

# The Search for Galactic Spaghetti

Quantifying the amount of substructure in the  
Galactic halo

Figure by P. Harding

Else Starkenburg

Supervisor: dr. A. Helmi

Co-supervisors: prof. H. Morrison, dr. P. Harding

Second and third reader: prof. H. van Woerden, prof. S. Zaroubi

Afstudeerscriptie Natuur- en Sterrenkunde

Rijks*universiteit* Groningen

July 10, 2007



## Abstract

We present the analysis and interpretation of the Spaghetti Survey data, a pencil-beam high latitude survey designed to test the amount of kinematic substructure in the Galactic halo. Photometrically, red giant candidates are selected which are additionally followed-up spectroscopically. The advantage of red giants is that they can be traced out to large distances up to  $\sim 100$  kpc. The final giant sample contains 102 giants, for which distances, radial velocities and metallicities have been determined.

Using a distance measure combining spatial and velocity information we find 7 significant groups within this dataset. One group of 7 members can confidently be matched to tidal debris of the Sagittarius dwarf galaxy. Three other groups can also be associated to Sagittarius debris, provided they were stripped off relatively early. However, two of these groups as well as an additional third group also match to known Virgo structures. We discuss the limitations of linking these found groups to larger substructures.

Two stars in the dataset are identified as candidate members of the Orphan Stream. One of these candidates confirms earlier radial velocity measurements from Belokurov et al. (2007).

In total, we have measured 22.5% of the stars in the Spaghetti dataset to be in substructures. From comparison with substructure in smooth random sets we have derived a very conservative lower limit for the amount of stars in the halo to be in substructures of 6%. Our results are consistent with a halo entirely built up from disrupted satellites, provided the dominating features are relatively broad due to early merging processes or relatively heavy progenitor satellites.



# Contents

<b>1</b>	<b>Introduction and Outline</b>	<b>3</b>
<b>2</b>	<b>Our Galaxy</b>	<b>5</b>
2.1	The structure of our galaxy . . . . .	5
2.1.1	The thin disk . . . . .	5
2.1.2	The thick disk . . . . .	6
2.1.3	The bulge . . . . .	6
2.1.4	The halo . . . . .	7
2.1.5	Globular Clusters . . . . .	8
2.2	The Formation . . . . .	9
2.2.1	Galaxy Formation models . . . . .	9
2.2.2	The formation history of the Milky Way . . . . .	10
2.3	Streams and Clumps . . . . .	12
<b>3</b>	<b>The Spaghetti Survey</b>	<b>15</b>
3.1	My part of the project . . . . .	16
3.2	Photometry . . . . .	16
3.3	Spectroscopy . . . . .	17
3.3.1	Spectroscopic Reduction . . . . .	18
3.3.2	Radial Velocity . . . . .	19
3.3.3	Indices . . . . .	21
3.3.4	Distances . . . . .	29
3.4	The final data set . . . . .	30
<b>4</b>	<b>Analysis of the data</b>	<b>31</b>
4.1	The 4distance . . . . .	32
4.1.1	Making a random set . . . . .	33
4.1.2	Choosing a relevant binsize . . . . .	34
4.1.3	Pairs and groups . . . . .	36
4.1.4	The effect of errors and our choice of 4distance . . . . .	40
4.1.5	Significance of the groups . . . . .	41
4.2	The Great Circle method . . . . .	47
4.2.1	Great Circle Family of orbital Poles . . . . .	48

---

4.2.2	Energy and angular momentum check . . . . .	50
4.2.3	Combination of both requirements: The full great circle method . . . . .	53
<b>5</b>	<b>Interpretation of the Data</b>	<b>59</b>
5.1	Are these substructures new? . . . . .	59
5.1.1	The Sagittarius Dwarf Galaxy . . . . .	60
5.1.2	The Virgo Substructures . . . . .	64
5.1.3	The Orphan Stream . . . . .	66
5.2	Constructing simulated datasets . . . . .	72
5.2.1	Substructure in the simulated datasets . . . . .	73
<b>6</b>	<b>Discussion and Conclusions</b>	<b>85</b>
6.1	The 4distance . . . . .	85
6.2	The great circle method . . . . .	87
6.3	Successes and limitations of the Spaghetti project . . . . .	88
6.4	The properties of the Galactic stellar halo . . . . .	90
6.5	Conclusions . . . . .	91
6.6	Future Work . . . . .	92
<b>A</b>	<b>Tables</b>	<b>95</b>

# Chapter 1

## Introduction and Outline

### The puzzle of the formation of the Milky Way

For my Groot Onderzoek, which is the finishing project of my master's degree in astronomy, I was looking for a project which dealt with a fascinating 'big issue' about the Universe and its history, but was still bounded enough that it could be done in 12 months of research. Also, the project I had in mind would combine observations with a more theoretical approach and, last but not least, give me the opportunity to visit an astronomy department outside the Netherlands. Although these requirements may sound very insistent, they were all fulfilled by the project described in this master thesis.

The 'big issue' addressed here is the process of Galaxy formation in general and the formation of 'our own' Galaxy in particular. The project enabled me to work with the dataset of the international collaboration of the Spaghetti Survey. The first part of the project, carried out in Cleveland, USA under supervision of Prof. Heather Morrison and Dr. Paul Harding consisted mainly of data reduction of 3 of the 13 runs in the total project. In the second part of the project, in Groningen under supervision of Dr. Amina Helmi, I analysed and interpreted this data. The direct aim of the Spaghetti Survey can roughly be described as to quantify the amount of 'Spaghetti' (streams of substructure) in the Milky Way halo.

Over the last decades, enormous progress has been made in our understanding of the way the Universe was formed. With improving resolution in numerical simulations, advanced knowledge of physical processes and very detailed observations of structures in our Galaxy and beyond, the field of galaxy evolution is booming. While more and more pieces of the big puzzle seem to be collected, the issue of how to put them all together still remains. Everywhere we look in the Universe we find galaxies in various shapes and sizes. But how did these highly complicated systems evolve?

Galaxy formation is thought to be a hierarchical process which means that smaller structures form first. These smaller building blocks will then

merge together to form a larger structure. In this theory every large galaxy we see today, such as the Milky Way galaxy we live in, is assembled out of smaller galaxies that merged due to their mutual gravitational attraction. Once a relatively large structure is formed, it will exert a larger gravitational attraction and can capture a lot more smaller systems. For the Milky Way the merging model finds support in observations. An example of a minor merger, a capture of a smaller (dwarf) galaxy by a big one like our own galaxy, can be observed today in the accretion of the Sagittarius dwarf galaxy by the Milky Way.

While we thus know that at least part of our Galaxy is build up out of smaller subsystems, we do not know whether accretion is the dominant or a minor factor in halo build-up and what the of properties the in-falling satellites were. The Spaghetti Survey tries to quantify the contribution of later merged systems to the outer parts of our Galaxy, the halo, by selecting halo stars randomly and searching for substructure within this dataset. Because merged systems remain detectable as distinct substructures in the outer parts of the Galaxy for many gigayears, the fraction of the total stars in substructures is a measure of the merging history of the Galaxy. Although the number of halo stars in the Spaghetti dataset is negligible small compared to the Galactic halo, close investigation of substructure within this dataset will give a first order answer to the major question for our survey: how much of the Galaxy's halo was built up by the accretion of small satellites?

### **Outline of this report**

In the second chapter, presented after this introduction, an overview is given of our Galaxy and its main components. Subsequently, the formation history scheme of the Milky Way is discussed in more detail including the current observational evidence for substructure in the Galactic halo. Chapter Three discusses the data selection and reduction. This also includes the determination of distance, radial velocity, metallicity and luminosity classification for every program star. The analysis of the final assembled Spaghetti dataset is presented in Chapter Four. Two different substructure finding methods are presented here and used to find substructures in the dataset. A further interpretation of the results is given in Chapter Five in which also our findings are compared to results obtained with several simulated datasets. Finally, the overall results are discussed and our conclusions and a brief outlook are presented.

## Chapter 2

# Our Galaxy

### 2.1 The structure of our galaxy

Our galaxy is a large spiral galaxy. It consists of a few distinct components being a disk (thick and thin), a bulge in its centre and a, more or less, spherical halo with both a stellar and a dark matter component. The different components of our Milky Way, which are shown schematically in Figure 2.1, will be discussed one by one in the following sections.

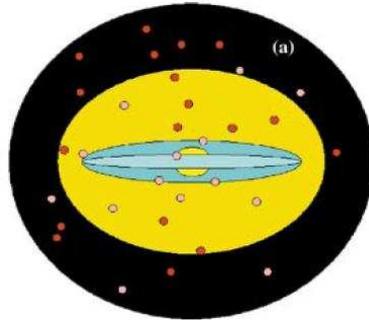


Figure 2.1: Sketch of the Milky Way showing the stellar disk (*light blue*), thick disk (*dark blue*), stellar bulge (*yellow*), stellar halo (*mustard yellow*), dark halo (*black*) and globular cluster system (*filled circles*). Figure from Freeman and Bland-Hawthorn, 2002.

#### 2.1.1 The thin disk

Looking at our Galaxy, the most obvious disk structure is the part we call the thin disk. Its surface brightness is exponential, both in radial direction and in height. It has a vertical scale height of 300 pc and extends until  $3.5 \pm 0.5$  kpc in the plane (Binney and Tremaine, 1994). Because its thought to

be the end product of quiescent dissipational collapse of baryons, the thin disk contains almost all of the baryonic angular momentum.

From radioactive dating, white dwarf cooling and isochrone estimates, the age of the oldest stars in the thin disk is estimated to be 10-12 Gyr (Freeman and Bland-Hawthorn, 2002). There is however also a large population of intermediate aged and young stars present. Thin disk stars are fairly metal-rich. Edvardsson et al.(1993) derived from their sample of 189 nearby stars a metallicity for the thin disk stars of  $\geq -0.2$  dex.

The rotation speed of the disk is hard to constrain from observations, especially because there is a degeneracy between the circular-speed and the distance to the centre of the Galaxy, but generally it is thought to be 220 km/s at the Solar radius, about 8.5 kpc from the centre (Binney and Merrifield, 1998).

### 2.1.2 The thick disk

The thick disk is a much more extended structure than the thin disk lying underneath, its scale height is about three times larger ( $\sim 1$  kpc). But its surface brightness is only about 10% of the surface brightness of the thin disk. Another remarkable difference is that the stellar population seems to be much older, ( $\sim 12$  Gyr) and significantly more metal poor ( $[\text{Fe}/\text{H}] \sim -0.6$  (Norris, 1999)). There is some evidence for a metal-weak thick disk (material having disk-like kinematics and  $[\text{Fe}/\text{H}] < -1$ ) (Norris, 1999). Such a metal-weak component is estimated to have a much larger scale-length of  $\sim 4.5$  kpc (Chiba and Beers, 2000).

The current belief is that an old thin disk, which was formed during the early stages of the formation of our galaxy, was heated in some way to form the thick disk and that a new thin disk was established after that (see also section 2.2.2 for a more detailed description of the Milky Way formation process). Possible heating mechanisms accounting for the stirring of the (then thin) disk would be for instance accretion events or minor mergers. If this picture proves to be right, the thick disk as we see it now would provide an important piece of information for the Milky way formation process, since it really is a snap-frozen view of the conditions of the disk shortly after its formation (Freeman and Bland-Hawthorn, 2002).

### 2.1.3 The bulge

There are many possible processes (disk instabilities, satellite accretion and major mergers) that could trigger the formation of a bar or bulge in the centre of a galaxy. Although most of the more luminous disk galaxies have bulges, many of the fainter disk galaxies have not. Bulge formation is thus not essential in the formation process of disk galaxies (Freeman and Bland-Hawthorn, 2002). Our Galaxy does possess a bulge, although it is not clear

what processes took place to form it. Observational evidence about the properties of this bulge in the central region of our own Milky Way is hard to obtain because of the obscuring dust. Historically, the only information about this mysterious region was obtained via Baade's window, a small area in the sky which happens to be almost free of obscuring dust clouds. When looking through the dust in infrared wavelength with for instance the COBE satellite a very clear bulge is seen (see Figure 2.2) that is probably triaxial (Weiland, 1994). The bulge has a scale-height of about 0.4 pc and a velocity dispersion of about 100 km/s in radial direction. It is thought to contain about one third of the disk mass and, inferred from its observed non-axisymmetric structure, it might also contain a bar (Binney et al., 1997).

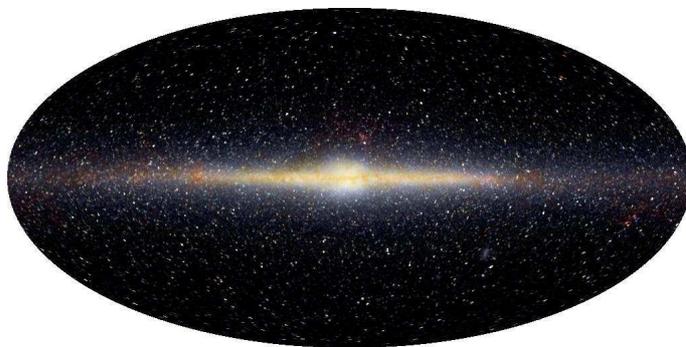


Figure 2.2: A picture of the Milky Way in infrared by COBE (credit: E.L. Wright, The COBE Project, NASA).

Some RR Lyrae stars first identified by Baade are present in the bulge and represent a metal-poor population. These RR Lyraes suggest that at least some fraction of the bulge is old. However, the bulge has quite a range of chemical abundances. McWilliam and Rich (1994) measured abundances for red giants in the bulge and found that, although there is a wide spread (-1.25 to +0.5 dex), the mean iron abundance of  $[\text{Fe}/\text{H}] \approx -0.25$  is closer to intermediate aged populations in the metal rich disk. The relative abundance of the  $\alpha$ -elements however, is much higher than observed in the disk and resembles more that of the metal-poor stars in the halo (McWilliam and Rich, 1994). While there is still a lot of information lacking, current best constraints on both age and formation of the bulge suggest that the bulge is as old as the globular clusters and formed relatively quickly in less than 1 Gyr (Rich, 2001).

#### 2.1.4 The halo

As stated before, the halo consists of two components: a stellar and a dark matter component. It was discovered by Zwicky in 1933 that just the gravi-

tational force of the luminous matter was not enough to account for the high velocity dispersions measured in the Coma Cluster and therefore a large part of the matter should be dark (Zwicky, 1933). In 1959, Kahn and Woltjer showed that also the system of the Local Group can only be dynamically stable if it contains an appreciable amount of intergalactic matter. This intergalactic dark matter is now thought to reside in the dark halos of the galaxies. For the Milky Way the dark halo is thought to extend until at least  $\sim 200$  kpc and to contain  $\geq 10^{12} M_{\odot}$  (Zaritsky, 1999) dark matter mass.

The stellar halo, on the other hand, is in luminosity probably the most insignificant part of our galaxy: its star density is 850 times less than the disk at the solar radius in the plane (Morrison, 1993), accounting in total for about 1% of the stellar mass. Although there are some claims that the outer halo should possess a net retrograde movement (Majewski, 1992), other datasets seem to deny any systematic rotation (Chiba and Beers, 2000). The halo is nearly spherical in its outer parts and can be modelled by means of a power-law profile with a slope of about -3.5 (Chiba and Beers, 2000; Morrison, 1993).

The main importance of the outer halo is its long dynamical timescale. At a radius of 100 kpc the dynamical timescale is already of the order of several Gyr (Freeman and Bland-Hawthorn, 2002) so mixing times are very long. This means that incoming satellites or satellite debris keep their original orbits for much longer than in the crowded disk. Numerical simulations indeed show that mergers would leave observable fossil structure in the stellar halo (Helmi and White, 1999; Harding et al., 2001). In velocity space, these structures can be recognised even when their spatial structure is no longer apparent. This theoretical picture is confirmed by the detection of debris in our Milky Way, of which the most obvious example is that of the Sagittarius dwarf galaxy (Ibata et al., 1994). The disruption of the Sagittarius dwarf is relatively recent however. Helmi et al. (1999) show evidence that the process of in-fall from various satellites and debris has, over the lifetime of the Milky Way, provided quite a large fraction of the total mass the Milky Way's stellar halo.

### 2.1.5 Globular Clusters

The globular cluster system of the Milky Way can be divided into at least two distinct subsystems. This distinction can be made based on their spatial distribution (clusters are associated to the halo or the disk), but also on the basis of metallicity arguments. The clusters with  $[Fe/H] \geq -0.8$  are the metal-rich subsystem which can be associated with the disk globular clusters, while the halo globular clusters, have  $[Fe/H] < -0.8$  (Binney and Merrifield, 1998).

Besides a distinction in metallicity and spatial distribution, these two subsystems also show differences in kinematics. The disk globular clusters

have a rotational velocity of  $193 \pm 29$  km/s and a relatively low velocity dispersion of  $59 \pm 14$  km/s. Within error-bars, these values are identical to the kinematics derived for thick disk stars. The scale height of the disk globular clusters is found to be between 800 and 1500 pc and is also consistent with scale height determinations of the thick disk. There is some evidence for metallicity gradients in the disk globular clusters, both with distance from the Galactic plane and distance from the Galactic centre (Armandroff, 1989).

The halo globular clusters, on the other hand, have a spherical distribution about the galactic centre with a small rotational velocity of  $50 \pm 23$  km/s and a large velocity dispersion of 114 km/s. As for the disk globular clusters, a metallicity gradient with radius can be found until  $R \approx R_{\odot}$  where it becomes either too shallow to measure, or it disappears (Zinn, 1985).

Besides the distinction in two components, there are claims that there exist at least 3 globular clusters which can be dynamically associated with the bulge, making up a third distinct subsystem (Minniti, 1996).

With its oldest members being about 12 Gyr old (Salaris and Weiss, 1997), the globular cluster system is among the oldest components of our galaxy. Globally the system seems to have an age spread correlating with metallicity, but this is not a linear relation. The metal-poor population seems to be almost coeval, while more spread in ages is found in the metal-rich members of the system (Salaris and Weiss, 1997).

## 2.2 The Formation

### 2.2.1 Galaxy Formation models

One of the first and certainly one of the most influential models of Milky Way formation was presented by Eggen, Lynden-Bell and Sandage (ELS) in 1962. Using two different data sets, they found that stars with a higher ultraviolet excess ( $\delta(U - B)$ ) moved in more eccentric orbits. Linking this ultraviolet excess with metallicity, they deduced that the metal-poor stars possessed different orbits than the metal-rich ones. A collapsing galaxy would give an explanation for the observed phenomenon. This collapse would have to be very rapid, in just a few  $10^8$  years the gas would relax to circular orbits in equilibrium and thus form a disk. The older metal-poor stars, which were formed before the collapse would still be expected on more eccentric, slower orbits, while the younger metal-rich stars formed during and after the collapse would be in more circular and faster orbits because of the disk formation process. While their galaxy formation model would be the dominant model for a few decades, it was shown later that the apparent correlation between the orbital eccentricity of halo stars with metallicity (or ultraviolet excess), was basically a result of their proper-motion selection bias (Chiba and Beers, 2000) and that thus their conclusions drawn from

these results were incorrect.

A different scheme was proposed by Searle and Zinn (SZ) in 1978. Based on observations of globular clusters in the outer part of the Galaxy, and in particular on the absence of a radial abundance gradient for those systems, they argued that a simple, monolithic collapse model would be very unlikely. Instead they proposed a model in which the process of halo formation is dominated by the merging of distinct subsystems.

More recent observations point out that the best model might be a combination of the two models described above (Bullock and Johnston, 2005). While the central body of the Galaxy might have formed in a rapid collapse or merger, the outer halo was subsequently build up by later merging. The models preferred nowadays to show this process are *hierarchical*. They basically assume that the Universe is dominated by massive, weakly interacting particles and that the primeval density fluctuations were nearly scale-invariant (Peebles, 1982). A result of these assumptions is that small structures will collapse first and then grow together to form larger structures. Observational evidence for a hierarchical formation of our galaxy and late merging was found in the discovery of structures like the Magellanic Stream (Mathewson et al., 1974) and the Sagittarius dwarf galaxy (Ibata et al., 1994) that is being tidally stripped by our Milky Way.

### 2.2.2 The formation history of the Milky Way

In the theories on Milky Way formation a lot of processes are still poorly understood. There are however some scenarios that seem to match the current observational evidence reasonably well. I will discuss here a formation scenario as sketched by Freeman and Hawthorn (2002) complemented with the formation scenario described by Amina Helmi in her thesis (2000).

The inner dark halo of our galaxy, consisting of Cold Dark Matter, was the first to assemble. The early stages (before  $z \sim 2$ ) of baryonic galaxy evolution were probably dominated by violent gas dynamics and accretion events. In this time, which was also called the Golden Age because of the peak in both star formation and accretion disk activity, the stellar bulge and the massive black hole in the centre of the Galaxy assembled. The evolution of these two components are believed to be closely linked together, as is expected also from the observed relation between black hole mass and stellar bulge dispersion in other galaxies.

The first globular clusters and also the first halo stars might have formed in this period. These stars would probably have been very metal poor, with  $[\text{Fe}/\text{H}] \approx -5$  until  $-2.5$ . Due to very rapid metal enrichment in the core of the galaxy, a very strong metal gradient was established quickly. The high speed at which the enrichment process took place helps explaining the observed properties of our bulge today, which is both old and moderately metal-rich.

While some of the halo stars and globular clusters might have formed in the Golden Age as described above, the current theories state that many of the halo stars and globular clusters are debris from early satellite galaxies. These satellites experienced independent evolution and can thus have very different chemical histories. In a relatively early merger – between  $z = 5$  and  $z = 3$  – with a spiral galaxy like our own, both galaxies would already have a large gas content and even a small population of stars. This kind of merger could account for the more metal-rich and slightly younger part of the globular cluster population in our galaxy, while the globular clusters associated more with the stellar halo could well have come with smaller galaxies that merged with our galaxy at different times. The observed dynamics of the globular cluster population seems to support this formation scheme; while the young systems are rapidly rotating due to the total angular momentum of the merged galaxies; the globular clusters associated with the stellar halo show a quite large velocity dispersion because they all reflect the orbits of their parent galaxies (Unavane et al., 1996). Comparison of halo stars or globular clusters with present day populations in satellite galaxies will not be valid though. Because of the strong effects of tidal forces upon these systems after merging, star formation is shut off almost immediately after the merging event, while in present day satellite galaxies populations might look quite different due to later star formation (Venn et al., 2004; Bullock and Johnston, 2004).

During the latter stages of the Golden Age, the baryons started to settle to a disk. If indeed our Galaxy contains not only a bulge but also a bar, this would be a dynamic object formed from the disk. In this case we expect old disk stars in the inner regions also and not just stars that were formed in the very early violent relaxation time when the bulge was formed. A possible formation scenario for the two components of the disk, thick and thin, would be one where in the early stages of disk formation an object with a mass of about 20% of the (proto)Milky Way fell in, heated up the (then thin) disk into the thick disk as we know it now. The thin disk, as we know it now, formed later, partly from gas that had already been polluted by supernovae Type I. This model is supported by both the observed age-spread as the dichotomy in metallicity in the disk. While the thin disk is much younger, the thick disk seems very comparable in age with the youngest globular clusters. Also the thin disk stars are much more metal-rich ( $[\text{Fe}/\text{H}] \approx -0.5$  until  $+0.3$ ) compared to the thick disk population ( $[\text{Fe}/\text{H}] \approx -0.6$ ).

### 2.3 Streams and Clumps

If the Milky Way was indeed formed by the merging of several galaxies, and this merging process continues today, we should be able to see the remnants of these events at the present epoch. The most remarkable example of this is the disrupting Sagittarius dwarf spheroidal galaxy (Sgr) which was serendipitously discovered by Ibata, Gilmore and Irwin (Ibata et al., 1994). They found this dwarf galaxy to be very similar to the other eight known dwarf spheroidal galaxies surrounding the Milky Way, but it is much closer and is showing clear signs of tidal disruption as it is elongated towards the plane of the Milky Way galaxy. Other large-scale features found in the Galaxy are the Monoceros stream, or Low-Latitude stream, a relatively broad stream of stars of unknown origin discovered by Newberg et al. (2002) and the Virgo substructures, among which are a reported stellar overdensity (the Virgo Over-Density (VOD)) which expands over  $1000^\circ$  (Juric, 2005; Duffau et al., 2006), an excess of RR Lyrae variables which is called the Virgo Stellar Stream (VSS) and an overdensity of stars near  $(l,b) = (297^\circ, 63^\circ)$  all in the direction of the stellar constellation of Virgo. Whether all of these substructures could be part of the same large structure is yet a matter of debate (this is discussed further in Section 5.1.2). Additional substructures known are several tidal tails from globular clusters (e.g. Odenkirchen et al., 2003, Grillmair and Johnson, 2006), dwarf galaxies (e.g. Martinez-Delgado et al., 2001, Irwin and Hatzidimitriou, 1995) and unknown origin of which the most clear example is the so-called “Orphan Stream” (Belokurov, 2007; Grillmair and Dionatos, 2006).

The existence of substructure as a sign of galaxy evolution is more generally supported by the observations of substructure in the stellar halos of other galaxies (Shang, 1998; Ibata et al., 2001). For the Milky Way the evidence for debris from merging systems is not just originating from the observations of distinct stellar streams and structures. Bell et al. (2007) analysed the amount of substructure in the stellar halo using  $\sim 4$  million colour-selected main sequence turn-off stars in the Sloan Digital Sky Survey (SDSS). They found that the fractional RMS deviations on scales  $\geq 100$  pc from the best fitting oblate/triaxial smooth model is  $\geq 40\%$ . Hence they conclude that the stellar halo is highly structured.

#### The Field of Streams

Mapping all stars satisfying  $g - r < 0.4$  in the SDSS Data Release 5, Belokurov et al. (Belokurov, 2006), obtained a panorama of part of the Sagittarius stream in the northern hemisphere. In addition to the streams associated with Sagittarius, Figure 2.3 shows much more substructure. It is therefore also called “Field of Streams”. The forked Sagittarius arms are clearly seen almost stretching through the field in right ascension. On the

right of the plot the Monoceros stream can be identified. Another remarkable feature is the much tighter Orphan Stream of which also no progenitor is found (hence its name).

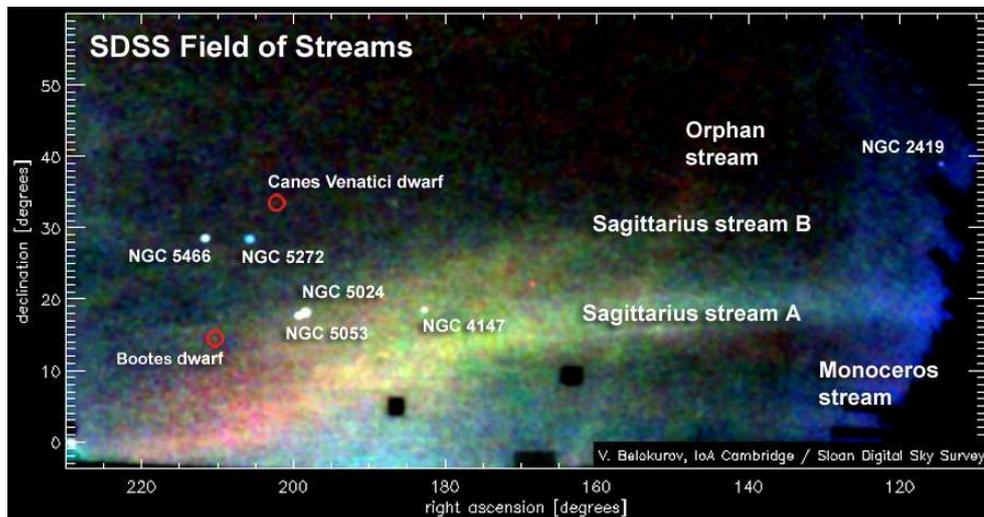


Figure 2.3: The “Field of Streams” with the marked features. Original image from Belokurov et al. 2006b. This image from astronomy.com.

The Field of Streams does not cover all known substructure today, it just maps 20% of the full sky and many features might still be missed due to their low surface brightness. Still, the Field of Streams gives an idea of the vast amount of substructure already discovered in the Galaxy today.



## Chapter 3

# The Spaghetti Survey

We expect that a large part of especially the outer halo was not formed with the collapse of the central body of the Galaxy, but was build up by subsequent merging processes. We do not know how large the contribution of these mergers was, however, and what the properties of these merged systems were. The Spaghetti Survey, as proposed in Morrison et al. (2000), aims to give a quantitative answer to the question “How much of the halo was accreted?”. To this end, a sample of halo stars is investigated out to great distances. Because the substructure due to disrupted satellites will survive longest in the outer Galaxy, distant halo tracers are particularly important to study the amount of substructure in the outer halo. The sub-sample discussed in this report consists solely of red giants. These stars are rare, which limits the methods available for substructure detection. Using a simple model and a range of assumptions about the halo density distribution Morrison (1993) estimated there are of the order of 1-10 halo giants per square degree down to  $V = 20$ . It is therefore impossible to obtain the sample sizes per field needed to use velocity histograms for detection. On the other hand, these stars provide excellent tracers for the outer halo because of their intrinsic luminosity. Red giants possess huge potential as outer halo tracers, for example: a metal-poor star near the giant branch tip with  $M_V = -2$  and  $V = 19.5$  has a distance of 200 kpc! (Morrison et al., 2000)

Unfortunately it was at the time of the start of this project not feasible to design a all-sky survey for halo substructure. The Spaghetti survey is therefore designed to be a pencil-beam survey of high-latitude fields, using CCD photometry. Because the halo provides only a very small fraction of stars ( $\sim 800:1$  locally for disk stars versus halo stars (Morrison et al., 2000)) it is important to use an efficient preselection method for the red giants. Because a classification by photometry is not 100% accurate and misclassifications will lead to serious errors in the interpretation of the data

(Dohm-Palmer et al., 2000), the next step to take is follow-up spectroscopy, which is also clearly needed to get a clean sample of K giants in the halo.

### 3.1 My part of the project

My research project, which is my contribution in the greater scheme of the Spaghetti Survey, actually consists of two parts. My first three months I spent in Cleveland, USA to work with Prof Heather Morrison and dr. Paul Harding on reductions of the spectroscopic data. In these three months I reduced 3 runs out of the total 13 spectroscopic runs in the project, which were taken at December 2002, May 2003 and June 2003 at the Magellan 6.5m telescope. This work also included the determination of radial velocities, distances and metallicities as well as making the dwarf/giant distinction. While the main technique for these reductions and calibrations had been set up for the largest part, as discussed in sections 3.3.1, 3.3.2, 3.3.3 and 3.3.4, we managed to improve a certain number of steps. We started to use SrII  $\lambda$  4077 lines, discussed in section 3.3.3, as an extra luminosity measurement for high S/N spectra. This method was subsequently further quantified by Heather Morrison. Also we had a closer look at the impact of the metallicity calibrations which eventually resulted in a change of the globular clusters metallicity scale to that of Kraft and Ivans (2003).

Once the spectra were reduced, we ended up with a sample over a hundred confirmed K giants from the galactic halo. The second part of my project consisted of the analysis and interpretation of this data set and was carried out at Groningen, the Netherlands under supervision of dr. Amina Helmi. This part of the project is described in chapters 4 and 5.

This chapter aims to give an overview of the data reduction and calibration within the Spaghetti project which in the end led to a data set of confirmed giants with spatial, radial velocity and metallicity information. It thus discusses not only my work, but that of the whole Spaghetti team. Most attention will however be given to the parts of the process in which I was (partly) involved.

### 3.2 Photometry

The preselection of giant candidates by photometry as summarised here is described in more detail in Morrison et al. (2000), Dohm-Palmer et al. (2000) and Morrison et al. (2001). The Washington colour system is particularly fit for this purpose. The system consists of four filters  $M$ ,  $T_2$ ,  $C$  and David Dunlap Observatory “51” filter (51). The  $M - T_2$  colour can be used as a temperature indicator, which transforms well to  $V - I$ . The 51 filter is especially designed to make the distinction between giants and dwarfs for G and K stars. It is centred around the Mg  $b$ /MgH region near 5170 Å,

which is much stronger in dwarfs than in giants, although it begins to lose sensitivity to luminosity blue-ward of  $M - T_2 = 1.2$ . The Spaghetti data's measuring errors for  $M - 51$  are  $0.02 - 0.04$  mag (Morrison et al., 2000), which allows easy discrimination between giants and dwarfs except for the most metal poor ones. The giant selection region in the  $M - T_2$  versus  $M - 51$  diagram is bounded by  $M - T_2 = 1.1$  and  $1.8$  and  $M - 51 = -0.02$  and  $0.09$ , also excluding the area below the line between  $(M - T_2, M - 51) = (1.1, 0.02)$  and  $(1.2, -0.02)$ .

Because all but a few percent of the halo stars are known to have metallicities  $[Fe/H] < -1$  metallicity is used as an extra selection criterion. Since the  $C - M$  colour can be used as a metallicity indicator, the extra requirement is that the candidates are living in the  $[Fe/H] < -1$  region of the  $M - T_2$  versus  $C - M$  diagram (Morrison et al., 2000).

Using these selection criteria, most of the foreground dwarfs can be excluded from the sample. The remaining contamination is due to two effects:

- Very metal-poor foreground dwarfs (also called subdwarfs) are photometrically indistinguishable from halo giants, they fall in the same regime in the  $M - T_2$  versus  $M - 51$  diagram.
- Photometric errors can scatter disk dwarfs into the halo region of the  $M - T_2$  versus  $M - 51$  diagram.

At  $M < 18.5$  spectroscopic follow-up observations show that the photometry is good enough to reach a very high efficiency in the elimination of the second group, the normal dwarfs. While this efficiency goes slightly down for fainter sources, the most serious concern is the contamination by the very metal-poor halo dwarfs whose weak lines make them indistinguishable with the Washington photometric system from the giants. Figure 3.1 shows a  $M - T_2$  versus  $M - 51$  diagram including the giant selection box and the various giant and dwarf populations from a subsample of the photometric data in the survey.

### 3.3 Spectroscopy

The large intrinsic luminosity of red giants in principle enabled the Spaghetti team to restrict the spectroscopic follow-up to 4 m-class telescopes, although larger telescopes require less integration time per object so are still favoured. Follow-up spectra were taken during runs on the Kitt Peak National Observatory (KPNO) 4 m, Cerro Tololo Inter-American Observatory (CTIO) and the Magellan 6.5 m telescope. At KPNO the RC spectrograph and KP007 grating were used, giving a spectral range of 3500 to 5900 Å and a resolution of 3.5 Å. At CTIO the RC spectrograph was combined with the KPLG1 grating which resulted in a spectral range of 3500 to 6450 Å and a resolution of 2.8 Å. The Magellan data were taken with the B&C spectrograph and

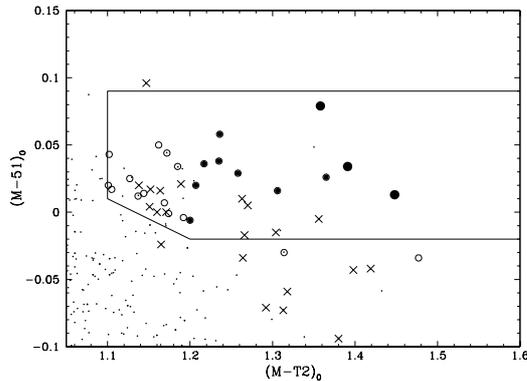


Figure 3.1: Position of the giant selection box and the most luminous giants in the M-51 vs. M-T<sub>2</sub> diagram showing photometry from a subset of the survey field (dots), subdwarfs (crosses), giants with  $M_V > 0.5$  (open circles),  $-0.5 < M_V < 0.5$  (small filled circles) and  $M_V < -0.5$  (large filled circles). It can clearly be seen that the different types of stars have their own position in this diagram. Figure from Morrison et al. 2001.

a 600 line grating, giving a wavelength range from 3850 to 5300 Å and a resolution of 2.5 Å. In general the aim is to get a S/N of 15 at the Ca I  $\lambda 4227$  line for the program stars.

The runs I reduced were three runs all taken with the Magellan 6.5 m telescope in December 2002, May 2003 and June 2003. In the following sections I will describe the general process of reduction together with some examples from the runs I reduced.

### 3.3.1 Spectroscopic Reduction

The spectroscopic data was reduced using Image Reduction and Analysis Facility (IRAF) packages. IRAF is written and supported by the IRAF programming group at the National Optical Astronomy Observatories (NOAO) in Tucson, Arizona.

First, the image is trimmed and over-scan strip correction is applied to get rid of the positive direct current offset on the CCD. Also bad pixels are interpolated. To get rid of the 2-D bias variation the taken zero frames are combined, cosmic rays are excluded by rejecting the highest (and lowest to avoid extra bias) pixel values and the resulting frame is subtracted from the images. Secondly, the images are corrected for the sensitivity of the CCD pixels by dividing by the average pattern of flat fields. To subtract the overall shape of the flats, which is not a true feature of the CCD pixels, the

response task is used. The rest of the images is divided by this response. A final step in the CCD reduction process is the illumination correction, correcting for the pattern parallel to the slit on the spatial axis. This is done using a combined image of the twilight images, because they are bright and uniform. The images are divided by this correction frame, which can consist of a number of bins to correct for variations perpendicular to the slit.

The next step is the extraction of the star spectrum from the 2-D CCD information. The aperture is selected using the IRAF task `apall`. The extraction of the arc images is a point of extra concern in particular runs of the Magellan Telescope data. These arc spectra are taken after each observation using a HeNeAr lamp and are used to correct for shifts in the spectrum as the telescope changes position. To construct the spectrum of the HeNeAr lamp, two prisms are pushed into the light beam. These two prisms unfortunately do not cover the whole slit and they also distort the light. In order to recover the needed accuracy of 0.1 pixel, the arc images are extracted carefully by averaging the lines over a large area along the slit using a previous extraction of a bright star as a standard template at the most trusted regions of the arc images. Once the arc images are extracted properly they can be used to transform the scale of the image spectra from pixel numbers to physical wavelengths. First the lines are identified in a combined image of HeNeAr spectra and secondly in the separate arc spectra taken after each observation. Using the `dispcor` routine in IRAF the images are transformed to a wavelength scale by the use of their corresponding arc images. Finally, all spectra taken of the same object are combined to improve the signal to noise ratio.

### 3.3.2 Radial Velocity

In order to get the true radial velocities for our program stars, we compare them to radial velocities standard stars observed during the same runs. To account for changes in the setup usually every observing run is treated separately. In one Magellan run (May 2003) the observing nights within one run had to be treated separately, because the slit width was changed in between nights. Two spectra, of for example a program star and a radial velocity standard star, can be cross correlated in Fourier space to obtain a wavelength shift which can be transformed to a velocity shift. In this comparison the continuum is subtracted and cuts are made at both low and high frequency to discard broad features and features above the spectrograph's resolution. First, the standards are cross-correlated with each other to determine which standards are best to be used. Figure 3.2 shows the cross correlations for the standards in the Magellan run of June 2003. Some of the radial velocity standards used in this run are multiple observed by several observers which agree within a few km/s. Two standards, HD81713 and GPEC1834 are only observed once and have a much larger error of around 10 km/s. Properties

Standard	dwarf/giant	[Fe/H] (dex)	$M_V$ (mag)	$M - T_2$	$V_r$ (km/s)
HD 134440	d	-1.5	9.4	1.22	$308 \pm 2$
HD 81713	g	-0.56	8.9	1.21	$42 \pm 10$
GPEC 1834	g	-0.99	11.8	1.12	$74 \pm 10$
HD 97	g	-1.21	9.6	1.10	$76 \pm 1$
HD 83212	g	-1.47	8.3	1.41	$110 \pm 1$
HD 165195	g	-2.14	7.3	1.61	$0 \pm 5$

Table 3.1: Properties of the radial velocity standards observed in the Magellan Jun 2003 run

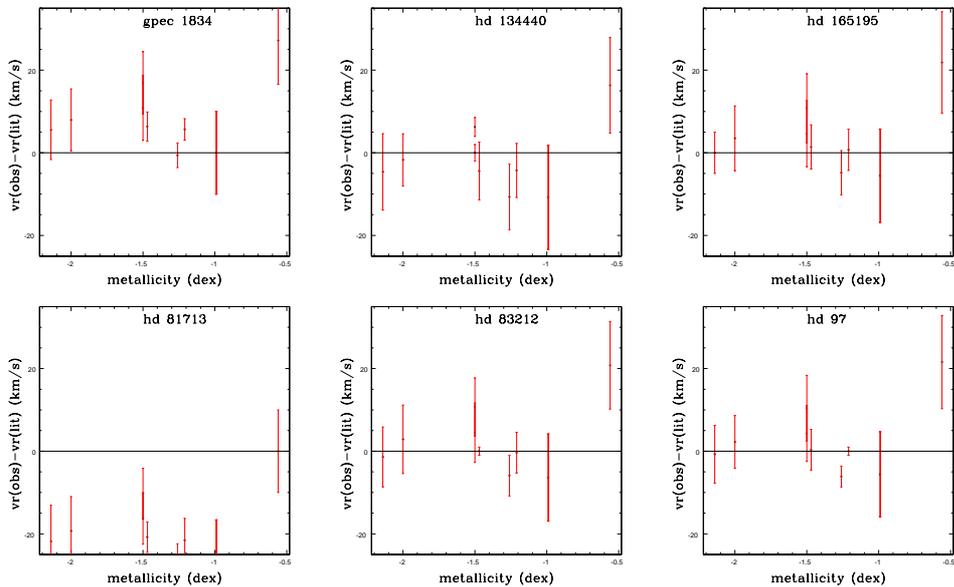


Figure 3.2: Cross correlation of the radial velocity standards in Table 3.1 with each other. Although correlations between stars with very different properties (metallicity and colour) are not expected to match very well, systematic errors like in HD81713 are not expected and probably the result of a wrong alignment of the star on the slit.

of the standards are shown in Table 3.1.

While some scatter is expected in the correlations, some of the standards show quite systematic offsets. The published uncertainty in its radial velocity is not always sufficient to account for the whole effect. In the Magellan June 03 run this is for example the case for HD81713 as can be seen in Figure 3.2. Other than an inaccurate literature value, this systematic offset might also be explained by a wrong placement of the star on the slit (Tonry and Davis, 1979). For every run only the standards with the best results in predicting the other standards radial velocities are used, also taking into ac-

Index	Line Band	Blue Sideband	Red Sideband
K'	3930.7-3936.7	3908-3918	4010-4025
Ca I	4221.7-4231.7	4147-4164	4240-4247
Mg	5130-5200	4935-4975	5217-5258

Table 3.2: Line bands and side bands for the three main indices.

count that a wide as possible range of standards is chosen concerning colour, luminosity and metallicity. As all program stars are correlated with the chosen list of radial velocity standards no a-priori assumptions have to be made regarding their properties. The match with the highest Tonry-Davis ratio (TDR (Tonry and Davis, 1979)), which scales approximately with the signal to noise ratio is selected. IRAF automatically makes the heliocentric correction (for the Earth's rotation and orbital motion around the Sun) by reading the appropriate header information.

The error introduced by flexure of the spectrograph as the telescope moves was also checked and found negligible. Close examination of the corresponding arc images shows the maximum arc shift is about 0.03 pixel, which typically corresponds to an error in the velocity of 1.6 km/s, significantly below the observational error determined by the tightness of the correlation, which is typically 15-20 km/s.

### 3.3.3 Indices

In order to get information on the metallicity and luminosity of the star, we measure indices for several lines. The indices used here are pseudo equivalent widths, they do not depend on the fitting of the continuum but just on a simple approximation of the continuum via linear interpolation between two adjacent and relatively line-free continuum bands.

The main indices we use for luminosity classification and metallicity determination are the Ca II K line, Ca I  $\lambda$ 4227 and the Mg *b*/H features. The line bands and sidebands I used on the Magellan data to calculate these indices are given in Table 3.2. In order to correct for different continuum spectrum shapes which are caused by spectrograph and detector response, a very metal-poor blue star spectrum is chosen. Its spectrum is fitted by a smooth cubic spline (order 5-7) function. The program star spectra are subsequently divided by this smoothed spectrum and moved to rest wavelength before the indices are measured. This division is necessary because the spectra were not flux calibrated. In the Magellan data that I reduced no night-to-night differences were noted in multiple observations of the same star and all indices did agree well.

Because Ca II K has little dependence in luminosity it acts merely as a metallicity indicator. Ca I  $\lambda$ 4227 and the Mg*b*/H features on the other hand,

Standard ID	$(M-T_2)_0$	[Fe/H]	K'	Ca I	Mg
G 56 48	1.16	-2.00	8.787	-0.254	3.797
G 160 30	1.67	-3.00	6.929	-0.280	3.344
G 194 37	1.20	-2.03	10.204	-0.183	6.892
HD 46663	1.37	-2.11	9.578	-0.042	12.356
HD 98281	1.01	-0.50	10.456	-0.199	7.597
HD 108564	1.34	-1.20	11.171	0.204	18.978
HD 117635	1.08	-0.70	10.425	-0.135	10.449
HD 134440	1.22	-1.50	11.210	-0.060	9.554

Table 3.3: Spectral indices for the standard dwarf stars observed in the three Magellan runs I reduced.

are sensitive to both luminosity and metallicity. Most ‘normal’ disk dwarfs can be excluded by their high Ca I  $\lambda 4227$  indices and strong Mg  $b/H$  features for their  $M - T_2$  value. Although metal-rich giants can show MgH features, they are much broader in dwarfs, due to the increased ease of molecule formation in the denser atmosphere. Subdwarfs (metal-poor dwarfs) can be selected because of their high Ca I  $\lambda 4227$  indices and strong Mg  $b/H$  features compared to the strength of the Ca II K line. To quantify the difference between the various indices we derive metallicity calibrations for all three indices using only globular cluster giants as calibration stars. The metallicity scale for globular clusters of Kraft and Ivans (2003) is used for this purpose. The behaviour of the Ca K, Ca I Mg $b/H$  index with temperature, metallicity and luminosity of the globular cluster and field standard stars used for the Magellan runs is shown in Figures 3.3, 3.4 and 3.5. The lines drawn through the globular cluster stars are serving as interpolation boundaries for the metallicity, which can then be calculated for each index separately for all the program stars. All the exact index values for the standard stars are given in Tables 3.3 and 3.4 for the standard dwarf and giant stars respectively. If a standard was observed multiple times, the values for the index are averaged.

Subsequently the metallicities of the separate indices are compared to determine whether a star is a giant or a dwarf. Because the metallicity scale used is calibrated for giants, dwarfs will show discrepant results for their MgH and Ca I metallicities as compared to their Ca II K metallicity estimate. Examples using this classification scheme for dwarf/giant discrimination for a giant, dwarf and subdwarf program star are plotted in Figure 3.6.

In addition to the use of these three indices, which are described in more detail in Morrison et al. (2003), the strength of the Sr II line at  $4077 \text{ \AA}$  is compared to three nearby Fe I lines. Because Sr is easily ionised the giants show strong Sr II lines. In the dwarfs however the collisional recombination process turns a portion of the Sr II in Sr I. The absolute strength of the lines is temperature and metallicity dependent and thus the dwarf/giant

Standard ID	$(M-T_2)_0$	[Fe/H]	K'	Ca I	Mg
47Tuc 5636	1.44	-0.62	12.235	-0.137	5.798
47Tuc 5640	1.20	-0.62	11.142	-0.247	5.638
47Tuc 5645	1.31	-0.62	11.443	-0.220	5.796
47Tuc 6603	1.07	-0.62	10.603	-0.316	3.187
BD 6 648	1.59	-2.04	9.707	-0.321	2.036
BD 9 2574	1.14	-1.95	8.465	-0.367	1.696
BD 9 2860	0.95	-1.67	7.711	-0.347	1.387
GPEC 1834	1.12	-0.99	10.335	-0.326	3.329
GPEC 3672	1.13	-0.66	10.794	-0.301	3.658
HD 97	1.10	-1.19	10.478	-0.321	2.641
HD 35179	1.22	-0.67	10.659	-0.296	4.953
HD 81713	1.21	-0.56	11.034	-0.269	4.134
HD 83212	1.41	-1.49	11.099	-0.287	2.558
HD 107752	1.22	-2.74	5.855	-0.367	0.842
HD 111721	1.08	-1.54	10.314	-0.325	2.810
HD 165195	1.61	-2.31	9.413	-0.312	1.626
NGC 1851 173	1.21	-1.22	11.422	-0.298	2.456
NGC 1851 293	1.31	-1.22	11.282	-0.290	3.436
NGC 1851 315	1.04	-1.22	9.103	-0.333	1.838
NGC 1851 319	1.41	-1.22	11.861	-0.239	3.811
NGC 1851 324	1.04	-1.22	9.612	-0.334	2.305
NGC 4590 71	1.13	-2.34	6.828	-0.354	1.243
NGC 4590 73	1.43	-2.34	8.663	-0.333	1.443
NGC 6397 33	1.13	-2.02	7.949	-0.353	1.439
NGC 6397 685	1.22	-2.02	8.648	-0.346	1.445
NGC 6397 468	1.31	-2.02	9.491	-0.317	2.883
NGC 6397 669	1.50	-2.02	10.272	-0.304	2.055
NGC 6752 3	1.45	-1.54	11.295	-0.267	2.938
NGC 6752 4	1.19	-1.54	10.096	-0.347	2.167
NGC 6752 10	1.17	-1.54	8.909	-0.345	1.775
NGC 6752 78	1.13	-1.54	9.675	-0.332	2.234
NGC 6752 4396	1.25	-1.54	10.291	-0.335	2.345
NGC 4590 20	1.24	-2.34	7.260	-0.376	1.077
NGC 4590 96	1.47	-2.34	8.894	-0.336	1.211

Table 3.4: Spectral indices for the standard stars observed in the three Magellan runs I reduced.

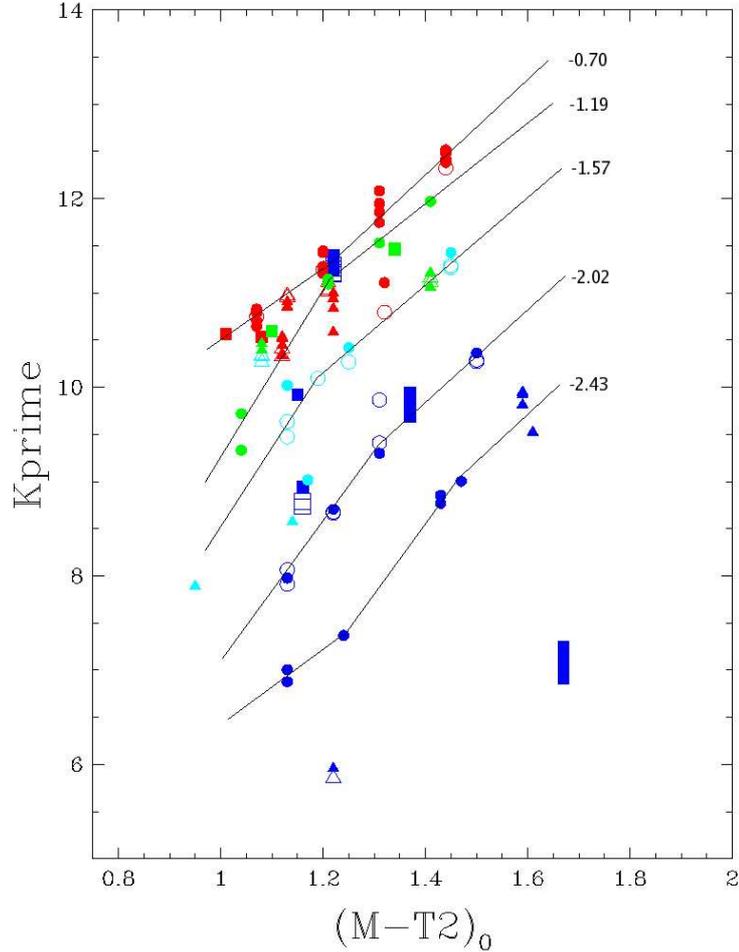


Figure 3.3: Dependence of Ca II K index on luminosity temperature and metallicity. Filled symbols are denoting data from runs feb01, dec02, jun03 on Magellan, open symbols are the Magellan may03 standards. Shown here are field giants (triangles), dwarfs (squares) and globular cluster giants (circles). The colours code for different metallicity bins: red  $> -1$ ,  $-1 > -1.5$ ,  $-1.5 > -2$ ,  $-2 > -2.43$ . The lines represent metallicity boundaries used for interpolation. Large outliers in the repeated observations are checked manually and are usually caused by misalignment with parallactic angle. In order from low to high metallicity, globular clusters used are: 47Tuc, NGC 1851, NGC 6752, NGC 6397 and NGC4590.

distinction can only be made by comparing the strength of the Sr II line to lines that are less dependent on luminosity (like the Ca II K line in the classification method described above). For this purpose the Fe I lines are very useful. Three Fe I lines are found in the direct vicinity of the Sr II  $\lambda$

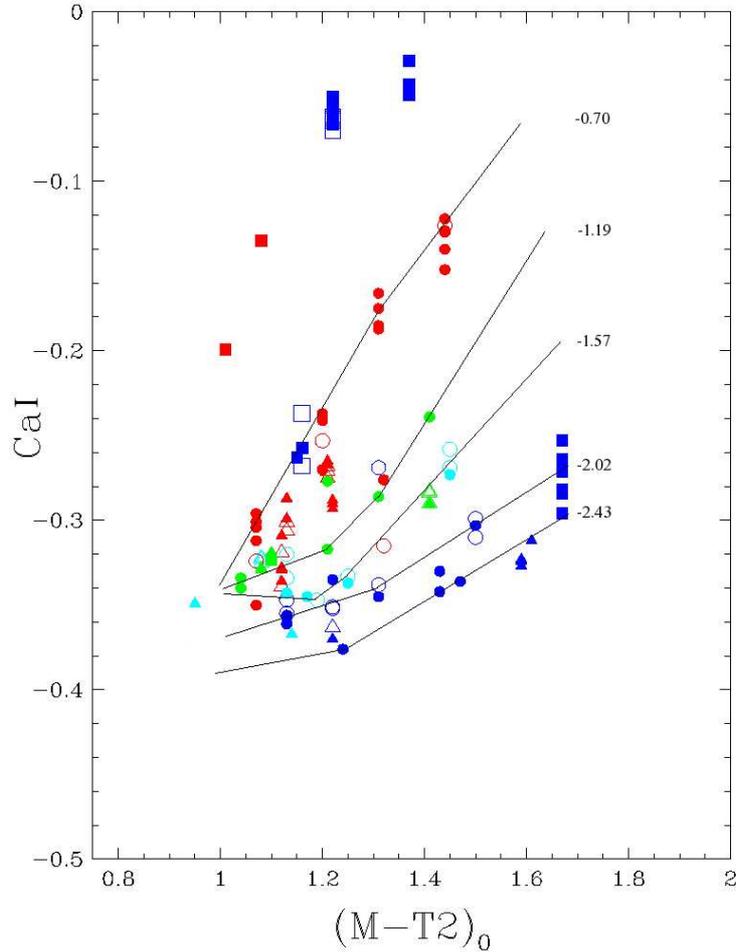


Figure 3.4: Same as Figure 3.3 for the Ca I index.

4078 line, so a comparison can be easily made. An example of the spectrum of a giant and a subdwarf star at the relevant wavelengths showing clearly different features is shown in Figure 3.7.

Close visual examination of the spectra of all the standard stars used shows that although the Sr II  $\lambda$  4078 feature has a weak dependence on colour it is very luminosity dependent, even showing clear distinctions between giants and subgiants. Also in the metal-poor standards the difference between the Sr II and Fe I features is strong enough to tell the difference between a metal-poor giant and a subdwarf. The disadvantage of this classification is however that a certain S/N ratio is required to evaluate the line strengths: at least a S/N of 6 at 4100Å. If the S/N ratio of the program

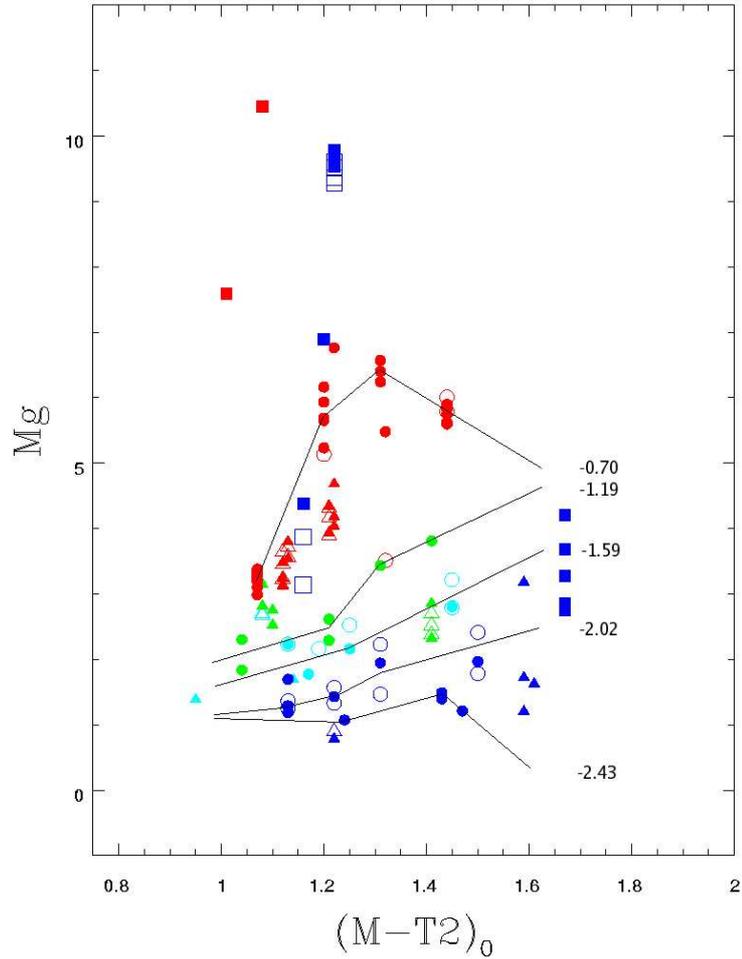


Figure 3.5: Same as Figure 3.3 for the Mg index

star allows this however, we perform the Sr II/Fe I measurement as an extra, sometimes even decisive, test.

### Metallicities of program giants

Once a program star is classified as a giant, its metallicity is determined using the metallicity estimates of the three indices, Ca II K, Ca I  $\lambda$  4227 and Mg  $b/H$ . Figure 3.4 shows that the Ca I index has a large scatter for basically the whole metallicity range. This index has very narrow line and continuum bands and is therefore not fit for accurate metallicity determination purposes. The right panel shows that the Mg index is more reliable at the relatively higher metallicities. On the other hand, the Ca II K index

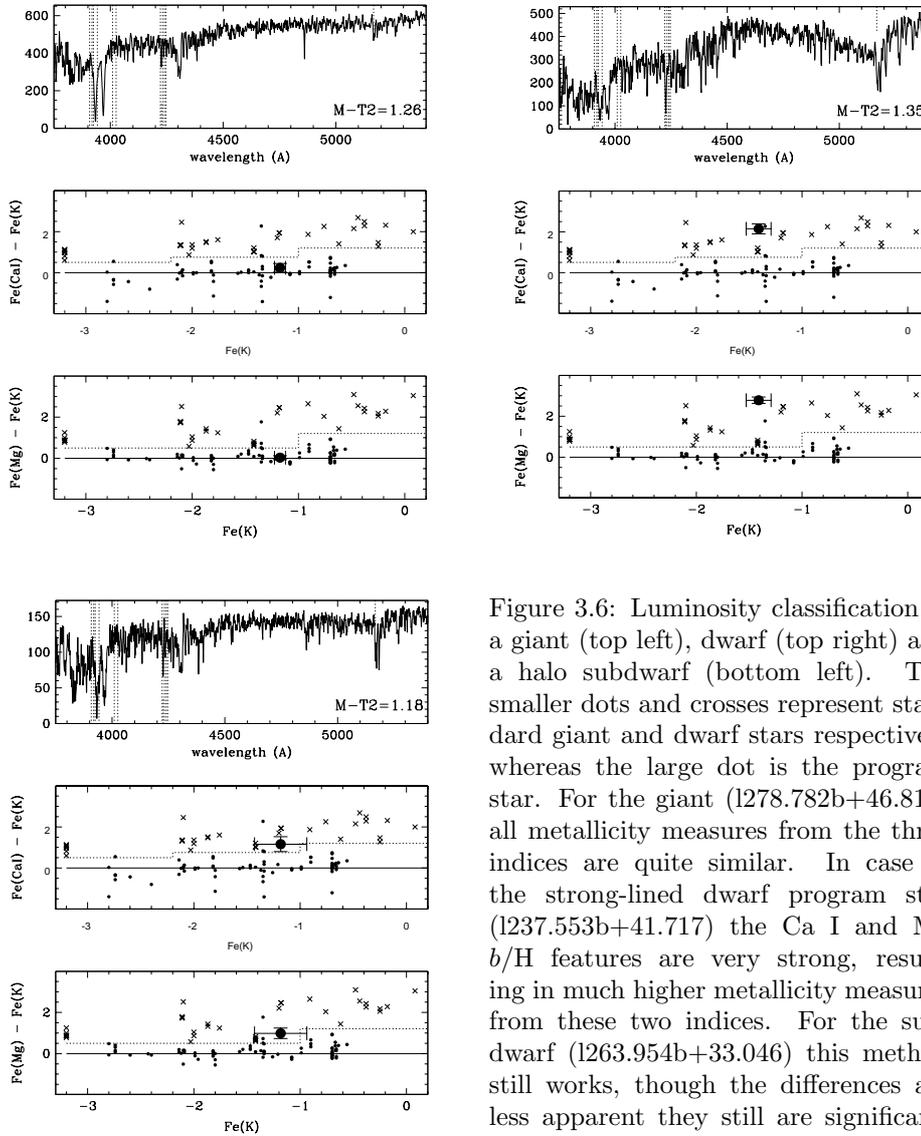


Figure 3.6: Luminosity classification of a giant (top left), dwarf (top right) and a halo subdwarf (bottom left). The smaller dots and crosses represent standard giant and dwarf stars respectively whereas the large dot is the program star. For the giant (1278.782b+46.819) all metallicity measures from the three indices are quite similar. In case of the strong-lined dwarf program star (1237.553b+41.717) the Ca I and Mg  $b/H$  features are very strong, resulting in much higher metallicity measures from these two indices. For the subdwarf (1263.954b+33.046) this method still works, though the differences are less apparent they still are significant. Figures from Morrison et. al 2003.

loses sensitivity exactly in that domain as the isometallicity lines get closer and closer together for  $[\text{Fe}/\text{H}] > -1$ . For program stars with a metallicity in Mg  $> -1.2$  we trust the interpolation of the MgH index. If the Ca II K is below  $[\text{Fe}/\text{H}] = -1$  however, we trust the Ca II K value. In the case both metallicities are to be trusted, the two values are averaged. If both are not to be trusted however, the values are still averaged, but the error is raised to 0.5 dex.

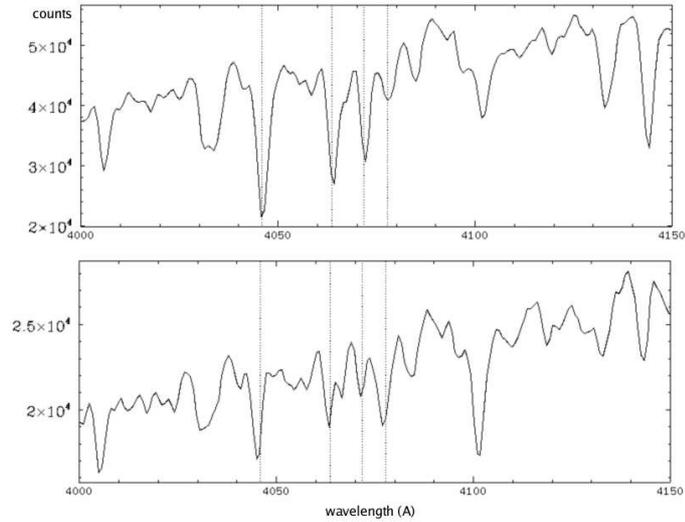


Figure 3.7: The strontium line (most right) and three of the Fe I lines for a subdwarf (HD 134440) at the top panel and a giant (HD 97) at the bottom panel.

Metallicity errors are calculated from two different contributions:

- A systematic error taking into account errors in the measurement of  $[\text{Fe}/\text{H}]$  in the standards, offsets between metallicity scales for globular clusters and field stars and measurement errors of the indices for our stars. At a S/N of  $10\text{-}15 \text{ pixel}^{-1}$  at  $\lambda 4227$  the  $[\text{Fe}/\text{H}]$  can be measured to a photon-statistic accuracy of  $0.2\text{-}0.3$  dex using the Ca II K and the Mg indices.
- Using the APSUM task in IRAF, pixel-by-pixel estimates of the standard deviation,  $\sigma$  are derived. Analytically then the errors are calculated for each spectral index and a Monte Carlo simulation is performed to calculate the error on the line index. This semi analytic index error subsequently acts as  $\sigma$  in a 1000 Monte Carlo simulations to calculate the effect of the changing line effect on the calculated metallicity. The random error on the derived metallicity is the standard deviation of this simulation.

More details on the calculation of the metallicity error can be found in Morrison et al. (2003). Errors in the continuum placement, due to variations over large scales between spectra taken on different nights is one of the major sources of error in our method. The final error on metallicity is obtained by adding quadratically both the systematic error and random error.

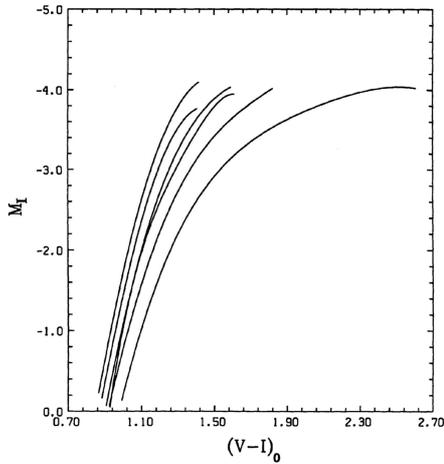


Figure 3.8: Giant branch loci in the  $(M_I, (V-I)_0)$  plane for the clusters (left to right) M15, NGC 6397, M2, NGC 6752, NGC 1851 and 47 Tuc respectively. The absolute I magnitudes are on the distance scale of Lee, Demarque and Zinn (1990). Figure from Da Costa and Armandroff (1990).

### 3.3.4 Distances

Distances were calculated for our giant stars estimating their absolute magnitude using the V-I globular cluster giant branches of Da Costa and Armandroff (1990), as shown in Figure 3.8. These branches return absolute magnitude in I as a function of V-I colour and metallicity. From the spectroscopic data the metallicity estimates are used and the  $M - T_2$  photometric colour can be transformed to V-I colour by the simple linear relation:

$$M - T_2 = 1.264(V - I) \quad (3.1)$$

The standard deviation of the residuals from this linear transformation is only 0.025 mag (Morrison et al., 2000). Subsequently, the Washington photometry colours can also be transformed to the magnitude in the V band using:

$$V = T_2 + 0.8(M - T_2) \quad (3.2)$$

Since now both the absolute magnitude and the apparent magnitude in the V band are known the distance to the program star can be calculated using:

$${}^{10}\log D = \frac{V - M_V + 5}{5} \quad (3.3)$$

Where D is the distance of the star to us in pc. Distance errors are calculated using a Monte Carlo technique on the errors on metallicity and  $M - T_2$  colours. The metallicity error is found to be the most significant contributor to the distance error, as the branches in Figure 3.8 have a strong dependence on metallicity (Morrison et al., 2003).

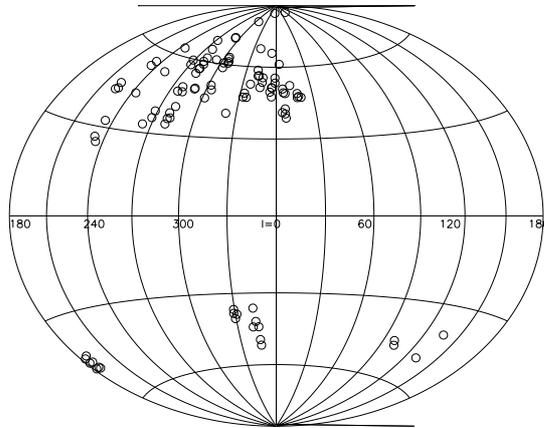


Figure 3.9: The distribution of the dataset on the sky. Plotted here are galactic longitude versus galactic latitude.

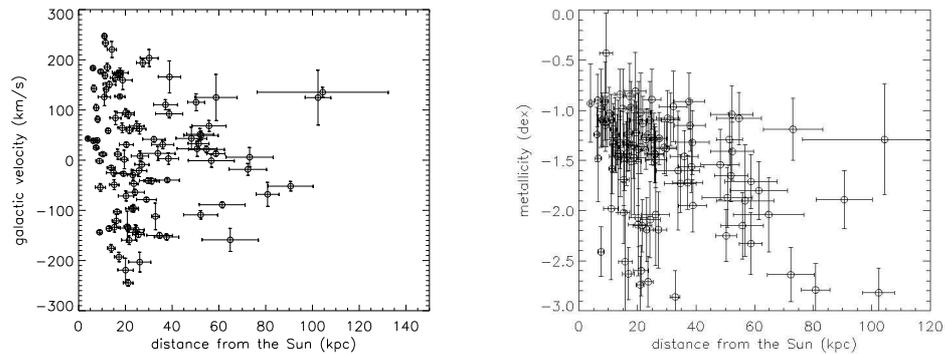


Figure 3.10: The distribution of the dataset in distance, radial galactocentric velocity and metallicity.

### 3.4 The final data set

All stars which are confirmed to be giants by the dwarf/giant distinction methods as described in section 3.3.3 are included in the Spaghetti data set. In total this data set consists of 102 giants, from 13 separate spectroscopic runs. Two giants have distances of more than 100 kpc, 33 of them have distances over 30 kpc. The typical errors on distance are 15%, on the radial velocity the typical errors are 15-20 km/s and the typical metallicity error is 0.25-0.3 dex. Distribution of this data set on the sky, on a distance versus radial galactocentric velocity scale and distance versus metallicity are plotted in Figures 3.9 and 3.10.

All program stars confirmed by luminosity classification as giants and their radial velocities, distances, metallicities and corresponding errors can be found in Table A.1.

## Chapter 4

# Analysis of the data

In order to answer the major question of the Spaghetti survey “How much of the Galaxy’s halo was built up by the accretion of small satellites?”, we want to quantify the amount of substructure we can find in this subset of 102 halo stars . Assuming that our dataset can be seen as random and can be extrapolated to larger scales, the amount of substructure in our dataset gives us a first order answer to how much substructure at similar scales we expect to find in the halo. Although several methods are generally used to find substructure, not all of them can be applied to this dataset. As discussed before in chapter 3, the Spaghetti survey K giants are not observed in large enough densities to apply a detection method based on velocity histograms per field. Instead, we have to rely on other group-finding methods.

In this chapter two methods are discussed, the first one is a simple distance in 4-parameter space (the 4distance) which uses the spatial and velocity information we have on the giants. This method is fit for finding small substructures with similar sky position, distance and radial velocities (clumps).

To complement this searching algorithm to larger structures on the sky, we implement a great circle method as developed by Lynden-Bell and Lynden-Bell (1995) and Palma et al. (2002). This method is based on the assumption that stars from a single parent conserve their angular momentum pole and this is used, in combination with a conservation of energy (which can be expressed as a function of distance and radial velocity) to find structures which might have had the same origin. This method was designed to discover structures with large separations on the sky, but its disadvantage is that it can only reliably be used for a larger amount of structures with a range of distances. This second group-finding method is discussed in section 4.2.

## 4.1 The 4distance

If a certain number of stars could be associated to a stream caused by a merger of a former satellite for instance, we expect the full phase space information, consisting of three components in space and three in velocity space, of these stars to be initially (almost) the same. During the merger of the satellite and in the halo these structures will become more and more disturbed by the Milky Way potential and therefore differentiated. Still, we expect the structures to be coherent for a long time. It is shown by Helmi and White (1999) that even when the spatial structure is no longer apparent, the structure of the merged satellite can still be recognised in velocity space. Numerical models of galactic accretion show that structures can remain coherent for many gigayears in the halo (Helmi and White, 1999; Johnston et al., 1996).

For the 102 giants in our dataset we possess information on four of the six phase space components. Three of these are the spatial components, measured as galactic longitude, galactic latitude and distance. The fourth component is the radial velocity of the star. No further proper motion information is available for stars at these large distances. With the four components we do possess however, we can define a distance measure in a four dimensional space for every pair of stars in our data. We define:

$l$  = galactic longitude

$b$  = galactic latitude

$d$  = distance to the Sun

$vr$  = radial velocity with respect to the galactic centre

$\phi$  = angular distance on the sky between the two stars (denoted as 1 and 2 in the formula), given by

$$\phi(1, 2) = a\cos(\cos(b_1) * \cos(b_2) * \cos(l_2 - l_1) + \sin(b_1) * \sin(b_2)) \quad (4.1)$$

We now define our four component distance between two stars in the following way:

$$4dist(1, 2) = \sqrt{a1(\phi(1, 2))^2 + a2(d_1 - d_2)^2 + a3(vr_1 - vr_2)^2} \quad (4.2)$$

One can easily see that his definition of distance is in its essence not very different from the ordinary definition of distance in a three-dimensional grid:

$$distance(1, 2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (4.3)$$

While the galactic longitude and latitude are incorporated within the measure of the angular distance, the other components are used totally in-

dependently in the final 4distance calculation. The main difference between this method and calculating the 3-D distance and difference in the radial velocity component separately, is that all four components are treated equally. Because we only evaluate the final distance we also pick up groups that have larger differences in one component, but are really close in the other three. The constants  $a_1$ ,  $a_2$  and  $a_3$  can be used to normalise or give relative weights to the different components. We choose to use the constants so that they normalise every contribution. Another possibility is to choose the weighting factors such that they are inverse proportional to the measurement errors in the stars. The advantages of this weighting method is that matches between stars with large errors are less likely and you take the errors into account within the method itself. However, a disadvantage of weighting with errors is that you will throw away pairs which appear to be very close in 4distance but have large errors, while you can not compensate for the pairs that are physically very close, but appear farther away because of their large measurement errors. In the end you will thus end up with less pairs on average. Another concern is you might favour pairs that have small error bars, but are not as close in 4distance. Therefore we choose to include every pair within a certain 4distance independent of its measurement errors. Every substructure we find is subsequently evaluated, in which the individual measurement errors are taken into account (see Section 4.1.5). The general effect of the measurement errors on our method is investigated in Section 4.1.4.

Every observable for every pair of stars is normalised in the dataset by choosing the largest physically possible value for each constant. The largest possible angular separation, for instance, is  $\pi$ . For the distance we normalise using the virial radius, which is about 250 kpc. For the velocities we use the escape velocity from the galaxy, approximately 500 km/s. Because velocities can be both negative and positive, the total weighting factor is 1000 km/s. Every component now will return a value between zero and one. The exact values used for the three constants are:

$$a_1 = \frac{1}{\pi^2}, \quad a_2 = \frac{1}{250^2}, \quad a_3 = \frac{1}{1000^2} \quad (4.4)$$

The sensitivity of the group-finding algorithm to these weighting values is discussed in section 4.1.4.

#### 4.1.1 Making a random set

We have defined a new distance measure between every two stars in our sample, the 4distance. We would expect that pairs with small 4distance may be possible stream members. However it is not clear within which 4distance stars are likely to have a common background.

In order to get a better grasp on when our defined 4distance starts to become

significant, a random sample is defined and compared with our data set. This random sample should not possess any streams or groups to allow us to get an idea of how many groups will be formed accidentally in a sample the size of our data sample. The most obvious random sample would be an isotropic distribution of the sky, with a certain distance distribution and a random Gaussian velocity distribution. But due to our way of observing, comparing with a random distribution on the sky will give false results. This is highlighted in Figure 3.9, where the distribution of the dataset giants over the sky is shown. This figure shows a large amount of clumpiness in the spatial distribution. This clumpiness is however not so much due to the true distribution of the red giants over the sky, as it is to the pointing of the telescope, the visibility of certain regions of the sky on the observing nights and the distinct photometric runs used to select giant candidates. Because also the angular separation on the sky enters in the  $4\text{distance}$ , the comparison of this clumpy distribution to a smoothed random sample will obviously result in a lot more small  $4\text{distance}$  values in the data sample independently of whether the data contains any true streams. This is a spurious result due to the poor choice of a random sample and we therefore would like to get rid of this effect.

One way to exclude this effect of apparent clumpiness over the sky is to give the random set the exact same sky distribution. The random set should also have to have the same size as the data set, as more stars will automatically increase the chance of finding pairs within a certain  $4\text{distance}$ . A third criterion to fulfil is that the range of distances and velocities should be approximately the same. We therefore choose to create a random set with the same galactic longitude and latitude for all the stars, but to re-shuffle their velocities and distances in a random manner. For every star in the sample the galactic longitude and latitude coordinates are preserved, but the star is randomly supplied with a different observed velocity and independently also with a different observed distance.

#### 4.1.2 Choosing a relevant binsize

We say two stars that are within a certain  $4\text{distance}$  will form a *pair*. By comparing the total number of pairs formed at a certain  $4\text{distance}$  in both the data and the random sets, the significance of the data pairs can be investigated. The number of pairs formed with various  $4\text{distances}$  for our data set compared to the average outcome of a thousand randomised sets is shown in Figure 4.1. In Table 4.1 the absolute values are given.

The number of pairs within a certain  $4\text{distance}$  can be seen as a measure for the clumpiness of the set at that particular scale. For all scales up to a  $4\text{dist}$  of 0.12 plotted in Figure 4.1, the amount of clumpiness in the data is larger than in the random set.

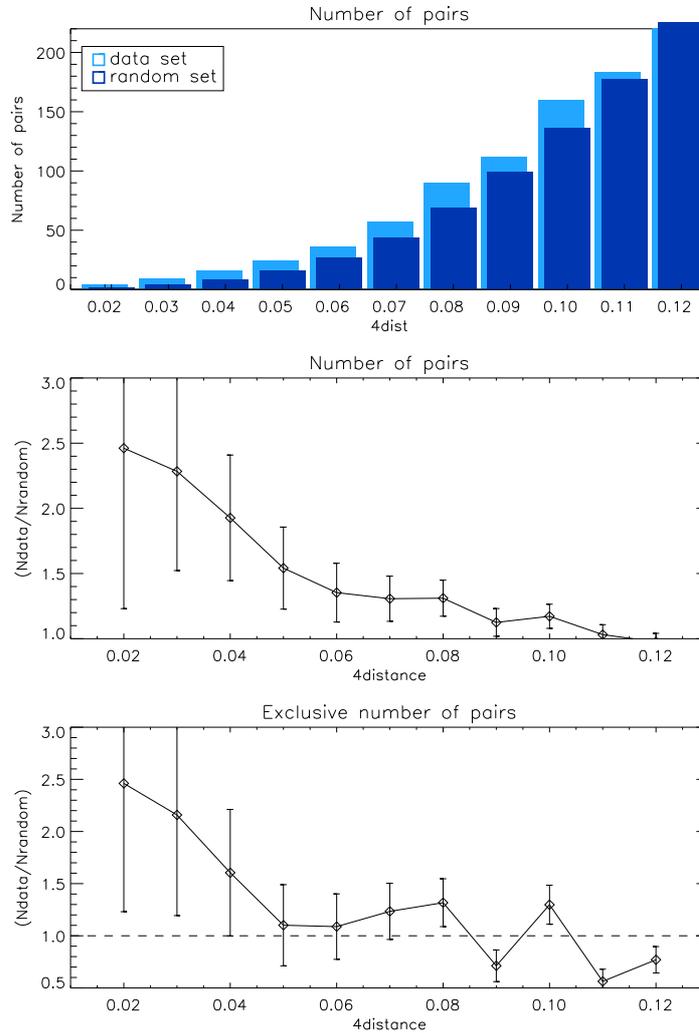


Figure 4.1: In the top panel the cumulative numbers of pairs found as function of  $4dist$  are shown. The cumulative number of random pairs is obtained by averaging 1000 random sets. The middle panel shows the cumulative correlation function defined as the number of pairs in the data divided by the average number of pairs in the random below a certain  $4dist$ . The bottom panel also shows the correlation function, but only the stars added by making one step in binsize are shown. Error bars are poissonian.

Based on Figure 4.1 shown above, we decide to investigate in more detail data pairs at two different scales. Our first choice is to focus on structures below  $4dist \leq 0.04$ . The bottom panel in Figure 4.1 shows that for  $4dist \geq 0.04$  relatively more noise is added to the pairs in the dataset. How-

4dist	Number of groups in data set	Number of of groups in random set
0.01	0	0.35
0.02	4	1.63
0.03	9	3.94
0.04	16	8.30
0.05	24	15.57
0.06	36	26.60
0.07	57	43.62
0.08	90	68.67
0.09	112	99.55
0.10	160	136.55
0.11	183	177.44
0.12	220	225.49

Table 4.1: Number of data pairs as well as the average of a 1000 random set for various values of  $4dist$ .

ever, it can also be seen that at a level of  $4dist \leq 0.08$  more data pairs are added at approximately the same level of significance as at 0.04. At even larger 4distances the correlation function shows very irregular patterns, always oscillating around the expected asymptotic value of 1. We therefore decide to first look at  $4dist \leq 0.04$ , but also take into account the information that is added to these structures for  $4dist \leq 0.04$  and  $4dist \leq 0.08$ . The choice to consider two different scales is a result of two considerations. First, we would like to choose a  $4dist$  within which the ratio of data pairs to random pairs is sufficiently large, such that our data pairs have a high chance of being real. Secondly, we would like to have enough data pairs and we would like to avoid to throw out real information in the dataset by being too selective. Our final choice takes both considerations into account.

### 4.1.3 Pairs and groups

#### Pairs and groups for $4dist \leq 0.04$

Every pair of stars in the data set within a 4distance of 0.04 is considered to be meaningful although, based on Table 4.1, we expect roughly 8 of them to be by chance. In the dataset 16 pairs are found within this 4distance. Some of these pairs can be combined to form larger groups. We distinguish two different ways to define a group (the properties of the two criteria are discussed further in Section 4.1.5). The first and less restrictive criterion states that a group of stars form a group when every star within the group has a  $4dist$  smaller than a certain limit with at least one other member of

the group. This is called this the *friends-of-friends* (FOF) criterion. Using this criterion we find a group of 3 and even one of 6 members below 0.04. This leaves 6 pairs that cannot be extended to groups with more members. The second and more restrictive criterion states that a group is found when all the  $4dist$  measures of the stars within the group are all within a certain limit. We call this the *all-friends* (AF) criterion. Using this more restrictive criterion we find just 2 groups of three and 8 pairs in our dataset below a  $4dist$  of 0.04.

The sky distribution of the dataset and pairs and groups found with the friends-of friends criterion below  $4dist = 0.04$  (denoted as FOF4 from now on) are shown in Figure 4.2. Figure 4.3 and Figure 4.4 show their velocity versus distance and metallicity versus distance distribution. Note that the metallicity component is not used as a criterion to select the groups, its value is not incorporated in the value of 4distance. While a spread in metallicity can be expected in the groups, this metallicity spread may give information about the likeliness of the group being real. If the metallicity range within a group is really large (e.g. larger than 1 dex), the members of the group are less likely to be originating from the same progenitor unless the progenitor was heavy enough to possess such a large internal metallicity range. To visualise the formation of groups in the dataset, Figure 4.5 shows the four groups in a linear multidimensional plot. Table A.2 gives an overview of all the values for the pairs and groups of FOF4. The different groups formed using the all-friends criterion below  $4dist = 0.04$  (AF4) are given in Table A.3.

### Pairs and groups for $4dist \leq 0.08$

We also look at the amount of substructure found with the two group-finding criteria at the level of  $4dist \leq 0.08$ . There are in total 90 pairs found within  $4dist = 0.08$  in the dataset. Using the friends-of-friends criterion with  $4dist \leq 0.08$  (FOF8) these can be clustered to a total of 17 groups. Of these groups there are 7 pairs, 3 groups of three, 2 of four, 1 of five, 1 of six, 1 of seven, 1 of eleven and 1 of thirteen members. The amount of large groups reduces when we look at the groups formed with the all-friends criterion below  $4dist = 0.08$  (AF8). Then the 90 pairs can be regrouped into 1 group of seven members, 1 of five, 2 of four, 8 of three and there are still 31 pairs left.

The distribution of the friends-of-friends groups over the sky is plotted in Figure 4.6. Subsequent information on all the groups and pairs found with both methods below 0.08 can be found in the appendix, Tables A.4 and A.5 for the friends-of-friends and all-friends criteria respectively.

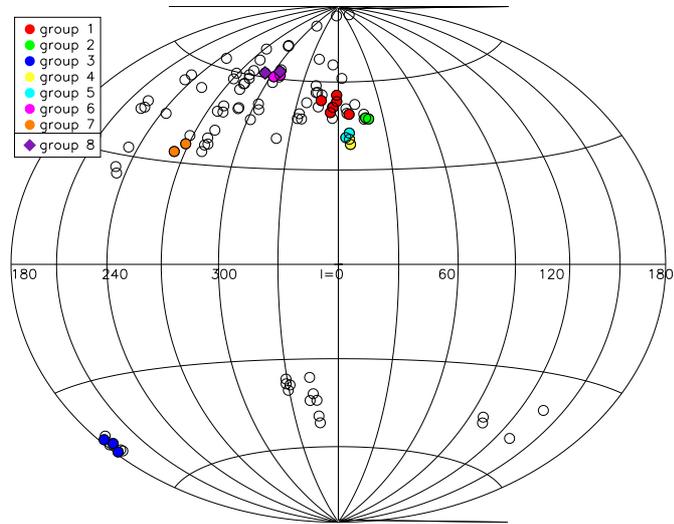


Figure 4.2: The dataset plotted on the sky in galactic latitude and longitude coordinates. Aitoff projection is used. Coloured stars are the groups found with the friends-of-friends criterion below  $4\text{dist} = 0.04$  (FOF4).

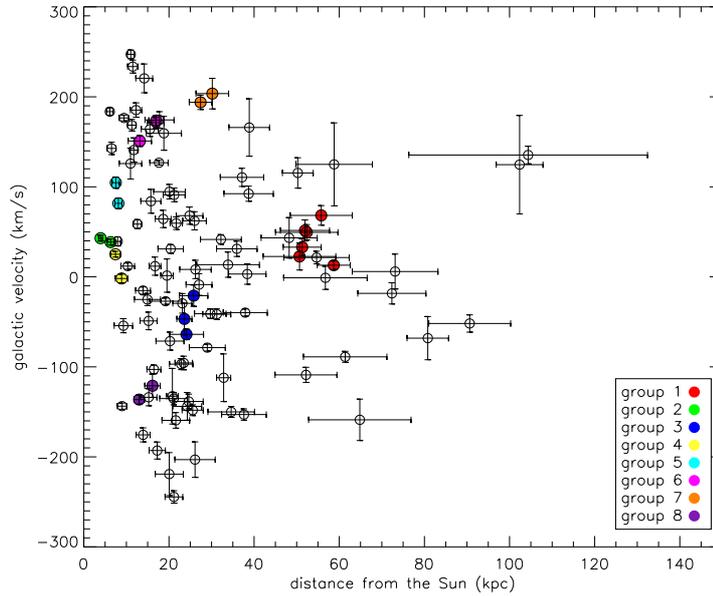


Figure 4.3: Distance versus galactic radial velocity plot for the whole dataset including the groups found with FOF4 (coloured stars).

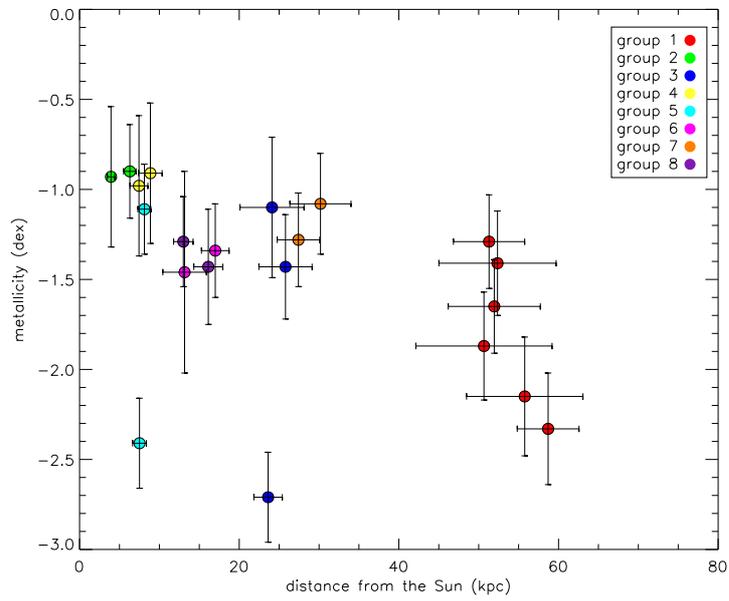


Figure 4.4: Distance versus metallicity plot for the stars in groups found with FOF4.

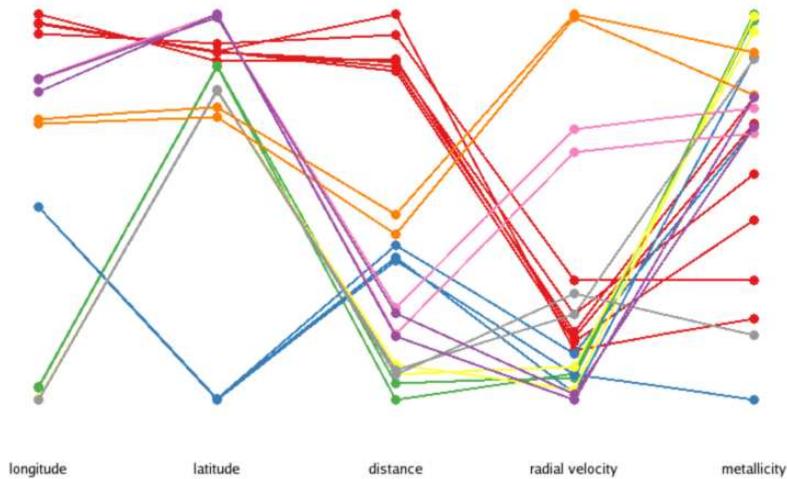


Figure 4.5: Visualisation of the grouping in five parameter space using the plotting tool gobi. Every colour denotes a different group that was found using the FOF4 group-finding criterion. The colour-coding for the groups is similar to that in Figures 4.2, 4.3 and 4.4 except that grey was used instead of light blue for group 5.

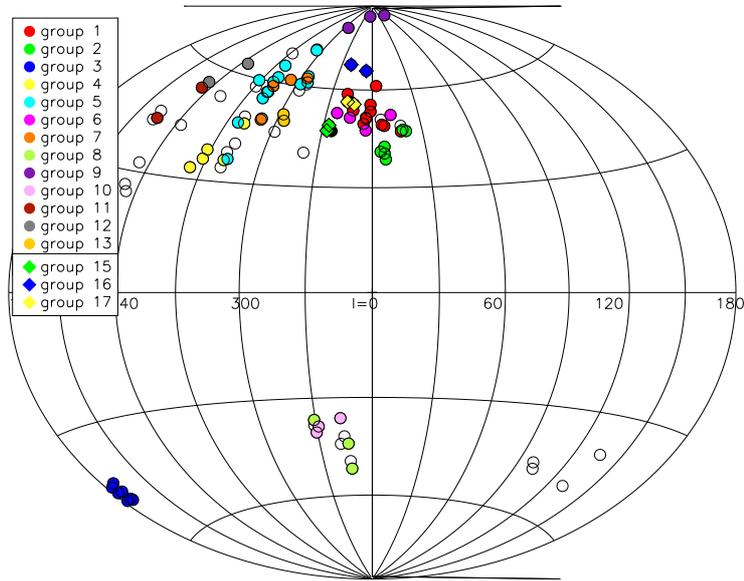


Figure 4.6: Coloured stars are the groups found with the friends-of-friends criterion below  $4dist = 0.08$

#### 4.1.4 The effect of errors and our choice of $4distance$

In the determination of the  $4dist$  measures in order to define whether a star belongs to a pair, the observational errors on the different quantities (in particular distance and radial velocity) are not taken into account. To quantify the effect of errors we performed a set of 1000 Monte Carlo simulations. In each simulation we convolve the observed distance and radial velocity with their estimated errors, assuming Gaussian statistics. Each of these 1000 “new data sets” are subject to the same analysis as performed in section 4.1.2; the number of pairs in each set is compared to 500 random sets. The results are shown in Figure 4.7. Comparison with Figure 4.1 shows that the datasets which were randomly convolved with errors show a much smoother distribution over the different binsizes. Overall the significance has decreased, but the results still have over  $1\sigma$  significance assuming Poissonian statistics. On average about 13.1 pairs are found in the datasets below  $4dist$  0.04. From the analysis of 10 datasets randomly convolved with errors we expect about 2 pairs per dataset which are not picked up below  $4dist$  0.04 in the original dataset. All these pairs are found below  $4dist$  0.08 in the original dataset however. Though the significance of the results clearly suffers from the small number counts, our previous choice to look first at a level of  $4dist = 0.04$  and later at added structures at  $4dist = 0.08$  is still justified.

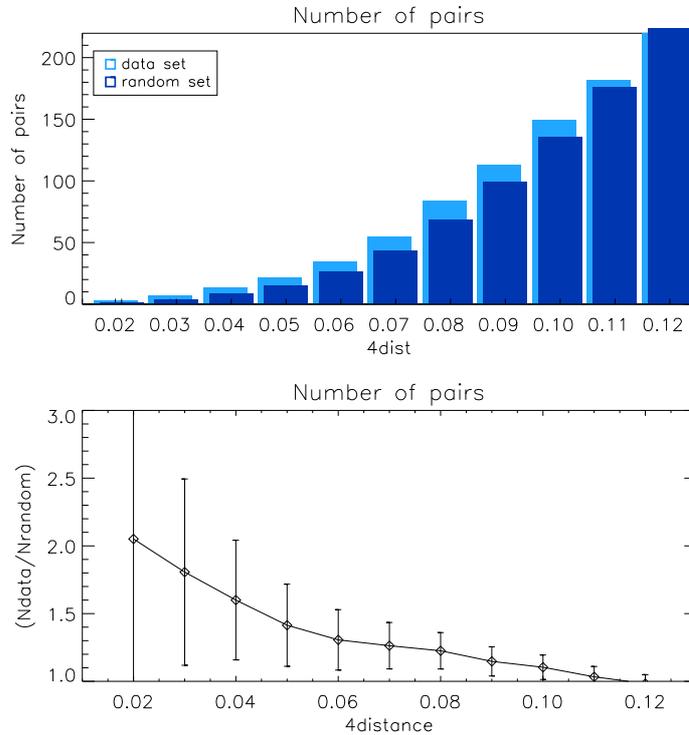


Figure 4.7: Averaged number of pairs found 1000 data sets with parameters normally convolved with errors using a Monte Carlo code. The number of pairs in the data sets from 1000 Monte Carlo simulations are each compared to 500 random sets with the same values. Top panel shows the absolute number of average pairs, second panel the ratio of data and random pairs and the bottom panel displays the number of pairs added in every bin. Error bars are poissonian.

The results obtained prove to be quite insensitive to the exact setting of the weights  $a_1$ ,  $a_2$  and  $a_3$ . Very similar results are obtained for example with the weights defined as the difference between the largest and smallest value in the dataset; the same pairs are found with the exception for the pairs which make up the largest group of 6 stars, this group then only consists of 4 stars: 2, 96, 97 and 98.

#### 4.1.5 Significance of the groups

We would also like to quantify the significance of the groups of varying sizes. Another question to answer is whether there exists some quantitative difference between the two group finding criteria: the friends-of-friends and the all-friends criterion. Tables 4.2 and 4.3 give the number of stars found in groups of certain group sizes compared to the number of stars found in

Friends-of-friends criterion			
Groupsize	$4dist$	Number of stars in data set	Number of stars in random sets
2	0.04	12	9.483
2	0.08	14	7.748
3	0.04	3	1.932
3	0.08	9	3.891
4	0.04	0	0.288
4	0.08	8	3.464
$\geq 5$	0.04	6	0.211
$\geq 5$	0.08	43	36.248

Table 4.2: Number of stars found in groups using the friends-of-friends group finding criterion and  $4dist \leq 0.04$  and  $4dist \leq 0.08$  varying in size from 2 to  $\geq 5$  group members in the data and an average of 1000 random sets.

All-friends criterion			
Groupsize	$4dist$	Number of stars in data set	Number of stars in random sets
2	0.04	16	12.418
2	0.08	42	25.287
3	0.04	5	0.986
3	0.08	19	15.658
4	0.04	0	0.036
4	0.08	5	4.830
5	0.04	0	0
$\geq 5$	0.08	8	1.699

Table 4.3: Number of stars found in groups using the all-friends group finding criterion and  $4dist \leq 0.04$  and  $4dist \leq 0.08$  varying in size from 2 to  $\geq 5$  group members in the data and an average of 1000 random sets.

similar groups in an average of 1000 random sets. Because groups can be really large and we can not evaluate all groupsizes, stars in groups with more than 5 members are counted in the last bin. Figure 4.8 shows the absolute number of groups as given in the tables in a bar plot as well as the ratio of the number of groups found in the data and random sets.

At all levels, more substructure is found in the dataset compared to the random set. There are a few points which deserve extra attention. One interesting point is that the group of 6 of FOF4 is very significant as only on average 0.2 stars are found in groups with comparable sizes in the random sets. When choosing a higher level of acceptance,  $4dist = 0.08$ , there seems to be a much higher chance that groups this size or larger

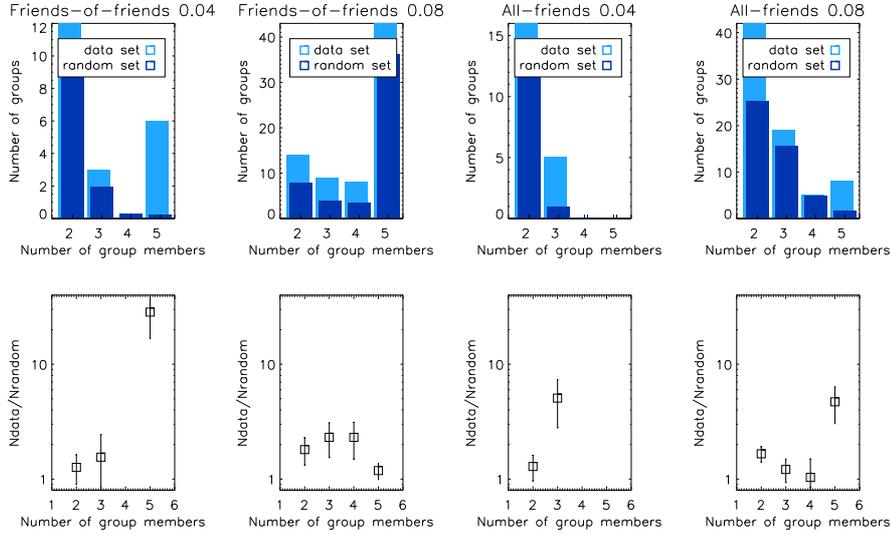


Figure 4.8: The number of groups found of different sizes per criterion and choice of  $4dist$ . Top panels show the absolute number of groups, bottom panels the ratio of the number of groups in data and random sets. Error bars are poissonian.

are formed (consistent at the  $1\sigma$  level). The other two points in the lower panels of Figure 4.8 which really stand out are the groups of 3 members of AF4 and the larger groups ( $\geq 5$  members) of the AF8 criterion. Closer inspection reveals that approximately the same stars are grouped in all these highly significant structures. The group of 6 with FOF4 consists of stars 2,93,96,97,98 and 99 in Table A.1. The 3 groups of 3 members with AF4 are formed out of different combinations of stars 2,96,97,98 and 99 and the large groups in AF8 has as its members stars 0,2,3,93,96,97,98 and 99. This group of stars, of which core members are at least 2,96,97,98 and 99 thus really stands out in the data set to a high significance level.

For the other stars which possibly form groups, we decide to take into account all groups at FOF4 and additional and smaller groups of FOF8 and AF8. The most restrictive group-finding criterion, AF4, is neglected here because just two groups larger than a pair formed using this criterion. Also, the starting point of the FOF criterion simplifies the definition of core groups, since stars cannot belong to several groups using this criterion.

An overview of the possibly extended groups found is given in Table 4.4. The groups are all plotted on the sky and in radial velocity versus distance and metallicity versus distance plots in Figures 4.9 and 4.10.

From the Figures 4.9 and 4.10 it is clear that in general the stars from FOF8 add more noise to the structure. These stars, in the figure plotted

Group	Members at FOF4	Extended with AF8	Extended with FOF 8
1c	2, 93, 96, 97, 98, 99	0, 3	10, 83, 101
2c	18, 21, 24	23	20, 22
3c	4, 7	11, 12	5, 6
4c	68, 69	51	39, 42, 45, 47, 50, 51, 54, 55, 56, 57, 61, 70
5c	36, 38	37	43, 49
6c	59, 66	-	53, 58

Table 4.4: Possible members of the same group. The ‘c’ after the group number emphasises this is a combined group from more than one group-finding criterion. The criteria used here are FOF4, AF8 and FOF8.

as open squares, are often no members of the group within their error bars. Also the systematic disadvantage of the friends-of-friends group-finding criterion becomes apparent; in the case of group 4 (bottom panel) clearly a runaway process can be observed as the stars form a long stream in all parameters. This is possible because a star is included in the group if it is close enough in  $4dist$  of only one other star. This included star can then have again one other star within this  $4dist$  and so more and more stars are added which are not necessarily close to the initial core group. On this basis we decide for a general group finding criterion which takes the groups found in FOF4 as a core, with a possible extension of the stars added in the AF8 criterion. Whether these stars are indeed likely to be a member of the group can be discussed for every group separately, based on diagrams as plotted in 4.9 and 4.10. In addition to the four parameters used in the  $4dist$  parameter space, the metallicity of the stars can be used as an extra check to include or exclude additional or even core members. Using the group selection process as described above, we get the following groups:

**Group 1.** The core members in this group are the highly significant stars found in the group of 6 with the FOF4 criterion. From Figure 4.9 we decide one of the stars from the AF8 criterion, star 0 in Table A.1, can be added to the group.

**Group 2.** From this core group of three stars, one giant has such a low and deviating metallicity that we can exclude it from the group if we assume it should be a smaller substructure. There is just one star which is within  $4dist = 0.08$  from all stars in the core group. But this star has a deviating radial velocity with the core group of at most 94 km/s. From a simple comparison with a sample of 10.000 stars drawn from a Gaussian distribution with the global velocity dispersion at that distance from the Galactic Centre (Battaglia et al., 2005) we find  $\approx 42\%$  of all stars will have a velocity difference equal or smaller. We conclude from this that star 23 is not very likely to be a member. No extra stars are thus added to the group.

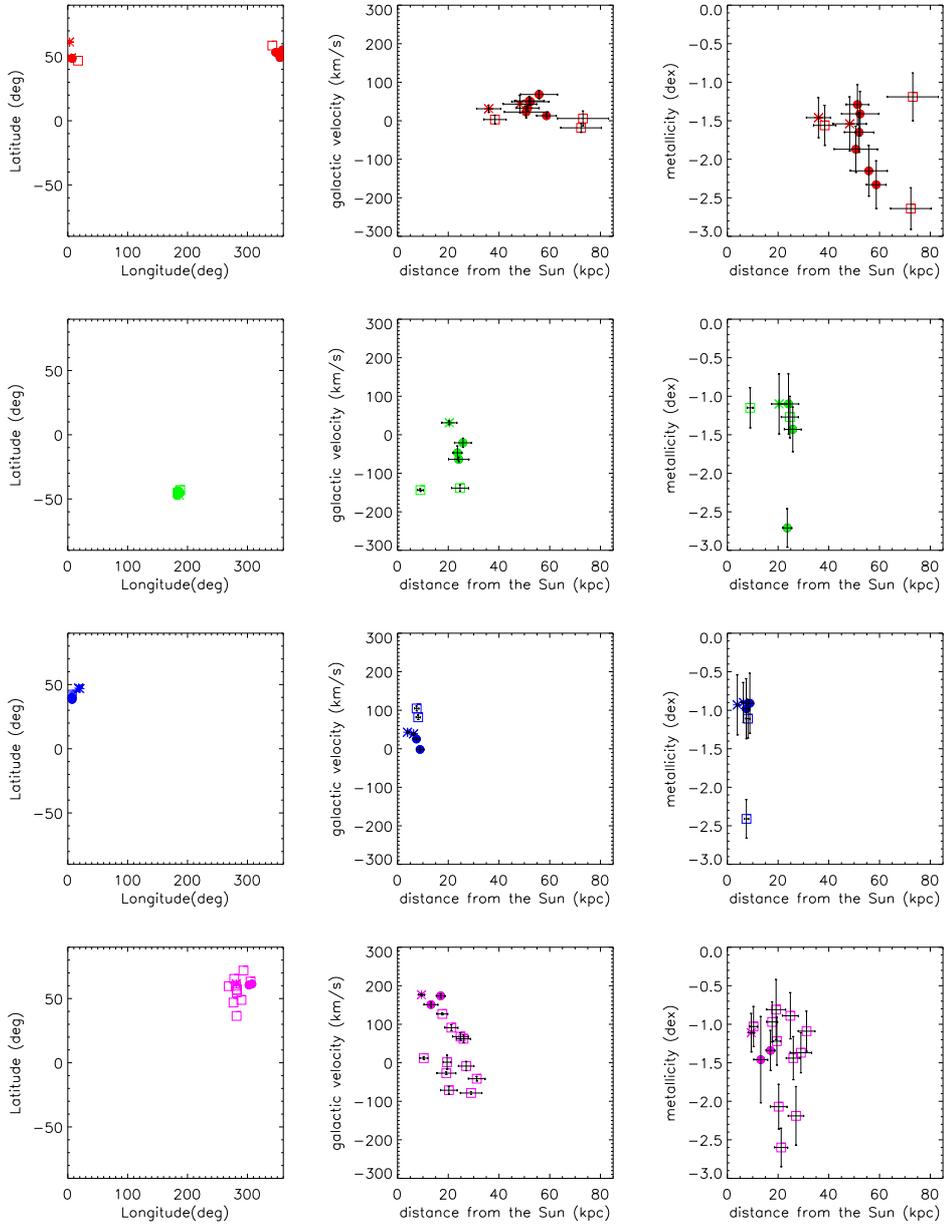


Figure 4.9: Galactic longitude and latitude, distance, galactocentric radial velocity and metallicity information for the stars which are possible members of groups 1c to 4c (top to bottom). The filled circles are the core of the group, from the FOF4 group-finding criterion. Asterisk symbols denote the stars which are additional members to the group if the AF8 groups are taken into account and open squares are giants which are added with the FOF8 criterion.

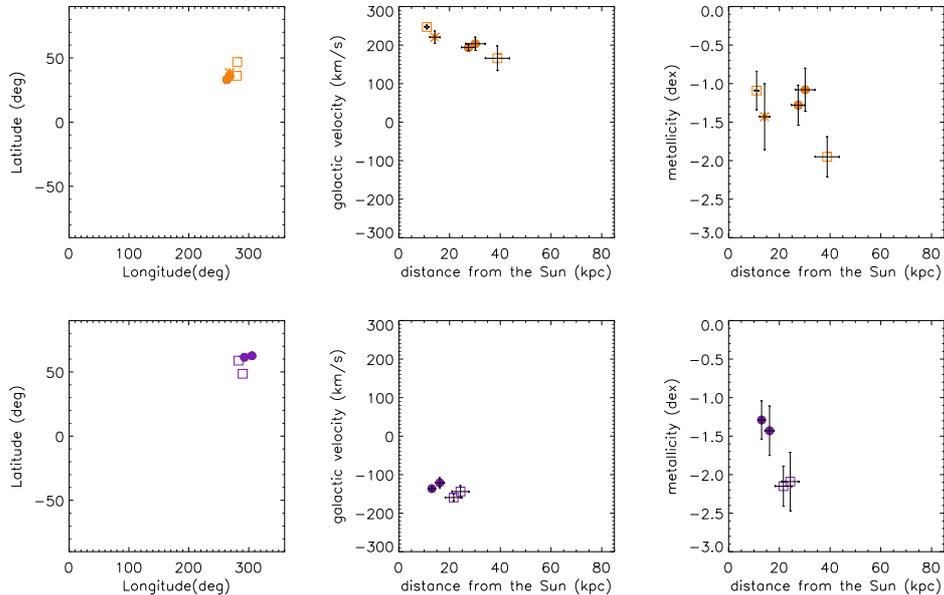


Figure 4.10: Galactic longitude and latitude, distance, galactocentric radial velocity and metallicity information for the stars which are possible members of groups 5c and 6c (top and bottom panels respectively). The filled circles are the core of the group, from the FOF4 group-finding criterion. Asterisk symbols denote the stars which are additional members to the group if the AF8 groups are taken into account and open squares are giants which are added with the FOF8 criterion.

**Group 3.** Basically this group exists of three pairs that are already picked up at a level of  $4dist = 0.04$ . The first pair: stars 4 and 7 in the dataset, match really well in velocity, distance and metallicity with the second pair from the AF8 criterion: stars 11 and 12. The third pair, 5 and 6, added with the FOF8 criterion seems to match not so well, especially in radial velocity. Again, we use a Gaussian sample to see what the chances are for stars at this distance to be so close in radial velocity. Because the chance is 37% to find a star this close in velocity, we think it is very likely this is a chance match between groups and we reject the third pair. We do not take the pair, 5 and 6, any further into account because of the very deviating metallicity between the two members. This case is a good example of our decision to take the AF8 stars into consideration, but not the stars added with the FOF8 criterion. The third pair is even a very questionable pair in itself since the metallicities of the two stars are very different. We decide to regard this as a group of four, using the first two pairs only.

**Group 4.** The core members are stars 68 and 69, which are a pair at  $4dist$  0.04. The only giant which is within a  $4dist$  of 0.08 of both stars in the

pair is star 51. Because this giant is consistent within error bars with the pair in all four parameters and the additional metallicity component, we regard this group as a group of three stars: 68, 69 and 51. Although a lot of the stars added from FOF8 form groups and pairs on their own at the  $4dist = 0.08$  level, we disregard these because none of these are present in pairs at  $4dist = 0.04$ .

**Group 5.** Core members of this group are stars 36 and 38. Although star 37 is within  $4dist$  0.08 from both of these stars, we choose not to include it because its large discrepancy in distance. The reason it is picked up is because of its similarity in the other observables: angular distance and radial velocity.

**Group 6.** Although the two pairs from the FOF4 and FOF8 criterion are relatively close in distance and radial velocity information it can be seen from the additional metallicity information that they probably do not belong to the same structure after all. Although the additional pair, stars 53 and 58, is just picked up at  $4dist$  0.08 we take it into account as a pair. The two stars of this pair are quite far apart in angular distance, but they are remarkable close in distance, velocity and metallicity. The chance of finding another star this close in radial velocity at this distance (using a Gaussian distribution with the global velocity dispersion) is only 7%.

All final groups which are considered to be meaningful structures after individual examination as performed above of their members and possible additional members are given in Table 4.5. Their range in sky positions, velocities, distances and metallicities is given in table 4.6. In total, these final groups consist of 23 stars, which is 22.5% of the complete Spaghetti survey dataset.

Group	Members from Table A.1
1f	0, 2, 93, 96, 97, 98, 99
2f	18, 21, 24
3f	4, 7, 11, 12
4f	51, 68, 69
5f	36, 38
6f	59, 66
7f	53, 58

Table 4.5: Final group classification.

## 4.2 The Great Circle method

The 4distance method is suitable to look for substructures on small scales; predominantly clump like structures. On the other hand, we also expect to

Longitude (deg)	Latitude (deg)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km s <sup>-1</sup> )	Metallicity (dex)
[347.3, 5.3]	[48.6, 53.3]	[50.7, 58.7]	[13.0, 68.3]	[-2.3, -1.4]
[185.1, 186.0]	[-45.5, -45.2]	[23.6, 25.8]	[-63.9, -20.9]	[-2.7, -1.1]
[6.8, 18.3]	[40.1, 47.8]	[3.9, 8.9]	[-1.8, 42.9]	[-0.9, -1.0]
[281.4, 305.4]	[59.5, 61.3]	[9.4, 17.0]	[150.9, 176.5]	[-1.5, -1.1]
[263.4, 268.1]	[33.1, 35.8]	[27.4, 30.2]	[193.9, 203.6]	[-1.3, -1.1]
[292.8, 305.3]	[60.6, 61.4]	[13.0, 16.1]	[-136.4, -121.1]	[-1.4, -1.3]
[283.0, 289.8]	[48.5, 56.8]	[21.6, 24.3]	[-159.5, -143.9]	[-2.2, -2.1]

Table 4.6: Properties of the final groups from top to bottom 1f - 7f.

see a significant amount of large-scale stream-like structures particularly in the outer halo. To search for these streams, we adopt a method first discussed by Lynden-Bell and Lynden-Bell (1995) and further refined by Palma et al. (2002): The great circle method. The main assumption underlying this method is that each accreted satellite orbits in a plane containing both the current position of the satellite and the Galactic centre, whose intersection with the celestial sphere is a great circle on the sky. Further it is assumed that any set of objects from a common progenitor, no matter how far apart on the sky they are today, will still have the same orbital plane. Both assumptions are valid if the gravitational potential of the Galaxy is approximately spherical.

#### 4.2.1 Great Circle Family of orbital Poles

For all objects orbiting in the same orbital plane it is possible to define an associated ‘orbital pole direction’, which defines the direction of their angular momentum pole. The direction of this orbital pole will be perpendicular to the great circle described by the intersection of the orbital plane with the celestial sphere. For each individual star in our dataset we do not know its full orbital plane, but we do know this plane has to include the current position of the star. The direction of the orbital pole therefore has to be perpendicular to the vector drawn from the Galactic centre to the star’s current position. The family of possible orbital poles for this star forms another great circle on the sky which includes all perpendicular directions to the star’s current position. This circle is called the ‘Great Circle Pole Family’ (GCPF) and is illustrated in Figure 4.11.

For every two objects, the great circles tracing their possible orbital pole orientations, or GCPFs, will intersect twice (180° apart), as is illustrated in Figure 4.12. If these two objects would come from the same progenitor, this pair of intersection points would represent the orbital pole and orbital anti-pole of the orbit of the original satellite. An indication of *possible* linkage in dynamical history in our dataset would be to find several objects with

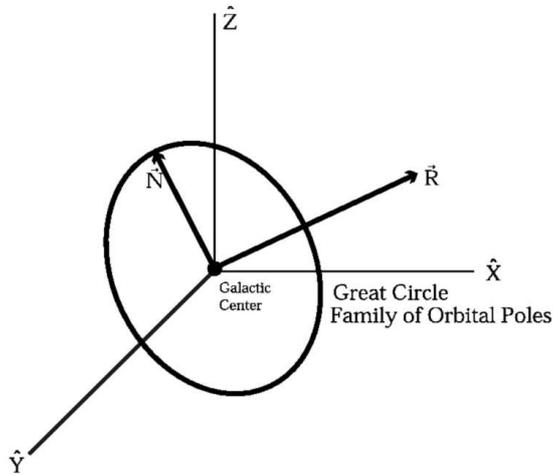


Figure 4.11: The geometry of the orbital pole family for an object that has a radius vector  $R$ . The vector  $N$  here is just one of the possible normal vector to the radius vector, all other possible endpoints of normal vectors are represented by the circle. This is the so-called ‘Great Circle Pole Family’ (GCPF) of  $R$ . Figure from Palma et al. (2002).

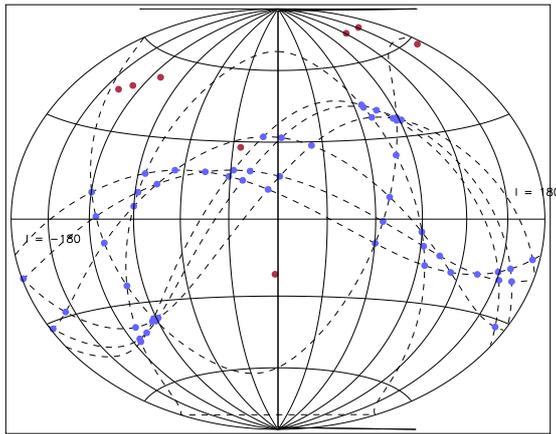


Figure 4.12: Example of the great circle method plotted for 8 giants. In an airtoff projection representing the whole sky, plotted here are the actual giants in a galactocentric perspective on the sky (red dots), their corresponding GCPFs perpendicular to the vector from the Galactic centre to the giant and the crossing points of the GCPFs for every two stars (blue dots).

similar crossing points in their circles of orbital pole families. However, the GCPF is just constrained by the current position of the object. Therefore there is little information to start from and determine the orbital plane. For every object the possible number of orbital planes is infinite and so chance alignments may not be uncommon. Even if more objects have a common crossing point for their GCPFs their true orbital poles could lie on any point on the circle for all of these objects and do not lie necessarily near the crossing point.

For several reasons we do not expect to find a perfect match in the orbital poles of tidal remnant objects from the same parent satellite (Palma et al., 2002):

- Even in a perfectly spherical potential, there will be a spread in energy in the debris orbits, because the finite size of the infalling system.

- If the host galaxy is not perfectly spherical, the satellite and its debris will not be on a single plane and differential precession will be induced. Especially inner halo debris out to 10-15 kpc from the Galactic Centre is not expected to stay in one plane (Helmi and White, 1999).
- Lumps in the halo and time evolution in the galactic potential, due to for instance accretion or disk growth could also induce changes in angular momentum directions.

In the outer halo, the region probed by Spaghetti project, these effects are likely to be less important.

#### 4.2.2 Energy and angular momentum check

A complementary way of deducing a possible membership is to use the stars' specific<sup>1</sup> energy and angular momentum as described in Lynden-Bell and Lynden-Bell (1995). The idea behind this is that stars from the same progenitor should have similar energies. This will hold provided (Lynden-Bell and Lynden-Bell, 1995):

- The objects were torn off at one or several close passages during a period in which the angular momentum and the energy of the progenitor's orbit did not change very much.
- The initial orbits of the objects torn off had approximately the same specific energy and specific angular momentum as the progenitor.
- The gravitational potential in the Galaxy has remained roughly constant since accretion time.

Because we do not possess full phase-space information for our stars, it is not possible to calculate the exact values of their specific energy and angular momentum. We do know, however, the radial velocity along the line of sight,  $v_l$  and this is for stars at large distances from the Galactic Centre (after correction for the motion of the Local Standard of Rest) approximately equal to the radial velocity as seen from the Galactic Centre,  $v_r$ . The specific energy of the star's motion is given by:

$$\frac{1}{2}v_r^2 + \frac{1}{2}h^2r^{-2} + \Psi = E \quad (4.5)$$

Unknowns in this equations are the specific angular momentum,  $h$ , the gravitational potential,  $\Psi$ , and the specific energy of the stars,  $E$ . Although we do not know the values for  $h$  and  $E$ , we may assume they are constant for debris from the same parent satellite. The distance from the Galactic centre,  $r$ , and a first approximation for  $v_r$  ( $v_l$ ) are measured for our stars and we

---

<sup>1</sup>per unit mass

may use prior knowledge on the gravitational potential,  $\Psi$ . We then rewrite this equation as follows:

$$E_r = \frac{1}{2}v_r^2 + \Psi = E - \frac{1}{2}h^2r^{-2} \quad (4.6)$$

For every star in our dataset we can calculate a first approximation for  $E_r$ . For the Galactic potential we use a Johnston, Spergel and Hernquist potential, consisting of three components: a disk, spherical and logarithmic potential. Because  $E$  and  $h$  are constants, we expect to see a linear dependence in a  $E_r$  versus  $r^{-2}$  plot for all the stars that are originating from a single parent satellite. The  $E_r$  versus  $r^{-2}$  diagram for all the stars in the dataset is given in Figure 4.13. The Galactic potential for 10.000 stars from a smooth halo simulation is over-plotted as a blue solid line. The scatter in potential for stars at a similar distance originates from the contribution of the disk potential component, which introduces a dependence on height above the plane as well. This scatter is stronger at smaller distances. From Figure 4.13 it becomes clear that the potential of the galaxy is dominating the trend; most of the stars follow neatly the shape of the overall potential. This means that the potential of the Galaxy is much stronger than the kinetic energy associated to the radial velocity component of the stars. Although by eye it appears possible to fit straight lines through many points, this is therefore not necessarily a sign of substructure.

To quantify the amount of substructure in the dataset, we plot a random set to compare the structures in both. The random set is constructed in the same manner as in Section 4.1.1, by re-shuffling distance and velocity information for the stars in the Spaghetti dataset. One  $E_r$  versus  $r^{-2}$  diagram for a random set is shown in Figure 4.14. The difference between Figures 4.14 and 4.13 is hard to see. This result casts doubts on the suitability of this method when applied blindly to a given dataset. Another concern are the extensive error bars in the dataset. For the closest stars (righthandside of the diagrams) the error bars are so large, that essentially any straight line can be fitted through. As discussed before, we do not expect inner debris at a radius smaller than 10 kpc from the Galactic Centre to have constant  $h$ . Therefore we decide to not consider any further all stars in the dataset closer than 10 kpc from the Galactic Centre.

As described by Lynden-Bell and Lynden-Bell (Lynden-Bell and Lynden-Bell, 1995), it is possible to improve the first approximation of a straight line in the  $E_r$  versus  $r^{-2}$  diagram once a possible association between objects is found. Because the method has provided a first order estimate for both the angular momentum magnitude and direction for the association of groups, the real value of  $v_r$  (before taken as equal to  $v_l$  while in reality the angles are different, certainly for objects close to the Galactic Centre) can be better approximated. Iteration on  $v_r$  and the angular momentum of the groups, which will influence eachother, will then eventually return the proper value.

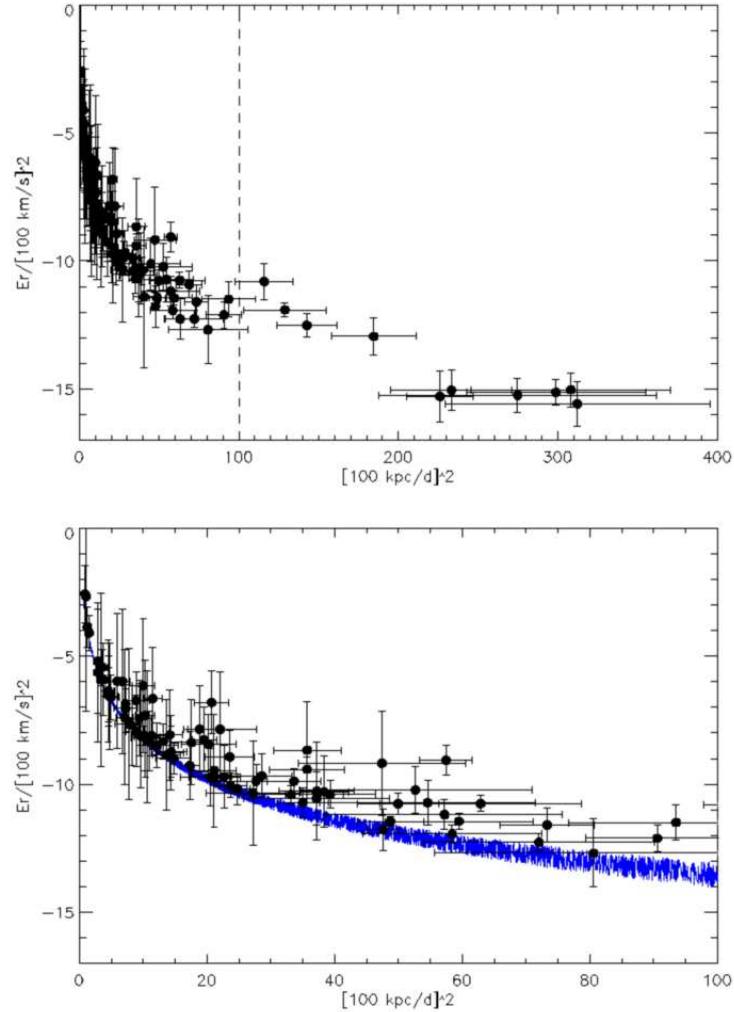


Figure 4.13:  $E_r$  versus  $r^{-2}$  diagram for all stars in the dataset including error bars on distance and in the calculation of the gravitational potential. The top panel shows all stars in the dataset. Towards the right of the vertical line stars located at a distance smaller than 10 kpc from the Galactic Centre are found. The bottom panel only shows the stars in the dataset further than 10 kpc from the Galactic Centre. The solid blue line represents the contribution of the Galactic potential for 10.000 stars from a smooth halo simulation.

However, we will not go into this analysis for our data, because our groups possess relatively large distances ( $> 10$  kpc) at which the effect will be negligible in comparison to the observational error bars which we expect to have a much stronger influence the slope of the straight line.

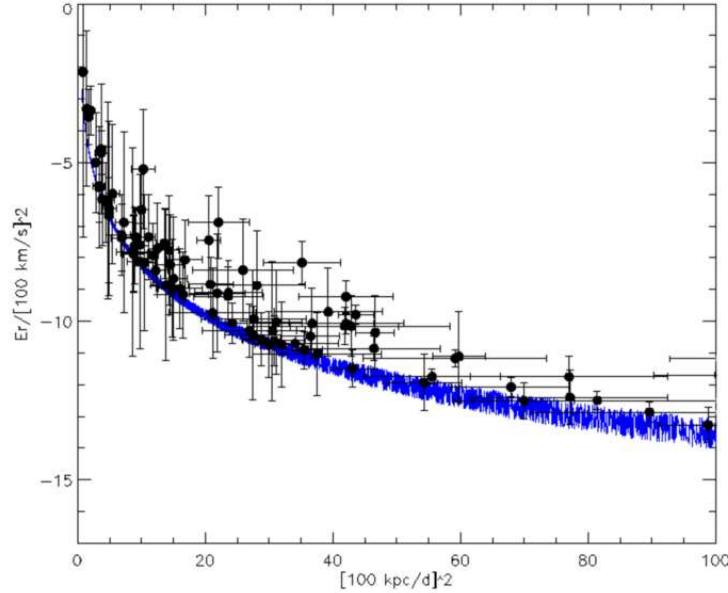


Figure 4.14:  $E_r$  versus  $r^{-2}$  diagram for all stars in a random set including error bars on distance and in the calculation of the gravitational potential for all stars more distant than 10 kpc from the Galactic Centre. The solid blue line represents the contribution of the Galactic potential for 10.000 stars from a smooth halo simulation.

### 4.2.3 Combination of both requirements: The full great circle method

Lynden-Bell and Lynden-Bell applied the above method by hand on a sample of 22 globular cluster objects. Clearly, this is not feasible for our dataset, nor for forthcoming ones. Therefore we wish to automatize this method to search for groups that have possible similar orbital poles and lie on a straight line in the  $E_r$  versus  $r^{-2}$  diagram.

For two adjacent stars on the sky, their GCPF's will be very close to each other. This will automatically mean that if they cross a third GCPF, or a pair of those, a common angular pole (crossing point of more than two GCPF's) will easily be found although it will have no physical meaning. After exploring the data set and various simulations (discussed in Section 5.2) for crossing points for the GCPF's of the giants, we decide that probably the best way to avoid finding a lot spurious of 'hits' due to the clumpy sky distribution of the data set or the inaccuracy of the  $E_r$  versus  $r^{-2}$  diagram (shown in Figures 3.9 and 4.13 respectively) is to start from the already found substructures using the 4distance method. The great circle method is complementary to the 4distance method in order to determine whether any

of our found substructures could, on the basis of sky position, energy and angular momentum considerations, be linked to one of the other structures found on other parts of the sky. Compared to the requirement of a straight line in the  $E_r$  versus  $r^{-2}$ , the direction of the possible orbital pole on the sky is the more restrictive requirement. Because of the extensive error bars and the general trend of the overall potential a straight line can easily be fit to a group of stars in the  $E_r$  versus  $r^{-2}$  diagram. We therefore choose to start by defining an orbital plane i.e. a crossing point for the GCPFs of three groups or more. A straight line is subsequently fit to the groups only if their GCPFs do meet within the required angular distance.

Every group of giants picked up by the 4distance method which is on average farther out than 10 kpc from the Galactic Centre (shown in Table 4.5), is averaged over its position and treated as one object in the search for the angular momentum pole. If two groups are within  $10^\circ$  from each other in sky position they are also taken together in the determination of their GCPF. For the dataset all the groups and their GCPFs are plotted in Figure 4.15. A crossing of three or more orbital pole family circles means the crossing point might be the orientation of their mutual angular momentum pole. If such a point is found, a subsequent check is performed on the position of the individual stars in the the  $E_r$  versus  $r^{-2}$  diagram. If indeed all the stars in these three or more substructures can be modelled by a straight line, they might be debris from the same parent satellite. For all stars considered in the discovered substructures we start off with, the  $E_r$  versus  $r^{-2}$  diagram is shown in Figure 4.16.

If a group is added to another pair of groups, because their GCPF crossing-points are close enough, a least square method is used to define the likelihood that their corresponding members in the  $E_r$  versus  $r^{-2}$  diagram can be fitted by a straight line. First, the best fitting linear equation is determined from the points using a least square fit, by minimising the  $\chi^2$  function with respect to the slope,  $b$ , and the cross-point with the y-axis,  $a$ , of the fitted straight line.

$$\chi^2(a, b) = \sum_{i=1}^N \left( \frac{y_i - a - bx_i}{\sigma_i} \right)^2 \quad (4.7)$$

In doing this, errors in both x and y have to be taken into account. For this purpose we use the *fitxy* routine as described in *Numerical Recipes in C* (Press, 1988). Subsequently the goodness-of-fit of this data to the model (the straight line fitted) is calculated using incomplete gamma function  $Q(a, x)$ . The probability  $Q$  that a value of chi-square is as *poor* as the  $\chi^2$  value calculated is given by:

$$Q(k, x) = \frac{1}{\Gamma(k)} \int_0^x e^{-t} t^{k-1} dt \quad (k > 0), \quad \text{with} \quad (4.8)$$

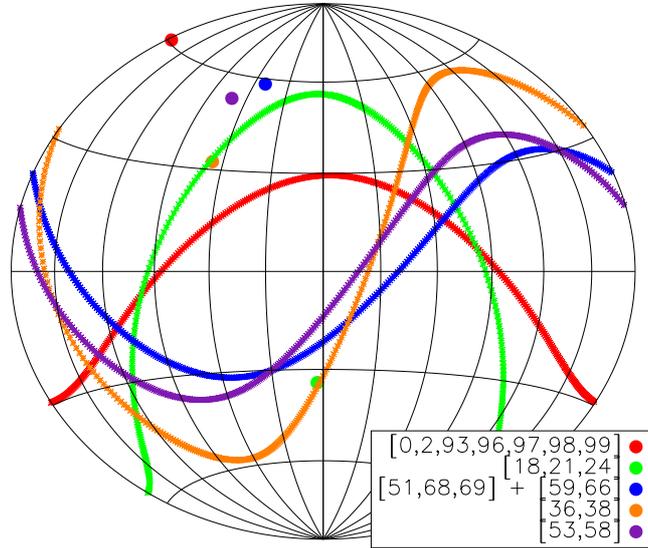


Figure 4.15: Groups in Table 4.5 are represented by one galactocentric position on the sky and one GCPF. The label gives the exact numbers of the stars as they can be found in Table A.1. Although stars 51,68 and 69 and stars 59 and 66 possess very different velocities they are here represented as one group, because their closeness in sky position would link the two groups to any other group.

$$a = \frac{N - 2}{2}, \quad x = \frac{\chi^2}{2} \quad (4.9)$$

If  $Q$  is larger than 0.1, the goodness-of-fit is believable (Press, 1988). We choose to consider for straight line fits with at least  $Q = 0.99$  however, because generally our  $\chi^2$  are small, due to the large internal (observational) errors. We also require the errors in both  $a$  and  $b$  to be smaller than 10% to avoid fitting “clumps” in the diagram too easily.

We require the crossing-point on the sky between multiple groups to be smaller than  $10^\circ$ . This requirement is somewhat arbitrary, but is chosen such that it will include the tidal radius of smaller objects, like Fornax which has a tidal radius of Fornax is less than  $1^\circ$  (Lynden-Bell and Lynden-Bell, 1995), and larger objects that are thought to be from a common origin. The separation of the Small and Large Magellanic Stream, for instance, is  $15^\circ$

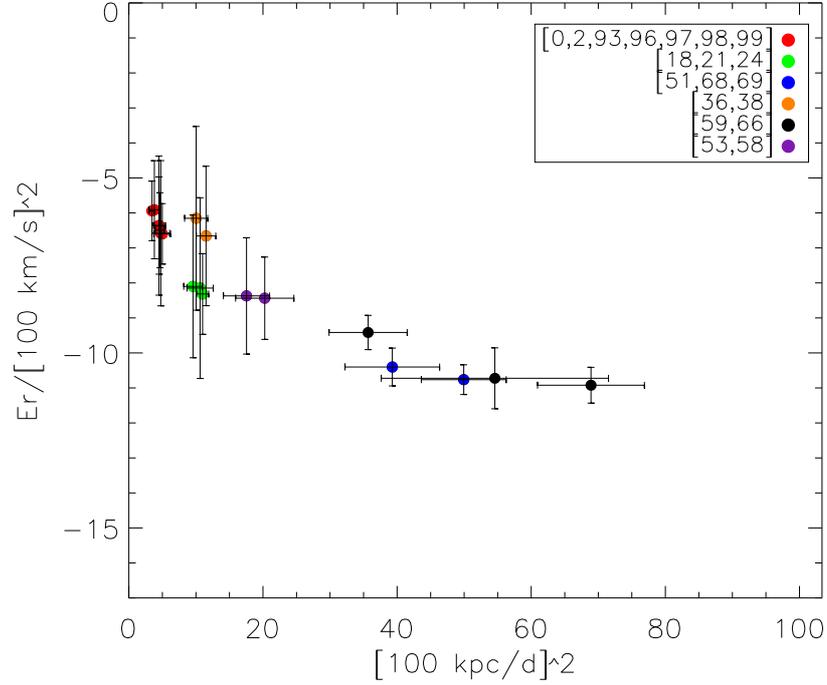


Figure 4.16: The  $E_r$  versus  $r^{-2}$  diagram for all stars in the different groups in Table 4.5.

Same origin?		
Matching Group 1	Matching Group 2	Matching Group 3
0, 2, 93, 96, 97, 98, 99	18, 21, 24	59, 66
0, 2, 93, 96, 97, 98, 99	36, 38	53, 58

Table 4.7: Already found groups with the 4distance method that are *possibly* dynamically linked.

along the line of the stream. However, this structure is complicated due to the binary nature of the Clouds and a single object should produce a narrower stream (Lynden-Bell and Lynden-Bell, 1995).

Considering all cross-points closer than  $10^\circ$  to be meaningful, we find two different possibilities for common origin for our groups. The possible associations of groups are given in Table 4.7. Figure 4.17 shows the crossing point of possible angular momentum poles and the  $E_r$  versus  $r^{-2}$  diagram for all the stars in the associated groups.

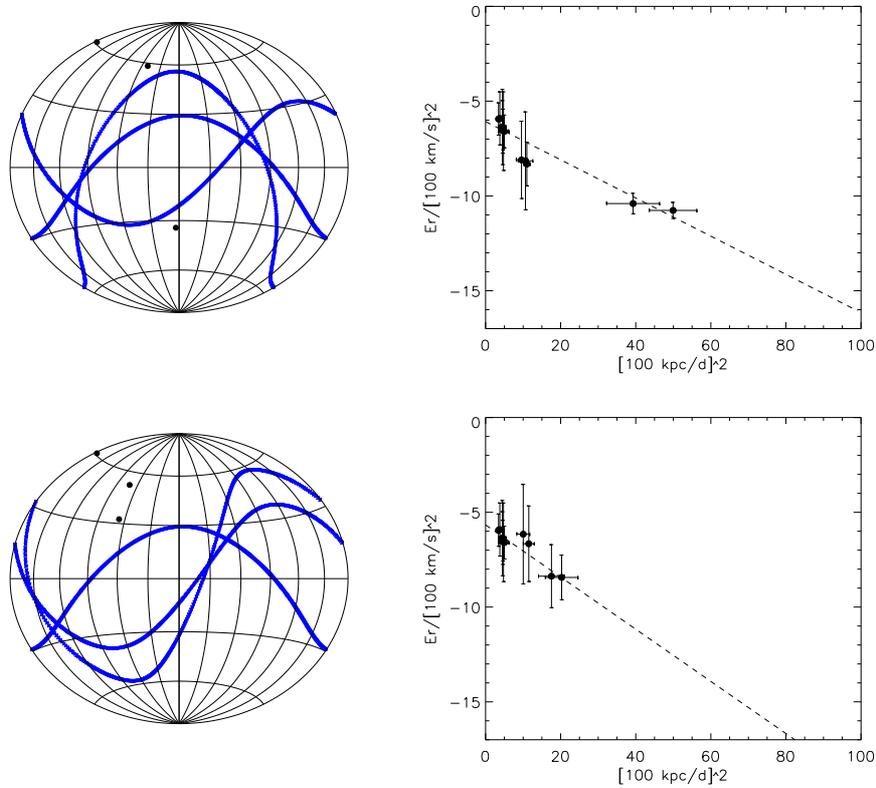


Figure 4.17: The plots on the left show the crossing point of possible angular momentum poles for two times three groups that were already found to be substructures from the 4distance method while the plots on the right show the position in the  $E_r$  versus  $r^{-2}$  diagram for all the stars in these groups.

For both associations the large group 1f with seven members is part of the linked substructure. Since all of the groups together do not fit well enough on one straight line in the  $E_r$  and  $r^{-2}$  diagram and their GCMs do not all cross within  $10^\circ$ , we do not expect both possibilities to be valid. For a further discussion of the possible link between these several groups we refer the reader to Section 5.1.1. In Section 5.2, we compare these results to several simulations.



## Chapter 5

# Interpretation of the Data

### 5.1 Are these substructures new?

Using the 4distance method, we defined 7 groups of stars which are likely to be real substructures in the Milky Way halo. A subsequent question to answer is of course whether these structures can be related to any structure previously discovered. Figure 5.1 shows the found substructure over-plotted with the regions of the sky which are most extensively researched for substructure; the area of the Sloan Digital Sky Survey (SDSS) covered so far (data release 5). From this figure it can be seen that the SDSS covers a large number of the stars in our data set, but not all of them. The structure that does show overlap can subsequently be matched with the substructure already observed in for example the Field of Streams.

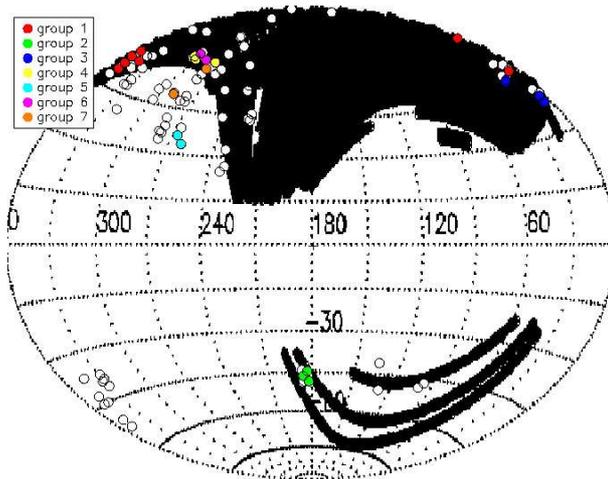


Figure 5.1: The found substructures on the sky compared with the area searched by SDSS in data release 5.

### 5.1.1 The Sagittarius Dwarf Galaxy

For the Spaghetti collaboration Dohm-Palmer et al. (2001), mentioned already that there was a detected concentration of giant stars which stood well above the expectations of a smooth halo model. These Spaghetti giants could be matched to a simulation model of the debris from the disrupting Sagittarius spheroidal dwarf galaxy. By then, only four stars possessed radial velocity information, which were matching well with the predicted velocity of the debris stars in the models.

In our examination of the full dataset from the Spaghetti Survey we do find this same over-density to be very prominent substructure of a striking 7 members, namely group 1f. This group was standing out remarkably from the comparison with the random sets as shown before in Figure 4.8. The substructure, at  $l = -12.7^\circ$  until  $l = 5.3^\circ$ ,  $b = 48.6 - 53.3^\circ$  and distances between 50.7 - 58.8 kpc, is in excellent agreement with the debris predictions of models as can be seen from Figures 5.2, 5.3 and 5.4. These figures show models for the Sagittarius dwarf galaxy within different galactic halos from Helmi (2004) over-plotted with the Spaghetti giants from this substructure. The Sagittarius dwarf and its debris are modelled using a fixed potential for the galaxy a bulge, disk and halo component. For the stellar distribution a King profile is chosen and self-gravity is modelled. The orbit of Sagittarius is generated constrained by observations. A chosen orbit is first integrated 10 Gyr backwards, after which the parameters are set up so that the present day model represents the main body and the debris observed. For the spherical, oblate and prolate potentials, orbits are chosen which have approximately the same apocentric and pericentric distance as well as  $L_z$  and it is required that after 10 Gyr the satellite has reached a similar degree of disruption, typically 10 to 20%. (Helmi, 2004)

Our method did not pick up the additional proposed structures at distances of 20 and 80 kpc in Dohm-Palmer et al. (2001) as significant.

The metallicity of the giants in group 1f matched to the Sagittarius model is  $[Fe/H] \approx -1.75$  which is about 0.7 dex lower than the mean for stars in the main body of the Sagittarius dwarf galaxy. Because outer stars of a galaxy are thought to be stripped off first, the streams could have a different stellar population than the remnant main body. The metallicity difference between the two parts would then reflect a metallicity gradient in the dwarf galaxy itself. We know from observations that dwarf spheroidal galaxies often possess metallicity gradients (e.g. Tolstoy et al., 2004) and also a population gradient has already been detected in the Sagittarius stream (Bellazzini et al., 2006).

The spread in metallicity in the group itself is even more likely to be real once we take into account multiple wraps of the satellite. From Figure 5.4 it can clearly be seen that the group is coinciding with an overlap between different wraps in all three panels. If the stars in the groups would originate

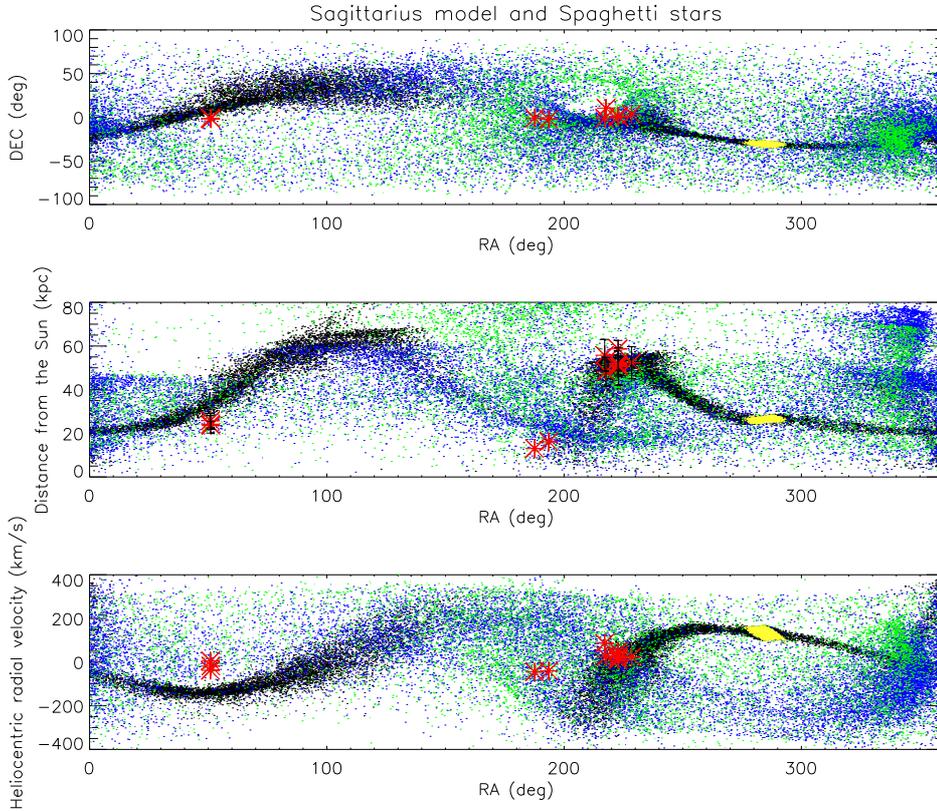


Figure 5.2: Model of the stream of debris left behind by the Sagittarius dwarf spheroidal galaxy while it is disrupted by the Milky Way in an oblate halo ( $q=0.80$ ). Plotted here are right ascension against declination, heliocentric distance and heliocentric radial velocity. Black dots represent particles that became unbound less than 3 Gyr ago, blue and green particles became unbound between 3 - 6 and 6 - 9 Gyr ago respectively. The core of the galaxy with still bound particles is plotted in yellow dots. Red asterisks denote the Spaghetti giants from group 2f and 6f and 1f at  $RA = 50, 190$  and  $220^\circ$  respectively.

from different wraps and thus became unbound at different times and if the satellite would possess a metallicity gradient, we would indeed expect to see a quite significant metallicity difference between them.

### Substructures from the great circle method

By applying the great circle method to the found groups from the 4distance method, two possible dynamically linked associations of each three groups were found. Both of these associations contained group 1f, which can very likely be associated with the disrupting Sagittarius dwarf galaxy as discussed above. Provided this group is indeed debris from the Sagittarius dwarf

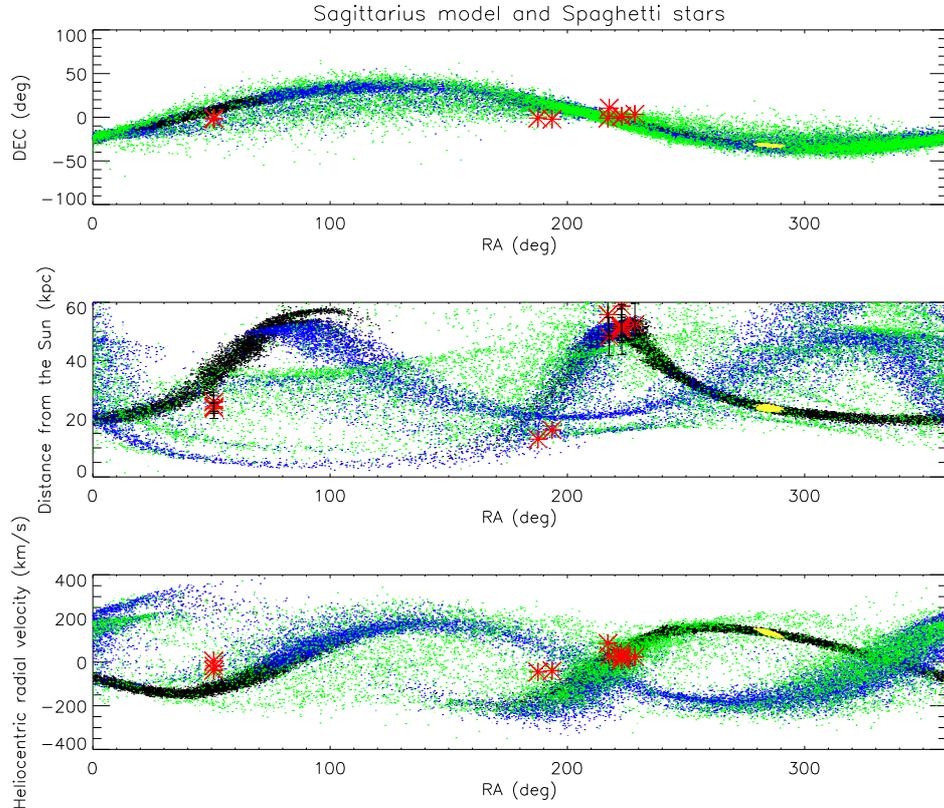


Figure 5.3: Same figure as Figure 5.2 for a spherical dark matter halo ( $q=1.00$ ).

galaxy it follows that if these groups would indeed be dynamically associated, they would all have to be debris from this satellite.

Figures 5.2, 5.3 and 5.4 show the first association of three groups, group 1f, 2f and 6f, over-plotted on the models. From these figures it follows that the three groups could indeed all three be associated with Sagittarius debris, provided that the debris would have been stripped off at different times. Both groups 2f and 6f have to have been stripped off relatively early, at least 3 - 6 Gyr ago. Group 2f consists of three stars. While it was not clear before if the third member of the group could be an actual member because of its deviating metallicity, its membership becomes more likely as debris from a large object such as Sagittarius is expected to show a large metallicity spread. The metallicities of the stars in group 6f are  $[-1.4, -1.3]$  dex which is even a little higher than the metallicities for the other groups which might be associated to Sagittarius. We would expect it to be the other way around since the most metal-poor material is stripped off first. The metallicities are still lower than the metallicities of the core of the dwarf

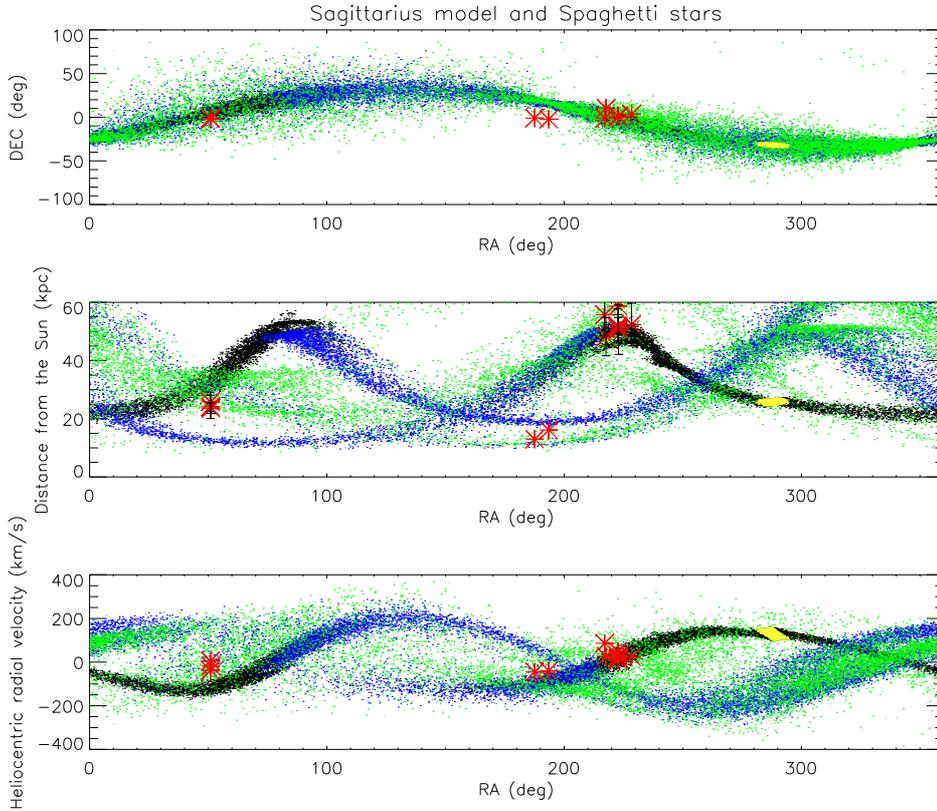


Figure 5.4: Same as Figure 5.2 and 5.3 for a prolate halo of  $q=1.25$ .

galaxy however.

The association of group 6f with the debris becomes more likely as a more oblate halo potential is used. In a prolate potential the debris is much more confined on the sky. The most prolate halo model shown here ( $q=1.25$ ) does not match group 6f in sky position at all. On the other hand, a higher density Sagittarius debris is matching group 2f in a more prolate halo, especially when considering the velocities.

The second association of groups found with the great circle method is over-plotted on the model of Sagittarius debris in Figure 5.5. The two additional groups associated with group 1f, group 5f and 7f, can only be matched with debris from the stream in a very oblate halo model. Again also, this material would have to have been stripped off early. Even then, the radial velocities of group 5f are too high match with the model. Group 7f could be considered to be early Sagittarius debris if the Galactic halo would be oblate.

None of the other groups found with the 4distance method matches well with the Sagittarius models.

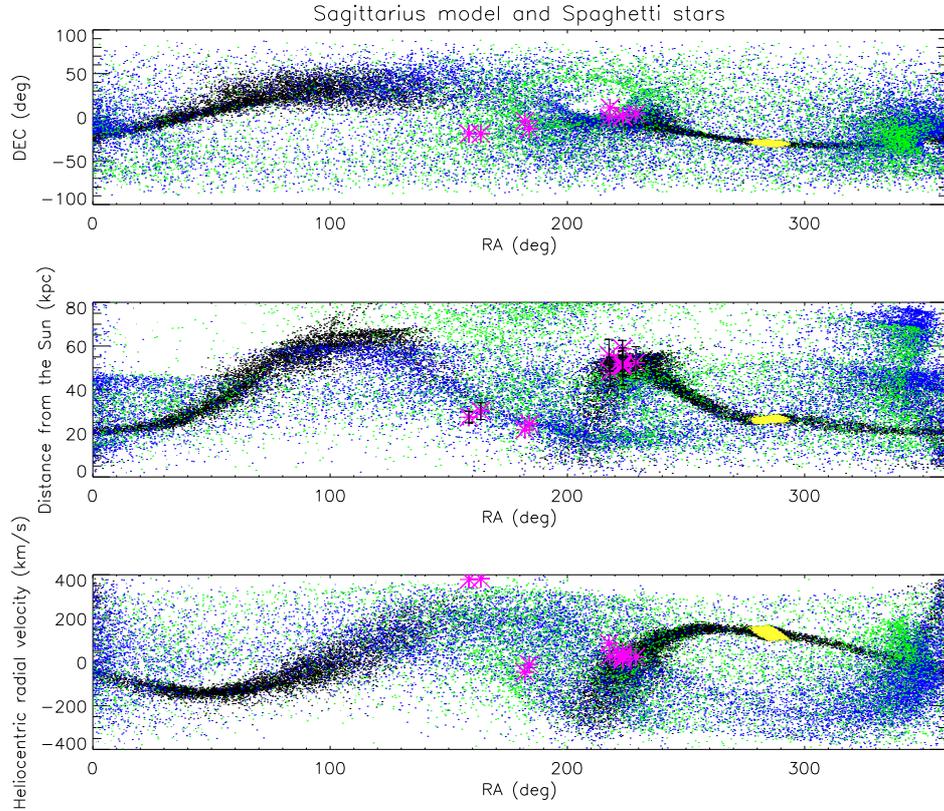


Figure 5.5: The Sagittarius model as plotted in Figure 5.2 using a oblate dark matter halo ( $q=0.80$ ). The over-plotted purple asterisks at  $RA = 160, 185$  and  $220^\circ$  are the groups 5f, 7f and 1f from Table 4.5 respectively.

### 5.1.2 The Virgo Substructures

Three over densities are known today towards the constellation of Virgo. The QUEST collaboration identify an excess of 23 RR Lyrae variables with apparent magnitudes of  $16.5 < V_0 < 17.5$  and within galactic coordinates  $279^\circ < l < 317^\circ$ ,  $60^\circ < b < 63^\circ$ . Spectra of the RR Lyrae and BHB stars towards  $(l, b) = (288^\circ, 62^\circ)$  revealed a moving group with Galactic standard of rest velocity along the line of sight of  $V_{rGal} = 99.8 \pm 10 \text{ km s}^{-1}$  and a distance of approximately 19 kpc from the Sun (Duffau et al., 2006). This group is called the Virgo Stellar Stream (VSS).

Independently an overdensity of stars near  $(l, b) = (297^\circ, 63^\circ)$  was confirmed by Newberg et al. (2002) using blue turn-off stars from the Sloan Digital Sky Survey (SDSS). This overdensity is known as S297+63-20.5. The estimated distance of this structure is 18 kpc from the Sun, but the broadness of the main sequence indicates that this structure is likely dis-

persed in distance. The faint turn-off stars near S297+63-20.5 measured in the Sloan Extension for Galactic Understanding and Exploration (SEGUE) project in fields  $(l, b) = (300^\circ, 55^\circ)$  and  $(l, b) = (288^\circ, 62^\circ)$  have a dominant radial line-of-sight velocity corrected for the Galactic centre of rest of  $130 \text{ km s}^{-1}$  and two additional moving groups are picked up near S297+63-20.5 with velocities  $V_{rGal} = -168 \pm 10 \text{ km s}^{-1}$  and  $V_{rGal} = -76 \pm 10 \text{ km s}^{-1}$  (Newberg et al., 2007).

Subsequently Jurić et al. identify a large overdensity (the Virgo Overdensity, VOD) covering  $\sim 1000 \text{ deg}^2$  towards direction  $(l, b) = (300^\circ, 65^\circ)$ . This stellar overdensity shows a remarkably extended vertical structure perpendicular to the Galactic plane ( $\Delta Z \approx 8 \text{ kpc}$  with heliocentric distances  $5\text{--}17 \text{ kpc}$ ). The density of VOD main-sequence stars peaks at  $\sim 16 \text{ kpc}$  from the Sun (Martinez-Delgado et al., 2007). From star counts in the Hess diagram of colour vs. magnitude for the Virgo overdensity field and the control field, Jurić et al. suggest that the Virgo overdensity stars have metallicities lower than (thick) disk stars and similar to those of halo stars.

The origin of these features and whether they are all part of the same large substructure is still unknown. Jurić et al. suggest that the VOD might be related with a yet unknown merger of a low metallicity dwarf galaxy system. Martinez-Delgado et al. (2007) investigate the possible association between the VOD and the leading tidal tail of the Sagittarius dwarf galaxy. They predict large negative radial velocities for this overdensity if the connection with the Sagittarius leading tidal tail would be true. However, Newberg et al. (2007) show that it is unlikely that S297+63-20.5 is part of the leading tidal tail, because the turn-off stars are much brighter than those in the leading tail debris in the same direction of the sky, and the S297+63-20.5 have highly positive radial velocities which is contrary to the expectations if the overdensity would be related to the Sagittarius leading tail. The group of Newberg et al. expect that although the VSS, VOD and S297+63-20.5 have slightly different measured positions and radial velocities, there might be a relationship between the three substructures. In particular they expect that as more data is analysed, the overdensity S297+63-20.5 will become known as the Virgo Stellar Stream (VSS) (Newberg et al., 2007).

### Connecting the Spaghetti dataset and the Virgo overdensities

We have found three substructures in the Spaghetti dataset that are near the constellation of Virgo and have distances that agree with the distance ranges measured for the Virgo overdensities. The properties of these three groups, group 4f, 6f and 7f, are given in Table 5.1.

While none of these groups matches exactly the found overdensities they all have similar sky positions and distances. The radial velocities of group 4f are  $2\text{--}4.5\sigma$  off the radial velocity measured for S297+63-20.5. This offset is of the same order as the difference in radial velocity measured between

Group	Longitude (deg)	Latitude (deg)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km s <sup>-1</sup> )
4f	281.4 - 305.4	59.5 - 61.3	9.4 - 17.0	150.9 - 176.5
6f	292.8 - 305.3	60.6 - 61.4	13.0 - 16.1	-136.4 - -121.1
7f	283.0 - 289.8	48.5 - 56.8	21.6 - 24.3	-159.5 - -143.9

Table 5.1: Properties of the groups in the direction of the constellation Virgo.

VSS and S297+63-20.5.

Although group 7f has a large distance compared to the cores of the discussed Virgo substructures, its radial velocity measurements agree well with one of the moving groups picked up by Newberg et al. who also detect high negative radial velocities around  $-168 \text{ km s}^{-1}$ .

Fields 4f, 6f and 7f seem too low in latitude to be members of the leading tail stream of Sagittarius, as can be seen from Figure 5.1.2 which shows the leading tail and S297+63-20.5 from the SDSS photometry of F stars with distances 14.5 - 23 kpc from the Sun.

With the current evidence we can present no conclusive statement on whether groups 4f, 6f and 7f do indeed belong to any Virgo substructure. We expect in the future (some) of the groups picked up by Spaghetti will be linked to these larger Virgo substructures. To do so, or prove them to be separate, more radial velocity and accurate distance information is needed however.

### 5.1.3 The Orphan Stream

The Orphan Stream is one of the most intriguing streams we know today. It is very well confined, but no progenitor has yet been found. Although Belokurov et al. (2007) and Fellhauer et al. (2007) have attempted to link the stream to the newly discovered dwarf spheroidal galaxy Uma Major II as well as two globular clusters, Ruprecht 106 and Palomar 1 and possibly the linear association of high-velocity clouds, Complex A, there is no conclusive evidence that these objects are dynamically related.

#### An orbit for the Orphan Stream

Using the end-points of the stream whose radial velocities and distances are given in Fellhauer et al. (2007), see table 5.2, we have found a possible orbit of the Orphan stream. This orbit integrated in a galactic potential including disk, bulge and halo is plotted in Figure 5.7. The orbit is integrated for 2 Gyr both forwards and backwards. The x versus y and x versus z projections of this orbit are shown in Figure 5.9.

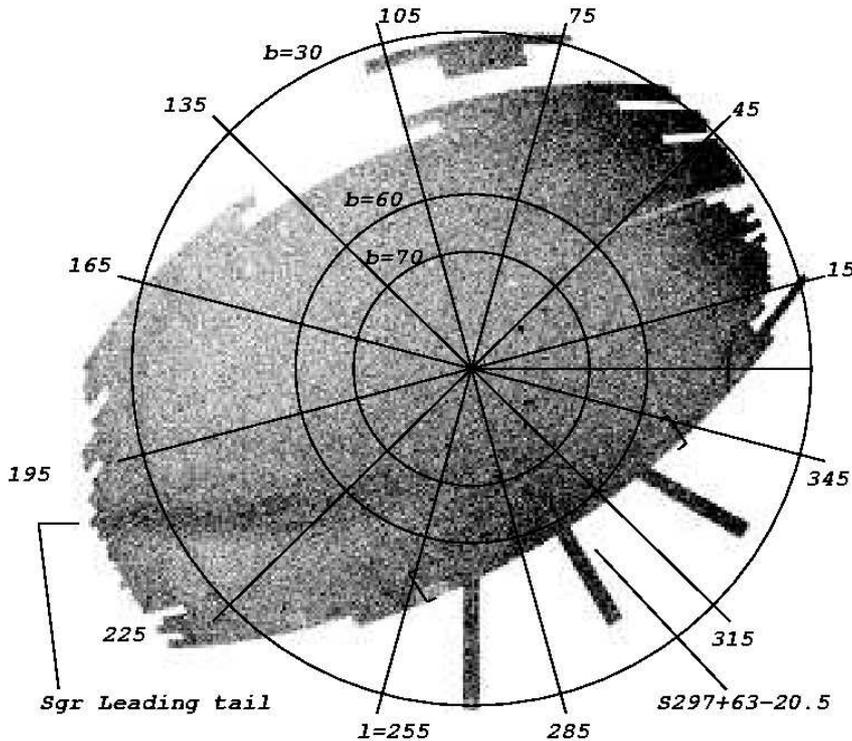


Figure 5.6: Structure in the Milky Way at distances  $\sim 14.5 - 23$  kpc from the Sun. A larger density of F stars is represented by a darker shading in the equal-area plot, centred on the North Galactic cap. The Sagittarius leading arm is clearly visible running from  $(l, b) = (205^\circ, 25^\circ)$  to  $(305^\circ, 65^\circ)$ . The prominent S297+63-20.5 overdensity is centred around  $l = 300^\circ$ . Figure from Newberg et al. (2007).

Right Ascension (deg)	Declination (deg)	Distance <sub>Sun</sub> (kpc)	V <sub>Sun</sub> (km/s)
162.1	-0.5	$20^{+7}_{-5}$	$-35 \pm 10$
149.4	32.0	$32^{+15}_{-12}$	$+105 \pm 10$

Table 5.2: The properties of the endpoints of the stream from Fellhauer et al.(2007).

Of all the stars in the Spaghetti dataset we have found 7 candidates that match the integrated orbit both in spatial, distance and velocity coordinates. These candidates are over-plotted as coloured dots in Figure 5.7. The properties of the stars which agree with the model are given in Table 5.3.

Especially star 31 is has a high chance of being a real member of the Orphan Stream, because its sky position is precisely where the observed part

of stream is going through.

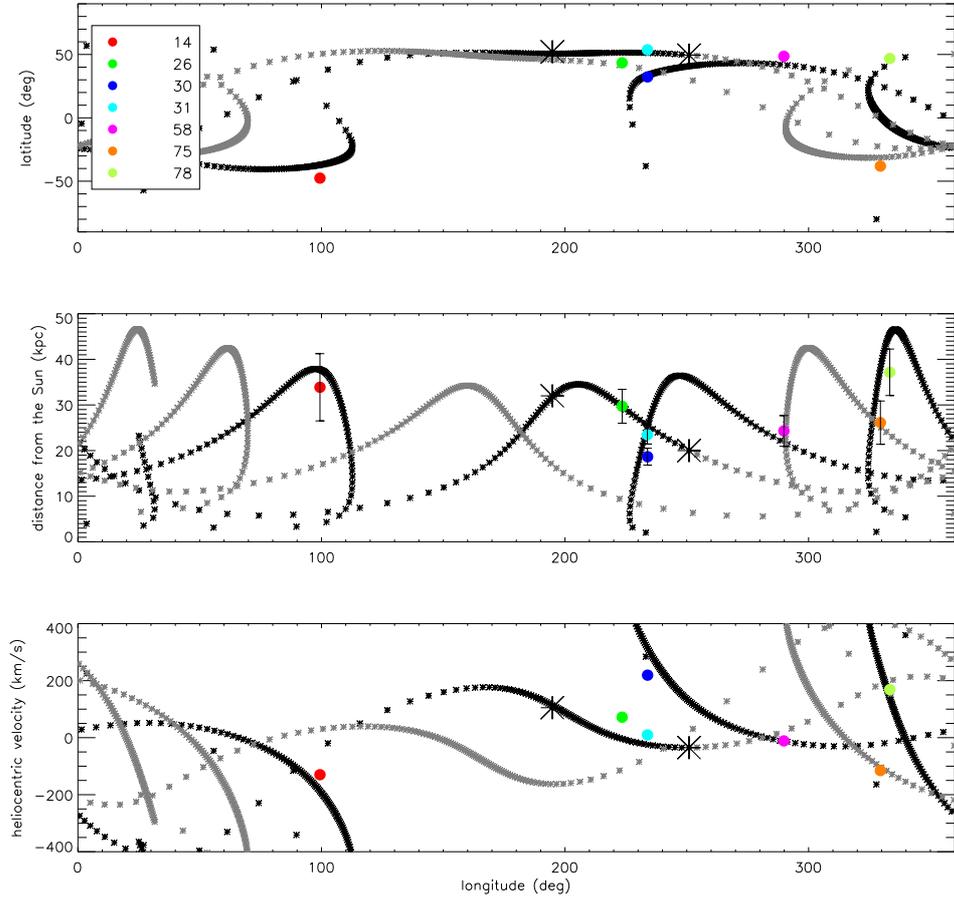


Figure 5.7: The candidates (coloured dots) plotted over the modelled orbit of the Orphan Stream integrated for 2 Gyr backwards (black) and forwards (grey). The bigger asterisks represent the endpoints of the stream from Fellhauer et al.(2007)

### Radial velocities of the Orphan Stream

The sky position of the stream and the distances of several observed parts of the stream are well constrained by the Sloan data, but the radial velocities are very uncertain. The published values for the endpoints of stream as given in table 5.2 are obtained by selecting all stars with the right magnitude to be red giants at the distance of the stream. But basically these stars could also be much nearer to us. Also the velocity histograms of these stars compared

Number	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	V <sub>Sun</sub> (km/s)
14	99.3799	-47.5892	-1.60	33.88	-129.9
26	223.446	43.4386	-1.38	29.71	71.8
30	233.896	32.3214	-1.35	18.65	219.2
31	233.832	53.7094	-1.24	23.46	9.6
58	289.842	48.5477	-2.09	24.30	-11.3
75	329.487	-38.0880	-2.04	26.12	-115.4
78	333.341	46.5056	-1.72	37.16	168.7

Table 5.3: The properties of the stars that match with the modelled orphan stream.

to off-stream field stars, shown in figure 5.8, are not very convincing as no more than 4 stars are counted per bin. As Belokurov et al. put it: “these detections are suggestive rather than conclusive”.

Provided star 31 would be a member of the Orphan Stream it would give a better radial velocity constraint. Because this star is confirmed to be a red giant it has a robust distance measurement (within error bars) whereas the SDSS stars might turn out to be foreground dwarfs.

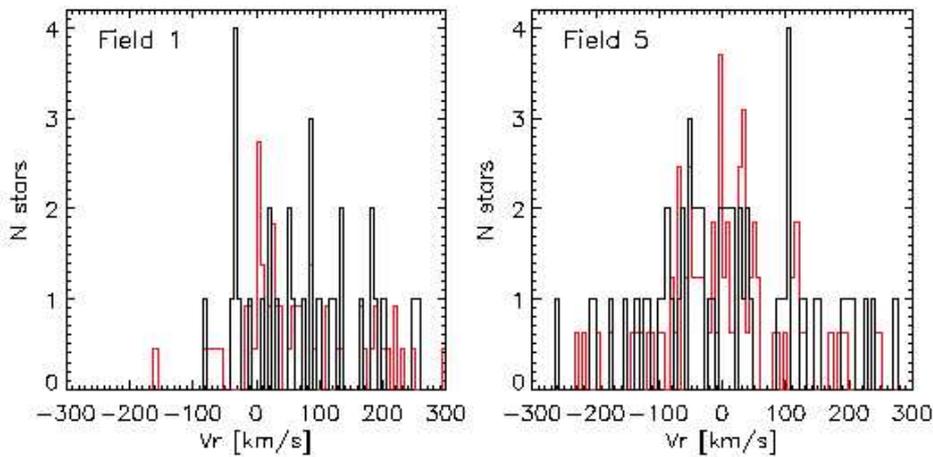


Figure 5.8: Velocity histograms of stars selected with M92 and M13 ridgeline masks at distance moduli satisfying  $16.5 \leq m - M \leq 17.5$  in field 1 and  $17.5 \leq m - M \leq 18.5$  in field 5. Black represents on-stream and red off-stream field stars. Figure from Belokurov et al. (2007).

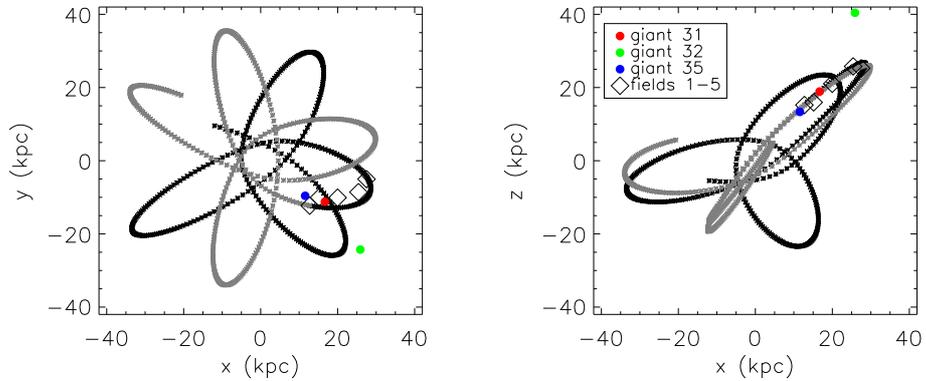


Figure 5.9: Modelled orbit of the Orphan Stream integrated for 2 Gyr backwards (black) and forwards (grey) in  $x$  versus  $y$  and  $x$  versus  $z$ . Three Spaghetti stars which are right on the stream in sky position are over-plotted in red, blue and green circles. The diamonds represent the fields on the stream which were studied by Belokurov et al. (2007).

### Without radial velocity constraint

If we do not use the constraint that the radial velocities of stars have to match the Orphan Stream model which is based on the radial velocities of the stream's endpoints as published by Belokurov et al. (2007), we find two more stars in the Spaghetti dataset that are near or right on top of the stream. One of these additional matches, star 35, does agree also in distance with the observed fields of the Orphan Stream by SDSS. The other extra candidate, star 32, does not match in distance with these fields, as can clearly be seen from the second panel of Figure 5.10.

However, the radial velocity of star 35 is very different from the first candidate found, star 31, and differs also substantially from the radial velocities obtained by Belokurov et al. We now have two possible candidates for Orphan Stream members that cannot be both members if the stream is considered to be one single wrap. From the SDSS spectra, Paul Harding and Heather Morrison (private communication) find more stars that are likely to be giants in both their colour range and Mg line strength and that are less than  $3^\circ$  from the reported sky position of the stream. In Figure 5.10 these stars are plotted together with the observed fields by Belokurov et al., the model and the Spaghetti member candidates. Although we do not have distance information for the stars found by Harding and Morrison in SDSS, it is clear from the bottom panel that the issue of the radial velocities is not solved by adding these extra stars. The SDSS subsample has new giant candidates with radial velocities close to both that of star 31 and star 35.

We think star 31 has a higher chance of being an actual member of the stream, since it also agrees with the (suggestive) radial velocity determined by Belokurov et al. The evidence is not conclusive however.

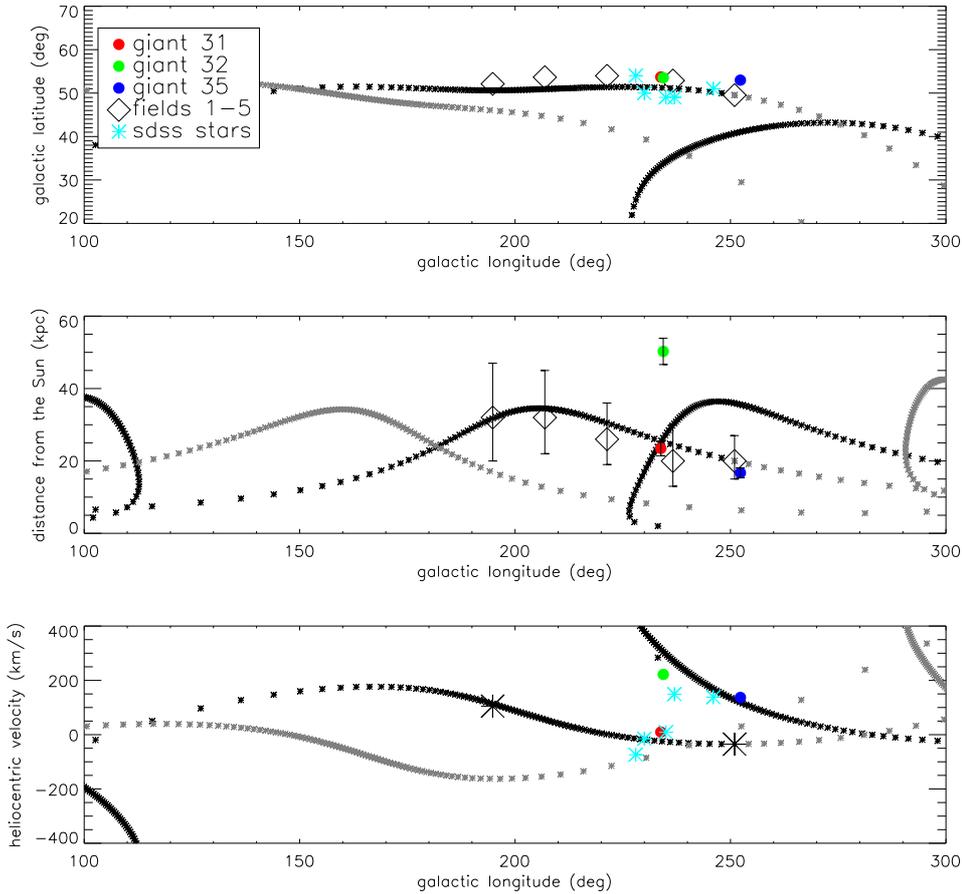


Figure 5.10: Modelled orbit of the Orphan Stream integrated for 2 Gyr backwards (black) and forwards (grey) in sky position (top panel), longitude versus distance from the Sun (middle panel) and heliocentric radial velocity (bottom panel). The three Spaghetti stars which are right on the stream in sky position, star 31, 32 and 35, are over-plotted in red, blue and green circles as well as the giant candidates in SDSS represented here by light blue asterisks. For these SDSS stars no distance information is available however. The black diamonds represent the fields on the stream which were studied by Belokurov et al. (2007).

A first step to better constrain the radial velocity of the Orphan Stream would be to determine distances (which involves proper dwarf/giant classification) for all candidate Orphan stream members in the spectroscopy data of SDSS. This will result in better constraints on the measured radial velocity

of the stream. A second objective of further research would be to investigate the flexibility of the modelling to several values of the radial velocity.

## 5.2 Constructing simulated datasets

We would like to constrain what fraction of the stellar halo has been built from accreted satellites using the Spaghetti data and the results from the previous section. To this end we compare our dataset statistically to a simulation of a halo which is entirely the result of disrupted dwarf galaxies. For this purpose we use the simulations from Harding et al. (2001) that model the destruction of a  $10^7 M_{\odot}$  satellite on different orbits by the Milky Way galaxy.

The Milky Way galaxy is modelled using a Johnston, Spergel and Hernquist potential, consisting of three components: a disk, spherical and logarithmic potential. The satellite follows a truncated Plummer model populated with 200,000 particles and has properties similar to the Milky Way dwarf spheroidals, with a core radius of 0.1 kpc and a tidal radius of 2.0 kpc. The velocity dispersion of all particles in the satellite is  $6.5 \text{ km s}^{-1}$ . The evolution of the particles in the satellites were followed for  $10^{10}$  years (Harding et al., 2001).

Thirty of the original simulations were re-sampled by Heather Morrison so that each particle corresponds to one halo K giant. This leaves about 8000 particles per simulation. A lower limit to distance from the Sun at 15 kpc has been imposed to almost all simulations to reproduce the sampling of the Spaghetti survey. However, three simulations are included which do have stars closer than 15 kpc. The sky distribution of these K-giant simulations is shown in Figure 5.11.

To create a halo built up completely out of disrupted galaxies, the end-points of the simulations (i.e. evolved for 10 Gyr) are used. From this sample of over 23,000 simulated ‘giants’ we draw subsets resembling the Spaghetti survey dataset. This is done by re-sampling the “K-giant” simulations such that the observed sky distribution, distance and radial velocity distribution are matched. In this manner, 10 simulated datasets are drawn which closely resemble the Spaghetti dataset. Unfortunately we are not able to match exactly the velocity distribution, especially the high velocity regime, where a smaller number of simulated stars are present compared to the observed Spaghetti dataset. This lack of high velocity stars is related to the lower limit imposed on the heliocentric distance in most simulations. Therefore, our simulated datasets have a slightly smaller number of stars. They contain on average 93 ‘giants’, while the original Spaghetti dataset contains 102 confirmed giant stars.

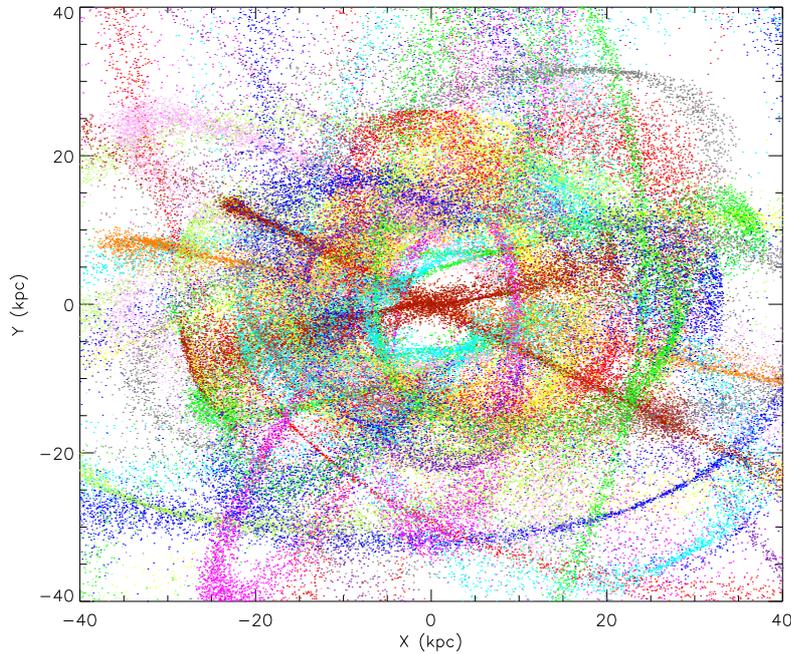


Figure 5.11: All the streams in the 30 re-sampled simulations plotted on the sky in galactic  $x$  and  $y$  coordinates. The hole in the middle is caused by the 15 kpc limit which is imposed on the streams.

### 5.2.1 Substructure in the simulated datasets

#### The 4distance method

We subsequently look for substructure in the simulated datasets in the same way as we did in the dataset. All structures found with the friends-of-friends method below a  $4dist$  of 0.04 (FOF4) and 0.08 (FOF8) are considered. In order to make a fair comparison with the dataset, the ‘giants’ from the simulation sets are convolved with errors, to mimic the observational errors. For the distance a relative error of 15% is used, while the velocities are convolved with errors similar to the errors of the matching star number in the dataset. The errors are then multiplied with a random number drawn from a normal distribution with a mean of 0 and a standard deviation of 1 and added to the value of the corresponding parameter. The substructure detected in each of the simulated datasets, before and after error convolution, are compared to the amount of pairs found in the dataset with FOF4 and FOF8 are shown in Figure 5.12. The results have been corrected for the difference in number of stars in the different sets. We were concerned that the three simulations which are followed at distances below 15 kpc would

provide a bias in the amount of substructures found, because all the stars in the simulated datasets closer than 15 kpc would originate from only three different satellites. However, the amount of pairs found below  $4dist = 0.04$  within 15 kpc from the Sun in the 10 simulated datasets is just 4% of the overall number of pairs.

In Figure 5.12 we show the number of pairs found below a certain  $4dist$  in every simulated dataset as well as the fraction of stars found to be in substructures. These two measures are related, but not the same since a given star may belong to more than one pair (e.g. would be part of a larger structure). This explains the difference between top and bottom panels.

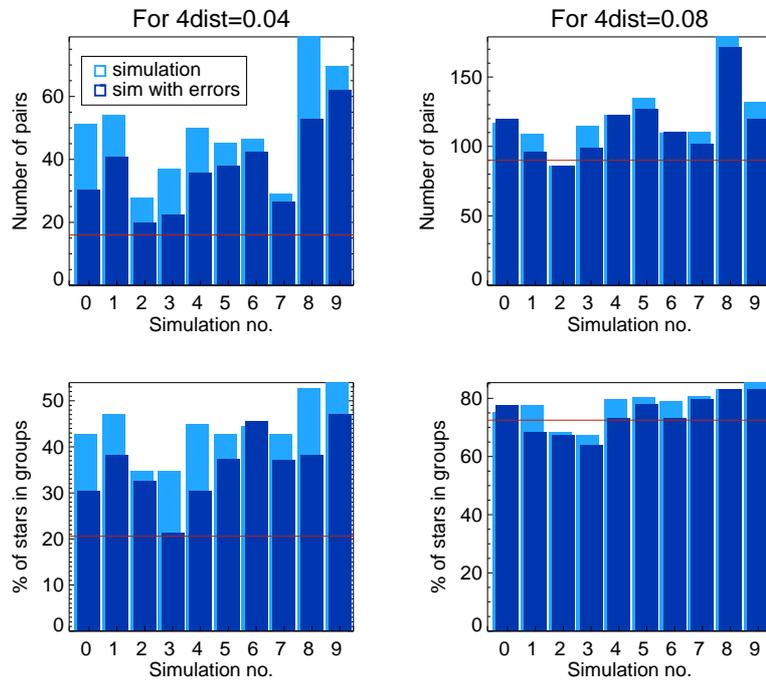


Figure 5.12: Top panels show the number of pairs detected with FOF4 and FOF8 in ten subsets of the “K-giants” simulations. The light blue bars correspond to the simulated subsets, while the dark blue bars denote the number of pairs found in the subsets convolved with observational errors. The red horizontal line represents the number of pairs found in the Spaghetti dataset. The bottom panels show, for the same simulations, the fraction of stars found to be in substructures.

From Figure 5.12 it becomes clear that the number of pairs found in a simulation dataset can vary significantly from dataset to dataset. For example, simulated dataset no. 3 has a comparable number of pairs to the dataset (if observational errors are taken into account). On the other hand some other simulated datasets like no. 9 show much more substructure at all levels.

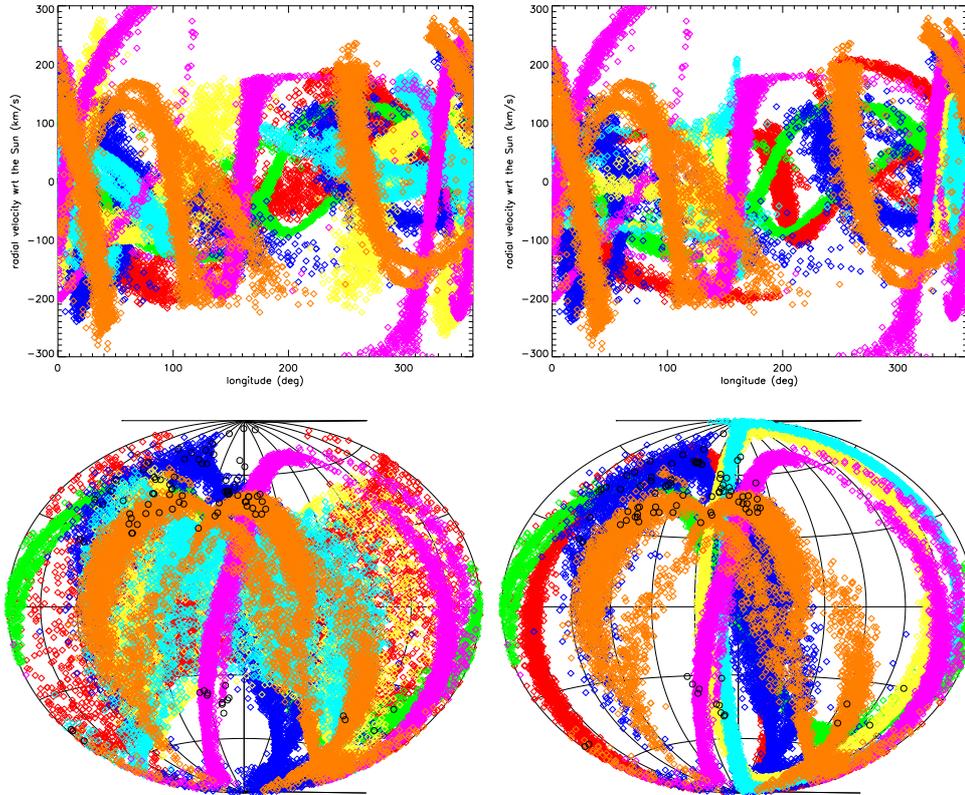


Figure 5.13: A longitude versus radial velocity plot and a skyplot for all the streams of which were found groups in simulations 3 (left) and 9 (right) with the friends-of-friends method for a 4distance below 0.04. Over-plotted black dots on the lower panels represent the stars in the simulation subsamples.

Figure 5.13 may be used to understand the origin of the difference between the two subsets. Plotted in this figure are longitude versus radial velocity (top panel) and the distribution on the sky of the various satellites that contributed to the simulated datasets. The stars in the two simulated datasets are over-plotted as black dots in the bottom panels. The distribution of the two datasets on the sky differs slightly, because the stars are drawn from the large simulation set using sky coverage bins which are ten square degrees wide. A simple comparison of the two simulated datasets reveals that although both sets contain some broad star streams, subset 9 also has a relatively large number of stars originated in much narrower features. Because these features are more confined, this substructure is easily picked up by our *4dist* method. The substructure found in a particular subset is therefore correlated to the confinement of the streams from which the stars in the subset belong.

To further quantify the nature of the substructure, we examine now the number of larger groups that can be formed from the previous found pairs. Figure 5.14 shows the distribution of the various groupsizes for our simulated subsets and compared to that of the Spaghetti dataset. Again, simulated dataset no. 3 appears to be quite similar to the Spaghetti sample. Like the Spaghetti dataset one large group of six members is found below  $4dist = 0.04$  and one group of three members. Additional to these larger structures the simulated dataset 3 has five pairs below  $4dist$  compared to six in the Spaghetti sample, but the simulated dataset only has 89 stars while the Spaghetti sample has 102. Other simulation subsamples show very different grouping of substructure. For example simulated dataset 2 has no substructure with more than three members and simulated dataset 8 has just pairs and one very large group.

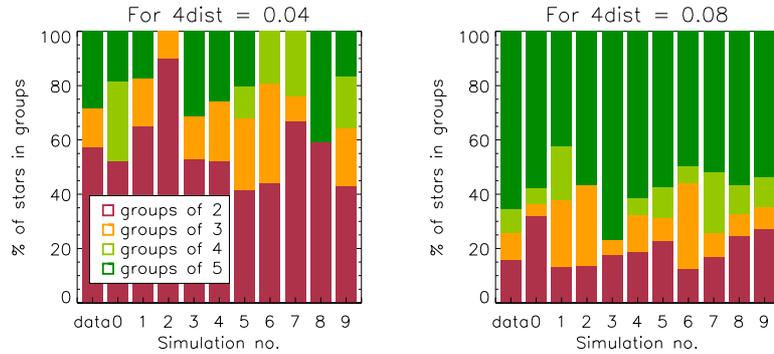


Figure 5.14: The distribution of the detected substructure over number of stars in a single group. The first bin shows the distribution over groupsizes in the Spaghetti dataset.

Because of its similarity to the Spaghetti dataset, we now focus in more detail into simulated dataset 3. The distribution of its stars on the sky and in radial velocity and distance are shown in Figure 5.15. Over-plotted in coloured dots are the five pairs, one group of three members and one group of six members of this simulated dataset. Comparison with Figures 4.2 and 4.3 show that this subset of the simulation is indeed very similar to the Spaghetti dataset. We have therefore decided to apply the  $4dist$  method to this set exactly in the same way as we have done with the Spaghetti data. We have generated random sets for this subset and compared the number of random simulation pairs to the simulation dataset 3 to determine the significance level of the substructure. Figure 5.16 shows that for most values of  $4dist$ , the two sets have not only the same amount of substructure, but that also, within error bars, the substructure has the same level of significance.

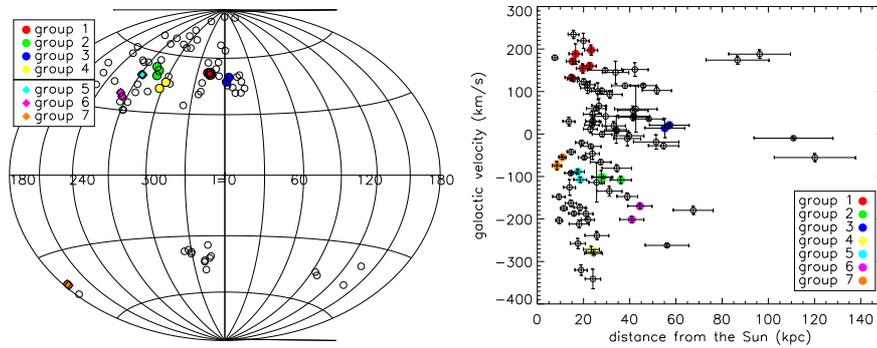


Figure 5.15: Skyplot and distance versus radial galactic velocity plot for all the groups found with the friends-of-friends method below 4distance 0.04 in simulations number 3.

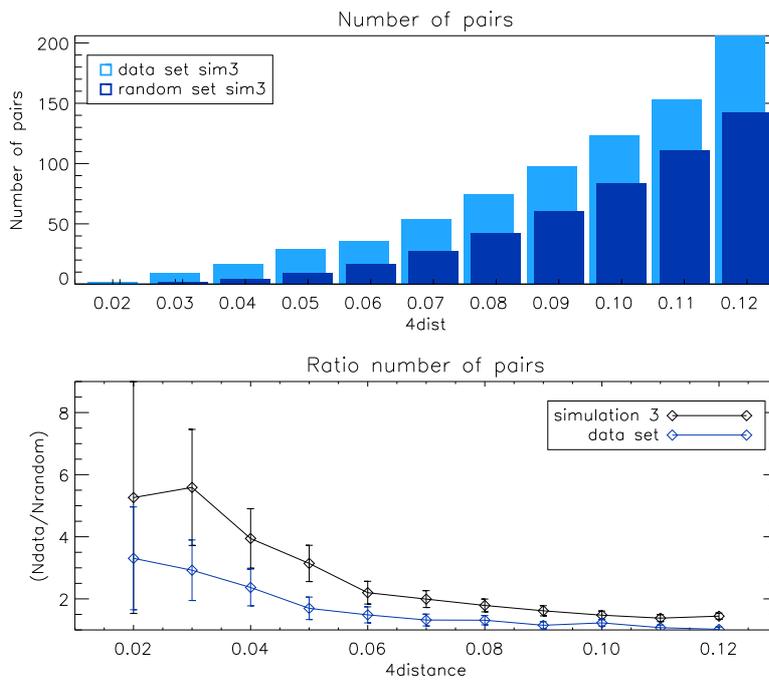


Figure 5.16: The significance of the groups found in simulation number 3, in the lower panel compared to the data set (blue).

### Evaluation of the 4distance method

One of the greatest advantages of working with simulated datasets is that we can check whether the substructure found comes indeed from the same

parent satellite. By tagging the stars given to each subset, it is easy to trace their progenitor. For each group we have checked whether all of its members originate from the same parent. Figure 5.17 displays the results of these tests separated for the different group sizes and two group-finding criteria. Using FOF4, it turns out that almost 92% of the pair members in all the simulated datasets do share common parent satellites! For larger groups this percentage slowly decreases as it becomes more and more likely one mismatched pair will be part of a larger group. Still, even for groups with 5 members or more over 50% of the members are a correct match. The number of mismatches is larger when we look at  $4dist$  of 0.08, but even at this criterion still 62% of the pair members share common progenitors. In simulated dataset 3 alone, which resembles most closely the Spaghetti dataset, 100% of the found pairs are originating from the same progenitor.

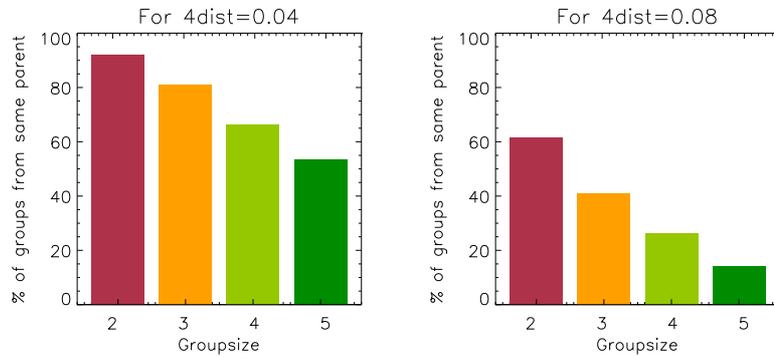


Figure 5.17: Panels show the percentage of groups of a particular groupsize for which all of the members have the same progenitor for both the FOF4 and FOF8 method.

### The Great Circle method

We use the great circle method on two simulated datasets, number 3 and 9 to find connections between groups in the same way as we did on the Spaghetti dataset. Simulated dataset 3 is chosen, because of its similarity in amount of substructure to the Spaghetti dataset and simulated dataset 9, because it possesses much more substructure and as such it can serve as an extreme example. Analogously to the great circle method for the Spaghetti dataset (see Section 4.2) the substructure found from the 4distance method, which is discussed in the previous section, is used as a starting point. For the simulated datasets we take all substructure found within  $4dist = 0.04$  into account.

The Great Circle Pole Families on the sky and position of the individual ‘giants’ of these substructures for both the simulated datasets is shown in

Figure 5.18. The datasets are convolved with errors, as described in Section 5.2.1, and all groups that are within  $10^\circ$  from each other are taken as one in the determination of their Great Circle Pole Family. The number of groups that could be linked because they shared a GCPF crossing point within  $10^\circ$  and their ‘giants’ were on a straight line in the  $E_r$  versus  $r^{-2}$  diagram with a goodness-of-fit parameter  $Q \geq 0.99$  is given in Table 5.4. The number of links found between groups is normalised with respect to the Spaghetti dataset by the ratio of the number of GCPF crossing points in the simulated and the Spaghetti dataset. Because simulated dataset 9 possesses a lot more substructure from the 4distance method, we expect more of these groups will be connected by chance. The normalising factor used is compensating for this effect.

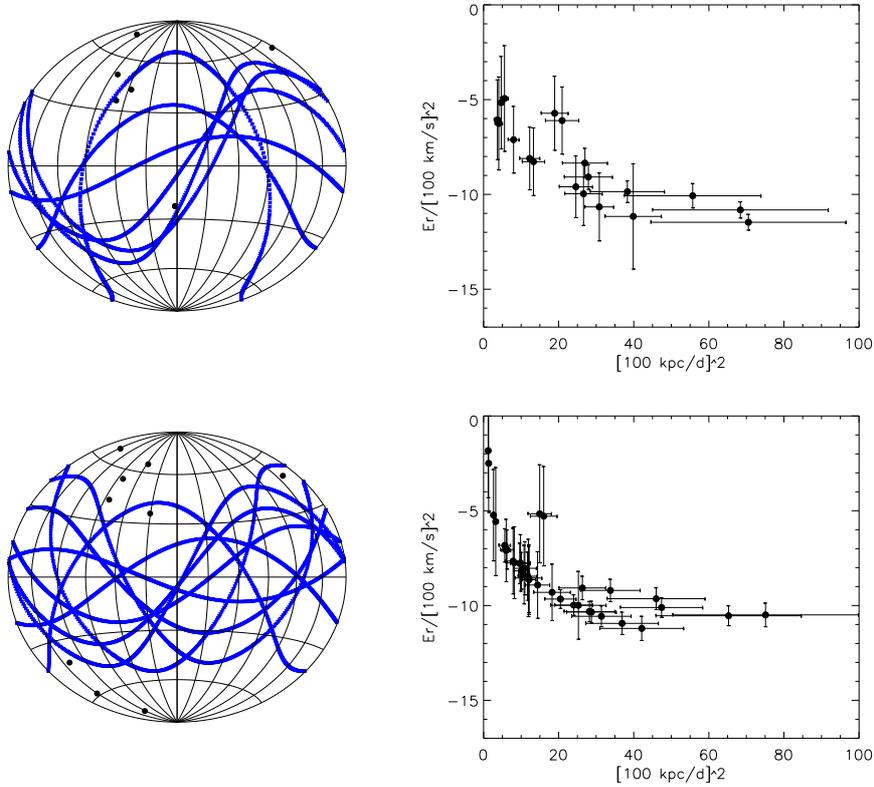


Figure 5.18: The plots on the left show the possible angular momentum poles for all the groups in the two simulated datasets (simulation 3 and 9) that were already found to be substructures from the 4distance method while the plots on the right show the position in the  $E_r$  versus  $r^{-2}$  diagram for all the stars in these groups.

Dataset	# Links between 3 groups	# Links between 4 groups	Fraction of groups linked
Spaghetti	2	-	5/6
Sim3	2.9	0.7	7/7
Sim9	2.7	0.7	11/12

Table 5.4: Number of groups found from the 4distance method that can be linked together by the great circle method. The number of links between groups in the simulated datasets is normalised by the ratio of the number GCPF crossing points started from in the Spaghetti and simulated dataset. Links between 3 and 4 different groups are evaluated separately.

### Evaluation of the great circle method

In the simulated datasets, we can trace the progenitors of the ‘giants’ in our simulated datasets. As shown in the previous section, most members of groups found with the 4distance are indeed originating from a common progenitor. Because we know which parent satellites the groups are originating from, we can also evaluate how well our great circle method is linking structures from common progenitor. For simulated dataset 3, only one of the found connections contains 2 groups from the same satellite. For simulated dataset 9, just two links between two groups from the same progenitor are found.

This bad result can partly be due to the fact that neither of the simulated datasets possess enough substructures from the same progenitor. In neither simulated dataset 3 nor 9 three groups are found which originate from the same progenitor. In our implementation of the great circle method, we can link only three or more groups, which in our simulated datasets never arises from the same progenitor. However, the great circle method matches a substantial number of groups together, in both the Spaghetti dataset and the simulated datasets. In all three datasets almost all groups we started from are linked to other groups, despite the very strict requirements on both the angular distance between the GCPFs and the goodness-of-fit in the  $E_r$  versus  $r^{-2}$  diagram. This result casts serious doubts on the use of the great circle method on datasets which are as small and have as “large” distance errors as the Spaghetti dataset. Since an equal amount of matches between the substructures are found in both the Spaghetti dataset and the simulated datasets (normalised to the number of GCPF crossing points we are starting from) and almost all matches in the simulated datasets turned out to be accidental, we should treat also the great circle results on the Spaghetti dataset with much care.

### Conservation of angular momentum?

In order to get a better understanding of how well our initial assumptions, in particular conservation of angular momentum, hold in the outer halo, we have a closer look at all the stars from five streams originating from five different infalling satellites in the simulation. The five streams are chosen such that their orbits have a range of different properties. The streams from the five disrupted satellites as well as their angular momentum magnitude and orientation specified by the angles  $(\theta, \phi)$  are shown in Figure 5.19. From this figure, it is clear that satellites which come closer to the Galactic Centre in their orbit are much more affected in their angular momentum properties. The effect is most evident in  $\phi$  as the galactic disk potential is causing a precessing motion. From this figure it is clear that we can not expect all satellite debris in the halo to remain on the same orbital plane with a similar direction of the angular momentum pole. Also, there is some scatter in the angular momentum magnitude in the satellite debris, although this effect is far less dominant.

To get a better understanding on the position of ‘giants’ in the simulated datasets in the  $E_r$  versus  $r^{-2}$  diagram, we construct another simulated dataset from these five streams matching the sky position, distance and radial velocity distribution from the Spaghetti dataset. This simulated dataset has 69 stars. We subsequently plot all ‘giants’ from the five streams with different properties described above that are part of this new simulated dataset in a  $E_r$  versus  $r^{-2}$  diagram. Figures 5.20 and 5.21 show these diagrams for the simulated dataset and simulated dataset convolved with errors. In addition also all the stars in this new simulated dataset are plotted in the right lower panel. From these figures it is clear that there will be some scatter around the straight line of the simulated giants from a common progenitor. Also it is difficult to reconstruct the straight lines from the  $E_r$  versus  $r^{-2}$  diagram with all the stars. Another remarkable thing is that in the dataset which is convolved with errors, it is not necessarily harder to fit the ‘giants’ by a straight line. In some cases it is even the other way around, because of the scattering and the extensive error bars, the points can be fit much easier by a straight line. This implies however that also structures which do not possess the same angular momentum will be easier connected by a linear dependence.

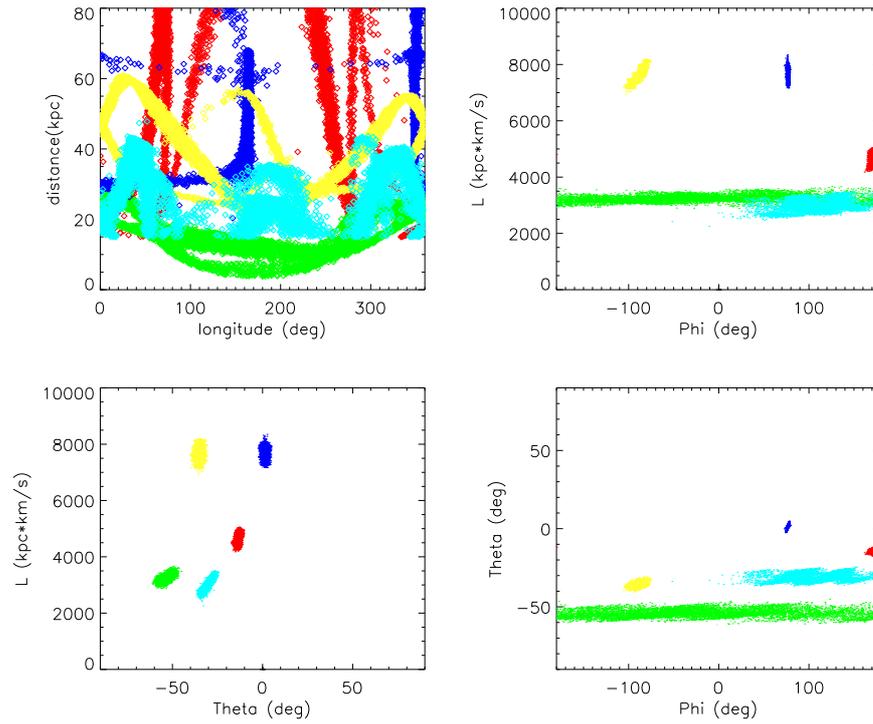


Figure 5.19: Five streams from the simulation by P. Harding re-sampled by H. Morrison such that each particle resembles a halo K-giant. Shown here are streams 0,3,6,7 and 26. The left upper panel shows the streams in distance versus galactic longitude. The left lower panel shows the one angle of the angular momentum vector,  $\theta$ , the angle which denotes the height above the plane, versus the magnitude of angular momentum for the same streams in the same colour coding. The right upper and lower panels show the direction of the angular momentum vector in  $\phi$ , which is defined as the azimuthal angle in the xy plane, versus the magnitude and  $\theta$  respectively.

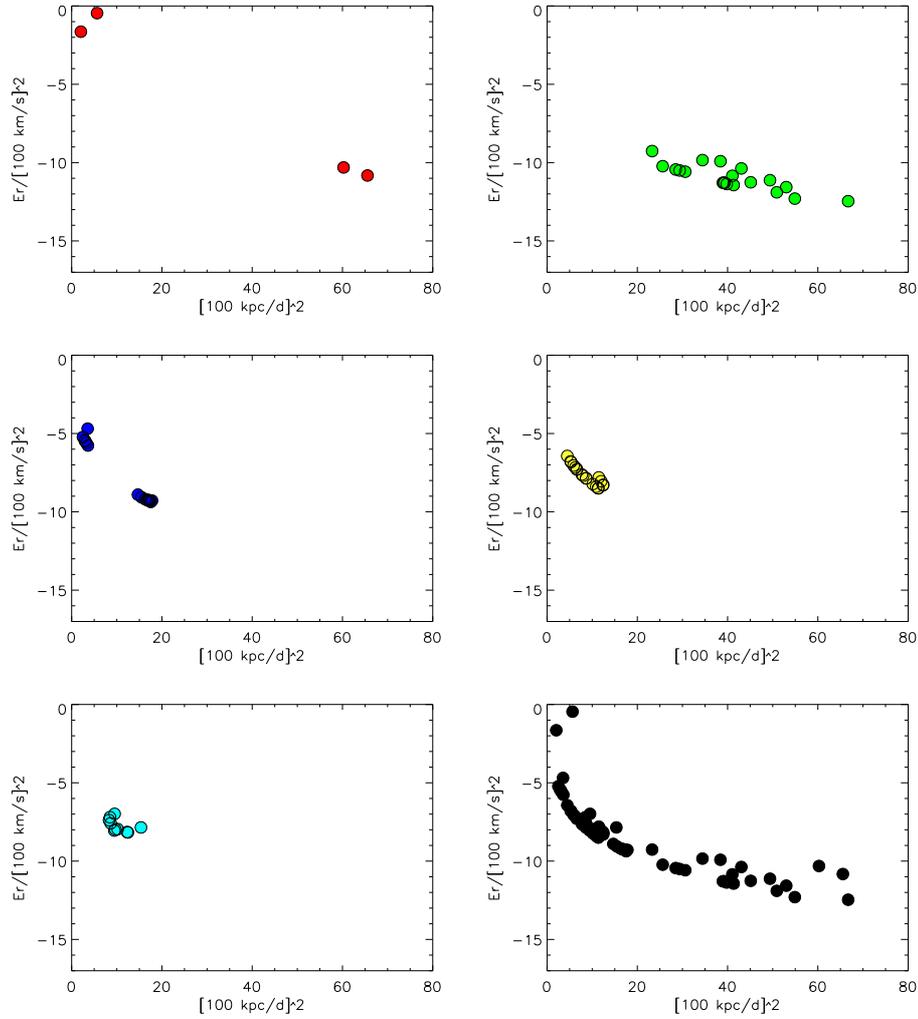


Figure 5.20: All ‘giants’ from the five streams shown in Figure 5.19 that are part of the simulated dataset constructed from the five streams, plotted in a  $E_r$  versus  $r^{-2}$  diagram. In all panels ‘giants’ from one particular parent satellite are plotted, using the same colour coding as in Figure 5.19. The last panels shows all giants from all of the five streams in the simulated dataset. The simulated subset is not convolved with errors.

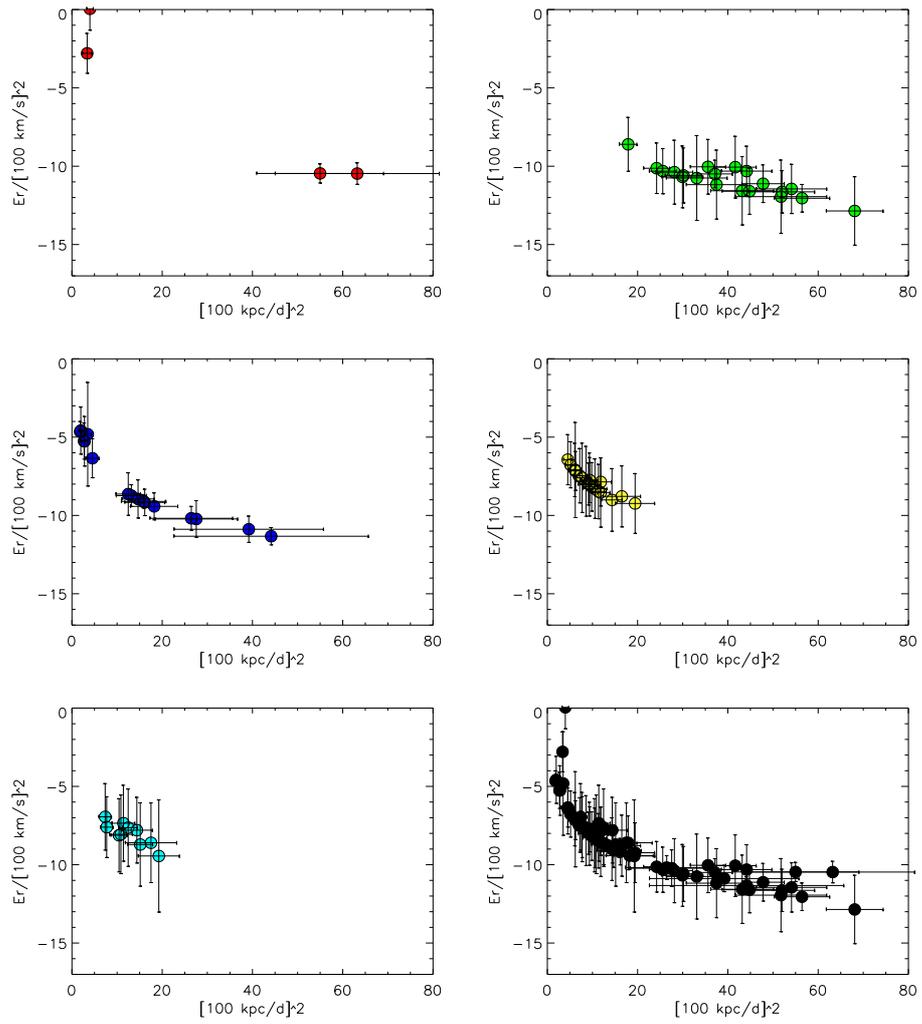


Figure 5.21: Same figure as Figure 5.20, but now the dataset is convolved with errors.

## Chapter 6

# Discussion and Conclusions

The Spaghetti survey (Morrison et al., 2001) was undertaken to quantify the amount of substructure in the Galactic halo. The survey is designed as a pencil-beam survey of high latitude fields using red giants as tracers for substructure. Using CCD-photometry giant candidates are preselected, which are later followed up spectroscopically. From comparison of their spectra to spectra of standard stars radial velocities are measured. Subsequently a careful luminosity classification and metallicity measurements are performed using the line strengths of Ca II K, Ca I  $\lambda$ 4227, Mg *b*/H as well as Sr II  $\lambda$  4078. Distances for the stars are calculated by estimating their absolute magnitudes using the V-I globular cluster giant branches of Da Costa and Armandroff (1990).

A final dataset is constructed containing 102 giants with distances ranging from  $\sim 5$  until  $\sim 100$  kpc from the Sun. The typical errors on distance are 15%, on the radial velocity the typical errors are 15-20 km/s and the typical metallicity error is 0.25-0.3 dex.

### 6.1 The 4distance

To quantify the amount of substructure in the Spaghetti dataset we define the 4distance, a distance measure using normalised contributions of the angular distance, radial distance and radial velocity differences between the stars in the dataset. First, the stars are combined in pairs and for each pair the 4distance is measured. These pairs can later be combined into larger groups if they are close enough in 4distance. The scale at which we take substructures (both pairs and larger groups) into account is defined by comparing the dataset to random sets, which are created by shuffling the distance and radial velocity values for the stars. We take two different scales into account (0.04 and 0.08) and use two criteria, the friends-of-friends criterion (FOF) and the all-friends criterion (AF) to define our groups. The final groups are put together taking all groups found with FOF4 potentially

extended with stars from the AF8 criterion. We find 7 groups which we consider to be significant.

Of these seven groups one really stands out in its significance. This group with a striking 7 members can confidently be associated with debris from the disrupting Sagittarius dwarf spheroidal galaxy. Closer examination of models of the Sagittarius galaxy debris reveals three other groups, one of three, and two with two members, which might be associated with the disrupting dwarf. From comparison with the Sagittarius models it follows these groups might be associated with the Sagittarius debris if these stars were stripped off earlier, between 3 and 6 Gyr ago. The membership of two of the groups is more likely if the galactic halo potential is considered to be more oblate than prolate. However, the third group matches best in a prolate halo.

In addition to the found Sagittarius substructure three groups can be related to one of the known Virgo substructures from their sky position and distance information. Given the large number of known overdensities in Virgo and the unknown origin of these structures, it is possible to find matches, but our data alone do not allow us to say whether there is any physical relation between them. Of the three groups possibly related to Virgo, two are also possible matches with the Sagittarius debris discussed above. SDSS photometry (see Figure 5.1.2) suggests that due to their low latitude the two groups are probably not related to the *leading* arm of the Sagittarius dwarf. If we associate these two groups with Virgo overdensities and not with Sagittarius debris, the only extra group still associated with Sagittarius debris (apart from the group of 7 members) favours a prolate halo potential. However, from our data alone, we can give no conclusive answer on what substructure the groups belong to.

There are only two other groups in the Spaghetti dataset which can not be related to any known substructures today. The first group is a relatively close group of 4 members, its heliocentric distance is only 4 - 9 kpc and it is less metal-poor than the average halo metallicity (-0.9 dex). Because it is such a large group and its metallicities are strikingly similar, we do not consider it very likely that this group is just a chance match. The second group, of just two members, is also special, since it is the only substructure we found outside the scope of the SDSS photometry survey (see Figure 5.1). Therefore the chance for it to be connected to already known substructure is smaller, since this region of the sky is less extensively studied.

From evaluation of the random sets we expect roughly 8 of the 16 pairs to be chance matches. We also compared the substructure in the Spaghetti dataset to a set of simulations in which a halo was built up completely from 23 disrupted satellites of  $\sim 10^7 M_{\odot}$ . We randomly sampled these simulations according to the Spaghetti fields distributed on the sky. In this way, we produced 10 simulated datasets which we analysed with the same statistical methods as the Spaghetti sample. If we evaluate the substructure found

in these simulated datasets with FOF4, over 92% of the matches of giant star particles to pairs is real, meaning that the two ‘giants’ matched did originate from the same progenitor in the simulation. For example in one of our simulated datasets which is closely resembling the Spaghetti dataset we expected roughly 4 of the 17 pairs to be chance matches, yet all the substructure found in this dataset was indeed originating from common progenitors.

## 6.2 The great circle method

The great circle method is used to search for more extensive substructure over the sky. The underlying assumption for the use of this method is that satellite debris from a common progenitor will continue to orbit on a great circle in a common orbital plane no matter how far apart on the sky the debris is today. Therefore, the direction of the angular momentum pole (or great circle pole) will remain the same for all debris from a common progenitor. Also the magnitude of angular momentum and the total energy are taken to be conserved quantities.

Because of the clumpy distribution of stars in the dataset and the dominating galactic potential shape and large error bars in the  $E_r$  vs.  $r^{-2}$  diagram, in which all objects with similar total energy and angular momentum magnitude will lie on a straight line, we chose to apply the great circle method not to the individual stars but to connect the already found clumpy substructures with the 4distance method. Only the substructures farther out than 10 kpc are considered. Applying the criteria of a crossing point of Great Circle Pole Families (GCPF, the possible orientations of the angular momentum pole for one object) within  $10^\circ$  and a goodness-of-fit for the straight line fit of  $Q = 0.99$  we found two substructures of each three groups which were possibly dynamically related. For both associations of groups, one group is strongly associated with debris from the disrupting dwarf galaxy Sagittarius. A further comparison with models of the disrupting Sagittarius dwarf galaxy favours one of the two found associations. From these models we can not exclude the possibility all three groups within this association are substructure originating from this disrupting satellite, but it would have to be stripped off at different times and both extra groups favour different halo potentials (oblate and prolate). Favouring a prolate potential would mean that only one of the extra groups can be linked to Sagittarius debris. These results are consistent with the two other groups belonging to Virgo substructures (see discussion in the previous section).

Another indication that this association might not be real is that no three groups from one parent satellite were found in any of the investigated simulated datasets, although one of the investigated simulated datasets possessed much more substructure in groups than the Spaghetti dataset. While

it statistically seems more likely that this match of three is coincidental, an explanation for this could be that estimated mass for the original Sagittarius dwarf galaxy is 10 times larger than the masses for the satellites in the simulations (Velazquez and White, 1995). Because the mass of Sagittarius is larger than the masses of the progenitor satellites in the simulated datasets, it will produce more debris in broader tails and longer streams. This opens the possibility that the Spaghetti dataset may contain more than three groups from one common progenitor (e.g. Sagittarius), whereas no three groups from a common progenitor can be found in the simulated datasets.

From the analysis of the great circle method on the simulated datasets, it is clear however that this method is not working as well as we would have liked. By far most structures matched in GCPF crossing point and straight lines in the  $E_r$  versus  $r^{-2}$  diagram do not originate from a common progenitor. The failure to connect groups from the same progenitor in the simulated datasets can easily be explained, since there is not enough substructure from common progenitors in these datasets. It is problematic however, that despite the lack of real structures still almost every group is linked to some other groups.

The amount of groups linked using this method in both the simulated datasets and the Spaghetti dataset is similar (normalised to the number of crossing points started from).

Overall, the requirement to link at least three groups together, which is inherent to the method of searching for GCPF crossing points (two GCPFs will always cross), might be too ambitious for datasets the size of the Spaghetti dataset. Also, the dominating potential shape and extensive error bars in both distance and radial velocity make it much harder to throw out possible groups from the requirement that they have to be on a straight line in the  $E_r$  versus  $r^{-2}$  diagram. From the analysis of the simulated subsets we have learned that even without observational errors not all satellite debris can be expected to stay on the same orbital plane and be on a straight line in the  $E_r$  versus  $r^{-2}$  diagram.

### 6.3 Successes and limitations of the Spaghetti project

Based on the number of stars probed, 102 giants, the Spaghetti Survey is a small survey, certainly compared to very extensive survey projects like SDSS. The main aspects that make the Spaghetti project unique is the high quality of its data and amount of information for every object (distances with ‘just’ 20% error bars, a thorough luminosity classification and radial velocity information and metallicity measures for every giant) combined with the large distances it probes (giants out to  $\sim 100$  kpc). The Spaghetti survey is really probing the outer parts of the halo of our Galaxy.

In this report we have shown that it is very well possible to trace larger substructures in the outer halo with this survey. We confidently identified a clump of debris from the disrupting Sagittarius dwarf galaxy as well as other smaller groups that might also be part of the satellite’s debris. Subsequently we identified three groups which might be associated with the Virgo substructures. Two other groups are left, which might or might not be real substructures.

The information we have on four of the six phase space coordinates for all Spaghetti giants also makes it suitable data to further constrain the properties of streams, once membership of a giant to the stream can be confirmed. An example shown in this report are the two giants we found that might be members of the Orphan Stream. Membership of either of these giants, which possess very different radial velocities, would have resulted in a more reliable measurement of the Orphan Stream’s radial velocity. Earlier measurements performed using SDSS spectroscopy (Belokurov, 2007) suffer from the uncertainty that the candidate stars might not be giants, they could as well be foreground dwarfs in the disk of the Milky Way. In the Spaghetti dataset, we find one Orphan stream candidate which is right on the stream in sky position and distance and confirms the radial velocity as tentatively measured by the SDSS team. Unfortunately however, we do not find enough Orphan Stream candidates to conclusively determine the radial velocity of the stream.

While the Spaghetti survey has two fields right on the Orphan stream we only find a small number of candidates which match the stream properties. The Orphan Stream has an approximate surface brightness of  $\sim 32.4$  mag arcsec $^{-1}$  (Belokurov, 2007), which using the M92 luminosity function from Paust et al. (2007) transforms to 1.3 giants per square degree. Considering the 0.25 square degree fields the Spaghetti project uses for the photometry, it is not surprising just one candidate per field was found even though we were tracing right on the stream. This example clearly marks the limitations of the survey. While the Spaghetti survey is very well suited to find and trace substructures especially farther out in the halo, as any red giant survey it will have limitations on surface brightness of the substructures it can detect. The Orphan Stream seems to be right on this boundary. With more fields on the stream and a more careful analysis of the existing SDSS spectra for giant candidates in fields on the stream, it should be possible to constrain the radial velocities better. However, fainter substructures might need other tracers such as main sequence stars to determine their properties. This also means that “pure” red giant surveys may not be able to put any constraints on the low mass end of the luminosity function for very early on accreted objects.

## 6.4 The properties of the Galactic stellar halo

The main question that the Spaghetti survey aims to answer is: “What is the amount of substructure in the stellar halo?” From a careful analysis of our dataset of 102 giants we have found 7 significant groups of giants. In these seven groups are 23 giants, which is 22.5% of all the giants in the Spaghetti dataset. From the analysis of random sets we expect somewhat over 8 pairs to be chance matches. Making the most conservative estimate this would leave 6 ‘real’ matched giants in the dataset. Nevertheless, this very conservative argument gives us a lower limit for substructure we expect in the outer halo of 6%. We think this measure is very conservative for two reasons. Firstly, there are many groups found with more than 2 members, which means there are a lot of stars which are in more than one pair whereas in this measure all ‘random’ pairs are treated as if they consist of two unique giants. Secondly, both the amount of groups we found which can be linked to already known substructures and the analysis of the groups found in the simulated datasets indicate that more than 6 giants in all the matches might be real.

Because of the limit we have on the detection of debris from small mass satellites due to their low surface brightness (see discussion in the previous section) we can not put an upper limit on the total amount of substructure in the halo. In our dataset we find no more than 22.5% to be in large substructures well above our detection limit, which would be around the mass of the Orphan stream progenitor. This upper limit percentage is very tentative however and can not easily be generalised. Although we are able to trace the Sagittarius dwarf galaxy and maybe even the Virgo overdensity, the two only known large substructures in the part of the sky probed by the Spaghetti survey, it is very well possible we missed even relatively large substructure at larger distances, or other directions, because the survey has only probed a small number of directions on the sky. Whether the distribution of (larger) substructures would be really isotropic on the sky is a question which remains yet unanswered. It is remarkable however that although  $\sim 36\%$  of the Spaghetti dataset is outside the coverage of SDSS (see Figure 5.1), just one pair of stars (and one star belonging to a pair) is found in that region!

We compare the substructure in the Spaghetti dataset to 10 simulated datasets which we analysed with the same statistical methods as the Spaghetti sample. The variations in the 10 simulated datasets are quite large, predominantly due to the fact that some datasets consisted of stars which mainly originated in very tight streams, while other stars came from more dispersed, broader debris. Nevertheless, one simulated datasets was found which, within error bars, resembled the Spaghetti dataset and even showed the same amount of stars in larger groups. We thus conclude from this comparison of substructure in our dataset that our results are consistent with

the whole stellar halo being built by accretion. If so, the halo would be dominated by broader structures and not by very narrow streams, which would give an even larger amount of substructure with the 4distance method provided they have a surface brightness well above our detection limit. Such a large amount of substructure has not been measured. Another possibility is that the accreted satellites in the stellar halo are more massive than  $10^7 M_{\odot}$  which would also result in broader features. For the only substructure we trace with a known progenitor, the Sagittarius stream, this is indeed the case, the mass of the initial Sagittarius satellite is estimated to be  $10^8 M_{\odot}$  (Velazquez and White, 1995).

## 6.5 Conclusions

We have shown that it is possible to detect substructure from the Spaghetti Survey, a pencil-beam photometric survey followed up by spectroscopy for red giant candidates. In total, 102 giants are classified for which we have distance, radial velocity and metallicity information. Using a distance measure combining spatial and radial velocity information of the stars, we find 7 groups which contain a total of 23 stars representing 22.5% of the total dataset. The most outstanding group, which contains a striking 7 members, can confidently be associated with debris from the disrupting Sagittarius dwarf galaxy. Three other groups found might be associated to Sagittarius debris, provided they were stripped off earlier. These groups also require the halo potential to have a particular shape (oblate or prolate). Subsequently three groups can be associated with known Virgo overdensities. Two of these groups were previously also associated to Sagittarius. From SDSS photometry it seems unlikely that they will be associated to the Sagittarius *leading* arm and they can more likely be associated with the Virgo overdensities as a separate entity or entities.

We have attempted to connect the previously found groups into larger structures over the sky by looking for associations which possibly share the same total energy and angular momentum as well as one orbital plane on the sky. Two associations of three groups are found, but neither of them show a confident match to debris models. From a similar analysis on simulated datasets, we find that no real associations can be made because of the lack of structures from a common parent satellite, but still almost every group is linked. The requirement to link at least three groups together, inherent to this method, seems too ambitious. A second problem are the dominating galactic potential and the large measurement errors which hamper the search for structures with similar angular momentum and total energy.

We have found two stars in the Spaghetti dataset which are candidate members of the Orphan Stream. One of the candidates supports the radial velocities as measured by Belokurov et al. (2007). However, we can not

conclusively determine the radial velocity of the stream from our data. The Orphan stream's surface brightness is on the boundary of what any red giants survey can detect. Any substructures on the lower mass end of the luminosity function will need additional tracers such as main sequence stars.

From our analysis we have measured the amount of substructure to be 22.5% in the Spaghetti dataset. From comparison with random sets we have derived a very conservative lower limit of 6%. No conclusive upper limit can be given. From comparison with datasets drawn from a simulated halo built up entirely of disrupted satellites it follows that the Spaghetti dataset is consistent with the whole stellar halo built by accretion. In this case, the characteristics of the substructure found seem to imply that broad streams dominate our dataset. This would suggest early merging and/or relatively heavy progenitors.

## 6.6 Future Work

A first objective for further research is to look for more substructure in other datasets (for example the sample of Blue Horizontal-Branch (BHB) stars in SDSS (Sirko et al., 2004)) or take more data in adjacent fields to investigate whether any stars could be added to the substructures found. Extra stars found will strengthen the hypothesis that these substructures are real and help to constrain the properties (radial velocities etc.) of the groups. A similar approach was already suggested in Section 5.1.3 to better constrain the radial velocity of the Orphan Stream. But also the groups associated with the Virgo overdensities and the groups found that could not be linked to previous known substructures would be interesting starting points. For the Orphan Stream also the modelling possibilities need to be investigated further, in order to restrict the range of possible radial velocities or even the properties of the progenitors.

Overall, the successes of the 4distance method within the Spaghetti dataset show the clear possibilities for even more extensive spectroscopic datasets, like SEGUE and eventually GAIA. Because GAIA will possess full phase-space information for a lot of stars, groups will be constrained even better. Also, we expect the great circle method to be much more effective once proper motion information is available. The proper motions already confine the stars into one orbital plane which will dramatically reduce the amount of chance matches.

# Acknowledgements

First of all I like to thank Amina Helmi for taking up this project with me and supervising me in such a great way. I've really enjoyed working together. We had many nice and encouraging discussions during the course of this project and she was always there to answer questions or discuss anything. In the process of writing, she somehow always found time and patience to read my work multiple times very carefully. I have learned a great deal from her in a very wide range of topics and I'm sure there is still much more to learn!

Heather Morrison and Paul Harding have been great supervisors during my time in Cleveland. I have learned a lot from them regarding astronomy, reduction techniques and (academic) life in general. In particular I would also like to mention their hospitality, taking me around to the telescope, the desert museum, dinners, etc. In short, they just made me feel at home, together with the rest of the CWRU astronomy institute, at my first trip ever on my own. Back in Groningen I received many encouraging and stimulating emails that helped to keep the project going.

There are a few other people I would like to thank. Huug van Woerden has made this project and my trip to Cleveland possible and helped to get the project started off. My fellow students and roommates I'd like to thank for all the support, discussions, tea and relaxing times. Last but definitely not least I would like to thank the people outside the institute who were always there. Joost, my friends, parents and family have been great moral support in the course and finishing of this project.

Thanks.



# Appendix A

## Tables

Star	Longitude. (deg).	Latitude. (deg).	Met. (dex).	Met <sub>err</sub> . (dex).	Dist <sub>Sun</sub> . (kpc).	Dist <sub>err</sub> . (kpc).	V <sub>rSun</sub> . (km/s).	V <sub>rGal</sub> . (km/s).	V <sub>rerr</sub> . (km/s).
0	3.0595	61.2956	-1.54	0.35	48.23	6.58	26.4	43.3	22.5
1	5.3649	48.5070	-1.29	0.55	104.37	28.03	109.4	135.3	9.8
2	5.3000	48.5550	-1.41	0.29	52.36	7.34	23.7	49.5	8.6
3	5.8410	48.9100	-1.46	0.26	35.94	4.76	4.1	31.1	8.5
4	6.7570	40.1090	-0.91	0.39	8.87	1.47	-34.3	-1.8	4.8
5	6.8330	40.1749	-1.1	0.25	8.12	0.82	48.9	81.6	5.9
6	6.6199	40.6699	-2.4	0.25	7.51	0.86	72.7	104.6	6.3
7	7.2620	40.3769	-0.98	0.39	7.44	1.12	-8.8	25.1	3.2
8	11.8532	51.9497	-1.29	0.30	22.88	2.75	-137.2	-97.0	6.1
9	17.3533	46.5049	-1.51	0.56	17.78	3.41	116.3	174.3	9.2
10	17.1388	46.7802	-1.56	0.26	38.40	4.39	-54.1	3.1	11.4
11	18.2689	47.2389	-0.90	0.26	6.30	0.78	-21.1	38.6	3.4
12	18.1800	46.8380	-0.93	0.39	3.94	0.46	-16.9	42.9	3.2
13	57.4989	85.9453	-1.48	0.43	6.51	1.05	121.7	142.7	6.9
14	99.3798	-47.5891	-1.60	0.59	33.88	7.37	-129.9	13.5	14.0
15	99.8480	-47.3042	-1.90	0.56	56.74	9.81	-145.1	-1.1	12.9
16	130.7297	-49.5592	-2.86	0.26	32.84	1.66	-213.1	-112.1	26.5
17	136.0932	-39.9921	-1.09	0.26	7.95	0.70	-70.1	39.4	5.1
18	186.0314	-45.5497	-2.71	0.25	23.62	1.78	-18.1	-46.8	17.7
19	185.2599	-45.6510	-1.73	0.29	34.62	5.47	-123.5	-150.0	6.1
20	185.5700	-45.0579	-1.15	0.26	8.99	1.13	-116.1	-143.7	3.5
21	185.0500	-45.2270	-1.10	0.39	24.11	4.01	-37.8	-63.9	5.9
22	185.7135	-45.0197	-1.27	0.27	24.65	3.38	-110.5	-138.5	8.2
23	184.6799	-45.1980	-1.10	0.39	20.40	2.98	56.1	31.0	4.9
24	185.0500	-45.2270	-1.43	0.29	25.80	3.33	5.2	-20.9	11.9
25	223.2323	43.5773	-1.02	0.39	21.08	3.27	-23.2	-135.3	6.9

continued on next page..

Star	Longitude. (deg).	Latitude. (deg).	Met. (dex).	Met <sub>err</sub> . (dex).	Dist <sub>Sun</sub> . (kpc).	Dist <sub>err</sub> . (kpc).	Vr <sub>Sun</sub> . (km/s).	Vr <sub>Gal</sub> . (km/s).	Vr <sub>err</sub> . (km/s).
26	223.4463	43.4385	-1.38	0.27	29.71	3.71	71.8	-41.0	5.5
27	223.4545	42.9744	-2.63	0.26	16.90	1.98	284.9	171.2	5.8
28	232.5257	24.9164	-0.86	0.33	17.24	1.91	-28.4	-193.0	9.5
29	232.2556	24.5973	-0.43	0.40	9.29	2.16	110.2	-54.3	7.8
30	233.8961	32.3214	-1.35	0.33	18.65	1.85	219.2	64.3	9.6
31	233.8318	53.7094	-1.24	0.26	23.46	2.03	9.6	-95.7	7.5
32	234.3959	53.5383	-2.25	0.26	50.26	3.62	221.9	115.4	16.9
33	243.2023	43.4476	-2.82	0.25	102.36	5.50	268.9	124.6	54.7
34	245.3997	63.0568	-1.71	0.27	58.79	8.98	213.2	125.0	46.1
35	252.2822	52.9838	-1.24	0.26	16.72	1.39	137.1	11.9	10.1
36	263.3617	33.1050	-1.28	0.26	27.42	2.66	378.3	193.9	8.0
37	268.0660	36.3806	-1.43	0.43	14.19	1.98	397.7	220.5	16.0
38	268.0987	35.7882	-1.08	0.28	30.18	3.83	382.2	203.6	17.0
39	268.7179	57.6161	-0.89	0.30	24.84	3.14	182.6	68.0	9.6
40	268.6530	57.6818	-2.79	0.26	80.80	4.88	46.3	-68.1	24.0
41	271.9984	69.2651	-1.04	0.28	52.19	7.27	-36.1	-108.9	8.4
42	277.9750	65.3109	-1.37	0.26	29.00	4.23	7.4	-78.6	3.8
43	278.9021	46.8680	-1.09	0.25	11.03	0.80	392.9	247.1	4.9
44	278.7818	46.8193	-1.36	0.27	21.16	2.08	-98.5	-244.5	6.9
45	278.7710	46.8755	-1.44	0.28	26.00	2.77	208.1	62.3	10.1
46	279.8414	36.3392	-1.69	0.26	15.24	1.93	39.3	-133.8	9.5
47	279.7568	36.4641	-2.60	0.25	21.22	2.57	263.9	91.0	7.6
48	279.9107	35.9953	-1.08	0.26	54.67	7.67	195.3	21.4	7.0
49	279.9245	36.0660	-1.95	0.26	38.86	4.77	339.6	165.9	31.9
50	280.6848	59.9341	-1.22	0.31	19.53	1.47	105.1	1.4	18.5

continued from previous page

continued on next page..

Star	Longitude. (deg).	Latitude. (deg).	Met. (dex).	Met <sub>err</sub> . (dex).	Dist <sub>Sun</sub> . (kpc).	Dist <sub>err</sub> . (kpc).	V <sub>rSun</sub> . (km/s).	V <sub>rGal</sub> . (km/s).	V <sub>err</sub> . (km/s).
51	281.4016	59.5485	-1.11	0.25	9.43	1.13	281.2	176.5	3.2
52	282.0216	40.9632	-1.15	0.26	37.92	5.18	120.3	-39.7	4.0
53	283.0020	56.7739	-2.15	0.26	21.60	3.24	-46.5	-159.5	8.7
54	281.9819	56.7220	-0.97	0.26	17.65	2.15	240.5	126.8	2.5
55	282.1380	56.7659	-1.03	0.26	10.32	1.60	125.4	11.9	3.5
56	282.5409	56.9109	-2.07	0.29	20.24	3.29	41.4	-71.4	10.0
57	290.0141	48.9221	-0.81	0.39	19.21	3.75	104.3	-27.1	3.5
58	289.8425	48.5476	-2.09	0.38	24.30	3.37	-11.3	-143.9	15.4
59	292.8320	61.4314	-1.29	0.25	13.00	1.22	-45.3	-136.4	5.0
60	292.9249	72.0907	-1.37	0.25	12.59	0.95	114.3	58.6	5.0
61	293.4002	71.7936	-1.09	0.26	31.23	3.29	15.0	-41.5	5.9
62	301.7750	45.4631	-1.19	0.26	11.74	0.92	266.3	140.9	13.5
63	302.5268	48.9441	-2.04	0.37	64.83	12.03	-43.1	-158.8	23.0
64	302.4345	48.8302	-1.80	0.29	61.39	9.80	27.3	-88.9	6.1
65	305.5008	60.6474	-1.32	0.32	38.72	5.80	173.1	92.4	8.4
66	305.3244	60.5765	-1.43	0.32	16.13	1.82	-39.9	-121.1	13.0
67	304.4910	60.5138	-1.89	0.29	90.64	9.62	30.5	-51.8	9.6
68	304.6944	60.5183	-1.34	0.26	17.00	1.73	255.7	173.6	5.4
69	305.4382	61.3433	-1.46	0.56	13.16	2.73	229.8	150.9	6.0
70	305.2203	61.2388	-2.19	0.38	27.06	3.04	70.8	-8.6	12.1
71	307.2927	80.8958	-1.98	0.71	10.97	2.63	146.1	125.8	17.1
72	322.1151	39.9070	-1.33	0.29	11.51	1.23	329.0	233.6	7.2
73	329.2918	-38.3684	-1.58	0.25	11.27	0.98	256.4	168.5	6.5
74	329.0885	-38.3419	-1.32	0.26	13.93	1.64	-87.1	-175.6	7.8
75	329.4865	-38.0880	-2.04	0.51	26.12	4.74	-115.4	-203.1	19.7

continued on next page..

Star	Longitude. (deg).	Latitude. (deg).	Met. (dex).	Met <sub>err.</sub> (dex).	Dist <sub>sun.</sub> (kpc).	Dist <sub>err.</sub> (kpc).	Vr <sub>sun.</sub> (km/s).	Vr <sub>Gal.</sub> (km/s).	Vr <sub>err.</sub> (km/s).
continued from previous page									
76	328.8904	-37.9206	-0.98	0.39	14.95	3.06	64.3	-25.2	6.8
77	333.4978	46.7542	-1.49	0.28	26.22	3.70	65.7	8.2	10.2
78	333.3410	46.5056	-1.72	0.27	37.16	5.10	168.7	110.5	10.1
79	332.7088	46.8377	-1.39	0.56	20.12	3.72	153.9	94.6	8.1
80	336.9020	52.3950	-0.91	0.28	37.61	5.25	-110.4	-153.0	6.3
81	338.8492	68.2734	-2.19	0.32	23.14	2.24	-9.5	-29.4	10.9
82	341.7229	56.1619	-0.96	0.35	32.20	4.80	69.6	41.5	5.4
83	341.5910	56.5540	-2.64	0.27	72.36	8.01	9.6	-18.4	11.8
84	343.2569	-36.0453	-2.14	0.51	20.07	3.29	-170.2	-219.1	23.9
85	342.5582	56.0200	-1.24	0.27	6.11	0.77	210.0	183.6	4.0
86	344.3840	-43.7165	-1.24	0.37	12.27	1.34	227.1	185.3	8.0
87	344.4641	-43.4882	-1.13	0.27	21.70	1.43	101.3	59.6	7.0
88	344.0347	-43.7415	-1.29	0.26	25.57	2.49	-105.4	-148.2	6.0
89	347.0898	-49.4523	-1.47	0.57	18.80	4.12	191.3	159.5	18.8
90	346.6747	-49.0919	-2.02	0.31	15.21	2.00	-15.9	-49.0	9.8
91	347.6834	53.0556	-2.74	0.28	20.81	1.37	-115.4	-132.6	30.9
92	347.4208	53.3060	-1.47	0.33	15.51	2.00	181.7	164.0	8.1
93	347.2796	53.2974	-2.15	0.33	55.77	7.29	86.3	68.3	10.9
94	350.7265	85.8784	-2.51	0.51	15.75	2.36	78.7	84.0	13.3
95	354.4082	66.3066	-0.84	0.25	13.91	1.75	-17.0	-15.2	4.0
96	356.8100	51.0607	-1.87	0.30	50.66	8.53	18.6	22.6	15.0
97	356.5444	51.1776	-1.29	0.26	51.30	4.46	29.6	33.0	9.2
98	356.7022	51.2281	-2.33	0.31	58.70	3.87	9.2	13.0	4.7
99	356.1505	50.9519	-1.65	0.26	51.95	5.77	49.2	51.6	11.7
100	355.9858	51.1621	-1.32	0.29	16.46	1.63	-104.8	-102.8	4.8
101	355.8893	51.0989	-1.19	0.31	73.12	10.07	4.2	5.9	19.4

Table A.1: Positional, velocity and metallicity information for all giants in the dataset.

Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km/s)
1	2	5.300	48.5550	-1.41	52.36	49.5
1	93	347.280	53.2974	-2.15	55.77	68.3
1	96	356.810	51.0608	-1.87	50.66	22.7
1	97	356.544	51.1776	-1.29	51.30	33.0
1	98	356.702	51.2282	-2.33	58.70	13.0
1	99	356.151	50.9519	-1.65	51.95	51.6
2	11	18.269	47.2390	-0.90	6.30	38.6
2	12	18.180	46.8380	-0.93	3.94	42.9
3	18	186.031	-45.5497	-2.71	23.62	-46.8
3	21	185.050	-45.2270	-1.10	24.11	-63.9
3	24	185.050	-45.2270	-1.43	25.80	-20.9
4	4	6.757	40.1090	-0.91	8.87	-1.8
4	7	7.262	40.3770	-0.98	7.44	25.1
5	5	6.833	40.1750	-1.11	8.12	81.6
5	6	6.620	40.6700	-2.41	7.51	104.6
6	68	304.694	60.5184	-1.34	17.00	173.6
6	69	305.438	61.3434	-1.46	13.16	150.9
7	36	263.362	33.1050	-1.28	27.42	193.9
7	38	268.099	35.7882	-1.08	30.18	203.6
8	59	292.832	61.4314	-1.29	13.00	-136.4
8	66	305.325	60.5766	-1.43	16.13	-121.1

Table A.2: Positional, velocity and metallicity information for all pairs and groups below  $4dist = 0.04$  found with the friends-of-friends method (FOF4).

Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km/s)
1	2	5.300	48.5550	-1.41	52.36	49.5
1	97	356.544	51.1776	-1.29	51.30	33.0
1	99	356.151	50.9519	-1.65	51.95	51.6
2	96	356.810	51.0608	-1.87	50.66	22.6
2	97	356.544	51.1776	-1.29	51.30	33.0
2	98	356.702	51.2282	-2.33	58.70	13.0
3	96	356.810	51.0608	-1.87	50.66	22.6
3	97	356.544	51.1776	-1.29	51.30	33.0
3	99	356.151	50.9519	-1.65	51.95	51.6
4	11	18.269	47.2390	-0.90	6.30	38.6
4	12	18.180	46.8380	-0.93	3.94	42.9
5	18	186.031	-45.5497	-2.71	23.62	-46.8
5	21	185.050	-45.2270	-1.10	24.11	-63.9
6	18	186.031	-45.5497	-2.71	23.62	-46.8
6	24	185.050	-45.2270	-1.43	25.80	-20.9
7	4	6.757	40.1090	-0.91	8.87	-1.8
7	7	7.262	40.3770	-0.98	7.44	25.1
8	5	6.833	40.1750	-1.11	8.12	81.6
8	6	6.620	40.6700	-2.41	7.51	104.6
9	68	304.694	60.5184	-1.34	17.00	173.6
9	69	305.438	61.3434	-1.46	13.16	150.9
10	36	263.362	33.1050	-1.28	27.42	193.9
10	38	268.099	35.7882	-1.08	30.18	203.6
11	59	292.832	61.4314	-1.29	13.00	-136.4
11	66	305.325	60.5766	-1.43	16.13	-121.1
12	93	347.280	53.2974	-2.15	55.77	68.3
12	99	356.151	50.9519	-1.65	51.95	51.6

Table A.3: Positional, velocity and metallicity information for all pairs and groups below  $4dist = 0.04$  found with the all-friends method (AF4).

Table A.4: Positional, velocity and metallicity information for all pairs and groups below  $4dist = 0.08$  found with the friends-of-friends method (FOF8).

Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km/s)
1	0	3.0595	61.2956	-1.54	48.23	43.3
1	2	5.3000	48.5550	-1.41	52.36	49.5
1	3	5.8410	48.9100	-1.46	35.94	31.1
1	10	17.1388	46.7802	-1.56	38.40	3.1
continued on next page..						

continued from previous page						
Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km/s)
1	83	341.591	56.5540	-2.64	72.36	-18.4
1	93	347.280	53.2974	-2.15	55.77	68.3
1	96	356.810	51.0608	-1.87	50.66	22.6
1	97	356.544	51.1776	-1.29	51.30	33.0
1	98	356.702	51.2282	-2.33	58.70	13.0
1	99	356.151	50.9519	-1.65	51.95	51.6
1	101	355.889	51.0989	-1.19	73.12	5.9
2	4	6.7570	40.1090	-0.91	8.87	-1.8
2	5	6.8330	40.1750	-1.11	8.12	81.6
2	6	6.6200	40.6700	-2.41	7.51	104.6
2	7	7.2620	40.3770	-0.98	7.44	25.1
2	11	18.2690	47.2390	-0.90	6.30	38.6
2	12	18.1800	46.8380	-0.93	3.94	42.9
3	18	186.031	-45.5497	-2.71	23.62	-46.8
3	19	185.260	-45.6510	-1.73	34.62	-150.0
3	20	185.570	-45.0580	-1.15	8.99	-143.7
3	21	185.050	-45.2270	-1.10	24.11	-63.9
3	22	185.713	-45.0198	-1.27	24.65	-138.5
3	23	184.680	-45.1980	-1.10	20.40	31.0
3	24	185.050	-45.2270	-1.43	25.80	-20.9
4	36	263.362	33.1050	-1.28	27.42	193.9
4	37	268.066	36.3807	-1.43	14.19	220.5
4	38	268.099	35.7882	-1.08	30.18	203.6
4	43	278.902	46.8681	-1.09	11.03	247.1
4	49	279.925	36.0660	-1.95	38.86	165.9
5	39	268.718	57.6161	-0.89	24.84	68.0
5	42	277.975	65.3110	-1.37	29.00	-78.6
5	45	278.771	46.8755	-1.44	26.00	62.3
5	47	279.757	36.4641	-2.60	21.22	91.0
5	50	280.685	59.9341	-1.22	19.53	1.4
5	51	281.402	59.5485	-1.11	9.43	176.5
5	54	281.982	56.7220	-0.97	17.65	126.8
5	55	282.138	56.7660	-1.03	10.32	11.9
5	56	282.541	56.9110	-2.07	20.24	-71.4
5	57	290.014	48.9222	-0.81	19.21	-27.1
5	61	293.400	71.7936	-1.09	31.23	-41.5
5	68	304.694	60.5184	-1.34	17.00	173.6
5	69	305.438	61.3434	-1.46	13.16	150.9
continued on next page..						

continued from previous page						
Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km/s)
5	70	305.220	61.2388	-2.19	27.06	-8.6
6	8	11.8532	51.9498	-1.29	22.88	-97.0
6	80	336.902	52.3950	-0.91	37.61	-153.0
6	91	347.683	53.0557	-2.74	20.81	-132.6
6	100	355.986	51.1621	-1.32	16.46	-102.8
7	53	283.002	56.7740	-2.15	21.60	-159.5
7	58	289.842	48.5477	-2.09	24.30	-143.9
7	59	292.832	61.4314	-1.29	13.00	-136.4
7	66	305.325	60.5766	-1.43	16.13	-121.1
8	73	329.292	-38.3684	-1.58	11.27	168.5
8	86	344.384	-43.7166	-1.24	12.27	185.3
8	89	347.090	-49.4523	-1.47	18.80	159.5
9	13	57.4989	85.9454	-1.48	6.510	142.7
9	71	307.293	80.8959	-1.98	10.97	125.8
9	94	350.727	85.8784	-2.51	15.75	84.0
10	74	329.089	-38.3419	-1.32	13.93	-175.6
10	75	329.487	-38.0880	-2.04	26.12	-203.1
10	84	343.257	-36.0453	-2.14	20.07	-219.1
11	25	223.232	43.5773	-1.02	21.08	-135.2
11	31	233.832	53.7094	-1.24	23.46	-95.7
12	32	234.396	53.5384	-2.25	50.26	115.4
12	34	245.400	63.0568	-1.71	58.79	125.0
13	63	302.527	48.9442	-2.04	64.83	-158.8
13	64	302.435	48.8302	-1.80	61.39	-88.9
14	77	333.498	46.7542	-1.49	26.22	8.2
14	82	341.723	56.1620	-0.96	32.20	41.5
15	78	333.341	46.5056	-1.72	37.16	110.5
15	79	332.709	46.8378	-1.39	20.12	94.6
16	81	338.849	68.2735	-2.19	23.14	-29.4
16	95	354.408	66.3066	-0.84	13.91	-15.2
17	85	342.558	56.0200	-1.24	6.11	183.6
17	92	347.421	53.3061	-1.47	15.51	164.0

Table A.5: Positional, velocity and metallicity information for all pairs and groups below  $4dist = 0.08$  found with the all-friends method (AF8).

Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km/s)
1	0	3.0595	61.2956	-1.54	48.23	43.3
1	2	5.3000	48.5550	-1.41	52.36	49.5
1	93	347.280	53.2974	-2.15	55.77	68.3
1	96	356.810	51.0608	-1.87	50.66	22.6
1	97	356.544	51.1776	-1.29	51.30	33.0
1	98	356.702	51.2282	-2.33	58.70	13.0
1	99	356.151	50.9519	-1.65	51.95	51.6
2	2	5.30000	48.5550	-1.41	52.36	49.5
2	3	5.84100	48.9100	-1.46	35.94	31.1
2	96	356.810	51.0608	-1.87	50.66	22.6
2	97	356.544	51.1776	-1.29	51.30	33.0
2	99	356.151	50.9519	-1.65	51.95	51.6
3	4	6.75700	40.1090	-0.91	8.87	-1.8
3	7	7.26200	40.3770	-0.98	7.44	25.1
3	11	18.2690	47.2390	-0.90	6.30	38.6
3	12	18.1800	46.8380	-0.93	3.94	42.9
4	5	6.83300	40.1750	-1.11	8.12	81.6
4	7	7.26200	40.3770	-0.98	7.44	25.1
4	11	18.2690	47.2390	-0.90	6.30	38.6
4	12	18.1800	46.8380	-0.93	3.94	42.9
2	5	6.83300	40.1750	-1.11	8.12	81.6
5	6	6.62000	40.6700	-2.41	7.51	104.6
5	7	7.26200	40.3770	-0.98	7.44	25.1
6	13	57.4989	85.9454	-1.48	6.51	142.7
6	71	307.293	80.8959	-1.98	10.97	125.8
6	94	350.727	85.8784	-2.51	15.75	84.0
7	18	186.031	-45.5497	-2.71	23.62	-46.8
7	21	185.050	-45.2270	-1.10	24.11	-63.9
7	24	185.050	-45.2270	-1.43	25.80	-20.9
8	18	186.031	-45.5497	-2.71	23.62	-46.8
8	23	184.680	-45.1980	-1.10	20.40	31.0
8	24	185.050	-45.2270	-1.43	25.80	-20.9
9	36	263.362	33.1050	-1.28	27.42	193.9
9	37	268.066	36.3807	-1.43	14.19	220.5
9	38	268.099	35.7882	-1.08	30.18	203.6
continued on next page..						

continued from previous page						
Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	Vr <sub>Gal</sub> (km/s)
10	50	280.685	59.9341	-1.22	19.53	1.4
10	55	282.138	56.7660	-1.03	10.32	11.9
10	57	290.014	48.9222	-0.81	19.21	-27.1
11	50	280.685	59.9341	-1.22	19.53	1.4
11	56	282.541	56.9110	-2.07	20.24	-71.4
11	57	290.014	48.9222	-0.81	19.21	-27.1
12	51	281.402	59.5485	-1.11	9.43	176.5
12	54	281.982	56.7220	-0.97	17.65	126.8
12	69	305.438	61.3434	-1.46	13.16	150.9
13	51	281.402	59.5485	-1.11	9.43	176.5
13	68	304.694	60.5184	-1.34	17.00	173.6
13	69	305.438	61.3434	-1.46	13.16	150.9
14	3	5.84100	48.9100	-1.46	35.94	31.1
14	10	17.1388	46.7802	-1.56	38.40	3.1
15	8	11.8532	51.9498	-1.29	22.88	-97.0
15	100	355.986	51.1621	-1.32	16.46	-102.8
16	19	185.260	-45.6510	-1.73	34.62	-150.0
16	22	185.713	-45.0198	-1.27	24.65	-138.5
17	20	185.570	-45.0580	-1.15	8.99	-143.7
17	22	185.713	-45.0198	-1.27	24.65	-138.5
18	21	185.050	-45.2270	-1.10	24.11	-63.9
18	22	185.713	-45.0198	-1.27	24.65	-138.5
19	25	223.232	43.5773	-1.02	21.08	-135.3
19	31	233.832	53.7094	-1.24	23.46	-95.7
20	32	234.396	53.5384	-2.25	50.26	115.4
20	34	245.400	63.0568	-1.71	58.79	125.0
21	37	268.066	36.3807	-1.43	14.19	220.5
21	43	278.902	46.8681	-1.09	11.03	247.1
22	38	268.099	35.7882	-1.08	30.18	203.6
22	49	279.925	36.0660	-1.95	38.86	165.9
23	39	268.718	57.6161	-0.89	24.84	68.0
23	45	278.771	46.8755	-1.44	26.00	62.3
24	39	268.718	57.6161	-0.89	24.84	68.0
24	50	280.685	59.9341	-1.22	19.53	1.4
25	42	277.975	65.3110	-1.37	29.00	-78.6
25	56	282.541	56.9110	-2.07	20.24	-71.4
26	42	277.975	65.3110	-1.37	29.00	-78.6
26	61	293.400	71.7936	-1.09	31.23	-41.5
continued on next page..						

continued from previous page						
Group	Star	Longitude (deg)	Latitude (deg)	Metallicity (dex)	Distance <sub>Sun</sub> (kpc)	V <sub>rGal</sub> (km/s)
27	45	278.771	46.8755	-1.44	26.00	62.3
27	47	279.757	36.4641	-2.60	21.22	91.0
28	50	280.685	59.9341	-1.22	19.53	1.4
28	70	305.220	61.2388	-2.19	27.06	-8.6
29	53	283.002	56.7740	-2.15	21.60	-159.5
29	58	289.842	48.5477	-2.09	24.30	-143.9
30	53	283.002	56.7740	-2.15	21.60	-159.5
30	59	292.832	61.4314	-1.29	13.00	-136.4
31	59	292.832	61.4314	-1.29	13.00	-136.4
31	66	305.325	60.5766	-1.43	16.13	-121.1
32	61	293.400	71.7936	-1.09	31.23	-41.5
32	70	305.220	61.2388	-2.19	27.06	-8.6
33	63	302.527	48.9442	-2.04	64.83	-158.8
33	64	302.435	48.8302	-1.80	61.39	-88.9
34	73	329.292	-38.3684	-1.58	11.27	168.5
34	86	344.384	-43.7166	-1.24	12.27	185.3
35	74	329.089	-38.3419	-1.32	13.93	-175.6
35	75	329.487	-38.0880	-2.04	26.12	-203.1
36	75	329.487	-38.0880	-2.04	26.12	-203.1
36	84	343.257	-36.0453	-2.14	20.07	-219.1
37	77	333.498	46.7542	-1.49	26.22	8.2
37	82	341.723	56.1620	-0.96	32.20	41.5
38	78	333.341	46.5056	-1.72	37.16	110.5
38	79	332.709	46.8378	-1.39	20.12	94.6
39	80	336.902	52.3950	-0.91	37.61	-153.0
39	91	347.683	53.0557	-2.74	20.81	-132.6
40	81	338.849	68.2735	-2.19	23.14	-29.4
40	95	354.408	66.3066	-0.84	13.91	-15.2
41	83	341.591	56.5540	-2.64	72.36	-18.4
41	101	355.889	51.0989	-1.19	73.12	5.9
42	85	342.558	56.0200	-1.24	6.110	183.6
42	92	347.421	53.3061	-1.47	15.51	164.0
43	86	344.384	-43.7166	-1.24	12.27	185.3
43	89	347.090	-49.4523	-1.47	18.80	159.5
44	91	347.683	53.0557	-2.74	20.81	-132.6
44	100	355.986	51.1621	-1.32	16.46	-102.8

# Bibliography

- Armandroff, T.: 1989, *AJ* **97**(2), 375
- Battaglia, G., Helmi, A., Morrison, H., Harding, P., Olszewski, E., Mateo, M., Freeman, K., Norris, J., and Shectman, S.: 2005, *Mon. Not. R. Soc. Astron.* **364**, 433
- Bell, E. e. a.: 2007, *The accretion origin of the Milky Way's stellar halo*, To appear in *ApJ* (astro-ph/07060004)
- Bellazzini, M., Newberg, H., Correnti, M., Ferraro, F., and Monaco, L.: 2006, *Astron. Astrophys.* **457**(2), L21
- Belokurov, V. e. a.: 2006, *ApJ* **642**, L137
- Belokurov, V. e. a.: 2007, *ApJ* **658**, 337
- Binney, J., Gerhard, O., and Spergel, D.: 1997, *Mon. Not. R. Astron. Soc.* **288**, 365
- Binney, J. and Merrifield, M.: 1998, *Galactic Astronomy*, Princeton Series in Astrophysics, Princeton University Press, 1st edition
- Binney, J. and Tremaine, S.: 1994, *Galactic Dynamics*, Princeton Series in Astrophysics, Princeton University Press, 3rd edition
- Bullock, J. S. and Johnston, K. V.: 2005, *Tracing Galaxy formation with stellar halos I: methods*, Draft version September 9, 2005
- Bullock, James, S. and Johnston, K. V.: 2004, *Astronomical Society of the Pacific* **327**, 80
- Chiba, M. and Beers, T.: 2000, *AJ* **119**, 2843
- Da Costa, G. and Armandroff, T.: 1990, *AJ* **100**, 162
- Dohm-Palmer, R., Helmi, A., Morrison, H., Mateo, M., Olszewski, E., P., H., Freeman, C., Norris, J., and Shectman, S.: 2001, *ApJ* **555**, L37
- Dohm-Palmer, R., Mateo, M., Olszewski, E., Morrison, H., Harding, P., Freeman, K., and Norris, J.: 2000, *AJ* **120**(3), 2496
- Duffau, S., Zinn, R., Vivas, A., Mendez, R., Carraro, G., Winnick, R., and Gallart, C.: 2006, in *RevMexAA (Serie de Conferencias)*, Vol. 26, pp 70–71
- Edvardsson, B., Andersen, J., Gustafsson, B., Lambert, D., Nissen, P., and Tomkin, J.: 1993, *Astron. Astrophys.* **275**, 101
- Eggen, O., Lynden-Bell, D., and Sandage, A.: 1962, *ApJ* **136**, 748
- Fellhauer, M., Evans, N., Belokurov, V., Z., Yanny, B., Wilkinson, M., Gilmore, G., Irwin, M., Bramich, D., Vidrih, S., Hewett, P., and Beers,

- T.: 2007, *Mon. Not. R. Astron. Soc.* **375**(4), 1171
- Freeman, K. and Bland-Hawthorn, J.: 2002, *Annu. Rev. Astron. Astrophys.* **40**, 487
- Grillmair, C. and Dionatos, O.: 2006, *ApJ* **643**, L17
- Grillmair, C. and Johnson, R.: 2006, *ApJ* **639**(1), L17
- Harding, P., Morrison, H. L., Olszewski, E. W., Arabadjis, J., Mateo, M., Dohm-Palmer, R., Freeman, K. C., and Norris, J.: 2001, *AJ* **122**, 1397
- Helmi, A.: 2000, *Ph.D. thesis*, Universiteit Leiden
- Helmi, A.: 2004, *Mon. Not. R. Astron. Soc.* **351**, 643
- Helmi, A. and White, S.: 1999, *Mon. Not. R. Astron. Soc.* **307**, 495
- Helmi, A., White, S., de Zeeuw, P., and H., Z.: 1999, *Nature* **402**, 53
- Ibata, R., Gilmore, G., and Irwin, M.: 1994, *Nature* **370**, 194
- Ibata, R., Irwin, M., Lewis, G., Ferguson, A., and Tanvir, N.: 2001, *Nature* **412**, 49
- Irwin, M. and Hatzidimitriou, D.: 1995, *MNRAS* **277**(4), 1354
- Johnston, K. V., Hernquist, L., and Bolte, M.: 1996, *ApJ* **465**, 278
- Juric, M. e. a.: 2005, *The Milky Way tomography with SDSS*, To appear in *ApJ* (astro-ph/0510520)
- Kahn, F. and Woltjer, L.: 1959, *ApJ* **130**(3), 705
- Kraft, R. and Ivans, I.: 2003, *Publications of the Astronomical Society of the Pacific* **115**, 143
- Lee, Y., Demarque, P., and Zinn, R.: 1990, *AJ* **350**(1), 155
- Lynden-Bell, D. and Lynden-Bell, R.: 1995, *Mon. Not. R. Astron. Soc.* **275**, 429
- Majewski, S.: 1992, *ApJS* **78**, 87
- Martinez-Delgado, D., Aparicio, A., Gomez-Flechoso, M., and Carrera, R.: 2001, *ApJ* **549**(2), L199
- Martinez-Delgado, D., Penarrubia, J., Juric, M., Alfaro, E., and Ivezić, Z.: 2007, *The Virgo Stellar Over-Density: Mapping the infall of the Sagittarius tidal stream onto the Milky Way disk*, astro-ph/0609104
- Mathewson, D., Cleary, M., and Murray, J.: 1974, *ApJ* **195**, 97
- McWilliam, A. and Rich, R.: 1994, *ApJS* **91**, 749
- Minniti, D.: 1996, *ApJ* **459**, 175
- Morrison, H.: 1993, *AJ* **106**(2), 578
- Morrison, H., Mateo, M., Olszewski, E., Harding, P., Dohm-Palmer, R., Freeman, K., Norris, J., and Morita, M.: 2000, *AJ* **119**, 2254
- Morrison, H., Olszewski, E., Mateo, M., Norris, J., Harding, P., Dohm-Palmer, R., and Freeman, K.: 2001, *AJ* **121**(1), 283
- Morrison, H. L., Norris, J., Mateo, M., Harding, P., Olszewski, E. W., Shectman, S. A., Dohm-Palmer, R., Helmi, A., and Freeman, K. C.: 2003, *AJ* **125**, 2502
- Newberg, H., Yanny, B., Cole, N., Beers, T., Fiorentin, P., Schneider, D., and Wilhelm, R.: 2007, *The Overdensity in Virgo, Sagittarius Debris and the Asymmetric Spheroid*, astro-ph/07063391

- Newberg, H. e. a.: 2002, *ApJ* **569**(1), 245
- Norris, J. E.: 1999, *Astrophysics and Space Science* **265**, 213
- Odenkirchen, M., Grebel, E., Dehnen, W., Rix, H., Yanny, B., Newberg, H., Rockosi, C., and Martinez-Delgado, D.: 2003, *AJ* **126**(5), 2385
- Palma, C., Majewski, S. R., and Johnston, K. V.: 2002, *ApJ* **564**, 736
- Paust, N., Chaboyer, B., and Sarajedini, A.: 2007, *AJ* **133**(6), 2787
- Peebles, P.: 1982, *ApJ* **263**, 1
- Press, W. e. a.: 1988, *Numerical Recipes in C, The art of scientific computing, Second Edition*, Cambridge University Press
- Rich, R.: 2001, in von Hippel et al. (ed.), *Astrophysical Ages and Timescales*, Vol. TBD of *ASP Conference Series*, pp 216–225
- Salaris, M. and Weiss, A.: 1997, *A&A* **327**, 107
- Searle, L. and Zinn, R.: 1978, *ApJ* **225**, 357
- Shang, Z. e. a.: 1998, *ApJ Letters* **504**, L23
- Sirko, E., Goodman, J., Knapp, G., Brinkmann, J., Ivezić, Z., Knerr, E., Schlegel, D., Schneider, D., and York, D.: 2004, *AJ* **127**, 899
- Tolstoy, E., Irwin, M., Helmi, A., Battaglia, G., Jablonka, P., Hill, V., Venn, K., Shetrone, M., Letarte, B., Cole, A., Primas, F., Francois, P., Armito, N., Sadakane, K., Kaufer, A., Szeifert, T., and Abel, T.: 2004, *ApJ* **617**, L119
- Tonry, J. and Davis, M.: 1979, *AJ* **84**, 1511
- Unavane, M., Wyse, R., and Gilmore, G.: 1996, *Mon. Not. R. Astron. Soc.* **278**, 727
- Velazquez, H. and White, S.: 1995, *Mon. Not. R. Astron. Soc.* **275**(3), L23
- Venn, K. A., Irwin, M., Shetrone, M. D., Tout, C. A., Hill, V., and Tolstoy, E.: 2004, *AJ* **128**(3), 1177
- Weiland, J. e. a.: 1994, *ApJ* **425**(2), L81
- Zaritsky, D.: 1999, in B. Gibson, T. Axelrod, and M. Putnam (eds.), *The third Stromlo Symposium: the Galactic Halo*, Vol. 165 of *ASP Conference Series*, pp 34–45
- Zinn, R.: 1985, *ApJ* **293**, 424
- Zwicky, F.: 1933, *Helvetica Physica Acta* **6**, 110