Bias Effects Study in Open Clusters Parallaxes using Gaia

Author: Daniel Castellà Santos.
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisor: Carme Jordi Nebot

Abstract: This is a research study on how the open clusters parallax measurements can be affected by different biases using the precision of Gaia satellite. In order to carry on this project, NGC6705 and NGC2682 data has been used. As a result, there are some deviations between observed and true parallax when magnitude and σ/π_o cutoffs are applied from different observation distances.

I. INTRODUCTION

A. Gaia mission

Gaia is one of the most ambitious missions of ESA (European Space Agency). Its objective is to register astrometric, photometric, and spectroscopic data for more than one billion stars in the Milky Way with a never seen accuracy until now [9]. This information will be used to build a tridimensional map of our cosmic neighborhood even though this one billion stars only represents 1% of our entire Galaxy.

Besides this stellar data, the discovery of new objects is expected, thanks to the great sensitivity and precision of the instruments being used. These can reach magnitude 20. New brown dwarfs, exoplanets, supernovas, and quasars are expected to be found, as well as asteroids and comets in our solar system.

The satellite was launched on 19 December, 2013 from Kourou Spaceport in French Guiana by a Soyuz-Fregat launcher. On 8 January, 2014 the satellite arrived to the second Lagrange point (L2) located about 1.5 million kilometers from Earth. There it orbits thanks to the gravitational influence of the Sun and our planet.

B. Biases Study

This work intends to derive which biases are introduced when measuring the observed parallax (π_o) of an open cluster from a sample of its individual stars parallaxes. Despite Gaia can only observe and measure stars with magnitude $G \leq 20$, this will not produce a bias. It will be introduced when a σ/π_o cutoff is applied. When an astrometric standard error of Gaia (σ) become significant in relation to the observed parallax, this star will be eliminated from the sample. Consequently, different biases will be introduced depending on the number of stars removed, that will be regulated by the tolerance of σ/π_o cutoff.

For many years, bias influences in astronomical measurements have been studied. As example, Lutz & Kelker [1] proved that the mean observed parallaxes of field stars are larger than their mean true parallaxes because there are more stars in the volume shells at farther distance, introducing a bias. In a parallax Gaussian distribution, due to the above mentioned bias, it can be observed that the larger the σ/π_o cutoff is applied the more the Gaussian center is displaced. This is the reason why the mean observed parallaxes measurements are larger than the mean true ones. With Hipparcos satellite, the predecessor of Gaia, studies were also carried out on how the resulting measurements were affected by biases [2].

C. Clusters Selection

For this work, two different stellar clusters data from Gaia simulators [8] have been used. However, the goal is to study a greater number of clusters, so NGC6705 and NGC2682 have been chosen as models. They will be set at farther distances to recreate other clusters with similar properties to them. This strategy will allow to study biases as a function of distance.

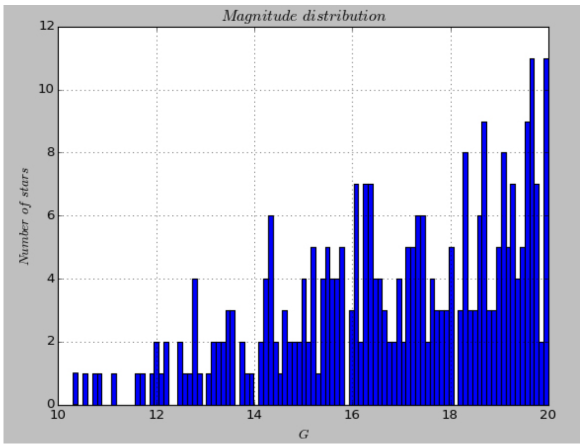

FIG. 1: Distribution of stars in NGC6705 as a function of G magnitude.

Simulations are available for the true distance of the cluster and contain stars up to the observing limit of Gaia ($G = 20$). If a shorter than true cluster distance

*Electronic address: danicas10489@hotmail.com

is assumed, stars at the faint limit would be observable, and the simulations do not contain them. So, the sample would be incomplete. For this reason, we only consider distances equal or larger than the true distance.

NGC6705 and NGC2682 have been chosen because they are very different from one another. The first cluster is a well known one, also called Wild Duck Cluster or M11 and it is located at the Scutum constellation (the Shield). It is located at about 1877 pc and its estimated age is about 0.2 Gyr [4]. NGC6705 is rich in stars and they have a lot of different magnitudes as shown in Fig.(1).

The second cluster does not have a proper name but it is also known as M67. It is located at the Cancer constellation (the Crab) at about 908 pc. In contrast with NGC6705, this cluster is much older. Its age is about 2.6 Gyr [5], and it is less rich and faint stars are numerous as per Fig.(2).

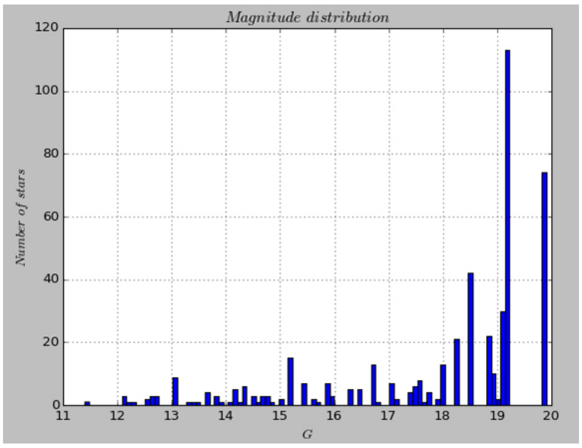

FIG. 2: Distribution of stars in NGC2682 as a function of G magnitude. The strong clumpy distribution is an artifact of the simulated data.

II. DESIGN AND DEVELOPMENT OF CALCULATIONS WITH PYTHON SOFTWARE

For this work, a set of new programs have been designed and developed using free Python software [6]. The main goal is to study the clusters at several different distances with several σ/π_o cutoffs applied. Therefore, the main program has been developed to perform a distance scan and a σ/π_o scan with some control parameters (start, end, and step) eligible by the user. For each scan, the program allows for the calculations of the observed mean parallax $\langle\pi_o\rangle$ and the true mean parallax $\langle\pi\rangle$. The results of the difference $\langle\pi_o\rangle - \langle\pi\rangle$ are shown in two different graphs, one for the distance scan, and the other for the σ/π_o scan. Moreover, to better visualize the results of the study, it is possible to select four different values of σ/π_o to compare different cutoffs at the distance scan. Likewise, four different distances are elegible at the σ/π_o scan.

A. Collecting and processing data

With the data available [8], the program reads the Right Ascension (α) , Declination (δ), Distance (d), Gaia Magnitude (G), and Color Index ($V - I_c$) of the cluster's stars. These are the true values of each star, so the observed values are required. To simulate them, the program defines three functions:

- *Error* function. The inputs are Magnitude G and Color Index. The output is the astrometric standard error σ of parallax as measured by Gaia [3],

$$\sigma = (-1.631 + 680.766 \cdot z + 32.732 \cdot z^2)^{1/2} \cdot \\ \cdot[0.986 + (1 - 0.986) \cdot (V - I_c)] \quad (1)$$

where,

$$z = MAX[10^{0.4\cdot(12.09-15)}, 10^{0.4\cdot(G-15)}] \quad (2)$$

- *Cluster1* function. First of all, the function calculates the tridimensional distribution of the stars in relation to the cluster center ($\vec{r}$) using Eq. (3),

$$\vec{r} = \vec{r}_* - \vec{r}_c \quad (3)$$

where $\vec{r}_*$ is the original stars position in relation to the observer and $\vec{r}_c$ is the original cluster center position in relation to the observer,

$$\vec{r}_* = \begin{pmatrix} d \cdot \cos(\delta) \cdot \cos(\alpha) \\ d \cdot \cos(\delta) \cdot \sin(\alpha) \\ d \cdot \sin(\delta) \end{pmatrix} \quad (4)$$

$$\vec{r}_c = \begin{pmatrix} d_c \cdot \cos(\delta_c) \cdot \cos(\alpha_c) \\ d_c \cdot \cos(\delta_c) \cdot \sin(\alpha_c) \\ d_c \cdot \sin(\delta_c) \end{pmatrix} \quad (5)$$

the values of d_c, δ_c, and α_c are obtained using the means of the distances, right ascensions and declinations of all cluster's stars.

After that, the new true parallaxes (π') and magnitudes (G') of stars are calculated using Eq. (6) and (7) when the cluster center is placed at a new distance (d'_c).

$$\pi' = \frac{1}{|\vec{r'}_*|} = \frac{1}{|\vec{r} + \vec{r'}_c|} \quad (6)$$

$$G' = G + 5 \cdot \log \frac{|\vec{r'}_*|}{|\vec{r}_*|} \quad (7)$$

Finally, it also assigns an observed parallax (π_o) to each star that results from a Gaussian random distribution defined by the true parallax (π or π') and the error (σ).

- *Cluster2* function. This one provides an output of true and observed parallaxes with the applied σ/π_o and magnitude cutoffs. All values are kept in arrays with each star having an assigned element. For stars above the cutoff no number is assigned to the element, yet the element remains and this helps to determine which stars are acceptable for the cutoff and which ones are not.

A statistic increase is also applied to avoid any problems when a low number of stars is in place (like the case of NGC2682). In essence, all stellar values are used repeated as many times as needed without affecting the bias. This helps to better draw the graph curves and avoid unstable behavior typical for small statistic samples.

B. Calculations and iterations

Once all tools are ready, it is time to obtain the results. Calculations are done in the distance loop, which is the main loop, with specific values assigned from the beginning. The program uses *Cluster1* and *Cluster2* functions. The first one, provides G, π, π_o, σ and σ/π_o arrays which will vary according to predetermined distance loop. The second one provides the observed parallaxes and the true parallaxes arrays with the first applied cutoff.

For each loop iteration, $\langle\pi_o\rangle - \langle\pi\rangle$ is calculated and the plot in distance is generated. Then, *Cluster2* function gives the observed parallaxes and the true parallaxes for each remaining σ/π_o cutoffs and those calculations are repeated and plotted in the graph.

Before finishing the distance loop iteration, there is an additional step. As previously mentioned, the program allows for σ/π_o scan for a specific set of distances. In the main loop it is necessary to set a sequence of conditional assignments that trigger sub-loops. Once an iteration reaches the first given distance in the main loop, this triggers the first σ/π_o loop. If this does not happen, π_o values change because *Cluster1* function would create new random values. Within the triggered error loop, *Cluster2* function provides cut parallaxes. The difference here is that instead of using a fixed cutoff it includes all cutoffs in the σ/π_o scan. Then, it calculates the difference between the mean observed and the true parallax for each iteration of the first σ/π_o loop, and provides the full plot. This sequenced process is repeated with the second, third, and fourth σ/π_o loop.

III. ANALYZING RESULTS

With all the results, the next step is to draw them in graphs to depict how the biases affect the mean calculated in all different cases. Usually, the means are used for this kind of calculations but other statistic averages will be studied in the last subsection.

A. NGC2682

Generally, the cluster mean observed parallax is larger than the true parallax because of the biased data. As in Fig.(3), the behaviors are very different depending on the σ/π_o cutoff. This is explained in Fig.(2) because the stars magnitude distribution has a well defined gaps. Since stars are all set together in big groups of similar magnitude (and σ mainly depends on G), these groups are eliminated at relatively the same distance. Suddenly losing a big number of stars results on a sharp decrease on the graph curve. For example, the peaks observed almost reaching 1000 pc (except in $\sigma/\pi_o = 0.1$) correspond to the stars group of magnitude ~ 20. Those are the first ones to be eliminated for being closer to the magnitude limit and therefore, when moving further the cluster, those stars are fainter than the limit. As per Fig.(3), in the $\sigma/\pi_o = 0.1$ case, the peak is not shown due to those stars have been eliminated by σ/π_o cutoff as shown in Fig.(4). In the $\sigma/\pi_o = 0.25$ case, the decrease is smooth due to this cutoff only takes a few stars of that group, so their elimination is not critical.

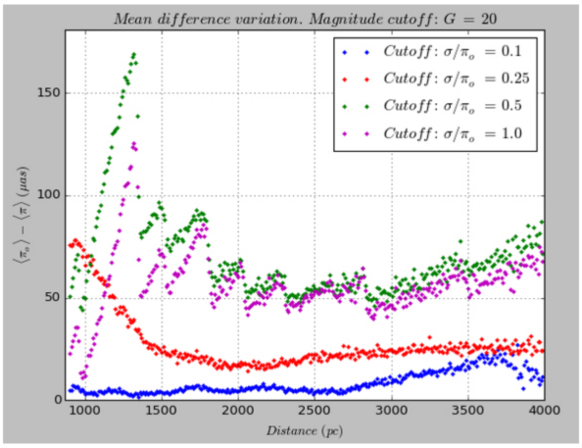

FIG. 3: Mean difference variation as per NGC2682-like cluster distance.

Likewise, the peak approximately situated at 1300 pc is due to elimination of stars set of magnitude ~ 19 in Fig.(2). It does not appear in cases $\sigma/\pi_o = 0.1$ and $\sigma/\pi_o = 0.25$ because they have been eliminated by the σ/π_o cutoff. The other visible peaks in cases $\sigma/\pi_o = 0.5$ and $\sigma/\pi_o = 1.0$ represent other groups.

For the σ/π_o scan, Fig.(5), the bias provokes the mean observed parallax to be larger than the true parallax again. This is explained in Fig.(4). The black line represents the true parallax of the cluster center (π_c), so more stars from the right side ($\pi_o > \pi_c$) are taken than from the left side ($\pi_o < \pi_c$). The peaks resulting from different distances correspond to the cutoff with the maximum number of stars taken with large and positive π_o. For tolerance with larger σ/π_o, more $\pi_o < \pi_c$ stars are being accounted for. Consequently, there is a balance that makes the value of $\langle\pi_o\rangle$ to be reduced and to be

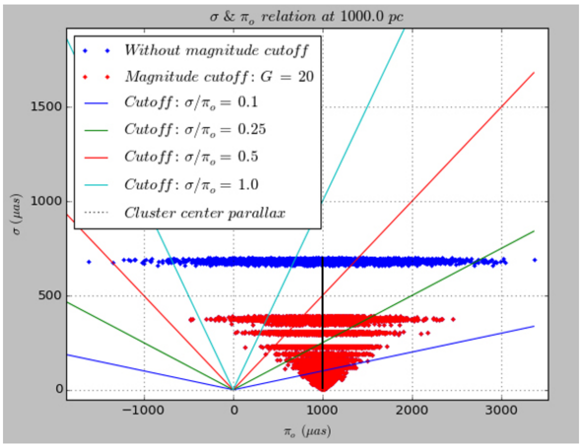

FIG. 4: σ of each star according to the observed parallax in NGC2682-like cluster at 1000 pc. The stars used to calculate the average are the red dots situated under the color lines depending on the applied cutoff.

closer to $\langle\pi\rangle$. Fig.(5) also helps to understand why in Fig.(3) the mean difference on $\sigma/\pi_o = 0.5$ case is larger to the one in $\sigma/\pi_o = 1.0$ case since the maximum point of the curve is found on $\sigma/\pi_o \sim 0.6$ (at longer distances). The difference between the 1000 pc case and the other ones is because the magnitude ~ 19 group at 1000 pc has not been eliminated.

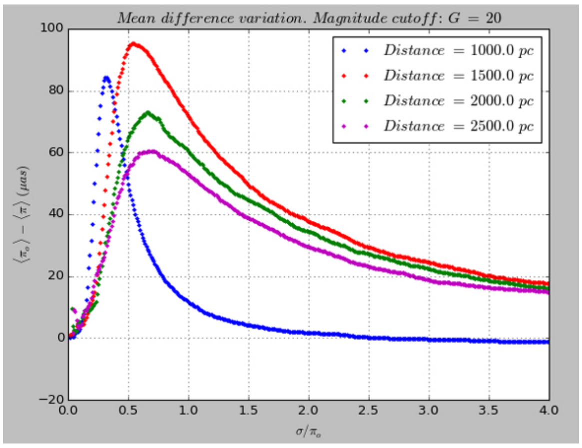

FIG. 5: Mean difference variation in relation to σ/π_o scan in NGC2682-like cluster.

B. NGC6705

It can be seen in Fig.(6) that, for NGC6705, mean deviation is small and fairly constant with the distance for very strict σ/π_o cutoffs. However, for larger cutoffs, $\langle\pi_o\rangle - \langle\pi\rangle$ deviation is larger and also increases with distance. Contrary to NGC2682, in this cluster there are not sharp decreases because the distribution of stars in Fig.(1) does not have gaps, so big stellar groups are not suddenly eliminated.

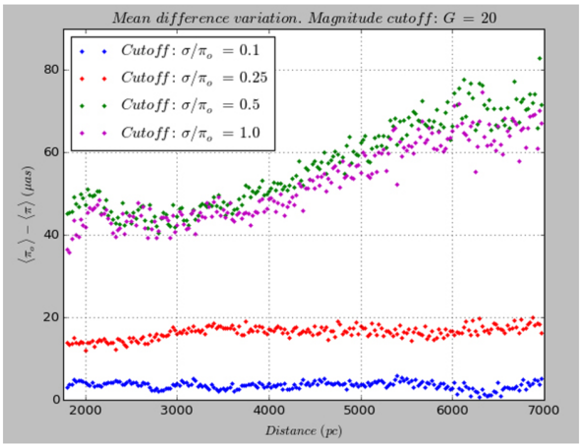

FIG. 6: Mean difference variation as per NGC6705-like cluster distance.

On Fig.(7), the behavior is very similar to NGC2682. For both type of clusters, when a specific distance is given, the maximum point appears when the σ/π_o cutoff is smaller than 1.0. This allows to think that in order to avoid the bias the best way to proceed is to choose either extremes, very restrictive σ/π_o values or very permissive σ/π_o values.

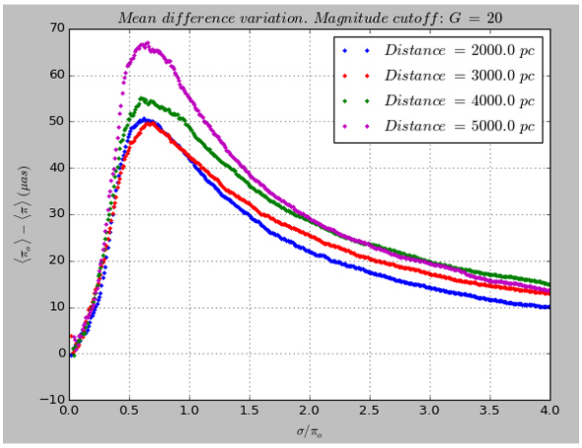

FIG. 7: Mean difference variation in relation to σ/π_o scan in NGC6705-like cluster.

The relative bias can also be studied if $\frac{\langle\pi_o\rangle-\langle\pi\rangle}{\langle\pi\rangle}$ calculation is done. For NGC6705-like cluster, the maximum relative error at 2000 pc is about 10%, at 3000 pc is about 15%, at 4000 pc is about 22%, and at 5000 pc is about 33%.

C. Other statistic averages

Even if this project has been mainly based in parallax mean calculations, the Python program also allows to do statistics of medians and weighted means by using $1/\sigma_i^2$ as a weight in this last one. The obtained results, Fig.(8), indicate that bias affects much less if, instead of measuring the means, the medians are used and are almost unnoticeable with the weighted means. Even if the graph only shows NGC6705 cluster at 5000 pc, in all other cases the behavior is the same. Consequently, when dealing with an open cluster, the best way to correct the bias is by using weighted means since, as proved in Fig.(8), by only taken the means this results on larger deviations between observed and true values.

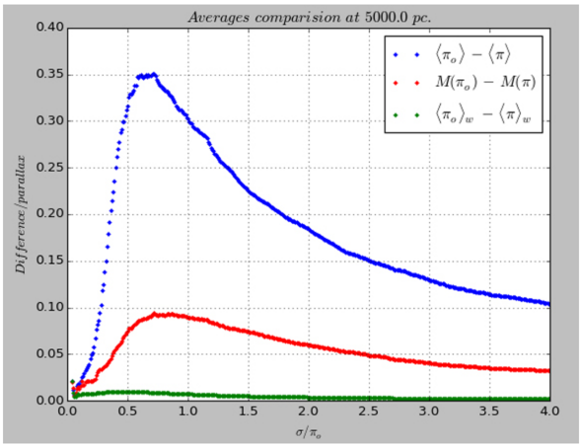

FIG. 8: NGC6705-like cluster at 5000 pc averages comparison when calculating cluster parallax according to means (blue), medians (red) and weighted means (green).

IV. CONCLUSIONS

- When applying σ/π_o, the $\langle\pi_o\rangle - \langle\pi\rangle$ is positive because more $\pi_o > \pi_c$ stars are taken than $\pi_o < \pi_c$ stars. σ vs π_o graph, Fig.(4), is very helpful to understand progression of deviations because it shows how elimination process is done.

- For both types of clusters, it is better to use either very restrictive σ/π_o values or very permissive σ/π_o values so that $\langle\pi_o\rangle$ is closer to $\langle\pi\rangle$, regardless of the cluster distance.

- For clusters with traits similar to NGC6705, distance progressing of $\langle\pi_o\rangle - \langle\pi\rangle$ is more stable because it is richer and the brighter stars are more noticeable. However, for clusters like NGC2682, there are sharp decreases due to the gaps at the magnitude distribution and because the small population of bright stars in relation to faint stars (typical from an old cluster).

- With Gaia precision it is observed that, for clusters like NGC6705 situated at 7000 pc, there will be biases smaller than 70 μas. So, the maximum relative error will be smaller than 49%. For clusters similar to NGC2682, there will be below 180 μas. So, below 126%.

- Overall, in order to avoid biased calculations by a σ/π_o cutoff, the best way to proceed is to do a statistic calculation based on the weighted means. Instead of being based only with the arithmetic mean direct values. For clusters with individual parallaxes with very large relative errors, maximum likelihood methods would be more suitable (like the one used in Palmer et al [10]).

Acknowledgments

First of all, I want to thank my advisor Dr. Carme Jordi for all her help and her wise guidance while working in this project. Also, I want to thank my family and friends for their unconditional support and attention. Last but not least, I want to express my gratitude to the UB rugby team, that has been a wonderful way to distress on the most challenging roadblocks during my studies.

[1] T. Lutz and D. Kelker, "On the Use of Trigonometric Parallaxes for the Calibration of Luminosity Systems: Theory" Astron. Soc. Pacific, **85**: 573-578 (1973).
[2] A.G.A. Brown, F. Arenou, F. van Leeuwen, L. Lindegren and X. Luri, "Some Considerations in Making Full Use of the Hipparcos Catalogue". Proceedings of meeting "Hipparcos Venice '97".
[3] Gaia Science Performance from Cosmos Portal: http://www.cosmos.esa.int/web/gaia/science-performance.
[4] NGC6705 data from WEBDA: http://www.univie.ac.at/webda/cgi-bin/ocl_page.cgi?cluster=ngc6705.
[5] NGC2682 data from WEBDA: http://www.univie.ac.at/webda/cgi-bin/ocl_page.cgi?cluster=ngc2682.
[6] Python development from Python Software Foundation: https://www.python.org.
[7] Python development from Stack Overflow: http://stackoverflow.com.
[8] C. Jordi, X. Luri and L. Balaguer, "Open Clusters as seen by Gaia". Simulated Gaia observations produced by the Gaia Object Generator (GOG).
[9] M.A.C. Perryman et al. "GAIA: Composition, formation and evolution of the Galaxy". A&A, **369**: 339-363 (2001).
[10] M. Palmer et al. "An updated maximum likelihood approach to open cluster distance determination". A&A, **564**: A49 (2014).