UNIVERSITY OF CAMBRIDGE

MPhil in Data Intensive Science

Easter Term 2025

Galactic Archaeology Module Coursework

A. Ardern-Arentsen, E. Sola, G. Myeong, V. Belokurov

*Attempt **all** parts of the coursework.*

The anticipated marks for each part are not explicitly listed here; instead, you should address each point of assessment thoroughly in both the code and the report. The relative weights between Task 1 (Constructing the all-sky Gaia map) and Task 2 (Identify and characterise substructures) will be 40% and 60%, respectively.

This coursework should be submitted via a GitLab repository which will be created for you. Place all of your code and your report into this repository. The report should be in PDF format and placed in a folder named `report/`*. You will be provided access to your repository until 11:59pm on Friday April 4, after which your work will be deemed ready to assess.*

Please do not commit any large data files to the repository. Instead, create a `data/` *folder and include within it a text file listing the data files needed to run your code. Also, do not commit large temporary/output files. Instead, create an* `outputs/` *folder for them.*

You are expected to submit code and associated material that demonstrates good software development practices, as covered in the Research Computing module.

Your report should not exceed 3000 words (including tables, figure captions, and appendices, but excluding references). You are reminded to comply with the requirements given in the Course Handbook regarding the use, and declaration of use, of any autogeneration tools.

© 2025 University of Cambridge

1 Introduction

The Aim

Optimise the construction of an all-sky Gaia map to highlight distant substructures in the Milky Way's halo. Identify and characterise the visible substructures.

Background

In class, we have discussed how to use Gaia data to create an all-sky map of the Galactic stellar halo. This coursework focuses on improving the implementation of the basic map introduced in class. Given the depth of Gaia data, it is not possible to directly replicate the approach of the original *Field of Streams* map, which revealed faint, low-surface-brightness structures in the Milky Way's halo (see, for example, the introduction lectures by V. Belokurov). Additionally, the original SDSS-based map only covered high-latitude regions, minimising contamination from disk stars and Galactic dust. In contrast, an all-sky Gaia map presents additional challenges, including contamination from nearby, younger stars and the effects of dust extinction, both of which need to be mitigated.

Mapping Approach

In the original SDSS map, the false-colour composite used Red/Green/Blue (R/G/B) channels to represent the density of Main Sequence (MS) and Main Sequence Turn-Off (MSTO) stars in three apparent magnitude bins. Since MS/MSTO stars have similar absolute magnitudes, their apparent magnitudes serve as proxies for distance, meaning the colour of each pixel in the composite reflects the distance of halo substructures. For this assignment, MS/MSTO stars are too faint for Gaia at the distances probed by SDSS, so Red Giant Branch (RGB) stars must be used instead. However, selecting a clean and efficient sample of RGB stars in Gaia is non-trivial, especially when aiming to cover a broad range of heliocentric distances.

Tasks

1. ***Construct the All-Sky Gaia Map***
   - Implement an optimised method to visualise the Galactic stellar halo while minimising contamination and dust effects.
   - Ensure the map effectively highlights distant substructures.
2. ***Identify and Characterise Substructures***
   - Automatically extract, identify, and label as many known substructures as possible.
   - Investigate their metallicities.
   - Include satellites (such as globular clusters and dwarf galaxies) and the Sagittarius stream.

Important Note

The Gaia database is extremely large. Attempting to download the entire catalog is highly inefficient and can cause your code to grind to a halt. Optimise your data access strategies accordingly.

© 2025 University of Cambridge

2 Assessment criteria

Big Picture

The goal of this coursework is to demonstrate your familiarity with large astronomical datasets and your understanding of stellar evolution and the structure of the Milky Way. Ideally, your work should also showcase an intuitive grasp of the fundamental components required for Galactic Archaeology.

In the final image you create, key halo substructures—including the Sagittarius Dwarf Galaxy, the Sagittarius Stream, the Large and Small Magellanic Clouds, and various small satellites (such as globular clusters and dwarf galaxies)—should be clearly visible.

Detailed points of assessment

Recreating the Halo Density Map

- Using Gaia data, reproduce a halo density map following the example discussed in class.
- Divide the selected stars into three groups based on your chosen criterion and use them to construct a three-channel false-colour R/G/B composite (each channel representing a different density map).
- Apply this approach to the original Field of Streams region before extending it to the full sky.

Optimizing Red Giant Selection

- Improve the identification of red giant stars.
- Clearly explain and justify your selection criteria, which may involve photometry (colours, magnitudes), astrometry (parallaxes, proper motions), or a combination of both.
- Estimate the purity and completeness of your selection.
- Determine the distance range probed by your selected red giants and explain the reasoning behind your choices.

Enhancing Map Visualization

- Optimise the density contrast to reveal substructures at different surface brightness levels.
- Experiment with and apply different colour schemes for the R/G/B mapping.
- Identify the colour mapping that best highlights the Galactic halo substructures and discuss your optimisations in the report.

Detecting and Identifying Substructures

- Automatically detect small-scale over-densities corresponding to globular clusters and dwarf galaxies.
- Compare your detections with established databases of Galactic satellites.

© 2025 University of Cambridge

Identifying Stellar Streams

- Develop an effective method for identifying stellar streams.
- Clearly explain your detection strategy in the report.

Comparing with APOGEE DR17 and Literature

- Cross-match your identified structures with the APOGEE DR17 catalog.
- Compare and discuss the metallicities of your detected structures using literature sources (e.g., the Harris 2010 catalog for globular clusters).
- Your findings should include at least NGC 1851, M3, NGC 5053, and the Sagittarius Dwarf Galaxy.

END OF PAPER

© 2025 University of Cambridge