NGC 6397: An Analysis of the Colour-Magnitude Diagram Using Gaia Data

Kamran Jornacion
Queen's University
22qfc4@queensu.ca

I. Introduction

The objective of this project is to leverage observation data from Gaia to construct a color-magnitude diagram of the NGC 6397 globular cluster. A statistical analysis of the data will be conducted via python and several common libraries such as `pandas` and `matplotlib`. The members of the globular cluster will be primarily identified via distance relative to the center of the globular cluster.

Gaia, or the *Global Astrometric Interferometer for Astrophysics*, is a telescope launched by the European Space Agency (ESA) in 2013 [1]. Its primary objective is to map our galaxy, and it has been very successful in doing so. It measures key information like distance, magnitude, and overall motion of about 2 billion stars [2]. The telescope orbits the Sun about 1.5 million kilometers beyond Earth and consists of two telescopes to collect its precise data. One of its most notable achievements is enabling astrophysicists at the ESA to model what the Milky Way would look like to an external observer, edge-on [1]. Other major contributions include identifying and tracking numerous planets, black holes, and star clusters within the Milky Way.

This project focuses on one of those clusters in particular: NGC 6397. This cluster is reported to be approximately 7800 light-years from Earth, making it the second nearest globular cluster to our solar system [3]. Globular clusters are defined by their spherical, densely packed nature, in contrast to open clusters, which contain a looser collection of younger stars. According to a report by NASA and ESA, NGC 6397 is estimated to be 13.4 billion years old [4]. To put that into perspective, the universe itself is thought to be 13.8 billion years old [5]. This aligns with the nature of globular clusters, which tend to contain some of the oldest stars in any galaxy. Through this project, we aim to gain a better understanding of NGC 6397, with particular focus on the color-magnitude diagram and its main sequence turnoff, which should reflect the cluster's ancient age.

II. Constructing the Colour Magnitude Diagram (CMD)

A. Star Catalogue and Filtering

The star catalogue came from the Gaia telescope dataset. After submitting a search request for NGC 6397 with a 15 arcmin radius, 66321 stars were returned and stored as a csv file. Then by leveraging python pandas, the csv file was converted to a dataframe. To maximize the amount of accurate data in the star catalogue, any records containing null values were removed. It came to my attention that some parallax values were negative due to measurement errors and so these stars were also removed. The final step in filtration was calculating distance (in pc) for each star and then removing any stars that are too far from 2500 pc. Initially a range of +/-200pc was selected as the tolerance but it lead to a CMD with too much noise, abstracting the shape of graph. So after iterating through smaller tolerances, a final range of +/- 100 pc was chosen. Therefore, any star whose distance is less than 2400 or more than 2600 was excluded from the catalogue. The resulting 2393 stars are estimated to make up the cluster.

B. Example Calculations

1) Calculation of Star Distance: To calculate the distance of each star, the following relation was used:

$$\text{distance (pc)} = \frac{1}{\text{parallax ('')}} \tag{1}$$

Example:

Star ID: 1636148068921376768

Parallax [mas]: 0.409363

Step 1: Convert mas to arcseconds ("):

$$\begin{aligned} 0.409363 \text{ mas} &= \frac{0.409363}{1000} \text{ ('')} \\ &= 0.000409363 \text{ ('')} \end{aligned}$$

Step 2: Convert to distance (pc):

$$\begin{aligned} \text{distance (pc)} &= \frac{1}{0.000409363} \\ &= 2442.821 \text{ pc} \end{aligned}$$

2) Determining Star Membership: To determine if a star was within the distance range of 2500 pc ± 100 pc, the following calculation was applied:

Step 1: Calculate upper and lower limits:

$$\text{Upper Limit} = 2500 \text{ pc} + 100 \text{ pc} = 2600 \text{ pc}$$
$$\text{Lower Limit} = 2500 \text{ pc} - 100 \text{ pc} = 2400 \text{ pc}$$

Step 2: Compare star distance with calculated bounds:

Star ID: 1636148068921376768

Star Distance: 2442.821 pc

Conclusion: Since 2400 pc < 2442.821 pc < 2600 pc, this star is retained in the cluster's star catalogue.

C. Average Distance Calculation

By storing all of the data in a python pandas dataframe, the average distance was calculated by using the "mean" method on the distance column. As the name implies it calculates the mean of all values in the column. The results indicate an average distance of **2501.918 pc**.

D. Proper Motion Analysis

An alternative method to identify which stars are located in the cluster is to observe the stars' proper motion by plotting their proper motion in the right ascention (PMRA) and declination (PMDEC) directions. Such plots are shown below, contrasting the data pre and post filtration.

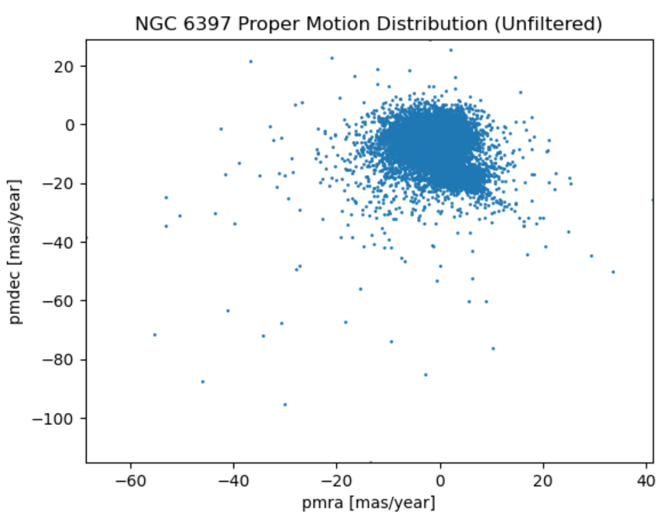

Fig. 1. Proper motion of all stars collected from Gaia dataset prior to any form of filtering

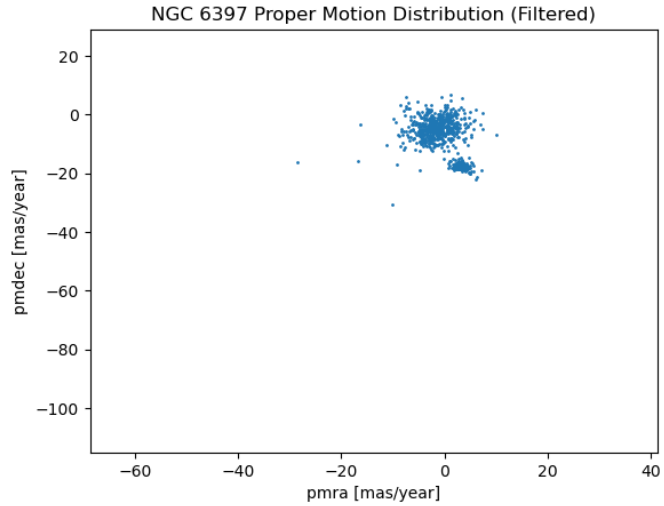

Fig. 2. Proper motion of all stars collected from Gaia dataset after completing filtering process

The key difference between the filtered plot and the unfiltered plot is the change in noise and decrease in size. The filtered plot of proper motion has significantly less outliers (ie. noise) and is distributed over a much smaller radius of varying proper motion. This implies the stars are more similar in motion than the unfiltered plot. Hence it is apparent that proper motion is a sufficient alternative method of identifying which stars are in a cluster.

E. Colour Magnitude Diagrams

The primary method of identifying stars from NGC 6397 and the focus of this study is to analyze the Colour-Magnitude Diagram. Below is two implementations of the diagram- one using absolute magnitude and the other using apparent magnitude.

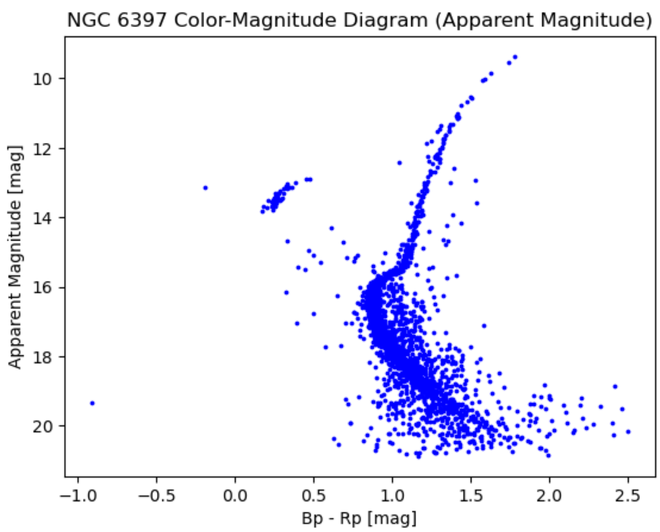

Fig. 3. Colour-Magnitude Diagram of the the calculated star catalog for NGC 6397. This chart plots the BP-RP Colour of each star againts its apparent magnitude.

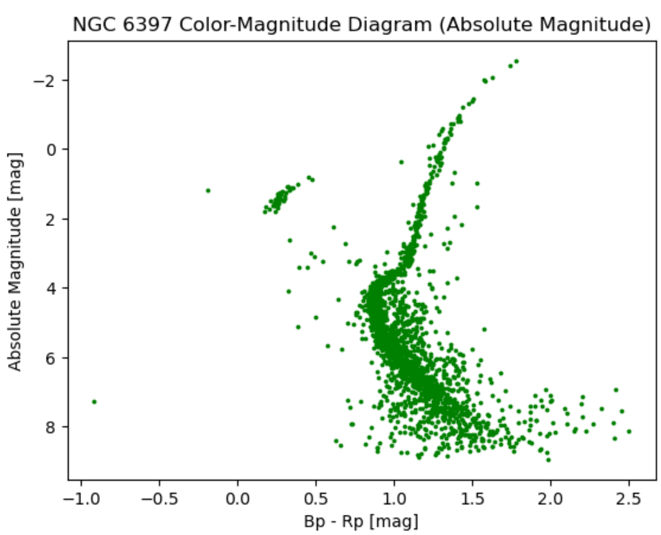

Fig. 4. Colour-Magnitude Diagram of the the calculated star catalog for NGC 6397. This chart plots the BP-RP Colour of each star againts its absolute magnitude.

Sample Calculation for Absolute Magnitude: To calculate the absolute magnitude of each star, the following relation was used:

$$m - M = 5\log_{10}\left(\frac{d}{10\,\text{pc}}\right)$$

Example:

Star ID: 1636148068921376768

Apparent Magnitude (m): 15.3

Distance (d): 2442.821 pc

Step 1: Calculate the logarithmic term:

$$\begin{aligned} 5\log_{10}\left(\frac{2442.821}{10}\right) &= 5\log_{10}(244.2821) \\ &= 5 \times 2.388 \\ &= 11.94 \end{aligned}$$

Step 2: Solve for absolute magnitude (M):

$$\begin{aligned} M &= m - 11.94 \\ &= 15.3 - 11.94 \\ &= 3.36 \end{aligned}$$

Thus, the absolute magnitude of this star is $M = 3.36$.

F. Identification of Stellar Evolution Stages

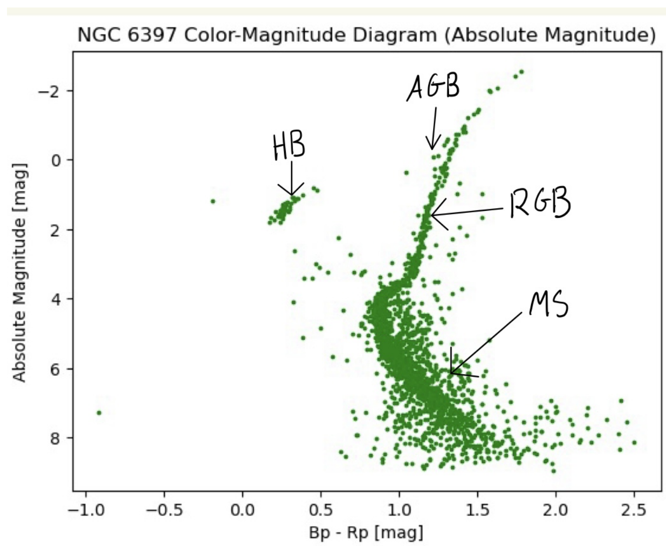

Fig. 5. Labelled Colour-Magnitude Diagram of the the calculated star catalog for NGC 6397. Here the Main Sequence (MS), Red Giant Branch (RGB), Asymptotic Giant Branch (AGB), and Horizontal Branch (HB) are identified.

As shown above all of the expected branches are easily identifiable by the naked eye. The majority of stars appear to still be in the main sequence, however it is also clear that this cluster is arguably older as the horizontal branch is also relatively populated. This is as expected from lecture content as its mentioned that globular clusters tend to be much older clusters and so one would expect to see stars in later stages of stellar evolution.

III. Analysis of the Globular Cluster

A. Main Sequence Turnoff (MSTO)

The main sequence turnoff on a CMD is where the stars from the main sequence begin to branch off to the right (i.e., higher magnitude and greater BP-RP values) into the RGB. For NGC 6397 — as shown in Figures 4 and 5 — the MSTO occurs at an absolute magnitude of about 3.9 mag and a BP-RP colour of about 0.9 mag.

B. Age Estimation

To deduce the age of NGC 6397 using the absolute magnitude CMD, the relationship between flux and magnitude is given by:

$$M - M_\odot = -2.5\log_{10}\left(\frac{F}{F_\odot}\right)$$

Step 1: Rearranging in terms of luminosity, using the relation $F \propto \frac{L}{d^2}$, and noting that both M and $M_\odot$ are absolute magnitudes (i.e., measured at the same standard distance of 10 pc), we get:

$$M - M_\odot = -2.5\log_{10}\left(\frac{L}{L_\odot}\right)$$

Step 2: Solving for $\frac{L}{L_\odot}$:

$$\frac{L}{L_\odot} = 10^{(M_\odot - M)/2.5}$$

Step 3: Using the mass-luminosity relation for main sequence stars:

$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^4$$

Step 4: Equating the two expressions and solving for $\frac{M}{M_\odot}$:

$$10^{(M_\odot - M)/2.5} = \left(\frac{M}{M_\odot}\right)^4$$

$$\frac{M}{M_\odot} = \left(10^{(M_\odot - M)/2.5}\right)^{1/4}$$

Step 5: The main sequence lifetime is approximated by:

$$t_{\text{MS}} = 10\,\text{Gyr}\left(\frac{M}{M_\odot}\right)^{-3}$$

Substituting the expression for $\frac{M}{M_\odot}$:

$$t_{\text{MS}} = 10\,\text{Gyr}\cdot\left(\left(10^{(M_\odot - M)/2.5}\right)^{1/4}\right)^{-3}$$

$$t_{\text{MS}} = 10\,\text{Gyr}\cdot 10^{-3(M_\odot - M)/10}$$

Step 6: Plugging in the known values:

$$M_\odot = 4.68 \quad \text{and} \quad M_{\text{MSTO}} = 3.9$$

$$t_{\text{MS}} = 10\,\text{Gyr}\cdot 10^{-3(4.68-3.9)/10}$$

$$t_{\text{MS}} = 10\,\text{Gyr}\cdot 10^{-0.234}$$

$$t_{\text{MS}} \approx 10\,\text{Gyr}\cdot 0.583$$

$$\boxed{t_{\text{MS}} \approx 5.83\,\text{Gyr}}$$

Thus, the estimated age of the globular cluster NGC 6397 is approximately **5.83 Gyr**. According to NASA, NGC 6397 is calculated to be approximately 13.4 Gyr old [4]. Therefore, our calculated age is only about half of the true age of this cluster.

To put this in perspective, consider the reverse calculation assuming an age of 13.4 Gyr, starting with the following relation:

$$t_{\text{MS}} = 10\,\text{Gyr}\cdot 10^{-3(M_\odot - M)/10}$$

Step 1: Divide both sides by 10 Gyr:

$$\frac{t_{\text{MS}}}{10} = 10^{-3(M_\odot - M)/10}$$

$$\frac{13.4}{10} = 10^{-3(M_\odot - M)/10}$$

$$1.34 = 10^{-3(M_\odot - M)/10}$$

Step 2: Take the base-10 logarithm of both sides:

$$\log_{10}(1.34) = -\frac{3(M_\odot - M)}{10}$$

$$0.1271 = -\frac{3(M_\odot - M)}{10}$$

Step 3: Solve for M:

$$M_\odot - M = -0.4236$$

$$M = M_\odot + 0.4236$$

$$M = 4.68 + 0.4236 = 5.10$$

Therefore, if the globular cluster NGC 6397 is approximately **13.4 Gyr** old, then the corresponding Main Sequence Turnoff should occur at an absolute magnitude of approximately:

$$\boxed{M_{\text{MSTO}} \approx 5.10\,\text{mag}}$$

This is about 1 mag greater than the approximate MSTO observed in figures 4 and 5.

C. Distance Using TRGB

Another method to estimate the distance to a globular cluster is through the Tip of the Red Giant Branch (TRGB). As the name implies, this is the lowest apparent magnitude star in the RGB. Here it is assumed to have an absolute magnitude of $M_{\text{TRGB}} \approx -3$ mag in the G-band.

Step 1: From the CMD of NGC 6397, the apparent magnitude of the TRGB is approximated as:

$$m_{\text{TRGB}} = 9$$

Step 2: Apply the distance modulus formula:

$$m - M = 5\log_{10}(\frac{d}{10(pc)})$$

Step 3: Rearranged to solve for distance d:

$$d = 10^{\frac{m-M}{5}+1}$$

Substituting values:

$$d = 10^{\frac{9+3}{5}+1} = 10^{\frac{12}{5}+1} = 10^{3.4}$$

$$d \approx 2511.886\,\text{pc}$$

Step 4: Compare with the distance calculated from Gaia parallax data:

$$d_{\text{parallax}} = 2501.918\,\text{pc}$$

The average distance for stars in NGC 6397 is 2511.886 as calculated in section C. This means that the TRGB method implies a distance of only 10 pc greater than the average. This minor discrepancy makes sense since the average distance is only an average an so a 0.4% difference in distance is still in line with expectations.

IV. Conclusions

The objective of this report was to analyze NGC 6397 and drive insights about its age, CMD, and general attributes. The CMD created for NGC 6397 in both apparent and absolute magnitude relfect the general shape of a CMD as it is very similar to examples shown in course content. The MS, RGB, AGB, and HB are very easily identified at first glance, with the slight exception to the AGB as it seems to be less densly populated. The data used for this analysis is certainly credible as it comes from well established professional organizations with advanced technology, experience, and telescopes. However, the age calculated based on the MSTO is significantly different from NASA's reported age. Our calculated age is about 5.8 billion years old, while NASA reports 13.4 billion years old [4]. Therefore, it is important to investigate what potential sources of error could cause this discrepancy.

By working backwards to isolate for the magnitude of the MSTO according to NASA's reported age, we expect the MSTO to occur at 5.1 mag. For visual comparsion, consult the following image.

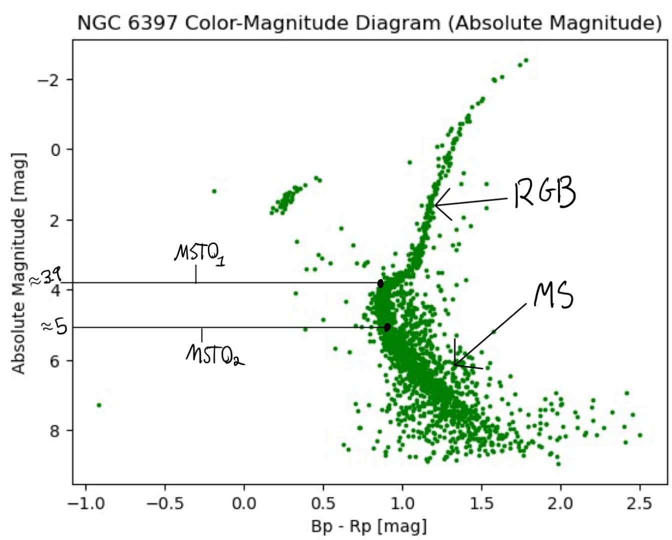

Fig. 6. Labeled Colour-Magnitude Diagram of the calculated star catalog for NGC 6397. Here, $MSTO_1$ represents the MSTO approximated from the CMD, and $MSTO_2$ corresponds to the MSTO derived using the literature age from NASA.

There are two main reasons that could be the cause of this discrepancy.

Firstly, there is apparent noise in the dataset. Although the filtration process attempts to remove many of these stars with a large degree of error in their parallax by removing negative parallax values, there are likely still many stars with large measurement errors. So including stars with parallax values that have large magnitudes of error could mean we've included some stars in the CMD that aren't actually part of NGC 6397.

The magnitude and colour also have error in measurement and so this also can contribute to skewing the true location of the MSTO in the CMD. Therefore, noise could be a large contribution to observing a seemingly incorrect MSTO with a seemingly lower absolute magnitude than.

Contrarily, the MSTO method of aging a cluster is also an approximation itself. It assumes perfect proportionalities and power laws. In reality, various other factors can affect the overall geometry of the CMD and stellar evolution. A paper from Cornell University highlights some examples of imperfections in this methodology by looking at extended MSTOs [6]. Therefore, it's possible that NASA used other calculations and methodologies to get a more accurate age of 13.4 billion years. An example of other methods to date a cluster via CMDs is fitting isochrones to the data. It's possible that these models pick up on other features of the data missed by the MSTO alone.

Therefore in the future there are some key changes that can be made to the study that could result in more favourable findings.

Firstly, a more rigorous approach could be taken in the data cleaning process. Stars with large relative error in apparent magnitude, parallax, and/or colour could be removed based on some margin of tolerated error. This could aid in removing inaccurate values that skew the location of the MSTO. Additionally, the filtering process could also incorporate considering proper motion. This could be used in conjunction with more rigorous data cleaning by using clustering algorithms to select stars with similar proper motion.

Secondly, isochrones or other alternative dating methods could be applied to the data. By fitting isochrone models, its possible that more accurate aging results could occur beyond approximations using the MSTO alone.

The results of this study drew key insights into the process of aging a globular cluster such as NGC 6397 using public accessible data from professional astro physicists and organizations. However, if the study were to be re conducted implementing a more rigorous data cleaning process and experimenting with other cluster dating methodologies could prove to return more accurate star cluster results.

References

[1] E. S. Agency, "Gaia overview," https://www.esa.int/Science_Exploration/Space_Science/Gaia/Overview, 2022, accessed: 2025-04-04.

[2] C. Radio, "The gaia spacecraft is orbiting into retirement," https://www.cbc.ca/radio/quirks/the-gaia-spacecraft-is-orbiting-into-retirement-1.7129944, 2024, accessed: 2025-04-04.

[3] E. S. Agency, "Euclid's view of globular cluster ngc 6397," https://www.esa.int/Science_Exploration/Space_Science/Euclid/Euclid_s_view_of_globular_cluster_NGC_6397, 2024, accessed: 2025-04-04.

[4] NASA, "Hubble finds ancient globular cluster 12.4 billion years old," https://www.nasa.gov/feature/goddard/2021/hubble-finds-ancient-globular-cluster-12-4-billion-years-old, 2021, accessed: 2025-04-04.

[5] N. S. Place, "What is the big bang?" https://spaceplace.nasa.gov/big-bang/en/, 2022, accessed: 2025-04-04.

[6] R. C. Paul Goudfrooij, "Extended main sequence turnoffs in intermediate-age star clusters: Stellar rotation diminishes but does not eliminate age spreads," *arXiv preprint arXiv:1707.03847*, 2017, accessed: 2025-04-04. [Online]. Available: https://arxiv.org/abs/1707.03847