



Contribution ID: 18

Type: Hackathon project proposal

## STARFINDER - Machine Learning Techniques for Evaluating Globular Cluster's Stars Membership Probabilities

Friday, 4 October 2024 10:40 (20 minutes)

### Abstract

Globular clusters (GCs), spheroidal conglomeration of stars tightly bound together by means of gravitational force, are among the oldest objects that live within our galaxy. A key characteristic of these objects is their high density, significantly greater than the average galactic star density (between  $\sim 10^4$  to  $\sim 10^6$  stars within a spheroid of radius up to  $\sim 100 pc$ , in stark contrast to the local average stellar density of about  $\sim 1 - 2 \frac{\text{stars}}{pc^3}$ ), so that they can be considered collisional systems. The ESA's Gaia (Global Astrometric Interferometer for Astrophysics) mission, which has mapped nearly 2 billion stars in our galaxy up to its third data release, provides the largest set of high-resolution data available, enabling the detailed study of GCs' internal dynamics.

However, the high density of these regions presents a challenge for Gaia's 1.45-meter primary mirror, often resulting in compromised data quality and insufficient resolution. Consequently, accurately associating stars with clusters becomes difficult due to poor estimates and high errors in the parameters.

Machine Learning (ML) algorithms offer a promising solution to this problem. As demonstrated in referenced paper [1], techniques inspired by ML such as Mixture Modelling, which uses Markov-Chain Monte Carlo, Extreme Deconvolution and Maximum Likelihood Estimation, can be employed to infer the general distribution properties of the cluster, distinguishing them from field star distributions. Enhancing these methodologies with neural networks such as Generative Adversarial Networks, which could be used to simulate stellar populations based on observational data, would allow for the assignment of membership probabilities to each source in the sample, significantly increasing the number of sources available, up to a factor of  $10^2$ , and thereby enhancing the statistical robustness of subsequent astrophysical analyses.

### References

[1] Vasiliev, Baumgardt (2021). \emph{Gaia EDR3 view on Galactic globular clusters}; MNRAS 505, 5978–6002

### Project proposal: general context

Globular clusters (GCs) are among the oldest and most densely populated stellar systems in our galaxy, offering unique opportunities to study stellar dynamics and galactic evolution. The European Space Agency's Gaia mission has provided extensive high-resolution data on nearly 2 billion stars, enabling detailed investigation of these clusters. However, the high density of stars within GCs presents significant observational challenges, mainly for the quality of the data.

### Project proposal: description of the problem

The primary challenge in studying GCs using Gaia data is the high stellar density, which often results in compromised data quality and insufficient resolution. This limitation makes it difficult to accurately associate

stars with their respective clusters, leading to poor parameter estimates and high error margins. Addressing these issues is crucial for enhancing the reliability of astrophysical analyses and improving our understanding of stellar dynamics within GCs.

## **Machine learning methods**

To overcome these challenges, the project proposes the application of advanced Machine Learning (ML) techniques. Specifically, Generative Adversarial Networks simulations, Markov-Chain Monte Carlo (MCMC) simulations within Mixture modelling, together, with Extreme Deconvolution and Maximum Likelihood Estimation, will be employed to analyze Gaia's data. These methods will help infer the distribution properties of clusters and distinguish cluster stars from field stars.

## **Input dataset**

The necessary data can be easily accessed through the 'astroquery' Python library, which can set up for Gaia data. Alternatively, a Python package that I am currently developing, which will be ready or in the final stages of completion by the time of the event, can be used. This package will be specialized in acquiring and analyzing Gaia's Globular Cluster Data.

The number of sources can vary from  $10^3$  to  $10^5$  entries, based on the search parameters and the chosen cluster, and all the data about these are packed and stored in 'astropy tables'.

**Primary author:** FERRAIUOLO, Pietro (INAF - Osservatorio Astrofisico di Arcetri)

**Co-author:** Mr MENESSINI, Matteo (INAF - Osservatorio Astrofisico di Arcetri)

**Session Classification:** Hackathon