Estimation of the distance and the properties of open clusters with Gaia

Author: Laura Barrio Hernandez.
Advisor: Xavier Luri
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Abstract: The purpose of this research work is to determine the distance and the properties of several open clusters using the Tycho-Gaia Astrometric Solution (TGAS) of the first Gaia Data Release. We have analysed the problem of the membership selection and the data collection using TGAS and the information in other catalogues. We have studied the properties of different clusters using a membership selection done by van Leeuwen (2017) and using the Palmer et al. (2015) algorithm to derive the different properties of the clusters. We have also compared our results with previous values from the bibliography for the studied clusters.

I. INTRODUCTION

An open cluster is a group of stars that were formed from the same molecular cloud and therefore have similar age, chemical composition and properties (such as proper motion and radial velocity). The study of open clusters is very important in the study of the stellar formation, structure and evolution and also provides information about the evolution and formation of the Galaxy.

Gaia is a satellite of the European Space Agency (ESA) whose mission is to compose a three-dimensional map of the Milky Way with an unprecedented accuracy in the measurements of distance, proper motions, radial velocities (available in future data releases), magnitudes, etc. The precision of the measurements of the Gaia satellite, in particular the TGAS catalogue, makes them very useful for the study of open clusters.

The main purpose of this work is to obtain the properties of open clusters using an algorithm specifically developed to use the Gaia data. This algorithm has been used in this work with Gaia data for the first time.

II. DATA COLLECTION

To collect the data for our study of open clusters we have used the Tycho-Gaia Astrometric Solution (TGAS) [1] combined with other catalogues, such Tycho, RAVE and GES, to complement it with additional data (colour indices, radial velocities, etc.).

TGAS has been used for the main data collection: position, parallax, proper motion, magnitude and stars IDs. But TGAS has not all the information needed for our analysis of clusters, so we have cross-matched it with these other catalogues in order to find extra information: colour indices from Tycho catalogue and radial velocities form RAVE [2] and GES [3]. But neither RAVE nor GES have information about all the stars in all the clusters we have analysed and, therefore the radial velocity information in our study is limited.

The clusters studied are the same 19 clusters analysed by van Leeuwen [4]: Alpha Persei, Blanco 1, Coll140, Coma berenices, the Hyades, IC2391, IC2602, IC4665, NGC2232, NGC2422, NGC2451, NGC2516, NGC2547, NGC3532, NGC6475, NGC6633, NGC7195, the Pleiades and Praesepe.

III. MEMBERSHIP SELECTION

To determine the properties of a cluster, it is first essential to correctly select its members, that is, identify the stars belonging to it and separate them from background stars. Due to the homogeneity in distance and proper motion, an initial approximation to identify possible members is to consider stars at a similar distance from the Sun and with a similar proper motion. We have implemented a simple algorithm that selects stars from a manually fixed range of proper motions in right ascension and declination. In a similar way, with the help of a histogram of distances we can reject stars too far from the cluster centre. We have used this dual selection process with 3 clusters: Alpha Persei, the Hyades and the Pleiades.

Cluster	van Leeuwen members	Manual selection members	Members in common
Alpha persei	116	72	65
Hyades	103	34	29
Pleiades	154	101	94

TABLE I: comparison between the members selected by van Leeuwen's algorithm and the manual selection showing the members in common. The criteria for the manual selection are the following: 17mas≤μα≤23mas, -48mas≤μδ≤-42mas and 6.67mas≤ϖ≤8.33mas for the Pleiades in a radius of 2deg, 95mas≤μα≤125mas, -40mas≤μδ≤-15mas and 16.67mas≤ϖ≤25.00mas for the Hyades in a radius of 3deg and 20mas≤μα≤30mas, -28mas≤μδ≤-23mas and 5.56mas≤ϖ≤6.25mas for Alpha persei for a radius of 2deg.

Cluster	F. van Leeuwen			Manual selection		
	ϖ σϖ	μα σμα	μδ σμδ	ϖ σϖ	μα σμα	μδ σμδ
Alpha Persei	5.91 0.03	23.06 0.06	-25.36 0.07	5.60 0.10	23.15 0.24	-25.13 0.16
Hyades	21.39 0.21	104.92 0.12	-28.00 0.09	21.36 0.26	107.11 1.13	-25.32 0.82
Pleiades	7.48 0.03	20.38 0.07	-45.39 0.08	7.39 0.07	19.99 0.12	-45.13 0.12

TABLE II: Comparison between the results derived by F. van Leeuwen and the results derived with the manual membership selection algorithm. The results derived with the manual selection algorithm are averages.

* Electronic address: lbarrihe8@alumnes.ub.edu

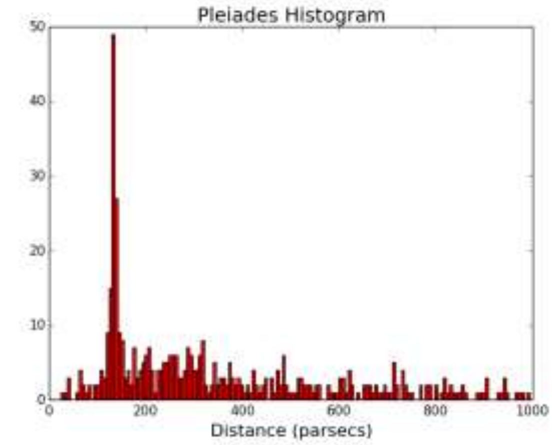

FIG 1: Example of histogram of distances for the sky region around the Pleiades, showing the density peak corresponding to the cluster.

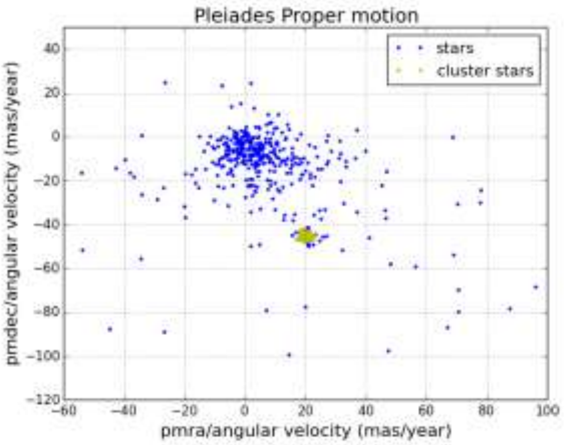

FIG 2: Example of the proper motion homogeneity and the manual selection of the cluster members in the case of the Pleiades. In this case, a file of 488 stars in a radius of 2 deg from the centre of the cluster were used, which 101 were selected as members of the cluster.

This manual procedure is just a first approximation, implemented in this work for learning purposes, and uses only a part of the available information. A more rigorous membership selection is needed to obtain more reliable results. Also, manual selection in more distant clusters could be problematic due to the difficulty of the visual identification of the distance in the histogram and the proper motions in the proper motions plot.

The implementation of an accurate membership selection algorithm is a difficult problem beyond the scope of this work, and therefore we have used the membership selection done by van Leeuwen [4], made using different methods depending on the distance of the cluster. But, as we have mentioned before, the key variables to use to determine the members of a cluster still are position, distance, proper motion and, in some cases, radial velocities.

Another example of membership selection method is Clusterix [5] a software that uses a non-parametric method to determine the membership of open clusters. We could not use Clusterix for this report but it can be used for future work extending our analysis.

IV. STUDIED CLUSTERS

Once the problem of the cluster membership is solved we can determine the properties of our clusters.

We have used for the first time with real Gaia data an algorithm developed by Palmer (2014) [6] based on the maximum likehood estimation (MLE) to derive the properties of the selected clusters. This algorithm combines parallax, positions, apparent magnitude, colour, proper motions and radial velocity (this last can be omitted from the calculations) information to estimate the parameters characterizing an open cluster.

The MLE method estimates the parameters of a statistical model (chosen by the user) by maximising the a-posteriori probability of the observed values as a function of these parameters. In this case, the probability distribution function describes an open cluster assuming a Gaussian spherical distribution for the spatial distribution, an ellipsoid for its velocity distribution and a distribution around an isochrone in the colour-magnitude plane. A benefit of this method is the possibility of include data with negative parallaxes, thus avoiding selection biases.

We have used a version of the algorithm not using the radial velocity information for the bulk of this work, although we have also tested a version with radial velocities in a few clusters. However, this second version was not fully functional and reliable at the time of completion of this work so we decided not to include any result from the version with radial velocities.

Thanks to this work we were able fix some bugs and to improve the implementation of the algorithm, making it more generally applicable to a wider range of clusters.

V. RESULTS

We have derived the main properties of the studied clusters: mean distance (parallax), mean proper motions, the variance of the spatial distribution of the stars (the cluster size) as a function of the colour index, observational isochrones, the dispersion of the absolute magnitudes around these isochrones and the variance of the proper motion distribution.

Since this was the first time that the algorithm was used, some errors and problems were found. The main one, which could not be fixed by the author of the code in time for the completion of this work, is the computation of formal errors of the cluster parameters derived by the program, therefore, these are not available for the moment. This limits the comparison with results from other authors since we cannot be sure of the level of confidence of our estimations.

The results presented below have been derived with the version of the algorithm without radial velocities.

A. Parallax

Cluster	$\varpi_{F.\ van\ Leeuwen}$	$\sigma\varpi_{F.\ van\ Leeuwen}$	ϖ
Alpha Persei	5.91	0.03	5.79
Blanco 1	4.34	0.11	4.41
Coll140	2.86	0.11	2.83
Coma berenices	11.73	0.05	14.34
Hyades	21.39	0.21	22.18

IC2391	6.97	0.13	6.92
IC2602	6.74	0.05	6.74
IC4665	2.83	0.05	2.79
NGC2232	3.00	0.06	3.01
NGC2422	2.28	0.09	2.21
NGC2451	5.59	0.11	5.45
NGC2516	3.01	0.10	2.84
NGC2547	2.75	0.08	2.75
NGC3532	2.42	0.04	2.25
NGC6475	3.57	0.02	3.58
NGC6633	2.41	0.04	2.47
NGC7092	2.99	0.12	3.22
Pleiades	7.48	0.03	7.51
Praesepe	5.47	0.05	5.46

TABLE III: Comparison between the results derived by F. van Leeuwen and the results derived with M. Palmer's algorithm for the parallax.

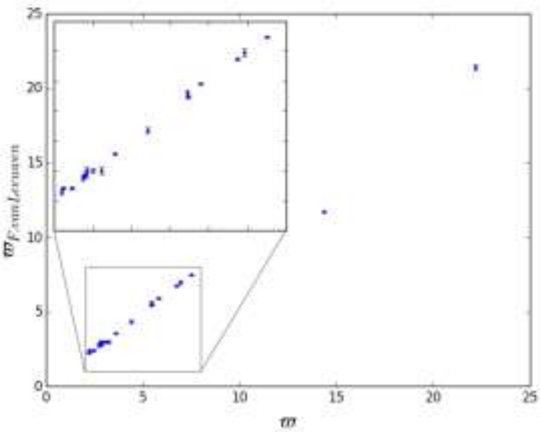

FIG 2: Parallax comparison. The zoomed region corresponds to the region with parallax between 2 mas and 8 mas.

As we can see in Table III and in Fig. 3, all values obtained for the mean parallax of the clusters are inside the errors for the value found by F. van Leeuwen, except for Alpha persei, Coma berenices, NGC2526, NGC3532 and NGC6633.

B. Proper motions

Cluster	F. van Leeuwen		MLE	
	μα σμα	μδ σμδ	μα	μδ
Alpha Persei	23.06 0.06	-25.36 0.07	23.04	-26.24
Blanco 1	18.20 0.12	2.66 0.11	18.60	2.86
Coll140	-8.36 0.09	4.95 0.10	-8.98	4.88
Coma berenices	-12.14 0.14	-8.90 0.16	-12.15	-9.11
Hyades	104.92 0.12	-28.00 0.09	103.55	-26.48
IC2391	-24.15 0.13	23.83 0.25	-32.60	23.25
IC2602	-17.67 0.09	11.06 0.13	-16.89	10.99
IC4665	-0.78 0.07	-8.37 0.06	-0.75	-8.52
NGC2232	-4.34 0.10	-1.71 0.08	-4.42	-1.73
NGC2422	-6.80 0.08	0.99 0.08	-6.76	1.09
NGC2451	-21.82 0.11	15.59 0.16	-21.48	15.08
NGC2516	-4.06 0.07	11.16 0.08	-4.70	10.92
NGC2547	-8.92 0.07	4.07 0.09	-8.91	4.23
NGC3532	-10.65 0.04	5.27 0.04	-10.44	5.26
NGC6475	3.10 0.06	-5.32 0.04	3.18	-5.48
NGC6633	1.45 0.06	-1.78 0.05	1.32	-1.78
NGC7092	-7.34 0.11	-19.94 0.13	-8.04	-19.67
Pleiades	20.38 0.07	-45.39 0.08	20.25	-45.37
Praesepe	-36.06 0.07	-13.15 0.08	-35.89	-6.17

TABLE IV: Comparison between the results derived by F. van Leeuwen and the results derived with M. Palmer algorithm for the proper motions.

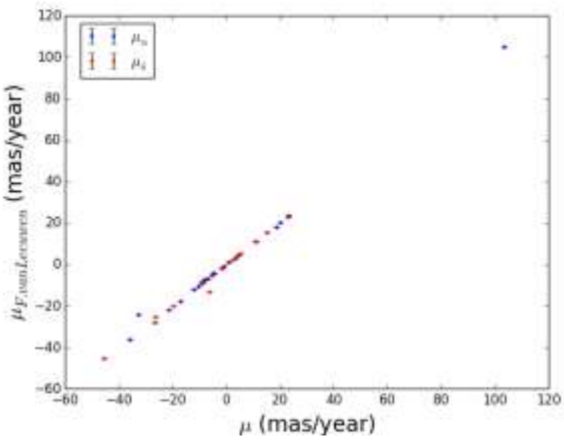

FIG 3: Proper motions comparison in right ascension (blue) and in declination (red).

As we can see in Table IV and in Fig. 4, the values derived for the mean proper motions are of the clusters are mostly compatible with the ones in van Leeuwen paper, but in a few cases show significant discrepancies (i.e IC2391 in right ascension and Praesepe in declination). Also, the results are more accurate in right ascension than in declination.

C. Other results

Another product of Palmer's code is the generation of the observational isochrones. The program uses a series of points in the Colour-Magnitude diagram to fit a spline representing the isochrones. A limitation of the code is that it can only handle main sequence stars, and therefore requires that the giants present in the cluster (if any) are removed beforehand. This has been done manually by analysing the clusters in the

Colour-Magnitude plane. In most cases the isochrones trace very well the main sequence of the cluster, but there are deviations in several cases that show the limitations of the program and the need to further refine the implementation of the method.

The isochrone fitting requires a colour index, not available in the TGAS data. Tycho catalogue photometry (B-V index) has been used for this purpose because of its homogeneity. But as the precision of this photometry is low for the fainter stars, the lower part of the main sequence has larger error bars and the results derived for the isochrones in this region are of limited reliability.

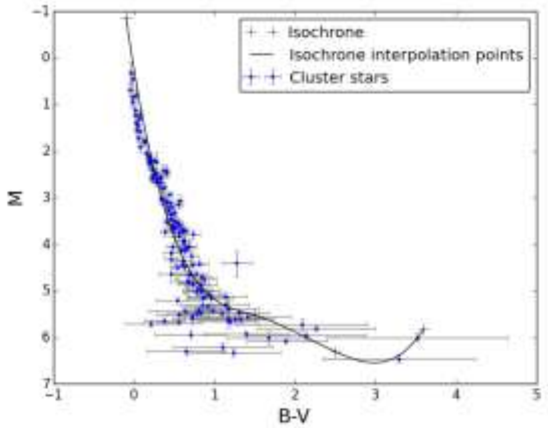

FIG 4: Isochrone interpolation example for the Pleiades. Notice the degradation of the precision of the Tycho photometry in the faint end.

The cluster size (radius) as a function of the B-V colour has also been derived. This colour-depending fitting allows to show the cluster's mass segregation: the hotter and more massive members of a cluster tend to be found at the centre of the cluster (smaller radius) while older and less massive members tend to be found at the periphery of the cluster (larger radius).

Evidence of mass segregation is shown in Fig. 6 but, because of the formal errors are not derived by the current version of the program we cannot validate some of the results for the radius of the cluster as a function of the colour index.

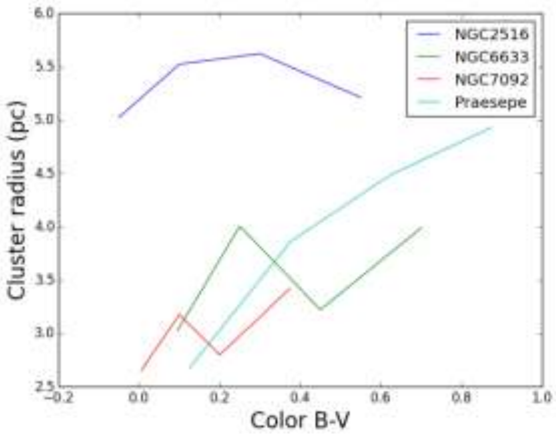

FIG 5: Cluster size as a function of the colour index for 4 examples: NGC2516, NGC6633, NGC7092 and Praesepe.

The variance of the proper motion distribution has also been derived. The values found are very large in some cases, so the proper motion distribution could be less homogeneous in these cases.

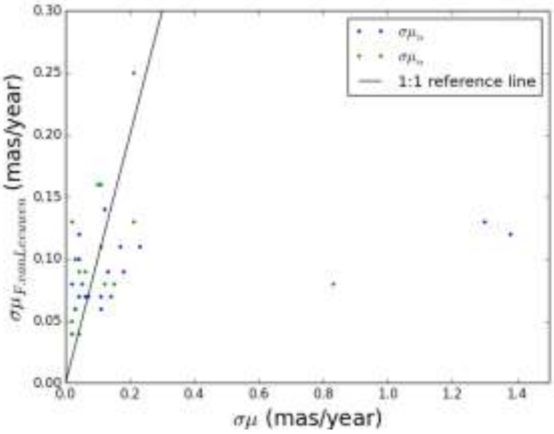

FIG 6: Variance of the proper motion comparison.

As we can see in Fig. 7, the variances of the proper motions do not match with the result derived by F. van Leeuwen and a further analysis of these results must be done in order to find the reason.

VI. CONCLUSIONS

- In this work, we were able to use for the first time the MLE algorithm developed by Palmer (2014), helping to improve it, to obtain the mean parallax, mean proper motions and other parameters describing an open cluster for the same 19 clusters studied by van Leeuwen (2017). As the algorithm was initially made for the Pleiades and the Hyades some problems were found when we tested it with other clusters.
- The first results derived with this method are accurate and compatible, in most cases, with the results derived by other authors.
- The results of this work can be improved in the future with the Gaia DR2 which will include more stars (implying a better membership selection with Clusterix), radial velocities and more photometric data. The results could also be improved if Geneva photometry is used instead of Tycho photometry, providing more precision at the faint end of the main sequence of the clusters.
- We expect to solve the remaining algorithm limitations, test it using radial velocities and compare results with those derived in this work, in order apply it to more clusters using the upcoming Gaia DR2.

Acknowledgments

I especially want to acknowledge my advisor Xavier Luri for guiding and teaching me in this work. I would also like to thank Erika Antiche for her constant help and for helping me with Max Palmer's algorithm. Finally, I want to thank Francesca Figueras, Carme Jordi, Teresa Antoja and all the Gaia Science group for their advice, help and interpretation of the results derived.

[1] "The Gaia Archive," [Online]. Available: https://gea.esac.esa.int/archive/.

[2] "Rave - The Radial Velocity Experiment," [Online]. Available: https://www.rave-survey.org.

[3] "GES - Gaia Eso Survey," [Online]. Available: http://ges.roe.ac.uk/.

[4] F. van Leeuwen et al, "Gaia Data Release 1. Open cluster astrometry: performance, limitations, and future prospects," *Astronomy & Astrophysics, in press (*2017).

[5] "Clusterix," [Online]. Available: http://clusterix.cerit-sc.cz/.

[6] Palmer, M, Luri, X. et al. , "An updated maximum likelihood approach to open cluster distance determination," *Astronomy & Astrophysics,564, 49, (*2014).