Physical parameters of Open Clusters with Gaia DR2

Author: Judith Santos Torres*.
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Advisor: Carme Jordi Nebot

Abstract: The aim of this work was to determine physical parameters (mean distances, interstellar absorption and metallicity) of the known Open Clusters using recently published data: (a) the *Gaia* Data Release 2 for the identification of the members of the group, and (b) the individual stellar information given in *StarHorse* by combining *Gaia* and additional photometry. First, we verified the reliability of *StarHorse* determinations searching for possible systematic errors across the HR diagram. 59 well populated clusters were selected for this statistical analysis, which revealed a trend in the derivation of absorption, metallicity and distance as a function of color. Finally, we calculated the physical parameters of a set of 1340 clusters for which *StarHorse* provides data, and compared them with previous values in the literature.

I. INTRODUCTION

Many stars are born in groups and subsequently they partially or totally disperse to become individual field stars. Before they dissolve, those groups of stars are assembled together in what we define as clusters. Star clusters are large structures of stellar congregations held together by the mutual gravitational attraction of its members, which are physically related through a common origin. Two different types of clusters can be distinguished: Globular Clusters, which are tight groups of hundreds to millions of old stars, and Open Clusters (OCs), which, on the contrary, are more loosely bound and contain a few thousand young stars. This project focuses on the latter ones. OCs member stars are thought to be formed in the same molecular cloud, have common motions and therefore, are at the same distance, have the same age, initial chemical composition and interestellar absorption. They become a valuable astrophysics laboratory and are the key objects to study and trace the stellar and Galactic disc structure, formation and evolution.

The recent results from the *Gaia* mission have allowed to confirm the discovery of a new population of OCs. *Gaia* is a 5 year ESA mission, launched in December 2013, aiming to create the most accurate three-dimensional map of the Milky Way so far, by measuring the positions and motions of more than one billion stars in our Galaxy. The first data release (DR1, 2016 [1]) provides astrometric information of two million stars, whereas the second data release (DR2, 2018 [2]), with significantly improved data precision, widely increases that number, reaching over 1.6 billions of stars catalogued. After the publication of *Gaia* DR1 & DR2, the existence of the previous catalogued clusters has been reanalyzed and new clusters have been discovered. Besides, *Gaia* is gradually unveiling a better understanding of our home Galaxy and its neighbourhood.

StarHorse is a Bayesian parameter estimation code that compares a set of observed quantities, mainly parallaxes and multi-band photometry (*Gaia* DR2, 2MASS, Pan-STARRS1 and A11WISE [2–4]), to stellar evolutionary models given a number of priors that include the stellar initial mass function, density laws, metallicity and age priors for the main components of the Milky Way [5]. These results are obtained in the derivation of the most probable distance, metallicity an absorption of each individual star, although not all the photometry passbands are available on every star.

The goal of our work was to contribute to the characterisation of physical properties of as much Open Clusters as possible. We first examined the quality of the *StarHorse* determinations in order to check their consistency accross the Hertzsprung Russell Diagram (HRD). The aim here was to estimate the average value and its corresponding standard deviation of parameters such as the absorption (AV), the distance (Dist) and the metallicity (Met), from the *StarHorse* data. After applying the output flags determined by the *StarHorse* code that helped us decide the best subset to use for our particular case, we were able to proceed to our calculations. Afterwards, we restricted our analysis to the clusters with higher number of verified members to classify them according to their position in the HRD. In the end, we plotted all the results concluding that the *StarHorse* code does have some systematic trends. We also compared our derived parameters with those in the literature [12]. To finish, we gathered all our computations in a unique list for the candidate clusters.

The paper is structured as follows: In section II, we present a description of the data sources that have been used, the selection criteria adopted to choose the right star candidates and the considered parameters. Section III provides the information about the computer programming language chosen to develop all the calculations and the justified data restriction conditions. The results

---

*Electronic address: judith0606@gmail.com

are shown in section IV as plots of the three evaluated parameters of the clusters and finally, in section V, we discuss our conclusions from the whole analysis.

II. DATA COLLECTION

Our process of collecting information started linking together the data of the clusters in the *Gaia* Archive. We compiled a list of all known clusters with their star members from various papers published during the past two years [5–11].

The next step consisted in cross-matching star by star, with the *StarHorse* results (`gaia@aip.de`), to create a new database table. The final inventory amounts to 375 514 stars in 1347 clusters. The manipulation of the data tables and posterior plotting, has been carried out using the astronomical java software TOPCAT (Tool for OPerations on Catalogues And Tables [13]), an interactive graphical viewer and editor for tabular data providing a wide range of facilities for the analysis and usage of source catalogues and other tables.

At that point, we made use of the quality flags of the *StarHorse* output code. For some cases, the code was not able to reach a convergence and, for others, it was of bad quality, meaning that the several observational data was not coherent enough and did not match a theoretical prediction. In order to apply a particular quality cut as a guidance to discern the stars with the most accurate measurements, we took into consideration the flags SH_GAIAFLAG and the SH_OUTFLAG. On one hand the SH_GAIAFLAG describes astrometric and photometric quality of the *Gaia* DR2 for each star in a three digit flag. On the other hand, the SH_OUTFLAG consists on 5 digits, and informs about the quality of the *StarHorse* output parameters [5]. To guarantee the selection of the top-level quality stellar data, we imposed that both flags had to be "000" and "00000" respectively. The essence of this definition is that posterior mean statistics of the star parameters where any digit of the flags equals to "1", should be treated mindfully as their combination yields, as we already mentioned, unphysical results.

To develop our calculations, we selected some of the *StarHorse* parameters of our interest: the line-of-sight extinction at 5420 nm parameter (AV), the distance (Dist) and the metallicity (Met). Each of these values is in different percentiles so one can figure out the distribution function of the possible values of the parameter. For each parameter, we chose the 16th, the 50th and the 84th percentiles to characterise them in our flag-cleaned list of stars. Furthermore, we added two more columns to the list for each member, one including the numerical values of the *StarHorse* absolute magnitude (MG_0) and another with *StarHorse* dereddened $G_{BP} - G_{RP}$ colour ($(BP - RP)_0$). Moreover, for the recognition of the stars, there was a column with a unique source identifier (source_id), and another one with the cluster's common name that each star belongs to (Cluster).

III. DATA PROCESSING AND ANALYSIS

A. Programming method

The list of stars we generated was quite large so we had to program a code to execute all the calculations and apply the corresponding conditions of the flags mentioned above. To develop all the operations with the data, we used Fortran as the computer programming language.

The code allowed us to read the compiled table of non-filtered data extracted from the joined information of the OCs papers and the cross-match made with the *StarHorse*. In addition, we imposed the conditions of the flags and after that, only selecting the valid stars, we calculated the mean values and the standard deviation of all the parameters for each OC. The results were written in an output file in which we could also see explicitly the number of total stars of each cluster and the number of valid stars considered in the statistics. Doing so, we could get an approximate idea about the quantity of dismissed stars. This file had 1340 rows, so 7 clusters failed to pass our requirements.

We created another file to store the differences between the averages of the cluster and the individual values of each star member, expressed as follows for the case of the distance 50th percentile: *meanDist50-Dist50**; where we named *meanDist50* to the mean value of the distance parameter for the cluster and *Dist50** to the value of the distance for each star belonging to the cluster. With that, we could notice that the number of stars decreased to 219 451 members due to the omission of the clusters previously referred. The mean values of these differences should be 0 or close to 0.

B. Limiting data

So far, we had various clusters with different quantities of accepted stars, some of them had 50 verified stars and some other had hundreds. We decided to only use the clusters with the highest number of valid stars to analyse the systematic differences thereafter. The lower limit we set was 600 stars and we got 59 clusters. We examined in detail the TOPCAT plots of the earlier mentioned differences of distances, absorptions and metallicities versus the intrinsic color of every star and we observed a peculiarity in the $1 < (BP - RP)_0 < 2$ range of the distance plot. For some of the clusters, the stars from the main sequence were slightly dispersed while the giant stars showed greater dispersions. In consequence, the clusters with more significant dispersions had higher uncertainities in their parallaxes.

We decided to change a little bit our approach and instead of analysing star by star within each cluster, we decided to group the member stars according to their spectral types and analyze the differences of each region with respect to the mean of the cluster:

* **O-B:** $(BP - RP)_0 \leq 0.1$

* **A-F:** $0.1 < (BP - RP)_0 < 0.8$

* **G-K:** $0.8 \leq (BP - RP)_0 \leq 1.7$; $MG_0 > 4.6$

* **SUB-GIANT:** $0.8 \leq (BP - RP)_0 \leq 1.7$; $2 \leq MG_0 \leq 4.6$

* **GIANT:** $0.8 \leq (BP - RP)_0 \leq 1.7$; $MG_0 < 2$

* **TOP-GIANT:** $(BP - RP)_0 > 1.7$; $MG_0 < 2$

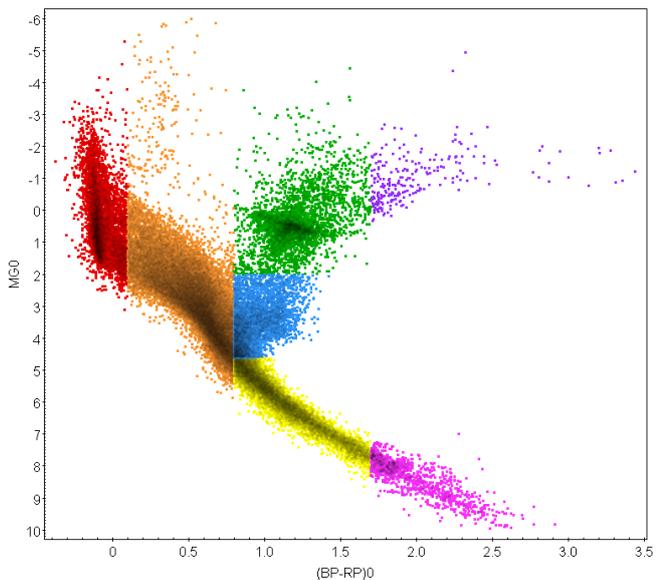* **M:** $(BP - RP)_0 > 1.7$; $MG_0 > 2$

FIG. 1: HRD divided in different coloured regions. The labels assigned to each group from left to right would be: O-B stars (red), A-F stars (orange), G-K (yellow), SUBGIANT (blue), GIANT (green), TOP-GIANT (up, purple) and M stars (down, pink).

Fig. 1 shows the 7 different areas in the HRD defined according to the conditions written above. Each region has been delimited in an approximate way and the denomination of all the families has been assigned concerning the spectral type of the stars included in each area.

This different categorisation and focusing allowed us to recalculate the new averages and the standard deviations of the stars from the 59 candidates in regards to the amount of members encompassed in each of the seven separate regions. The new file showed the statistics and the quantity of stars in every region per cluster. After that, we estimated, once again and in the same way we did before, the difference between the mean value of the cluster and each star in it, for all the parameters.

Referring to the parallax uncertainities of the more dispersed clusters mentioned before, we thought that another subdivision was required in the SUBGIANT and the GIANT regions. Our list of 59 clusters was re-scanned again dividing the mean standard deviations of the parallaxes of their members into their parallaxes. This time, we gathered all the clusters that presented a parallax error higher than the 10% in a group that we classified as 'distant clusters', otherwise, they were considered 'nearby clusters'. Both subsets were taken into account in the clusters that had GIANT and SUB-GIANT members to improve our visibility in the contrast between the *StarHorse* results and our calculations. Furthermore, they allowed us to confirm that the previously revealed clusters with the singularities in their distance vs. color diagrams, were effectively, distant clusters, with $> 10\%$ error parallax. Although, we chose the separation to be approximately 10%, we must say that all the clusters that showed a peculiar behaviour in their distances, were$> 10\%$ or even higher.

IV.   RESULTS

In order to verify the quality of the determinations of *StarHorse*, we have tested the mean values of AV50, Met50 and Dist50 to observe the systematic trends of each HRD area previously defined. We fixed a common zero that could be used to compare one cluster to another. The chosen reference was the A-F group due to its existence in the totality of the 59 clusters. Other than that, would have been wrong because no other HRD area was comprised in the entirety of our cluster set. Introducing a balance between the mean studied parameters for each family and the mean parameter for the A-F's in each case, allowed us to observe the deviations of every region from the zero. The error of each value has been calculated regarding the same chosen group of reference.

Fig. 2 shows the differences of the absorption, the distance and the metallicity of the 50th percentile, and the corresponding error bars for each HRD region. If the *StarHorse* determinations had no systematics, all the mean values would be mostly around zero and there would be no dispersion in the regions, contrary as what is seen in the three plots above.

At first glance, for the absorption, we observe that the dispersions from the zero of each region of the HRD are quite significant, the general trend are the negative mean values, except for the O-Bs. The metallicities show the opposite behaviour, the spreading is mostly to positive mean values, except for the O-B and the TOP-GIANT groups, which meet the expectations correctly. Last but not least important, the distance, which displays the distributions of most of the regions systematically to negative values apart from the O-Bs and both GIANT subdivisions, thus, these exceptions fulfill the predictions properly.
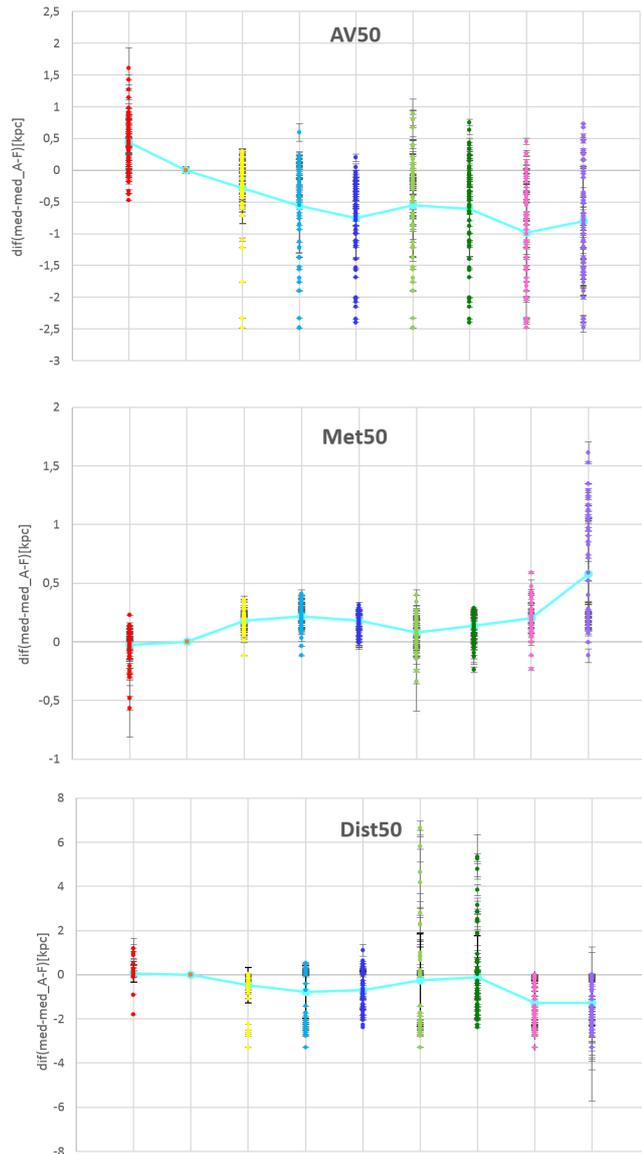
FIG. 2: Distributions of AV50, Met50 and Dist50 average differences (and error bars) for every HRD region for the 59 clusters with more than 600 members. The total average of all clusters (also with error bars), of each of the regions is shown in turquoise. The color code is consistent with the one in *Fig. 1*. For the split groups, the SUBGIANTS and the GIANTS, we chose light tones of blue and green for the nearby clusters and darker tones of the same colors for the distant ones. We have arranged the set of regions in the following order, from left to right: O-B, A-F, G-K, nearby SUBGIANT, distant SUBGIANT, nearby GIANT, distant GIANT, M and TOP-GIANT.

In short, the HRD inter-regions' differences are in the following ranges: $-1 < AV50 < 0.5$, $0 < Met50 < 0.5$ and $-1 < Dist50 < 0$ [kpc].

To complement our whole analysis, we compared the obtained mean values of the distance and absorption, for
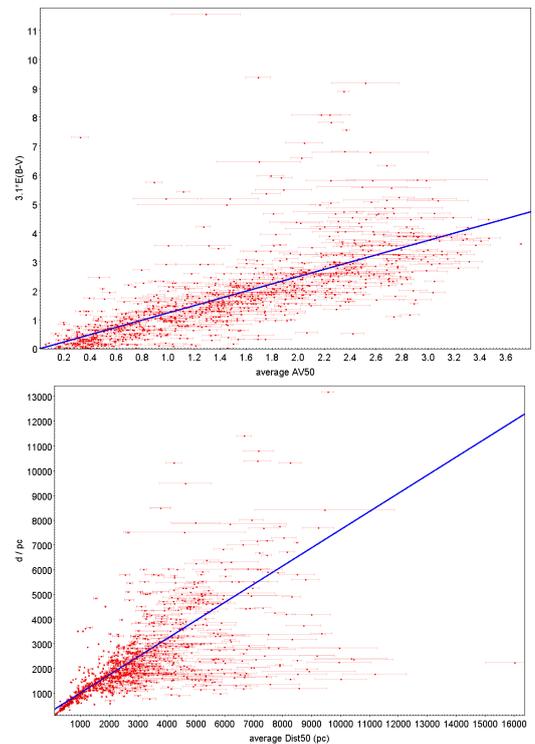

FIG. 3: Absorption-absorption and distance-distance (red) from the values on the literature [12] and *StarHorse* data with linear fit (blue) and error bars.

our set of 1340 clusters, with the ones in the literature [12]. It is worth mentioning that the star member determination of each cluster is more accurate with *Gaia* than it is with the chosen literature [12]. Besides, the *StarHorse* determinations are made star by star, disregarding their belonging to the same cluster, whereas [12] fits isochrones taking into consideration a more general vision of the cluster. In addition to that, [12] doesn't provide data of metallicities nor errors, so we did not make any comparison with this parameter. It is also fair to notice that [12] calculates the color excess, E(B-V) instead of the AV50, which equals to 3.1 times the E(B-V).

As can be noticed in *Fig. 3*, there is a visible linearity between the two sources of data. The coefficients of the linear fit are: $y = 1.253x - 0.017$ (correlation = 0.715), for the absorption and $y = 0.734x + 2082.251$ (correlation = 0.718), for the distance measured in *pc*. We assume that most of the clusters fit in the linear correlation, since their error bars, that we calculated, are favorable enough. Fewer other clusters, are clearly differing, which could be due to *StarHorse* or [12] inaccuracies. In fact, [6] resolved that most of the parallaxes from *Gaia* mismatch with the [12] data.

While the absorption comparison shows a relatively uniform linear tendency, the distance presents a better adjustment to the linear fit for the nearby clusters than for the distant ones.

V. CONCLUSIONS

We cross-matched lists of stars in clusters from *Gaia* DR2 with the results of *StarHorse* code with the purpose of deriving mean extinction, distances and metallicity parameters for each cluster. For that, we imposed some requirements of quality using the output flags from the code itself and rated each star member out of 375 514. 1340 clusters passed our selection. Our evaluation was based on the extinction (AV50), the distance (Dist50) and the metallicity (Met50).

Additionally, we limited the amount of clusters in order to reduce the study to the most populated clusters. The criteria that we established to select the candidates was that they had to contain more than 600 hundred valid stars, which involved a subsequent decrease to 59 clusters remaining. Furthermore, we also separated the HRD in 7 regions depending on the spectral type of the stars and their position in the diagram in order to have categorized each type of stars. Moreover, it was necessary to subdivide the SUBGIANT and the GIANT group in terms of the parallax uncertainities of the clusters that encompassed such stars. We noticed that every cluster had stars from the A-F group, so it was chosen to be our frame of reference in the subsequent analysis.

By the time we plotted the difference of each parameter, between the mean of every region in the HRD and the equivalent mean value for the A-F group, we saw the unexpected. Knowing that the stars from the same cluster have similar values on their parameters, the distributions of the mean should be near zero but, instead, the plots showed significant and systematic trends. In particular, the extinction and the distance had a negative distribution scattering, while the metallicity had it the inverse direction. The conclusion of all these considerations is that, although there are evident differences in the results, we must take into account that *StarHorse* doesn't identify star clusters but analyses stars individually. In addition, it is fairly efficient in terms of the actual available photometry passbands and its quality, which are not the same throughout all the stars. We also know that *StarHorse* compares the data with current models, and thus, the observed differences could be induced by those models.

Once we came to these interpretations, another last issue had to be solved. We wanted to contrast our calculations of the distance and the absorption with the literature and, if our numbers were right, we expected a linear correlation between the mean values. What we saw in the plots was close to our predictions, except for a few clusters spread apart. Under all these considerations, as a summary of our entire study, we could collect in a final list all the mean values and standard deviations of the 1340 clusters with validated star members and have reasonable certainty of their reliability.

[1] *Gaia* Collaboration, 2016, *A&A*, 595, A1
[2] *Gaia* Collaboration, 2018, *A&A*, 616, A1
[3] Maerrese, P.M, 2019A&A, 621, A144
[4] Scolnic, D., 2015ApJ, 815, 117
[5] Anders, F., et al. 2019, *A&A*, GDR2_SH
[6] Cantat-Gaudin, T., et al. 2018, *A &A*, 618, A93
[7] Cantat-Gaudin, T., et al. 2019, *A &A*, 624, A126
[8] Castro-Ginard, A., et al. 2018, *A&A*, 618, A59
[9] Castro-Ginard, A., et al. 2019, *A&A*, 627, A35
[10] Bossini, D., et al. 2019, *A&A*, 623, A108
[11] Gyuheon, S., et al. 2019, *Journal of the Korean Astronomical Society*, 52: 145 158
[12] Kharchenko, N. V, Piskunov, A. E., Schilbach, E., Röser, S., Scholz, R.D., 2016, A&A, 585, A101
[13] http://www.star.bris.ac.uk/ mbt/topcat/