efficient to strip the clusters from their outer stars than mass segregation to populate the peripheral areas of the clusters.

5.1.3 Kinematical properties of Open Clusters

Using the available membership lists for the overall OC population, we aimed at building the largest RV catalogue of OCs. By cross-matching these lists of members with the *Gaia*-RVS as well with several ground based surveys, we provide to the community the largest catalogue of RV for OCs. Out of the 2 017 clusters in the original list, 1 382 now have a determination of their RV, among which half are based on more than three stars and 38% also have an uncertainty lower than 3 $\mathrm{km\,s^{-1}}$.

By computing Galactocentric cylindrical velocities, we noticed significant differences between the kinematics of OCs and field stars. First, the OCs are not found to be aligned with the structures of the phase space recently discovered by Gaia Collaboration et al. (2018d). Secondly, the analysis of the AVR of OCs shows that OCs suffer from the same amount of kinematical heating as field stars in the radial and azimuthal direction but significantly less in the vertical direction. This can be due to a bias in the age distribution of the clusters as most clusters are disrupted after a hew hundreds of Myr. However, encounters with GMCs which are the dominant vertical heating process might also be more efficient to disrupt the clusters than to kinematically heat them.

From the computation of orbits, we obtain a clear dependence of the maximum height and eccentricity with age. Finally, the orbital characteristics of the sample of clusters as shown by the action variables, follow the distribution of field stars.

5.2 Limits

The study of OCs suffer from multiple biases among which some are inherent to these particular objects. The main important one is caused by the disruption of most of the clusters. Massive OCs, or those located on the outer parts of the Galaxy are more likely to survive than clusters composed of only a few tens of stars. This results in a bias of the age distribution of OCs with very few clusters older than 1 Gyr. This bias can then affect our results when looking at the dependance of the clusters' parameters with age.

The determination of the parameters of young clusters raises some issues as well. Very young OCs (a few Myr old) are still surrounded by the gas and dust which constitute the GMC progenitor of clusters (Lada and Lada, 2003) and they are still located in the spiral arms close to the Galactic plane. The observation of the members of such clusters therefore suffers from heavy extinction, especially in the optical wavelengths in which *Gaia* is observing. As clusters usually form at the intersection of the filamentary structures of GMCs (Lada, 2010), young clusters usually also have close neighbours. In Sect. 2.3, we chose to query the *Gaia* archive with cone search of 50 pc radius around each cluster in order to detect members of its halo. This choice had one inherent drawback as our queries contained in some cases several clusters with very similar parameters that our clustering algorithm was unable to separate. We had to discard clusters younger than 50 Myr for this reason in order to provide reliable membership lists.

The determination of the RV of young stars is also challenging as young stars can have high temperatures or rotate rapidly, making their RV less reliable. This is the reason why most of the young and distant clusters have been discarded from our HQS in Sect. 4.3.

Another limit of our clustering methodology is a direct consequence of the large relative parallax error highlighted in Sect. 3.2. As we directly use the parallax in our clustering procedure, the identification of tidal tails aligned (or slightly tilted) with the direction of the line of sight like the ones of Ruprecht 147 and NGC 2632/Praesepe is difficult. The tails of Ruprecht 147 have been discovered by Yeh et al. (2019) with a method similar to the one presented in Sect. 2.2 and the tails of Praesepe were identified by Röser and Schilbach (2019) thanks to te convergent point method. None of these au-

thors directly used the parallax in their procedure like we did. Their method is suitable for the anlaysis of such nearby clusters but it is difficult to adapt to a large sample of clusters further away than a few hundreds of pc. The fact that the parallax relative error increases rapidly with the distance made us decide to do the morphological analysis in two dimensions only. This decision allows the inclusion of clusters up to 1.5 kpc, while the limitation of the parallax errors would only allow a homogeneous study for 14 clusters inside the 250 pc radius from the sun.

5.3 Perspectives

The future DRs of *Gaia* will be more and more precise as they will be based on much more months of observation. With, more precise data, the study of the three dimensional structure of OCs will not necessarily be limited to the clusters closer than 250 pc. Besides, while the uncertainties on the astrometric parameters of faint stars is still high, an improved accuracy of the astrometric measurements will allow to go to fainter magnitudes and to consider stars fainter than 18 mag in clustering procedures like the one presented in Sect. 2.3. The study of dynamical process like mass segregation which depends on the census of low-mass objects will be improved by these additional informations. An improved determination of the astrometric parameters of OCs will also allow our clustering procedure to be efficient further than the 1.5 kpc limit we fixed.

The number of RV measurements is also about to increase dramatically with the release of the complete *Gaia*-DR3 (expected for the first semester of 2022) and the publication of RV information for ~ 33 million stars making it by far the largest RV catalogue. A number of ground-based surveys like WHT Enhanced Area Velocity Explorer (WEAVE) or 4-metre Multi-Object Spectroscopic Telescope (4MOST) will also dedicate part of their observing time to OCs and provide to the community RVs measurements. The study of the kinematics of OCs will take a leap forward with this amount of data.

Our clustering procedure described in Sect. 2.3.2 can also be exploited to evaluate the chemical homogeneity of OCs. In Casamiquela, Tarricq et al. (in preparation), we will redetermine the memberships of UBC 274, investigate the morphology and extent of its tidal tails and assess the chemical homogeneity of the cluster thanks to the spectroscopic observations carried out in January 2021 at the Magellan telescope (Las Campanas Observatory). Our membership catalogue will also be used for the OC case in the galactic archaeology survey of WEAVE. This catalogue is specially suited because the field of view of WEAVE is very large and the inter-fiber restrictions of the multi-object spectrograph has allocation limitations in crowded fields, making WEAVE unable to observe the central regions of clusters. In addition, WEAVE will be able to confirm the detection of tidal tails and stellar halos presented in this thesis using the wealth of RVs and chemical abundances. As more tidal tails will also be progressively discovered, a detailed three dimensional analysis of their members will be necessary to check if these tails are always oriented in the direction of the Galactic rotation or if this orientation depends on the age of the clusters.

As the clusters whose structure has been studied in this thesis are mainly old clusters,

a similar study dedicated to clusters younger than 50 million years would also allow to examine the differences between these two sample of clusters. Furthermore, our study has shown important differences between the kinematics of field stars and OCs that remain to be confirmed. These differences in vertical velocity dispersion raise questions about the real influence of GMCs on OCs: are GMCs more efficient to disrupt OCs than to vertically scatter them ?

Finnaly, we have made available for the community three catalogues: the largest catalogue of RV for OCs, extended membership lists at large radii for 389 clusters closer than 1.5 kpc and older than 50 Myr as well as a catalogue of the structural parameters of OCs based on the aforementionned membership analysis. These catalogues will be useful in the global effort to a better constrain of the theoretical models of cluster evolution and Galactic dynamics.

Bibliography

M. Agarwal, K. K. Rao, K. Vaidya, and S. Bhattacharya. ML-MOC: Machine Learning (kNN and GMM) based Membership determination for Open Clusters. MNRAS, 502 (2):2582–2599, 2021. (Cited on page 42.)

R. Ahumada, C. Allende Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, R. Arcodia, et al. The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. ApJS, 249(1):3, 2020. (Cited on page 85.)

B. S. Alessi, A. Moitinho, and W. S. Dias. Searching for unknown open clusters in the Tycho-2 catalog. A&A, 410:565–575, 2003. (Cited on pages 34 and 173.)

R. J. Allison, S. P. Goodwin, R. J. Parker, R. de Grijs, S. F. Portegies Zwart, and M. B. N. Kouwenhoven. Dynamical Mass Segregation on a Very Short Timescale. ApJ, 700(2): L99–L103, 2009a. (Cited on page 81.)

R. J. Allison, S. P. Goodwin, R. J. Parker, S. F. Portegies Zwart, R. de Grijs, and M. B. N. Kouwenhoven. Using the minimum spanning tree to trace mass segregation. MNRAS, 395(3):1449–1454, 2009b. (Cited on pages 11, 78, 80 and 82.)

R. J. Allison, S. P. Goodwin, R. J. Parker, S. F. Portegies Zwart, and R. de Grijs. The early dynamical evolution of cool, clumpy star clusters. MNRAS, 407(2):1098–1107, 2010. (Cited on page 81.)

J. Alves, C. Zucker, A. A. Goodman, J. S. Speagle, S. Meingast, T. Robitaille, D. P. Finkbeiner, E. F. Schlafly, and G. M. Green. A Galactic-scale gas wave in the solar neighbourhood. Nature, 578(7794):237–239, 2020. (Cited on page 56.)

M. S. Angelo, W. J. B. Corradi, J. Santos, J. F. C., F. F. S. Maia, and F. A. Ferreira. Characterizing dynamical states of Galactic open clusters with Gaia DR2. MNRAS, 500(4):4338–4353, 2021. (Cited on pages 18, 70 and 72.)

B. Anguiano, S. R. Majewski, K. C. Freeman, A. W. Mitschang, and M. C. Smith. The velocity ellipsoid in the Galactic disc using Gaia DR1. MNRAS, 474(1):854–865, 2018. (Cited on page 101.)

T. Antoja, F. Figueras, D. Fernández, and J. Torra. Origin and evolution of moving groups. I. Characterization in the observational kinematic-age-metallicity space. A&A, 490(1):135–150, 2008. (Cited on pages 107 and 108.)

T. Antoja, A. Helmi, O. Bienayme, J. Bland -Hawthorn, B. Famaey, K. Freeman, B. K. Gibson, G. Gilmore, et al. Kinematic groups beyond the solar neighbourhood with RAVE. MNRAS, 426(1):L1–L5, 2012. (Cited on page 84.)

T. Antoja, A. Helmi, M. Romero-Gómez, D. Katz, C. Babusiaux, R. Drimmel, D. W. Evans, F. Figueras, et al. A dynamically young and perturbed Milky Way disk. Nature, 561(7723):360–362, 2018. (Cited on pages 17, 37, 94 and 176.)

N. M. Artyukhina and P. N. Kholopov. The Open Cluster M 37 and the Coronas of Star Clusters. Soviet Ast., 7:840, 1964. (Cited on page 42.)

T. L. Astraatmadja and C. A. L. Bailer-Jones. Estimating Distances from Parallaxes. II. Performance of Bayesian Distance Estimators on a Gaia-like Catalogue. ApJ, 832(2): 137, 2016. (Cited on page 60.)

M. Aumer and J. J. Binney. Kinematics and history of the solar neighbourhood revisited. MNRAS, 397(3):1286–1301, 2009. (Cited on page 101.)

M. Aumer, J. Binney, and R. Schönrich. Age-velocity dispersion relations and heating histories in disc galaxies. MNRAS, 462(2):1697–1713, 2016. (Cited on page 101.)

C. A. L. Bailer-Jones, J. Rybizki, M. Fouesneau, G. Mantelet, and R. Andrae. Estimating Distance from Parallaxes. IV. Distances to 1.33 Billion Stars in Gaia Data Release 2. AJ, 156(2):58, 2018. (Cited on pages 57, 60 and 62.)

C. A. L. Bailer-Jones. Estimating Distances from Parallaxes. PASP, 127(956):994, 2015. (Cited on pages 56, 58 and 60.)

D. A. Barros, J. R. D. Lépine, and W. S. Dias. Models for the 3D axisymmetric gravitational potential of the Milky Way galaxy. A detailed modelling of the Galactic disk. A&A, 593:A108, 2016. (Cited on page 96.)

D. A. Barros, A. Pérez-Villegas, J. R. D. Lépine, T. A. Michtchenko, and R. S. S. Vieira. Exploring the Origin of Moving Groups and Diagonal Ridges by Simulations of Stellar Orbits and Birthplaces. ApJ, 888(2):75, 2020. (Cited on page 94.)

N. Bastian. Cluster Disruption: From Infant Mortality to Long Term Survival. In *Stellar Clusters & Associations: A RIA Workshop on Gaia*, pages 85–97, 2011. (Cited on pages 28 and 166.)

N. Bastian, M. Gieles, B. Ercolano, and R. Gutermuth. The spatial evolution of stellar structures in the Large Magellanic Cloud. MNRAS, 392(2):868–878, 2009. (Cited on page 78.)

H. Baumgardt and P. Kroupa. A comprehensive set of simulations studying the influence of gas expulsion on star cluster evolution. MNRAS, 380(4):1589–1598, 2007. (Cited on pages 29, 56 and 168.)

H. Baumgardt and J. Makino. Dynamical evolution of star clusters in tidal fields. MNRAS, 340(1):227–246, 2003. (Cited on pages 30 and 169.)

G. Beccari, H. M. J. Boffin, and T. Jerabkova. Uncovering a 260 pc wide, 35-Myr-old filamentary relic of star formation. MNRAS, 491(2):2205–2216, 2020. (Cited on page 44.)

S. Bhattacharya, M. Agarwal, K. K. Rao, and K. Vaidya. Tidal tails in the disintegrating open cluster NGC 752. MNRAS, 2021. (Cited on pages 42 and 77.)

J. Binney and S. Tremaine. *Galactic dynamics*. 1987. (Cited on pages 30, 68 and 168.)

J. Binney and S. Tremaine. *Galactic Dynamics: Second Edition*. 2008. (Cited on pages 33, 101, 107 and 172.)

C. Bonatto, L. O. Kerber, E. Bica, and B. X. Santiago. Probing disk properties with open clusters. A&A, 446(1):121–135, 2006. (Cited on pages 31, 38, 170 and 177.)

I. A. Bonnell and M. B. Davies. Mass segregation in young stellar clusters. MNRAS, 295 (3):691–698, 1998. (Cited on pages 56, 78 and 80.)

D. Bossini, A. Vallenari, A. Bragaglia, T. Cantat-Gaudin, R. Sordo, L. Balaguer-Núñez, C. Jordi, A. Moitinho, et al. Age determination for 269 Gaia DR2 open clusters. A&A, 623:A108, 2019. (Cited on pages 64 and 85.)

D. Boubert, J. Strader, D. Aguado, G. Seabroke, S. E. Koposov, J. L. Sanders, S. Swihart, L. Chomiuk, and N. W. Evans. Lessons from the curious case of the 'fastest' star in Gaia DR2. MNRAS, 486(2):2618–2630, 2019. (Cited on page 86.)

J. Bovy. galpy: A python Library for Galactic Dynamics. ApJS, 216(2):29, 2015. (Cited on page 104.)

A. Bressan, P. Marigo, L. Girardi, B. Salasnich, C. Dal Cero, S. Rubele, and A. Nanni. PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code. MNRAS, 427(1):127–145, 2012. (Cited on page 139.)

S. Buder, S. Sharma, J. Kos, A. M. Amarsi, T. Nordlander, K. Lind, S. L. Martell, M. Asplund, et al. The GALAH+ survey: Third data release. MNRAS, 506(1):150–201, 2021. (Cited on pages 85 and 87.)

R. J. G. B. Campello, D. Moulavi, and J. Sander. Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, editors, *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-37456-2. (Cited on page 48.)

T. Cantat-Gaudin and F. Anders. Clusters and mirages: cataloguing stellar aggregates in the Milky Way. A&A, 633:A99, 2020. (Cited on page 85.)

T. Cantat-Gaudin, C. Jordi, A. Vallenari, A. Bragaglia, L. Balaguer-Núñez, C. Soubiran, D. Bossini, A. Moitinho, et al. A Gaia DR2 view of the open cluster population in the Milky Way. A&A, 618:A93, 2018a. (Cited on pages 23, 31, 37, 47, 48, 85, 139, 170 and 176.)

T. Cantat-Gaudin, A. Vallenari, R. Sordo, F. Pensabene, A. Krone-Martins, A. Moitinho, C. Jordi, L. Casamiquela, et al. Characterising open clusters in the solar neighbourhood with the Tycho-Gaia Astrometric Solution. A&A, 615:A49, 2018b. (Cited on page 75.)

T. Cantat-Gaudin, C. Jordi, N. J. Wright, J. J. Armstrong, A. Vallenari, L. Balaguer-Núñez, P. Ramos, D. Bossini, et al. Expanding associations in the Vela-Puppis region. 3D structure and kinematics of the young population. A&A, 626:A17, 2019a. (Cited on page 44.)

T. Cantat-Gaudin, A. Krone-Martins, N. Sedaghat, A. Farahi, R. S. de Souza, R. Skalidis, A. I. Malz, S. Macêdo, et al. Gaia DR2 unravels incompleteness of nearby cluster population: new open clusters in the direction of Perseus. A&A, 624:A126, 2019b. (Cited on pages 37 and 176.)

T. Cantat-Gaudin, F. Anders, A. Castro-Ginard, C. Jordi, M. Romero-Gómez, C. Soubiran, L. Casamiquela, Y. Tarricq, et al. Painting a portrait of the Galactic disc with its stellar clusters. A&A, 640:A1, 2020. (Cited on pages 10, 12, 23, 38, 47, 177 and 178.)

R. Carrera, A. Bragaglia, T. Cantat-Gaudin, A. Vallenari, L. Balaguer-Núñez, D. Bossini, L. Casamiquela, C. Jordi, et al. Open clusters in APOGEE and GALAH. Combining Gaia and ground-based spectroscopic surveys. A&A, 623:A80, 2019a. (Cited on pages 49, 71, 85, 90, 112 and 182.)

R. Carrera, M. Pasquato, A. Vallenari, L. Balaguer-Núñez, T. Cantat-Gaudin, M. Mapelli, A. Bragaglia, D. Bossini, et al. Extended halo of NGC 2682 (M 67) from Gaia DR2. A&A, 627:A119, 2019b. (Cited on page 60.)

R. Carrera, L. Casamiquela, J. Carbajo-Hijarrubia, L. Balaguer-Núñez, C. Jordi, M. Romero-Gómez, S. Blanco-Cuaresma, T. Cantat-Gaudin, et al. OCCASO IV. Radial Velocities and Open Cluster Kinematics. *arXiv e-prints*, art. arXiv:2110.02110, 2021. (Cited on page 85.)

L. Casamiquela, R. Carrera, C. Jordi, L. Balaguer-Núñez, E. Pancino, S. L. Hidalgo, C. E. Martínez-Vázquez, S. Murabito, et al. The OCCASO survey: presentation and radial velocities of 12 Milky Way open clusters. MNRAS, 458(3):3150–3167, 2016. (Cited on pages 86 and 87.)

L. Casamiquela, Y. Tarricq, C. Soubiran, S. Blanco-Cuaresma, P. Jofré, U. Heiter, and M. Tucci Maia. Differential abundances of open clusters and their tidal tails: Chemical tagging and chemical homogeneity. A&A, 635:A8, 2020. (Cited on pages 20, 28, 34, 46, 139, 140, 167 and 173.)

A. Castro-Ginard, C. Jordi, X. Luri, F. Julbe, M. Morvan, L. Balaguer-Núñez, and T. Cantat-Gaudin. A new method for unveiling open clusters in Gaia. New nearby open clusters confirmed by DR2. A&A, 618:A59, 2018. (Cited on pages 37, 44, 47, 48, 66, 85 and 176.)

A. Castro-Ginard, C. Jordi, X. Luri, T. Cantat-Gaudin, and L. Balaguer-Núñez. Hunting for open clusters in Gaia DR2: the Galactic anticentre. A&A, 627:A35, 2019. (Cited on pages 37, 47, 48, 66, 85 and 176.)

A. Castro-Ginard, C. Jordi, X. Luri, J. Álvarez Cid-Fuentes, L. Casamiquela, F. Anders, T. Cantat-Gaudin, M. Monguió, et al. Hunting for open clusters in Gaia DR2: 582 new open clusters in the Galactic disc. A&A, 635:A45, 2020. (Cited on pages 37, 47, 48, 66, 85, 90, 111, 139, 176 and 181.)

A. Castro-Ginard, P. J. McMillan, X. Luri, C. Jordi, M. Romero-Gómez, T. Cantat-Gaudin, L. Casamiquela, Y. Tarricq, et al. Milky Way spiral arms from open clusters in Gaia EDR3. A&A, 652:A162, 2021. (Cited on pages 28, 38, 167 and 177.)

B. Q. Chen, Y. Huang, H. B. Yuan, C. Wang, D. W. Fan, M. S. Xiang, H. W. Zhang, Z. J. Tian, and X. W. Liu. Three-dimensional interstellar dust reddening maps of the Galactic plane. MNRAS, 483(4):4277–4289, 2019. (Cited on page 60.)

W. P. Chen, C. W. Chen, and C. G. Shu. Morphology of Galactic Open Clusters. AJ, 128(5):2306–2315, 2004. (Cited on page 76.)

E. Chereul, M. Crézé, and O. Bienaymé. The distribution of nearby stars in phase space mapped by Hipparcos. Clustering and streaming among A-F type stars. A&AS, 135: 5–28, 1999. (Cited on pages 84 and 93.)

P. C. Clark, I. A. Bonnell, and M. R. Bate. Star Formation in Unbound GMCs. In H. J. G. L. M. Lamers, L. J. Smith, and A. Nota, editors, *The Formation and Evolution of Massive Young Star Clusters*, volume 322 of *Astronomical Society of the Pacific Conference Series*, page 381, 2004. (Cited on pages 29 and 168.)

D. P. Clemens. Massachusetts-Stony Brook Galactic plane CO survey: the galactic disk rotation curve. ApJ, 295:422–436, 1985. (Cited on pages 33 and 171.)

X.-Q. Cui, Y.-H. Zhao, Y.-Q. Chu, G.-P. Li, Q. Li, L.-P. Zhang, H.-J. Su, Z.-Q. Yao, et al. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST). *Research in Astronomy and Astrophysics*, 12(9):1197–1242, 2012. (Cited on page 85.)

R. de Grijs, G. F. Gilmore, and R. Johnson. Primordial or Dynamical Mass Segregation in Young LMC Clusters? In G. Piotto, G. Meylan, S. G. Djorgovski, and M. Riello, editors, *New Horizons in Globular Cluster Astronomy*, volume 296 of *Astronomical Society of the Pacific Conference Series*, page 207, 2003. (Cited on page 80.)

R. de La Fuente Marcos. The initial mass function and the dynamical evolution of open clusters. III. With primordial binaries. A&A, 314:453–464, 1996. (Cited on pages 31, 56 and 169.)

W. Dehnen and J. J. Binney. Local stellar kinematics from HIPPARCOS data. MNRAS, 298(2):387–394, 1998. (Cited on pages 84 and 93.)

W. S. Dias, B. S. Alessi, A. Moitinho, and J. R. D. Lépine. New catalogue of optically visible open clusters and candidates. A&A, 389:871–873, 2002. (Cited on pages 29, 31, 33, 37, 42, 64, 70, 90, 167, 170, 172 and 176.)

W. S. Dias, H. Monteiro, T. C. Caetano, J. R. D. Lépine, M. Assafin, and A. F. Oliveira. Proper motions of the optically visible open clusters based on the UCAC4 catalog. A&A, 564:A79, 2014. (Cited on pages 34 and 173.)

W. S. Dias, H. Monteiro, J. R. D. Lépine, and D. A. Barros. The spiral pattern rotation speed of the Galaxy and the corotation radius with Gaia DR2. MNRAS, 486(4):5726–5736, 2019. (Cited on pages 94 and 96.)

W. S. Dias and J. R. D. Lépine. Direct Determination of the Spiral Pattern Rotation Speed of the Galaxy. ApJ, 629(2):825–831, 2005. (Cited on pages 17, 28, 31, 32, 33, 38, 167, 170, 171, 172 and 178.)

S. Dib, S. Schmeja, and R. J. Parker. Structure and mass segregation in Galactic stellar clusters. MNRAS, 473(1):849–859, 2018. (Cited on pages 72, 78, 79 and 80.)

F. Dinnbier and P. Kroupa. Tidal tails of open star clusters as probes of early gas expulsion. I. A semi-analytic model. A&A, 640:A84, 2020a. (Cited on page 56.)

F. Dinnbier and P. Kroupa. Tidal tails of open star clusters as probes to early gas expulsion. II. Predictions for Gaia. A&A, 640:A85, 2020b. (Cited on page 56.)

J. Donor, P. M. Frinchaboy, K. Cunha, J. E. O'Connell, C. Allende Prieto, A. Almeida, F. Anders, R. Beaton, et al. The Open Cluster Chemical Abundances and Mapping Survey. IV. Abundances for 128 Open Clusters Using SDSS/APOGEE DR16. AJ, 159 (5):199, 2020. (Cited on pages 34 and 173.)

J. L. E. Dreyer. A New General Catalogue of Nebulæ and Clusters of Stars, being the Catalogue of the late Sir John F. W. Herschel, Bart, revised, corrected, and enlarged. MmRAS, 49:1, 1888. (Cited on pages 26 and 164.)

J. L. E. Dreyer. Index Catalogue of Nebulæ found in the years 1888 to 1894, with Notes and Corrections to the New General Catalogue. MmRAS, 51:185, 1895. (Cited on pages 26 and 164.)

J. L. E. Dreyer. Second Index Catalogue of Nebulæ and Clusters of Stars, containing objects found in the years 1895 to 1907; with Notes and Corrections to the New General Catalogue and to the Index Catalogue for 1888-94. MmRAS, 59:105, 1910. (Cited on pages 26 and 164.)

O. J. Eggen. Stellar groups. I. The Hyades and Sirius groups. MNRAS, 118:65, 1958. (Cited on pages 84 and 93.)

O. J. Eggen. Star Streams and Galactic Structure. AJ, 112:1595, 1996. (Cited on page 84.)

R. A. W. Elson, S. M. Fall, and K. C. Freeman. The Structure of Young Star Clusters in the Large Magellanic Cloud. ApJ, 323:54, 1987. (Cited on pages 30, 61, 66, 67 and 168.)

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, page 226–231. AAAI Press, 1996. (Cited on page 48.)

C. Fabricius, X. Luri, F. Arenou, C. Babusiaux, A. Helmi, T. Muraveva, C. Reylé, F. Spoto, et al. Gaia Early Data Release 3. Catalogue validation. A&A, 649:A5, 2021. (Cited on page 47.)

B. Famaey, A. Siebert, and A. Jorissen. On the age heterogeneity of the Pleiades, Hyades, and Sirius moving groups. A&A, 483(2):453–459, 2008. (Cited on page 84.)

D. K. Feuillet, N. Frankel, K. Lind, P. M. Frinchaboy, D. A. García-Hernández, R. R. Lane, C. Nitschelm, and A. Roman-Lopes. Spatial variations in the Milky Way disc metallicity-age relation. MNRAS, 489(2):1742–1752, 2019. (Cited on page 60.)

D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. PASP, 125(925):306, 2013. (Cited on page 68.)

K. C. Freeman. The galactic spheroid and old disk. ARA&A, 25:603–632, 1987. (Cited on page 101.)

E. D. Friel. The Old Open Clusters Of The Milky Way. ARA&A, 33:381–414, 1995. (Cited on pages 28, 31, 34, 167, 170 and 172.)

Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, et al. The Gaia mission. A&A, 595:A1, 2016. (Cited on pages 35 and 174.)

Gaia Collaboration, C. Babusiaux, F. van Leeuwen, M. A. Barstow, C. Jordi, A. Vallenari, D. Bossini, A. Bressan, et al. Gaia Data Release 2. Observational Hertzsprung-Russell diagrams. A&A, 616:A10, 2018a. (Cited on pages 57, 58, 62, 64, 85 and 139.)

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, et al. Gaia Data Release 2. Summary of the contents and survey properties. A&A, 616:A1, 2018b. (Cited on pages 35, 42, 57, 84 and 174.)

Gaia Collaboration, A. Helmi, F. van Leeuwen, P. J. McMillan, D. Massari, T. Antoja, A. C. Robin, L. Lindegren, et al. Gaia Data Release 2. Kinematics of globular clusters and dwarf galaxies around the Milky Way. A&A, 616:A12, 2018c. (Cited on page 104.)

Gaia Collaboration, D. Katz, T. Antoja, M. Romero-Gómez, R. Drimmel, C. Reylé, G. M. Seabroke, C. Soubiran, et al. Gaia Data Release 2. Mapping the Milky Way disc kinematics. A&A, 616:A11, 2018d. (Cited on pages 84, 91, 93, 94, 95, 98, 107, 113 and 183.)

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, M. Biermann, O. L. Creevey, et al. Gaia Early Data Release 3. Summary of the contents and survey properties. A&A, 649:A1, 2021. (Cited on page 42.)

X. Gao. Discovery of Tidal Tails around the Old Open Cluster NGC 2506. ApJ, 894(1): 48, 2020. (Cited on pages 42 and 77.)

Y. M. Georgelin and Y. P. Georgelin. The spiral structure of our Galaxy determined from H II regions. A&A, 49:57–79, 1976. (Cited on pages 33 and 171.)

M. Gieles, S. F. Portegies Zwart, H. Baumgardt, E. Athanassoula, H. J. G. L. M. Lamers, M. Sipior, and J. Leenaarts. Star cluster disruption by giant molecular clouds. MNRAS, 371(2):793–804, 2006. (Cited on pages 30 and 168.)

M. Gieles, E. Athanassoula, and S. F. Portegies Zwart. The effect of spiral arm passages on the evolution of stellar clusters. MNRAS, 376(2):809–819, 2007. (Cited on pages 30 and 169.)

M. Gieles and S. F. Portegies Zwart. The distinction between star clusters and associations. MNRAS, 410(1):L6–L7, 2011. (Cited on pages 28 and 166.)

J. Goodman and J. Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, 2010. (Cited on page 68.)

R. J. J. Grand, V. Springel, F. A. Gómez, F. Marinacci, R. Pakmor, D. J. R. Campbell, and A. Jenkins. Vertical disc heating in Milky Way-sized galaxies in a cosmological context. MNRAS, 459(1):199–219, 2016. (Cited on page 101.)

R. Gratton, A. Bragaglia, E. Carretta, V. D'Orazi, S. Lucatello, and A. Sollima. What is a globular cluster? An observational perspective. A&A Rev., 27(1):8, 2019. (Cited on pages 27 and 165.)

R. G. Gratton, E. Carretta, and A. Bragaglia. Multiple populations in globular clusters. Lessons learned from the Milky Way globular clusters. A&A Rev., 20:50, 2012. (Cited on pages 27 and 166.)

N. A. Grice and D. W. Dawson. The Open Cluster NGC 6716. PASP, 102:881, 1990. (Cited on page 49.)

B. Gustafsson, R. P. Church, M. B. Davies, and H. Rickman. Gravitational scattering of stars and clusters and the heating of the Galactic disk. A&A, 593:A85, 2016. (Cited on page 103.)

R. A. Gutermuth, P. C. Myers, S. T. Megeath, L. E. Allen, J. L. Pipher, J. Muzerolle, A. Porras, E. Winston, and G. Fazio. Spitzer Observations of NGC 1333: A Study of Structure and Evolution in a Nearby Embedded Cluster. ApJ, 674(1):336–356, 2008. (Cited on pages 28, 29, 166 and 168.)

D. Heggie and P. Hut. *The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics*. 2003. (Cited on page 70.)

A. Helmi, C. Babusiaux, H. H. Koppelman, D. Massari, J. Veljanoski, and A. G. A. Brown. The merger that led to the formation of the Milky Way's inner stellar halo and thick disk. Nature, 563(7729):85–88, 2018. (Cited on pages 36 and 174.)

L. A. Hillenbrand and L. W. Hartmann. A Preliminary Study of the Orion Nebula Cluster Structure and Dynamics. ApJ, 492(2):540–553, 1998. (Cited on page 56.)

E. Høg, C. Fabricius, V. V. Makarov, S. Urban, T. Corbin, G. Wycoff, U. Bastian, P. Schwekendiek, and A. Wicenec. The Tycho-2 catalogue of the 2.5 million brightest stars. A&A, 355:L27–L30, 2000. (Cited on pages 34 and 173.)

M. Hoskin. John Herschel's Cosmology. *Journal for the History of Astronomy*, 18:1, 1987. (Cited on pages 26 and 164.)

L. G. Hou and J. L. Han. The observed spiral structure of the Milky Way. A&A, 569: A125, 2014. (Cited on pages 31 and 170.)

J. Hron. Kinematics of young open clusters and the rotation curve of our Galaxy. A&A, 176:34–52, 1987. (Cited on pages 33 and 171.)

Q. Hu, Y. Zhang, A. Esamdin, J. Liu, and X. Zeng. Deciphering Star Cluster Evolution by Shape Morphology. ApJ, 912(1):5, 2021. (Cited on pages 29, 75, 76 and 168.)

E. L. Hunt and S. Reffert. Improving the open cluster census. I. Comparison of clustering algorithms applied to Gaia DR2 data. A&A, 646:A104, 2021. (Cited on page 48.)

R. J. Jackson, R. D. Jeffries, J. Lewis, S. E. Koposov, G. G. Sacco, S. Randich, G. Gilmore, M. Asplund, et al. The Gaia-ESO Survey: Empirical determination of the precision of stellar radial velocities and projected rotation velocities. A&A, 580:A75, 2015. (Cited on page 86.)

K. A. Janes. Evidence for an abundance gradient in the galactic disk. ApJS, 39:135–156, 1979. (Cited on pages 34 and 172.)

J. H. Jeans. The Stability of a Spherical Nebula. *Philosophical Transactions of the Royal Society of London Series A*, 199:1–53, 1902. (Cited on pages 29 and 167.)

A. Jenkins. Heating of galactic discs with realistic vertical potentials. MNRAS, 257(4): 620–632, 1992. (Cited on page 101.)

A. Jenkins and J. Binney. Spiral heating of galactic discs. MNRAS, 245:305–317, 1990. (Cited on pages 31, 103, 104 and 170.)

T. Jerabkova, H. M. J. Boffin, G. Beccari, G. de Marchi, J. H. J. de Bruijne, and T. Prusti. The 800 pc long tidal tails of the Hyades star cluster. Possible discovery of candidate epicyclic overdensities from an open star cluster. A&A, 647:A137, 2021. (Cited on page 42.)

D. R. H. Johnson and D. R. Soderblom. Calculating Galactic Space Velocities and Their Uncertainties, with an Application to the Ursa Major Group. AJ, 93:864, 1987. (Cited on page 43.)

D. Katz, P. Sartoretti, M. Cropper, P. Panuzzo, G. M. Seabroke, Y. Viala, K. Benson, R. Blomme, et al. Gaia Data Release 2. Properties and validation of the radial velocities. A&A, 622:A205, 2019. (Cited on pages 35, 43, 85 and 174.)

D. Kawata, J. Baba, I. Ciucă, M. Cropper, R. J. J. Grand, J. A. S. Hunt, and G. Seabroke. Radial distribution of stellar motions in Gaia DR2. MNRAS, 479(1):L108–L112, 2018. (Cited on page 94.)

S. Khanna, S. Sharma, T. Tepper-Garcia, J. Bland -Hawthorn, M. Hayden, M. Asplund, S. Buder, B. Chen, et al. The GALAH survey and Gaia DR2: Linking ridges, arches, and vertical waves in the kinematics of the Milky Way. MNRAS, 489(4):4962–4979, 2019. (Cited on pages 84 and 94.)

N. V. Kharchenko, A. E. Piskunov, E. Schilbach, S. Röser, and R. D. Scholz. Global survey of star clusters in the Milky Way. II. The catalogue of basic parameters. A&A, 558:A53, 2013. (Cited on pages 18, 29, 34, 37, 42, 64, 70, 72, 90, 111, 167, 173, 176 and 181.)

I. King. The structure of star clusters. I. an empirical density law. AJ, 67:471, 1962. (Cited on pages 30, 61, 66 and 168.)

M. Kounkel and K. Covey. Untangling the Galaxy. I. Local Structure and Star Formation History of the Milky Way. AJ, 158(3):122, 2019. (Cited on pages 38, 48 and 177.)

D. A. Kovaleva, M. Ishchenko, E. Postnikova, P. Berczik, A. E. Piskunov, N. V. Kharchenko, E. Polyachenko, S. Reffert, et al. Collinder 135 and UBC 7: A physical pair of open clusters. A&A, 642:L4, 2020. (Cited on page 46.)

A. Krone-Martins and A. Moitinho. UPMASK: unsupervised photometric membership assignment in stellar clusters. A&A, 561:A57, 2014. (Cited on pages 37, 85 and 176.)

P. Kroupa. The Fundamental Building Blocks of Galaxies. In C. Turon, K. S. O'Flaherty, and M. A. C. Perryman, editors, *The Three-Dimensional Universe with Gaia*, volume 576 of *ESA Special Publication*, page 629, 2005. (Cited on pages 29 and 168.)

P. Kroupa. Star cluster evolution, dynamical age estimation and the kinematical signature of star formation. MNRAS, 277:1522, 1995. (Cited on pages 31, 56, 78 and 169.)

P. Kroupa. *Initial Conditions for Star Clusters*, volume 760, page 181. 2008. (Cited on page 64.)

M. R. Krumholz, C. F. McKee, and J. Bland-Hawthorn. Star Clusters Across Cosmic Time. ARA&A, 57:227–303, 2019. (Cited on pages 27, 56, 64 and 165.)

K. Kuijken and S. Tremaine. On the Ellipticity of the Galactic Disk. ApJ, 421:178, 1994. (Cited on page 101.)

A. H. W. Küpper, A. MacLeod, and D. C. Heggie. On the structure of tidal tails. MNRAS, 387(3):1248–1252, 2008. (Cited on page 56.)

A. H. W. Küpper, P. Kroupa, H. Baumgardt, and D. C. Heggie. Peculiarities in velocity dispersion and surface density profiles of star clusters. MNRAS, 407(4):2241–2260, 2010. (Cited on pages 66 and 71.)

A. H. W. Küpper, T. Maschberger, P. Kroupa, and H. Baumgardt. Mass segregation and fractal substructure in young massive clusters - I. The McLuster code and method calibration. MNRAS, 417(3):2300–2317, 2011. (Cited on page 68.)

C. G. Lacey. The influence of massive gas clouds on stellar velocity dispersions in galactic discs. MNRAS, 208:687–707, 1984. (Cited on pages 31, 103 and 170.)

C. J. Lada. The physics and modes of star cluster formation: observations. *Philosophical Transactions of the Royal Society of London Series A*, 368(1913):713–731, 2010. (Cited on pages 29, 113, 167 and 184.)

C. J. Lada, J. Alves, and E. A. Lada. Near-Infrared Imaging of Embedded Clusters: NGC 1333. AJ, 111:1964, 1996. (Cited on pages 29 and 168.)

C. J. Lada and E. A. Lada. Embedded Clusters in Molecular Clouds. ARA&A, 41: 57–115, 2003. (Cited on pages 27, 29, 30, 56, 78, 113, 164 and 168.)

C. J. Lada, E. T. Young, and T. P. Greene. Infrared Images of the Young Cluster NGC 2264. ApJ, 408:471, 1993. (Cited on pages 29 and 168.)

H. J. G. L. M. Lamers and M. Gieles. Clusters in the solar neighbourhood: how are they destroyed? A&A, 455(2):L17–L20, 2006. (Cited on pages 30, 56 and 169.)

C. C. Lin and F. H. Shu. On the Spiral Structure of Disk Galaxies. ApJ, 140:646, 1964. (Cited on pages 39 and 178.)

L. Lindegren, J. Hernández, A. Bombrun, S. Klioner, U. Bastian, M. Ramos-Lerate, A. de Torres, H. Steidelmüller, et al. Gaia Data Release 2. The astrometric solution. A&A, 616:A2, 2018. (Cited on page 60.)

L. Lindegren, U. Bastian, M. Biermann, A. Bombrun, A. de Torres, E. Gerlach, R. Geyer, J. Hernández, et al. Gaia Early Data Release 3. Parallax bias versus magnitude, colour, and position. A&A, 649:A4, 2021a. (Cited on page 57.)

L. Lindegren, S. A. Klioner, J. Hernández, A. Bombrun, M. Ramos-Lerate, H. Steidelmüller, U. Bastian, M. Biermann, et al. Gaia Early Data Release 3. The astrometric solution. A&A, 649:A2, 2021b. (Cited on pages 35 and 174.)

F. Liu, D. Yong, M. Asplund, I. Ramírez, and J. Meléndez. The Hyades open cluster is chemically inhomogeneous. MNRAS, 457(4):3934–3948, 2016. (Cited on pages 34 and 173.)

L. Liu and X. Pang. A Catalog of Newly Identified Star Clusters in Gaia DR2. ApJS, 245(2):32, 2019. (Cited on pages 38, 85, 90 and 177.)

X. Luri, A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, A. Castro-Ginard, J. de Bruijne, T. Prusti, et al. Gaia Data Release 2. Using Gaia parallaxes. A&A, 616:A9, 2018. (Cited on page 58.)

G. Lynga. Open clusters in our galaxy. A&A, 109:213–222, 1982. (Cited on pages 28 and 167.)

J. T. Mackereth, J. Bovy, H. W. Leung, R. P. Schiavon, W. H. Trick, W. J. Chaplin, K. Cunha, D. K. Feuillet, et al. Dynamical heating across the Milky Way disc using APOGEE and Gaia. MNRAS, 489(1):176–195, 2019. (Cited on pages 101, 102 and 103.)

J. Maindonald and W. J. Braun. *Styles of data analysis*, page 43–76. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 3 edition, 2010. (Cited on page 95.)

S. Maran and L. Marschall. *Galileo's New Universe: The Revolution in Our Understanding of the Cosmos*. BenBella Books, 2009. ISBN 9781935251866. (Cited on pages 26 and 164.)

M. Martig, I. Minchev, and C. Flynn. Dissecting simulated disc galaxies - II. The age-velocity relation. MNRAS, 443(3):2452–2462, 2014. (Cited on page 101.)

L. Martinez-Medina, B. Pichardo, A. Peimbert, and O. Valenzuela. From ridges in the velocity distribution to wiggles in the rotation curve. MNRAS, 485(1):L104–L108, 2019. (Cited on pages 96 and 97.)

T. Maschberger, C. J. Clarke, I. A. Bonnell, and P. Kroupa. Properties of hierarchically forming star clusters. MNRAS, 404(2):1061–1080, 2010. (Cited on pages 79 and 81.)

R. D. Mathieu. The structure and dynamics of the open cluster M 11. ApJ, 284:643–662, 1984. (Cited on pages 31, 56 and 169.)

S. S. McGaugh. The Imprint of Spiral Arms on the Galactic Rotation Curve. ApJ, 885 (1):87, 2019. (Cited on page 96.)

L. McInnes, J. Healy, and S. Astels. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205, 2017. URL `https://doi.org/10.21105/joss.00205`. (Cited on page 48.)

C. F. McKee and E. C. Ostriker. Theory of Star Formation. ARA&A, 45(1):565–687, 2007. (Cited on pages 29 and 167.)

P. J. McMillan. The mass distribution and gravitational potential of the Milky Way. MNRAS, 465(1):76–94, 2017. (Cited on page 104.)

P. J. McMillan and J. J. Binney. Disassembling the Galaxy with angle-action coordinates. MNRAS, 390(1):429–437, 2008. (Cited on page 105.)

S. L. W. McMillan, E. Vesperini, and S. F. Portegies Zwart. A Dynamical Origin for Early Mass Segregation in Young Star Clusters. ApJ, 655(1):L45–L49, 2007. (Cited on pages 56 and 81.)

S. Meingast and J. Alves. Extended stellar systems in the solar neighborhood. I. The tidal tails of the Hyades. A&A, 621:L3, 2019. (Cited on pages 42, 43, 46, 62 and 139.)

S. Meingast, J. Alves, and A. Rottensteiner. Extended stellar systems in the solar neighborhood. V. Discovery of coronae of nearby star clusters. A&A, 645:A84, 2021. (Cited on pages 9, 29, 42, 50, 65, 77, 78, 111, 168 and 181.)

J. C. Mermilliod, M. Mayor, and S. Udry. Red giants in open clusters. XIV. Mean radial velocities for 1309 stars and 166 open clusters. A&A, 485:303–314, 2008. (Cited on pages 86 and 87.)

J.-C. Mermilliod, M. Mayor, and S. Udry. Catalogues of radial and rotational velocities of 1253 F-K dwarfs in 13 nearby open clusters. A&A, 498:949–960, 2009. (Cited on pages 86 and 87.)

J.-C. Mermilliod. *The Database for Galactic Open Clusters (BDA)*, volume 203, page 127. 1995. (Cited on pages 29 and 167.)

M. Miyamoto and R. Nagai. Three-dimensional models for the distribution of mass in galaxies. PASJ, 27:533–543, 1975. (Cited on page 104.)

S. Molinari, J. Bally, S. Glover, T. Moore, A. Noriega-Crespo, R. Plume, L. Testi, E. Vázquez-Semadeni, et al. The Milky Way as a Star Formation Engine. In H. Beuther, R. S. Klessen, C. P. Dullemond, and T. Henning, editors, *Protostars and Planets VI*, page 125, 2014. (Cited on pages 29 and 167.)

G. Monari, B. Famaey, I. Minchev, T. Antoja, O. Bienaymé, B. K. Gibson, E. K. Grebel, G. Kordopatis, et al. Coma Berenices: The First Evidence for Incomplete Vertical Phase-mixing in Local Velocity Space with RAVE—Confirmed with Gaia DR2. *Research Notes of the American Astronomical Society*, 2(2):32, 2018. (Cited on page 84.)

P. Moore and R. Rees. *Patrick Moore's Data Book of Astronomy*. Cambridge University Press, 2 edition, 2011. (Cited on pages 26 and 164.)

E. Moraux. Open clusters and associations in the Gaia era. In *EAS Publications Series*, volume 80-81 of *EAS Publications Series*, pages 73–114, 2016. (Cited on pages 27, 28, 165 and 166.)

P. Mróz, A. Udalski, D. M. Skowron, J. Skowron, I. Soszyński, P. Pietrukowicz, M. K. Szymański, R. Poleski, et al. Rotation Curve of the Milky Way from Classical Cepheids. ApJ, 870(1):L10, 2019. (Cited on page 96.)

H. Nakanishi and Y. Sofue. Three-dimensional distribution of the ISM in the Milky Way galaxy. III. The total neutral gas disk. PASJ, 68(1):5, 2016. (Cited on page 96.)

J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. ApJ, 490(2):493–508, 1997. (Cited on page 104.)

Nilakshi, R. Sagar, A. K. Pandey, and V. Mohan. A study of spatial structure of galactic open star clusters. A&A, 383:153–162, 2002. (Cited on page 42.)

T. Nony, J. F. Robitaille, F. Motte, M. Gonzalez, I. Joncour, E. Moraux, A. Men'shchikov, P. Didelon, et al. Mass segregation and sequential star formation in NGC 2264 revealed by Herschel. A&A, 645:A94, 2021. (Cited on page 79.)

B. Nordström, M. Mayor, J. Andersen, J. Holmberg, F. Pont, B. R. Jørgensen, E. H. Olsen, S. Udry, and N. Mowlavi. The Geneva-Copenhagen survey of the Solar neighbourhood. Ages, metallicities, and kinematic properties of $\sim$ 14 000 F and G dwarfs. A&A, 418:989–1019, 2004. (Cited on pages 86, 87, 93 and 101.)

M. Odenkirchen, E. K. Grebel, C. M. Rockosi, W. Dehnen, R. Ibata, H.-W. Rix, A. Stolte, C. Wolf, et al. Detection of Massive Tidal Tails around the Globular Cluster Palomar 5 with Sloan Digital Sky Survey Commissioning Data. ApJ, 548(2):L165–L169, 2001. (Cited on pages 111 and 181.)

J. Olivares, E. Moraux, L. M. Sarro, H. Bouy, A. Berihuete, D. Barrado, N. Huelamo, E. Bertin, and J. Bouvier. The seven sisters DANCe. III. Projected spatial distribution. A&A, 612:A70, 2018a. (Cited on pages 67 and 71.)

J. Olivares, L. M. Sarro, E. Moraux, A. Berihuete, H. Bouy, S. Hernández-Jiménez, E. Bertin, P. A. B. Galli, et al. The seven sisters DANCe. IV. Bayesian hierarchical model. A&A, 617:A15, 2018b. (Cited on pages 34 and 173.)

J. Olivares, L. M. Sarro, H. Bouy, N. Miret-Roig, L. Casamiquela, P. A. B. Galli, A. Berihuete, and Y. Tarricq. Kalkayotl: A cluster distance inference code. A&A, 644: A7, 2020. (Cited on page 61.)

X. Pang, Y. Li, Z. Yu, S.-Y. Tang, F. Dinnbier, P. Kroupa, M. Pasquato, and M. B. N. Kouwenhoven. 3D Morphology of Open Clusters in the Solar Neighborhood with Gaia EDR 3: Its Relation to Cluster Dynamics. ApJ, 912(2):162, 2021. (Cited on pages 9, 78, 111 and 181.)

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL `http://jmlr.org/papers/v12/pedregosa11a.html`. (Cited on pages 63 and 74.)

M. A. C. Perryman, L. Lindegren, J. Kovalevsky, E. Hog, U. Bastian, P. L. Bernacca, M. Creze, F. Donati, et al. The Hipparcos Catalogue. A&A, 500:501–504, 1997. (Cited on pages 34 and 173.)

A. E. Piatti. Binary star sequence in the outskirts of the disrupting Galactic open cluster UBC 274. A&A, 639:A55, 2020. (Cited on page 90.)

M. Piecka and E. Paunzen. Structure of Open Clusters – Gaia DR2 and its limitations. *arXiv e-prints*, art. arXiv:2107.07230, 2021. (Cited on pages 18, 57, 58, 60, 62 and 63.)

G. Piotto, S. Villanova, L. R. Bedin, R. Gratton, S. Cassisi, Y. Momany, A. Recio-Blanco, S. Lucatello, et al. Metallicities on the Double Main Sequence of ω Centauri Imply Large Helium Enhancement. ApJ, 621(2):777–784, 2005. (Cited on pages 27 and 166.)

A. E. Piskunov, E. Schilbach, N. V. Kharchenko, S. Röser, and R. D. Scholz. Towards absolute scales for the radii and masses of open clusters. A&A, 468(1):151–161, 2007. (Cited on pages 18, 70 and 72.)

A. E. Piskunov, A. Just, N. V. Kharchenko, P. Berczik, R. D. Scholz, S. Reffert, and S. X. Yen. Global survey of star clusters in the Milky Way. VI. Age distribution and cluster formation history. A&A, 614:A22, 2018. (Cited on page 57.)

I. Platais, V. Kozhurina-Platais, and F. van Leeuwen. A Search for Star Clusters from the HIPPARCOS Data. AJ, 116(5):2423–2430, 1998. (Cited on pages 34 and 173.)

H. C. Plummer. On the problem of distribution in globular star clusters. MNRAS, 71: 460–470, 1911. (Cited on pages 66 and 67.)

A. L. Plunkett, M. Fernández-López, H. G. Arce, G. Busquet, D. Mardones, and M. M. Dunham. Distribution of Serpens South protostars revealed with ALMA. A&A, 615: A9, 2018. (Cited on page 79.)

S. F. Portegies Zwart, S. L. W. McMillan, and M. Gieles. Young Massive Star Clusters. ARA&A, 48:431–493, 2010. (Cited on pages 27 and 164.)

T. Preibisch and E. Mamajek. *The Nearest OB Association: Scorpius-Centaurus (Sco OB2)*, volume 5, page 235. 2008. (Cited on page 46.)

S. Randich, G. Gilmore, and Gaia-ESO Consortium. The Gaia-ESO Large Public Spectroscopic Survey. *The Messenger*, 154:47–49, 2013. (Cited on page 85.)

M. J. Reid, K. M. Menten, A. Brunthaler, X. W. Zheng, T. M. Dame, Y. Xu, Y. Wu, B. Zhang, et al. Trigonometric Parallaxes of High Mass Star Forming Regions: The Structure and Kinematics of the Milky Way. ApJ, 783(2):130, 2014. (Cited on pages 31, 99, 100 and 170.)

M. J. Reid, K. M. Menten, A. Brunthaler, X. W. Zheng, T. M. Dame, Y. Xu, J. Li, N. Sakai, et al. Trigonometric Parallaxes of High-mass Star-forming Regions: Our View of the Milky Way. ApJ, 885(2):131, 2019. (Cited on pages 95, 96, 99 and 100.)

W. W. Roberts. Large-Scale Shock Formation in Spiral Galaxies and its Implications on Star Formation. ApJ, 158:123, 1969. (Cited on pages 31 and 170.)

S. Roeser, M. Demleitner, and E. Schilbach. The PPMXL Catalog of Positions and Proper Motions on the ICRS. Combining USNO-B1.0 and the Two Micron All Sky Survey (2MASS). AJ, 139(6):2440–2447, 2010. (Cited on pages 29, 34, 42, 167 and 173.)

C. G. Román-Zúñiga, E. Alfaro, A. Palau, B. Hasenberger, J. F. Alves, M. Lombardi, and G. P. S. Sánchez. Not so different after all: properties and spatial structure of column density peaks in the pipe and Orion A clouds. MNRAS, 489(3):4429–4443, 2019. (Cited on page 79.)

S. Röser and E. Schilbach. Praesepe (NGC 2632) and its tidal tails. A&A, 627:A4, 2019. (Cited on pages 42, 77, 113, 139 and 184.)

S. Röser, E. Schilbach, and B. Goldman. Hyades tidal tails revealed by Gaia DR2. A&A, 621:L2, 2019. (Cited on pages 34, 42, 57, 58, 62, 64, 139 and 173.)

L. M. Sarro, H. Bouy, A. Berihuete, E. Bertin, E. Moraux, J. Bouvier, J. C. Cuillandre, D. Barrado, and E. Solano. Cluster membership probabilities from proper motions and multi-wavelength photometric catalogues. I. Method and application to the Pleiades cluster. A&A, 563:A45, 2014. (Cited on pages 34 and 173.)

P. Sartoretti, D. Katz, M. Cropper, P. Panuzzo, G. M. Seabroke, Y. Viala, K. Benson, R. Blomme, et al. Gaia Data Release 2. Processing the spectroscopic data. A&A, 616: A6, 2018. (Cited on pages 85 and 87.)

E. Schilbach, N. V. Kharchenko, A. E. Piskunov, S. Röser, and R. D. Scholz. Population analysis of open clusters: radii and mass segregation. A&A, 456(2):523–534, 2006. (Cited on page 71.)

R. Schönrich, J. Binney, and W. Dehnen. Local kinematics and the local standard of rest. MNRAS, 403(4):1829–1833, 2010. (Cited on page 92.)

J. E. Scott, E. D. Friel, and K. A. Janes. Kinematics of the Old Open Cluster System. AJ, 109:1706, 1995. (Cited on pages 33, 171 and 172.)

S. Sharma, M. R. Hayden, J. Bland-Hawthorn, D. Stello, S. Buder, J. C. Zinn, T. Kallinger, M. Asplund, et al. Fundamental relations for the velocity dispersion of stars in the Milky Way. MNRAS, 506(2):1761–1776, 2021. (Cited on pages 102 and 103.)

F. H. Shu, V. Milione, W. Gebel, C. Yuan, D. W. Goldsmith, and W. W. Roberts. Galactic Shocks in an Interstellar Medium with Two Stable Phases. ApJ, 173:557, 1972. (Cited on pages 29 and 167.)

G. Sim, S. H. Lee, H. B. Ann, and S. Kim. 207 New Open Star Clusters within 1 kpc from Gaia Data Release 2. *Journal of Korean Astronomical Society*, 52:145–158, 2019. (Cited on pages 37 and 176.)

M. F. Skrutskie, S. E. Schneider, R. Stiening, S. E. Strom, M. D. Weinberg, C. Beichman, T. Chester, R. Cutri, et al. The Two Micron All Sky Survey (2MASS): Overview and Status. In F. Garzon, N. Epchtein, A. Omont, B. Burton, and P. Persi, editors, *The Impact of Large Scale Near-IR Sky Surveys*, volume 210, page 25, 1997. (Cited on pages 29, 42 and 167.)

Y. Sofue. Rotation Curve of the Milky Way and the Dark Matter Density. *Galaxies review*, 8:37, 2020. (Cited on pages 95 and 97.)

Y. Sofue, M. Honma, and T. Omodaka. Unified Rotation Curve of the Galaxy – Decomposition into de Vaucouleurs Bulge, Disk, Dark Halo, and the 9-kpc Rotation Dip –. PASJ, 61:227, 2009. (Cited on page 96.)

C. Soubiran, O. Bienaymé, T. V. Mishenina, and V. V. Kovtyukh. Vertical distribution of Galactic disk stars. IV. AMR and AVR from clump giants. A&A, 480(1):91–101, 2008. (Cited on page 101.)

C. Soubiran, G. Jasniewicz, L. Chemin, F. Crifo, S. Udry, D. Hestroffer, and D. Katz. The catalogue of radial velocity standard stars for Gaia. I. Pre-launch release. A&A, 552:A64, 2013. (Cited on page 88.)

C. Soubiran, T. Cantat-Gaudin, M. Romero-Gómez, L. Casamiquela, C. Jordi, A. Vallenari, T. Antoja, L. Balaguer-Núñez, et al. Open cluster kinematics with Gaia DR2. A&A, 619:A155, 2018a. (Cited on pages 33, 43, 85, 88, 90, 91, 94, 112, 139, 172 and 182.)

C. Soubiran, G. Jasniewicz, L. Chemin, C. Zurbach, N. Brouillet, P. Panuzzo, P. Sartoretti, D. Katz, et al. Gaia Data Release 2. The catalogue of radial velocity standard stars. A&A, 616:A7, 2018b. (Cited on pages 86 and 87.)

J. S. Speagle. A Conceptual Introduction to Markov Chain Monte Carlo Methods. *arXiv e-prints*, art. arXiv:1909.12313, 2019. (Cited on page 59.)

J. Spitzer, Lyman. Equipartition and the Formation of Compact Nuclei in Spherical Stellar Systems. ApJ, 158:L139, 1969. (Cited on pages 31 and 169.)

J. Spitzer, Lyman and M. Schwarzschild. The Possible Influence of Interstellar Clouds on Stellar Velocities. ApJ, 114:385, 1951. (Cited on pages 31, 104 and 170.)

M. Steinmetz, T. Zwitter, G. Matijevic, A. Siviero, and U. Munari. A Comparison between RAVE DR5 and Gaia DR2 Radial Velocities. *Research Notes of the American Astronomical Society*, 2(4):194, 2018. (Cited on page 87.)

M. Steinmetz, G. Guiglion, P. J. McMillan, G. Matijevič, H. Enke, G. Kordopatis, T. Zwitter, M. Valentini, et al. The Sixth Data Release of the Radial Velocity Experiment (RAVE). II. Stellar Atmospheric Parameters, Chemical Abundances, and Distances. AJ, 160(2):83, 2020. (Cited on pages 85, 86 and 87.)

K. A. Strand. Hertzsprung's Contributions to the HR Diagram. *IAU Symposium*, 80(80): 55, 1977. (Cited on pages 26 and 164.)

S.-Y. Tang, X. Pang, Z. Yuan, W. P. Chen, J. Hong, B. Goldman, A. Just, B. Shukirgaliyev, and C.-C. Lin. Discovery of Tidal Tails in Disrupting Open Clusters: Coma Berenices and a Neighbor Stellar Group. ApJ, 877(1):12, 2019. (Cited on pages 42 and 77.)

Y. Tarricq, C. Soubiran, L. Casamiquela, T. Cantat-Gaudin, L. Chemin, F. Anders, T. Antoja, M. Romero-Gómez, et al. 3D kinematics and age distribution of the open cluster population. A&A, 647:A19, 2021a. (Cited on pages 28, 38, 43, 84, 167 and 178.)

Y. Tarricq, C. Soubiran, L. Casamiquela, A. Castro-Ginard, J. Olivares, N. Miret-Roig, and P. A. B. Galli. Structural parameters of 389 local Open Clusters. *arXiv e-prints*, art. arXiv:2111.05291, 2021b. (Cited on pages 47 and 66.)

A. Toomre. On the gravitational stability of a disk of stars. ApJ, 139:1217–1238, 1964. (Cited on pages 39 and 178.)

W. H. Trick, J. Coronado, and H.-W. Rix. The Galactic disc in action space as seen by Gaia DR2. MNRAS, 484(3):3291–3306, 2019. (Cited on pages 106, 107 and 108.)

B. A. Twarog, K. M. Ashman, and B. J. Anthony-Twarog. Some Revised Observational Constraints on the Formation and Evolution of the Galactic Disk. AJ, 114:2556, 1997. (Cited on pages 34 and 172.)

G. van de Ven, R. C. E. van den Bosch, E. K. Verolme, and P. T. de Zeeuw. The dynamical distance and intrinsic structure of the globular cluster ω Centauri. A&A, 445(2):513–543, 2006. (Cited on page 67.)

R. A. Vázquez, J. May, G. Carraro, L. Bronfman, A. Moitinho, and G. Baume. Spiral Structure in the Outer Galactic Disk. I. The Third Galactic Quadrant. ApJ, 672(2): 930–939, 2008. (Cited on pages 31 and 170.)

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. (Cited on page 79.)

R. Wielen. The Diffusion of Stellar Orbits Derived from the Observed Age-Dependence of the Velocity Dispersion. A&A, 60(2):263–275, 1977. (Cited on pages 33, 101 and 172.)

C. C. Worley, P. de Laverny, A. Recio-Blanco, V. Hill, A. Bijaoui, and C. Ordenovic. The AMBRE Project: Stellar parameterisation of the ESO:FEROS archived spectra. A&A, 542:A48, 2012. (Cited on pages 86 and 87.)

Z.-Y. Wu, X. Zhou, J. Ma, and C.-H. Du. The orbits of open clusters in the Galaxy. MNRAS, 399(4):2146–2164, 2009. (Cited on pages 33 and 172.)

M. S. Xiang, X. W. Liu, H. B. Yuan, Y. Huang, Z. Y. Huo, H. W. Zhang, B. Q. Chen, H. H. Zhang, et al. The LAMOST stellar parameter pipeline at Peking University - LSP3. MNRAS, 448(1):822–854, 2015. (Cited on page 85.)

F. C. Yeh, G. Carraro, M. Montalto, and A. F. Seleznev. Ruprecht 147: A Paradigm of Dissolving Star Cluster. AJ, 157(3):115, 2019. (Cited on pages 10, 42, 44, 53, 77, 113 and 184.)

J. Yu and C. Liu. The age-velocity dispersion relation of the Galactic discs from LAMOST-Gaia data. MNRAS, 475(1):1093–1103, 2018. (Cited on pages 101 and 103.)

Z. Yuan, J. Chang, P. Banerjee, J. Han, X. Kang, and M. C. Smith. StarGO: A New Method to Identify the Galactic Origins of Halo Stars. ApJ, 863(1):26, 2018. (Cited on page 42.)

N. Zacharias, C. T. Finch, T. M. Girard, A. Henden, J. L. Bartlett, D. G. Monet, and M. I. Zacharias. The Fourth US Naval Observatory CCD Astrograph Catalog (UCAC4). AJ, 145(2):44, 2013. (Cited on pages 34, 42 and 173.)

M. Zhai, H. Abt, G. Zhao, and C. Li. The Shape Evolution of Galactic Open Clusters from Observations Under Galactic External Forces. AJ, 153(2):57, 2017. (Cited on page 76.)

Y. Zhang, S.-Y. Tang, W. P. Chen, X. Pang, and J. Z. Liu. Diagnosing the Stellar Population and Tidal Structure of the Blanco 1 Star Cluster. ApJ, 889(2):99, 2020. (Cited on pages 42, 49 and 77.)

Appendices

Appendix A

Selection of clusters members for chemical tagging

Any chemical characterization of an OC starts by the selection of the best suited spectroscopic targets among the members of a cluster. We selected such targets for several clusters during this thesis. We first focused on three nearby OCs: the Hyades, NGC 2632/Praesepe and Ruprecht 147. To select the cluster members to be targetted for spectroscopic follow-up, we used the memberships lists provided by Gaia Collaboration et al. (2018a) and Cantat-Gaudin et al. (2018a). We complemented this list of targets by the members identified by Röser et al. (2019) and Meingast and Alves (2019) as part of the tidal tails of the Hyades, by the tidal tails stars of NGC 2632 identified by Röser and Schilbach (2019) and by the list of candidates identified for Ruprecht 147 in Sect. 2.2.2 (see Fig. 2.1). We selected red clump giants, sub-giants and main sequence stars which fall in the range of spectral types F5-G0, for which our method yields the most precise abundances. These stellar types were selected making use of PARSEC isochrones (Bressan et al., 2012) based on their Gaia photometry. For NGC 2632 and Ruprecht 147, we additionally used RVs from *Gaia* to reject some of the stars with discrepant values with respect to the cluster mean RV computed by Soubiran et al. (2018a).

These selection criteria gave us a list of 467 targets to be studied spectroscopically. As these clusters have been massively studied in the past, we looked for high resolution spectra $\lambda/\Delta\lambda \geq 45\,000$ of these stars in the public archives. When low Signal-to-Noise ratio (S/N) spectra of the same stars were retrieved, their spectra were co-added to reach S/N of ~ 50. We looked for spectra in the archives of several instruments: UVES, FEROS, HARPS, HARPS-N, FIES, ESPaDOnS, NARVAL and ELODIE. This allowed us to collect 848 spectra for 108 different stars: 62 from the Hyades, 22 from NGC 2632 and 24 from Ruprecht 147. The CMD of the three clusters with the stars for which we retrieved spectra is showed in Fig. A.1.

No spectra from the stars belonging to the structure apparented to Ruprecht 147's tidal tail were obtained. However, eight stars from the tidal tails of the Hyades and two from the tidal tails of NGC 2632 were analysed. The differential chemical abundances of these stars compared to the chemical signatures of the clusters showed that three stars from the Hyades and one from NGC 2632 were likely contaminants. The analysis of the spectra is explained in details in Casamiquela et al. (2020).

In the same spirit, we selected stars from the recently discovered cluster UBC 274 which exhibit a remarkable tidal tail (Castro-Ginard et al., 2020). We applied for three nights of observation with MIKE@Magellan telescope (Las Campanas Observatory) and got awarded for two nights. The observations have been carried out in January 2020.

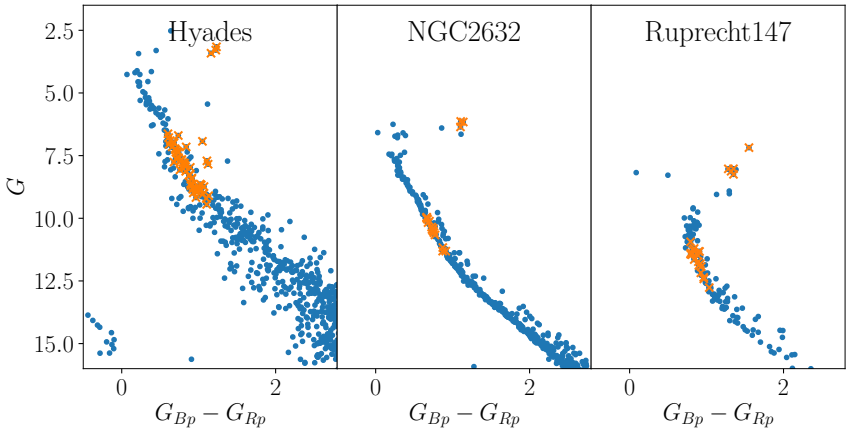

Figure A.1: From left to right: CMD of the Hyades, NGC 2632 and Ruprecht 147. We show in blue the initial list of members and in orange the stars for which we retrieved spectra. From Casamiquela et al. (2020).

The analysis of the observations is in progress and will be published soon (Casamiquela et al. 2021 , in prep.).

Appendix B

Additional figures

In this Appendix, we present additional figures showing the members selected by our clustering procedure described in Sect. 2.3. We show the results for 20 clusters for which we obtain remarkable results. We show the projection of the members in all the relevant planes. For the sake of comparison, we also represented both field stars and members recovered in CG+20.

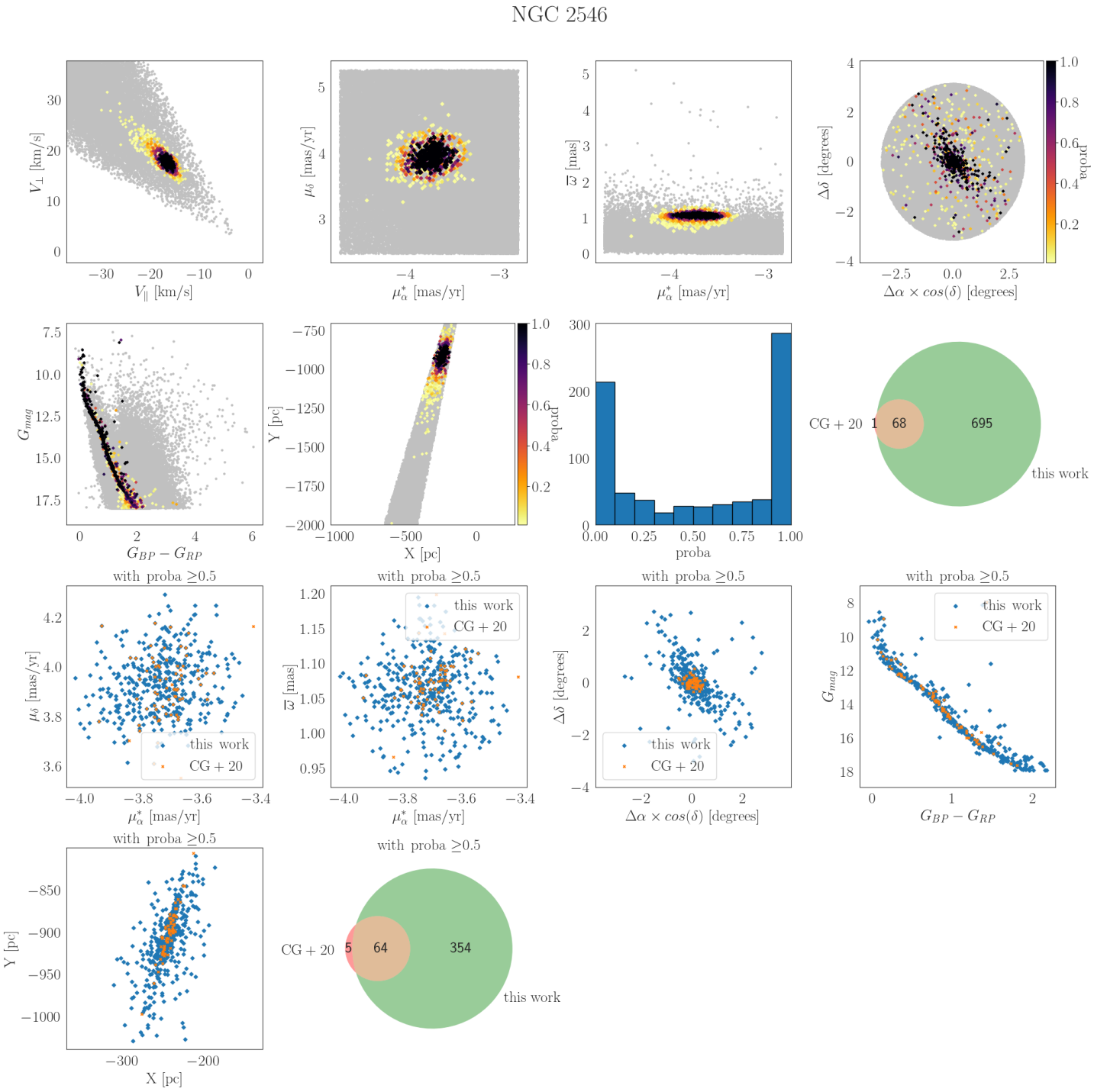

Figure B.1: Members of NGC 2546 recovered with HDBSCAN shown in different planes. On the two first rows, we show the members, colored according to their membership probability, overplotted on the field stars in the perpendicular-tangential velocity space, in the proper motion space, in the space of parallaxes versus μ_{α^*}, in the equatorial coordinate space, in the CMD and in the X-Y space. We then show the histogram of the membership probabilities of the recovered members and a Venn diagram comparing all our members with the ones from CG+20. On the third and fourth row, we show a comparison of our members having a membership probability higher than 0.5 with the one from CG+20 in the proper motion space, in the space of parallaxes versus μ_{α^*}, in the equatorial coordinate space, in the CMD and in the X-Y space. We finally show the Venn diagram comparing our members with proba≥0.5 with the ones from CG+20

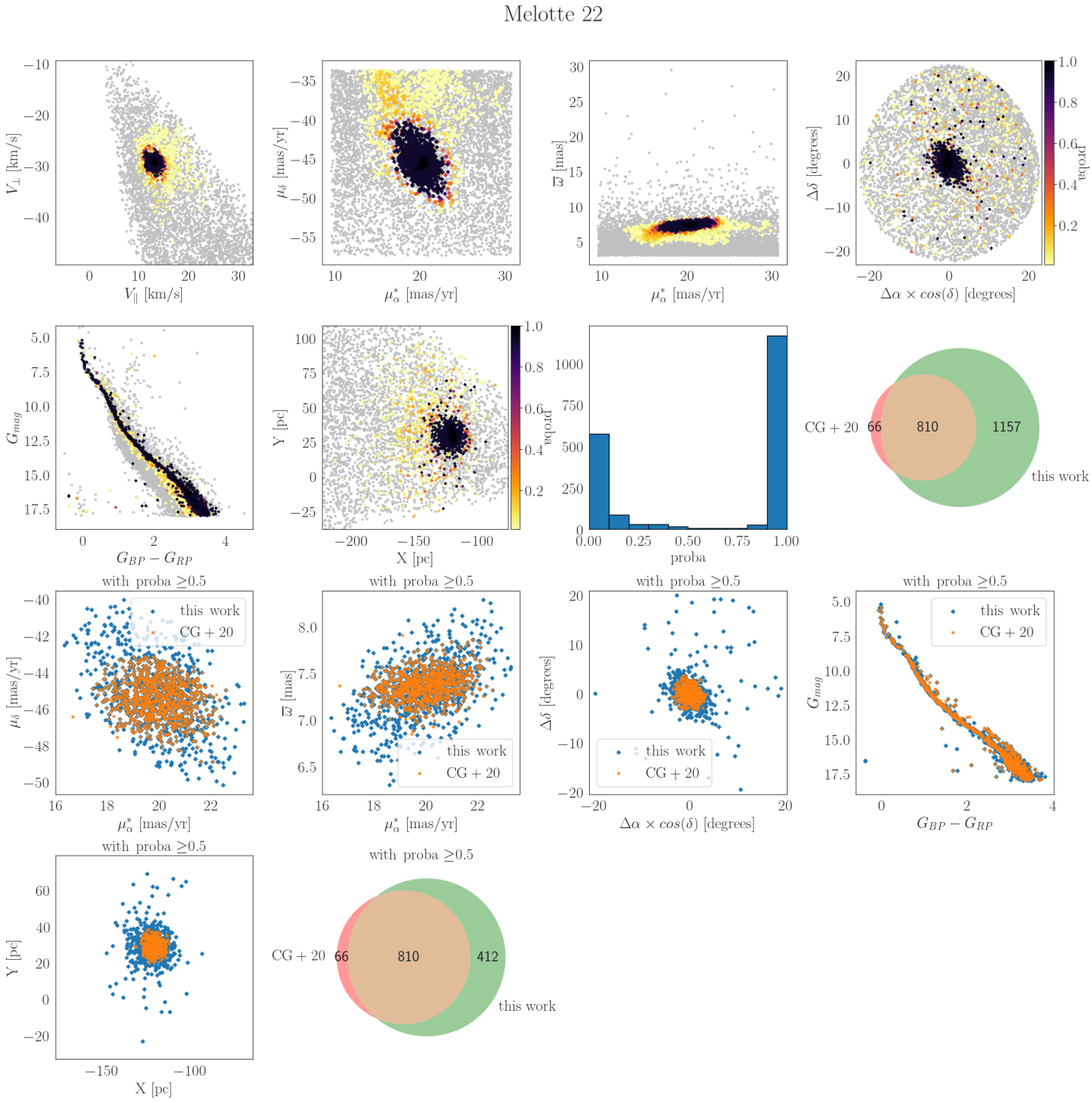

Figure B.2: Same as B.1 but with the cluster Melotte 22.

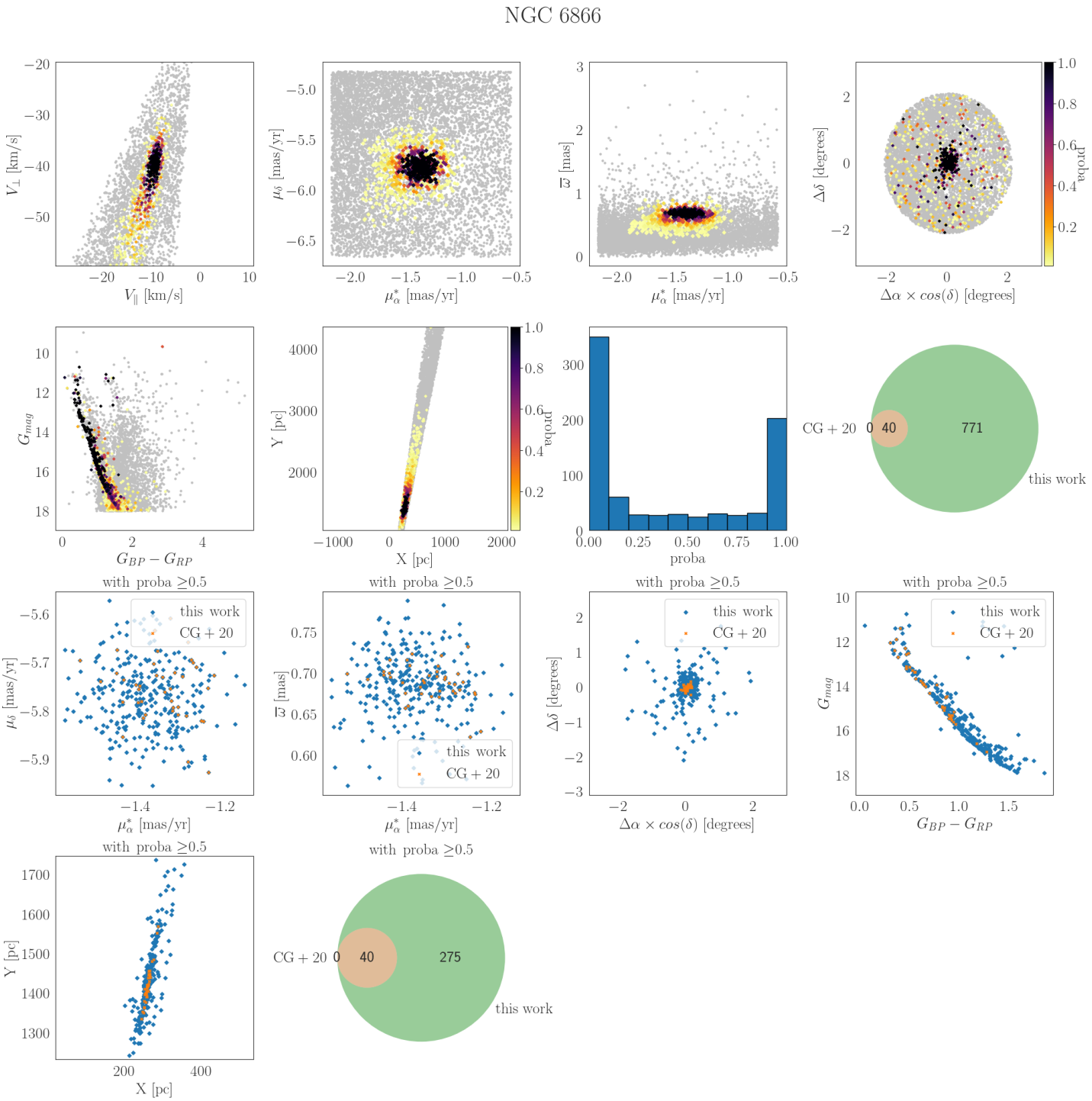

Figure B.3: Same as B.1 but with the cluster NGC 6866.

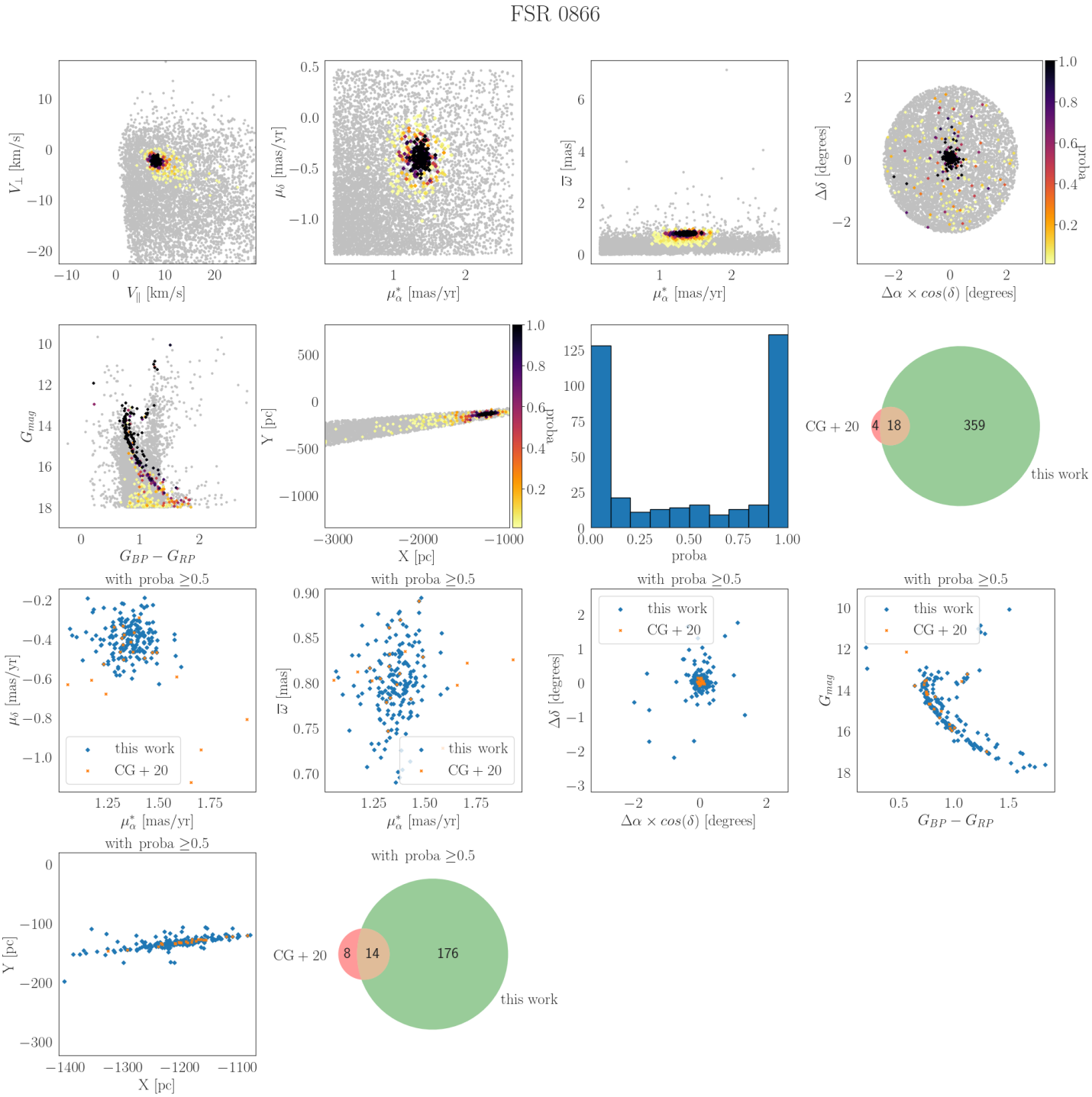

Figure B.4: Same as B.1 but with the cluster FSR 0866.

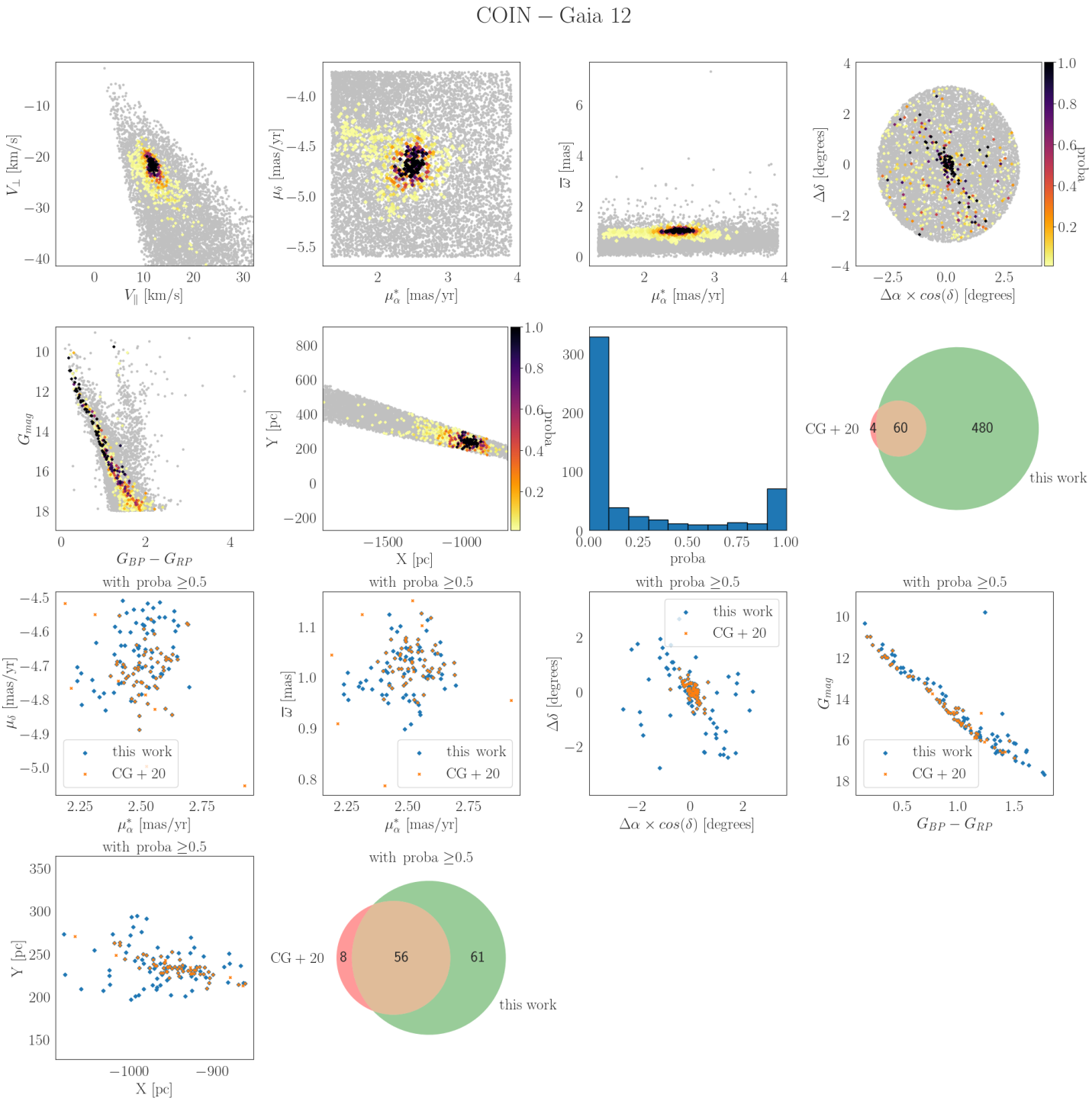

Figure B.5: Same as B.1 but with the cluster COIN-Gaia 12.

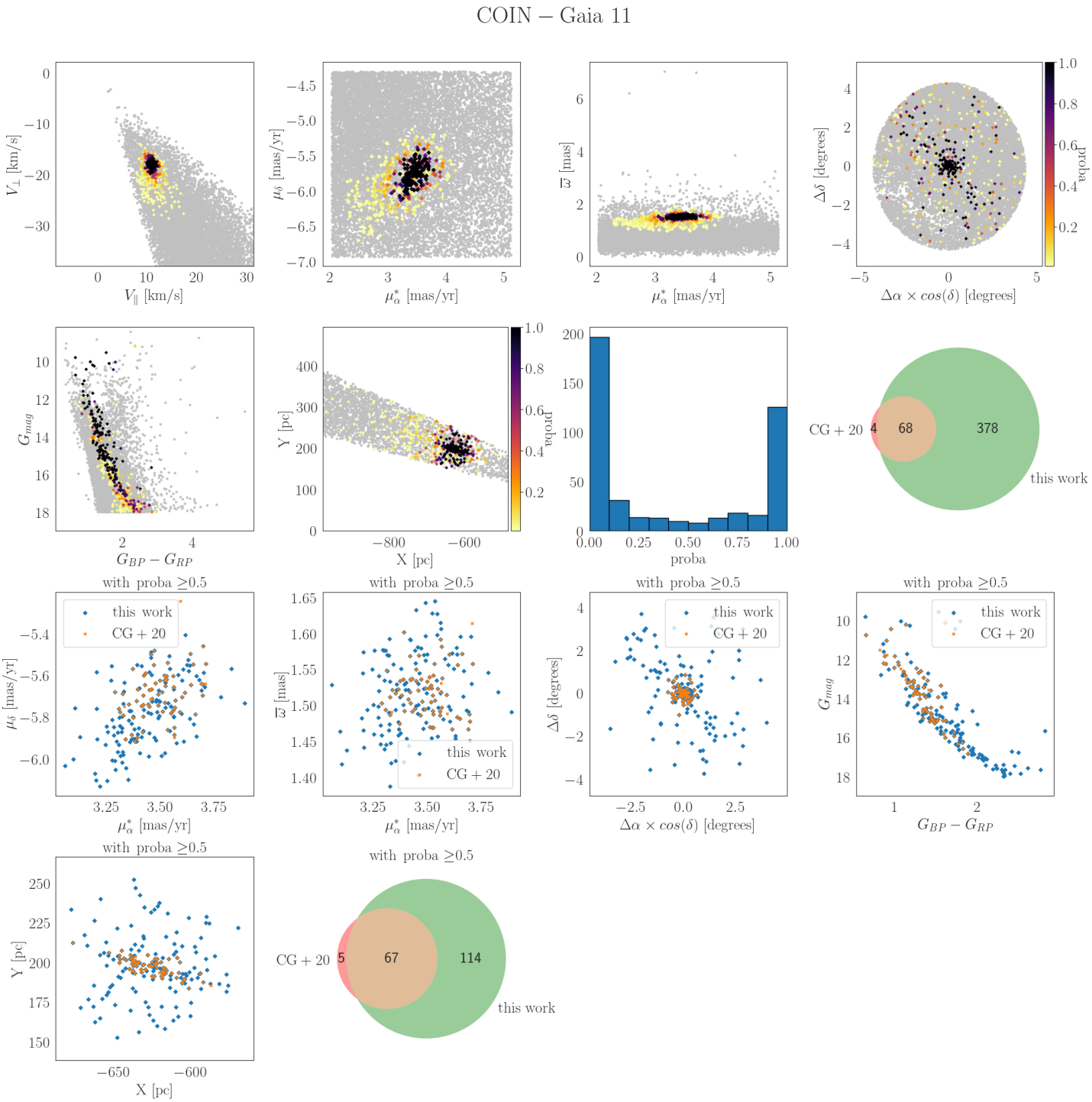

Figure B.6: Same as B.1 but with the cluster COIN-Gaia 11.

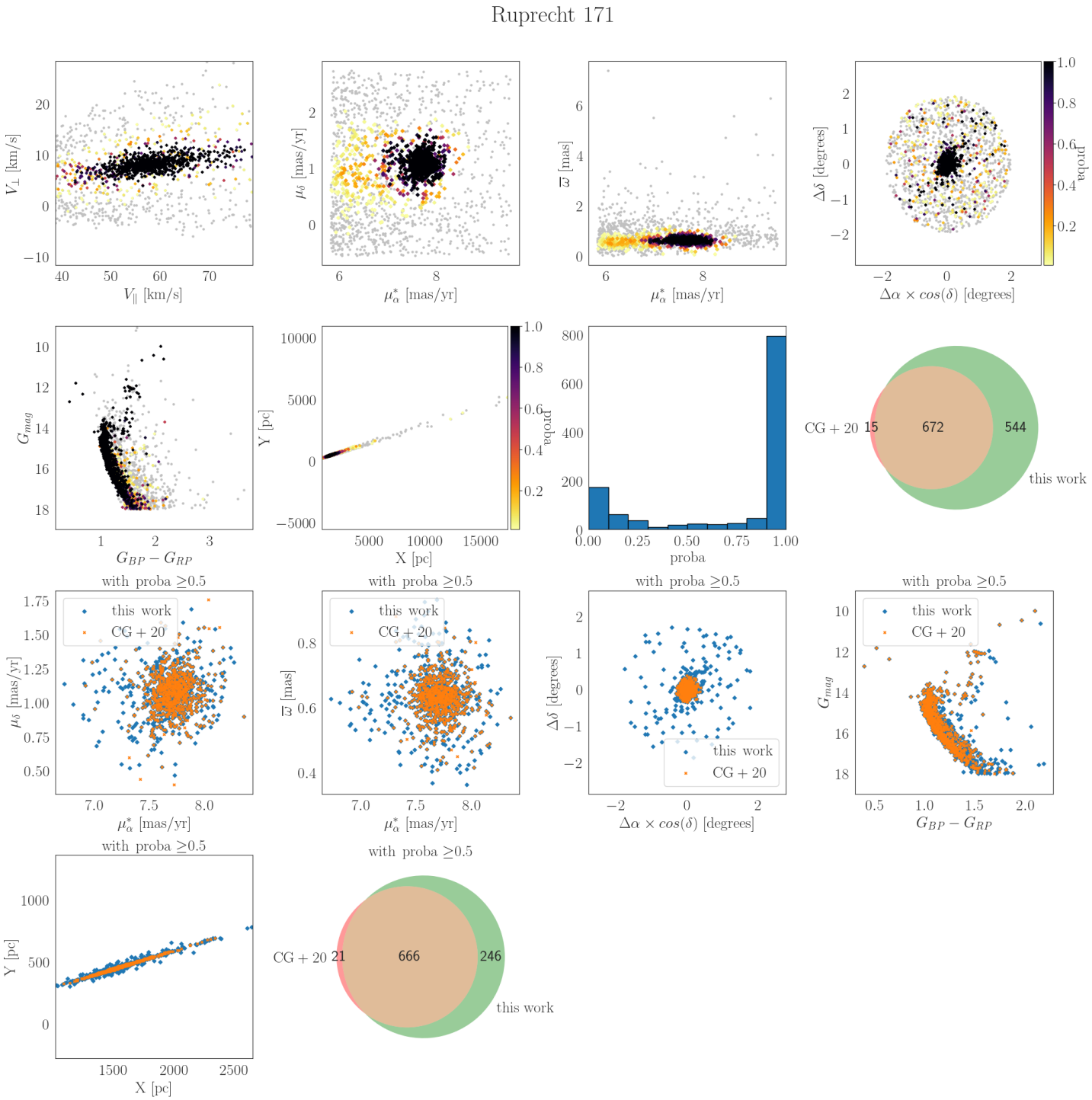

Figure B.7: Same as B.1 but with the cluster Ruprecht 171

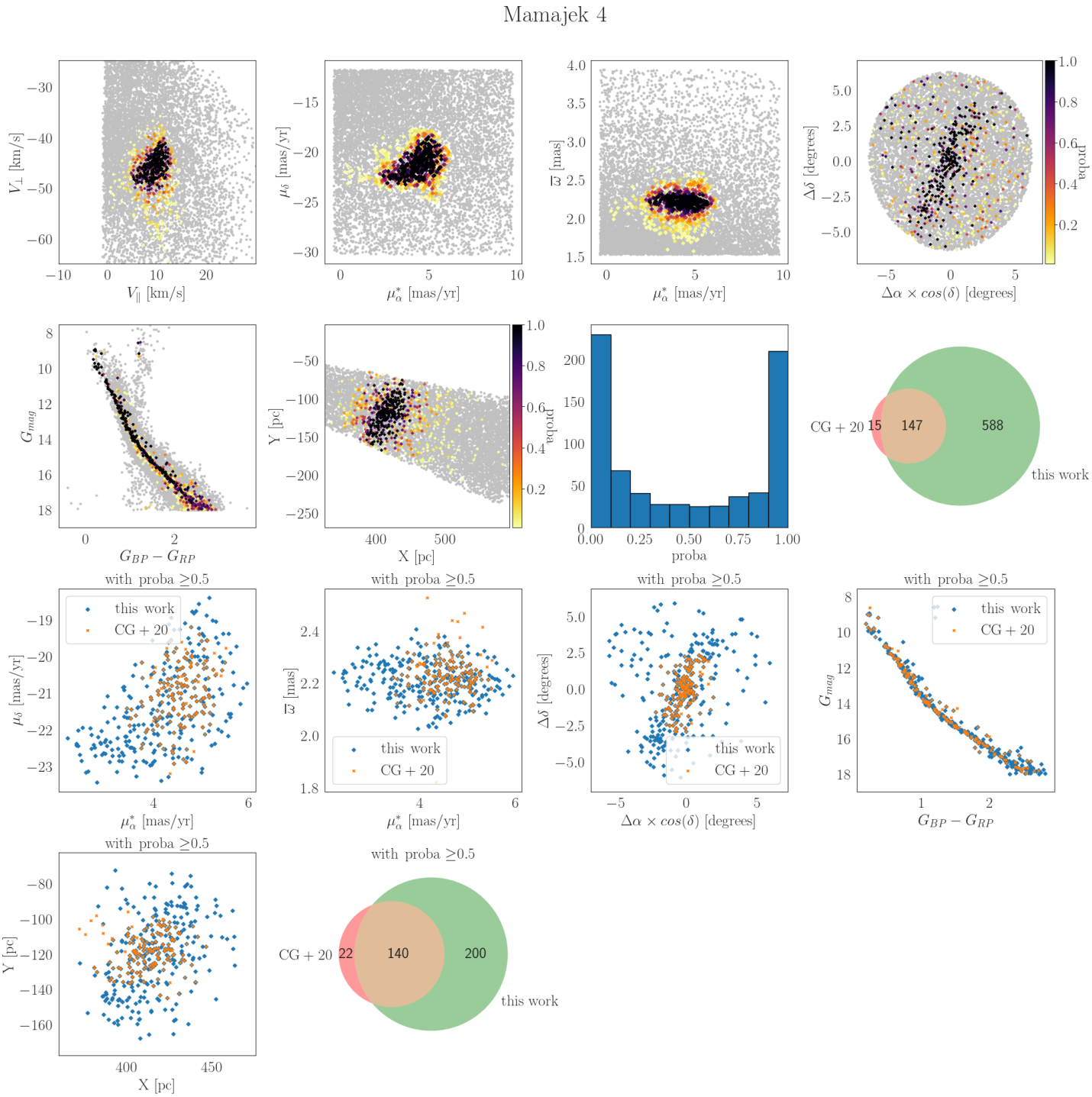

Figure B.8: Same as B.1 but with the cluster Mamajek 4.

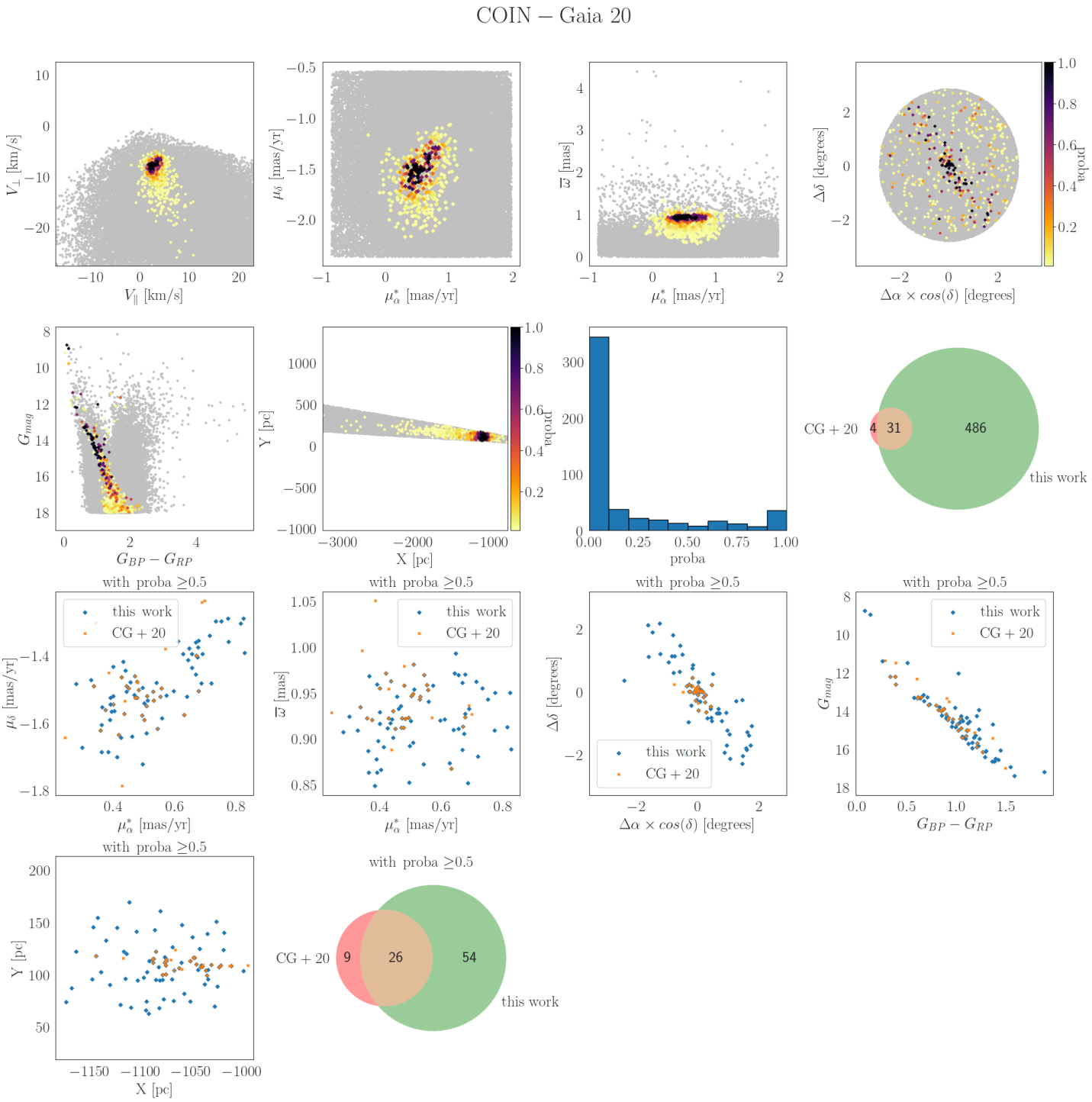

Figure B.9: Same as B.1 but with the cluster COIN-Gaia 20.

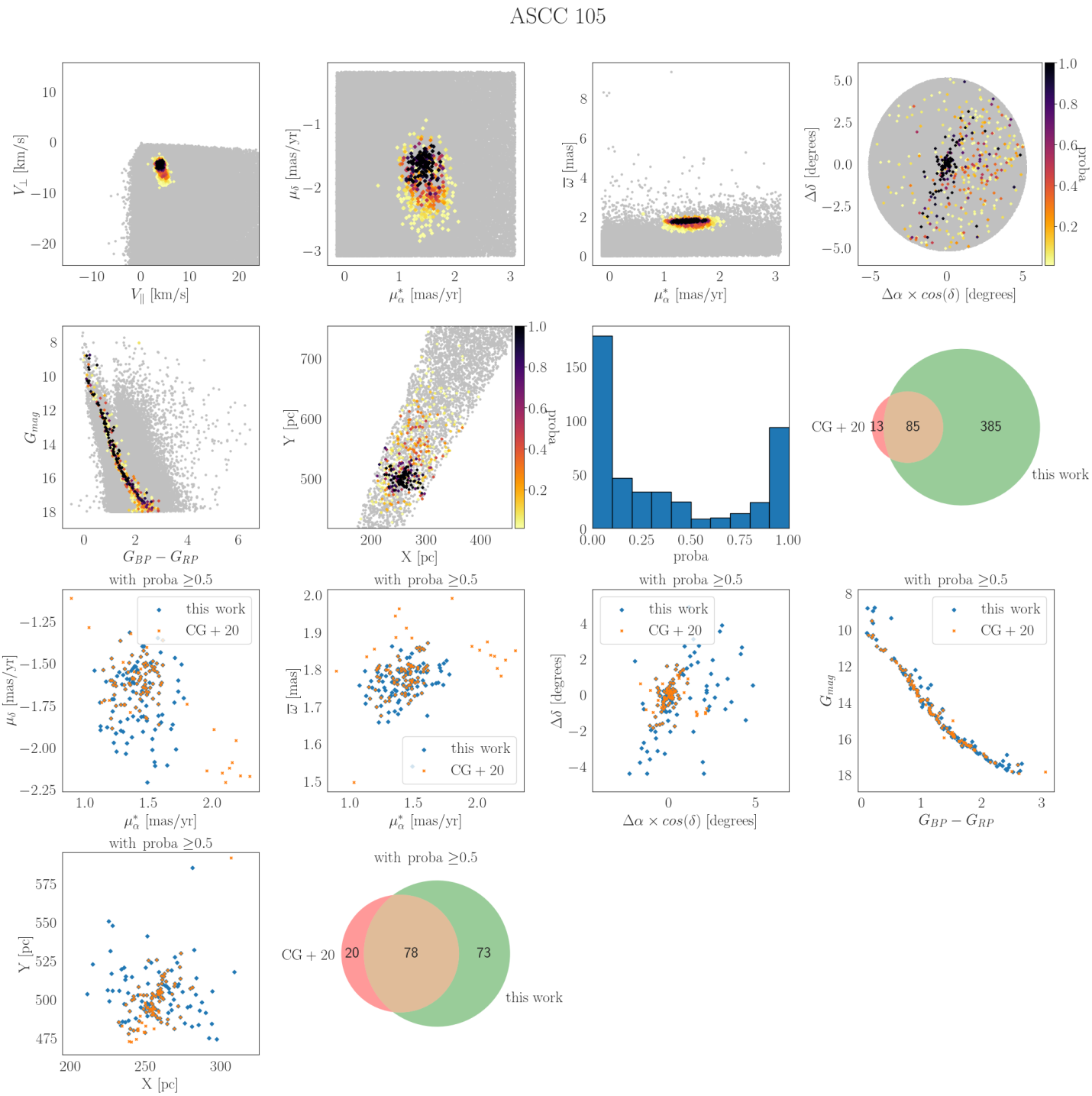

Figure B.10: Same as B.1 but with the cluster ASCC 105

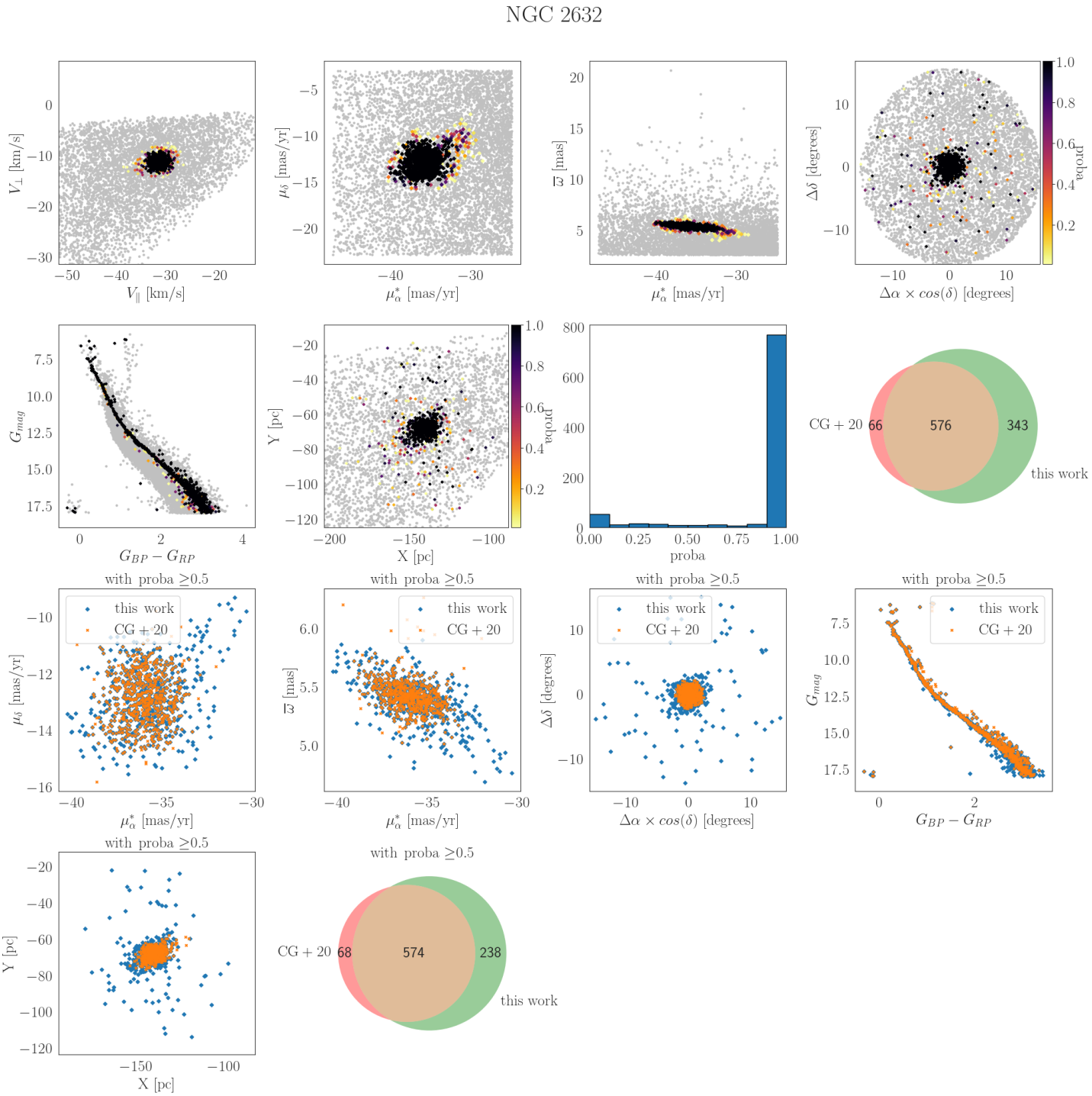

Figure B.11: Same as B.1 but with the cluster NGC 2632.

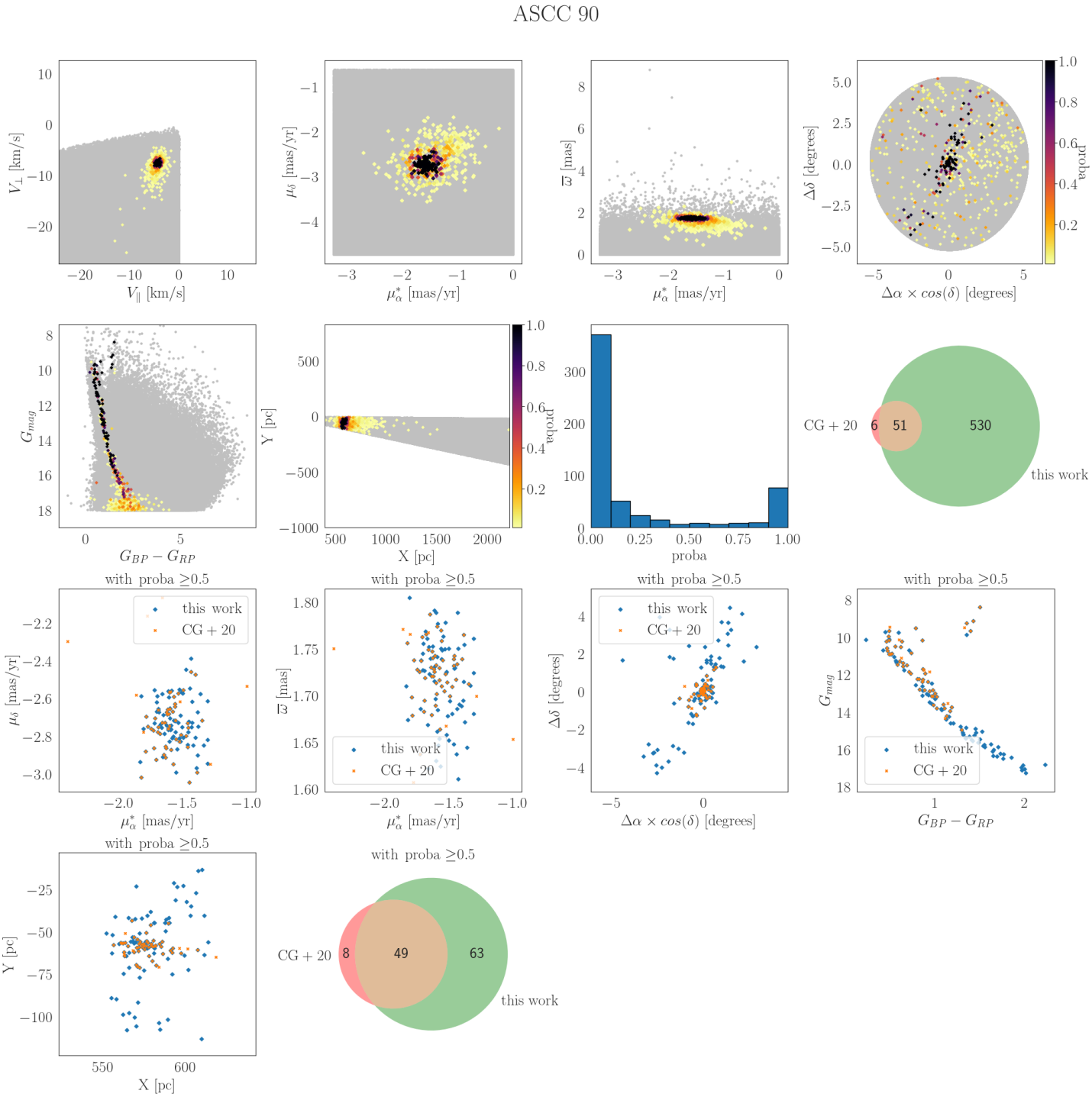

Figure B.12: Same as B.1 but with the cluster ASCC 90

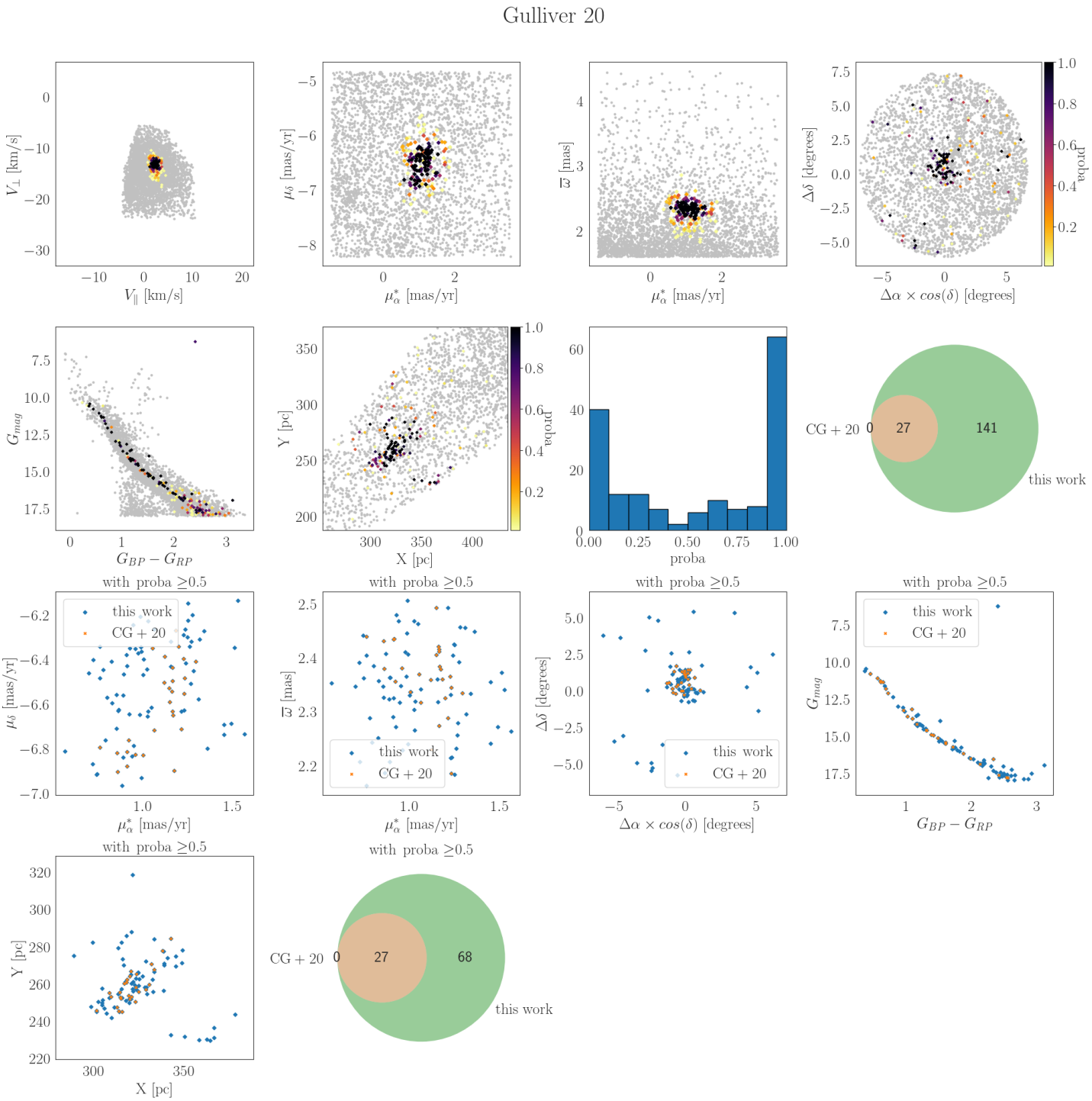

Figure B.13: Same as B.1 but with the cluster Gulliver 20.

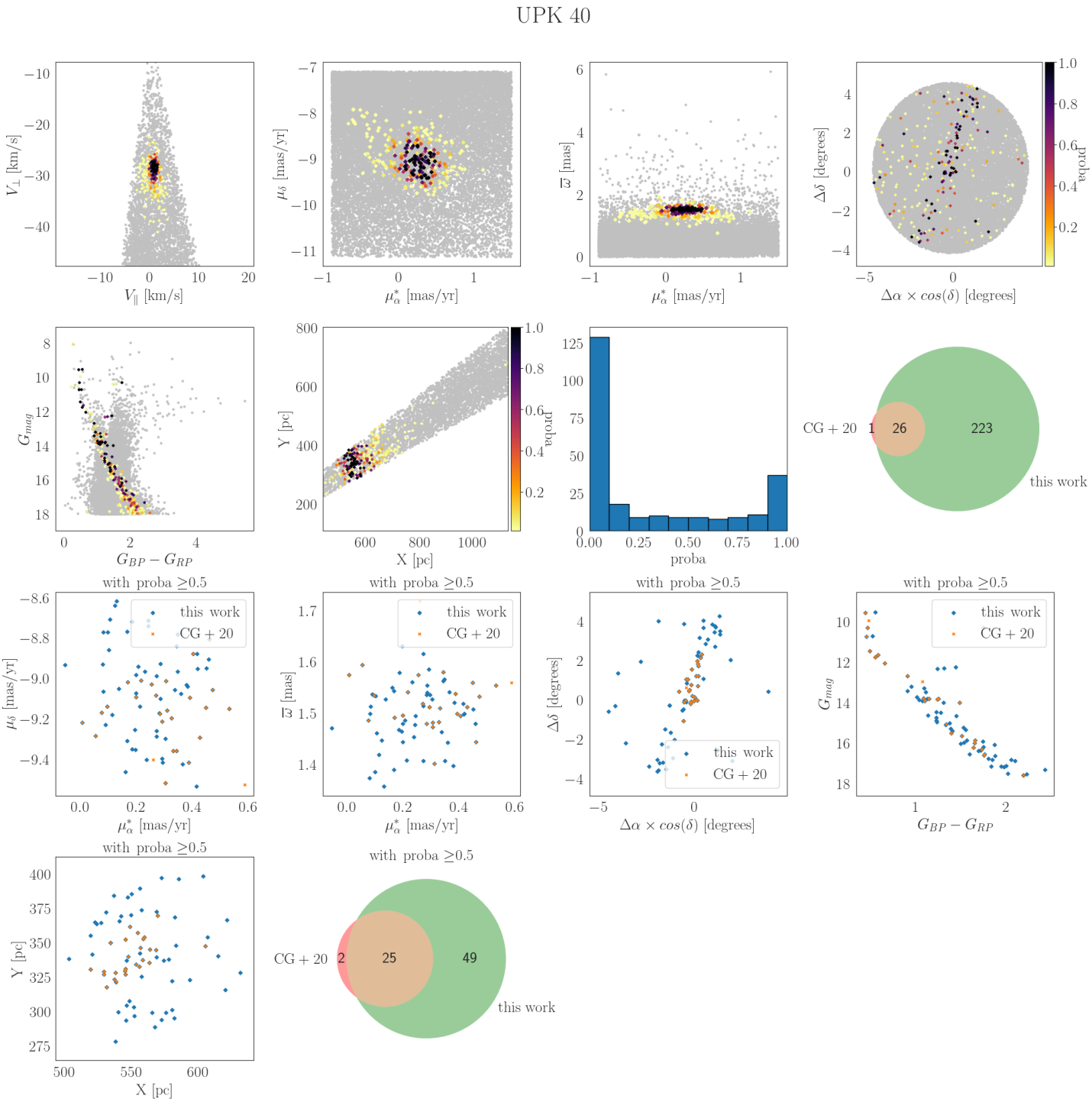

Figure B.14: Same as B.1 but with the cluster UPK 40.

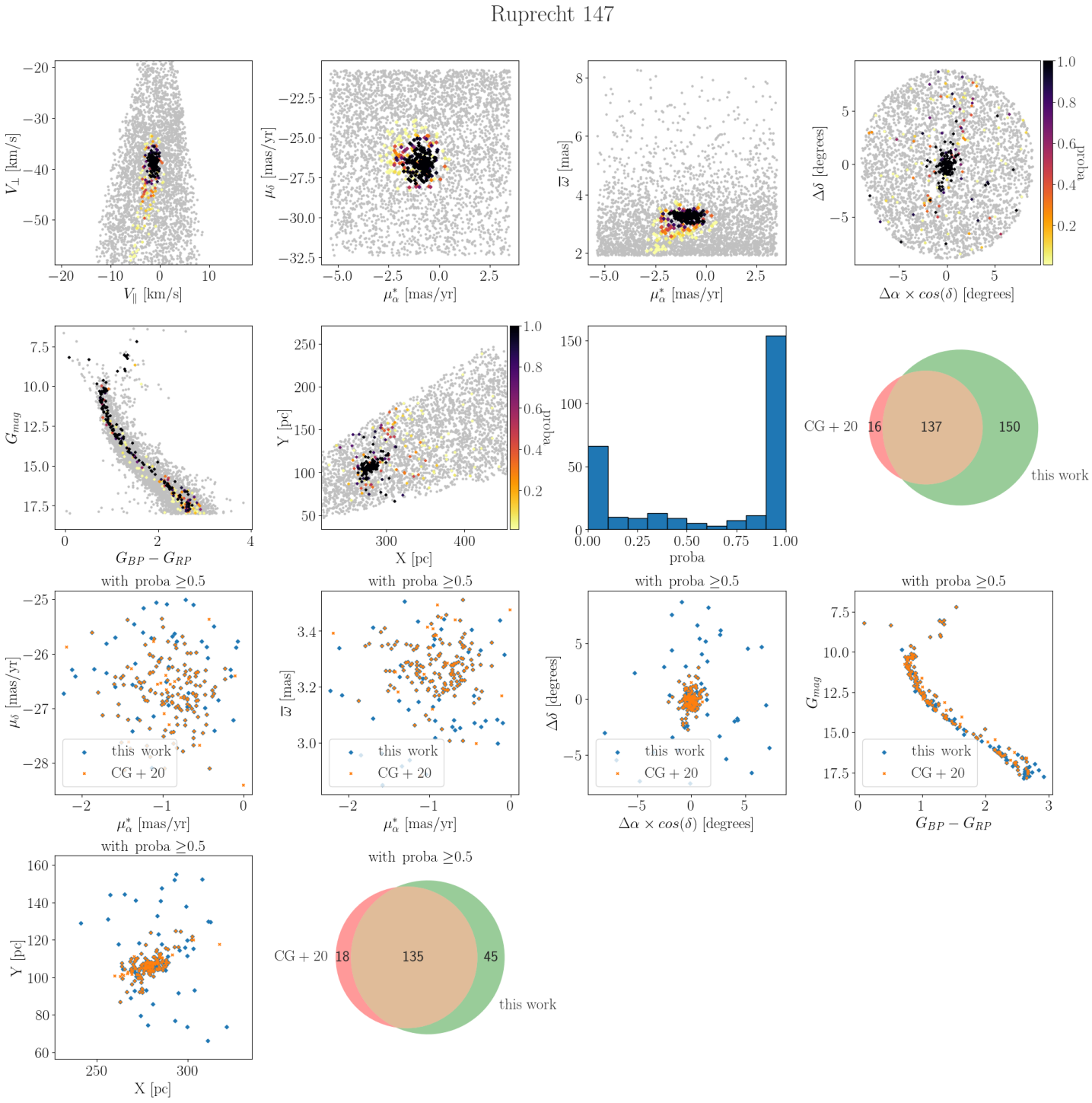

Figure B.15: Same as B.1 but with the cluster Ruprecht 147.

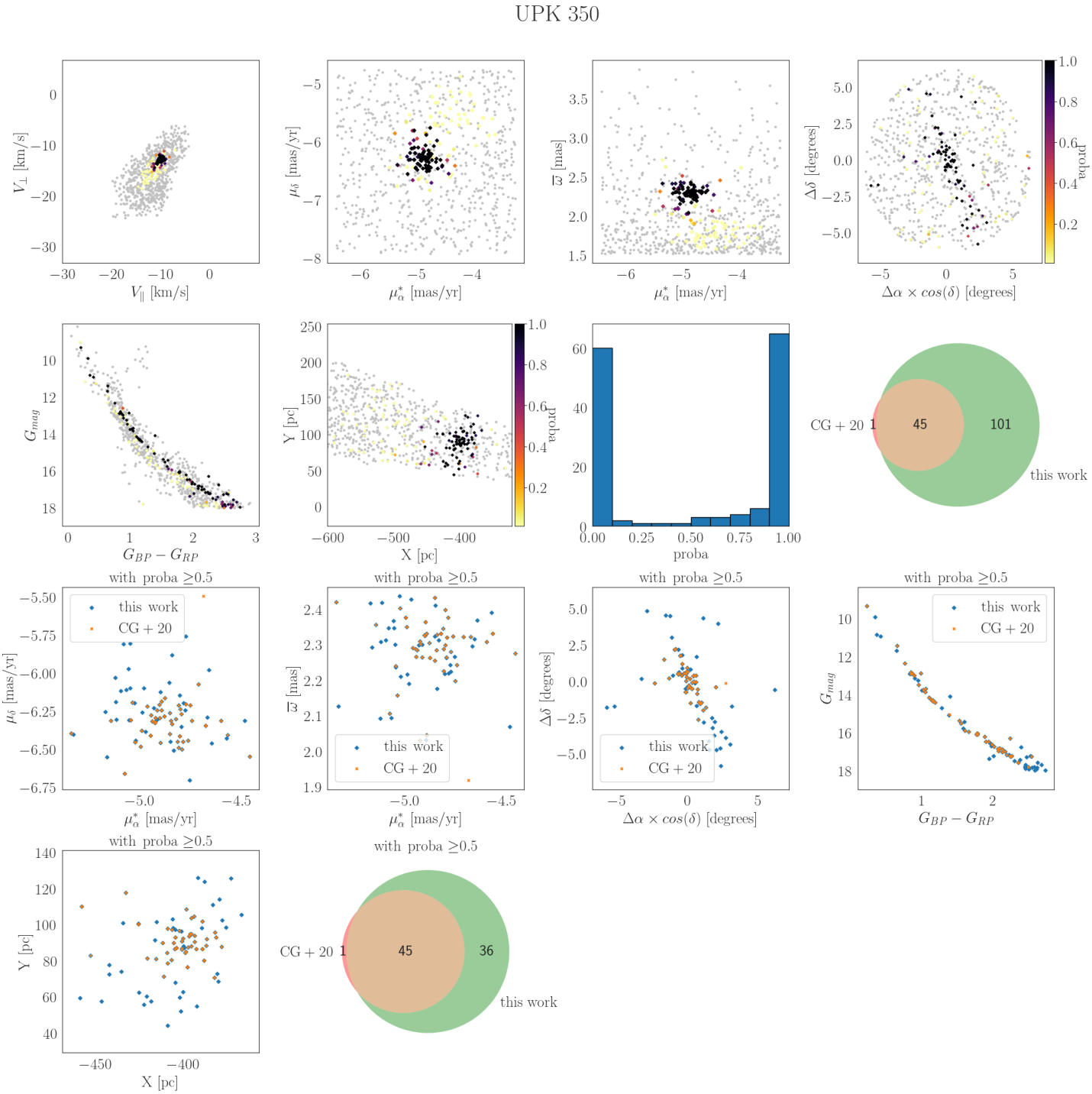

Figure B.16: Same as B.1 but with the cluster UPK 350.

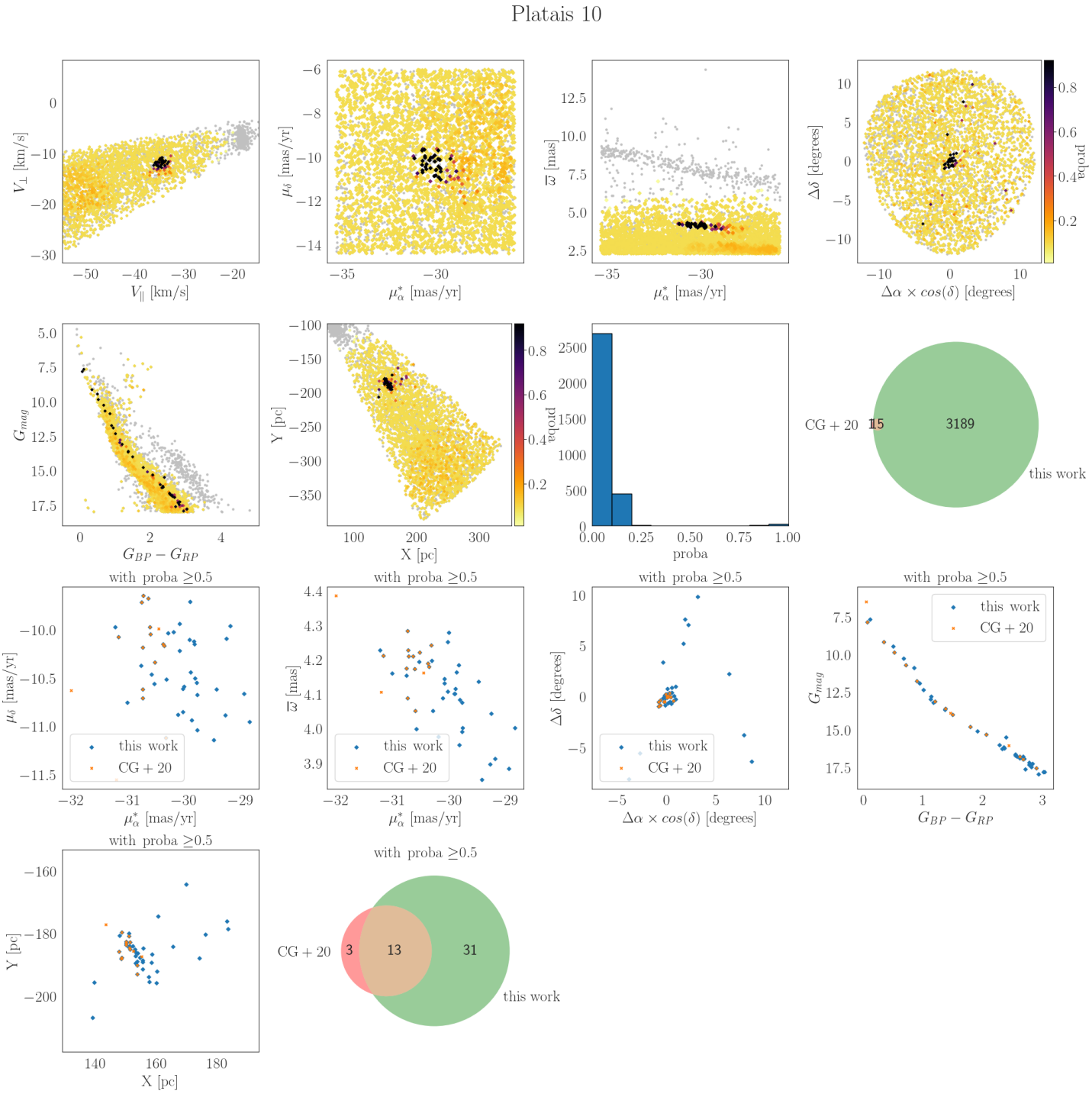

Figure B.17: Same as B.1 but with the cluster Platais 10.

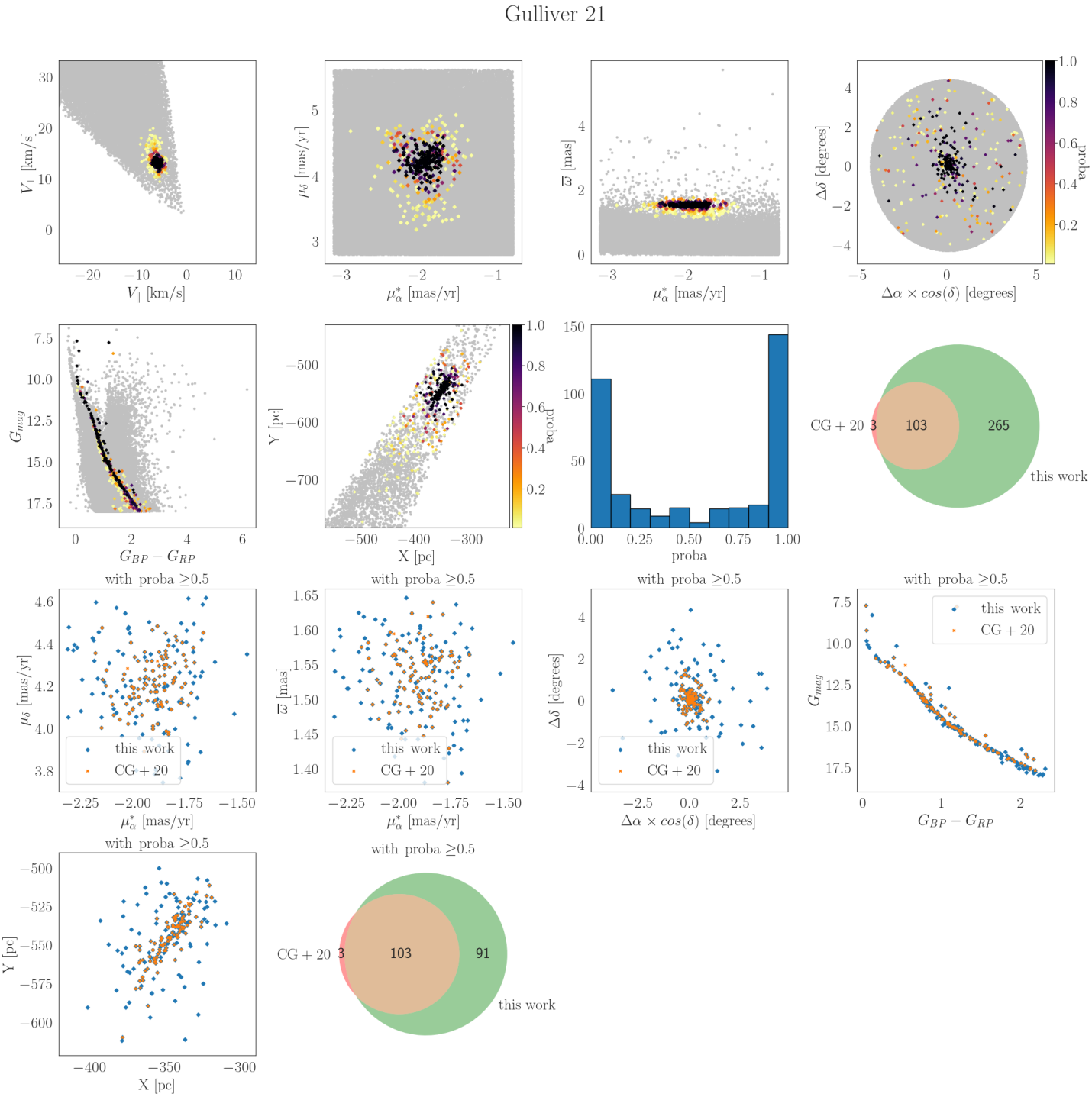

Figure B.18: Same as B.1 but with the cluster Gulliver 21.

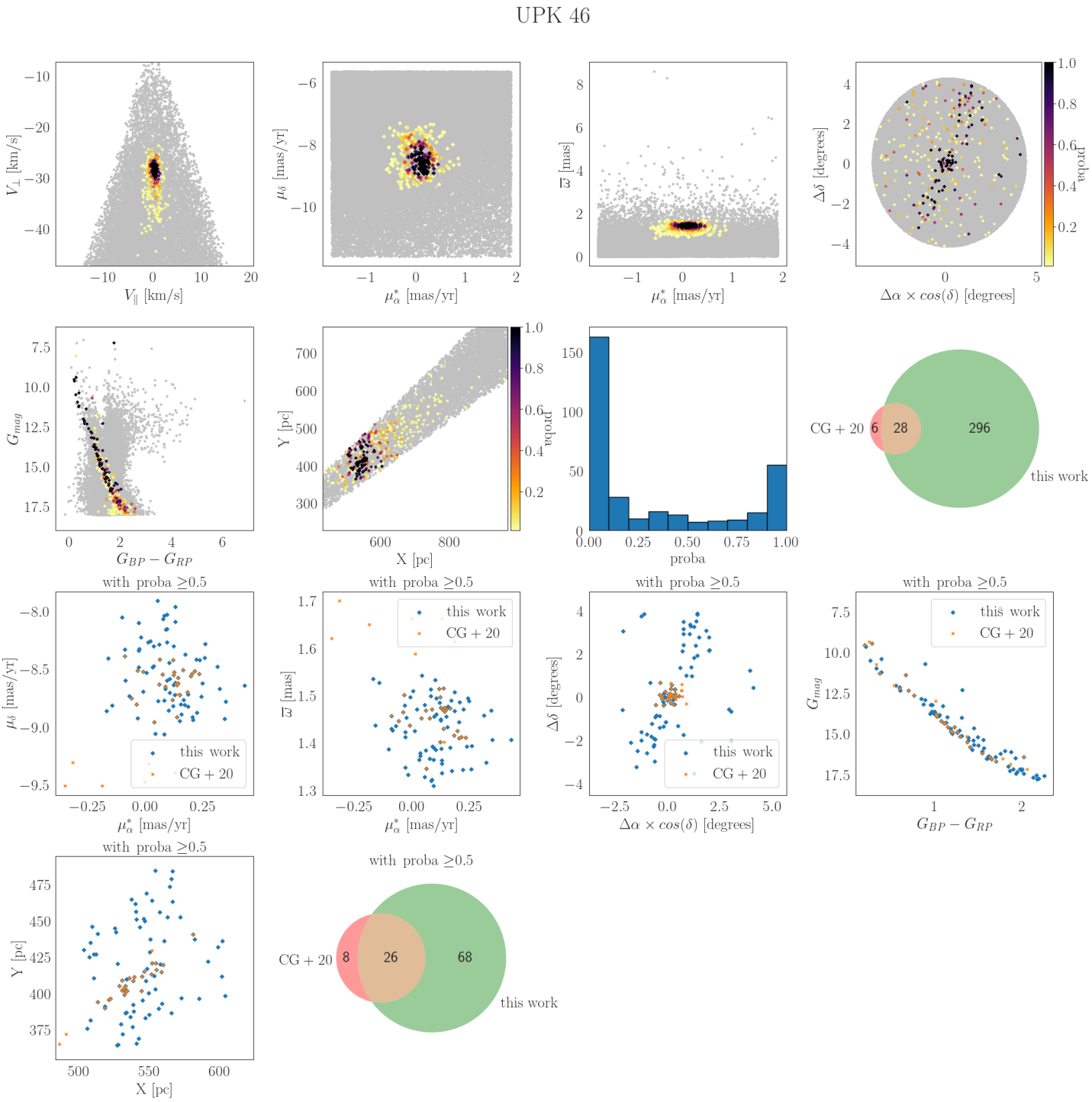

Figure B.19: Same as B.1 but with the cluster UPK 46.

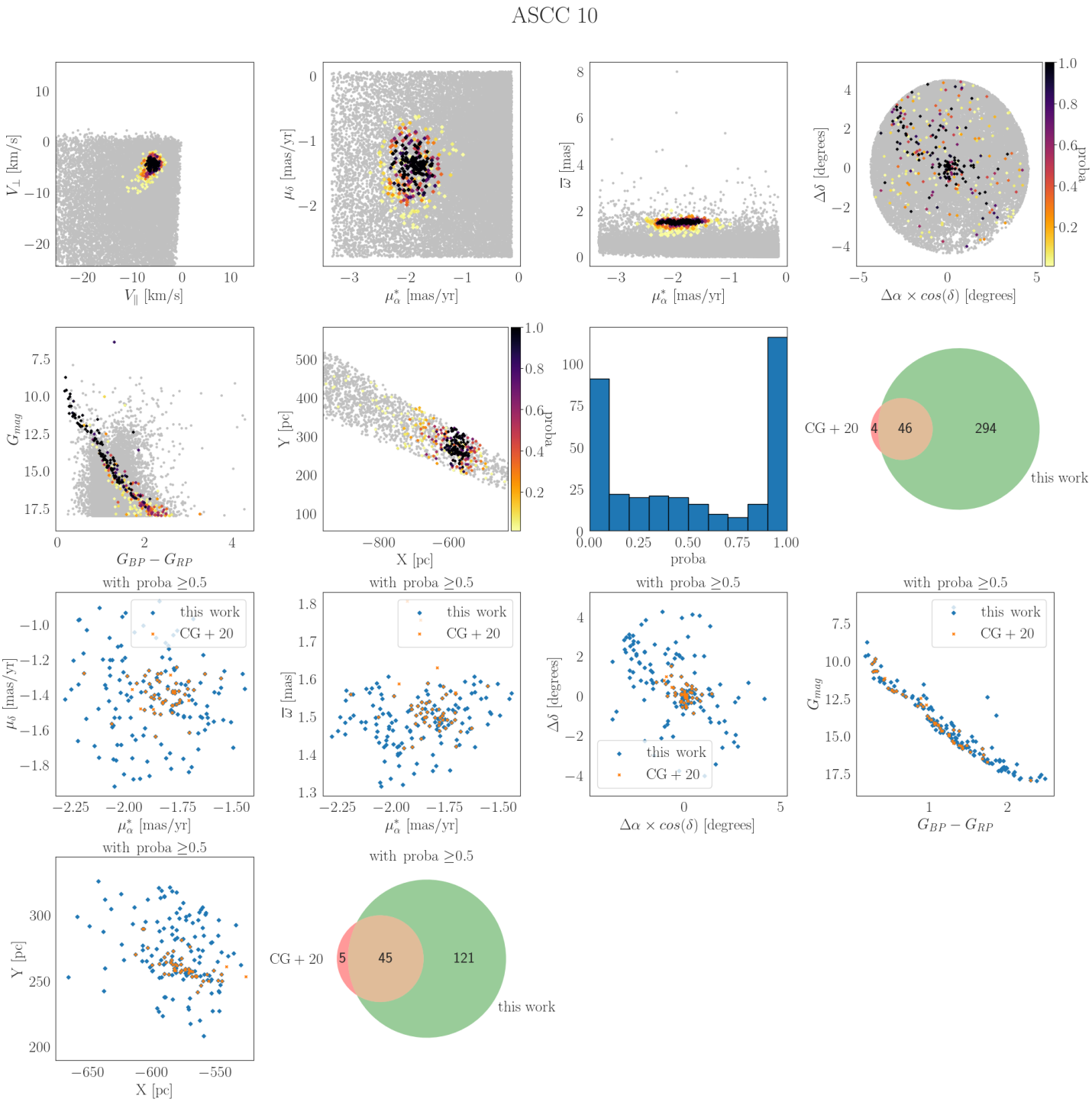

Figure B.20: Same as B.1 but with the cluster ASCC 10.

Appendix C

Introduction (en français)

Contents

C.1 Contexte historique

Depuis l'Antiquité, les amas d'étoiles sont scrutés par les astronomes. L'apparition des Pléiades dans le ciel nocturne était attendue chaque année car son apparition marquait le début des moissons dans l'hémisphère nord. Quelques siècles plus tard, l'astronome romain Ptolémée mentionna dans ses travaux plusieurs amas tels que Praesepe, le Double Amas ou NGC 6475 (également connu sous le nom d'amas de Ptolémée, voir Fig. C.1) et les classifia comme des nébuleuses (Moore and Rees, 2011). L'invention du télescope réfracteur au début du 17$^{\text{ème}}$ siècle permis à ses premiers utilisateurs de résoudre ces nébuleuses et distinguer leurs membres. En particulier, Galilée orienta son télescope vers certaines des nébuleuses identifiées par Ptolémée et identifia 50 étoiles appartenant à l'amas de Praesepe (Maran and Marschall, 2009). À la fin du 18$^{\text{ème}}$ siècle, les astronomes William et Caroline Herschel entreprirent une étude systématique des objets ayant un aspect nébuleux : le *Catalogue of Nebulae and Clusters of Stars* (Hoskin, 1987). Ils s'aperçurent que nombre de ces objets étaient composés d'étoiles individuelles. Un siècle plus tard, des centaines d'amas d'étoiles furent identifiés et classés dans le *Nouveau Catalogue Général* et dans le *Catalogue Index* compilé par Dreyer (1888, 1895, 1910). L'avènement des télescopes révéla que les amas d'étoiles pouvaient être divisés en deux catégories : les amas dits globulaires et les amas ouverts. Alors que les premiers sont composés de milliers d'étoiles, sont de forme sphérique et peuvent être observés dans tout le ciel, les seconds sont plus faiblement peuplés, ont une plus grande variété de formes et ont été généralement identifiés près du plan galactique. Au début du 20$^{\text{ème}}$ siècle, Ejnar Hertzsprung dessina le premier diagramme couleur-magnitude d'AO pour les Pléiades et les Hyades et remarqua que ces deux amas avaient des populations stellaires différentes (Strand, 1977). Ceci a été interprété plus tard comme une différence d'âge entre ces deux amas.

C.2 Amas Ouverts

C.2.1 Définition

Les AO sont des groupes d'étoiles formés à partir du même événement d'effondrement gravitationnel d'un Nuage Moléculaire Géant. Leurs populations stellaires sont liées par la gravitation et comme elles se sont formées à partir du même nuage de gaz et de poussière, elles ont le même âge, la même distance, le même mouvement et la même composition chimique.

Cela dit, il n'existe pas de définition claire et universelle des AO. Dans leur étude sur les jeunes amas stellaires massifs, Portegies Zwart et al. (2010) a simplement défini un amas d'étoiles comme un groupe d'étoiles liées par la gravitation. En se concentrant sur les jeunes amas encore entourés de gaz, Lada and Lada (2003) ont proposé qu'un amas soit un ensemble d'étoiles physiquement liées ayant une densité volumique de masse stellaire suffisamment grande ($\geq 1,0 M_{\odot}\text{pc}^{-3}$) pour résister à la perturbation par les forces de marées et suffisamment peuplé pour survivre pendant au moins 100 millions

Figure C.1: Images de six amas, montrés à des fins descriptives. De gauche à droite et de haut en bas sont montrés respectivement les Pleiades, NGC 6475, NGC 6705, NGC 3766, NGC 2547 et NGC 2422. Crédit: Davide de Martin & ESO.

d'années. Comme ils se sont concentrés sur des amas d'étoiles de tous âges et de toutes masses, Krumholz et al. (2019) n'ont pas pu adopter la même définition. Les jeunes amas ont des distributions spatiales complexes, généralement constituées de plusieurs sous-structures. Afin de ne pas exclure de leur étude les jeunes amas tels que l'amas de la nébuleuse d'Orion, ils ont défini un amas comme un groupe d'au moins 12 étoiles ayant une densité moyenne au moins quelques fois supérieure à la densité du champ et beaucoup plus grande que la densité locale de matière noire. Même si ces définitions ne se recoupent pas parfaitement, elles permettent de distinguer les amas d'étoiles des associations. Ces dernières ne sont pas liées par la gravitation et ont une densité stellaire d'au plus $\sim$0.01 étoiles/pc^3, avec des étoiles réparties sur des zones pouvant atteindre quelques centaines de parsecs. Par conséquent, elles ne survivent que pendant quelques dizaines de millions d'années car elles sont facilement perturbées par la nature dissipative du disque galactique (Moraux, 2016). Au contraire, les AO sont liés, beaucoup plus denses que les associations, ont une densité stellaire $\sim$10 fois plus grande et peuvent survivre pendant plusieurs milliards d'années.

Toutes ces définitions ne permettent pas de faire une distinction claire entre les AO et les Amas Globulaires, l'autre type d'amas que l'on trouve dans la Voie lactée. Plusieurs aspects permettent de distinguer ces deux types d'amas. Tout d'abord, selon Gratton et al. (2019), les Amas Globulaires se trouvent uniquement dans le bulbe, le halo et le disque épais de la Galaxie alors que les AO sont principalement situés dans le disque mince (voir Fig. C.2 pour une vue schématique de la Voie lactée). Cela implique que les Amas Globulaires sont âgés ($\sim$ 10 milliards d'années) alors que la plupart des AO sont plus jeunes qu'1 milliard d'années et se forment encore dans les bras spiraux de la Galaxie. Parce que les AO sont beaucoup plus jeunes que les Amas Globulaires, ils ont

été formées à partir d'un environnement déjà enrichi en métaux par la nucléosynthèse atomique des étoiles de population II. Les Amas Globulaires sont donc pauvres en métaux par rapport aux AO. Enfin, les AO sont composées de la même population d'étoiles, et sont chimiquement homogènes (Moraux, 2016) alors que les Amas Globulaires peuvent contenir plusieurs populations stellaires (Piotto et al., 2005) et peuvent être chimiquement inhomogènes (Gratton et al., 2012). Dans cette thèse, nous nous concentrons sur les AO de la Voie lactée.

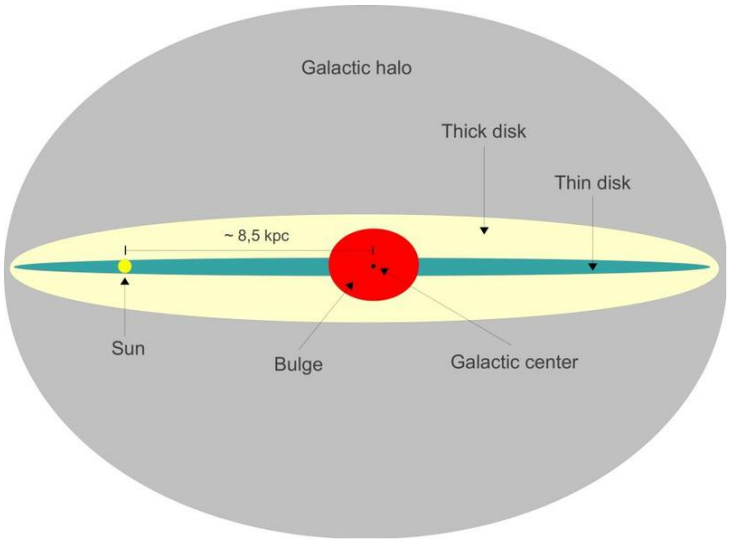

Figure C.2: Vue de profil schématisée de la Voie lactée, avec ses principales composantes indiquées. Crédit: Wikipedia

C.2.2 Propriétés des Amas Ouverts

Les AO ont longtemps été considéré comme le seul environnement dans lequel les étoiles se formaient. Cependant, les observations effectuées avec le télescope spatial *Spitzer* ont révélé la structure filamentaire de jeunes objets stellaires dans la région de formation stellaire NGC 1333 et ont mis en évidence le fait qu'ils suivent la distribution du gaz (Gutermuth et al., 2008). Cela montre que l'importance des amas dans le processus de formation des étoiles a peut-être été surestimée dans le passé, car les étoiles peuvent également se former de manière hiérarchique dans des environnements de densités variables, allant des associations lâches aux amas denses (Bastian, 2011; Gieles and Portegies Zwart, 2011). De nos jours, il serait plus exact de dire que les étoiles se forment en groupes plutôt qu'en amas (Moraux, 2016). Plus précisément, les amas liés se forment dans les noyaux denses des nuages moléculaires géants et les associations dans leurs régions de faible densité.

Les amas d'étoiles sont des objets clés afin de mieux comprendre les processus de formation stellaires. Ils constituent des sondes très utiles pour tester les théories d'évolution stellaire. En effet, les populations stellaires des AO couvrent généralement un large éventail de masses et de phases d'évolution stellaire, faisant de leurs diagrammes HR un aperçu

de l'évolution stellaire. De plus, leurs propriétés (principalement la distance et l'âge) sont déterminées avec une précision bien supérieure à celle des étoiles de champ. C'est pourquoi ils constituent des laboratoires idéaux afin de tester les théories d'évolution stellaire. Comme la formation d'étoiles se produit dans les bras spiraux, les AO jeunes peuvent aussi être utilisés pour tracer leurs positions et comprendre les mécanismes nécessaires à leur formation (Dias and Lépine, 2005; Castro-Ginard et al., 2021). Les amas anciens abritant des étoiles plus évoluées, ils constituent d'excellentes cibles spectroscopiques qui peuvent être utilisées pour déterminer avec précision leur VR ou leur composition chimique (Tarricq et al., 2021a; Casamiquela et al., 2020). De plus, comme leurs âges couvrent toute la durée de vie du disque galactique, retraçant les composantes jeunes comme les composantes anciennes du disque mince, ils sont aussi largement utilisés pour retracer l'histoire du disque galactique (Friel, 1995).

Des efforts répétés ont été conduits dans le passé pour construire de grands catalogues d'AO et pour fournir une estimation de leurs positions moyennes, mouvements propres, distance, etc. Le premier pas majeur dans cette direction correspond au catalogue de Lund (Lynga, 1982) qui rassemble environ 1200 entrées. Dans les années 1990, la base de données des AO galactiques (Mermilliod, 1995) couvre les données d'environ 100 000 étoiles appartenant à 500 amas, la plupart faisant partie du catalogue de Lund. Ensuite, Dias et al. (2002) a mis à jour le catalogue de Lund en rassemblant de nouvelles découvertes faites à l'aide des données 2MASS (Skrutskie et al., 1997). Ils ont fourni - lorsque suffisamment de données étaient disponibles - l'âge, la VR et le mouvement propre de plus de 1500 amas de manière systématique. Plus tard, Kharchenko et al. (2013) a publié un catalogue de $\sim$3 000 amas avec des estimations de leurs rayons, âge, position et mouvements propres en exploitant à la fois 2MASS et le relevé PPMXL (Roeser et al., 2010).

C.2.2.1 Formation & évolution des amas

Depuis la mission spatiale *Herschel*, on sait que les Nuages Moléculaires Géants sont composés d'un réseau de filaments et de sous-structures (Molinari et al., 2014). Au sein de ces nuages, la formation d'étoiles est un processus multi-échelle qui implique à chacune de ces échelles des processus physique différents. La formation d'étoiles se produit lorsque la masse d'un amas dépasse la masse de Jeans (Jeans, 1902), déclenchant l'effondrement gravitationnel du nuage. Cet effondrement gravitationnel peut également être déclenché par le passage d'un bras spiral (entre autres processus, Shu et al., 1972). Les nuages moléculaires sont en outre soumis à de la turbulence interne. À l'échelle des nuages, la turbulence agit généralement contre l'effondrement gravitationnel, mais à plus petite échelle, elle comprime le gaz et fragmente le nuage en amas et en filaments (McKee and Ostriker, 2007). Les amas massifs peuvent subir de multiples événements de fragmentation gravoturbulente, conduisant à la formation de plusieurs étoiles. Une description plus détaillée des nuages moléculaires ou du processus physique déclenchant le début de la formation d'étoiles n'entre toutefois pas dans le cadre de cette thèse.

Les AO se forment au sein des sous-structures massives et denses situés à l'intersection des longues structures filamentaires des Nuages Moléculaires Géants (Lada, 2010). L'efficacité

de la formation d'étoiles étant relativement faible (Lada and Lada, 2003; Clark et al., 2004), seule une fraction du gaz des Nuages Moléculaires Géants est effectivement consommée pour former des étoiles. Un amas récemment né est donc composé à la fois d'étoiles et de gaz en équilibre viriel (Kroupa, 2005). Sous l'effet des vents stellaires ou de l'explosion d'une supernova, le gaz est ensuite expulsé de l'amas, ce qui abaisse son potentiel gravitationnel. Ayant perdu son gaz, l'amas n'est plus en équilibre viriel et doit se relaxer dynamiquement pour atteindre un autre état d'équilibre. Au cours de ce processus, un nombre important d'étoiles de l'amas sont expulsées de celui-ci et la plupart des amas sont détruits. Ceux qui survivent souffrent d'une perte massive d'étoiles (Baumgardt and Kroupa, 2007). Lada and Lada (2003) a postulé que $\sim 90\%$ des amas sont ainsi détruits et a nommé ce processus la mortalité infantile. Dans les amas survivants, cette perte d'étoiles peut également conduire à la formation de structures en forme de queue de distribution (Hu et al., 2021; Meingast et al., 2021).

Les amas jeunes, toujours entourés de gaz et de poussière, sont obscurcis dans les bandes photométriques optiques et seules les observations infrarouges peuvent révéler leurs structures. Lada et al. (1993) et Lada et al. (1996) ont respectivement mené des observations infrarouges pionnières de NGC 2264 et NGC 1333 dans les bandes J, H et K. Ils ont montré que certains des amas encore entourés de gaz et de poussière ne sont pas nécessairement des systèmes présentant un unique centre, mais sont au contraire composés de nombreuses sous-structures montrant qu'ils ont gardé dans leur distribution spatiale l'empreinte de leur Nuage Moléculaire Géant progéniteur (Gutermuth et al., 2008). La densité de surface de ces amas hiérarchiquement formés ne présente pas de profils de densité radiale lisses et est mal décrite par des fonctions simples de type loi de puissance. Au contraire, les systèmes présentant un unique centre qui se sont formés à l'intersection des filaments de gaz sont bien décrits par des fonctions comme la fonction EFF (Elson et al., 1987) ou le profil de King (King, 1962). Le profil EFF ne suppose pas un équilibre énergétique de l'amas et il est donc plus adapté aux amas jeunes qui ne se sont pas encore complètement relaxés. Après une période de relaxation (i.e. le temps nécessaire à un système pour atteindre un équilibre énergétique, typiquement quelques dizaines de millions d'années), les amas ont perdu toute trace de leurs conditions initiales, ils sont complètement relaxés et leur profil de densité radiale devient bien ajusté par la fonction de King. La période de relaxation est définie telle que :

$$t_{relax} \simeq \frac{0.1N}{\ln N} t_{cr}, \tag{C.1}$$

(Binney and Tremaine, 1987; Lada and Lada, 2003) avec N le nombre de membres de l'amas et t_{cr} le temps de traversée dynamique qui réprésente le temps nécessaire pour une étoile typique afin de faire une révolution autour du système.

La cohésion interne des AO subit continuellement des perturbations dues à des processus physiques internes et externes. L'évolution stellaire fait perdre aux amas une partie de leur masse de par l'explosion de supernova et les vents stellaires. Les interactions gravitationnelles avec les Nuages Moléculaires Géants sont également connues pour perturber fortement l'équilibre des amas (Gieles et al., 2006). Le passage des bras spiraux

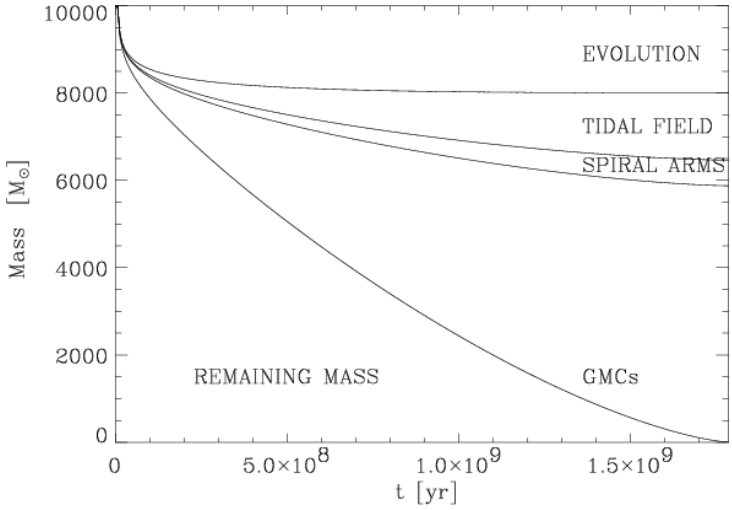

Figure C.3: Evolution de la masse d'un amas du voisinage solaire ayant une masse initiale de 10^4 $M_\odot$. La ligne continue représente la contribution à la perte de masse de respectivement (et de haut en bas) : l'évolution stellaire, le harcèlement galactique, le passage des bras spiraux et les rencontres avec des nuages moléculaires. Tiré de Lamers and Gieles (2006).

contribue également à l'évaporation des étoiles des amas stellaires (Gieles et al., 2007). Enfin, le harcèlement galactique, c'est-à-dire les interactions gravitationelle avec le potentiel galactique, peut également contribuer progressivement à la dislocation des amas (Baumgardt and Makino, 2003). Lamers and Gieles (2006) ont utilisé un modèle analytique approximatif pour estimer la perte de masse subie par les AO dans le voisinage solaire, en tenant compte de tous ces processus. La figure C.3 est tirée de leur travail. On peut y voir que le processus de dissolution d'un amas avec une masse initiale de 10^4 $M_\odot$ est dominé par les rencontres avec les Nuages Moléculaires Géants.

Ces événements sont complétés par un autre processus dynamique : la ségrégation de masse (Spitzer, 1969). En raison de l'équipartition de l'énergie cinétique dans les rencontres gravitationnelles entre les membres de l'amas, les étoiles massives se déplacent autour du centre des amas plus lentement que les membres moins massifs. Les étoiles massives plongent donc vers les centres des amas et orbitent sur des orbites plus basses que les étoiles moins massives. La ségrégation de masse a été très étudiée dans les AO et on pense qu'elle contribue à l'évaporation des amas (Mathieu, 1984; Kroupa, 1995; de La Fuente Marcos, 1996). Les étoiles de faible masse, sur des orbites extérieures, sont en effet plus susceptibles d'être arrachées au potentiel gravitationnel de l'amas par les forces de marée évoquées précédemment. La multitude de processus physique en jeu dans la dislocation d'un amas implique que la distribution d'âge des AO est fortement biaisée : très peu d'amas atteignent des âges supérieurs à 1 milliard d'années par rapport au nombre d'amas qui se forment.

C.2.2.2 Distribution spatiale des amas

L'étude de la structure de la Voie lactée n'est pas une tâche facile compte tenu de la position du Soleil, situé à l'intérieur du disque galactique. Ainsi, un aperçu détaillé des

caractéristiques des bras spiraux fait toujours défaut. L'origine, l'emplacement et même le nombre des bras spiraux restent encore incertains. Sachant que la formation d'étoiles se produit dans les bras spiraux, leur position peut être retracée par des populations jeunes, telles que les jeunes AO (Roberts, 1969). Dias and Lépine (2005) ont utilisé les jeunes amas de leur échantillon publié dans Dias et al. (2002) pour étudier la nature des bras spiraux. La figure C.4 montre les positions de leurs amas dans trois tranches d'âge différentes. Il est clair que les amas très jeunes de leur échantillon contraignent efficacement la position des bras spiraux. Nous notons également qu'ils remplissent rapidement la région inter-bras : les bras spiraux sont toujours distinguables dans la figure du milieu même si les amas ont déjà commencé à dériver. Aucune structure n'est visible dans la figure du bas, illustrant la dispersion des amas par les nuages moléculaires et par les composantes non-axisymétriques du disque. Des études plus approfondies ont tenté de cartographier les bras spiraux dans l'environnement du Soleil. Par exemple, (Vázquez et al., 2008) a remarqué une sous-densité de la distribution des amas dans le bras de Persée. Cette sous-densité, également observée dans les régions de formation d'étoiles massives de Reid et al. (2014) et dans la distribution des régions HII de Hou and Han (2014), pourrait changer notre compréhension du bras de Persée et des bras locaux, mais aucune conclusion n'a encore été tirée.

La dispersion des AO depuis leur lieu de naissance peut également être observée dans la direction verticale en étudiant l'échelle de hauteur du disque. Les AO sont principalement situés dans le disque mince même si les amas âgés qui ont été dispersés par leur rencontre avec les Nuages Moléculaires Géants et par les bras spiraux (Spitzer and Schwarzschild, 1951; Lacey, 1984; Jenkins and Binney, 1990) peuvent également s'aventurer dans le disque épais. Les vieux amas ont donc tendance à être situés à des altitudes galactiques plus élevées que les jeunes (Friel, 1995; Cantat-Gaudin et al., 2018a). De plus, les amas qui restent longtemps dans le plan médian galactique sont plus susceptibles d'être perturbés par les différents processus discutés dans la Partie C.2.2.1. Au contraire, les amas qui s'éloignent du plan médian survivent plus longtemps car ils souffrent moins de ces effets, en particulier des rencontres avec les Nuages Moléculaires Géants. Le manque d'amas âgés à basse altitude est également un produit de cet effet de sélection.

Bonatto et al. (2006) ont examiné la distribution verticale d'un échantillon de plus de 600 AO. En supposant que leur distribution pouvait être décrite par un simple profil de décroissance exponentielle et que les AO jeunes étaient distribués symétriquement au-dessus et au-dessous du plan, ils ont calculé l'échelle de hauteur des amas dans différentes tranches d'âge ainsi que la position du Soleil au-dessus du plan galactique. Ils ont trouvé que le Soleil était à une altitude de 14 pc au-dessus du plan médian galactique. En ce qui concerne l'échelle de hauteur des AO, ils rapportent une augmentation d'un facteur trois entre les amas plus jeunes que 200 millions d'années et les amas d'âge intermédiaire (ayant un âge compris entre 200 millions et 1 milliard d'années) mais une distribution uniforme des amas plus vieux qu'1 milliard d'années. Ils signalent également une nette augmentation de l'échelle de hauteur des amas situés dans la partie externe de la Galaxie par rapport à ceux situés dans la partie interne. Comme précédemment, il pourrait s'agir d'un simple effet de sélection dû à un taux de survie plus élevé des AO à grand rayon

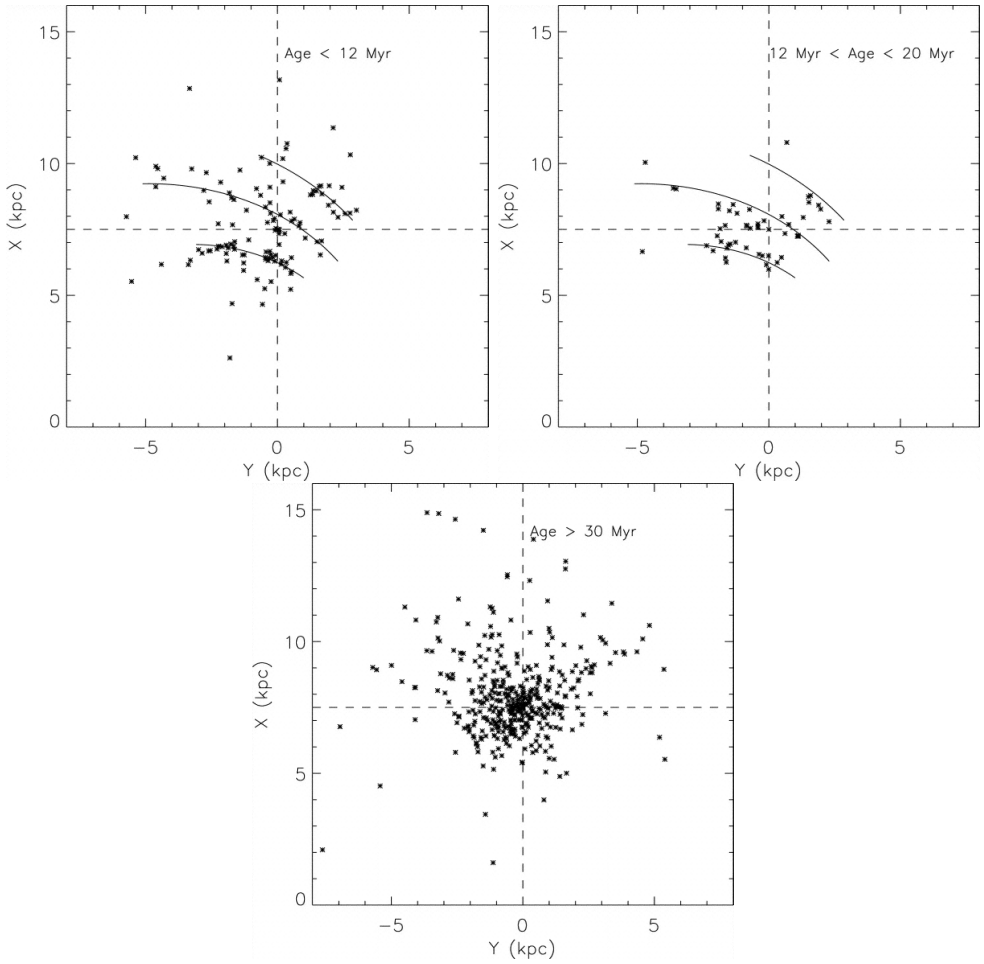

Figure C.4: Distribution dans le plan $(X - Y)$ des amas de Dias and Lépine (2005). Dans le panneau situé en haut à gauche sont représentés les amas plus jeunes que 12 Myrs et dans le panneau situé au haut à droite sont représentés les amas avec un âge compris entre 12 et 20 Myrs. La ligne continue représente la position actuelle des bras spiraux. Le panneau du bas montre la distribution des amas plus vieux que 30 Myr. Le Soleil est situé aux coordonnées (0, 7.5) dans tous les panneaux.

galactocentrique.

C.2.2.3 Cinématique des amas

Les premières études de la cinématique des AO étaient généralement consacrées à la compréhension du champ cinématique local. Hron (1987) a mené l'une de ces études en compilant les VR de 105 jeunes amas. Il a utilisé ces données pour mesurer le mouvement solaire et l'une des constantes de Oorts. Il a aussi utilisé son échantillon d'amas afin de tracer la courbe de rotation galactique et a trouvé que la distribution de vitesses azimutales des amas est en accord au niveau du rayon solaire avec celles des régions HII et des nuages moléculaires obtenus respectivement par Georgelin and Georgelin (1976) et Clemens (1985). Plus tard, Scott et al. (1995) ont étudié la cinématique des amas de plus d'1 milliard d'années et ont trouvé qu'elle était cohérente avec celle de Hron (1987) mais

avec une dispersion beaucoup plus grande (29 kms^{-1} au lieu de 10 kms^{-1}) pour les jeunes AO. Scott et al. (1995) sont ainsi parvenu à constater, sans aucune information sur les mouvements propres de leur échantillon d'amas, l'augmentation de la relation âge-vitesse des AO.

Dans les années qui ont suivies, la publication des catalogues *Hipparcos*, Tycho-2, UCAC et PPMXL a permis d'ajouter 2 dimensions supplémentaires à l'étude des propriétés cinématiques des amas. Nous avons vu dans la Partie C.2.2.2 que les amas pouvaient être utilisés afin d'étudier la position des bras spiraux. En ajoutant l'information cinématique 3D complète, Dias and Lépine (2005) ont déterminé la vitesse de rotation des bras spiraux. En intégrant les orbites de leurs jeunes amas, ils ont trouvé une vitesse angulaire unique pour tous les bras spiraux étudiés, indiquant que ceux-ci tournent comme des corps rigides.

Les informations sur le mouvement propre combinées aux VR pour de grandes quantités d'AO ont rendu possible l'étude tridimensionnelle de la relation âge-vitesse des amas d'étoiles. En divisant leur échantillon de 488 amas en différentes tranches d'âge, Wu et al. (2009) ont constaté que la dispersion augmentait avec l'âge pour chaque composante de la vitesse. Cette augmentation de la dispersion de la vitesse avec l'âge, déjà observée avec les étoiles de champ, est une preuve du chauffage cinématique des amas dans le disque galactique (Wielen, 1977; Binney and Tremaine, 2008). Plus tard, Soubiran et al. (2018a) ont fourni les informations dans l'espace de phase en 6 dimensions pour 861 amas et ont également examiné la relation âge-vitesse. Ils ont mesuré une augmentation significative de la dispersion verticale des vitesses pour leur population la plus ancienne avec une dispersion des vitesses représentative du disque mince.

C.2.2.4 Composition chimique des amas

Comme expliqué dans la Partie C.2.1, les AO sont des groupes d'étoiles chimiquement homogènes. La détermination de la composition chimique d'un amas bénéficie donc des mêmes avantages que pour les autres paramètres des amas : comme elle peut être basée sur plusieurs étoiles, la détermination de la composition chimique d'un amas est beaucoup plus précise que celle d'une seule étoile. La signature chimique des AO rappelle l'abondance chimique des Nuages Moléculaires Géants à partir desquels ils se sont formés. Comme on trouve des amas à tous les âges et à tous les endroits du disque galactique, ils constituent des objets clés afin de retracer son évolution et sa composition chimique et en particulier son enrichissement en éléments lourds par nucléosynthèse.

Dans le catalogue d'amas de Dias et al. (2002), la métallicité des AO s'étend sur la plage $-1,0 < [\mathrm{Fe/H}] < 0,5$ dex avec $\sim 75\%$ des amas ayant une métallicité comprise dans la plage $-0,4 < [\mathrm{Fe/H}] < 0,08$ dex. La première étude de la signature chimique du disque à l'aide des AO a été menée par Janes (1979) qui a trouvé une diminution de la distribution de la métallicité lorsque le rayon Galactocentrique croît. Ces résultats ont été confirmés par de nombreuses autres études comme celle de Friel (1995). Twarog et al. (1997) ont proposé que le gradient de métallicité des AO n'était pas une simple fonction linéaire du rayon galactocentrique mais était en fait mieux décrit par deux fonctions linéaires avec des pentes différentes et une transition aux alentours du rayon solaire. Ceci

semble être confirmé par les données les plus récentes de relevés spectroscopiques à grande échelle comme APOGEE (Donor et al., 2020).

Des déterminations précises des abondances des AO permettent également d'évaluer le niveau d'homogénéité chimique au sein des différentes étoiles d'un amas (Liu et al., 2016). Casamiquela et al. (2020), grâce à l'analyse différentielle de spectres haute résolution et à haut rapport signal/bruit d'étoiles jumelles, ont utilisé cette approche. Ils ont détecté un léger niveau d'inhomogénéité (0,02-0,03 dex) dans les trois amas étudiés : Ruprecht 147, les Hyades et NGC 2632. Bien que le niveau d'inhomogénéité soit trop faible pour tirer une quelconque conclusion sur la formation d'étoiles, certaines des étoiles à queue de marée identifiées par Röser et al. (2019) ont été écartées des listes de membres sur la base de leur signature chimique.

C.2.3 Détermination de listes de membres

L'identification des membres d'un amas nécessite des données d'observation à grande échelle et la combinaison de plusieurs types d'observations. Comme nous l'avons vu dans la Partie C.2.1, les étoiles appartenant à un même amas sont nées au même moment et partagent le même âge et la même métallicité. Des données photométriques peuvent donc être utilisées pour dessiner le diagramme HR des étoiles pouvant appartenir à l'amas et vérifier si un isochrone peut s'adapter à la distribution des étoiles. Cependant, les étoiles peuvent aussi suivre la séquence principale d'un amas par simple alignement fortuit. Des observations spectroscopiques de suivi des étoiles peuvent alors être effectuées afin de comparer la métallicité ou la VR des étoiles.

D'autres propriétés partagées par les membres d'un amas peuvent également être utilisées pour établir des listes de membres d'AO : la distance et les vitesses galactiques des étoiles d'un même amas sont très similaires. Avec l'avènement de la mission astrométrique à grande échelle *Hipparcos* et les publications successives des catalogues Hipparcos et Tycho-2 (Perryman et al., 1997; Høg et al., 2000), plusieurs amas ont été identifiés par des recherches systématiques conçues afin de traquer les surdensités d'étoiles dans l'espace des position ou des mouvements propres, complétées par des critères photométriques (Platais et al., 1998; Alessi et al., 2003). Lorsque les catalogues de mouvement propre du ciel entier tels que le PPMXL (Roeser et al., 2010) et le quatrième UCAC (Zacharias et al., 2013) ont été publiés, les données astrométriques sont devenues systématiquement utilisées dans l'analyse des membres des AO (Kharchenko et al., 2013; Dias et al., 2014).

Des méthodes statistiques puissantes capables de combiner les informations astrométriques et photométriques ont également été développées (Sarro et al., 2014; Olivares et al., 2018b). L'utilisation des magnitudes et des couleurs des étoiles ainsi que de leurs mouvements propres et de leurs positions permet à ces méthodes d'être moins biaisées par les incertitudes élevées sur les mouvements propres des étoiles peu lumineuses. Elles permettent également une sélection plus complète et plus exhaustive des membres, mais sont moins faciles à adapter à un grand nombre d'amas. Avec la mission *Gaia*, la chasse aux AO et à leurs membres est rentrée dans une nouvelle ère. La suite de cette thèse se concentrera et apportera plus de détails sur les découvertes liées à *Gaia*.

C.3 Les Amas Ouverts à l'ère de *Gaia*

C.3.1 *Gaia*

La mission *Gaia* est un projet de l'Agence Spatiale Européenne conçu pour cartographier la Voie lactée afin de mettre en lumière sa structure, sa formation et son évolution (Gaia Collaboration et al., 2016). Ceci étant ses principaux objectifs scientifiques, *Gaia* fournit également des informations cruciales sur la formation et l'évolution des étoiles et permet de détecter des exoplanètes, des supernovae, des naines brunes et blanches. Pour ce faire, le satellite mesure l'astrométrie (c'est-à-dire la position, les parallaxes et les mouvements propres) de plus d'un milliard d'étoiles avec une précision sans précédent, il met à disposition leur spectrophotométrie optique ainsi que la VR de plusieurs millions d'étoiles. Cela représente des informations astrophysiques importantes pour environ 1% des étoiles de la Voie lactée. Alors qu'*Hipparcos*, la précédente mission astrométrique de l'Agence Spatiale Européenne avait publié un catalogue de ∼100 000 objets avec une solution astométrique complète, *Gaia* a multiplié ce nombre par plus de 10 000 et a amélioré la précision astrométrique des mesures par un facteur ∼ 100.

Le consortium *Gaia* publie périodiquement un catalogue rassemblant toutes les données disponibles. Son deuxième catalogue (la DR2), publié en avril 2018 était basé sur 22 mois d'observations (Gaia Collaboration et al., 2018b). Un aperçu de son contenu est présenté dans le panneau supérieur de la figure C.5. Une solution astométrique complète est fournie pour plus d'un milliard d'étoiles avec une gamme de magnitude de $3 < G < 21$ mag. Les incertitudes dépendent fortement de la magnitude des étoiles mais sont typiquement inférieures à 1 mas en parallaxe et à 1 mas/an en mouvements propres, même pour les cibles les plus faibles. Les VR fournis par *Gaia* pour plus de 7 millions d'étoiles d'une magnitude $4 < G < 13$ mag constituent déjà le plus grand catalogue de VR (Katz et al., 2019). La DR suivante a été divisé en deux versions : l'eDR3 (pour early Data Release) publiée en décembre 2020 et la DR3 attendu pour le premier semestre 2022. Comme le montre le panneau inférieur de la figure C.5, le nombre d'étoiles avec une solution astrométrique complète est ∼ 10% plus élevé dans l'eDR3 que dans la DR2. Basée sur 34 mois d'observations, les précisions des mesures de positions et de parallaxes ont été améliorées d'un facteur 0,8 tandis que les incertitudes sur les mouvements propres ont été améliorées d'un facteur 0,5 (Lindegren et al., 2021b). Le nombre de sources ayant une photométrie en bande G, G_{BP} et G_{RP} a également augmenté d'environ ∼ 10%. Dans la DR3 complète, les informations sur les sources variables, les températures effectives, l'extinction, le rougissement, le rayon et la VR sont attendues pour beaucoup plus d'étoiles en comparaison avec la DR2. Les différents catalogues *Gaia* constituent les relevés les plus grands et les plus homogènes à ce jour dans le domaine de la physique stellaire et galactique.

C.3.2 Quelques résultats obtenus avec *Gaia*

Depuis la sortie de la DR2, *Gaia* a permis de comprendre l'origine de certaines étoiles dans la Galaxie. Par exemple, Helmi et al. (2018), en examinant les signatures cinématiques

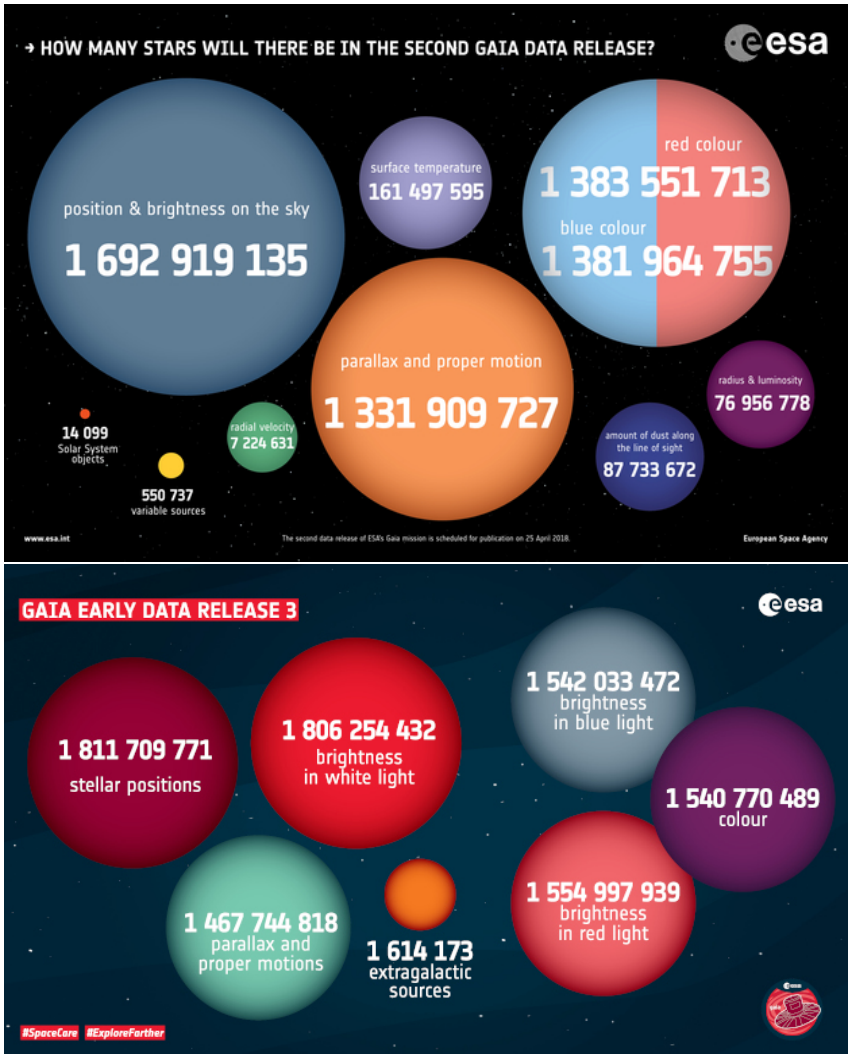

Figure C.5: Aperçu du contenu de *Gaia* DR2 (en haut) et eDR3 (en bas). Crédit: ESA

des étoiles dans un rayon de 2,5 kpc autour du Soleil a détecté qu'une fraction significative d'entre elles avait un mouvement rétrograde. L'analyse photométrique et chimique de ces étoiles a confirmé que ces étoiles situées dans le halo proche ne pouvaient pas avoir été formées dans le disque épais mais provenaient d'un système différent qui a été accrété à la Voie lactée. Ils ont appelé ce système, qui a fusionné avec notre galaxie il y a environ 10 milliards d'années, le système *Gaia*-Enceladus. La précision sans précédent et la couverture de tout le ciel des données de *Gaia* ont également dévoilé des structures inédites. De fines crêtes diagonales saillantes apparaissent dans l'espace de la vitesse azimutale en fonction du rayon galactocentrique (voir Fig. C.6). Ces crêtes diagonales pourraient avoir plusieurs origines : elles pourraient être le signe d'un mélange de phase en cours dans le disque galactique ou elles pourraient aussi être induites par la barre et les bras spiraux. En effet, la barre et les bras spiraux créent des structures orbitales résonantes dans le disque, c'est-à-dire des régions d'orbites stables et instables. Les régions d'orbites instables se traduisent par des sous-densités dans la distribution des étoiles tandis que les régions d'orbites stables se traduisent par des surdensités d'étoiles identifiables comme

les crêtes de la Fig. C.6. Ces crêtes diagonales pourraient également être créées par une combinaison de mélange de phases et de structures orbitales résonantes.

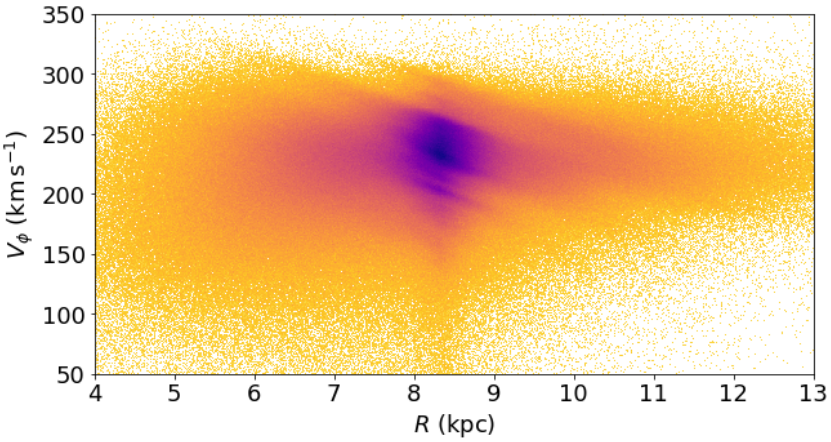

Figure C.6: Histogramme en deux dimensions des vitesses azimutales en fonction du rayon galactocentrique, avec des intervalles de 1 km s^{-1} en V_ϕ et de 0,01 kpc en R. Tiré de (Antoja et al., 2018).

Parallèlement, la vision des AO s'est considérablement améliorée depuis la publication de la DR2 de *Gaia*. Les catalogues d'AO précédents, comme ceux de Kharchenko et al. (2013) et Dias et al. (2002), étaient des collections hétérogènes de différentes études, chacune ayant ses propres caractéristiques, biais et incertitudes, rendant très difficile une caractérisation uniforme des amas. La publication du catalogue *Gaia*-DR2 a permis de revisiter le recensement des AO dans la Voie lactée et de déterminer leurs paramètres moyens de manière homogène. Sur la base de la liste de plus de 3 000 amas et candidats connus, provenant pour la plupart des deux catalogues susmentionnés, Cantat-Gaudin et al. (2018a) a obtenu une liste de membres et de paramètres moyens pour seulement 1 229 d'entre eux en utilisant l'algorithme de clustering UPMASK (Krone-Martins and Moitinho, 2014). Même si certains des amas signalés avant l'étude *Gaia* étaient soit encore entourés de gaz et de poussière, soit observables uniquement aux longueurs d'onde infrarouges, un grand nombre d'objets considérés comme des amas se sont avérés ne pas être de véritables groupes d'étoiles. Cela souligne, selon les auteurs, la nécessité de réexaminer ces objets à la lumière des récents catalogues *Gaia*. En outre, ils ont découvert par hasard 60 AO dans le voisinage immédiat d'objets connus, ce qui montre que le recensement des AO, que l'on croyait complet jusqu'à 1,8 kpc, était loin de l'être.

Des centaines de nouveaux AO et leurs membres ont également été identifiés par Castro-Ginard et al. (2018, 2019, 2020) qui ont développé une approche d'apprentissage automatique destinée à repérer les surdensités dans l'espace à cinq dimensions des positions, parallaxes et mouvements propres. Ces travaux ont permis d'établir un catalogue de plus de 600 nouveaux AO. Cantat-Gaudin et al. (2019b) ont également recherché des surdensités dans la direction du bras de Persée et a détecté 41 nouveaux amas. Sim et al. (2019) ont inspecté visuellement les distributions stellaires dans l'espace des coordonnées

galactiques et des mouvements propres et ont identifié 207 nouveaux potentiels amas. De plus, Liu and Pang (2019) ont utilisé un algorithme *friend-of-friend* déjà largement utilisée dans la communauté étudiant les amas de galaxies pour identifier 76 amas précédémment inconnus. Kounkel and Covey (2019) ont appliqué l'algorithme d'apprentissage automatique non supervisé HDBSCAN afin d'identifier non seulement les amas mais aussi les groupes et associations mobiles dans un rayon de 1 kpc. Plus récemment, Cantat-Gaudin et al. (2020), ont également publié un catalogue de 2 017 AO précédemment identifiés par les auteurs susmentionnés et ont déterminé pour la plupart d'entre eux et de manière homogène leurs membres, distances et âges. Cetté étude constitue le catalogue le plus complet d'AO.

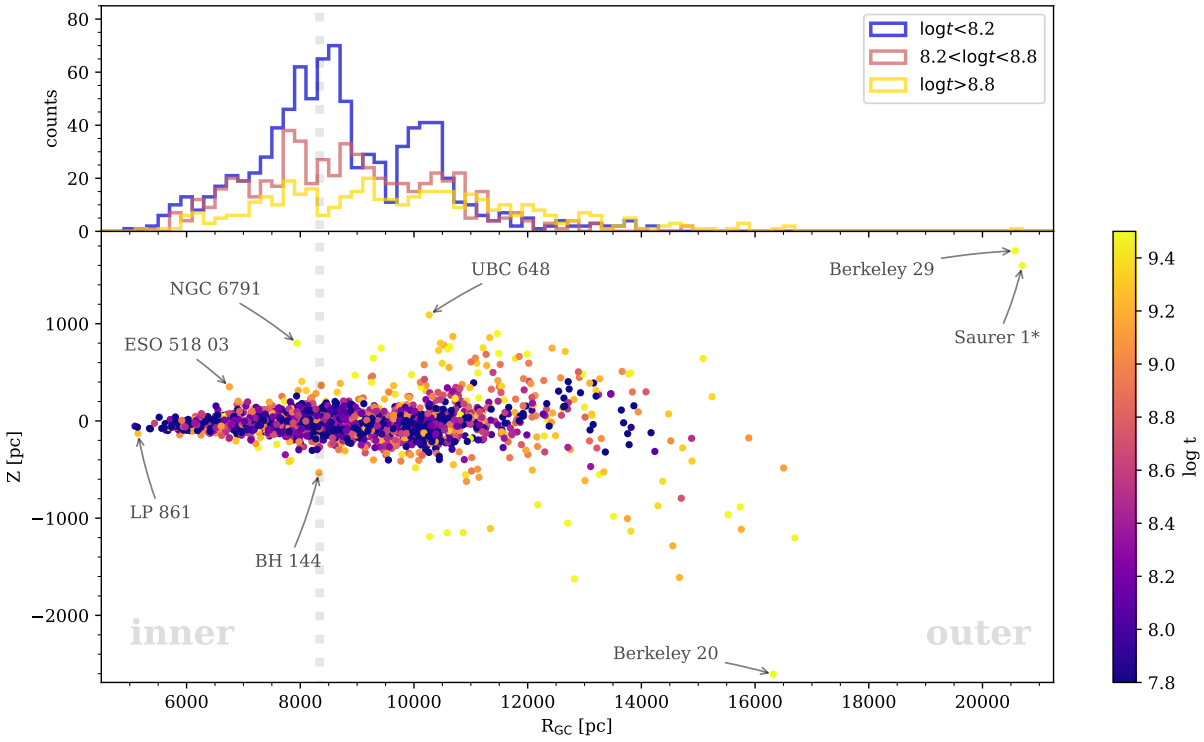

Figure C.7: En haut : distribution galactocentrique de l'échantillon d'amas de Cantat-Gaudin et al. (2020) dans trois tranches d'âge. En bas : altitude au-dessus du plan médian galactique de ces amas en fonction de leur rayon galactocentrique, les point étant colorés en fonction de l'âge des amas. Dans les deux panneaux, la ligne pointillée verticale représente le rayon galactocentrique du Soleil. Tiré de Cantat-Gaudin et al. (2020).

Ce catalogue homogène a permis aux auteurs d'étudier, entre autres, la hauteur d'échelle du disque galactique et de revisiter les résultats de Bonatto et al. (2006) discutés dans la Sect. C.2.2.2 mais avec un nombre sans précédent d'amas et des déterminations de leurs âges plus fiable. La Fig. C.7 montre l'augmentation de l'altitude des amas en fonction de leurs âges ainsi que de leur rayon Galactocentrique. Les auteurs ont trouvé une altitude du Soleil de 23±3 pc. Alors que Bonatto et al. (2006) n'a pas pu constater une augmentation de la hauteur d'échelle pour les très vieux amas, Cantat-Gaudin et al. (2020) rapportent qu'elle atteint plusieurs centaines de parsecs.

Castro-Ginard et al. (2021) ont également revisité le travail effectué par Dias and Lépine (2005) sur la nature des bras spiraux avec les données *Gaia*. En utilisant le plus grand échantillon d'amas avec des informations cinématiques complètes et des déterminations d'âge (mouvements propres et âges ainsi que VR provenant respectivement de: Cantat-Gaudin et al., 2020; Tarricq et al., 2021a), ils ont mesuré différentes vitesses angulaires pour chacun des bras spiraux étudiés. Comme ils ne trouvent pas que les bras spiraux tournent comme des corps rigides, ces résultats sont en meilleur accord avec les simulations et les études observationnelles basées sur la cinématique des étoiles de champ. Ceci impliquerait que, plutôt que d'être des structures à longue durée de vie selon la théorie des ondes de densité (Lin and Shu, 1964), les bras spiraux auraient plutôt de courtes durées de vie et un comportement transitoire (Toomre, 1964).

C.4 Aperçu de la thèse

L'objectif de cette thèse est de mettre en évidence certains des processus physiques à l'œuvre dans notre galaxie, la Voie lactée, en étudiant en particulier les AO, parfois appelés les "pierres angulaires" des galaxies. De nombreux phénomènes restent mal compris dans la Voie lactée : le nombre et la nature des bras spiraux comme son histoire dynamique. Pour dévoiler certaines de ses propriétés, les modèles théoriques doivent être enrichis et mieux contraints afin qu'ils soient en mesure de reproduire pleinement la complexité et la très grande variété des phénomènes physiques en jeu. L'observation des AO est l'un des meilleurs moyens d'y parvenir, car la détermination des paramètres des amas est beaucoup plus facile et précise que pour tout autre objet de la Galaxie. La mission *Gaia*, avec son volume sans précédent de données astrométriques, contribue à contraindre les modèles théoriques en nous fournissant des informations cinématiques pour plus d'un milliard d'étoiles. Dans le domaine des amas stellaires, elle a permis de revoir le recensement des AO et de fournir de nouvelles listes de membres extrêmement fiables.

Nous nous sommes principalement concentrés sur l'étude de la relation entre la morphologie et la cinématique des AO avec leurs propriétés telles que leur âge ou leur position dans la Galaxie. Une attention particulière a été portée sur les régions externes des amas afin de dévoiler les différents mécanismes en jeu dans leur dislocation.

Ce manuscrit est organisé comme suit. Dans le chapitre 2, nous présentons la méthodologie utilisée afin de détecter des nouveaux membres à grande distance des centres des amas. Nous présentons les biais inévitables de chacune des méthodes utilisées et montrons les résultats de cette approche de regroupement. Le Chapitre 3 est consacré à l'étude de la morphologie des AO. Nous montrons ici une étude tridimensionnelle des amas proches ainsi que les limites de cette approche dues aux incertitudes sur la parallaxe de *Gaia*. Dans la dernière partie de ce chapitre (Partie 3.5), nous présentons l'analyse morphologique des amas en utilisant les listes de membres obtenues dans le chapitre 2 et évaluons comment leurs propriétés sont corrélées avec l'âge et la position galactique des amas. Dans le chapitre 4, nous expliquons les différentes étapes nécéssaires à la constitution de notre catalogue de VR et nous utilisons ces VR pour calculer les coordonnées galactocentriques

en six dimensions du plus grand échantillon d'AO à ce jour. Nous étudions ensuite les propriétés cinématiques de notre échantillon d'amas ainsi que leurs paramètres orbitaux. Enfin, dans le chapitre 5, nous présentons les principaux résultats de cette thèse ainsi que ses limites et perspectives.

Appendix D

Conclusion, limites et perspectives (en français)

D.1 Résumé et Conclusions

Cette thèse porte sur l'étude de la cinématique et de la morphologie de la population des AO connus de la Galaxie. Dans ce but, nous avons développé une méthodologie capable de détecter des structures étendues autour de plusieurs centaines d'entre eux. Nous avons revisité leurs paramètres structurels à la lumière de ces nouvelles listes de membres. Nous avons également combiné des études au sol avec *Gaia* et les nouvelles listes de membres afin d'étudier en détail les propriétés cinématiques des AO.

D.1.1 Contexte

Les AO constituent des sondes idéales de l'histoire et des caractéristiques du disque galactique. La détermination de leurs paramètres physiques tels que leur âge, leur distance, leur composition chimique et leur cinématique est plus précise que pour les étoiles de champ, ce qui en fait des objets clés dans les études d'archéologie galactique. Depuis la publication de la DR2 de la mission *Gaia*, le recensement des AO a radicalement changé. La quantité massive de données, la couverture totale du ciel et la précision inégalée de la mission *Gaia* a permis de revisiter les listes de membres des amas connus de la Voie lactée.

En particulier, CG+18 a publié de nouvelles listes de membres pour plus de 1100 amas, a écarté des centaines d'astérismes précédemment considérés comme des AO et a découvert 60 nouveaux amas. Certains d'entre eux se trouvaient à de faibles distances alors que l'on pensait avant *Gaia* que le recensement des AO était complet jusqu'à 1,8 kpc (Kharchenko et al., 2013). La quantité de données *Gaia* à analyser était si énorme que les découvertes d'AO se sont succédées. De plus, alors que les queues de marée de certains GC avaient déjà été identifiées (comme pour Palomar 5 : Odenkirchen et al., 2001), aucune structure de ce type n'était connue pour les AO, et ce même pour les plus proches. Avec les données *Gaia*-DR2, certaines études dédiées à l'étude de petits échantillons d'AO ont révélé des structures à la périphérie de certains d'entre eux. Les deux études principales sur ce sujet sont celles de Pang et al. (2021) et Meingast et al. (2021) qui ont détecté de telles structures pour plusieurs amas proches.

Cependant, à l'exception de Castro-Ginard et al. (2020) qui a identifié la queue de marée d'UBC 274/LP 5, un amas situé à 1,8 kpc du Soleil, toutes les études dédiées à la

détection de structures étendues autour des AO se sont concentrées sur de petits échantillons d'amas. Une étude homogène et complète de la périphérie de la population connue d'AO fait encore défaut alors que la richesse, l'exactitude, la précision et l'homogénéité des données de *Gaia* permettent ce type d'étude.

L'étude cinématique des AO a également été grandement améliorée par la publication de la DR2. Les plus grands catalogues de VR pour AO ont vu leur taille augmenter deux fois l'année suivant la publication du DR2 avec les études de Soubiran et al. (2018a) et Carrera et al. (2019a) qui ont pleinement exploité les dernières déterminations de membres. La combinaison des données *Gaia*-RVS avec les études au sol permet également de vérifier la cohérence et les décalages entre les différentes études et de comparer la cinématique des AO avec les étoiles de champ.

D.1.2 Listes de membres et structure des Amas Ouverts

Nous avons développé une méthodologie capable d'identifier les membres d'amas jusqu'à 50 pc de leurs centres et nous publions ces listes de membres pour plusieurs centaines d'amas. Pour ce faire, nous avons tiré parti de l'algorithme d'apprentissage automatique HDBSCAN. Nous fournissons de nouvelles listes de membres pour 389 amas jusqu'à 1,5 kpc du Soleil, ce qui représente la plus grande étude homogène de la périphérie des AO. Ces listes de membres nous permettent d'identifier 71 amas avec une queue de marée. Parmi ces 71 amas, plus de 60 sont identifiés pour la première fois. De plus, alors que seule une poignée d'amas (qui avaient fait l'objet d'études spécifiques) étaient connus pour avoir des halos étendus, notre méthodologie met en évidence que la majorité des AO de notre échantillon sont entourés de halos proéminents qui peuvent contenir jusqu'à plusieurs centaines d'étoiles.

Notre nouvelle compilation de membres nous permet de revisiter la morphologie en deux dimensions des AO. Avec cette étude bidimensionnelle, nous exploitons les meilleures performances de *Gaia*. Nous constatons que les amas sont beaucoup plus grands qu'on ne le pensait auparavant, et qu'ils s'étendent jusqu'à la limite de notre cône de recherche (50 pc). Nous observons également ce qui semble être une corrélation négative entre les tailles des cœur des AO et leur âge et une corrélation positive entre leurs tailles globales et leurs âges. Nous étudions également la proportion d'étoiles peuplant les noyaux et les halos des amas et montrons que les amas âgés ont tendance à être proportionnellement moins peuplés que les jeunes. Si nous ne trouvons pas de preuve que la ségrégation de masse augmente avec l'âge, nous constatons néanmoins que les amas montrant des signes significatifs de ségrégation de masse sont en moyenne plus vieux que les amas qui ne présentent pas ces mêmes signes.

L'augmentation de la taille globale des amas avec l'âge ainsi que la diminution de la proportion d'étoiles peuplant leurs halos reflètent les deux principaux processus dynamiques en jeu durant la vie des amas : la ségrégation de masse et l'évaporation. Nous observons en même temps que les halos des vieux amas semblent être peu peuplés et que les vieux amas semblent être plus fortement ségrégués en masse que les jeunes. Cela pourrait indiquer que les processus de perturbation déclenchés par les rencontres avec les

Nuages Moléculaires Géants, le harcèlement galactique et les passages des bras spiraux sont plus efficaces pour dépouiller les amas de leurs étoiles extérieures que la ségrégation de masse pour peupler les zones périphériques des amas.

D.1.3 Propriétés cinématique des amas ouverts

En utilisant les listes de membres disponibles pour la population globale des AO, nous avons cherché à construire le plus grand catalogue de VR pour les AO. En croisant ces listes de membres avec le *Gaia*-RVS ainsi qu'avec plusieurs études au sol, nous fournissons à la communauté le plus grand catalogue de VR pour les AO. Sur les 2 017 amas de la liste originale, 1 382 ont maintenant une détermination de leur VR. La moitié de ces déterminations est basée sur plus de trois étoiles et 38% d'entre elles ont également une incertitude sur la VR inférieure à 3 $\mathrm{km\,s^{-1}}$.

En calculant les vitesses cylindriques galactocentriques, nous avons remarqué des différences significatives entre la cinématique des AO et des étoiles de champ. Premièrement, les AO ne sont pas alignées avec les structures de l'espace des phases récemment découvertes par Gaia Collaboration et al. (2018d). Deuxièmement, l'analyse de la relation age-vitesse des AO montre que les AO souffrent de la même quantité de chauffage cinématique que les étoiles de champ dans la direction radiale et azimutale mais que ce chauffage cinématique est beaucoup moins efficace sur les AO que sur les étoiles de champ dans la direction verticale. Cela peut être dû à un biais dans la distribution d'âge des amas, car la plupart d'entre eux sont dissous après des centaines de millions d'années. Cependant, les rencontres avec les Nuages Moléculaires Géants qui constituent le processus dominant de chauffage vertical pourraient également être plus efficaces pour perturber les amas que pour les chauffer cinématiquement.

A partir du calcul des orbites des amas, nous observons une dépendance claire de la hauteur maximale et de l'excentricité des orbites avec l'âge. Enfin, les caractéristiques orbitales de l'échantillon d'amas, telles que montrées par les variables d'action, suivent la distribution des étoiles de champ.

D.2 Limites

L'étude des AO souffre de multiples biais parmi lesquels certains sont inhérents à ces objets particuliers. Le plus important d'entre eux est dû à la dissolution de la plupart des amas. Les AO massifs, ou ceux situés dans les parties extérieures de la Galaxie ont plus de chances de survivre que les amas composés de seulement quelques dizaines d'étoiles. Il en résulte un biais de la distribution d'âge des AO avec très peu d'amas plus vieux qu'1 milliard d'années. Ce biais peut alors affecter nos résultats lorsque nous examinons la dépendance des paramètres des amas en fonction de leur âge.

La détermination des paramètres des jeunes amas soulève également quelques questions. Les AO très jeunes (quelques millions d'années) sont encore entourés par le gaz et la poussière constituant leur Nuage Moléculaire Géant progéniteur. En outre, ils sont encore situés dans les bras spiraux, proches du plan galactique. L'observation des

membres de ces amas souffre donc d'une forte extinction, surtout dans les longueurs d'onde optiques dans lesquelles *Gaia* observe. Comme les amas se forment généralement à l'intersection des structures filamentaires des Nuages Moléculaires Géants (Lada, 2010), les jeunes amas ont aussi généralement des voisins proches. Dans la Sec. 2.3, nous avons choisi d'interroger l'archive *Gaia* avec une recherche conique de 50 pc de rayon autour de chaque amas afin de détecter les membres de son halo. L'inconvenient inhérent à ce choix étant que nos requêtes contenaient dans certains cas plusieurs amas avec des paramètres très similaires que notre algorithme de partitionnement des données était incapable de séparer. Pour cette raison, nous avons dû écarter de notre échantillon les amas plus jeunes que 50 millions d'années afin de fournir des listes de membres fiables.

La détermination de la VR des jeunes étoiles est également un défi car les étoiles jeunes peuvent avoir des températures élevées ou être en rotation rapide, rendant leur VR moins fiable. C'est la raison pour laquelle la plupart des amas jeunes et distants ont été écartés de notre HQS dans la Sec. 4.3.

Une autre limite de notre méthodologie de partitionnement est une conséquence directe des importantes erreurs relatives de parallaxe mise en évidence dans la Sec. 3.2. Comme nous utilisons directement la parallaxe dans notre procédure de regroupement, l'identification des queues de marée alignées (ou légèrement inclinées) avec la direction de la ligne de visée comme celles de Ruprecht 147 et NGC 2632/Praesepe est difficile. Les queues de marées de Ruprecht 147 ont été découvertes par Yeh et al. (2019) en utilisant une méthode similaire à celle présentée dans la Sec. 2.2 et les queues de marées de Praesepe ont été identifiées par Röser and Schilbach (2019) grâce à la méthode des points convergents. Aucun de ces auteurs n'a utilisé directement la parallaxe dans leur procédure comme nous l'avons fait. Leur méthode convient à l'analyse d'amas proches mais elle est difficile à adapter à un large échantillon d'amas éloignés de plus de quelques centaines de pc. L'augmentation rapide de l'erreur relative de parallaxe avec la distance nous a décidé de conduire notre analyse morphologique en deux dimensions seulement. Cette décision permet d'inclure dans notre échantillon des amas ayant une distance allant jusqu'à 1.5 kpc alors que la limitation des erreurs de parallaxe ne permettrait de conduire une étude homogène que pour 14 amas dans un rayon de 250 pc du soleil.

D.3 Perspectives

Les futures DRs de *Gaia* seront de plus en plus précises car elles seront basées sur beaucoup plus de mois d'observation. Avec des données plus précises, l'étude de la structure tridimensionnelle des AO ne sera pas nécessairement limitée aux amas situés à moins de 250 pc. De plus, alors que les incertitudes sur les paramètres astrométriques des étoiles faibles sont encore élevées, une meilleure précision des mesures astrométriques permettra d'inclure des étoiles de magnitude plus faible que 18 mag dans les procédures de regroupement comme celle présentée dans la Sec. 2.3. L'étude de processus dynamiques comme la ségrégation de masse qui dépend du recensement des objets de faible masse sera améliorée par ces informations supplémentaires. Une meilleure détermination des paramètres astrométriques des AO permettra également à notre procédure de regroupe-

ment d'être efficace au-delà de la limite de 1,5 kpc que nous avons fixée.

Le nombre de mesures de VR est également sur le point d'augmenter de façon spectaculaire avec la publication de la version complète de *Gaia*-DR3 (prévue pour le premier semestre 2022) et la publication des VR pour $\sim$ 33 millions d'étoiles, ce qui en fera de loin le plus grand catalogue de VR. Un certain nombre d'études au sol, comme WEAVE ou 4MOST, consacreront également une partie de leur temps d'observation aux AO et fourniront à la communauté des mesures de VR. L'étude de la cinématique des AO fera un bond en avant avec cette quantité de données.

Notre procédure de regroupement décrite dans la Sec. 2.3.2 peut également être exploitée pour évaluer l'homogénéité chimique des AO. Dans Casamiquela, Tarricq et al. (en préparation), nous allons redéterminer les membres de UBC 274, étudier la morphologie et l'étendue de ses queues de marée et évaluer l'homogénéité chimique de l'amas grâce aux observations spectroscopiques effectuées en janvier 2021 au télescope Magellan (Observatoire de Las Campanas). Notre catalogue de membres d'amas sera également utilisé pour l'étude des AO dans le cadre du relevé dédié à l'archéologie galactique du projet WEAVE. Ce catalogue est particulièrement adapté car le champ de vision de WEAVE est très grand et les restrictions inter-fibres du spectrographe multi-objets ont des limites d'attribution dans les champs encombrés, rendant WEAVE incapable d'observer les régions centrales des amas. En outre, WEAVE sera en mesure de confirmer la détection des queues de marée et des halos stellaires présentée dans cette thèse en utilisant la richesse de ses mesures de VR et d'abondances chimiques. Tandis que d'autres queues de marée seront également progressivement découvertes, une analyse tridimensionnelle détaillée de leurs membres sera nécessaire pour vérifier si ces queues sont toujours orientées dans la direction de la rotation Galactique ou si cette orientation dépend de l'âge des amas.

Les amas dont la structure a été étudiée au cours de cette thèse étant principalement des amas vieux, une étude similaire d'amas âgés de moins de 50 millions d'années permettrait également d'examiner les différences entre ces deux échantillons d'amas. En outre, notre étude a montré des différences importantes entre la cinématique des étoiles de champs et des AO qui restent à confirmer. Ces différences de dispersion de vitesse verticale interrogent sur l'influence réelle des Nuages Moléculaires Géants sur les AO : les Nuages Moléculaires Géants sont-ils plus efficace pour dissoudre les AO que pour les disperser verticalement ?

Enfin, nous avons mis à la disposition de la communauté trois catalogues : le plus grand catalogue de VR pour AO, des listes d'appartenance étendues à de grands rayons pour 389 amas situés à moins d'1,5 kpc et plus vieux que 50 millions d'années ainsi qu'un catalogue de paramètres structurels d'amas basé sur l'analyse d'appartenance mentionnée ci-dessus. Ces catalogues seront utiles dans l'effort global destiné à mieux contraindre les modèles théoriques d'évolution des amas et de la dynamique galactique.