Open clusters and the structure of the Milky Way

Author: Irene Quadrino Lodoso
Advisor: Friedrich Anders
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain**.

Abstract: Our galaxy contains a large number of stellar open clusters still to be identified and characterized, but thanks to the new *Gaia* DR2 release, it is possible to identify new ones that are unknown to date and discard earlier reported clusters that turn out to be spurious. The aim of this work is to do a statistical analysis of the spatial and the age distribution as well as the completeness, using the most recent and comprehensive sample of open clusters available today from the *Gaia* satellite. Using these data and comparing them with old catalogues, we redetermine fundamental parameters of the cluster age function: $\log T_{\mathrm{break}} = 8.52 \pm 0.18$, $\alpha_1 = -0.61 \pm 0.28$, $\alpha_2 = -1.82 \pm 0.15$ for the broken power-law fit and $\log T_* = 8.89 \pm 0.15$, $\alpha_T = -0.55 \pm 0.10$ for the Schechter fit.

I. INTRODUCTION

Since time immemorial, humans have wondered their place in the Universe, and more concretely in the Galaxy. According to star formation and evolution theories, most stars are born and live for part of their lives in groups. We distinguish two types: clusters and associations. Depending on the conditions in which the birth of the group occurs, stars can be linked to each other by gravitational potential and constitute clusters or form scattered aggregations, not gravitationally linked, called stellar associations.

There are two types of clusters: Globular Clusters, and Open Clusters (OCs). The main differences between them are shape, location and the number and age of the stars they contain. In order to understand the structure of the Milky Way in this work, we focus on OCs that are very useful tracers of the Galactic disc structure of the Galaxy. OCs are young groups of stars mostly located in the Galactic disc, held together by mutual gravitation, and with similar properties such as position, age, chemical composition and proper motion.

Since Hipparchus of Nicea made the first stellar catalogue, thanks to new technologies we have progressed in mapping our galaxy. However, the OCs census remains extremely incomplete. A new era in the cartography of the Milky Way has been introduced with the *Gaia* mission [7]. The *Gaia* satellite has two telescopes that focus light to the same focal plane made up of 106 charge-coupled devices (CCD) detectors, with a total of 938 million pixels [7]. The second release of data from the *Gaia* mission (DR2[8]) represents an important advance compared to *Gaia* DR1[7], presenting enhanced data for more than 1.3 billion sources. As a result of this improvement, hundreds of new clusters have been detected (e.g. [4], [5], [6]) and a lot of false clusters have been discarded thus providing an update of the old catalogues.

The basic purpose of the present study is to reanalyse the distribution of open star clusters in position-age space, thus improving on previous works relying on data from the pre-*Gaia* era. To this end, we use established data analysis concepts from the OC literature, in particular the cluster age function. Firstly, we have selected pre-*Gaia* catalogues to compare with the new *Gaia* DR2 census (post-*Gaia* DR2). After examining the literature, we have chosen some plots that better represent the topics we want to deal with throughout the study. Finally, through programming in Python, we have reproduced them with the data of the catalogues that we have chosen.

The structure of this work is the following: In section II we present a description of the input data, corresponding to the recent OCs *Gaia* based catalogue published by [3] and the literature we use to compare with it. In section III we examine the spatial distribution of OCs. Section IV describes the age distribution: the cluster age function correction is computed in subsection A, and completeness is treated in subsection B. Finally, our conclusions are presented in section V.

II. DATA

In this study, we use the OCs catalogue from *Gaia* DR2 published by [3]. This catalogue represents the largest sample of cluster parameters performed with *Gaia* data, obtained from the sum of various catalogues. Most of the entries of this catalogue were taken from [2] and [6].

The recent OCs *Gaia* based catalogue [3] contains several parameters (age, distance modulus, and extinction) and in this study we use age and distance. The cluster parameters reported in this catalogue have been obtained from the OC Color-Magnitude Diagrams (CMDs) with an artificial neural network (ANN) trained on the CMDs of well-characterised clusters. An artificial neural network is a computational model that consists of a set of units, called neurons, that receive inputs and deliver outputs based on their activation functions. In this case, the ANN is used as a regressor. In total, their sample contains 2017 clusters, out of which 1867 have sufficiently good CMDs to derive reliable parameters.

In order to compare the most recent OCs sample of *Gaia* DR2 with the literature, we use different catalogues. We work with the pre-*Gaia* MWSC (Milky Way Star Clusters catalogue) data [10] as well as the later additions from [17] and [18] (which could not be confirmed by *Gaia*, see [3]). We also use the data from [2] and [6] catalogues that are *Gaia* DR2 derived but do not have homogeneous parameters and do not contain ages.

* Electronic address: iquadrino@hotmail.com

III. SPATIAL DISTRIBUTION

The spatial distribution of clusters provides us with very valuable information to determine the structure of the galactic disc. Stars tend to form in the spiral arms of the Galaxy and therefore the younger clusters draw the spiral structure.

With the purpose of estimating the spatial completeness of the samples, we studied the distribution of OCs in the X-Y plane. In Figure 1 we represent the data obtained from [5] and [6]. We have not made a distinction between young or old clusters, we have represented all together. We observe that the distribution of OCs is not uniform. OCs are not evenly distributed, we find regions where there is a greater concentration of objects and other zones where there is a deficiency of OCs. The spiral arms would be located at the highest concentration of young OCs, therefore we confirm their location with the new data sample from *Gaia* DR2. We also confirm that beyond 8 kpc there are few OCs detected.

Now we focus on a smaller scale. For comparison, we illustrate in Figure 2 the results obtained with the pre-*Gaia* data MWSC [10], the list of known clusters [2] together with the list of reliable candidates from [6], and neural-network parameters [3]. We appreciate some important differences between the three plots. If we set the completeness limit commonly assumed for MWSC i.e. 1.8 kpc from the Sun, we notice that the left diagram seems to be more complete than the other two. We observe a major density of objects in the MWSC sample and a more or less uniform spatial distribution. On the other hand [2] plus [6] plot looks more like the representation of the recent sample of *Gaia* DR2 catalogue.

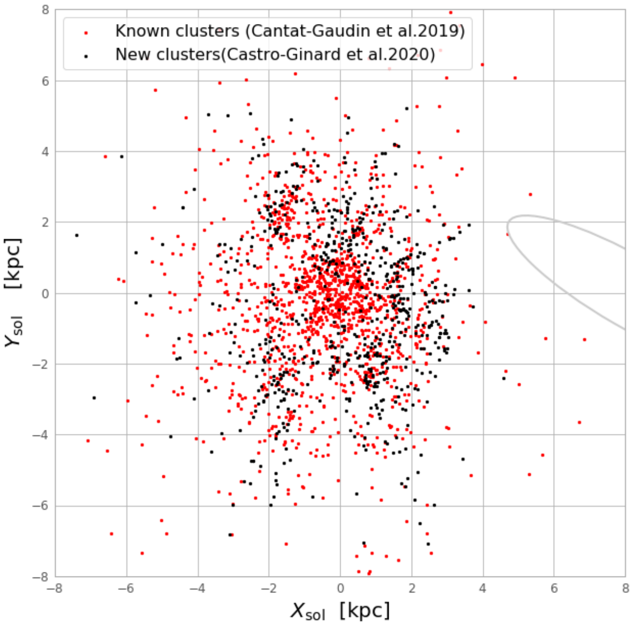

FIG. 1: Distribution of OCs in heliocentric Cartesian coordinates. “Known clusters” are the OCs from [5] and “New clusters” from [6]. The Sun is located at (0,0) and the ellipse gives the approximate extent and orientation of the Galactic bar.

If we analyze the region 1 kpc from the Sun both plots have a similar distribution; we found that the density in this region is higher than the rest of the map. Nevertheless, we observe that in the [3] sample there are more scattered clusters than in the middle plot and the distribution is a little bit different.
As we have discussed in Figure 1, we notice that for middle and right panels the spiral arms of the Galaxy can be more clearly distinguished than in MWSC where the arms are poorly defined.

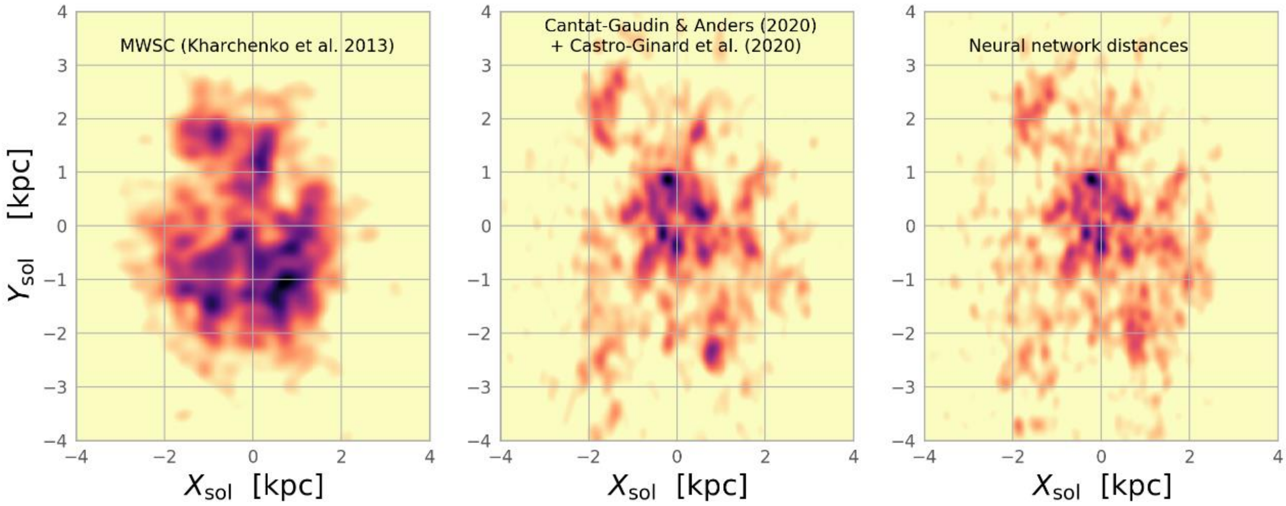

FIG. 2: X-Y maps of OCs (pre-*Gaia* vs post-*Gaia* DR2). Left: MWSC [10]. Middle: [2] plus [6]. Right: Post-*Gaia* DR2 [3]. Middle and right plots show the same objects (except for the 150 that don't have ANN parameters). The difference between left and middle is driven by largely different samples (MWSC including lots of noisy and dubious clusters).

IV. AGE DISTRIBUTION

A. Completeness

Figure 3 is related to Figure 4, both give us information about the completeness of the samples.

In Figure 3 we compare the distribution of OCs in a d_{xy} (projected distance from the Sun onto the Galactic disc plane) vs $\log T$ plot using pre-*Gaia* (MWSC) and post-*Gaia* DR2 census [3]. We note that in the case of the pre-*Gaia* data, more distant groups appear for $\log T \cong 6$ and $\log T \cong 10$, however, in the representation of the most recent data these groups are discarded.

For the *Gaia* data, the youngest detected clusters tend to be at short distances while the oldest ones are the most distant. The reason why we hardly find clusters at far distances at an early age is due to the fact that their detection is complicated because they often reside in molecular clouds and the interstellar dust makes the detection difficult. However, we observe that there are objects located at a greater distance for higher ages. The relative scarcity of old OCs is due to the fact that these objects have a low stellar density, therefore their structure is vulnerable to the effects caused by the heterogeneities of the gravitational field in the galactic disc. Along their orbits in the plane of the galaxy, these clusters are suffering encounters with giant molecular clouds and experience various perturbations in colliding with the spiral arms. In this way, OCs expand and disintegrate over time.

However, OCs are not all concentrated in the Galactic plane and there are some old OCs located at high altitudes.

If we analyse the catalogue of *Gaia* DR2 we can detect some old OCs at $d_{xy} > 6$ kpc that are composed by stars of intermediate-mass and high metallicity in one of the late phases of their evolution, called red clump phase. We can see that both samples are more complete at around 1.8 kpc.

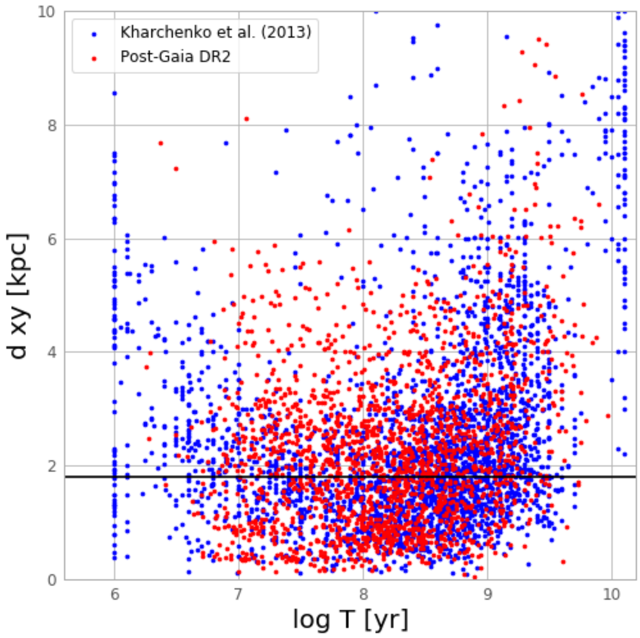

FIG. 3: Disc-projected distances of OCs against age. Blue dots represent members of [10] (pre-*Gaia*) and members of [3] (post-*Gaia* DR2) are shown in red. The black line indicates the general completeness limit, situated at 1.8 kpc [15].

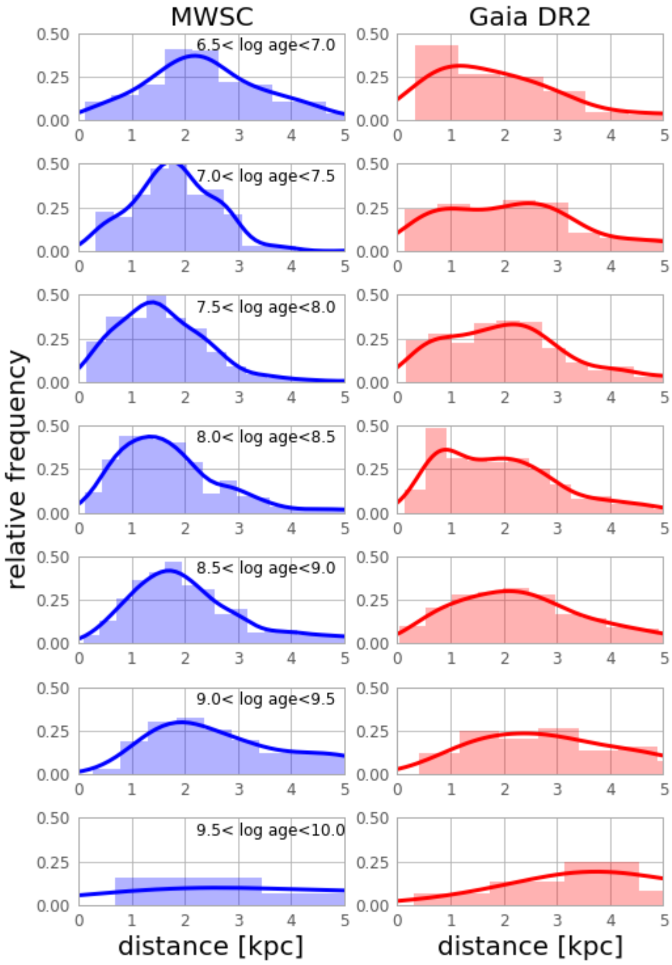

FIG. 4: Number of OCs against distance in log age bins. We use pre-*Gaia* data from [10] (blue histograms) and post-*Gaia* DR2 from [3] (red histograms). In the MWSC sample, the histograms of the number of OCs as a function of distance has a Gaussian profile, while the red histograms follow an exponential profile.

In Figure 4, we see that *Gaia* DR2 histograms deviate significantly from MWSC histograms. In the MWSC sample, the peak of the distribution is higher than in the *Gaia* sample because of the presence of lots of non-existing putatively old clusters. In the MWSC sample, the distribution presents a Gaussian profile except for the older groups. However, in the *Gaia* DR2 sample, the curve flattens and indeed we see that it can be approximated by an exponential disc profile.

As we have discussed in Figure 3, in the histograms for older ages the number of OCs is much smaller than for intermediate ages. Around a distance of 1.8 kpc both samples are probably complete and the censuses are significantly incomplete for early ages.

A proper completeness analysis is deferred to a future publication [1].

B. Cluster age function correction

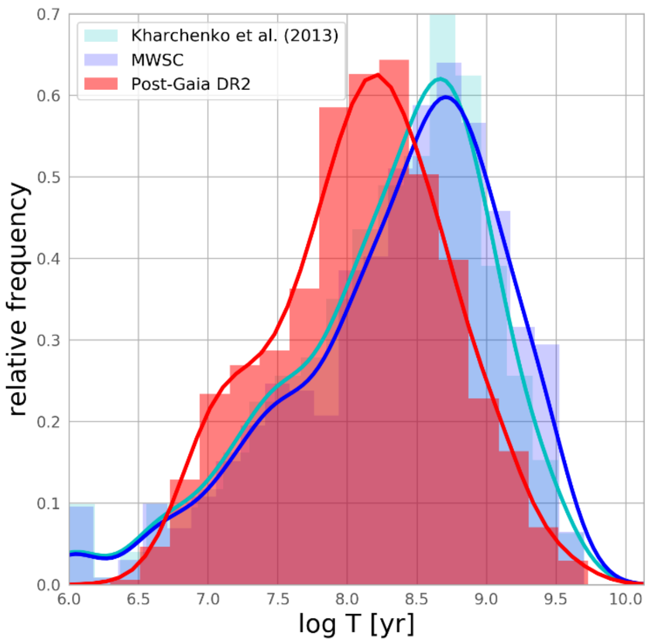

FIG. 5: Age distribution for OCs within 1.8 kpc. Members of MWSC are shown in blue and cyan. The red histogram corresponds to the data from [3].

Figure 5 shows the relative frequency as a function of log age. The relative frequency gives us the probability that the clusters are in certain age bins. A higher relative frequency implies a greater probability of finding the clusters in those ranges of age.

Analyzing the plot, we observe that for the MWSC sample there is a higher number of old objects because of the presence of lots of non-existing putatively old clusters. Likewise, we see that for *Gaia* data in the logT <7 regions, our sample is significantly incomplete, since many of the OCs may still be enveloped by gas clouds in star formation regions. Also, for logT> 9.5 we observe that there is a notable deficiency in the number of objects. A possible explanation for the lack of very old objects is that most clusters are dissolved before they can reach older ages. Another reason for the incompleteness in the *Gaia* DR2 catalogue is due to the extinction. This phenomenon is produced by the scattering and absorption of light by dust particles and causes a weakening of the apparent luminosity of the stars. Extinction is more accentuated for short wavelengths, so an interstellar redness of the light is observed more important the more amount of interstellar material passes through. Thus, the most distant stars will be more affected by this phenomenon than not the closest. In summary, the data completeness is not uniform, the statistics of very young OCs is insufficiently known.

In [13] studied how clusters are destroyed considering four effects: tidal stripping, stellar evolution, spiral arm shocks and encounters with giant molecular clouds. The dissolution time of clusters depends on the mass of the OCs which can be difficult to determine. Until estimates become available from *Gaia* data, we can work with the cluster age function (CAF) in order to study the formation and destruction of OCs.

The number of clusters in the younger bins gives us the formation rate and the slope of the CAF gives us a typical destruction time scale.

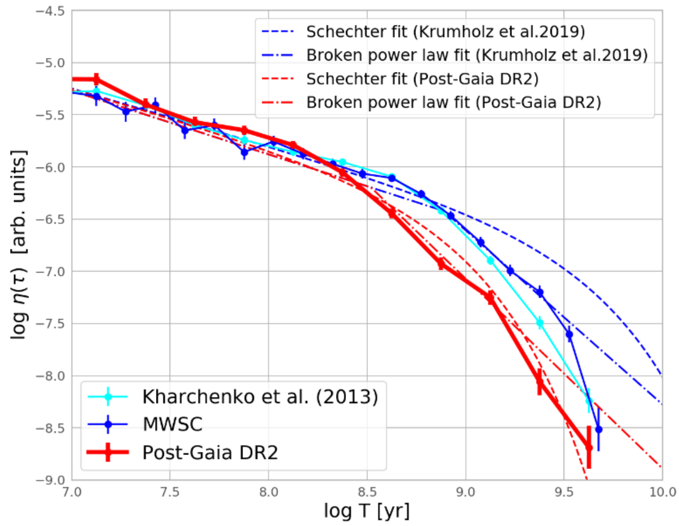

FIG. 6: Comparison of cluster ages functions for star clusters (pre-*Gaia* vs. post-*Gaia* DR2) within 1.8 kpc. Error bars show the Poisson uncertainty of the bins. For reference, we also show Schechter function and a broken-power law fit.

In order to compute the CAF for the 3 samples we follow the method described in [15]:

$$\eta(t) = \frac{\log e}{t} \cdot \nu(t) \tag{1}$$

With

$$\nu(t) = \frac{1}{S(t)} \cdot \frac{dN(t)}{dt} \tag{2}$$

$S(t) \equiv S_0 \equiv$ constant and Equation (2) can be written in logarithmic form, more convenient. $\Delta_k N$ is the distribution of cluster numbers and $\Delta_k \log T$ represents the age interval.

$$\nu(t) = \frac{1}{S_0} \cdot \frac{\Delta_k N}{\Delta_k \log t} \tag{3}$$

For clusters in galaxy discs, we can fit the CAF as a broken power- law $\frac{dN}{dT} \propto T^{\alpha_T}$[12], where T is the age of the cluster and α_T represents an index of cluster distribution. In this case, α_T will have different values depending on the age range. In Figure 3 we show the resulting functions for $\log T_{break} = 8.89$, $\alpha_1 = -0.61$ for $\log T < \log T_{break}$ and $\alpha_2 = -1.67$ for $\log T > \log T_{break}$ (obtained by [12]).

Schechter function is computed as $\frac{dN}{dT} \propto T^{\alpha_T} e^{\frac{-T}{T_*}}$ with $\alpha_T = -0.55$ and $\log T_* = 9.59$ [12].
We observe that CAFs, despite following the same profile are quite different. The fits were only performed with clusters older than logT=7 because interstellar extinction due to dust in the Galactic plane affects the completeness of the youngest cluster sample. The best-fit parameters we obtain with the *Gaia* clusters are significantly different from those reported [12]. We obtain: $\log T_{break} = 8.52 \pm 0.18$,

α_1 = -0.61 $\pm$ 0.28, α_2 = -1.82 $\pm$ 0.15 for the broken-power law fit, and logT* = 8.89 $\pm$ 0.15, α_T = -0.55 $\pm$ 0.10 for the Schechter fit.

V. CONCLUSIONS

In this work, we have studied and analyzed the spatial and age distributions and the completeness of the most recent sample of OCs based on *Gaia* DR2. Our main conclusions are summarized as follows:

● The result obtained from the analysis of the spatial distribution is that young OCs effectively distribute in a spiral pattern. With the new data, we observe that OCs extend to further distances, show more fine structure, and are not as concentrated as in the pre-*Gaia* sample.

● Incompleteness is a difficult problem: the probability of finding an OC depends on various parameters as distance, age, mass, metallicity. The OCs sample by [3] suffers from incompleteness as any other survey. As we have said in the study, we have little statistics for very young clusters. This fact means that young clusters may be unnoticed in the optic due to high extinction, and therefore the sample is unlikely to be 100% complete. We have shown that the sample is probably complete within 1.8 kpc, although this claim might be challenged by future detections. We discuss the completeness in a more quantitative way in a future publication [1].

● Following the method described in [15], we computed the CAF for the pre-*Gaia* and post-*Gaia* DR2 samples. After examining Figure 3 we confirm the results of [12], Milky Way CAF is best fitted by a Schechter function or a broken-power law. However, it is necessary to review and correct the fit parameters. For the *Gaia* data, we obtain best-fit values of $\log T_{\mathrm{break}}$ = 8.52 $\pm$ 0.18, α_1 = -0.61 $\pm$ 0.28, α_2 = -1.82 $\pm$ 0.15 for the broken power-law fit, and logT* = 8.89 $\pm$ 0.15, α_T = -0.55 $\pm$ 0.10 for the Schechter fit. These results indicate that the Milky Way contains a smaller proportion of old clusters than previously thought. Although the profiles are similar, there is a difference in the representation of the CAFs. We have to consider that the number of objects in the different samples is not the same and this fact affects the representation of the CAFs and the completeness of the samples. However old models like [14], continue to adjust well to the new data.

Acknowledgements

I wish to express my deepest gratitude to my advisor Dr. Friedrich Anders for the knowledge he has transmitted to me, for his guidance and supervision. Also, I wish to thank Dr. Carme Jordi and Dr. Tristan Cantat-Gaudin. Finally, I would like to thank my whole family and friends for their support and encouragement.

[1] Anders, F., Cantat-Gaudin, T., Quadrino-Lodoso, I., et al.2020. A&A, submitted, arXiv:2006.01690
[2] Cantat-Gaudin, T.& Anders,F.,2020, A&A, 633, A99
[3] Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, accepted, arXiv:2004.0727
[4] Cantat-Gaudin, T., Jordi,C., Vallenari, A., et al. 2018, A&A, 618, A93
[5] Cantat-Gaudin, T., Krone-Martins, A., Sedaghat, N., et al. 2019, A&A, 624, A126
[6] Castro-Ginard, A., Jordi, C., Luri, X., et al. 2020, A&A, 635, A45
[7] Gaia Collaboration, 2016, A&A, 595, A1
[8] Gaia Collaboration, 2018, A&A, 616, A1
[9] Joshi, Y.C., Dambis, A. K., Pandey, A. K., Joshi, S. 2016, A&A, 593, A116
[10] Kharchenko, N. V., Piskunov, A. E., Roeser, S., Schilbach, E., Scholz, R.D. 2013, A&A, 558, A53
[11] Kharchenko, N. V, Piskunov, A. E., Schilbach, E., Röser,S., Scholz, R.D. 2016, A&A, 585, A101
[12] Krumholz, M.R., McKee, C.F. & Bland-Hawthorn, J. 2019, ARA&A, 57, 227
[13] Lamers, H. J. G. L. & Gieles, M. 2006, A&A, 455, L17
[14] Lamers, H. J. G. L., Gieles, M., Bastian, N., et al. 2005, A&A, 441, 117
[15] Piskunov, A.E., Just, A., Kharchenko, N.V., et al. 2018, A&A, 614, A22
[16] Piskunov, A.E., Kharchenko, N.V., Röser, S., Schilbach, E., & Scholz, R.D. 2016, A&A, 445, 545
[17] Schmeja, S., Kharchenko, N.V., Piskunov, A.E., Röser, S., Schilbach, E., Froebrich D., Scholz, R.D. 2014, A&A, 568, A51
[18] Scholz, R.D., Kharchenko, N.V., Piskunov, A.E., Röser, S., Schilbach, E. 2015, A&A, 581, A39