Finding stars that are likely to be from accreted satellite galaxies

Helena Cronstedt

Division of Astrophysics
Department of Physics
Lund University

2023-EXA204

Degree project of 15 higher education credits
June 2023

Supervisor: Paul McMillan

Division of Astrophysics
Department of Physics
Box 43
SE-221 00 Lund
Sweden

Abstract

The Milky Way galaxy has through its lifetime merged with several smaller galaxies in its surroundings. Merger events like these play an important role in the formation of galaxies and are therefore of great interest to better our understanding of galaxy formation. In this project we create candidate lists of stars that are likely to originate from past merger events experienced by the Galaxy, using the method and results of Mikkola et al. (2023).

Merging stars tend to follow the trajectory of their original galaxy, even as this system is pulled apart by the larger galaxy. With time the merged galaxy will no longer be spatially coherent but its stars can instead be found as overdensities in velocity space. The full 3D velocity of stars is needed to identify stars within these overdensities, but only a small number of the observed stars have measured radial velocities. Mikkola et al. (2023) therefore used a maximum-likelihood method to obtain a full 3D velocity distribution of stars in the solar neighbourhood, without using radial velocities.

In this project we use this velocity distribution to identify overdensities, corresponding to the four known moving groups Gaia Sausage Enceladus (GSE), ED-2, MMH-1 and the Helmi streams, and find stars that are likely to be members of them. These lists of probable candidates can be used as a tool for future spectroscopic surveys to efficiently target stars and gain further knowledge about these moving groups, such as their detailed chemical abundances.

Populärvetenskaplig beskrivning

Precis som arkeologer gräver i jordens historia finns det astrofysiker som gräver i Vintergatans förflutna. Det finns ingen fullständig bild över Vintergatans evolution. Därför är galaxformationen ett relevant ämne inom astrofysiken. När man grävt i ljusspektra och rörelsedata av stjärnor i vår galax har man funnit att många stjärnor har sitt ursprung utanför Vintergatan.

En galax som rör sig genom rymden kan genom gravitationskraften sammanfogas med andra galaxer. Mindre galaxer som är gravitationellt bundna till en större galax, ungefär som jorden är bunden till solen, kallas satellitgalaxer och Vintergatan har genom sin livstid slukat ett antal sådana här galaxer. Ett exempel på en sådan här satellit-galax är Gaia-Enceladus som slukades av Vintergatan för ungefär 10 miljarder år sedan och vars stjärnor är spridda över en stor del av Vintergatans så kallade halo.

Det är möjligt att koppla stjärnor till deras ursprung genom deras rörelse genom galaxen. En mindre galax som slukas av en större galax dras isär av den större galaxen men trots detta tenderar stjärnorna att fortsätta i samma kurs som deras ursprungsgalax. Stjärnorna från den slukade galaxen bildar så kallade strömmar i den större galaxen. Med tiden dras stjärnorna längre och längre isär men deras hastighet förblir väldigt snarlik. Stjärnorna kan kopplas till varandra genom deras rörelse och hastighet även när de efter en tid inte är rumsligt sammanhängande längre.

Det här arbetet syftar till att identifiera stjärnor i solens omgivning som har sitt ursprung utanför Vintergatan. Identifiering av sådana stjärnor kan underlätta framtida forskning om hur slukade sattelitgalaxer har bidragit till utformningen av vår galax.

Contents

Chapter 1

Introduction

As objects travel through space they can merge through gravitational attraction in a process called accretion. In the early Universe, when galaxies were located more closely together, these accretion events were more frequent (Helmi 2020). From the current understanding of galaxy formation it is believed that galaxies form inside dark matter halos and evolve through accretion of gas, dust, stars and other baryonic materials (White & Rees 1978). Gravitationally bound satellite galaxies can be pulled into a more massive galaxy (e.g., Quinn et al. 1993) and these merger events also play a key role in the shaping of a galaxy.

Disk galaxies like the Milky Way consist of a number of distinct components which are represented in Fig. 1.1. There is the central bulge, the thin- and thick disks, the stellar halo and, evidence also suggests, an invisible dark matter halo (Ryden & Peterson 2021). Depending on its size and mass ratio, a merger between two galaxies could significantly change the formation of the larger galaxy. An accreted satellite galaxy could for example thicken the galactic disk of the larger galaxy or even destroy the disk if the accreted galaxy is massive enough (Quinn et al. 1993). It is also believed that a large merger could trigger star formation in a young galaxy (Helmi 2020).

Evidence shows that the Milky Way galaxy has experienced many merger events through its lifetime. Most debris from these accreted satellite galaxies can be found in the stellar halo (Helmi 2008). The halo is an almost spherical distribution of stars that extends from the centre of the galaxy out far beyond the other components. The halo stars have large orbital timescales and tend to have velocities very different from the Sun when passing the solar neighbourhood. External galaxies that have merged with the Milky Way in the past are believed to have been less massive and more metal-poor than the early Milky Way disk (e.g., Helmi 2008). This explains why some of the oldest and most metal-poor stars can be found in the galactic halo (e.g., Howes et al. 2015). Investigating the nature of these mergers experienced by the Milky Way is a necessary step toward learning the history of our home galaxy and understanding galaxy formation.

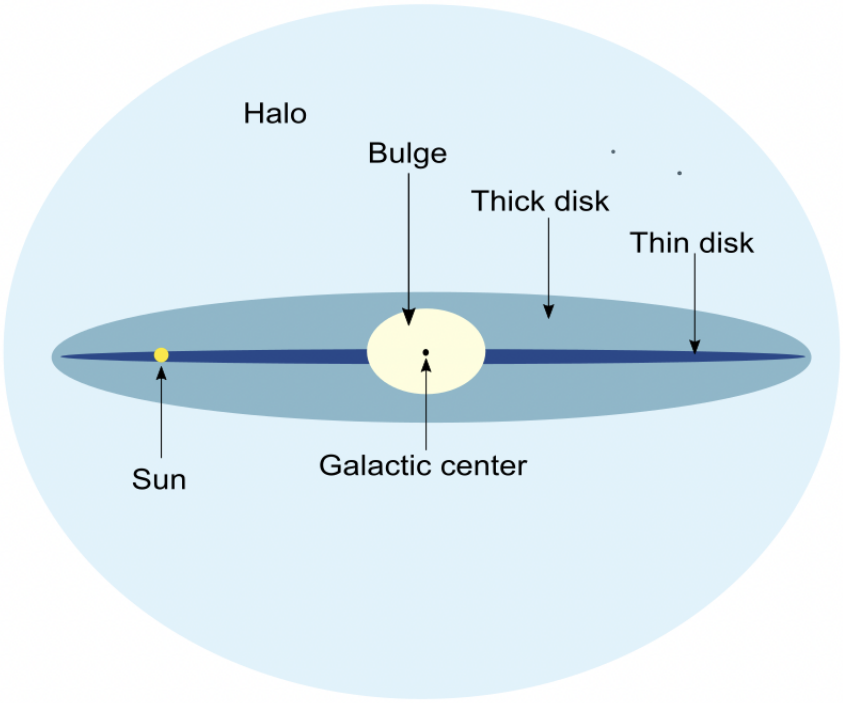


Figure 1.1: The galactic components of the Milky Way.

The archaeological digging in the Galaxy is performed in stellar spectra and kinematic data. This is because stars retain information about their past. Stars form in clouds of dust and gas and their atmospheres will reflect the chemistry of this environment in which they were born. Since some stars have lived for very long their atmospheres can contain information from the very early Universe, and spectroscopy allows us to measure their chemical contents. Stars with similar chemistry tend to share origin (Freeman & Bland-Hawthorn 2002).

The motion of stars also reflects their origin. As a system of stars merges into another galaxy the stars tend to continue on the path of their original system (e.g., Eggen et al. 1962). The system will be stripped apart by the larger galaxy but as the stars continue on the same trajectory they will form streams. The size of the parent system, the time since the merger and the orbital timescales will determine the properties of the stream (Helmi & White 1999). Small systems form very narrow streams while large merging systems form broader streams. A stream can easily be detected as an overdensity in the early stages of a merger when the system is still spatially coherent. Figure 1.2 shows spatially coherent streams in the northern and southern hemisphere. The stars forming a stream have similar velocities and therefore also form overdensities in velocity space (Helmi 2020). Investigating these overdensities can provide information about the merger events and the objects involved.

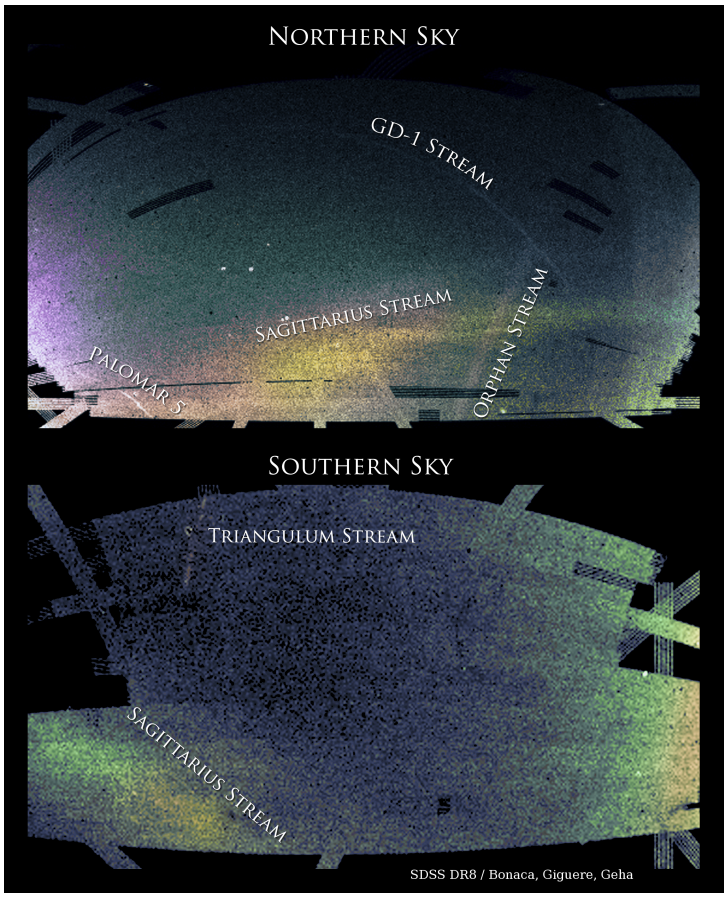


Figure 1.2: Spatially coherent streams from past merger events seen in the northern and southern hemisphere. Figure credit: SDSS DR8 / Bonaca, Giguere, Geha.

1.1 Background

In 2013 the European Space Agency launched a space observatory called Gaia (Gaia Collaboration et al. 2016). The goal of the Gaia mission is to build a large and precise six-dimensional map of the Milky Way galaxy. In 2022 Gaia delivered its third data release *Gaia DR3* (ESA 2023) which provided a large set of accurate kinematic data. Gaia obtains radial velocities from Doppler shifts in stellar spectra (ESA 2019). This requires high resolution spectra and therefore only a fraction of the data have radial velocity measurements.

The data set quickly decreases in size when the selection is restricted to stars with measured radial velocity. Mikkola et al. (2023) therefore implemented the penalized maximum-likelihood method first used by Dehnen (1998), to determine a velocity distribution of stars in the solar neighbourhood without radial velocities. Without the radial velocity restriction a larger set of data could be used. The method can provide a full three dimensional velocity distribution, using only proper motion. In the velocity distribution found by Mikkola et al. (2023) a number of overdensities were found. Many of these overdensities are debris

from known mergers experienced by the Galaxy, but Mikkola et al. (2023) also identified two previously unrecognised overdensities named MMH-1 and MMH-2. The penalized maximum-likelihood method will be further explained in the method section.

1.2 Overdensities

As mentioned above, stars that belong to the same merger event tend to have similar velocities and therefore form high density regions in velocity space. We now present a few, already identified, overdensities in velocity space that will be further investigated in this project.

1.2.1 Gaia-Sausage Enceladus

Around 10 Gyr ago the Milky Way experienced its last known significant merger event (Helmi et al. 2018; Belokurov et al. 2018). From Gaia DR2 data it could be seen that a large fraction of the halo stars in the solar neighbourhood can be connected to one big structure. It was suggested that the stars associated with this structure originate from another galaxy that had mass $\sim 6 \times 10^8 M_\odot$ and that merged with the Milky Way in the past. The chemical abundance and metallicity of stars associated with this structure was investigated with APOGEE data which supported the hypothesis that the stars share origin. The accreted dwarf galaxy got the name Gaia-Enceladus, also called Gaia-Sausage Enceladus, *GSE*, (Helmi et al. 2018). The debris of GSE dominate the galactic halo in the solar neighbourhood and have a mean metallicity, [Fe/H]$\sim$-1.15 (Feuillet et al. 2021).

1.2.2 The Helmi streams

The Helmi streams are another trace of the Milky Way's accretion history. The streams are not spatially coherent but can be seen as overdensities in velocity space (Roederer et al. 2010). Evidence suggests that the streams originate from a dwarf galaxy that had a mass of $\sim 10^8 M_\odot$ and merged around 5-8 Gyr ago (Koppelman et al. 2019). One character of the Helmi streams is that they have two moving groups, one moving with positive- and the other with negative v_Z velocity in the (v_Y, v_Z) plane. In spherical coordinates these two moving groups can be seen in the v_θ range, one moving with positive v_θ and the other with negative v_θ (Mikkola et al. 2023). The stars in the Helmi streams have a metallicity range of approximately [Fe/H]$\sim$[-2.3,-1.0] and the metallicity distribution peaks at [Fe/H]$\sim$-1.5 (Koppelman et al. 2019).

1.2.3 ED-2

ED-2 is a small substructure that can be found in the galactic halo velocity space (Dodd et al. 2023). One defining character of ED-2 is that it is very tightly bound in velocity space. Dodd et al. (2023) identified 32 members of ED-2 but only three of these members had metallicity data. These metallicities are [Fe/H]= -1.88, -2.07 and -2.19. Based on the chemical abundance of the stars and their location in velocity space, ED-2 is believed to come from an accreted system (Dodd et al. 2023).

1.2.4 MMH-1

Two new overdensities were found in the velocity distribution produced by Mikkola et al. (2023). These moving groups got the names MMH-1 and MMH-2 as previously mentioned. MMH-2 can be seen in the (v_ϕ, v_θ) plane at (150,100) km/s but is very difficult to see in the (v_r, v_ϕ) and (v_r, v_θ) planes. MMH-2 will therefore not be discussed further in this project. MMH-1 can be seen at (v_ϕ, v_θ)=(25,300) km/s and is split in two in the v_r range, (v_r, v_θ)=(-250,300) km/s and (v_r, v_θ)=(200,350) km/s (Mikkola et al. 2023). [1]

1.3 This Project

To further investigate the accretion history of the Milky Way, one must look at the accreted stars. This project makes use of the methods and results of Mikkola et al. (2023) to identify members of the overdensities found in the velocity distribution. The goal is to produce candidate lists of probable members to find what stars should be targeted in further investigation of the accretion debris by spectroscopic surveys such as WEAVE (Jin et al. 2023) or 4MOST (de Jong et al. 2019).

[1] Mikkola et al. (2023) use a different sign convention for v_θ and v_ϕ. There will therefore be a sign difference between the velocities given in this project and the velocities given by Mikkola et al. (2023).

Chapter 2

Method

2.1 Selection of data

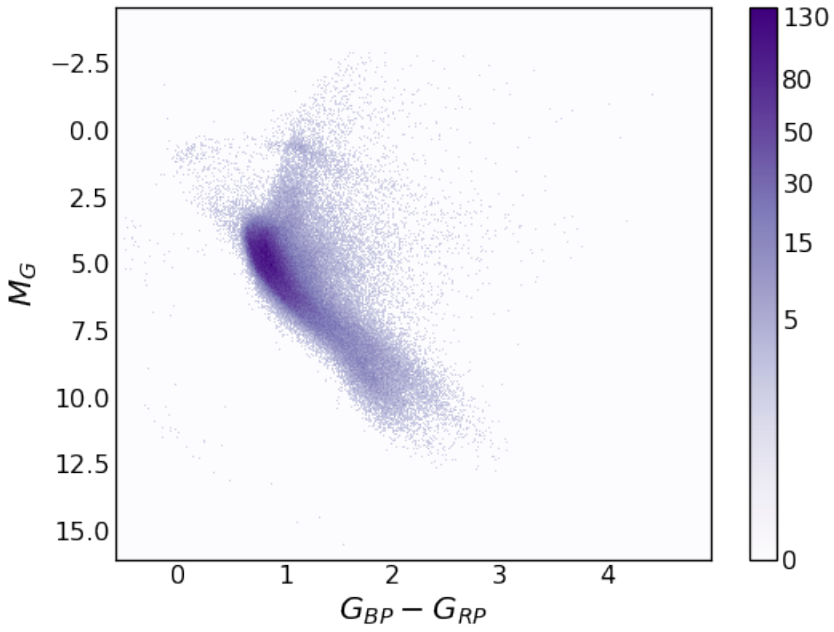


Figure 2.1: Colour-Magnitude diagram of the Solar neighbourhood halo sample. The colour shows the number density of stars.

The data used in this project was collected from the Gaia archive through a similar query as that which was used for the halo sample by Mikkola et al. (2023). The query can be found in Appendix A. The stars selected are all within 3 kpc of the Sun. We make this cut because distance estimations become less accurate the further away we go from the Sun and because we want the velocity distribution to be approximately the same across

the whole volume. Mikkola et al. (2023) required the Gaia measured parallax over error $\varpi/\sigma_\varpi > 10$, to be able to accurately determine absolute magnitudes of stars and divide the data in two colour-magnitude sequences. In this project however, I do not make this division, so accurate parallax data is not as important. We can therefore set a greater parallax uncertainty, $\varpi/\sigma_\varpi > 5$, to select a larger set of halo stars. Figure 2.1 shows a colour-magnitude diagram of all stars in our data set. In this colour-magnitude diagram we can very faintly see these two sequences that Mikkola et al. (2023) wanted to separate.

The halo sample is selected since most of the debris from past mergers can be found in the halo. The halo stars are selected to have a transverse velocity of $v_T > 200$ km/s which removes the majority of the disk stars. This set of data contains 1003656 stars (compared to 433622 stars in the Mikkola et al. (2023) sample) and 70420 of these stars have radial velocities.

607030 of the stars in our sample have measured metallicities from Andrae et al. (2023)[1], obtained through low-resolution spectrophotometry published in Gaia DR3. These stars can be used to look at the metallicity distribution of the stars in the overdensities.

2.2 Penalized maximum-likelihood method

This project builds on the penalized maximum-likelihood method used by Mikkola et al. (2023). In the implementation of this method, velocity space with spherical polar directions (v_r, v_θ, v_ϕ)[2] relative to the galactic centre is divided into a three dimensional grid of cells with widths h_{v_r}, h_{v_θ} and h_{v_ϕ} in the respective directions. Each cell l has an associated element of the velocity distribution, $f_l(v) = e^{\phi_l(v)}$, where $\phi_l(v)$ is the logarithm of the probability density. The tangential motion of a star, k, is projected into 3D space creating a point in the grid. Radial velocities, $v_\parallel$, of most stars in the set are unknown which means that the star can be anywhere along the corresponding direction in velocity space. A stars motion can therefore be seen as a line, $v = p_k + v_\parallel \hat{r}$, through the range of the velocity grid. Here p_k is the tangential motion of a star projected into 3D and $\hat{r}$ is the line-of-sight velocity direction. All velocities are converted to their Galactocentric components using the Astropy default values (Astropy Collaboration et al. 2022).

The line formed by a star's velocity will have a specific length $K(k|l)$ in each cell of the grid. Most of these length values are zero since the stars possible motion only crosses a fraction of the cells. The product of the probability density and the length of the line formed by a stars motion through a cell summed over all cells gives a probability distribution for the

[1]Available from: https://doi.org/10.5281/zenodo.7599788

[2]In this project, we both talk about the radial component of a star's velocity in these coordinates (towards or away from the Galactic centre) and the measured radial velocity of a star (toward or away from the Sun). This can be very confusing and we will therefore use different notations: v_r for velocity relative to the Galactic centre and $v_\parallel$ for velocity relative to the Sun.

stars velocity,

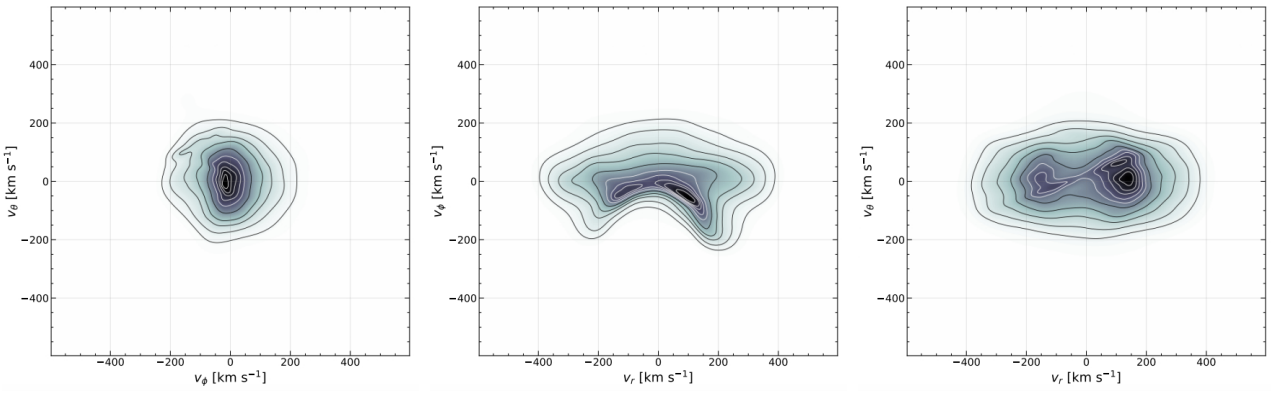
$$\sum_l e^{\phi_l} K(k|l). \tag{2.1}$$

This probability distribution is calculated for all stars in the sample. The penalized maximum-likelihood estimation for the velocity distribution of the full data set is then described by,

$$\mathscr{L}_\alpha(\phi) = N^{-1} \sum_k ln \left[\sum_l e^{\phi_l} K(k|l) \right] - \sum_l e^{\phi_l} - \frac{1}{2} \alpha h_{v_r} h_{v_\theta} h_{v_\phi} \sum_l \left(\sum_n \phi_n \Xi_{nl} \right). \tag{2.2}$$

The first term in Eq. 2.2 is the sum of the probability distribution over all stars in the set divided by the sample size N. The second term in Eq. 2.2 is a normalizing term to ensure that the total probability distribution adds up to one. The last term in the estimation is a penalizing term to smoothen the probability distribution. In this penalizing term, Ξ is related to the second order derivative of $\phi(v)$ in a given cell and α is a smoothing parameter.

In short, the method uses the length of the lines formed by the unknown radial velocity and the probability density of the cells to calculate the probability that a star has a certain velocity and thus belongs to a specific bin in the grid. Through these calculations a three dimensional velocity distribution can be obtained. In the velocity distribution obtained by Mikkola et al. (2023) a number of overdensities were found. Projections of the full velocity distribution of the data sample used in this project can be seen in Fig. 2.2.



((a)) The probability distribution in the (v_θ, v_ϕ) plane.

((b)) The probability distribution in the (v_ϕ, v_r) plane.

((c)) The probability distribution in the (v_θ, v_r) plane.

Figure 2.2: The full probability distribution of the data set in spherical coordinates.

2.3 Candidate lists

An overdensity associated with a single merger, can be found over a range of velocities in velocity space. The overdensity will in other words, occupy a number of bins in the 3D-

grid. The probability distribution of a stars velocity within the range of the overdensity, denoted $*$, is thus,

$$\sum_{l^*} e^{\phi_l^*} K(k|l^*). \tag{2.3}$$

By dividing Eq. 2.3 by the probability distribution of the full velocity grid one can find the probability that the star belongs to the velocity range of the overdensity (Mikkola et al. 2023).

$$P_k(l^*) = \frac{\sum_{l^*} e^{\phi_l^*} K(k|l^*)}{\sum_l e^{\phi_l} K(k|l)}. \tag{2.4}$$

The numerator represents the probability distribution restricted to the cells enclosing the overdensity in velocity space and the denominator is the probability distribution of a star's velocity over all cells in the grid. Stars that have a high probability, $P_k(l^*)$, to belong to the velocity range of the overdensity are possible members of the velocity feature, that can be further investigated.

To use Eq. 2.4 the dimensions of the overdensity must be known. The velocity distribution is plotted in two-dimensional planes, but since we have a full 3D distribution we can put these planes together as slices of the third dimension. By varying the velocity range and the size of the slices it is possible to study the distribution in closer detail. Through these density maps we can find where a high density region can be seen in the sliced dimension and thus put limits to its velocity range in that dimension. Determining this range in all three dimensions encloses the overdensity in velocity space. With the velocity ranges of the overdensity it is possible to locate the feature in the 3D grid and obtain the numerator of Eq. 2.4. By finding the probability that a star belongs to these velocity ranges, I will create candidate lists containing probable members of the remains of the merged systems.

2.3.1 Result checkup

We want to ensure that the candidate lists we create are trustworthy and that the results are comparable to earlier research about the moving groups.

One concern regarding our approach of creating the candidate lists is that the method builds on a circular argument. The candidate lists are created by calculating the probability that a star belongs to a certain feature which itself is a probability distribution in velocity space in which the star is included. It is therefore important to test the results to see if the method is valid. One way of testing the results is to see how dependent the method is on individual stars. This can be done by first removing some of the most probable members of the overdensities from the data set and exclude them from the probability distribution calculations. The stars can then be put back into the membership probability

calculations to see if this changes the results significantly.

The penalized maximum-likelihood method is implemented to not be restricted to stars with measured radial velocities. Without radial velocities we can only find the probability that a star belongs to an overdensity and not say if it is really true. But the stars that *do* have a measured radial velocity can be used as a comparison to the overdensity candidates. These stars have a known location in velocity space and not just a probability distribution. This makes it possible to say if the stars truly do lie within the velocity range of the overdensity.

Even if the stars have a high probability to be within the range of the overdensity, it is not for sure that they are members of the merger event. It is therefore of great interest to not only look at the kinematics of the stars, but also look at some chemical properties. The stars in our data set that have measured metallicities from Andrae et al. (2023) can be used to calculate a metallicity distribution of probable overdensity candidates. This metallicity distribution can be compared to the metallicity of the overdensities obtained from previous research.

Chapter 3

Results

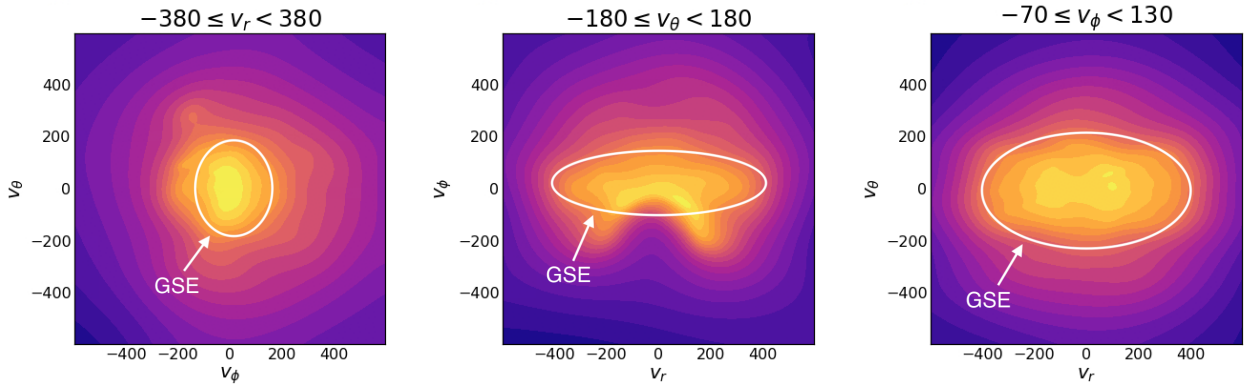
3.1 Velocity ranges of overdensities

Since we have a full 3D velocity distribution, we can divide it into slices and study the distribution in more detail. The velocity distribution is plotted in two-dimensional planes that can be summed together in the third dimension. Each slice can then be studied individually. I changed between different initial velocities and varied the size of the slices to find high density regions. By carefully looking through these slices I found start- and endpoints of the overdensities in the velocity distribution.

Through this process of looking through slices of the velocity distribution it was possible to find the velocity ranges of the overdensities presented in the introduction. These ranges are presented in Table 3.1. It should be noted that the Helmi streams have two separate v_θ ranges and MMH-1 have two separate v_r ranges. Fig. 3.1 to Fig. 3.4 show the summed slices of the velocity distribution in the ranges in which the different overdensities were visible. In these plots, bright areas represent a high density of stars in velocity space. The v_θ range of MMH-1 and one of the v_θ ranges of the Helmi streams were not possible to distinguish in these plots and these ranges were therefore obtained from Mikkola et al. (2023). These ranges were not distinguishable because of how the data was cut and because high densities of stars in these ranges made it impossible to tell where the specific overdensity started and ended. This is the same reason why only a few of the velocity features presented by Mikkola et al. (2023) are investigated in this project. Moving groups like Thamnos and Sequioa (Koppelman et al. 2019) could not be told apart from other high density regions.

Velocity feature	v_r	v_θ	v_ϕ
GSE	$-380 < v_r < 380$	$-180 < v_\theta < 180$	$-70 < v_\phi < 130$
Helmi	$-180 < v_r < 180$	$210 < v_\theta < 330$ & $-300 < v_\theta < -180$	$-210 < v_\phi < -110$
ED-2	$-230 < v_r < -70$	$90 < v_\theta < 185$	$250 < v_\phi < 360$
MMH-1	$-300 < v_r < -150$ & $140 < v_r < 300$	$-350 < v_\theta < -250$	$-90 < v_\phi < -10$

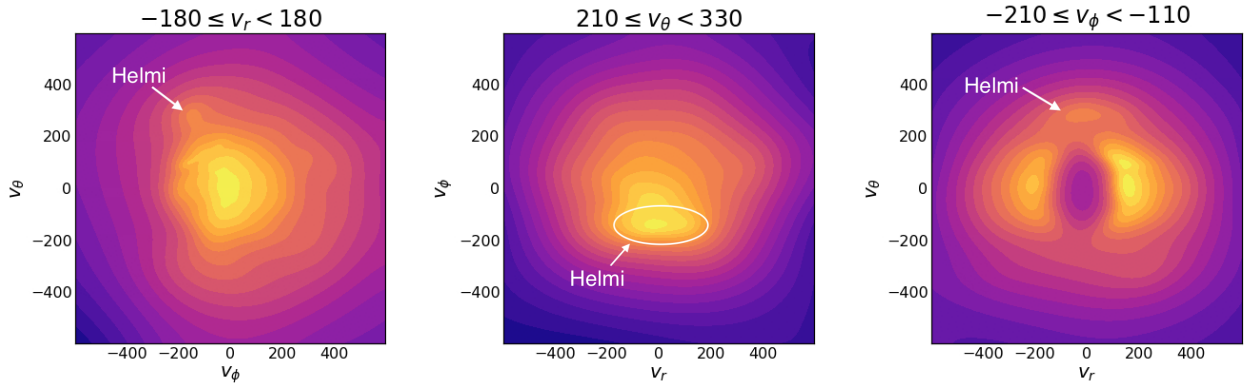
Table 3.1: Velocity ranges in v_r, v_θ and v_ϕ for the moving groups Gaia-Enceladus, Helmi streams, MMH-1 and ED-2.



((a)) The v_r range of GSE. ((b)) The v_θ range of GSE. ((c)) The v_ϕ range of GSE.

Figure 3.1: The velocity dimensions of Gaia-Enceladus.

The bright areas encircled in Fig. 3.1 show high densities of stars in velocity space that correspond to GSE. The debris of GSE make up a large fraction of the galactic halo stars and are distributed over a large range of velocities. Gaia-Sausage-Enceladus is very extended in the v_r range which makes the feature look like a sausage in velocity space, hence the name. This implies that there are GSE stars that travel with very high velocities radially in or out of the centre of the Galaxy.



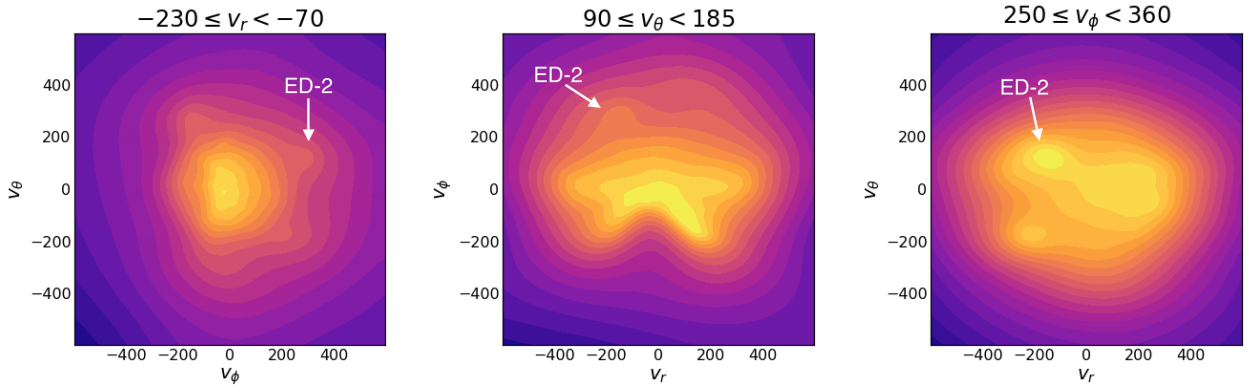
((a)) The v_r range of the Helmi streams.

((b)) One of the v_θ ranges of the Helmi streams.

((c)) The v_ϕ range of the Helmi streams.

Figure 3.2: The velocity dimensions of the Helmi streams. Only one of the v_θ ranges could be identified in these plots, the second v_θ range was taken from Mikkola et al. (2023).

The arrows in Fig. 3.2 point at high densities of stars in velocity space that correspond to the Helmi streams. Only the positive v_θ range could be distinguished in these plots. The Helmi streams are spread over relatively small ranges in v_θ and v_ϕ but is more extended in v_r. These stars move with velocities close to the velocity of the Sun, except in the v_θ-range where the Helmi stream stars have much higher velocities. The stars thus move with high velocities perpendicular to the galactic plane.



((a)) The v_r range of ED-2.

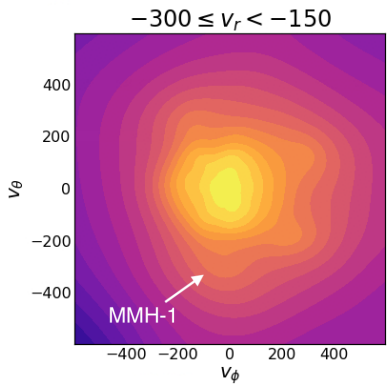
((b)) The v_θ range of ED-2.

((c)) The v_ϕ range of ED-2.

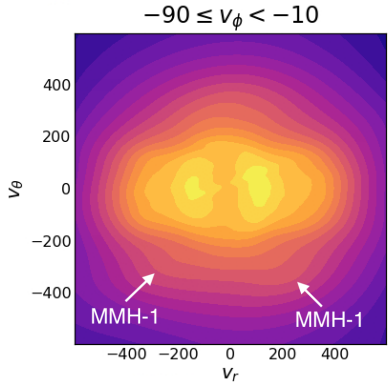
Figure 3.3: The velocity dimensions of ED-2.

The arrows in Fig. 3.3 point at high densities of stars in velocity space that correspond to ED-2. ED-2 is spread over a relatively small range of velocities in all three velocity directions but the feature has a relatively high density of stars. As Dodd et al. (2023) pointed out, we can see that ED-2 is very tightly clustered in velocity space.

The arrows in Fig. 3.4 point at high densities of stars in velocity space that correspond



((a)) One of the v_r ranges of MMH-1.



((b)) The v_ϕ range of MMH-1.

Figure 3.4: Two of the velocity dimensions of MMH-1. MMH-1 could not be found in the v_θ range.

to MMH-1. This velocity feature is also relatively small in velocity space but it is weaker than ED-2. MMH-1 could not be distinguished in the v_θ range and we therefore only see it in the (v_θ, v_ϕ)- and (v_θ, v_r) plane. In the v_r range there are two high density regions that correspond to MMH-1. These two v_r ranges were identified by Mikkola et al. (2023) separately, but because of their overlap in the full 3D distribution they are considered to be the same feature.

3.2 Candidates

With the velocity ranges of the overdensities given in Table 3.1 I can calculate the probability that a star belongs to a specific overdensity through Eq. 2.4. Through this we have created candidate lists[1], ordered from most to least probable member. These candidate lists contain source-id, apparent magnitude, right ascension, declination and galactic longitude and latitude of the stars. The five most probable members from these candidate lists, for each overdensity, can be found in Appendix B.

The number of stars with a specific probability to belong to an overdensity is presented in the histograms seen in Fig. 3.5. The stars are distributed in bins of width 0.025 over the probability range.

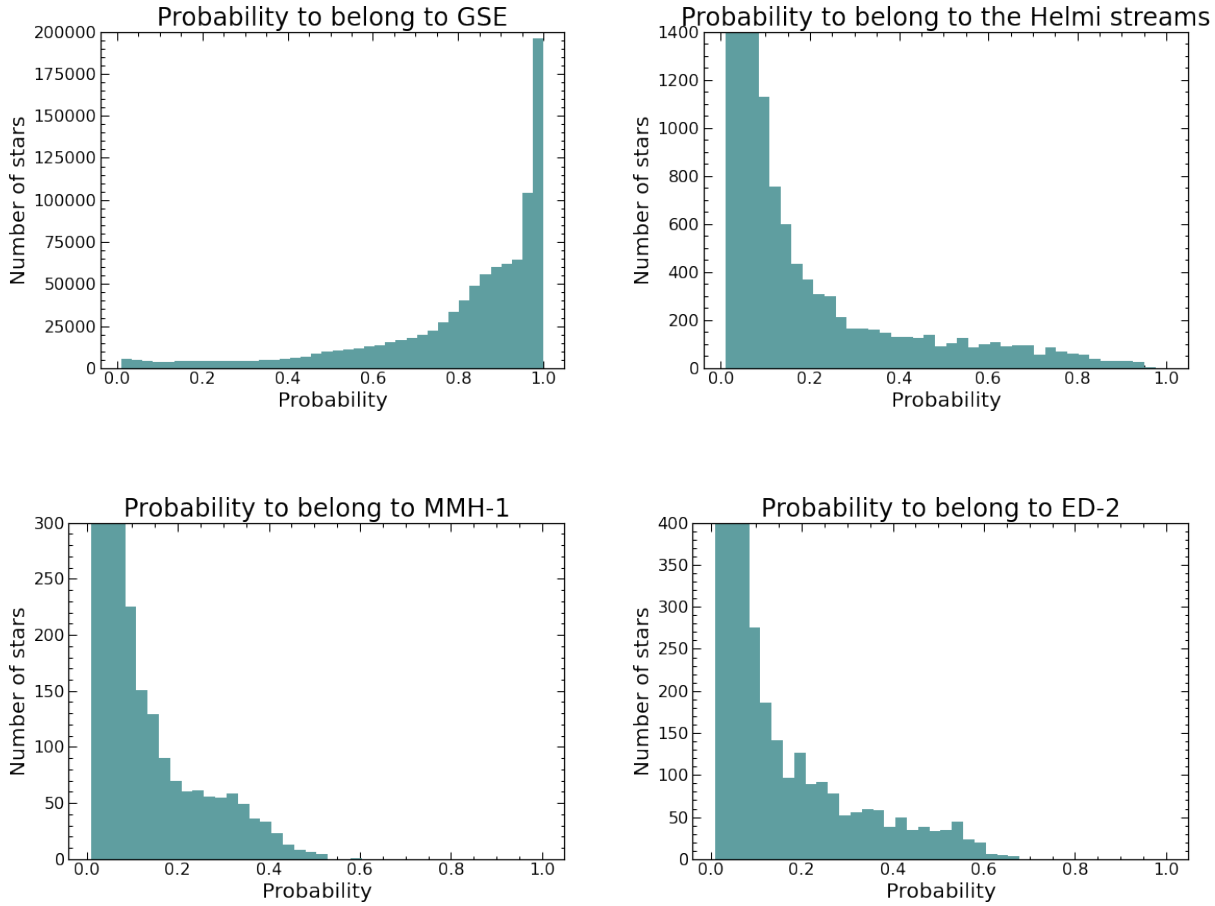


Figure 3.5: Histograms showing the number of stars with a certain probability to belong to the different overdensities.

[1]Available from: https://zenodo.org/record/7944688

In the histograms shown in Fig. 3.5 it can be seen that most stars have zero probability to belong to either MMH-1, ED-2 or the Helmi streams. In the histogram of GSE however, there is a significantly larger portion of stars that have almost 100% probability to belong to the feature than stars with close to zero probability. This is because GSE has a very large distribution of velocities and therefore a large fraction of the stars are likely to be found within the boundaries of GSE in velocity space. As I noted in the introduction, GSE is known to dominate the stellar halo near the solar neighbourhood. Since the other overdensities have much smaller velocity ranges there are fewer stars that are likely to belong to them.

The source-id and probability of the stars with the highest probability to belong to each velocity feature can be found under "All stars included" in Table 3.2. For the larger velocity features, GSE and the Helmi streams, there are stars with over 90% probability to belong to them. The most probable members of these overdensities have probabilities of 99.9% and 95.1% respectively. The probability of stars belonging to the smaller velocity features, ED-2 and MMH-1, also reach relatively high probabilities, around 60%. The most probable members of these smaller overdensities have probabilities of 67.2% and 59.1% respectively. Since these velocity features are relatively small in velocity space it is reasonable that the probability for stars to be within their ranges is lower than for the larger velocity features.

		Probability	
Velocity feature	Source-id	All stars included	Most probable stars excluded
GSE	6765638853548101760	0.999691	0.999693
Helmi	3455136709565458176	0.95101	0.950459
ED-2	582061801974487808	0.672031	0.669164
MMH-1	4333216011294413696	0.591686	0.588111

Table 3.2: Source-id and probability of the most probable members of the overdensities. The probabilities were first calculated with all stars in the data set included in the probability density calculations and then with the most probable stars of MMH-1 and ED-2 excluded from the probability density calculations.

3.2.1 Verification of assumptions

As discussed in section 2.3.1, the membership probability estimation is based on circular logic. The membership probabilities are calculated from a velocity distribution which itself is a probability distribution determined using all observations, including the ones for which we are trying to determine membership probabilities. Therefore, we now show that this does not affect our results.

I first calculated the membership probabilities with all stars in the data set included in the initial probability distribution calculations. I then removed the most probable members of MMH-1 and ED-2 from the probability distribution calculations and then put them back into the sample when calculating the membership probabilities. These two stars are thus not included in the velocity distribution but their probability to belong to a specific overdensity is still calculated. The probabilities of the most probable members after this change can be found under "Most probable stars excluded" in Table 3.2. These probabilities are very similar to the probabilities found under "All stars included" which means that excluding some of the most probable stars from the probability distribution did not affect the results.

In Table 3.3 the number of stars with a membership probability greater than 40% can be found for each overdensity. Again, these numbers were calculated both for the full data set included in the probability density and with the most probable members of MMH-1 and ED-2 excluded. It can be seen from Table 3.3 that the number of stars with a membership probability greater than 40% barely changed when excluding the most probable members. These two results reassure us that the membership probabilities do not depend too heavily on the velocity distribution of individual stars.

	Number of stars with probability greater than 40%	
Velocity feature	All stars included	Most probable stars excluded
GSE	877103	877007
Helmi	1794	1794
ED-2	304	301
MMH-1	64	62

Table 3.3: Number of stars with probability greater than 40% to belong to each overdensity. This was first calculated with all stars included in the probability density and then with the most probable members of MMH-1 and ED-2 excluded from the probability density.

From Table 3.3 it can also be seen that Gaia-Enceladus is a considerably more massive overdensity than the other ones. There are 877103 stars with probability greater than 40% to belong to the velocity range of GSE while there are only 64 stars with probability greater than 40% to belong to the velocity range of MMH-1. This is expected considering their size difference in velocity space.

3.2.2 Radial velocity stars

Stars that have a measured radial velocity have a precisely known position in velocity space. For these stars it is therefore possible to determine if they truly do lie within the

velocity range of the overdensities and not just the probability that this is true. This makes it possible for us to compare "true" members of the overdensity in velocity space to the probable members found in the membership probability calculations.

I selected stars with measured radial velocities from the sample and determined what stars were truly within the velocity ranges of the overdensities. In our data sample there are 70420 stars with radial velocities. From these calculations I found that there are 46490 stars in this sample that truly belong to the velocity range of GSE, 222 stars that truly belong to the velocity range of the Helmi streams, 39 stars that truly belong to the velocity range of ED-2 and 10 stars that truly belong to the velocity range of MMH-1.

With stars in this sample we compare the expected members (light blue) to the actual members (dark blue) in Fig. 3.6. The expected members are plotted as a cumulative sum of their membership probabilities from most- to least probable. The true members are paired to their corresponding membership probability and then plotted as a step-function that increase for each true member from high to low probabilities. In the ideal case, the true members would perfectly correspond to the expected members of our analysis.

From Fig. 3.6 it can be seen that the probable candidates of ED-2 and the Helmi streams correspond very well to the true members of these overdensities. For Gaia-Enceladus and MMH-1 there is a large overshoot in probable candidates versus true members. For GSE this is not surprising since the feature is so extended in velocity space which results in many stars having high membership probabilities. From what is seen in Fig 3.6 the true members of MMH-1 does not correspond well to the probable members. MMH-1 is a very small feature in velocity space and was only newly discovered as a rather weak overdensity in the velocity distribution of Mikkola et al. (2023). Since it only contains 10 radial velocity stars it is hard to do a fair comparison to the probable members.

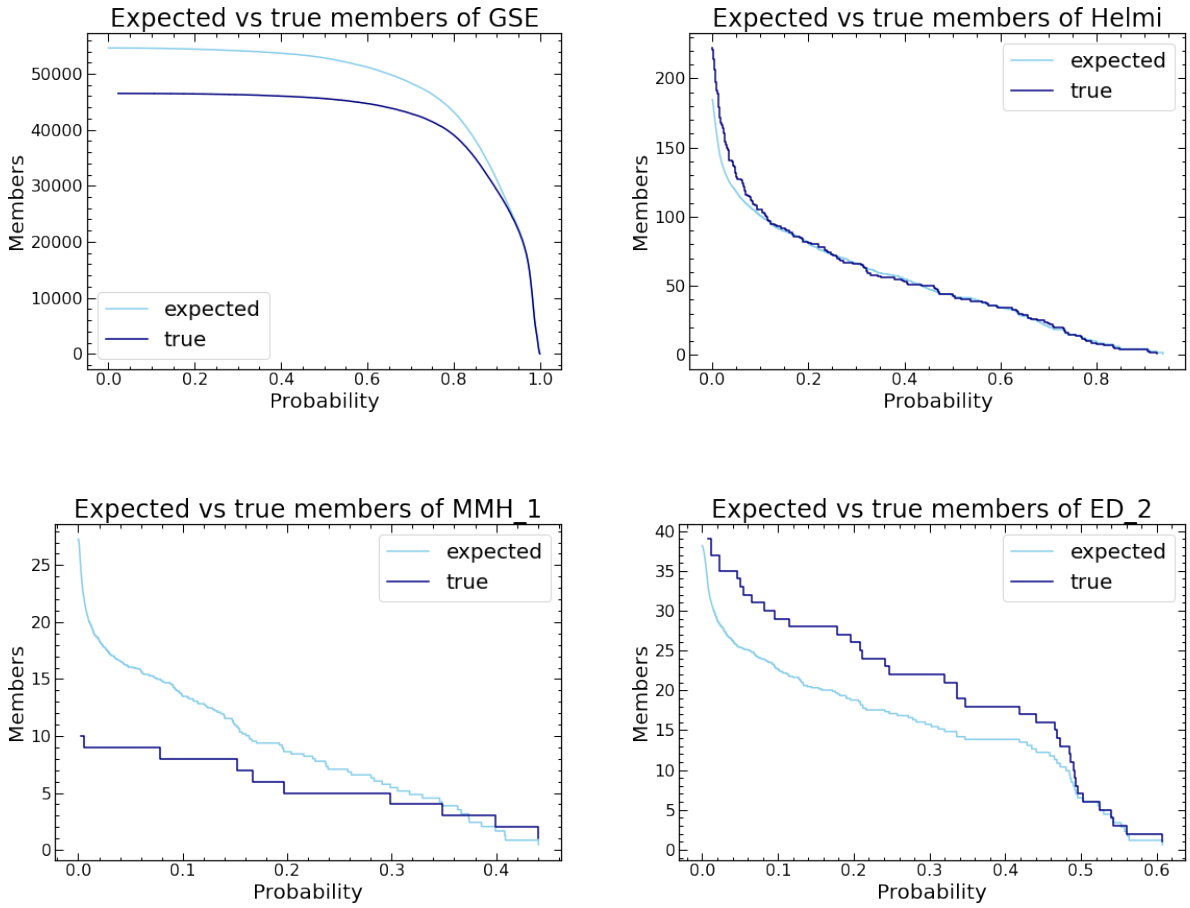


Figure 3.6: Comparison between probable members and true members of the overdensities. The light-blue curve is a cumulative sum of the membership probabilities of the radial velocity stars and the dark blue plot is a step function representing the stars that truly belong to the velocity range of the overdensity.

3.2.3 Metallicities

Around 60% of the stars in the data set (607030 stars) have measured metallicities from Andrae et al. (2023). I used these stars to calculate the metallicity distribution of the overdensities presented in this project. Previous research have found values for the mean metallicity of GSE, the Helmi streams and ED-2 which makes it possible for us to compare earlier results to the probable members found in this project. The metallicity distribution of the entire data set is shown in the histogram in Fig. 3.7 and the mean metallicity of this set is approximately -0.8. In Table 3.4 I show the mean metallicity of high probability candidates where the threshold for "high" probability is set at values between 40% to 90%.

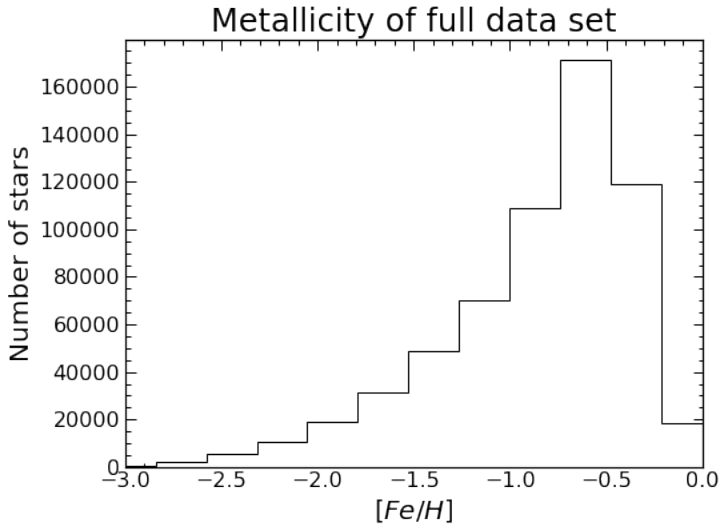


Figure 3.7: Metallicity distribution of all stars in the data set with measured metallicities from Andrae et al. (2023).

	Mean metallicity of stars with probability greater than...						
Velocity feature	40%	50%	60%	70%	80%	90%	Literature value
GSE	-0.854	-0.858	-0.864	-0.864	-0.877	-0.884	-1.15
Helmi	-1.065	-1.085	-1.088	-1.088	-1.064	-1.129	-1.5
ED-2	-1.366	-1.349	-1.803	-1.8034			-2.05
MMH-1	-1.189	-1.363					

Table 3.4: The mean metallicity of the overdensities for different sets of probable candidates compared to the mean metallicity from earlier research.

From Table 3.4 it can be seen that the most probable candidates of GSE have a mean metallicity of -0.88, which is very similar to the mean metallicity of the entire data set. This is however very different from the mean metallicity [Fe/H]$\sim$ -1.15 from Feuillet et al. (2021). The mean metallicity of the Helmi stream candidates with probability greater than 90% is $\sim$ -1.13, which is in the higher range of [Fe/H]$\sim$[-2.3,-1.0] shown by Koppelman et al. (2019). When only looking at the 10 most probable stars of the Helmi streams (probability greater than 93%) the mean metallicity is a little bit lower at a value of -1.2. ED-2 candidates with probability greater than 70% have a mean metallicity of -1.80, which is also in the higher range of what was shown by Dodd et al. (2023). The mean metallicity of MMH-1 candidates with probability greater than 50% is -1.36. Since this is a newly discovered velocity feature there are no metallicity values to compare to.

Metallicity distributions of probable members of the overdensities are shown in the histograms in Fig. 3.8.

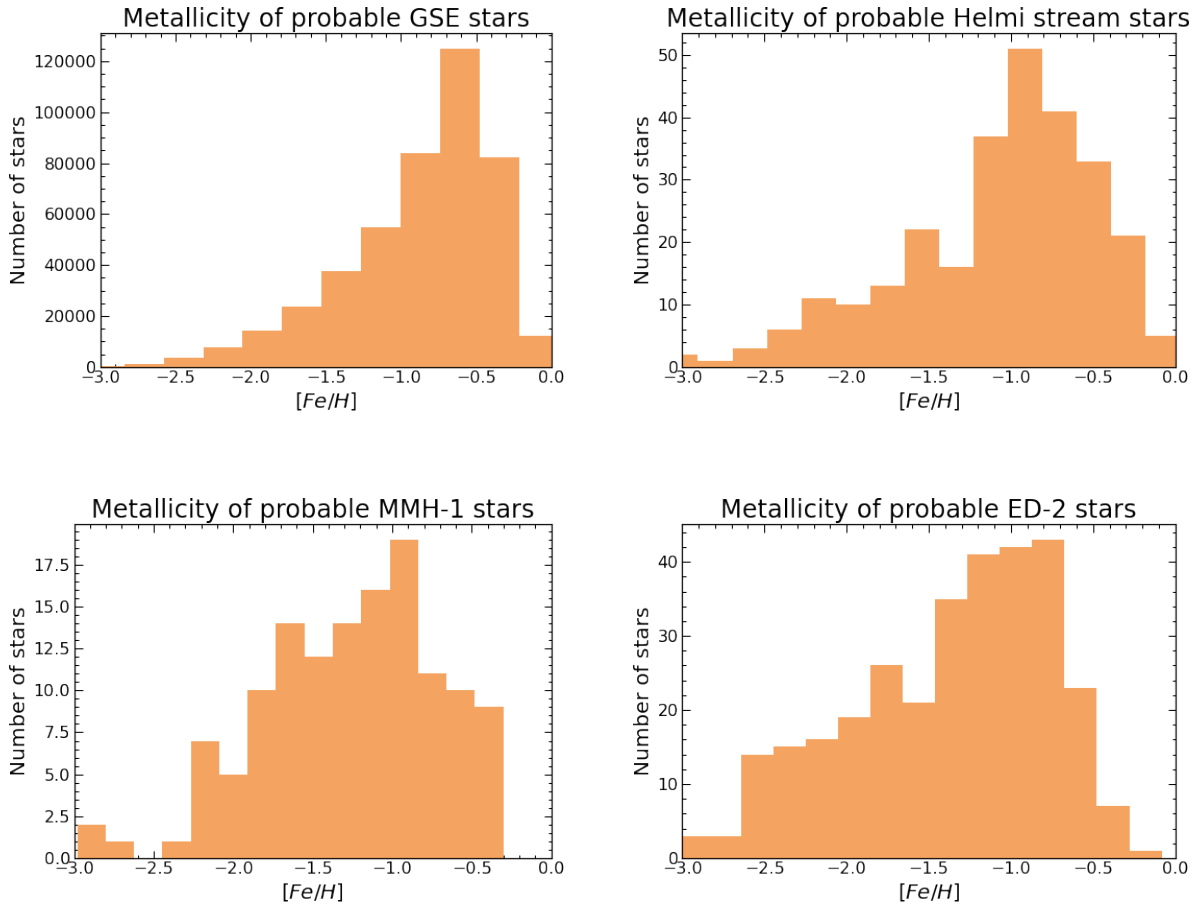


Figure 3.8: Histograms showing the metallicity distribution of probable candidates in the different overdensities. For GSE and the Helmi streams, stars with probability greater than 70% have been used. For ED-2 and MMH-1, stars with probability greater than 30% have been used. The stars are distributed in bins of size 0.2 over the metallicity range.

The GSE candidates have very similar metallicities to the metallicity distribution of the entire data set with a peak at [Fe/H]$\sim$ -0.6. The fact that the metallicity of the GSE candidates is very similar to the total data set is not surprising when considering the large number of stars that are likely to belong to this velocity feature. The metallicity distribution of the Helmi stream candidates have a peak at [Fe/H]$\sim$ -0.9. In this distribution

we can also see a much smaller but very clear peak around [Fe/H]$\sim$ -1.5 which is the peak metallicity shown by Koppelman et al. (2019). The metallicity distribution of MMH-1 candidates also peaks at [Fe/H]$\sim$ -0.9, but there are also two smaller but clear peaks around [Fe/H]$\sim$ -1.65 and [Fe/H]$\sim$ -2.2. The metallicity distribution of ED-2 has a peak at [Fe/H]$\sim$ -0.8 but there is also another clear peak at [Fe/H]$\sim$ -1.8 which is closer to the metallicty of the ED-2 members found by Dodd et al. (2023).

It looks like the general set of probable candidates have relatively high metallicities which depends on the high metallicity of our total data set. Since we only find probable members we can expect contamination from stars that are not really in the moving groups. We also need to consider errors in the metallicities found by Andrae et al. (2023) since they were obtained from low-resolution spectrophotometry. In the metallicity distribution of the Helmi streams and ED-2, we can however see the lower metallicity results from earlier research.

3.2.4 Opportunities for follow-up observations

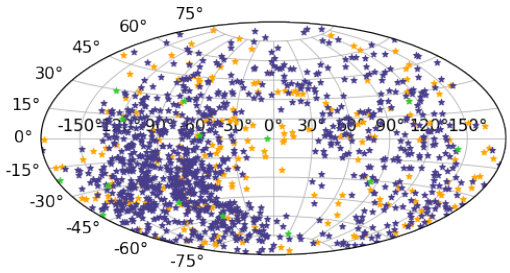
The idea behind the candidate lists is to make the selection of stars in future research, aimed at these moving groups, easier. These stars can for example be targeted by spectroscopic surveys such as WEAVE and 4MOST. Future investigations will also be performed by Gaia which is still gathering kinematic data of stars in the Galaxy.

WEAVE targets the northern hemisphere and can observe objects with declination $\gtrsim -30°$ (Jin et al. 2023). Stars in our data set that can be observed by WEAVE are found in the red area in Fig. 3.9 d). 4MOST targets stars in the southern hemisphere and can see objects with declination $\lesssim 20°$ (Chiappini et al. 2019). Stars that can be targeted by 4MOST can be found in the dark blue area in Fig. 3.9 d). There is an overlap in the stars that can be targeted by these surveys, so the stars that can be observed by both are shown as the purple area in Fig. 3.9 d).

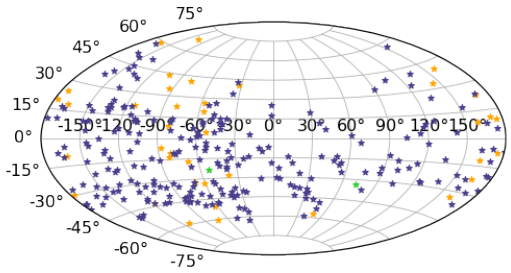
On board of Gaia there is a radial velocity spectrometer (RVS)(Gaia Collaboration et al. 2016). To measure the radial velocity of a star, the star must be bright enough. The magnitude in the wavelength range in which RVS works is denoted G_{RVS} and Gaia can detect stars with $G_{RVS} \lesssim 15$. In Fig. 3.9 a), b) and c), I have plotted the right ascension and declination of overdensity candidates. In these plots we see candidates that can be detected by Gaia's radial velocity spectrometer (dark blue), candidates that can not be detected by the spectrometer (green) and stars that already have measured radial velocities and are true members of the overdensities (yellow). GSE is not included in these plots because its large number of candidates completely covers the sky which makes it impossible to see all three sets of data.

From these plots we can see that these moving groups do not form clear overdensities in space but they are not completely scattered over the sky. We can also see that there are

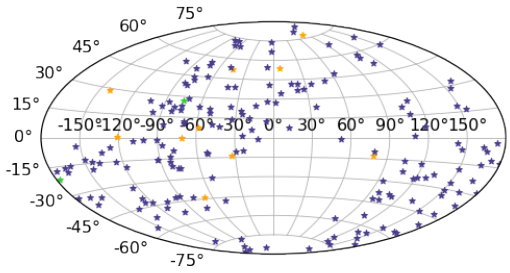
a lot of probable candidates for which it should be possible to measure radial velocities.



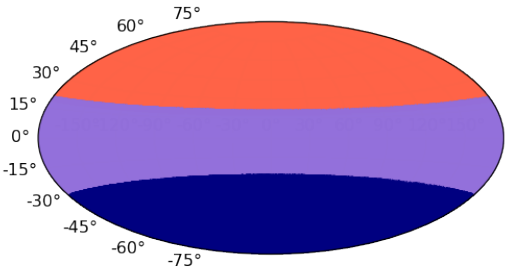
((a)) Position of Helmi stream stars on the sky.



((b)) Position of ED-2 stars on the sky.



((c)) Position of MMH-1 stars on the sky.



((d)) WEAVE and 4MOST

Figure 3.9: The three plots in a), b) and c) show the position of stars, connected to different overdensities, on the sky. Dark blue: stars from the candidate lists that can be detected by Gaia's RVS. Green: stars that can not be detected by Gaia's RVS. Yellow: stars that have measured radial velocities and are true members of the overdensity. The plot in d) shows what part of the sky can be observed by WEAVE and 4MOST. Red: stars that can be detected by WEAVE. Blue: stars that can be detected by 4MOST. Purple: stars that can be detected by both.

This distribution of candidates in space can also be visualised in terms of galactic longitude and latitude (l & b). In Fig. 3.10 I have plotted the galactic longitude and latitude of Helmi stream candidates. From this plot we can see that most of the Helmi stream candidates lie within the galactic plane and there are barely any stars with galactic latitudes around $b = 90°$ and $b = -90°$. This is because the Helmi stream stars have very similar velocities to the disk stars in the v_r and v_ϕ components, but their defining trait is that they have much higher v_θ velocities. When looking straight up above or below the galactic plane this v_θ component is very hard to measure since it corresponds to the line-of-sight velocity. Without this v_θ component these stars look just like disk stars at high/low latitudes and therefore these stars are not seen as likely members of the Helmi streams.

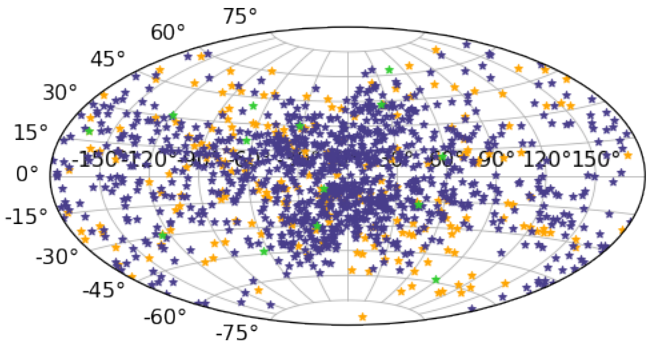


Figure 3.10: Galactic longitude and latitude (l & b) of Helmi stream candidates. These stars are colour coded in the same way as the plots in Fig. 3.9 a), b) and c).

Chapter 4

Discussion

In this project we implemented the penalized maximum-likelihood method used by Mikkola et al. (2023) to obtain a 3D velocity distribution of solar neighbourhood halo stars without using radial velocities. It was possible to identify overdensities in this velocity distribution that correspond to the moving groups GSE, ED-2, MMH-1 and the Helmi streams. These moving groups are believed to come from past merger events experienced by the Galaxy which makes them important targets within the field of galaxy formation. By determining the velocity ranges of these overdensities we could calculate the probability that the stars in our data set belong to the different groups. With this we have produced candidate lists with probable members of the different overdensities.

From the membership probability calculations we found that all of the overdensities have, at least a few, highly probable members with probabilities around 60% for MMH-1 to 99% for GSE. GSE have, by far, the largest amount of probable members and this is because of how extended it is in velocity space. The Helmi streams, which are the second largest feature in this project, also have a great number of probable members with the most probable member having 95% probability to belong to the streams. The two smallest features, ED-2 and MMH-1, have a significantly smaller number of probable members which is expected considering how small they are in velocity space.

Since the membership probability calculations are based on circular logic, we wanted to ensure that our candidate lists are trustworthy. We decided to leave out some of the most probable stars from the probability distribution to see how this would change the membership probabilities. From this we found that leaving out some of the most probable members from the velocity distribution barely changed the results. This reassures us that the method does not depend too heavily on the velocity of individual stars and that we can trust the membership probabilities.

Approximately 7% of the stars in our data set have measured radial velocities and from these stars we could find "true" members of the overdensities. When comparing these true members to the expected members, we found that the probable candidates of the Helmi

streams and ED-2 corresponded very well to the true members. For GSE and MMH-1 however, there was a large overshoot of expected members compared to true members. For GSE this is not surprising, since it is a very large feature there are a large number of stars with high membership probabilities. One possible explanation for the overshoot in expected members for MMH-1 is that the velocity ranges of the overdensity might not be correct. MMH-1 is a small and relatively weak feature in velocity space which makes it difficult to notice. It could not even be seen in the v_θ range of the slice plots used to set the ranges.

Since around 60% of the stars in our sample have measured metallicities from Andrae et al. (2023), we decided to compare the metallicity of our overdensity candidates to metallicities found by previous research about the moving groups. The general theme of these results was that the mean metallicity of the overdensities is higher than what has been shown from earlier research. One explanation for this is that the mean metallicity of our entire data set is relatively high. This might be because of possible disk stars within our sample despite our attempt to cut them out. It could also be because of uncertainties in the measured metallicities since they are obtained from low resolution spectra (Andrae et al. 2023).

The metallicity of our GSE candidates was much higher than what was expected from previous research but we could also see that the metallicity distribution of GSE was almost identical to the distribution of the whole sample. This is not a surprising result since most of the stars in our sample have a very high probability to belong to GSE. The metallicity distributions of the Helmi streams and ED-2 were also shifted toward higher metallicities than expected. But in both of these distributions we could also see very clear, but smaller, peaks at metallicity values found by earlier research. Since MMH-1 is a newly discovered feature we have no metallicitiy measurements to compare to.

Satellite galaxies that have merged with the Milky Way in the past are believed to have had more metal poor stars than our young Milky Way. Our results should not be thought to point toward higher metallicities than what has been previously shown. They rather show that both kinematic and chemical analysis is needed to identify members of past merger events.

There are some improvements that can be made in future work regarding the candidate lists. The measured velocities of the stars in our sample have uncertainties that could be taken into account in future studies when calculating the membership probabilities. The velocity ranges of the overdensities obtained in this project is also a source of uncertainty. It was sometimes very difficult to tell where a high density region started and ended in the slice plots used to obtain these ranges. A more accurate method of obtaining the velocity ranges of the moving groups would be an advantage in the search for probable members.

4.1 Conclusions

Our understanding of how galaxies form is still incomplete. It is however clear from the current knowledge of galaxy formation that accretion of smaller galaxies play a crucial role in this process. The Milky Way galaxy is full of debris from past merger events and we have in this project looked at overdensities in velocity space formed by this debris. To trace out the history of our Galaxy a better understanding of these mergers is necessary. Identifying stars that belong to these past mergers requires accurate kinematic and chemical data. There are a lot of stars in the Galaxy and prioritisations must be made when selecting samples of data. The candidate lists created in this project are therefore a useful tool to navigate what stars to target in future research about these moving groups.

To determine true members of overdensities in velocity space, radial velocity measurements are needed. Since only a small set of stars have measured radial velocities, we decided to find membership probabilities instead. As we saw in Fig. 3.9 there is a relatively large number of probable candidates that can be observed by Gaia's radial velocity spectrometer, but do not currently have measured radial velocities. If the radial velocities of these stars can be measured it will be possible to determine if they really do belong to the overdensities.

We hope that future research can perform spectroscopic follow-ups on the stars in these candidate-lists. By analyzing their chemical properties, we can determine if they are members of the past merger events and maybe gain knowledge about the merge. We recommend that the spectroscopic surveys such as 4MOST Consortium survey 1 (Helmi et al. 2019) and 2 (Christlieb et al. 2019) use our membership probabilities to target these specific moving groups.

Acknowledgements

I am truly thankful for the opportunity to undertake this project. I would like to express my gratitude to Paul McMillan, for being an amazing supervisor and for guiding me through the project. I also want to thank Danial Mikkola for writing a well structured code and for letting me use it.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

Bibliography

Andrae, R., Rix, H.-W., & Chandra, V. 2023, Robust Data-driven Metallicities for 120 Million Stars from Gaia XP Spectra

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167

Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611

Chiappini, C., Minchev, I., Starkenburg, E., et al. 2019, The Messenger, 175, 30

Christlieb, N., Battistini, C., Bonifacio, P., et al. 2019, The Messenger, 175, 26

de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, The Messenger, 175, 3

Dehnen, W. 1998, AJ, 115, 2384

Dodd, E., Callingham, T. M., Helmi, A., et al. 2023, A&A, 670, L2

Eggen, O. J., Lynden-Bell, D., & Sandage, A. R. 1962, ApJ, 136, 748

ESA. 2019, Radial Velocity with Gaia, `https://sci.esa.int/web/gaia/-/31367-radial-velocity`

ESA. 2023, Gaia Data Release 3, `https://www.cosmos.esa.int/web/gaia/dr3`

Feuillet, D. K., Sahlholdt, C. L., Feltzing, S., & Casagrande, L. 2021, Royal Astronomical Society, 508, 1489–1508

Freeman, K. & Bland-Hawthorn, J. 2002, Annual Review of Astronomy and Astrophysics, 40, 487

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Helmi, A. 2008, The Astronomy and Astrophysics Review, 15, 145

Helmi, A. 2020, Annual Reviews of Astronomy and Astrophysics, 58

Helmi, A., Babusiau, C., Koppelman, H., et al. 2018, Nature, 563

Helmi, A., Irwin, M., Deason, A., et al. 2019, The Messenger, 175, 23

Helmi, A. & White, S. D. M. 1999, MNRAS, 307, 495

Howes, L. M., Casey, A. R., Asplund, M., et al. 2015, Nature, 527, 484

Jin, S., Trager, S. C., Dalton, G. B., et al. 2023, MNRAS

Koppelman, H. H., Helmi, A., Massari, D., Roelenga, S., & Bastian, U. 2019, A&A, 625, A5

Mikkola, D., McMillan, P. J., & Hobbs, D. 2023, MNRAS, 519, 1989

Quinn, P. J., Hernquist, L., & Fullagar, D. P. 1993, ApJ, 403, 74

Roederer, I. U., Sneden, C., Thompson, I. B., Preston, G. W., & Shectman, S. A. 2010, ApJ, 711, 573

Ryden, B. & Peterson, B. M. 2021, Foundations of Astrophysics (Cambridge University Press)

White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341

Appendix A

Gaia Archive Query

```
select source id, bp rp, phot g mean mag, phot bp rp excess factor, ruwe, ra, dec,
parallax, pmra, pmdec, parallax error, pmra error, pmdec error, parallax pmra corr,
parallax pmdec corr, pmra pmdec corr, visibility periods used, astrometric chi2 al,
astrometric n good obs al, radial velocity,
radial velocity error,
if then else(
 bp rp > -20,
 to real(case condition(
  phot bp rp excess factor - (1.162004 + 0.011464 bp rp + 0.049255power(bp rp,2)
                    - 0.005879power(bp rp,3)),
  bp rp < 0.5,
  phot bp rp excess factor - (1.154360 + 0.033772 bp rp + 0.032277power(bp rp,2)),
  bp rp > = 4.0,
  phot bp rp excess factor - (1.057572 + 0.140537bp rp)
)),
phot bp rp excess factor
) as excess flux
from gaiadr3.gaia source
where parallax over error > 5
and parallax > power(3, -1)
and ruwe < 1.15
and phot g mean flux over error > 50
and phot rp mean flux over error > 20
and phot bp mean flux over error > 20
and visibility periods used > 8
and astrometric chi2 al/(astrometric n good obs al-5)
< 1.44greatest(1,exp(-0.4(phot g mean mag-19.5)))
and 4.74sqrt(power(pmra, 2) + power(pmdec, 2))/parallax > 200
```

Appendix B

Most probable stars

The tables below show the first five rows of the candidate lists created in this project for each moving group. The stars are ordered from most to least probable and for each star the source-id, apparent magnitude, right ascension, declination and galactic longitude and latitude is included. The full candidate lists can be found at, https://zenodo.org/record/7944688.

Source-id	m_G	ra	dec	l	b
6765638853548101760	17.053	291.146	-27.866	10.682	-19.058
6758950627399445632	16.256	288.880	-30.124	7.702	-18.034
6757981270453488640	17.345	289.453	-31.429	6.605	-18.964
6758066654411293440	15.538	288.954	-30.915	6.951	-18.381
6757910038420496512	17.289	290.611	-31.018	7.383	-19.746

Table B.1: The five most probable members of GSE.

Source-id	m_G	ra	dec	l	b
3455136709565458176	18.376	87.857	35.478	175.0168	4.418
3426431591999115008	17.619	91.008	24.385	186.005	1.249
189031406301782272	18.313	78.606	40.605	166.746	1.101
3350508905351012736	16.658	89.300	18.391	190.424	-3.103
3444244161402830208	16.568	85.920	29.614	179.200	-0.005

Table B.2: The five most probable members of the Helmi streams.

Source-id	m_G	ra	dec	l	b
582061801974487808	17.330	133.611	5.552	222.615	29.805
6907543920169690880	17.217	308.717	-7.572	38.047	-26.415
6879590726822323200	16.570	303.143	-13.177	29.862	-23.850
6895076965842891648	17.160	318.251	-11.196	39.087	-36.454
6877427059736525952	15.080	300.883	-15.453	26.634	-22.763

Table B.3: The five most probable members of ED-2.

Source-id	m_G	ra	dec	l	b
4333216011294413696	17.127	255.623	-12.049	8.781	17.646
2982220059626504064	16.723	80.796	-16.556	218.680	-26.767
102518056152746112	15.182	35.354	24.074	147.973	-34.384
18721418846252288	16.522	41.545	6.712	166.492	-46.228
19566432187174272	16.664	38.048	7.352	161.832	-47.764

Table B.4: The five most probable members of MMH-1.