

Re-examining the dynamics of the Milky Way with new distance estimates

John Wimarsson

Lund Observatory
Lund University



2018-EXA137

Degree project of 15 higher education credits
June 2018

Supervisor: Paul McMillan

Lund Observatory
Box 43
SE-221 00 Lund
Sweden

Abstract

The use of reliable distance estimates is key to derive accurate three-dimensional velocities of celestial objects. Systematic overestimation and large uncertainties in the distances will thus provide misleading information of an object's kinematics. In turn, astronomers will get a flawed understanding of dynamical properties of substructures in the Milky Way.

In this paper, we employ the improved distance estimates of observations common to the TGAS and RAVE catalogues, made available in [McMillan et al. \(2018\)](#), to re-investigate claims of three previous papers on Milky Way dynamics where lower quality distance estimates have been used. We opt for a general approach where we use analysis in the form of statistical resampling as well as a velocity model of Gaussian nature adopted from [Kordopatis et al. \(2013\)](#) to evaluate the validity of the claims. Further, we use at least two different subsets of our data with varying quality in each investigation to gain a sense of how lower quality data points affect the observed dynamical phenomena.

We find that there are results in each paper that can be attributed to the use of poor distance estimates. Firstly, we find that there are no indications of the existence of the dearth of zero angular momentum stars in the Milky Way, which was proposed in [Hunt et al. \(2016\)](#). Secondly, we find that the excess of retrograde halo orbits in the Milky Way found in [Helmi et al. \(2017\)](#) is a product of the systematic overestimation of distances. Our results indicate that the fraction of retrograde halo orbits is substantially lower than they have claimed. Finally, we find that the contraction-like behaviour of the Galactic disc detected in [Williams et al. \(2013\)](#) can be connected to systematic overestimation and/or large uncertainties in their distance estimates.

Acknowledgements

I would like to express my gratitude to my supervisor, Paul McMillan, who has provided me with a lot of support, knowledge and advice during the course of this project. The relaxed atmosphere during our meetings has greatly contributed to an overall enjoyable and interesting experience. I also would like to thank Daniel Mikkola for being available to answer any questions I had and for proofreading my thesis. Lastly, I want to thank Anahi Carrasco and my siblings for their continued encouragement.

Populärvetenskaplig beskrivning

Solen är bara en utav miljardtals stjärnor i Vintergatan - vår hemgalax. Likt en myrstack kan Vintergatan som helhet vid en första anblick verka simpel. Men när man tar en närmare titt upptäcker man att varje invånare har sitt eget liv, vilket bidrar till en komplex natur. Därför blir det svårt att få information om Vintergatans omfattande egenskaper genom att undersöka enskilda stjärnor. Men precis som att vi lättare kan förstå en arbetsmyras rörelser genom att analysera andra arbetsmyror, så kan vi kategorisera stjärnor beroende på vart i Vintergatan de befinner sig och kartlägga hur stjärnorna i varje del rör sig som grupp.

Ett stort hinder i undersökningar av storskaliga rörelser i Vintergatans olika delar är bristen på noggranna avståndsbedömningar. Ett överskattat avstånd leder till en alltför stor observerad hastighet hos enskilda stjärnor. Därmed leder systematiskt överskattade avstånd till en felaktig bild av hur stjärnorna rör sig som grupp. Orsaken till att avstånd tenderar till att bli felaktigt bedömda kan tillskrivas brister hos de instrument vi använder för att mäta dem, samt otillräcklig behandling av dessa mätningar. De enorma distanserna mellan solen och andra stjärnor sträcker sig upp mot tiotusentals ljusår, vilket leder till att vi måste ha teknologiskt avancerade instrument för att kunna mäta dem tillräckligt bra.

Den 19 december 2013 skickades satelliten *Gaia* upp i rymden, ett datum som kom att förändra forskning om Vintergatan för all framtid. Denna satellit har de mest avancerade mätinstrumenten för kartläggning av stjärnor hittills. Den kommer under sitt uppdrag mäta positioner och rörelser för över en miljard stjärnor i Vintergatan. Faktum är att detta motsvarar 10 000 gånger fler stjärnor än vad som mättes av *Gaias* föregångare, satelliten *Hipparcos*. Dessutom kommer dessa mätningar att vara 200 gånger mer noggranna.

Galaxforskaren Paul McMillan vid Lunds Universitet har använt data från *Gaia* för att beräkna avståndsbedömningar för 219 565 av dessa stjärnor, vilka är betydligt mer noggranna än de som tidigare funnits tillgängliga för forskare. I vårt projekt använder vi dessa avstånd för att ta reda på om faktumet att man i tidigare forskning har använt sig av mindre noggranna avståndsbedömningar har lett till en förvrängd bild av olika storskaliga rörelsemönster i Vintergatans delar.

Med de bättre avståndsbedömningarna kan vi först och främst ta reda på om vi kan detektera samma rörelsemönster som föreslagits i artiklarna. Vi kan därefter använda oss av noggranna statistiska tester för undersöka hur de sämre avståndsbedömningarna påverkat de olika rörelsemönstrena. I slutändan är målet med denna undersökning att ge oss en bättre förståelse för olika typer av storskaliga rörelser i Vintergatan samt att påvisa hur viktigt det är att använda sig av noggrannt analyserad data för att få en sanningsenlig bild av vår galax.

Contents

1	Introduction	7
1.1	Background	8
1.2	The TGAS & RAVE dataset	8
1.2.1	Obtaining the distance estimates	10
1.2.2	Known issues with TGAS & RAVE	10
1.3	The Milky Way	12
1.3.1	The Galactic bulge	12
1.3.2	The Galactic disc	12
1.3.3	The Galactic halo	13
1.3.4	Coordinate systems	14
1.4	Three dynamical phenomena	16
2	Treating the data	18
2.0.1	Numerical values	18
2.0.2	TGAS & RAVE subsets	19
2.1	Statistical tests	20
2.1.1	The bootstrap	20
2.1.2	Uncertainty sampling	20
2.2	Velocity model	21
3	Dearth of stars	23
3.1	Method	24
3.1.1	Resampling the TGAS & RAVE subsets	25
3.1.2	Simulating a dearth of stars	25
3.2	Results	26
3.2.1	Sampling tests	26
3.2.2	Velocity model simulation	28
3.3	Discussion	30
4	Counter-rotating halo	33
4.1	Method	35
4.1.1	Galactic potential and angular momentum	36

4.1.2	Analyzing the halo sample	37
4.2	Results and analysis	38
4.2.1	Studying the default sample	39
4.3	Discussion	43
5	The wobbly disc	45
5.1	Method	46
5.1.1	Computing the velocity field	47
5.1.2	Including the velocity model	47
5.2	Results and analysis	48
5.2.1	The median v_R field	48
5.2.2	The median v_z field	50
5.2.3	Systematic distance shifts	52
5.3	Discussion	53
6	Conclusion	56
A	Re-investigation of modelled dip for higher quality data	61
B	Counter-rotating halo for <code>flag.any-lowlogg</code> sample	63
C	Standard deviations for scaled velocity field plots	65

List of Figures

1.1	The 219 565 stars common to the TGAS and RAVE datasets plotted in a Galactic frame of reference. The horizontal coordinate is the Galactic longitude l and the vertical is the Galactic latitude b	9
2.1	A histogram displaying the mean of 100 pseudo velocity distributions obtained using the model from Kordopatis et al. (2013) compared to the computed velocities for the default subset of the TGAS & RAVE data. The black bars indicate the standard deviation in bin counts for the computed pseudo-samples.	22
3.1	Histogram of the v_y distribution for stars in the solar neighbourhood which has been adapted from figure 1 in Hunt et al. (2016) . The dearth of zero angular momentum stars is centred around $v_y = -v_\odot$, which is indicated with an arrow. Reprinted with permission.	24
3.2	Histograms of the tangential velocity of the sample of stars common to the TGAS & RAVE catalogues in a Galactocentric frame of reference with bin widths of 10 km s^{-1} . We have chosen to only show the distributions within the range $-400 \leq v_\phi \leq 400 \text{ km s}^{-1}$. The left plot is the default sample with only duplicate stars removed. The right plot shows the <code>flag_any</code> with all known low-quality measurements removed.	27
3.3	The tangential velocity distributions for the default sample with bin widths of 10 km s^{-1} and the mean of the bootstrap samples. The standard deviation for each bin is shown by the black bars. The result from resampling of the uncertainties is to the left while the bootstrap resampling result is to the right. No dip in the bin counts can be seen.	28
3.4	The same plots provided in figure 3.3 but zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. We do not see any indication of a dip in counts.	28
3.5	The tangential velocity distributions for the sample with <code>flag_any=0</code> and the mean of the bootstrap samples computed. The bins have widths of 10 km s^{-1} . The standard deviation for each bin is shown by the black bars. The result from the resampling of the uncertainties is shown in the left plot while the bootstrap resampling result is to the right.	29

3.6	The same plots provided in figure 3.5 but zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. In agreement with figure 3.4 there is no indication of a dearth.	29
3.7	Histograms of the distribution of tangential velocities for the default sample data points when the velocities have been taken from the model described in section 2.2. For each case of λ , the left plot shows the major part of the distribution while the right is zoomed in around $v_\phi = 0 \text{ km s}^{-1}$	31
3.8	The same plots as in figure 3.7 but the simulated dearth now has a width of 60 km s^{-1} which would correspond to the findings of Hunt et al. (2016)	32
4.1	A simple diagram displaying the qualitative effect that distance overestimation has on the apparent motion of a stationary star with respect to the Sun.	34
4.2	The distributions of energy vs. angular momenta for our three subsets of data. The left plot displays the distribution for the default sample, while the middle plot shows the same quantities but for the <code>flag_any-lowlogg</code> sample. The rightmost plot shows $L_z - E$ space for the <code>flag_any</code> subset.	38
4.3	A visualisation of how the cuts were introduced in order to obtain the fractions of stars with negative angular momenta x . The given fraction in the figure was computed for the <code>flag_dup</code> sample using the uncertainty resampling method.	39
4.4	On the left, the distribution of E vs. L_z for the default sample is displayed. The right-hand plot contain the corresponding distribution for a pseudo-sample based on the same data points.	41
4.5	Four different cases where the distance estimates of the modelled pseudo-samples based on the <code>flag_dup</code> sample have been scaled by a factor k	42
5.1	An example median v_R distribution for a pseudo-sample created from the velocity model. It has been based on the observations and uncertainties of the default sample.	48
5.2	To the upper left is a map of the median velocity field in R for the default sample divided into $(0.2 \text{ kpc})^2$ bins. The upper middle plot displays the mean median velocities for 100 pseudo-samples created from the velocity model. The upper right plot displays the standard deviations of the mean modelled median v_R velocities. The bottom row shows the same plots but for the <code>flag_any</code> sample.	49
5.3	Same as in figure 5.2 but for v_z	51
5.4	Four different cases of mean modelled median velocity fields in z that display the impact of overestimated distances. Each plot contains a computed mean model median v_z field for distances scaled by a factor k	53
5.5	Same as in figure 5.4 but for the v_R component.	54

A.1	The results from computing the tangential velocity using the model with bin widths of 10 km s^{-1} , zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. The introduced dips in the plots are 10 km s^{-1} and 20 km s^{-1} , from left to right respectively.	62
A.2	The results from computing the tangential velocity using the model with a simulated dip of width 10 km s^{-1} , zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. The bin widths of the plots are 5 km s^{-1} and 10 km s^{-1} , from left to right respectively.	62
C.1	The standard deviations for the four cases displayed in figure 5.4	65
C.2	The standard deviations for the four cases displayed in figure 5.5	66

List of Tables

1.1	The quality flags that are included in the TGAS & RAVE dataset. The description explains what observations are removed when each flag is raised, i.e. set equal to 1. N_{stars} tells us how many of the TGAS & RAVE observations that fulfil the criteria.	11
1.2	Relevant variables and corresponding units for the ICRS frame with a spherical coordinate representation.	15
1.3	Relevant variables and corresponding units for the Galactocentric frame with a cylindrical coordinate representation.	15
2.1	The name of reference and description of the three datasets used during our investigations of the phenomena mentioned in this paper.	19
2.2	The chosen parameters for the modelled Gaussian velocity distributions of each substructure in the Milky Way.	21
4.1	The fractions of halo stars from the default dataset that have negative and positive angular momenta, found using the uncertainty sampling method.	40
4.2	The fractions of halo stars from the default dataset with modelled velocities that have negative and positive angular momenta. N.B. the model implies equal probability of positive and negative L_z	40
4.3	The fractions of default sample halo stars with modelled velocities that have negative and positive angular momenta. However, the distances have now been scaled by a factor k to see how the distribution changes when we include systematic errors.	43
B.1	The fractions of halo stars from the <code>flag_any-lowlogg</code> dataset that have negative and positive angular momenta, found using the uncertainty sampling method.	63
B.2	The fractions of halo stars from the <code>flag_any-lowlogg</code> subset with modelled velocities that have negative and positive angular momenta.	64
B.3	The fractions of halo stars from the <code>flag_any-lowlogg</code> subset with modelled velocities that have negative and positive angular momenta. However, the distances have now been scaled by a factor k to see how the distribution changes when we include systematic errors.	64

Chapter 1

Introduction

Acquiring precise measurements of celestial objects is one of the main challenges of Milky Way research. Especially in the field of astrometry, where the relative positions and motions of celestial bodies are studied. Stars mostly emit light in the visible spectrum, which means that the atmosphere on Earth limits the data quality attainable with ground-based measurements, as its particles interact with photons in the corresponding waveband. Astronomers therefore benefit substantially from the ability to launch sophisticated astrometric instruments into space, attached to satellites.

Given the high pace of technological advancement in society, there is also a steady increase in the quality of such tools. Better instruments yield more reliable data, which in turn allows astronomers to draw conclusions with higher confidence from the results of their research. Information gained using accurate astrometric instruments can subsequently be used to estimate distances between the Sun and the detected celestial objects. Knowing the distance to a celestial body is not only the key to determine where the object is, but also how fast it is moving and in what direction ([Perryman 2012](#)).

The methods involved in astrometry deal with large error margins that grow with distance. Computing positions of stellar objects is a complicated process, even for stars within the solar neighbourhood¹. Consequently, distance estimates benefit notably from more reliable data. Another aspect is the handling of information gathered from such observations. With continuous research involving data treatment comes constant improvement upon existing techniques. Hence, by combining datasets from more precise instruments with refined data treatment, it is possible to significantly reduce uncertainties in distance estimates, as shown by [McMillan et al. \(2018\)](#).

In this paper, we review the results of three previous studies of dynamical phenomena in the Milky Way, where less certain distance estimates have been employed. By re-examining the phenomena using the new improved distance estimates of [McMillan et al. \(2018\)](#), we are

¹Stars with distances up to a few kpc from the Sun.

able to reach a better understanding of the large-scale motions in the Milky Way and the influence that poor distance estimates have on different types of dynamical studies.

1.1 Background

The launch of the satellite *Gaia* in 2013 was the start of a new era in Milky Way research (Gaia Collaboration et al. 2016b). This currently operating spacecraft the most advanced tool for astrometric measurements available to galactic astronomers in present day and its mission results are expected to considerably aid research in various astrophysical fields. The primary objective of *Gaia* is to analyse the Galaxy by measuring physical properties, such as position and motion with respect to the Sun, of more than 1 billion stars². Furthermore, the satellite will provide astrometric data for stars that are fainter and farther away than any instrument before it. In contrast, *Gaia*'s predecessor *Hipparcos* that was launched in 1989 and conducted measurements until 1993 merely catalogued 100 000 stars with high precision. When its mission concludes, *Gaia* will have collected 200 times more accurate data from 10 000 times as many stars.

Despite that the final results of the mission are expected to be available after 2020, researchers have already found *Gaia* to be a potent source of information. This is due to the early release of results during late 2016 in *Gaia* Data Release 1 (GDR1, Gaia Collaboration et al. 2016a). GDR1 contains measurements of 2 million sources that already have been detected during previous survey missions such as HIPPARCOS. The latter has provided two main catalogues with astrometric data, HIPPARCOS which contains the high precision data and *Tycho*, which is larger but has lesser precision. By combining the data from the *Tycho-2* and HIPPARCOS catalogues with the new data from GDR1, Michalik, Lindegren, & Hobbs (2015) created the *Tycho-Gaia* astrometric solution (TGAS). TGAS contains 5-parameter³ astrometric data for 2 million Milky Way stars.

1.2 The TGAS & RAVE dataset

The potential of the TGAS catalogue truly shows when combined with measurements obtained from the Radial Velocity Experiment (RAVE). RAVE is a ground-based project that utilises a different data gathering method than *Gaia* (Steinmetz 2003) and therefore acquires other physical quantities. Hence, the two datasets complement each other in a way that allows astronomers to gain information about observations that appear in both TGAS and RAVE tables. TGAS provides the astrometric parameters: positions, parallaxes and proper motions, while RAVE contains the surface gravity values $\log g$, metallicities and

²When we use the word “stars” in this paper, we mainly refer to observed objects by astrometric instruments that are believed to be possible stars.

³Positions, parallaxes and proper motions

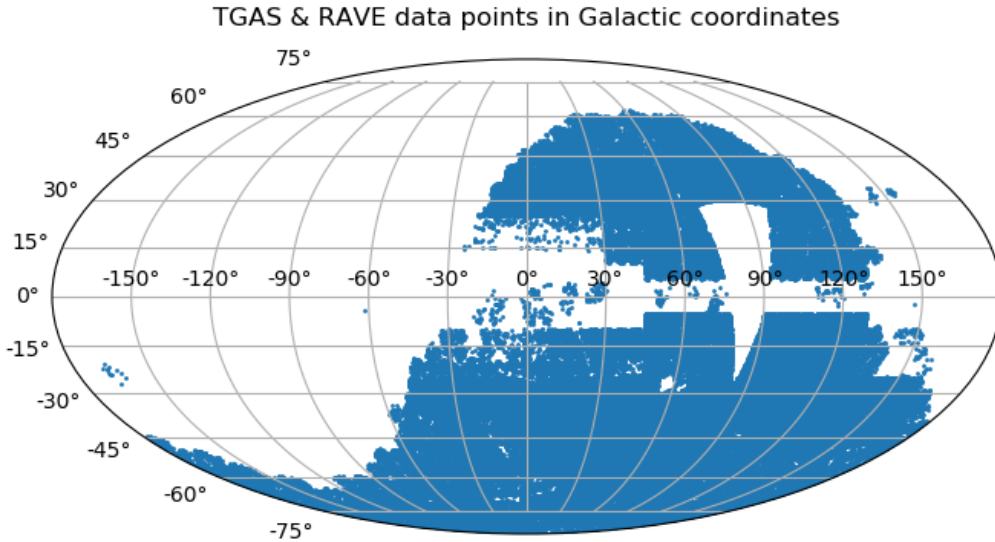


Figure 1.1: The 219 565 stars common to the TGAS and RAVE datasets plotted in a Galactic frame of reference. The horizontal coordinate is the Galactic longitude l and the vertical is the Galactic latitude b .

heliocentric radial velocities. The metallicity of a star, $[M/H]$, is a logarithmic measurement of its abundance of metals normalised to the abundance of hydrogen compared to the same fraction for the Sun (Schneider 2006).

The latest results released from RAVE was the fifth issue (DR5) and it contained distance estimates to the stars in the catalogue on top of the previously mentioned properties. Ideally, the combined RAVE and TGAS dataset should include the most precise of the DR5 distance estimates and the ones derived from the TGAS parallaxes. However, obtaining distance estimates from parallaxes is non-trivial.

After careful analysis of the two datasets, it was shown by McMillan et al. (2018) that it is possible to produce improved distance estimates for 219 565 stars by combining the two sets. The difference in the quality of the combined sample and the two isolated arrays of distance estimates is considerable, as the median fractional distance uncertainty drops from 31 % for DR5 only estimates, to 15 % when both DR5 and TGAS distance estimates are included. Further on we refer to these distance estimates as McMillan distance estimates. The data points of the sample are portrayed in a Galactic coordinate system (l, b) star map in figure 1.1.

1.2.1 Obtaining the distance estimates

In order to obtain the improved estimates by combining the RAVE DR5 distance estimates with the TGAS parallaxes, [McMillan et al. \(2018\)](#) used a Bayesian estimation method based on the work of [Burnett & Binney \(2010\)](#). Essentially, this method compares the observed luminosity of a star and the expected magnitude from a stellar model, which can be used to estimate the distance to the star from the distance modulus.

The Bayesian statement used is applied by a so-called distance pipeline which combines inputs from measured data for a single star, as well as a model star with known physical properties. The statement is as follows:

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}, \quad (1.1)$$

where P signifies the probability distribution function (pdf) in distance modulus for a given argument. The data parameters used to derive $P(\text{model}|\text{data})$ are: the effective temperature (RAVE DR5); $\log g$ (RAVE DR5); $[M/H]$ (RAVE DR5); various colour band magnitudes from other surveys of the Milky Way and parallaxes (TGAS). The theoretical physical quantities of the model star are determined with the use of isochrones⁴.

The pdf in distance modulus can be characterised by its expectation values and uncertainties in distance modulus, distance or parallax. It can also be represented by a single- or multi-Gaussian fit. The reason why several Gaussian components may be needed is the fact that it in some cases can be difficult to determine whether a star is to be classified as main sequence or giant. Hence for those cases, a single Gaussian component will not provide a good fit to the pdf.

The reason why the inclusion of the TGAS parallaxes reduces the uncertainties in the distance estimates is that it provides another parameter for the distance pipeline. This can easily be seen in the fact that it significantly reduces the need to use multi-Gaussian fits for stars, as it allows for higher confidence in the determination of a star's position within the Hertzsprung-Russell diagram. For RAVE-only distance estimates, multi-Gaussian representations are needed in $\sim 45\%$ of the cases, while this drops to $\sim 23\%$ when TGAS parallaxes are included.

1.2.2 Known issues with TGAS & RAVE

Despite the vast improvement in fractional distance uncertainty, the dataset containing observations common to TGAS and RAVE DR5 with McMillan distance estimates (henceforth referred to as TGAS & RAVE) still carries large uncertainties. Partly because of

⁴An isochrone is a curve in the Hertzsprung-Russell diagram indicating the positions of stars with the same age ([Schneider 2006](#)).

Table 1.1: The quality flags that are included in the TGAS & RAVE dataset. The description explains what observations are removed when each flag is raised, i.e. set equal to 1. N_{stars} tells us how many of the TGAS & RAVE observations that fulfil the criteria.

Flag name	Description	N_{stars}
<code>flag_dup</code>	Duplicate RAVE observations, keeps the best measurement.	25998
<code>flag_lowlogg</code>	Stars with $\log g < 2$.	28492
<code>flag_N</code>	Stars with abnormal spectra.	26726
<code>flag_outlier</code>	TGAS and RAVE-only parallaxes differ with more than 4σ .	2017
<code>flag_pole</code>	Observations in a problematic region near the ecliptic pole.	11286
<code>flag_any</code>	Removes stars for which any of the above is true.	81866

the known shortcomings of GDR1, which merely is a preliminary release of data, but also because of the dissimilarities between the TGAS and RAVE DR5 catalogues. RAVE and *Gaia* use fundamentally different techniques to obtain data, which leads to a divergence in precision for some types of stars.

Moreover, the RAVE DR5 data has known limitations such as abnormal spectra for some observations and systematically underestimated values for low $\log g$ stars compared to different catalogues. Therefore, TGAS & RAVE includes a set of quality flags which are introduced in table 1.1. They each identify stars that fulfil a certain criterion which makes its data unreliable. By removing data where a given flag statement is true, a higher quality sample can be obtained. Two flags that require further explanation are `flag_N` and `flag_lowlogg`, as for the rest we refer to [McMillan et al. \(2018\)](#) for a more in-depth discussion.

The abnormality in the spectra of the stars flagged with `flag_N` can be assigned to their morphological classification. Most of the stars removed when `flag_N` is raised either have spectra morphologically classified as spectroscopic binary stars or chromospherically active stars ([Matijević et al. 2012](#)). The latter class mostly include dwarf stars of spectral class K, which tend to be faint.

As for `flag_lowlogg`, the stars for which $\log g < 2$ is true are giants. Giants are intrinsically bright and are visible at significantly larger distances than the fainter dwarfs, which means that many of the observed giants are distant. Therefore, most of the low $\log g$ data points in TGAS have small parallaxes and thus carry large relative uncertainties. This is not the case for RAVE, which indicates that the low $\log g$ data points in RAVE have systematically underestimated parallaxes.

1.3 The Milky Way

The Milky Way is a barred spiral galaxy and its structure can essentially be divided into three main components: the bulge, the disc and the halo. The stars in the Milky Way are divided into two subgroups by their metallicity. The first subgroup is called population I (pop I) and contain young stars with a metallicity similar to that of the Sun, $[M/H] \sim 0$. The next subgroup is the old metal-poor population II (pop II) stars with $[M/H] \leq -1$.

1.3.1 The Galactic bulge

The bulge makes up the centre of our Galaxy and is a dense concentration of stars. It roughly has the shape of a bar with a scale-length of ~ 1 kpc, which explains the Milky Way's morphological classification. At the centre of this bar, there is a large concentration of mass in a small volume, referred to as the Galactic nucleus. Most studies have arrived at the conclusion that this region is inhabited by a supermassive black hole ([Schneider 2006](#)).

1.3.2 The Galactic disc

Viewing the Milky Way face-on would show us stars spread out in a luminous circular surface centred on the Galactic nucleus that stretches out to a radius of approximately 20 kpc. This so-called Galactic plane is an important reference for the various coordinate systems used to describe the positions of stars in the Galaxy. From an edge-on perspective, we would observe that this Galactic component is not merely a surface but also has a thickness of ~ 1 kpc, which roughly is symmetrically distributed above and below the Galactic plane. This characteristic has provided the substructure with its name: the Galactic disc.

A closer look tells us that the disc can be divided into two distinct sub-components with different densities: the thin and the thick disc. The thin disc has a typical scale-height of 300 pc, while the less dense thick component has a scale-height of ~ 1 kpc.

Disc stars in TGAS & RAVE

We consider a three-dimensional coordinate system where \hat{x} is parallel to the Galactic plane, pointing towards the Sun and \hat{y} is in the direction of motion of the Sun, also lying in the plane. The \hat{z} vector points upwards, perpendicular to the plane. The Sun, which belongs to pop I, is a member of the thin disc and can be found at a distance of $r_{\odot} \sim 8$ kpc from the Galactic centre (GC) at a height z_{\odot} on the order of 15 pc above the Galactic plane.

The position of our solar system within the thin disc of the Milky Way puts a restriction on the possible observations that can be made in the visible bandwidth. High densities of stars and extinction by dust in the interstellar medium limits how far we can see into the disc. Therefore TGAS & RAVE only contains precise measurements at optical wavelengths of stars within the solar neighbourhood.

A vast increase in the size of the sampled region is expected in future releases of *Gaia* data (Gaia Collaboration et al. 2016a).

Disc dynamics

To a first approximation, the Milky Way stars rotate around GC on simple circular orbits with a velocity $v(R)$, where R is the distance from GC. If this was the case, the Sun would have the orbital velocity v_0 and we would find that its three-dimensional velocity is $\mathbf{v} = (0, v_0, 0)$. However, this proves to be a poor approximation since previous studies have found that the Sun has non-negligible velocity components with respect to a completely circular orbit (e.g. Schönrich, Binney, & Dehnen 2010). Therefore, to gain a sense of the Sun's three-dimensional motion with respect to GC we have to include the solar motion relative to the circular reference orbit at a radius of r_\odot , which is commonly referred to as the Local Standard of Rest (LSR). The velocity of LSR is given by $(U_{LSR}, V_{LSR}, W_{LSR}) = (0, v_0, 0)$, where U_{LSR} , W_{LSR} , V_{LSR} are the velocity components along the x , y and z axes, respectively.

The non-zero velocity components along the z and y axes seem to be present for most stars within the disc. Thus, when taking a closer look at the dynamics of the Milky Way disc, it does not seem to be in a stable state of equilibrium. Instead, its stars show complex kinematic behaviour as non-zero velocity gradients have been detected in the solar neighbourhood using RAVE data (e.g. Siebert et al. 2011; Williams et al. 2013).

1.3.3 The Galactic halo

The entire Milky Way is enveloped by an approximately spherical component called the Galactic halo, which has a density that is comparably low in relation to the other substructures. The visible part of the halo, which is the region inhabited by so-called field stars, has a radius of ~ 50 kpc. The field stars are pop II stars that traverse the halo on a large variety of orbits far from the Sun, often with lower angular momenta than stars residing in the Galactic plane but with overall greater total velocities.

Halo dynamics

Conducting surveys on the kinematics of Galactic halo field stars poses quite the challenge. They are in general significantly fainter than the stars that belong to the thin and thick disc populations due to large distances from the Sun, so measurements of the physical properties of halo stars carry large uncertainties that manifest themselves in the halo star distance estimates. The complex distribution of motions combined with the on average more uncertain distance estimates of the halo stars makes it more difficult to find patterns and obtain a proper understanding of the dynamics involved.

There is an ambiguity regarding the origin of halo stars. Some findings suggest that halo stars have been formed in-situ from metal-poor gas in the early stages of the Milky Way’s evolution, while others indicate that a fraction of the halo stars has been accreted due to tidal forces in galactic mergers between low-mass satellite galaxies and the Milky Way (Helmi et al. 1999). A better understanding of their origin will help resolve this ambiguity.

Halo stars in TGAS & RAVE

On account of their large distances from the Sun, a major portion of the observed halo stars in the TGAS & RAVE dataset have to be intrinsically bright, i.e. giants. We recognise that the low $\log g$ stars in our sample are known to carry large uncertainties due to underestimated parallaxes in RAVE (see section 1.2.2). Therefore, many of the possible halo stars observations available in TGAS & RAVE will be affected when we cut the sample from data points with `flag_lowlogg` raised. This means that we either have to sacrifice sample size or data quality when studying halo subsets based on our data.

1.3.4 Coordinate systems

In our study of the Milky Way dynamics, we have used two different coordinate systems. The first system is the International Coordinate Reference System (ICRS) which has its origin in the barycentre of the Solar System.

TGAS & RAVE uses the coordinates (α, δ) to identify the positions of the observations (right ascension and declination). All the relevant astrometric properties from the stars common to TGAS & RAVE that has been used in the project can be found in table 1.2.

The second frame of reference used is a Galactocentric system with cylindrical representation (R, ϕ, z) , which in order are: the radius from the Galactic centre with $\hat{\mathbf{R}}$ pointing from GC towards the Sun; the azimuthal angle counterclockwise from the Cartesian x -axis, which points from the Sun towards GC; the height above the Galactic plane. The resulting

Table 1.2: Relevant variables and corresponding units for the ICRS frame with a spherical coordinate representation.

Variable	Notation	Unit
Right ascension (J2000)	α	$^{\circ}$
Declination (J2000)	δ	$^{\circ}$
Distance from the Sun	d	pc
Parallax	ϖ	mas
Proper motion in right ascension	μ_{α^*}	mas yr $^{-1}$
Proper motion in declination	μ_{δ}	mas yr $^{-1}$
Radial velocity (heliocentric)	v_r	km s $^{-1}$

Table 1.3: Relevant variables and corresponding units for the Galactocentric frame with a cylindrical coordinate representation.

Variable	Notation	Unit
Galactocentric radius	R	pc
Azimuth	ϕ	$^{\circ}$
Height	z	pc
Velocity in R	v_R	km s $^{-1}$
Velocity in ϕ	v_{ϕ}	km s $^{-1}$
Velocity in z	v_z	km s $^{-1}$

quantities obtained after converting to a Galactocentric frame of reference are given in table 1.3.

While we opt to use this default right-handed Galactocentric reference system in most of our computations, we have employed a left-handed system in chapter 4 where $\hat{\phi}$ is positive in the direction of the Sun’s rotation. This allows for easier comparison with the results of [Helmi et al. \(2017\)](#).

Transformation between the systems

In order to treat the RAVE & TGAS data, a code was written based on the `astropy` package of Python ([Astropy Collaboration et al. 2013](#)). Given that the initial values and units of the different quantities were based on the ICRS coordinate system, the dataset was initialised using containers of the `astropy.coord` library to allow for easy conversions to and from a Galactocentric frame of reference.

In order to transform between the three-dimensional ICRS frame (α, δ, d) and the Galactocentric frame, additional quantities are needed. First, we need to specify the position of the GC $(\alpha, \delta)_{GC}$ and its distance from the Sun r_{\odot} . The radial velocity of the stars with respect

to the Sun combined with the proper motions $(\mu_{\alpha^*}, \mu_{\delta})$ then yields a three-dimensional velocity. Said velocity can be transformed to a Galactocentric frame by specifying how the Solar system barycentre moves with respect to GC.

An important difference between the ICRS and the Galactocentric reference systems is that the three-dimensional velocities measured in the ICRS coordinates are independent of the distance estimates. As soon as we convert the data points to a Galactocentric frame, the velocities will be strongly coupled to the distances estimates. This will allow us to clearly see the impact of a dataset with more accurate distance estimates when re-examining the dynamical phenomena.

Details about the matrices used in the code to perform the transformations can be found in the `astropy` documentation ([Astropy 2018](#)).

1.4 Three dynamical phenomena

The papers chosen for our re-examination of the dynamical properties of the Milky Way have all been based on datasets with strong similarities to TGAS & RAVE, both when it comes to size and objects observed. The main differences between the samples are the distance estimates employed. Due to the significant increase in precision of the McMillan distances as compared to estimates previously available, we set out to investigate the implications that this might have on the results of the studies.

Moreover, the three articles touch on dynamical phenomena related to different substructures in the Galaxy. Accordingly, we cover the kinematics of a wide range of observations at varying distances in our investigation.

The first phenomenon originates from the interaction of zero angular momentum stars with the Galactic nucleus. It was found in [Hunt, Bovy, & Carlberg \(2016\)](#) that there is a 60 km s^{-1} dip in the tangential velocity distribution of neighbourhood stars. This indicates that there is a shortage of stars with velocities within a range of size 60 km s^{-1} , which makes the distribution diverge from its expected smooth nature. This dip was reported to be centred on a velocity corresponding to zero angular momentum. The suggested explanation for this observation is that stars on orbits heading directly towards the Galactic nucleus have been scattered away onto more chaotic orbits. We investigate if the dearth is still present for our improved distance estimates.

The second phenomenon is related to the kinematic behaviour of halo stars in the solar neighbourhood. As mentioned, there is an ambiguity regarding the origin of halo stars (see section 1.3.3). The detection of a second component of the halo that has a counter-rotating nature in terms of orbit - with respect to the Sun - speaks in favour of the theory that part of the halo has been built by accretion ([Carollo et al. 2007](#)). However, said detection led to a discussion of whether or not the observed retrograde motions of some halo stars were a

direct cause of uncertain distance estimates (Schönrich et al. 2011; Beers et al. 2012).

In their study of the possible halo stars in the observations common to TGAS and RAVE, Helmi et al. (2017) found that a significant portion of the solar neighbourhood halo stars are on such counter-rotating orbits, which further speaks in favour of the theory where halo stars originate from accretion. Our re-examination looks into the possibility that the fractions have been exaggerated due to uncertain distance estimates.

The third and final phenomenon involves the complex motions of the Galactic disc. As mentioned, non-zero velocity components and gradients have been detected for solar neighbourhood stars in the directions of $\hat{\mathbf{R}}$ and $\hat{\mathbf{z}}$, causing wave patterned motions in the disc.

The survey in Williams et al. (2013) shows that the gradients differ above and below the Galactic plane, from an edge-on perspective. They also detect a difference in signs of the median velocities inside and outside of the solar circle. For example, this pattern in v_z results in a rarefaction-like motion for stars with $R < r_\odot$ and a compression-like motion for $R > r_\odot$. These are known as breathing modes. They also found a similar behaviour in v_R , where the stars within the solar circle move radially outwards and stars outside move inwards. In our review of the phenomenon, we investigate if we detect the same patterns when utilising the improved distance estimates and how uncertainties in data can affect such results.

Given the known shortcomings of TGAS & RAVE, we acknowledge that we are limited in our ability to draw precise quantitative conclusions regarding e.g. the width of a possible dearth, the fraction of halo stars on a counter-rotating orbits or the magnitudes of the velocity gradients in the disc. Nevertheless, we are still able to discuss the qualitative aspects of the phenomena and determine how the change in the quality of distance estimates affect the observed dynamics of the Milky Way substructures. To clarify, our interests lie in determining whether the claims made are still valid for our dataset and how systematic and random errors in the data influence the outcome of our investigations.

The paper is structured as follows. In chapter 2 we introduce the methods employed to treat the data used in our investigations. In chapter 3 we investigate the presence of a dearth of zero angular momentum stars proposed in Hunt et al. (2016). In chapter 4 we re-examine the findings of Helmi et al. (2017) which suggests that the fractions of possible halo stars on counter-rotating orbits in the solar neighbourhood is significantly larger than the corresponding fraction for halo stars on prograde orbits. In chapter 5 we investigate the local dynamics of the disc to learn whether the breathing modes of the disc suggested in Williams et al. (2013) are present or not. Finally, we provide a general discussion and conclusion regarding the findings of the study in chapter 6.

Chapter 2

Treating the data

In the re-examination of the Milky Way dynamics, we have resorted to statistical resampling to make up for the lack of data points in our sample. Additionally, we have employed a velocity model to allow for the creation of samples with velocities independent of the distance estimates. These main methods are introduced after a definition of numerical values and the different samples used in our study.

2.0.1 Numerical values

In order to convert the relevant properties between the coordinate systems, the position and relative motion of the Sun with respect to the Galactic centre had to be known. Using the findings of [McMillan \(2017\)](#), it was assumed that the Sun is at a distance of $r_{\odot} = 8.20$ kpc from the Galactic centre. As for the position of the Sun with respect to the Galactic plane z_{\odot} , the value of 15 pc from [Binney, Gerhard, & Spergel \(1997\)](#) was used. The position of the Galactic centre was taken to be $(\alpha, \delta)_{GC} = (17 : 45 : 37.224 \text{ hr}, -28 : 56 : 10.23 \text{ deg})$ ([Reid & Brunthaler 2004](#)).

It was further assumed that the LSR moves along a circular orbit with a velocity of $v_0 = 232.8 \text{ km s}^{-1}$, such that $\mathbf{v}_{LSR} = (0, 232.8, 0) \text{ km s}^{-1}$. Due to the fact that the Sun's orbit around the Galactic centre is not completely circular, its peculiar motion relative to the LSR had to be included as well. Here the values from [Schönrich et al. \(2010\)](#) of $(U_{\odot}, V_{\odot}, W_{\odot}) = (11.10, 12.24, 7.25) \text{ km s}^{-1}$ were adopted. Hence, the total velocity of the Sun on its orbit around the Galactic centre is taken to be $\mathbf{v}_{\odot} = (11.10, 245.04, 7.25) \text{ km s}^{-1}$.

Table 2.1: The name of reference and description of the three datasets used during our investigations of the phenomena mentioned in this paper.

Subset name	Quality flags raised
Default sample, <code>flag_dup</code>	<code>flag_dup</code>
<code>flag_any</code>	<code>flag_dup</code> , <code>flag_N</code> , <code>flag_outlier</code> , <code>flag_pole</code> , <code>flag_lowlogg</code>
<code>flag_any-lowlogg</code>	<code>flag_dup</code> , <code>flag_N</code> , <code>flag_outlier</code> , <code>flag_pole</code>

2.0.2 TGAS & RAVE subsets

The original TGAS & RAVE dataset with improved distance estimates contained results from 219 565 RAVE observations. However, due to the fact that RAVE has measured the properties of some stars more than once, directly using this sample in computations would lead to double counting of some data points. Thus, the default sample size was reduced to 193 567 stars by removing all points for which the `flag_dup` statement is true (see table 1.1). This meant that only the highest quality measurement of the unique observations was kept.

We could obtain other more reliable subsets of the default sample by raising additional quality flags. Two additional sets were used during the re-examination of the three aforementioned dynamical phenomena in the Milky Way (see section 1.4). The first one being the highest quality data available to us, where all flags have been raised. In this paper we refer to this as the `flag_any` sample and it contains 137 699 data points. Furthermore, this sample is used to investigate results found when using the default sample, which allows us to learn whether said results still can be seen when the data points we know to carry large errors have been removed.

The third dataset we use is referred to as `flag_any-lowlogg`, where all flags except `flag_lowlogg` have been raised. This sample is used whenever we investigate dynamical features involving halo stars and has a size of 159 920. As mentioned in section 1.2.2, a large portion of the possible halo stars of the observations contained in the TGAS & RAVE have low $\log g$ values. They are thus removed when we raise `flag_lowlogg`, yielding a sample with an insufficient number of halo stars. Our three main samples and the related quality flags can be found in table 2.1.

After initiating a sample, it was transformed from a spherical ICRS to a cylindrical Galactocentric coordinate representation, which was our main frame of reference for the studies in this project.

2.1 Statistical tests

In order to draw conclusions based on the distribution of the quantities in the dataset, statistical tests had to be employed. The reason behind this is that there is a limited number of stars from the TGAS & RAVE data and the resulting distribution of a given variable ω in the sample cannot be taken to represent the population of all stars in the Milky Way. Since the real distribution is unknown, it is not possible to estimate the true error of the statistic. A way to get around this issue is to utilise statistical resampling methods.

2.1.1 The bootstrap

The first technique employed was the so-called bootstrap (Efron 1979). Given a sample with N data points $\mathbf{X} = X_1, X_2, \dots, X_N$ with corresponding variables ω_i for each data point i , the bootstrap draws a random sample $\hat{\mathbf{X}} = \hat{X}_1, \hat{X}_2, \dots, \hat{X}_N$ with replacement based on the original sample. The set of variables $\hat{\omega} = \hat{\omega}_1, \hat{\omega}_2, \dots, \hat{\omega}_N$ can subsequently be determined from the bootstrap sample. By repeating this many times, e.g. 1000 iterations, statistical properties of the bootstrap samples such as the mean or median values can be evaluated. By visualising the estimated mean or median of the random samples and the corresponding variation in each point, information regarding the validity of the original sample can be acquired.

An important statistical property for visualising the variation is the standard deviation σ_{SD} . For an array of K bootstrap samples, each of length N , the standard deviation in a variable $\hat{\omega}_i$ can be computed using the following formula:

$$\sigma_{SD,i}^2 = \frac{1}{K} \sum_j^K (\hat{\omega}_{i,j} - \bar{\omega})^2, \quad (2.1)$$

where $\bar{\omega}$ is the arithmetic mean of the K bootstrap sample variables computed for data point i of the total N data points.

2.1.2 Uncertainty sampling

The second technique was based on the known errors provided in the dataset, from both direct measurements and estimates of physical properties. The process was carried out in the original ICRS coordinate system, as converting uncertainties between different frames is non-trivial. Given a parameter ω with a known error ω_{err} a random number was sampled from a univariate Gaussian distribution and multiplied with ω_{err} , yielding a new error $\hat{\omega}_{err}$. The sampled error was then added to the original parameter, resulting in the resampled

Table 2.2: The chosen parameters for the modelled Gaussian velocity distributions of each substructure in the Milky Way.

Substructure	Relative weight	μ_v [km s ⁻¹]	σ_v [km s ⁻¹]
Thin disc	0.91	(0,-215,0)	(30,20,17)
Thick disc	0.08	(0,-180,0)	(80,60,55)
Halo	0.01	(0,0,0)	(160,100,100)

value $\hat{\omega}$. By repeating this procedure for all relevant quantities in the dataset (see table 1.2), a new sample could be obtained. After creating a large number of samples using this technique, the standard deviation for a given parameter was computed using equation (2.1).

2.2 Velocity model

To further reaffirm conclusions drawn from the results of resampling, new random samples were created from the original sample by retaining the estimated position and assigning velocities based on a model to each data point. We refer to these as pseudo-samples. These types of simulations are useful since they yield velocities for the stars in the sample that have known properties. Accordingly, we can learn how the distance errors affect the results of our dynamical analysis. By including the model in our computations to simulate various properties of the population, we could obtain a deeper understanding of how the presence of these phenomena could be found when investigating the distributions of the original sample.

Furthermore, we could study the impact that random errors had on the dataset distributions by adding random errors based on the uncertainties of the dataset to the relevant physical quantities. The velocity model also proved useful when testing how possible systematic shifts in the distances could affect the outcomes of the re-examination. This method is introduced and further discussed in chapter 4.

The applied model was an adaptation from [Kordopatis et al. \(2013\)](#), where it is assumed that the distribution of velocities in each of the Milky Way’s substructures can be represented by Gaussian distributions. This yields three different populations, each with a relative weight χ_i such that $\chi_{thin} + \chi_{thick} + \chi_{halo} = 1$. Each population has a given mean velocity, $\mu_v = (\mu_{v_R}, \mu_{v_\phi}, \mu_{v_z})$, and dispersion of velocities, $\sigma_v = (\sigma_{v_R}, \sigma_{v_\phi}, \sigma_{v_z})$, which together with the chosen relative weights are given in table 2.2.

The velocities were assigned to the N stars in the original sample by creating a random set of numbers $\kappa = \kappa_1, \kappa_2, \dots, \kappa_N$ in the range $\kappa_i \in [0, 1)$. Each value was compared to the relative weights of the substructure populations. If $\kappa_i \leq \chi_{thin}$, the velocity $\mathbf{v}_i = (v_R, v_\phi, v_z)_i$

was randomised from the thin disc Gaussian; if $\chi_{thin} < \kappa_i \leq \chi_{thin} + \chi_{thick}$, v_i was taken from the thick disc distribution. The rest of the stars were assigned a halo velocity.

The values for the relative weights of the substructures were chosen such that the shape of the velocity model distribution would resemble the default sample distribution, where only `flag_dup` has been raised. A comparison between the distribution of tangential velocities v_ϕ for the default sample and the model can be seen in figure 2.1

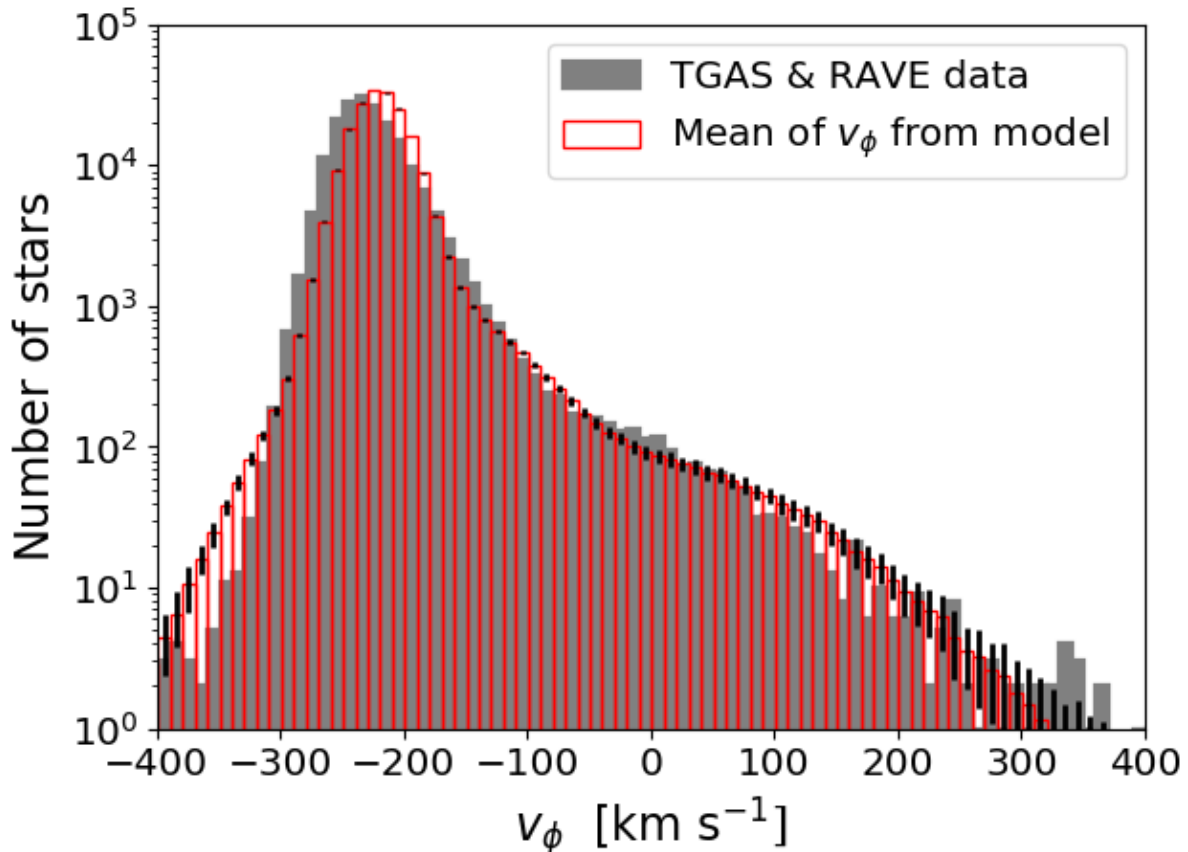


Figure 2.1: A histogram displaying the mean of 100 pseudo velocity distributions obtained using the model from [Kordopatis et al. \(2013\)](#) compared to the computed velocities for the default subset of the TGAS & RAVE data. The black bars indicate the standard deviation in bin counts for the computed pseudo-samples.

Chapter 3

Dearth of zero angular momentum stars

The first proposed phenomenon regarding the nature of Milky Way dynamics that was investigated was the claim that there is a lack of stars in the solar neighbourhood with angular momenta close to zero (Hunt et al. 2016). The paper suggests that the origin of the dearth is the scattering of stars with zero angular momentum onto halo orbits after interaction with the Galactic nucleus. Their findings show an approximately 60 km s^{-1} wide dip in the distribution of velocities around the Cartesian coordinate $v_y = -v_\odot$ in a Galactic frame of reference centred on the Sun, which is portrayed in figure 3.1. Here, v_y is positive in the direction of the Sun’s rotation around the Galactic centre. In this coordinate system, the Sun has a velocity $v_y = 0 \text{ km s}^{-1}$.

While the analysis that led to the conclusion was based on the combined TGAS and RAVE DR5 data, Hunt et al. (2016) opted to use the spectrometric distance estimates of DR5 when computing the velocities of the stars in question. TGAS parallaxes were only included when the DR5 distance estimate of a given star was not available. The large relative uncertainties in the DR5 distances compared to the improved estimates made available by McMillan et al. (2018) gave reason to review the findings anew.

In the re-examination of the presence of a zero-momentum-star dearth around $v_\phi = 0 \text{ km s}^{-1}$ we looked to answer three main questions using histograms of the velocity distribution, equipped with the statistical tools introduced in section 2.1. The main questions are as follows:

- Will we still detect a 60 km s^{-1} wide dip in the relevant velocity distribution when using the improved McMillan distance estimates?
- How large is the standard deviation in the number counts of the velocity bins around $v_\phi \sim 0 \text{ km s}^{-1}$?
- How wide must a dip be in order for us to detect it?

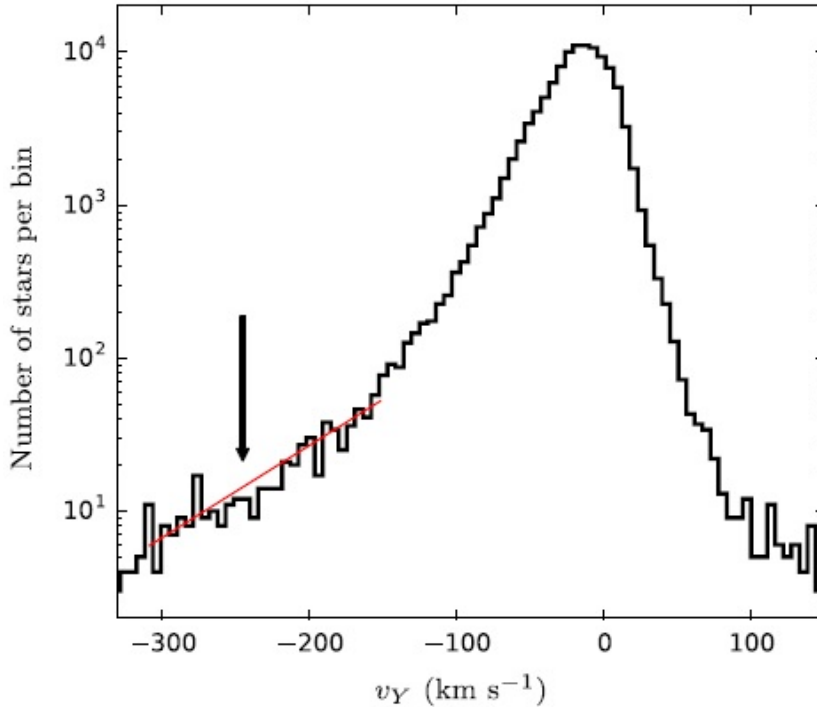


Figure 3.1: Histogram of the v_y distribution for stars in the solar neighbourhood which has been adapted from figure 1 in [Hunt et al. \(2016\)](#). The dearth of zero angular momentum stars is centred around $v_y = -v_\odot$, which is indicated with an arrow. Reprinted with permission.

3.1 Method

The re-examination of the dearth of zero angular momentum stars was done by computing the tangential velocity distribution using the dataset of stars common to TGAS and RAVE DR5 with the improved McMillan distance estimates and portraying it in histograms. The possible presence of a dip was evaluated by investigating the variation in counts of velocity bins around the proposed centre of the dip. Statistical analysis using the resampling tests introduced in section 2.1 provided a sense of the validity of our sample in the form of an estimate of the true error in the velocity distribution. We further investigated how the results changed when introducing more quality flag cuts to our sample. Finally, the [Kordopatis et al. \(2013\)](#) velocity model of section 2.2 was introduced to create pseudo-samples of data points based on the uncertainties and positions of the stars in the default sample, as well as the additional subsets obtained by raising quality flags. We then computed the mean velocity distribution along with the standard deviation for the pseudo-samples, which allowed us to evaluate the dearth further.

Rather than using a Galactic Cartesian coordinate system centred on the Sun, the re-

investigation was carried out based on the Galactocentric cylindrical coordinate system. Therefore, the dearth of stars - if any - was expected to be found around $v_\phi \sim 0 \text{ km s}^{-1}$. After initialising the data, the tangential velocity was visualised by computing and plotting a histogram, where the velocities between $v_{\phi,min}$ and $v_{\phi,max}$ were assigned to a number of bins with a given width. This width was by default chosen to be 10 km s^{-1} .

3.1.1 Resampling the TGAS & RAVE subsets

Both the uncertainty sampling and the bootstrap were used to create new samples based on a given subset. The new sample - created from either of the two methods - was converted to Galactocentric cylindrical coordinates and the tangential velocity was finally computed and assigned to a histogram bin.

To gain any sensible statistical information from the test, a number of bootstrap samples were needed. Therefore, the bootstrapping process was repeated K times (typically K is 1000 or 10 000), resulting in K bootstrap samples. By applying equation (2.1), the standard deviation for each velocity bin was evaluated and plotted together with the histogram for the arithmetic mean of all sets created from the resampling method. The corresponding velocity distribution for the original sample was added to the plots for easy comparison.

3.1.2 Simulating a dearth of stars

The [Kordopatis et al. \(2013\)](#) velocity model was implemented by providing every star with a random velocity based on the relative weight of the Milky Way substructures and altering its coordinates using the error sampling method. This made it possible to simulate an entirely new set of stars from the various sets in table 2.1.

This simple method could be utilised to gain insight into how wide a possible dip would have to be in order for us to observe it. To achieve this, the model was altered to create a lack of stars within a given limit λ around $v_\phi = 0 \text{ km s}^{-1}$. As completely removing the stars within the limit would be untrue to the phenomena involved and create a variation of the sample sizes, another approach was needed.

The reason why completely removing them is likely to be an inaccurate representation of our situation, is that it would correspond to each star with v_ϕ within the limit having a close encounter with the Galactic nucleus, obtaining extreme tangential velocities far from zero. Such a scenario would cause a sharp and deep dip in the distribution, which clearly is not present in the histogram of [Hunt et al. \(2016\)](#) (see figure 3.1). The issue with varying sample sizes is that it complicates the process of computing the histograms for each pseudo-sample and comparing them to each other.

Therefore, the absence of stars was instead created by pushing the stars within the limit to the sides. This approach led to a smoother dip with more reasonable depth and also kept the sample sizes the same. The algorithm used is as follows:

1. Given a data point i from the original sample, we take a random velocity from the model introduced in section 2.2.
2. We compare the absolute value of v_ϕ to the specified value of λ . If $|v_\phi| \leq \lambda$ it is within the dip and must be altered.
3. If $v_\phi < 0 \text{ km s}^{-1}$ the velocity was altered such that $v'_\phi = v_\phi - \lambda$ and for $v_\phi > 0 \text{ km s}^{-1}$ the new velocity became $v'_\phi = v_\phi + \lambda$
4. The velocities v_R , v'_ϕ and v_z were coupled to data point i .

This process was repeated for each data point in the original sample. By testing various values for λ , creating numerous pseudo-samples using the basic algorithm and plotting the resulting mean and standard deviation in a histogram, the presence of a dearth could be evaluated further.

3.2 Results

In the search for a dip in the v_ϕ distribution, a number of plots were produced using different resampling methods. Histograms with different bin widths were also computed for each method to rule out any possibilities that the patterns observed in the plots only were a direct cause of the number of bins used. The default value for the width of the bins was chosen to be 10 km s^{-1} , but the results were also produced for 5, 15 and 20 km s^{-1} wide bins. A sole plot of the Galactocentric velocity v_ϕ for the default and `flag.any` samples can be found in figure 3.2.

3.2.1 Sampling tests

The resampling of the dataset was first carried out without removing any stars from the original sample but the duplicates. Computing the number counts with standard deviation for each bin and plotting them provided a sense of how the counts varied for the new samples. The resulting histograms for 10 km s^{-1} wide bins are shown in figure 3.3 and a version zoomed in around $v_\phi = 0 \text{ km s}^{-1}$ where a possible dearth would emerge is found in figure 3.4.

Neither the uncertainty nor data point resampling methods show any indication of a scarcity of zero angular momentum stars. Any indications of a significant variation in bin height are well within the computed standard deviation, which was found to be roughly ~ 10 counts per bin around $v_\phi = 0 \text{ km s}^{-1}$. We also observe that there is no substantial

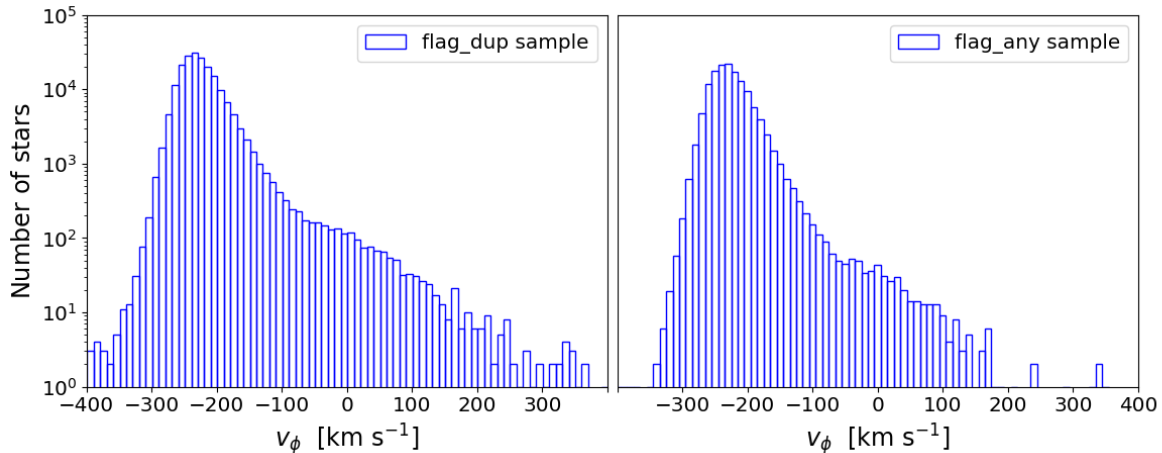


Figure 3.2: Histograms of the tangential velocity of the sample of stars common to the TGAS & RAVE catalogues in a Galactocentric frame of reference with bin widths of 10 km s⁻¹. We have chosen to only show the distributions within the range $-400 \leq v_\phi \leq 400$ km s⁻¹. The left plot is the default sample with only duplicate stars removed. The right plot shows the `flag_any` with all known low-quality measurements removed.

change to the plots if the bin width is altered. Thus, the possibility of bin selection having an impact on the shown distribution can be ruled out. Given that this result is based upon the most uncertain case, with only the duplicate stars removed, we are already led towards the conclusion that there is no dearth present.

The same procedure was thereafter repeated for the `flag_any` sample, meaning that the sample size was reduced to only include our highest quality data points (see section 2.0.2). The outputted histograms for both the uncertainty and the bootstrap resampling for the standard bin width of 10 km s⁻¹ are shown in figure 3.5 and a version zoomed in around $v_\phi = 0$ km s⁻¹ can be found in figure 3.6.

We observe that there is a larger variation in bin heights when employing the bootstrap sampling method for the given set of stars. Nevertheless, the fluctuations are still within the standard deviation for each bin and it is not possible to make out any indications of a dip in either of the plots. This also proved to be the case when reiterating using other bin widths.

Further testing was done for samples with different combinations of quality flags raised, but no apparent lack of zero momentum stars could be observed in the histograms for any of the scenarios. That being so, we are led to believe that the dearth detected by [Hunt et al. \(2016\)](#) was a direct result of the low quality of their distance estimates - especially since a ~ 60 km s⁻¹ wide dip should have been found, assuming their claim was true. However, to completely be able to rule out a possible shortage of stars we needed a better understanding of how such a dip would emerge for the given stars used in the study.

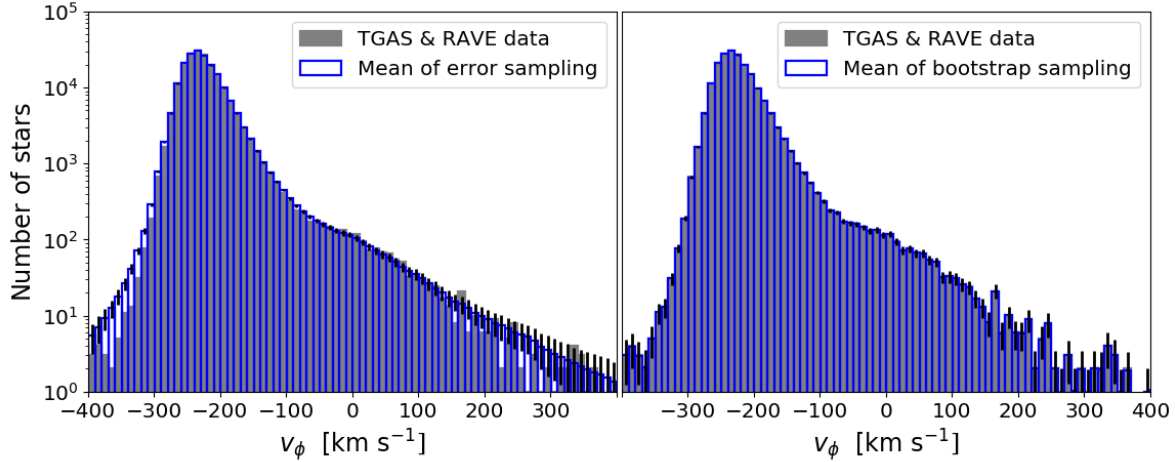


Figure 3.3: The tangential velocity distributions for the default sample with bin widths of 10 km s^{-1} and the mean of the bootstrap samples. The standard deviation for each bin is shown by the black bars. The result from resampling of the uncertainties is to the left while the bootstrap resampling result is to the right. No dip in the bin counts can be seen.

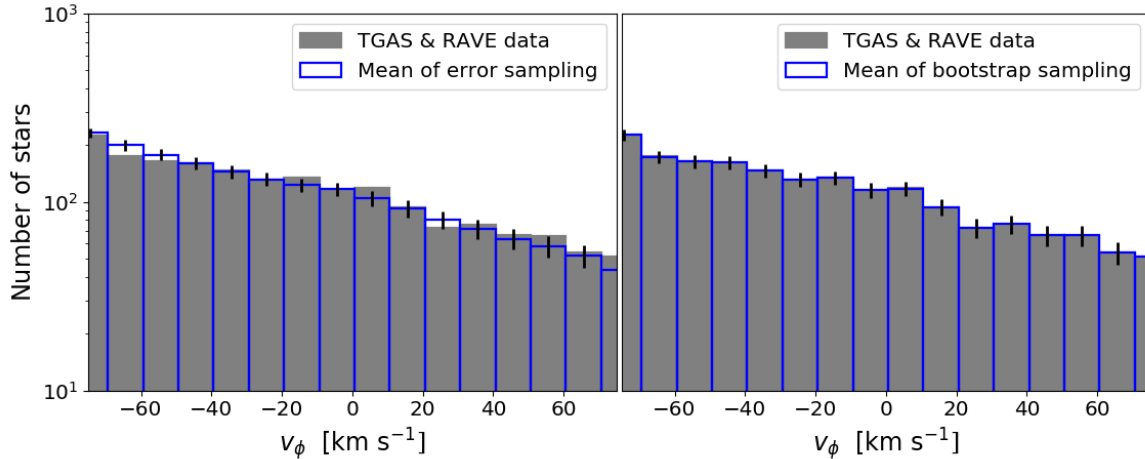


Figure 3.4: The same plots provided in figure 3.3 but zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. We do not see any indication of a dip in counts.

3.2.2 Velocity model simulation

The follow-up question to answer was how large of a dip was needed for us to observe it and if any patterns of dips could be found in the resampling results. Therefore, the next step in the analysis was to carry out the simulations based upon the velocity model with and without a dip centred around $v_\phi = 0 \text{ km s}^{-1}$. By studying various limits λ around $v_\phi = 0 \text{ km s}^{-1}$, we could gain the insight needed to conclude whether the presence of a dip was possible or not.

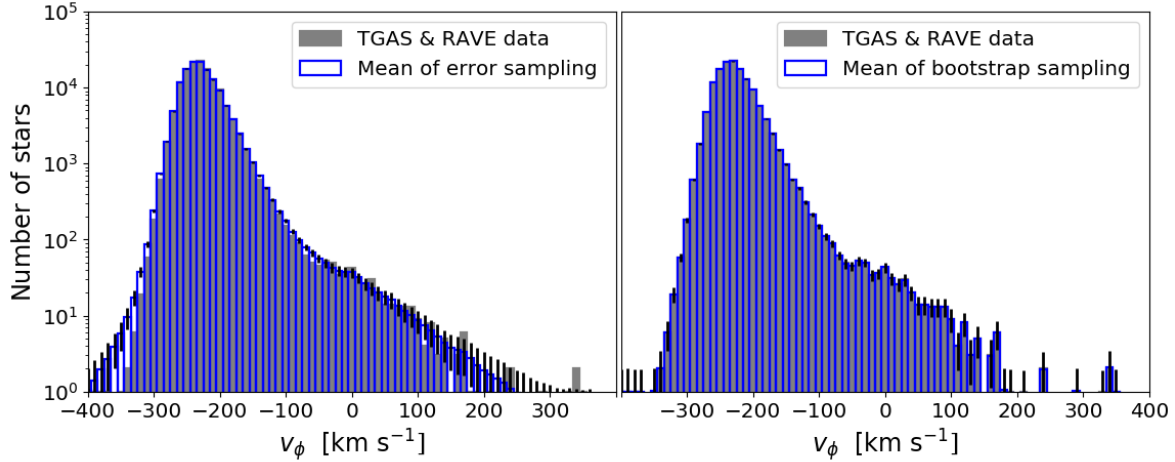


Figure 3.5: The tangential velocity distributions for the sample with `flag_any=0` and the mean of the bootstrap samples computed. The bins have widths of 10 km s^{-1} . The standard deviation for each bin is shown by the black bars. The result from the resampling of the uncertainties is shown in the left plot while the bootstrap resampling result is to the right.

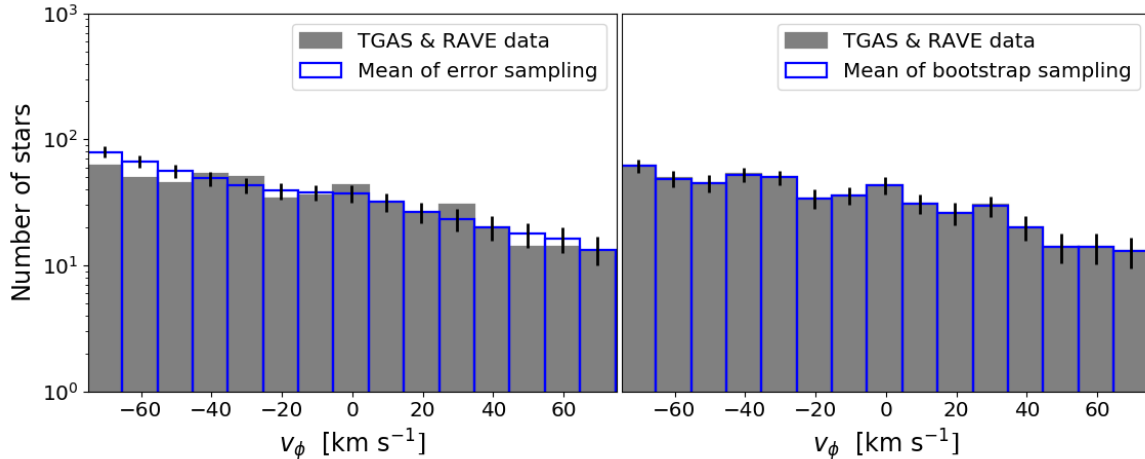


Figure 3.6: The same plots provided in figure 3.5 but zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. In agreement with figure 3.4 there is no indication of a dearth.

This was achieved by combining the velocity model with the uncertainty sampling method to create a large number of pseudo-samples. By producing histograms for each case, computing the standard deviation and comparing them to the resampling results, we could learn which dip widths could be dismissed. A histogram consisting of the standard 10 km s^{-1} wide bins, portraying the tangential velocity distribution of a pseudo-sample produced from the default sample without introducing a dip, is shown in the topmost plot of figure 3.7.

To cover the extreme cases for possible dips in the distribution we chose to run the tests with half-width values such that $\lambda \in [0, 5, 10, 15, 20, 30, 40]$. For each case, the velocity model was sampled 100 times and the mean count, as well as the standard deviation for each bin could be computed from the resulting pseudo-samples.

In addition to the histogram without a dip, we have chosen to display histograms with dip widths of 10, 20 and 30 km s⁻¹ which can be found in figure 3.7 for comparison with the resampling results. The case of a simulated 60 km s⁻¹ wide dip, which was the width of the dip found in [Hunt et al. \(2016\)](#), can be found in figure 3.8.

After a comparison with the error and bootstrap sampling results for the `flag_dup` sample shown in figures 3.3 and 3.4, we notice that the presence of dips with $\lambda = 10, 15, 30$ km s⁻¹ can be ruled out. These dips would correspond to depths of $\sim 50, 100$ and 200 counts, respectively. The standard deviation in the bin heights for the resampling results only covers a dip of depth ~ 10 counts.

Consequently, we cannot draw any conclusions regarding either the presence or the absence of a 10 km s⁻¹ wide dip from this result. The large standard deviations in the bin counts prevent us from making any further progress. To conduct further investigations we have to base our pseudo-samples on higher quality data and opt for more narrow bins. Hence, we chose to re-investigate the modelled dip using the `flag_any` subset. However, we were still unable to see any dip around $v_\phi \sim 0$ km s⁻¹ due to the uncertainties in the sample, which is shown in appendix A.

3.3 Discussion

In conclusion, we have investigated the distribution of tangential velocity v_ϕ for a Galactocentric frame of reference. This was done in order to assess the existence of a dearth of stars around $v_\phi = 0$ km s⁻¹. The results did not show any indication of the presence of a 60 km s⁻¹ wide dip, which was the claim made in [Hunt et al. \(2016\)](#).

It is noteworthy that we never tried to replicate the results showing the dip using the same data as in [Hunt et al. \(2016\)](#) for a Galactocentric coordinate system, which would have been an easy way of comparing the difference in distribution between the two datasets. Nevertheless, it was clear from the distribution of the tangential velocities that a possible dip in the bin counts would have to be significantly narrower than in [Hunt et al. \(2016\)](#).

Therefore, we used a velocity model to simulate a dip around $v_\phi = 0$ km s⁻¹ in the distribution to further assess the results, which would tell us how wide a dip would have to be in order for us to detect it in our histograms. This was done by pushing velocities within a given range $v_\phi \in [-\lambda, \lambda]$ around the zero angular momentum point to either side. It was found that we would likely be able to detect a dip of width ~ 20 km s⁻¹ and wider, which reinforces our claim that there is no dip of width 60 km s⁻¹ present in the v_ϕ

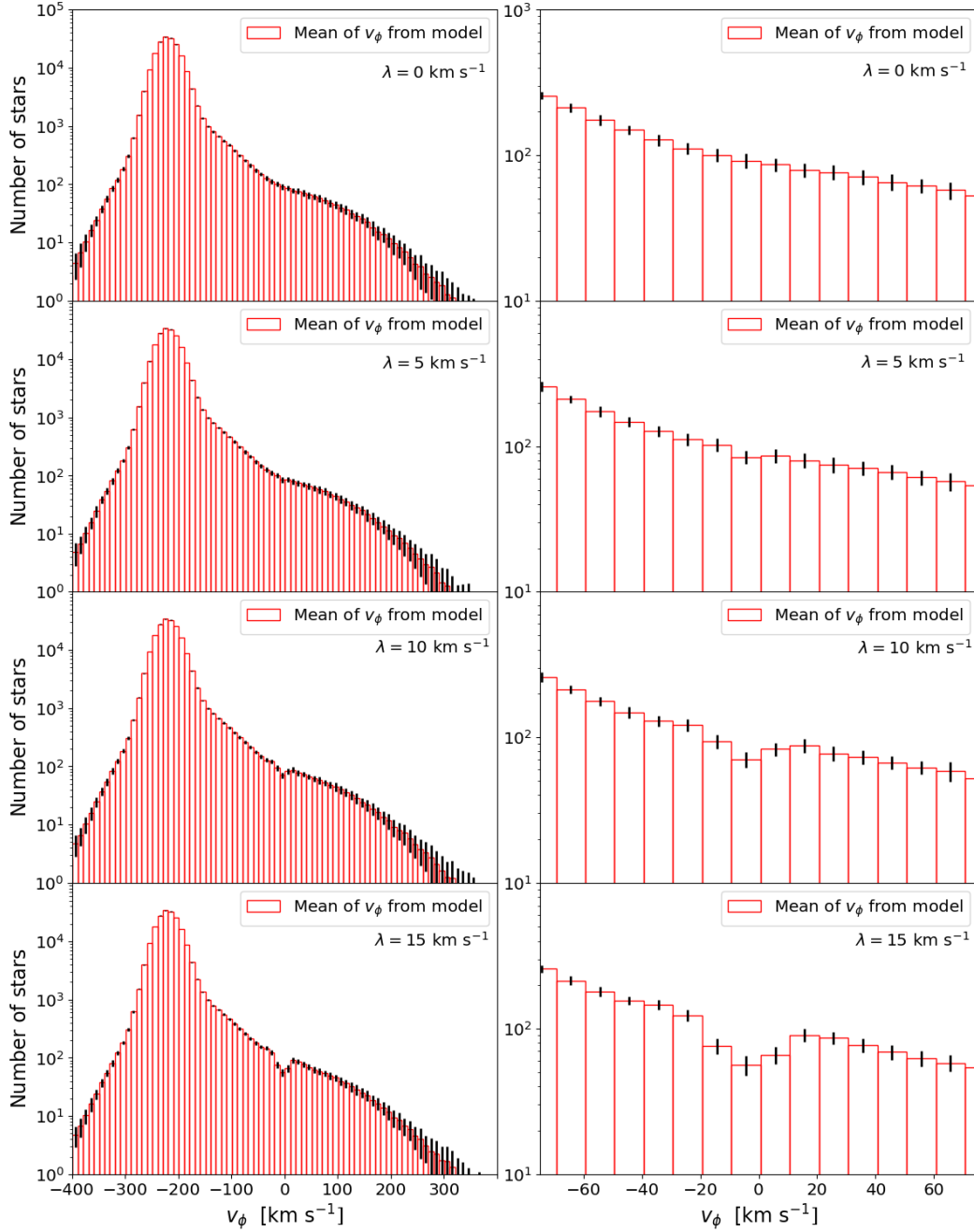


Figure 3.7: Histograms of the distribution of tangential velocities for the default sample data points when the velocities have been taken from the model described in section 2.2. For each case of λ , the left plot shows the major part of the distribution while the right is zoomed in around $v_\phi = 0 \text{ km s}^{-1}$.

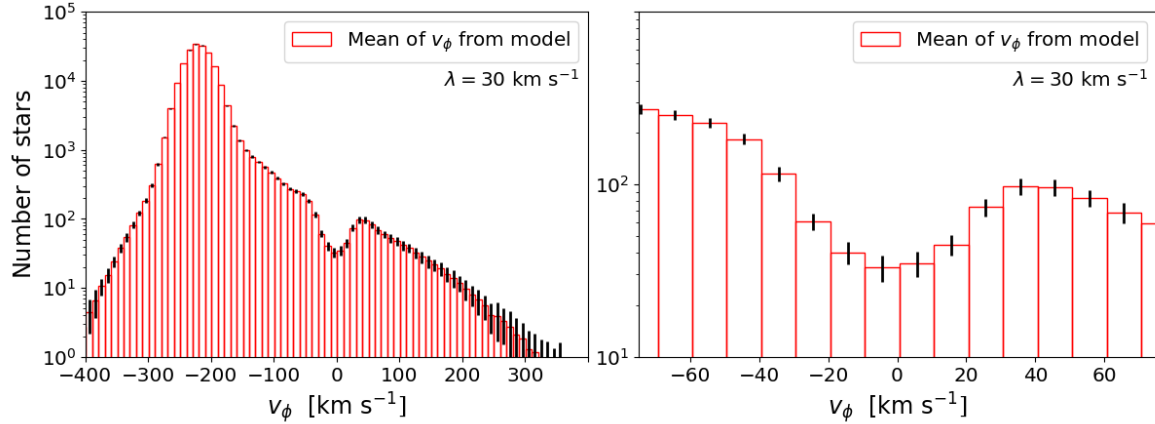


Figure 3.8: The same plots as in figure 3.7 but the simulated dearth now has a width of 60 km s $^{-1}$ which would correspond to the findings of [Hunt et al. \(2016\)](#).

distribution.

However, when examining the possibility of dips widths ~ 10 km s $^{-1}$ and narrower, the uncertainty in our sample restricts us from making any claims regarding its presence. This was shown to be the case for both the default and the high-quality samples.

Under the circumstances, we have to take another look at our method for simulating a dearth. The model chosen pushes away all of the stars within the given domain of $-\lambda \leq v_\phi \leq \lambda$ km s $^{-1}$ around $v_\phi = 0$ km s $^{-1}$. This approach most likely shows us an exaggerated case of the dearth for each λ . In reality, we would probably observe that only a fraction of the stars within the domain have interacted with the Galactic nucleus to be scattered onto halo orbits.

Nevertheless, we do not believe this to have any larger implications for our results and therefore choose not to further develop our model. In the end, we have not been able to confirm any indication of a dip in either of our investigations of the TGAS & RAVE dataset and we have shown that the existence of the proposed dip with a width a 60 km s $^{-1}$ is in fact not present.

The claim that we can make is that a possible dip in the number of stars with zero angular momentum would be significantly narrower than the width proposed by [Hunt et al. \(2016\)](#). Furthermore, we expect that performing our tests on a more accurate dataset will ultimately show that the proposed dearth is in fact non-existent.

Chapter 4

Counter-rotating halo

The second dynamical phenomenon that we have chosen to investigate with the new distance estimates at hand, is the claim that a large portion of the halo stars in the Solar neighbourhood have counter-rotating orbits with respect to the rotation of disc stars. This was found from a study by [Helmi et al. \(2017\)](#) and the results suggest that a larger portion of the Milky Way halo stars originate from accretion than previously believed.

The trend was detected when [Helmi et al. \(2017\)](#) investigated the distribution of energy and the angular momenta component perpendicular to the Galactic plane, L_z , for a dataset consisting of observations common to the TGAS and RAVE DR5 catalogues. The sample had been reduced to only include stars classified as possible halo members by removing the sources that did not fulfil two given criteria. The first cut was made by requiring that $[M/H] \leq -1.5$ and the second cut was made by only including stars that were at least at a distance of 100 pc from the Sun.

After additional treatment of the sample, they ended up with a subset of size 1116. By studying the sample stars with low binding energies¹, it was observed that a larger fraction of them had negative L_z as opposed to positive L_z , increasingly so for lower binding energies.

[Helmi et al. \(2017\)](#) computed the fractions introducing a cut in energy, E , to the sample, only keeping the stars with low binding energies. The cuts were introduced for various values and the percentages of supposed halo stars with retrograde orbits were calculated for each of these cuts. For example, it was found that these percentages were: 57.6 % for $E > -1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$, 72.7 % for $E > -1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$ and 84.9 % for $E > -1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$.

To exclude that the uncertainty in the parallaxes used could not explain the excess of negative L_z stars for low binding energies, they employed uncertainty resampling. By creating 1000 new samples, each with a unique set of randomised errors added to the

¹Low binding energy for stars indicates that they are unusually unbound to the Milky Way.

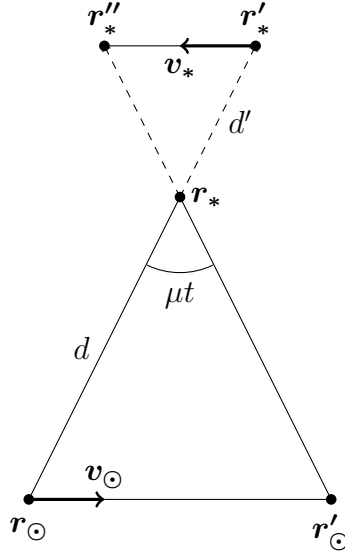


Figure 4.1: A simple diagram displaying the qualitative effect that distance overestimation has on the apparent motion of a stationary star with respect to the Sun.

distances, they found that the percentage was $57.1\% \pm 2.4\%$ for $E > -1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$ and $71.8\% \pm 4.2\%$ for $E > -1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$, which clearly is consistent with their findings. This, in turn, led to their conclusion that the observed phenomenon cannot merely be caused by errors.

Even so, given the uncertain nature of the halo star data and the vast improvement in distance estimates from [McMillan et al. \(2018\)](#), it is possible that re-examination of the halo star dynamics will show that the fractions of weakly bound halo stars on retrograde orbits have been overestimated. An overestimation of the distance to a celestial body will yield a misleading velocity when computing its corresponding three-dimensional motion.

This can be argued for with a simple scenario, which is illustrated in figure 4.1. We consider a stationary celestial object in a heliocentric coordinate system at a distance d , with an observed proper motion μ due to the three-dimensional velocity of the Sun $\mathbf{v}_{\odot} = (v_R, v_{\phi}, v_z)_{\odot}$. In a left-handed Galactocentric frame of reference the star would be at a position \mathbf{r}_* with a zero-length velocity vector (see section 1.3.4). Over the timespan t , the Sun will have travelled from its original Galactocentric position \mathbf{r}_{\odot} to its new position $\mathbf{r}'_{\odot} = \mathbf{r}_{\odot} + \mathbf{v}_{\odot}t$.

If we use the true distance in our transformation from a heliocentric frame to a left-handed Galactocentric frame of reference, the star will be in the same position \mathbf{r}_* with no apparent velocity. However, if we overestimate the distance such that we measure d' where $d < d'$, this will not be the case. The observed proper motion will still be the same for the star, resulting in an angular displacement μt on the celestial sphere from a heliocentric perspective. As a result, the star will appear to have moved in the opposite direction of

the Sun. Transformation to a Galactocentric frame will show a new position \mathbf{r}_*'' for the star, which indicates that the star has a non-zero three-dimensional velocity \mathbf{v}_* .

Consequently, an overestimation of the distance to a star will lead to a computed three-dimensional velocity relative to the Sun that is biased in the direction opposite of the Sun's motion. Since the largest velocity component of the Sun in our left-handed coordinate system is parallel to $\hat{\phi}$, the stars whose distances have been overestimated will have larger apparent velocity components in the direction of $-\hat{\phi}$. Ultimately the large negative velocities v_ϕ will yield large negative angular momenta L_z , which results in an inadequate perception of its kinematics. Since distances to halo stars in the RAVE DR5 data tend to be overestimated (McMillan et al. 2018), the fraction of halo stars with negative angular momenta is likely to decrease when re-examining the distribution using the more reliable McMillan distances.

In the re-examination of the counter-rotating halo, we have sought to answer the following questions:

- Will we see as large of a fraction of possible halo stars with negative angular momenta as in Helmi et al. (2017) when using the McMillan distances?
- How does sample size of possible halo stars change when we raise additional quality flags?
- How large of an impact does the random uncertainties and the systematic uncertainties have for the observed results?

4.1 Method

In the main strategy employed to answer the posed questions, we took one of the samples introduced in table 2.1 and applied cuts to it, yielding a subset of possible halo stars. We then computed the energy and angular momenta of each data point and could thereby obtain a sense of the distribution of negative and positive angular momenta in the subset for different energy cuts. Further, we investigated the impact of random and systematic errors in the data by employing the velocity model from section 2.2. The procedure of introducing systematic shifts in the distances to the pseudo-sample data points allowed us to determine whether or not the large fraction of negative angular momenta stars in Helmi et al. (2017) was a result of systematically overestimated distances.

After performing the same halo cuts on the default sample of data points as in Helmi et al. (2017), we were left with a subset consisting of 773 possible halo stars. As for the higher quality subsets, it was not possible to simply use our `flag_any` sample. If we perform the same cuts with `flag_any` raised, we end up with a subset of size 174. The amount of data points in this subset is too small for us to gain any reliable information from it. Again,

this follows from the fact that the observations from the TGAS & RAVE dataset that are believed to be Milky Way halo stars tend to have low $\log g$ values (see section 2.0.2).

Instead we had to opt to use the case where all quality flags except `flag_lowlogg` were raised, referred to as `flag_any-lowlogg` (see table 2.1). For this dataset 505 possible halo stars were found, a number deemed sufficiently large for our purposes. Keeping stars with `flag_lowlogg` = 1 means that we have been forced to include data points that are known to have systematically underestimated parallaxes in our re-examination of the counter-rotating halo.

4.1.1 Galactic potential and angular momentum

For our calculations of the energy, we opted to use the same Galactic potential as in [Helmi et al. \(2017\)](#) for consistency. It consists of components for the halo, disc and bulge of the Milky Way given in a Galactocentric frame of reference such that $\Phi = \Phi_{halo} + \Phi_{disc} + \Phi_{bulge}$. For the halo the following logarithmic potential was chosen:

$$\Phi_{halo} = v_{halo}^2 \ln(1 + r^2/d_c^2 + z^2/d_c^2), \quad (4.1)$$

where $v_{halo} = 173.2 \text{ km s}^{-1}$ is the circular velocity at $R \gg d_c$ with the assumed core radius $d_c = 12 \text{ kpc}$, r is the distance from the Galactic centre and z is the distance to the Galactic plane. As for the disc we used a Miyamoto-Nagai disc potential which is given by

$$\Phi_{disc} = -\frac{GM_{disc}}{\sqrt{r^2 + (a_d + \sqrt{z^2 + b_d^2})^2}}, \quad (4.2)$$

where G is the Gravitational constant, $M_{disc} = 6.3 \cdot 10^{10} M_\odot$ is the mass of the Galactic disc and $a_d = 6.5 \text{ kpc}$ as well as $b_d = 0.26 \text{ kpc}$ are scale lengths of the disc ([Binney & Tremaine 2008](#)). For the bulge we opted to employ a Hernquist bulge described by

$$\Phi_{bulge} = -\frac{GM_{bulge}}{R + c_b}, \quad (4.3)$$

where the mass of the bulge is $M_{bulge} = 2.1 \cdot 10^{10} M_\odot$ and $c_b = 0.7 \text{ kpc}$ is a scale length ([Hernquist 1990](#)). All the numerical values chosen for our three expressions correspond to the values used in [Helmi et al. \(2017\)](#).

The total energy E was computed for each star by combining the potential Φ with the specific kinetic energy $K = v^2/2$, where v is the magnitude of the three-dimensional velocity vector $\mathbf{v} = (v_R, v_\phi, v_z)$. Moreover, a constant term of appropriate dimensions, E_0 , was subtracted from the expression to avoid dealing with infinities originating from the

logarithmic nature of the halo potential. This value of this constant was chosen to be $1.75 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$ such that the energy versus angular momenta plots would be comparable to those of [Helmi et al. \(2017\)](#). This yielded the final expression for our total energy

$$E = \Phi + K - E_0. \quad (4.4)$$

The angular momentum component parallel to the z -axis for a star in a Galactocentric coordinate system is

$$\mathbf{L}_z = \mathbf{R} \times \mathbf{p}_\phi = Rmv_\phi(\hat{\mathbf{R}} \times \hat{\boldsymbol{\phi}}) = Rmv_\phi\hat{\mathbf{z}}, \quad (4.5)$$

where \mathbf{p}_ϕ is the component of the momentum vector of a star that is parallel to $\hat{\boldsymbol{\phi}}$. In our computations we use the specific angular momentum which is the magnitude of \mathbf{L}_z divided by the mass m . We have chosen to denote this quantity by L_z .

4.1.2 Analysing the halo sample

To assess the hypothesis that the large fractions of negative angular momenta stars are a direct result of the low-quality distance estimates in the TGAS and RAVE DR5 sample used in [Helmi et al. \(2017\)](#), we had to study the set of possible halo stars using statistical tests.

While we were able to visualise the distribution of energies versus angular momenta for each data point in the two sets used (`flag_dup` and `flag_any-lowlogg`) and computing the fraction of negative angular momenta stars above a given limit in total energy, there was still a lack of information regarding the true distribution. Again, this issue was solved using the uncertainty resampling method introduced in section 2.1 and the velocity model of section 2.2.

In the first test, the uncertainties in the measurements for each possible halo star were assumed to be trustworthy. This meant that a new sample could be obtained using uncertainty resampling. This allowed for computation of the fraction of halo stars with negative angular momenta above a given cut in energy for the resampled set. The process was repeated 1000 times, providing us with a mean value and a computed standard deviation for the fraction.

The second test method was based on the velocity model, where all possible halo stars were given a velocity from the Gaussian halo velocity distribution in table 2.2. By trusting the uncertainties in the quantities of each data point and using them to add a random error to each parameter, we could obtain a pseudo-sample for which the fraction could be recalculated. Reiteration ultimately led to a new mean and standard deviation.

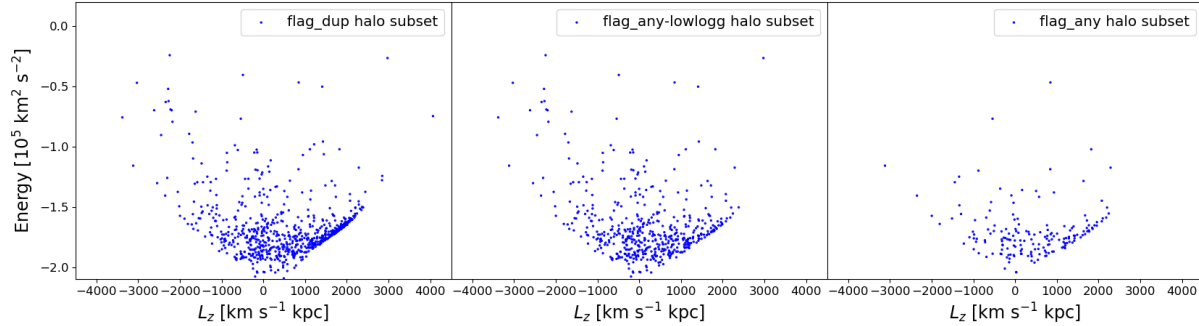


Figure 4.2: The distributions of energy vs. angular momenta for our three subsets of data. The left plot displays the distribution for the default sample, while the middle plot shows the same quantities but for the `flag_any-lowlogg` sample. The rightmost plot shows $L_z - E$ space for the `flag_any` subset.

The next step taken was to investigate how the modelled fraction changed if the distances were overestimated since this was believed to be the case for the sample used in [Helmi et al. \(2017\)](#). This was done by introducing a systematic error for all distance estimates in the modelled pseudo-sample in the form of a scaling factor k , such that our new distance for data point i was $d'_i = kd_i$.

4.2 Results and analysis

With the default and `flag_any-lowlogg` halo star samples at hand, we proceeded to compute the total energy and the specific angular momenta for each star using equations (4.4) and (4.5). The two distributions are displayed in figure 4.2. We have also included the `flag_any` halo sample in the figure, which visualises the substantial reduction in sample size caused by raising the quality flag `flag_lowlogg`. The `flag_any` halo star subset evidently lacks a large portion of the data points with low binding energies that are of interest to our investigation.

When comparing the plots for the default and `flag_any-lowlogg` sample, we observe a large density of stars in the distribution of two quantities for the default sample that appears to be missing in the corresponding `flag_any-lowlogg` plot. These stars have high binding energies and are situated in the region $L_z \in [1000, 2500]$ km s⁻¹ kpc, which indicates that some disc stars have made it through the criteria cut.

On account of the loss of stars in the high-density area, the impact that each quality flag had on the halo sample composition was examined. When excluding the `flag_lowlogg` quality indicator, which removes 331 possible halo stars, it was determined that the most significant reduction comes from `flag_N` (see table 1.1). The data points removed by this flag include dwarf stars of spectral class K, which tend to be faint. The halo stars in the TGAS & RAVE

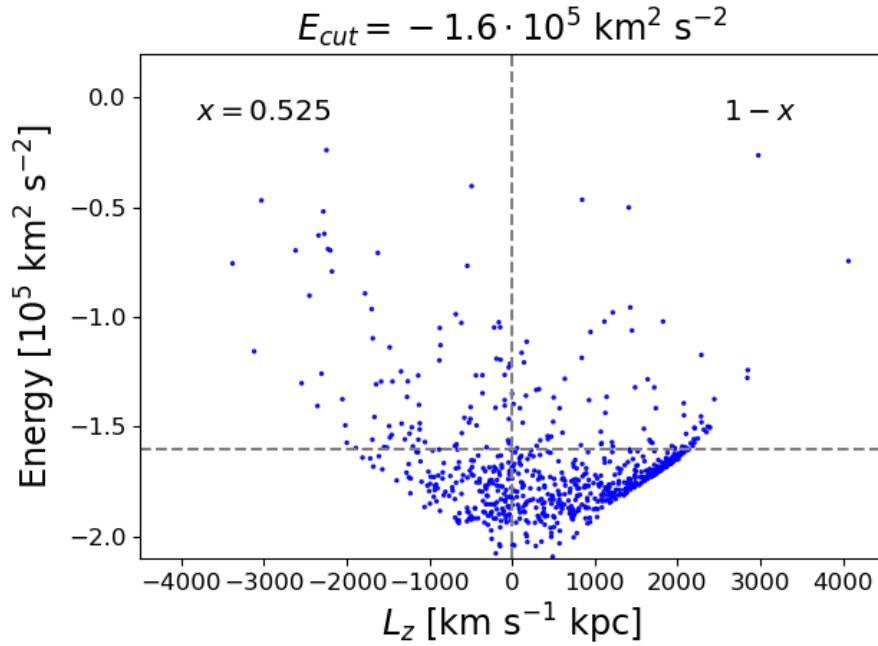


Figure 4.3: A visualisation of how the cuts were introduced in order to obtain the fractions of stars with negative angular momenta x . The given fraction in the figure was computed for the `flag_dup` sample using the uncertainty resampling method.

dataset are intrinsically bright, which provides us with an explanation of why there still are numerous halo stars with low binding energies left in our `flag_any-lowlogg` sample. This reaffirms our theory that the high-density area comes from contamination by disc stars.

4.2.1 Studying the default sample

Each statistical test was employed for three different energy cuts, which gave insight into how the percentage of counter-rotating stars varied with lower binding energies. Again, these values were chosen to coincide with the cuts used in [Helmi et al. \(2017\)](#) for consistency. The fraction of stars with a negative L_z with corresponding standard deviation, as well as the mean sample size from the resampling, $\langle N_{halo} \rangle$, was computed for each energy cut. These values are given in table 4.1 and a visualisation of a computation can be found in figure 4.3.

When comparing the resulting fractions from our resampling with the percentages found by [Helmi et al. \(2017\)](#), we observe that the fractions for the dataset with improved distance estimates are substantially lower. For the energy cuts $-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$, $-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$ and $-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$ we found that there are 52.5 %, 66.8 % and 69.5 % halo stars on

Table 4.1: The fractions of halo stars from the default dataset that have negative and positive angular momenta, found using the uncertainty sampling method.

Energy cut	Negative L_z	Positive L_z	Standard deviation	$\langle N_{halo} \rangle$
$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.525	0.475	0.023	178
$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.668	0.332	0.034	74
$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.695	0.305	0.036	60

retrograde orbits. This corresponds to a decrease of 5.1 %, 5.9 % and 15.4 % as compared to [Helmi et al. \(2017\)](#). Still, it is important to keep in mind that the halo samples used in the two surveys have different sizes (1116 in [Helmi et al. 2017](#), as opposed to our 773), which can affect the results.

Despite the variation in sample sizes, there is still an indication that the improved distance estimates had led to fewer observations of possible halo stars on retrograde orbits. That being so, we further wanted to investigate the impact of our random uncertainties and how the distribution of negative and positive angular momenta would change if systematic errors were introduced into the distance estimates. Therefore, the next step taken in the investigation was to include the [Kordopatis et al. \(2013\)](#) velocity model in the computations.

The model resampling was initially carried out without introducing the systematic errors. The computed mean fractions, found in table 4.2, show that the number of negative L_z stars still was higher than the positive counterpart. Further, the difference becomes larger for lower binding energies. This result is not to be expected from a statistical point of view. The Gaussian nature of the velocity distributions and random uncertainties of our model suggests that the numerous pseudo-samples computed should have a mean distribution of angular momenta such that ~ 50 % of the data points have negative L_z .

Consequently, we can draw the conclusion that including random distance errors in the quantities of the pseudo-sample data points shifts the percentages such that we observe a larger fraction of negative L_z halo stars. Further examination of the phenomenon revealed that the uncertainties in our dataset causing the shift towards higher percentages

Table 4.2: The fractions of halo stars from the default dataset with modelled velocities that have negative and positive angular momenta. N.B. the model implies equal probability of positive and negative L_z .

Energy cut	Negative L_z	Positive L_z	Standard deviation	$\langle N_{halo} \rangle$
$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.546	0.454	0.037	183
$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.601	0.399	0.066	55
$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.615	0.385	0.078	38

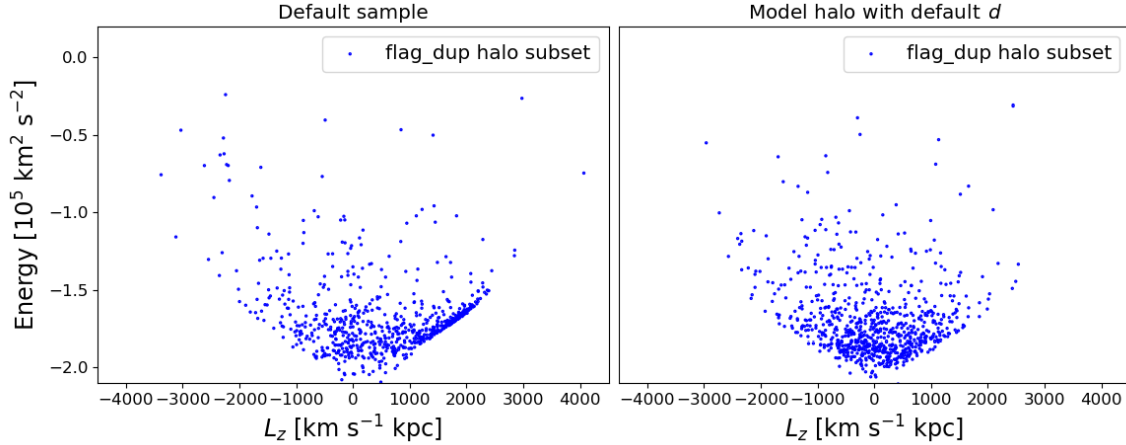


Figure 4.4: On the left, the distribution of E vs. L_z for the default sample is displayed. The right-hand plot contains the corresponding distribution for a pseudo-sample based on the same data points.

for negative angular momenta was the distance estimate errors.

When excluding the distance estimate uncertainties and adding random errors in proper motion and radial velocity to the pseudo-sample, the mean negative and positive fractions indeed aligns around $\sim 50\%$. This confirms that the added random errors that overestimate the distances can increase the observed fraction of negative L_z stars.

To elaborate, we expect a similar distribution between negative and positive errors for each pseudo-sample, since we draw errors from a univariate Gaussian distribution based on the uncertainty in our measurement. A probable explanation for the observed behaviour is that the underestimations caused by adding the negative distance errors and the overestimations caused by adding the positive distance errors have different impacts on the velocities. The reasoning is again explained in figure 4.1, where it is argued that overestimation of the distance to an object will yield a velocity that is biased in the direction opposite of the Sun’s motion.

Another point we have to discuss, is that the modelled halo distribution lacks the high-density region on the positive side of the default sample plot. This becomes more apparent in the direct comparison between the distribution of energy and specific angular momentum of the default sample and a pseudo-sample with modelled velocity, which is provided in figure 4.4.

This follows from the fact that we assume the halo subset in our sample to be solely made up of stars with halo velocities. The absence of the high-density area confirms our suspicion that it is caused by contamination of the halo sample by disc stars. After another look at figure 4.3 with this in mind, we note that a portion of the disc stars is included when performing the energy cut at $-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-1}$ on the default sample. However, we still

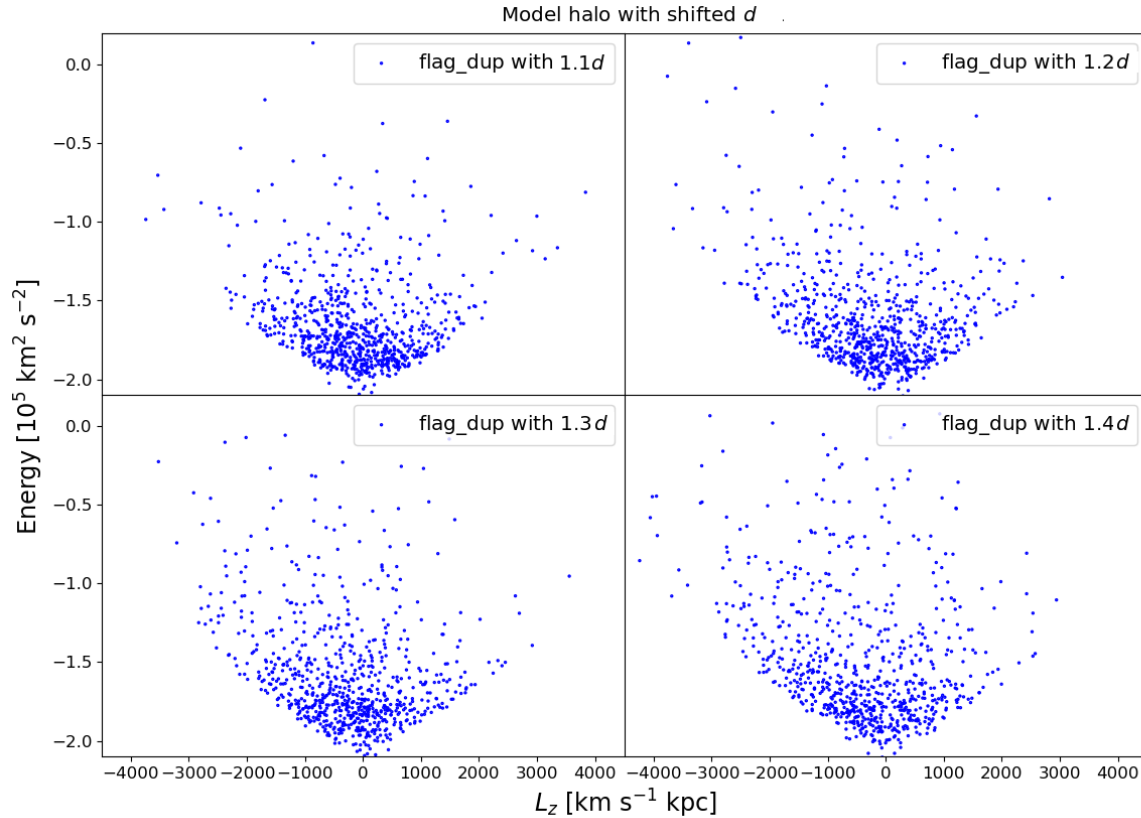


Figure 4.5: Four different cases where the distance estimates of the modelled pseudo-samples based on the `flag_dup` sample have been scaled by a factor k .

opt to use the same cuts as in [Helmi et al. \(2017\)](#) and choose to merely keep this fact in mind when discussing the results.

The random velocities drawn from the model’s Gaussian distribution are initially independent of the distance estimate, which yields fewer stars on extreme orbits in our pseudo-sample. Hence the considerable reduction in $\langle N_{halo} \rangle$. Together with the randomness of the model, this accounts for the large standard deviations in the fractions of the two more extreme energy cuts.

Nevertheless, the main purpose of the inclusion of the velocity model was as mentioned to examine how the positive and negative L_z fractions changed with systematic shifts of the distance. Hence, several scaling factors of varying magnitudes were used to overestimate the distances to each star in the pseudo-samples. Examples of distributions for the different scaling factors used are portrayed in figure 4.5. The fractions for each energy cut given a certain k are shown in table 4.3.

As expected, increasing systematic shifts of the distance estimates lead to larger observed fractions of negative specific angular momenta in the data points of the pseudo-sample.

Table 4.3: The fractions of default sample halo stars with modelled velocities that have negative and positive angular momenta. However, the distances have now been scaled by a factor k to see how the distribution changes when we include systematic errors.

k	Energy cut	Negative L_z	Positive L_z	Standard deviation	$\langle N_{halo} \rangle$
1.1	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.604	0.396	0.033	219
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.647	0.353	0.058	75
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.658	0.342	0.064	54
1.2	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.650	0.350	0.029	261
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.689	0.311	0.044	100
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.700	0.300	0.054	74
1.3	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.689	0.311	0.027	305
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.723	0.277	0.039	131
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.727	0.273	0.046	101
1.4	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.716	0.284	0.023	349
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.746	0.254	0.034	166
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.750	0.250	0.038	131

These results indicate that the large fractions of halo stars on retrograde orbits observed in [Helmi et al. \(2017\)](#) are a direct result of uncertain distance estimates.

We performed the same study for the `flag_any-lowlogg` subset, which reinforced our findings. The corresponding results can be found in Appendix B.

4.3 Discussion

In this study of the counter-rotating halo of the Milky Way, it was found that the negative L_z fractions were indeed lower for data with the improved distance estimates of [McMillan et al. \(2018\)](#). Yet it was still observed that this fraction accounts for a majority of the stars with low binding energies.

Further, the impact of the random and systematic uncertainties on the results was studied by implementing the halo component of the velocity model from section 2.2. We note that the random uncertainties of the RAVE & TGAS distances shift the distribution such that low binding energy stars tend to have a negative L_z (see table 4.2), while we would expect it to be around $\sim 50\%$ due to the Gaussian nature of our model. This speaks in favour of our hypothesis as the distances that are overestimated due to the addition of the random errors increases the observed fraction of negative L_z stars.

An analysis of the parallax errors was also carried out in [Helmi et al. \(2017\)](#) in the form of uncertainty resampling, creating 1000 new samples. The mean fractions of the distribution

in L_z - E space showed an excess of possible halo observations with negative L_z and the standard deviations were consistent with their original fractions for each energy cut. This led them to draw the conclusion that the errors cannot account for the observed excess of retrograde halo orbits.

Nevertheless, the testing was based on a sample which was already dependent on the initial distance estimates and not on a model. This indicates that any possible systematic errors in the distance estimates of the observations were still present in their new samples. For this reason, it seems that the effect displayed in figure 4.1 has not been taken into account in their error analysis.

By introducing systematic shifts to the distances of the modelled pseudo-samples, we could further validate this theory. As shown in figure 4.5 and table 4.3, larger shifts in distance corresponds to larger fractions of negative L_z for the possible halo stars. On the account of this, we argue that the large fractions of negative angular momenta for possible halo stars that were found in [Helmi et al. \(2017\)](#) have been directly caused by overestimated distances.

The tests were also performed for the higher quality dataset `flag_any-lowlogg` which also showed lower fractions of retrograde orbit halo stars than in [Helmi et al. \(2017\)](#). Furthermore, the results from the model velocity tests on the `flag_any-lowlogg` subset revealed patterns that coincided with the corresponding results for the default sample as the inclusion of random errors, as well as systematic errors, resulted in similar fractions of low binding energy stars with negative L_z which increased with larger systematic shifts of the distance.

While we have to acknowledge that our samples are unreliable due to them being dominated by stars with low $\log g$, it still stands that the results have validated our hypothesis. The described qualitative effect is without a doubt present in the data used by [Helmi et al. \(2017\)](#). However, we cannot confirm how large of an impact the systematic and random errors have had without access to a larger dataset with more precise measurements.

In conclusion, the re-examination has confirmed that the large fraction of negative L_z for possible halo stars with low binding energies that were found in [Helmi et al. \(2017\)](#) is an effect of overestimated distances. Since the samples available for our study are known to contain unreliable data points, we cannot draw any firm conclusion regarding the true fraction. However, due to the fact that we have seen that overestimated distances increases the observed fraction of negative L_z , we suggest that a re-investigation using a more reliable dataset will prove that the percentage will be in the range of 50-60 % for all three energy cuts used in our study.

Chapter 5

The wobbly disc

The third and final phenomenon that has been examined in this review of Milky Way dynamics are the proposed wave-like motions of the Galactic disc. In their investigation of the kinematics of solar neighbourhood disc stars, [Williams et al. \(2013\)](#) found that they have non-zero velocity components in R and z . They also determined that there are non-negligible velocity gradients present in every dimension, suggesting complex dynamical behaviour of the disc.

The study was based on an earlier data release of RAVE data and does not include the distance estimates from DR5. Furthermore, the parallaxes or proper motions from TGAS were not available at the time it was conducted. Hence the distance estimates utilised in the survey are of a different origin.

Instead of parallaxes, [Williams et al. \(2013\)](#) employed photometric distance estimates based upon measurements of so-called red-clump (RC) giants, which were part of the fourth data release of RAVE (RAVE DR4). These giants have solar-like metallicities and reside in a clustering of the horizontal branch in the Hertzsprung-Russell diagram, which is known as the red clump ([Karttunen et al. 2017](#)).

The use of the RC distances leads to a restriction in the data points available to the study. In other words, [Williams et al. \(2013\)](#) only examined the kinematics of RC stars. Due to the improved proper motions and distance estimates as well as the larger variety of sources available to us, we saw a reason to re-examine the results found in the paper.

As mentioned, the findings related to v_z shows that the velocities have different parities above and below the Galactic plane (section 1.4). The stars appear to be moving with speeds of $|v_z| \sim 5 \text{ km s}^{-1}$ in opposite directions (see figure 13 in [Williams et al. 2013](#)). This produces a wave-like behaviour of the disc stars which is known as a breathing mode. We have studied the velocities for the TGAS & RAVE data in order to determine whether we will detect the same types of motion patterns for a more reliable set of data with better distance estimates.

When it comes to the v_R component, the disc was found to display a behaviour that differs significantly inside and outside of the solar circle (see figure 12 in Williams et al. 2013). Within the radius of the Sun, $r_\odot = 8.20$ kpc, the disc stars that are situated a few hundred pc above or below the plane are moving outwards with velocities around $v_R \sim 8$ km s⁻¹. Around $z = 0$ pc, the velocities are lower but have the same sign. Outside of the solar radius, the stars instead appear to be moving inwards with velocities around $v_R \sim -5$ km s⁻¹. However, the velocities fluctuate and reach magnitudes of $v_z \sim 15$ km s⁻¹. The region in z for which the stars are showing this behaviour seems to grow wider with increasing radius and reach a maximum around $R = 9$ kpc.

It is noteworthy that a paper with the same aim was published a few months into our project (Carrillo et al. 2018). The published re-examination has made a thorough investigation and computed both velocity gradients and median velocities for the TGAS & RAVE dataset, taking various sources of errors into account. The results imply that the local parts of the disc do not exhibit two breathing modes, but one breathing mode and one bending mode, where stars move in the same direction above and below the Galactic plane. They have opted to use the `flag_lowlogg` cut when treating their data as well as a cut on uncertain data points with distance uncertainties larger than 25 %. They also included additional cuts based on the RAVE data. On account of this, there is still information to be gained by performing the computations for TGAS & RAVE subsets with different quality flags raised.

That being so, our study will mainly focus on the impact of improved distance estimates on the computations and how the random errors of our measurements will affect the results. To be more specific, we have aimed to answer the following questions in our study:

- Will we find the same kinematic patterns in R and z when using different subsets of the TGAS & RAVE data?
- Will we detect differences in the dynamics observed for low-quality and high-quality subsets?
- How much will the velocities v_z and v_R of a model pseudo-sample fluctuate when including random errors?
- What happens if we introduce systematic shifts to the distance estimates?

5.1 Method

The re-examination of the velocities of the solar neighbourhood stars was done in the following manner. First, we applied cuts to a given TGAS & RAVE subset, which allowed us to obtain a sample of disc stars. We subsequently computed the velocities for the disc star sub-sample in R and z . The median velocity component for a given region in the disc was then determined and plotted in a two-dimensional velocity map, which allowed us to

analyse the motions in the disc. The velocity model from section 2.2 was also introduced to create pseudo-samples with velocities independent of the distance estimates. The velocity maps for the mean modelled median velocities and the corresponding standard deviations gave insight into the fluctuations caused by the uncertainties of the sample.

The TGAS & RAVE dataset is fundamentally different from the RAVE DR4 used in Williams et al. (2013). Accordingly, it does not make sense to replicate the same cuts on the data to obtain our disc star sub-sample. However, we adopt the same cuts on position such that we investigate the same region of the Milky Way. Hence, we only include the stars in our chosen TGAS & RAVE subset that fulfil the two criteria: $-2 < z < 2$ kpc and $5 < R < 10$ kpc. For our default sample this leaves us with a sub-sample of 191 754 stars and for the `flag_any` sample we obtain a sub-sample of size 137 090.

5.1.1 Computing the velocity field

In order to obtain the velocity field for a given component, we divided the chosen region into bins with the area $(0.2 \text{ kpc})^2$. The median value for all velocities within each bin with a minimum of 50 counts was then determined. The resulting velocity distribution was then portrayed in the form of a two-dimensional velocity field map.

5.1.2 Including the velocity model

Given the uncertainties of the measurements and distance estimates of TGAS & RAVE, we had to determine whether or not they could explain any unexpected behaviour of the disc. By assessing the influence of the errors in the sample, we could gain a sense of the validity of our findings.

The approach chosen was to once more introduce the velocity model (see section 2.2). As mentioned it allows for the creation of a pseudo-sample that has velocities with known properties. Despite the fact that our TGAS & RAVE subset was reduced to only contain stars fairly close to the disc, our boundaries are sufficiently large to include stars on halo orbits. Therefore we included same relative weights as in table 2.2, meaning that circa one percent of our “disc” sample will have halo velocities. By adding random errors to each parameter, based on the TGAS & RAVE uncertainties, we could obtain different pseudo-samples with varying velocity fields. An example of such a pseudo-sample, which portrays a modelled velocity field for the v_R component can be found in figure 5.1.

Repetition of this method yielded a large number of samples, each with varying median v_R and v_z distributions. The corresponding arithmetic mean median values and counts in each bin for all samples were then computed to once more exclude the bins with on average less than 50 counts.

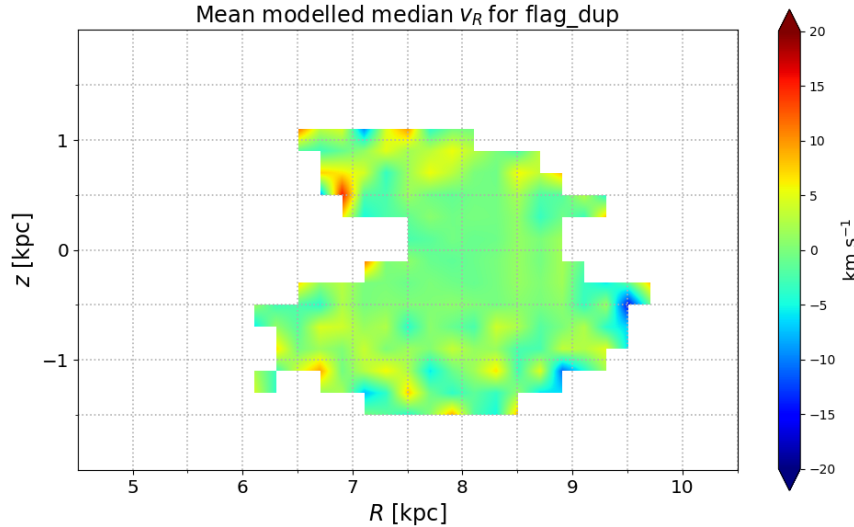


Figure 5.1: An example median v_R distribution for a pseudo-sample created from the velocity model. It has been based on the observations and uncertainties of the default sample.

To get a sense of the variations between the distribution for the many pseudo-samples, we opted to compute the standard deviation in each velocity bin using equation (2.1).

5.2 Results and analysis

We have produced maps in R and z for the median values of v_R and v_z by computing the velocity field and its corresponding modelled field with standard deviation for the default and `flag_any` samples. For each component, the two sets of plots are displayed in the same figure to allow for easy comparison.

5.2.1 The median v_R field

We began by investigating the v_R distribution for our main sample. We also generated 100 pseudo-samples with the velocity model and determined the mean, as well as standard deviation for the set. The plots portraying the measured velocity field, the mean velocity field for the pseudo-samples and the corresponding standard deviation are provided in the upper row of figure 5.2. We find many similarities between the velocity field for our TGAS & RAVE subset and the corresponding field computed by [Williams et al. \(2013\)](#). The pattern where stars within a few hundred pc of the disc are moving radially outwards inside the solar radius and inwards outside the solar radius can clearly be distinguished.

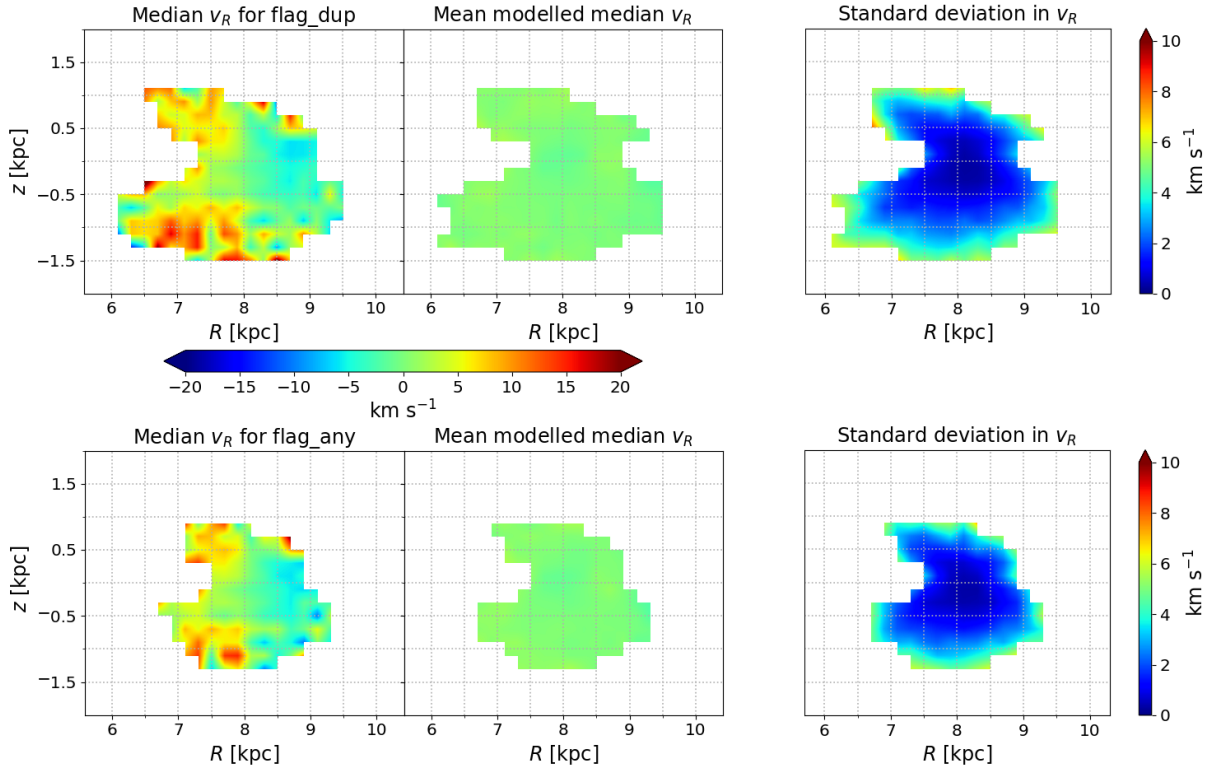


Figure 5.2: To the upper left is a map of the median velocity field in R for the default sample divided into $(0.2 \text{ kpc})^2$ bins. The upper middle plot displays the mean median velocities for 100 pseudo-samples created from the velocity model. The upper right plot displays the standard deviations of the mean modelled median v_R velocities. The bottom row shows the same plots but for the `flag_any` sample.

However, for stars situated outside $|z| \sim 0.5 \text{ kpc}$ we observe that the median velocities indicate that they are moving outwards, both for $R < r_\odot$ and $R > r_\odot$. Moreover, the median velocities appear to be slightly larger for the regions where $R < r_\odot$ and $z < -0.75 \text{ kpc}$ for our sample, as it reaches $v_R \sim 17 \text{ km s}^{-1}$ near the edges of the map. For comparison, the velocities in the same disc regions of [Williams et al. \(2013\)](#) reach magnitudes of $v_R \sim 8 \text{ km s}^{-1}$.

To identify any possible troublesome regions of our map we study the corresponding standard deviation. The fluctuations in observed velocity increase with distance from the Sun in the plot as the standard deviation goes from nearly zero within 250 pc of the Sun to $\sim 7 \text{ km s}^{-1}$ near the edges of the map. Hence, while some of the velocities farther from the Sun may be over- or underestimated, the overall behaviour observed is inconsistent with the sample error fluctuations.

When investigating the v_R distribution for the `flag_any` subset, which together with the modelled median velocity field and the standard deviation is provided in the bottom row

of figure 5.2, we observe that both velocity maps are significantly smaller in comparison to the maps for the default sample in the upper row. Hence, we can deduce that the edges of the v_R map for the default sample contained many observations with large distance estimate uncertainties.

Despite the reduced number of bins with more than 50 counts, we can still identify the same kinematic patterns as in the corresponding plot for the default sample. Even so, the median velocities of observations within the solar radius that are close to the Galactic disc appear to be lower for our more reliable subset.

The mean modelled median velocity field in the right plot remains smooth for both the lower and higher quality samples, which is to be expected from the modelled distribution since the velocity model implies an equal probability to get positive and negative velocities (see section 4.1.2). Furthermore, the standard deviation shows a similar behaviour as for the default sample. The map shows larger amplitudes near its edges, which reach values of up to 5 km s^{-1} and very small standard deviations in the vicinity of the Sun. We conclude that the standard deviation does not indicate that any of the observed outward and inward flows are due to uncertainty in the measurements of TGAS & RAVE.

In the end, there are no grounds to further investigate the dynamics along $\hat{\mathbf{R}}$. The lack of a reliable sample with many data points puts a restriction on the claims that we can make regarding the dynamical behaviour of the disc farther away from the Sun. We can, however, reaffirm the findings of Williams et al. (2013) as the kinematic patterns of our investigation coincides with the corresponding regions in plot a) of figure 12 in their paper.

5.2.2 The median v_z field

Figure 5.3 shows the resulting plots for v_z . Starting with the measured distribution for the default sample, which is the upper left velocity field map, we can identify the rarefaction behaviour of observations within the solar circle suggested by Williams et al. (2013). The stars above the Galactic plane appear to be moving upwards while the stars below it appear to be moving downwards. Even as positive median velocities of notable magnitudes are present in the lower left part of the map, the region as a whole appears to have a net motion away from the disc. The region of the upper left corner of the map exhibits similar behaviour as there are some bins with large negative velocities while the overall net motion of the stars appears to be in the positive z direction.

It is more difficult to make out the proposed compression of the disc outside of the solar circle. While we observe that the median velocities of the lower right corner suggest a motion towards the Galactic plane, the upper right region does not display any apparent net motion in the negative z direction. Investigating the standard deviations from the modelled pseudo-samples, we find that the random errors fluctuate with values around $\sim 3 - 4 \text{ km s}^{-1}$ by the outer edges of the map. This renders us unable to confirm either the

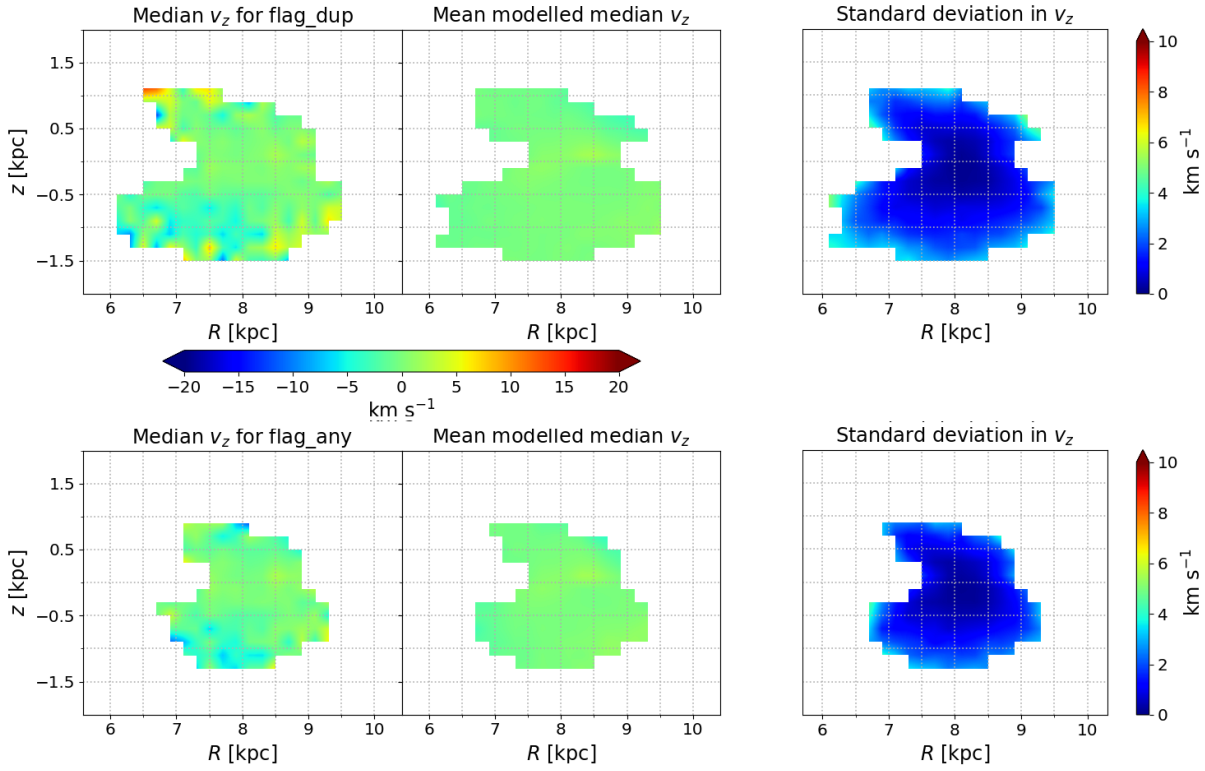


Figure 5.3: Same as in figure 5.2 but for v_z .

rarefaction or contraction behaviours of the disc. It is clear that a higher quality sample is required.

Therefore, we proceeded by reviewing the patterns in the z direction velocity for the sample based on the `flag_any` subset. The measured and modelled median velocity fields, as well as the standard deviation, can be found in the lower row of figure 5.3. The cut of uncertain observations has again shrunk the velocity map and revealed new features. The topmost part of the map, at a radius of 8 kpc shows a median velocity that is more negative than the corresponding velocity for the default sample. Besides this discrepancy, the amplitudes are generally smaller.

We observe that the rarefaction behaviour of the regions within the solar circle is prominent for the `flag_any` sample. However, it is still difficult to find any tendencies of the map that supports a contraction of the disc. The small size of the map restricts us from drawing conclusions other than that the regions below the Galactic disc at radii greater than 8.5 kpc exhibit an upwards flow. We note that the standard deviation reaches values of 3 - 4 km s^{-1} near the edges, which indicates that some of the larger median velocity magnitudes may be overestimated. However, the standard deviation due to random errors cannot account for the observed behaviour alone.

Unexpectedly, we find amplitudes in the upper right region of the mean modelled median velocities for both samples that deviate from the otherwise smooth distribution (see middle column of figure 5.3). We also detect a similar feature in the lower left corner around $R = 6.75$ kpc and $z = -0.75$ kpc. Further, these properties can also be distinguished in the left column with the measured velocity fields. This ambiguity resembles the situation in chapter 4 where it was found that overestimation of distances increases the apparent amount of stars on retrograde orbits.

Therefore, we opted to use the same approach as in 4.1.2, where systematic shifts of the distance estimates were introduced in an attempt to explain the observed phenomenon.

5.2.3 Systematic distance shifts

In order to investigate the effect of systematic shifts, we again simply scaled the distances for the velocity model with a factor $k > 1$ to gain a new larger distance $d' = kd$. We chose the values for k such that $k \in [1.1, 1.2, 1.3, 1.4]$. The resulting model velocity maps for the `flag_any` sample are displayed in figure 5.4. The corresponding standard deviations are found in figure C.1 of Appendix C.

As the distance is shifted by factors of increasing size, we see that the observed fluctuations grow into distinct patterns. When the distance is scaled by a factor $k = 1.4$ we can identify distortions outside of the solar circle that converges on the position of the Sun. Above the Galactic plane, the stars appear to be moving downwards and below the Galactic plane, the stars exhibit a net motion upwards. Inside of the solar circle, the flows are less explicit, but we still observe patterns with smaller amplitudes where there is an upward flow above and downward flow below the Galactic plane.

The detected dynamical motions for systematically shifted distances thus coincide with the wave-like behaviour found in [Williams et al. \(2013\)](#). Inside of the solar circle, we see the indications of a rarefaction pattern while the disc appears to be contracting outside of it.

The case of v_R

Due to our findings, it was in our interest to carry out the same simulations for our modelled v_R fields. We used the same scaling factors to obtain four cases which are portrayed in figure 5.5. The corresponding set of standard deviations is found in figure C.2

In the figure, we observe that systematic shifts produce patterns similar to what we see in figure 5.2. This explains why some of the features in the median v_R fields of the default sample are removed or less prominent when using the higher quality data. However, the effective increase in amplitude that could be caused by a realistic systematic overestimation of distances is not enough to account for the observed outwards/inwards flows

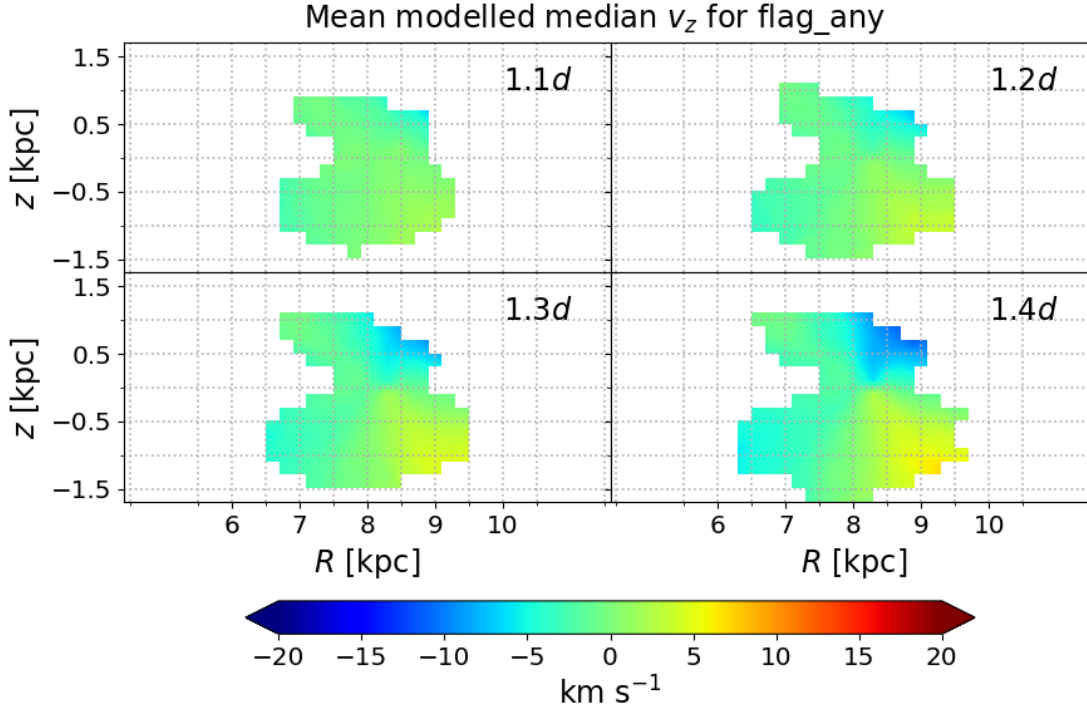


Figure 5.4: Four different cases of mean modelled median velocity fields in z that display the impact of overestimated distances. Each plot contains a computed mean model median v_z field for distances scaled by a factor k .

inside/outside of the solar circle which was found in [Williams et al. \(2013\)](#) and in this paper. Hence, we did not further investigate this discovery.

5.3 Discussion

Despite not being able to trust the outer edges of the computed velocity maps due to large uncertainties, we were able to reaffirm the proposed dynamical behaviour in v_R for both the default and the `flag_any` sample. It was shown that systematic shifts would not change the apparent patterns in the radial motion of the disc, which is a plausible reason why we did not detect any divergence in v_R from the results of [Williams et al. \(2013\)](#) despite the large differences in the certainty of the data used in the respective studies.

As for the v_z component fields, we saw a more complex distribution. The suggested rarefaction-like behaviour of the disc inside of the solar circle was prominent for both our datasets and the pattern became more explicit when removing data points with large uncertainties. However, we could not confirm the contraction pattern outside of r_\odot . It was also noticed that the random error fluctuations found with the velocity model only could account for over- or underestimations of the observed velocities and could not provide an

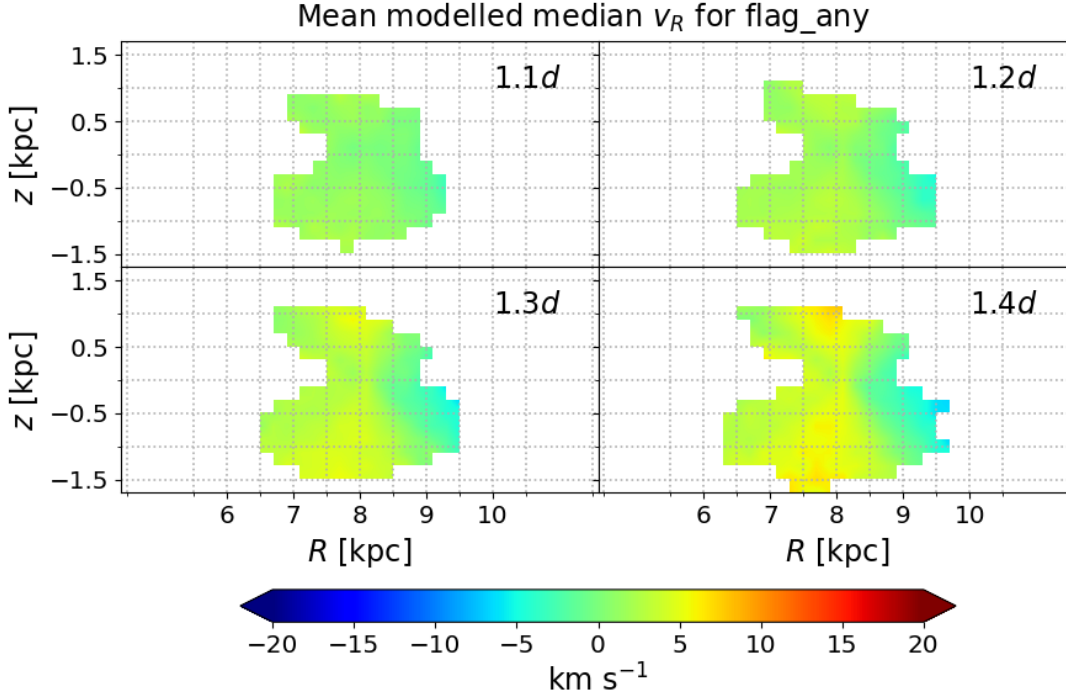


Figure 5.5: Same as in figure 5.4 but for the v_R component.

explanation for any of the patterns seen.

Considering the significant improvement in median fractional distance uncertainty for the McMillan distances, we argued that large uncertainties and/or systematic errors in the distance estimates is a likely explanation for the lack of the contraction-like feature in the TGAS & RAVE median v_z field (given it was found by [Williams et al. 2013](#)). This is supported by figure 5.4, which show that the systematic overestimation of the distances causes distortions outside of the solar circle which produce an apparent contracting behaviour. The phenomenon emerges for relatively small shifts where the distances have been overestimated by merely a few percent.

We note that these results were consistent with the findings of a similar experiment in [Carrillo et al. \(2018\)](#) (see their figure 7). Instead of introducing systematic shifts, they introduced large artificial random errors to their modelled star sample distances, such that $\sigma_d/d = 0.8$. To gain some clumping in the map, they also introduced an error in proper motion of $\sigma_\mu = 5 \text{ mas yr}^{-1}$, which produced a velocity field similar to that of figure 12 in [Williams et al. \(2013\)](#).

The origin of the systematically overestimated distance estimates of [Williams et al. \(2013\)](#) can be argued for from their choice of sources. The RC stars are giants and thereby have low $\log g$ values, which we know to have systematically underestimated distances in RAVE. In combination with the set of proper motions used which are inferior in quality to the TGAS

counterpart, this makes for a dataset that is known to carry large uncertainties.

It is therefore likely that the combination of random and systematic errors has led to the seemingly faulty conclusion regarding the dynamical features due to non-zero v_z components of disc stars. However, the low-quality dataset has not led to a misleading picture of the motions along R in their results. This is consistent with the findings in figure 5.5, where we observe that systematic shifts in d induce patterns similar to what we see in our measurements (leftmost column of figure 5.2).

Despite being able to detect the rarefaction motion in v_z and reaffirm the flows in v_R , we are unable to draw any other conclusions regarding the disc dynamics in this study. Mainly because of the limitations of our dataset. While we have used a minimum of 137 090 stars within the range $-2 \leq z \leq 2$ kpc and $5 \leq R \leq 10$ kpc, we can barely provide information on a smaller fraction in the region $-1.3 \leq z \leq 0.8$ kpc and $7 \leq R \leq 9$ kpc. Therefore, even if we chose to trust our distances and uncertainties to draw conclusions, any observed patterns outside of the solar circle in our small `flag_any` map are difficult to justify.

The fact that we did not find the bending mode pattern in v_z outside of the solar circle, that was suggested by Carrillo et al. (2018), can be attributed to the difference in data cuts. While they did not use any other quality flags available in TGAS & RAVE than `flag_lowlogg`, they performed a thorough selection of RAVE data and excluded observations with large relative errors in proper motion and distance.

To summarise, the study of the disc dynamics have further argued for the importance of precise astrometric data. We have found that observations in larger parts of the Galactic disc region studied in Williams et al. (2013) are unreliable. While we have reinforced the findings of an outward/inward flow in the median v_R field of the solar neighbourhood and a rarefaction behaviour in the v_z field inside of the solar circle, the comparably small size of our map restricts us from making further assessments of the dynamics involved. However, we have shown that systematic shifts in the distance estimates and random errors of the sample can cause misleading patterns in the v_z field and amplify the magnitudes of the median v_R values found in the observed flows. Nevertheless, it is clear that a larger high-quality dataset than what was available for this study is needed to properly assess the complex disc dynamics and its origins. With the data releases to come from *Gaia*, we will likely be able to understand its wave-like motions before long.

Chapter 6

Conclusion

The review of the papers chosen for this study has looked at large-scale motions of the three main substructures of the Milky Way. In chapter 3, we have evaluated the existence of a dearth of zero angular momentum stars which is linked to the interaction of stars with the Galactic nucleus. In chapter 4, we have discussed the proposed excess of halo stars on counter-rotating orbits. Lastly, we have studied the dynamics of the local disc in chapter 5. Summaries and brief discussions for the three investigations can be found in sections 3.3, 4.3 and 5.3, respectively.

Each of the investigated dynamical phenomena has been related to stars on fundamentally different orbits in the Galaxy, which has provided insight into the strengths as well as the shortcomings of the TGAS & RAVE dataset. There has been a general trend throughout the three studies where the small number of reliable data points in the dataset has impaired our ability to assess the quantitative aspects of the related dynamics. This was particularly apparent in the study of chapter 4. As mentioned in section 4.3, the small subsets of possible halo stars for the two samples used (sizes 773 and 505) hindered us from estimating the true fractions of observations on retrograde orbits.

This can also be attributed to the known uncertainty in the measurements of low $\log g$ stars and the systematic overestimation of their distances in RAVE (see section 1.2.2). Due to most of the possible halo stars being giants, we were forced to use observations that we knew to be unreliable in order to obtain a set of stars that was sufficiently large for statistical testing.

Nevertheless, this argument brings forth a major asset of TGAS & RAVE: even though the dataset contains many observations with unreliable physical quantities and overestimated distances, the quality flags tells us exactly which these are. On account of this, we have been able to effectively use various combinations of quality flags to create subsets with known shortcomings. Furthermore, by excluding observations that demonstrate any of the known issues we have obtained the `flag_any` subset, which is the highest quality version of the catalogue (see section 2.0.2). The results obtained using `flag_any` has when possible

served as a comparison with results from less reliable samples to make us aware of how lesser quality data affects the results.

In combination with statistical testing and the [Kordopatis et al. \(2013\)](#) velocity model, TGAS & RAVE has therefore proven to be a simple, yet powerful tool in the qualitative assessment of dynamical phenomena in the substructures of the Milky Way. The statistical resampling methods of section 2.1 that have been employed in the re-examination has allowed us to bypass the lack of data points in TGAS & RAVE and provided a sense of how the sample properties compare to the true population. The velocity model further aided the investigation by enabling the creation of pseudo-samples with velocities independent of the distance estimates as argued for in sections 2.2 and 4.3.

Using these resources we have shown that all three papers chosen for this study have produced results that have been affected by large uncertainties and/or systematic errors in their distance estimates. Firstly, the dearth of stars with zero angular momentum proposed in [Hunt et al. \(2016\)](#) could not be detected when employing better distance estimates. Secondly, the large fractions of possible halo stars on retrograde orbits found in [Helmi et al. \(2017\)](#) was shown to be directly related to systematically overestimated distances. Finally, the contracting motion in z of the disc stars outside of the solar circle detected by [Williams et al. \(2013\)](#) was also shown to be caused by overestimated distances.

These findings emphasise the importance of new catalogues based on precise instruments such as *Gaia* for the field of Milky Way research. However, we recognise that the data employed in both [Hunt et al. \(2016\)](#) and [Helmi et al. \(2017\)](#) are based on stars common to the TGAS and RAVE catalogues, yet we have still found that their data gave misleading results. The explanation lies in their choice to use DR5 distances when available and TGAS for the remainder of the data points. Hence, it is apparent that they have failed to take into account the effects of possible systematic errors and large uncertainties present in their distances.

Consequently, we conclude that while further development of instruments is a necessity to gain a deeper understanding of Milky Way dynamics, the proper treatment of astrometric measurements is all the more so. The work that has been done in [McMillan et al. \(2018\)](#), as well as the findings of this paper, has not only given reason to re-examine other previous findings where inferior distance estimates have been employed, they also make us acknowledge the need for careful analysis of distance estimates based on future catalogues.

It is clear that re-examinations of prior work will be made easier with future releases of *Gaia* data. The number of observations available will see an enormous increase and the certainty in measurements of astrometric properties will also be improved. This will allow for deeper examination of the dynamical features investigated in this paper as well. Moreover, the use of a larger and more accurate dataset will make it possible to find good estimations regarding the quantitative questions left unanswered in this study, such as: can we exclude a dearth of width of 10 km s^{-1} or lower?; how large is the fraction of the low binding energy halo stars on retrograde orbits?

At the time of writing, the second data release of *Gaia* has been made public, which contains 1 331 909 727 observations with 5-parameter astrometry (GDR2: [Gaia Collaboration et al. 2018a](#)). This catalogue will make TGAS & RAVE obsolete due to a vast improvement in measurement accuracy of proper motions and parallaxes, as well as its complementary survey of radial velocities.

The potential of GDR2 when it comes to studies of Milky Way dynamics has already been displayed in the performance verification papers which were part of the release. The paper [Gaia Collaboration et al. \(2018b\)](#) includes an analysis of the kinematic patterns in the disc, which have been discussed in this paper. The velocity maps computed using data from GDR2 cover a much greater region of the local disc, approximately in the radial range $4 \leq R \leq 13$ kpc spread out above and below the Galactic plane such that $-3 \leq z \leq 3$ kpc (see their figure 11).

Their findings coincides with our maps for v_R in figure 5.2. Furthermore, they also confirmed the rarefaction behaviour of the v_z field inside the solar circle, which we detected in chapter 5, as well as the bending mode outside of the solar circle which was suggested by [Carrillo et al. \(2018\)](#). The fact that we were unable to see any indications of a bending mode despite using the same distance estimates as [Carrillo et al. \(2018\)](#) emphasises the importance of careful data selection. Albeit we employed the most reliable dataset according to the quality flags of TGAS & RAVE - the `flag_any` sample - it is evident that some of its data points still carry large relative uncertainties. If we had opted for a similar approach as in [Carrillo et al. \(2018\)](#) and done a more thorough analysis of the subset, removing the data points with large relative uncertainties in distance and proper motion, we would likely have seen kinematic patterns that were more consistent with the true behaviour of the disc.

To conclude, we have shown that uncertain distance estimates have large implications on detected dynamical patterns related to the three main substructures in the Milky Way. That being so, it is a necessity that the GDR2 5-parameter astronomy catalogue is made subject to similar treatment as done in [McMillan et al. \(2018\)](#). While there already are known limitations of GDR2 as reported in [Gaia Collaboration et al. \(2018a\)](#), there might be shortcomings that are yet to be detected that will lead to a flawed understanding of the large-scale motions in our Galaxy. Nevertheless, in due time the GDR2 catalogue and its successors will without a doubt be able to unravel the complex dynamics of the Milky Way and provide explanations to the many questions we have yet to answer.

Bibliography

- Astropy. 2018, Description of Galactocentric coordinates transformation, <http://docs.astropy.org/en/stable/coordinates/galactocentric.html#coordinates-galactocentric>, [Online; accessed 2018-04-05]
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- Beers, T. C., Carollo, D., Ivezić, Ž., et al. 2012, *ApJ*, 746
- Binney, J., Gerhard, O., & Spergel, D. 1997, *MNRAS*, 288, 365
- Binney, J. & Tremaine, S. 2008, in *Galactic Dynamics*, 2nd edn. (Princeton: Princeton University Press), pp. 72–78
- Burnett, B. & Binney, J. 2010, *MNRAS*, 407, 339
- Carollo, D., Beers, T. C., Lee, Y. S., et al. 2007, *Nature*, 450, 1020
- Carrillo, I., Minchev, I., Kordopatis, G., et al. 2018, *MNRAS*, 475, 2679
- Efron, B. 1979, *The Annals of Statistics*, 7, 1
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018a, ArXiv e-prints, arXiv:1804.09365
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016a, *A&A*, 595
- Gaia Collaboration, Katz, D., Antoja, T., et al. 2018b, ArXiv e-prints, arXiv:1804.09380
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016b, *A&A*, 595
- Helmi, A., Veljanoski, J., Breddels, M. A., Tian, H., & Sales, L. V. 2017, *A&A*, 598, A58
- Helmi, A., White, S. D. M., de Zeeuw, P. T., & Zhao, H. 1999, *Nature*, 402, 53
- Hernquist, L. 1990, *ApJ*, 356, 359
- Hunt, J. A. S., Bovy, J., & Carlberg, R. G. 2016, *ApJ*, 832, L25

- Karttunen, H., Kröger, P., Oja, H., Poutanen, M., & Donner, K. J., eds. 2017, *Fundamental Astronomy*, 6th edn. (Berlin: Springer-Verlag)
- Kordopatis, G., Gilmore, G., Wyse, R. F. G., et al. 2013, *MNRAS*, 436, 3231
- Matijević, G., Zwitter, T., Bienaymé, O., et al. 2012, *The Astrophysical Journal Supplement Series*, 200
- McMillan, P. J. 2017, *MNRAS*, 465, 76
- McMillan, P. J., Kordopatis, G., Kunder, A., et al. 2018, *MNRAS*, 1016
- Michalik, D., Lindegren, L., & Hobbs, D. 2015, *A&A*, 574, A115
- Perryman, M. 2012, *European Physical Journal H*, 37, 745
- Reid, M. J. & Brunthaler, A. 2004, *ApJ*, 616, 872
- Schneider, P. 2006, in *Extragalactic Astronomy and Cosmology*, 2nd edn. (Berlin Heidelberg: Springer-Verlag), pp. 46–54, 66–67, 134
- Schönrich, R., Asplund, M., & Casagrande, L. 2011, *MNRAS*, 415, 3807
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, 403, 1829
- Siebert, A., Famaey, B., Minchev, I., et al. 2011, *MNRAS*, 412, 2026
- Steinmetz, M. 2003, in *GAIA Spectroscopy: Science and Technology*, ASP Conference Proceedings, Vol. 298, held 9-12 September 2002 at La Residenza del Sole Congress Center, Gressoney St. Jean, Aosta, Italy. Edited by Ulisse Munari. ISBN: 1-58381-145-1, 2003, p.381, Vol. 298, 381
- Williams, M. E. K., Steinmetz, M., Binney, J., et al. 2013, *MNRAS*, 436, 101

Appendix A

Re-investigation of modelled dip for higher quality data

It is noteworthy that the employed velocity model does not pose a good fit for the measured distribution of v_ϕ for `flag_any`. Most halo stars are, as mentioned in section 1.3.3, removed when raising `flag_lowlogg`. Hence, we saw reason to perform additional resampling tests using both methods on the `flag_any-lowlogg` subset, for 5 and 10 km s⁻¹ wide bins.

We cannot observe any apparent changes when introducing dipoles to the modelled pseudo-samples obtained from `flag_any`. Comparison of the `flag_any` resampling results in figure 3.6 with modelled distributions of 10 and 20 km s⁻¹ found in figure A.1, indicates that it should be possible to detect a dip of width 20 km s⁻¹, which is not the case for our bootstrapped and uncertainty sampled results. We observe that we are still not able to say anything regarding the presence of a dip with a width of 10 km s⁻¹.

Hence, despite this supplementary analysis of the 10 km s⁻¹ wide dip for the less uncertain data sample it was found that there is no way to draw any conclusion regarding its existence in the resampling histograms. It suffices to look at the zoomed in histograms of the modelled velocity distribution, provided in figure A.2, to deduce that the random errors of our data restrict further assessment.

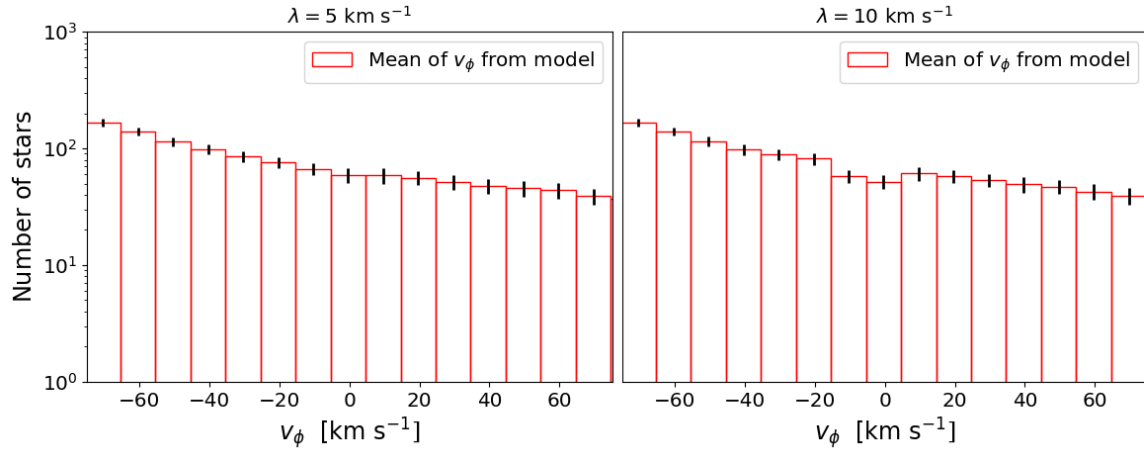


Figure A.1: The results from computing the tangential velocity using the model with bin widths of 10 km s^{-1} , zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. The introduced dips in the plots are 10 km s^{-1} and 20 km s^{-1} , from left to right respectively.

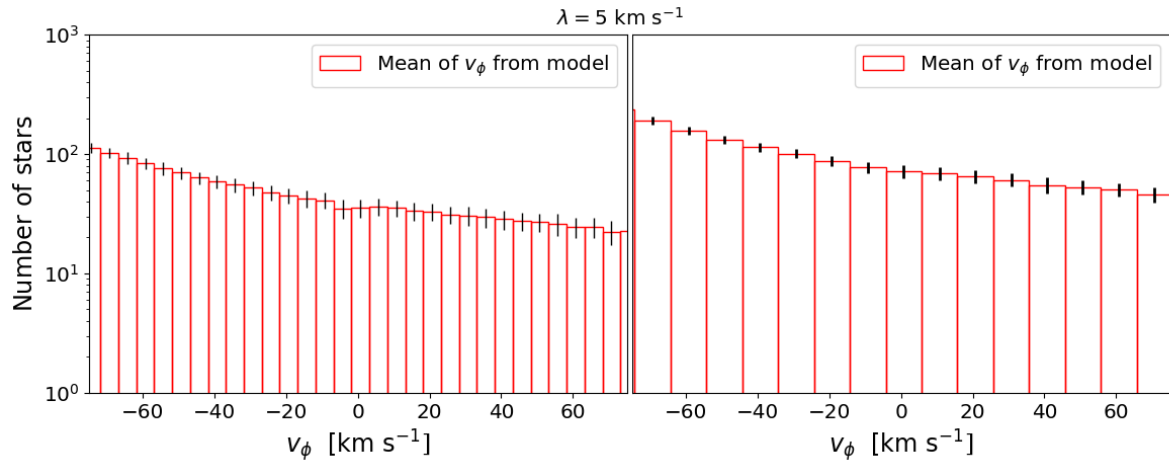


Figure A.2: The results from computing the tangential velocity using the model with a simulated dip of width 10 km s^{-1} , zoomed in around $v_\phi = 0 \text{ km s}^{-1}$. The bin widths of the plots are 5 km s^{-1} and 10 km s^{-1} , from left to right respectively.

Appendix B

Counter-rotating halo for flag_any-lowlogg sample

The statistical testing of the energy and angular momenta distributions of low binding energy observations was also performed for the `flag_any-lowlogg` dataset. As shown in table B.1, the computed fractions of counter-rotating stars using uncertainty resampling are larger than for the default sample.

Regarding the energy cut $E > -1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$, the increased percentage can be attributed to the loss of disc stars from the aforementioned region with high number densities, which overlaps slightly with the energy cut. As for the higher energy cuts, we observe that the mean sample sizes have dropped significantly in comparison to the cuts for the default dataset. In fact, the `flag_any-lowlogg` values for $\langle N_{halo} \rangle$ have decreased with more than 10 % for both cases. This factor, as well as the knowledge that this data still carries high uncertainties for the halo stars, restricts us from making any conclusions based on this result.

For this reason, we once more resorted to the testing method based on the velocity model. The results from the model with non-scaled distances can be found in table B.2 and the results from the model with systematic shifts are provided in table B.3. Comparison with the corresponding fractions obtained for the default sample shows similar trends. While

Table B.1: The fractions of halo stars from the `flag_any-lowlogg` dataset that have negative and positive angular momenta, found using the uncertainty sampling method.

Energy cut	Negative L_z	Positive L_z	Standard deviation	$\langle N_{halo} \rangle$
$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.600	0.400	0.046	145
$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.738	0.262	0.036	65
$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.777	0.223	0.039	51

Table B.2: The fractions of halo stars from the `flag_any-lowlogg` subset with modelled velocities that have negative and positive angular momenta.

Energy cut	Negative L_z	Positive L_z	Standard deviation	$\langle N_{halo} \rangle$
$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.556	0.444	0.046	123
$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.613	0.387	0.082	37
$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.630	0.370	0.099	26

Table B.3: The fractions of halo stars from the `flag_any-lowlogg` subset with modelled velocities that have negative and positive angular momenta. However, the distances have now been scaled by a factor k to see how the distribution changes when we include systematic errors.

k	Energy cut	Negative L_z	Positive L_z	Standard deviation	$\langle N_{halo} \rangle$
1.1	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.607	0.393	0.039	147
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.659	0.341	0.066	51
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.663	0.337	0.080	37
1.2	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.653	0.347	0.035	176
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.696	0.304	0.054	69
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.704	0.296	0.066	52
1.3	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.691	0.309	0.033	205
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.725	0.275	0.049	89
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.730	0.270	0.052	69
1.4	$-1.6 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.718	0.282	0.029	235
	$-1.3 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.747	0.253	0.040	114
	$-1.2 \cdot 10^5 \text{ km}^2 \text{ s}^{-2}$	0.751	0.249	0.043	91

the percentages found are slightly larger for the `flag_any-lowlogg` sample, this is again balanced by smaller halo subset sizes and greater standard deviations.

The qualitative trend where systematic shifts in the distances lead to enlargement of the observed fractions of counter-rotating specimen in our halo star pseudo-samples is evidently present for both datasets. This speaks in favour of our hypothesis that the fractions found in [Helmi et al. \(2017\)](#) have been overestimated due to the less accurate distance estimates used in their study. We cannot draw any further conclusions from this observation since the number of possible halo stars was reduced by 268 as compared to the corresponding subset of the `flag_dup` sample, which is more than 10 % of the default halo sample size. However, it is known that the `flag_any-lowlogg` sample contains numerous data points with overestimated distances in the form of the `flag_lowlogg` stars. If our theory based on figure 4.1 is correct, we should thus observe that there still is an excess of negative L_z stars. This was indeed the case, which is indicated by the values in table B.1.

Appendix C

Standard deviations for scaled velocity field plots

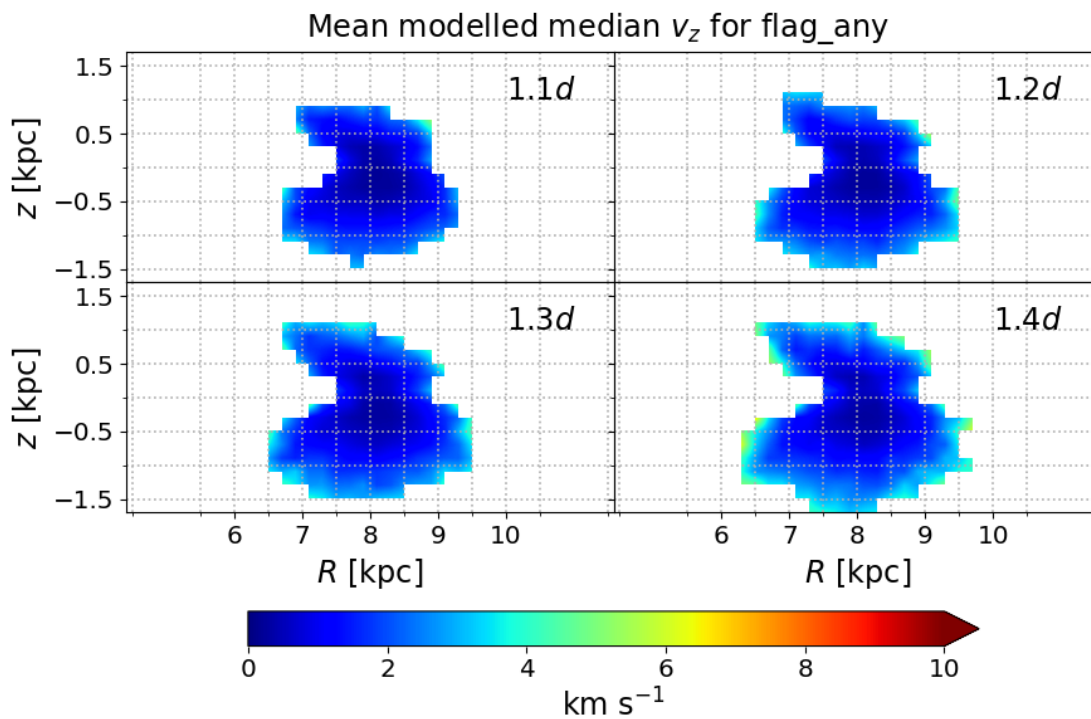


Figure C.1: The standard deviations for the four cases displayed in figure 5.4

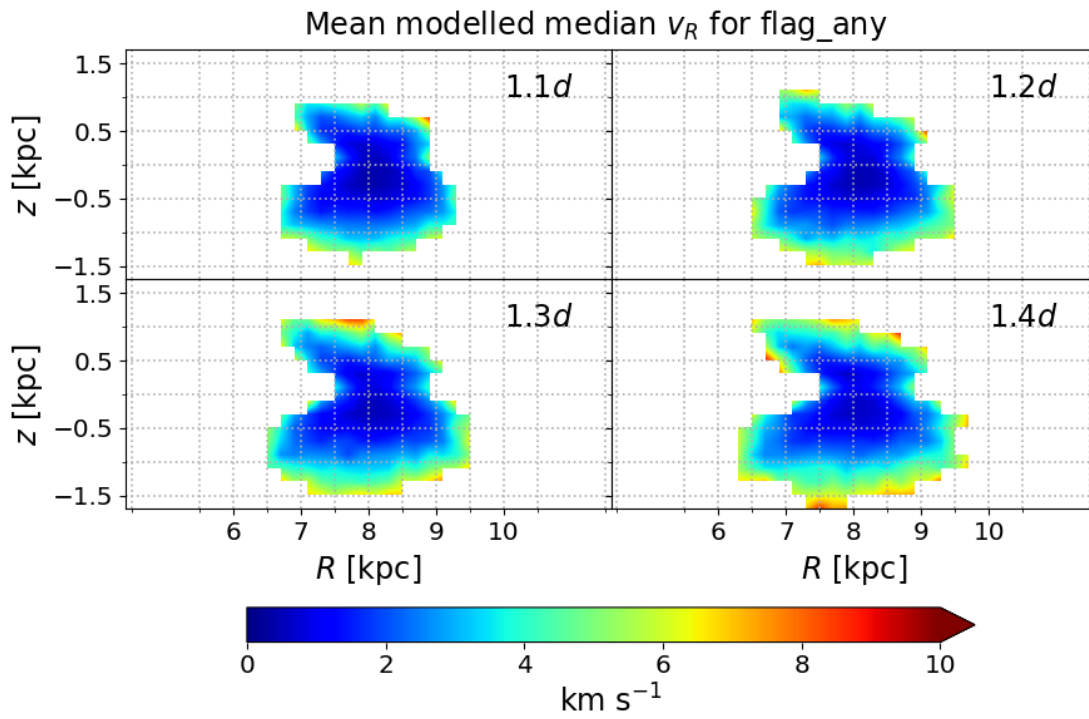


Figure C.2: The standard deviations for the four cases displayed in figure 5.5