

Lights in Dark Places:

Inferring the Milky Way Mass Profile using
Galactic Satellites and Hierarchical Bayes

LIGHTS IN DARK PLACES:
INFERRING THE MILKY WAY MASS PROFILE USING
GALACTIC SATELLITES AND HIERARCHICAL BAYES

BY

GWENDOLYN M. EADIE

M.Sc., Queen's University (Kingston, ON)

B.Sc., Simon Fraser University (Burnaby, BC)

a thesis submitted to

THE DEPARTMENT OF PHYSICS & ASTRONOMY

and

THE SCHOOL OF GRADUATE STUDIES OF MCMASTER UNIVERSITY

in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

© Copyright by Gwendolyn M. Eadie, August 2017

All Rights Reserved

Doctor of Philosophy (2017)
(Physics & Astronomy)

McMaster University
Hamilton, Ontario, Canada

TITLE: Lights in Dark Places:
Inferring the Milky Way Mass Profile using Galactic
Satellites and Hierarchical Bayes

AUTHOR: Gwendolyn M. Eadie

SUPERVISOR: Dr. William Harris

NUMBER OF PAGES: xx, 265

Dedicated to my parents

Abstract

Despite valiant effort by astronomers, the mass of the Milky Way (MW) Galaxy is poorly constrained, with estimates varying by a factor of two. A range of techniques have been developed and different types of data have been used to estimate the MW's mass. One of the most promising and popular techniques is to use the velocity and position information of satellite objects orbiting the Galaxy to infer the gravitational potential, and thus the total mass. Using these satellites, or Galactic *tracers*, presents a number of challenges: 1) much of the tracer velocity data are incomplete (i.e. only line-of-sight velocities have been measured), 2) our position in the Galaxy complicates how we quantify measurement uncertainties of mass estimates, and 3) the amount of available tracer data at large distances, where the dark matter halo dominates, is small. The latter challenge will improve with current and upcoming observational programs such as *Gaia* and the Large Synoptic Survey Telescope (*LSST*), but to properly prepare for these data sets we must overcome the former two. In this thesis work, we have created a hierarchical Bayesian framework to estimate the Galactic mass profile. The method includes incomplete and complete data simultaneously, and incorporates measurement uncertainties through a measurement model. The physical model relies on a distribution function for the tracers that allows the tracer and dark matter to have different spatial density

profiles. When the hierarchical Bayesian model is confronted with the kinematic data from satellites, a posterior distribution is acquired and used to infer the mass and mass profile of the Galaxy.

This thesis walks through the incremental steps that led to the development of the hierarchical Bayesian method, and presents MW mass estimates when the method is applied to the MW's globular cluster population. Our best estimate of the MW's virial mass is $M_{vir} = 0.87 (0.67, 1.09) \times 10^{12} M_{\odot}$. We also present preliminary results from a blind test on hydrodynamical, cosmological computer-simulated MW-type galaxies from the McMaster Unbiased Galaxy Simulations. These results suggest our method may be able to reliably recover the virial mass of the Galaxy.

Co-Authorship

Chapters 2, 3, 4, and 5 of this thesis are original research completed and written by myself, Gwendolyn Marie Eadie. The first three of these chapters have been published as peer-reviewed journal articles in *The Astrophysical Journal (ApJ)*, and have been reformatted to meet the McMaster thesis criteria. Chapter 6 is original research by me, in preparation for submission to *ApJ*.

The reference to Chapter 2 is Eadie, G., Harris, W., & Widrow, L. 2015, *ApJ*, 806, 54, DOI: 10.1088/0004 – 637X/806/1/54. The second and third authors on this publication are my PhD supervisor Dr. William Harris and my MSc supervisor Dr. Lawrence Widrow, respectively. The method that I developed during my MSc degree laid the groundwork for this study, and thus substantial parts of the introduction and theory (Sections 2.1 to 2.2.7) overlap with content in my Masters thesis (Eadie, G. M. 2013, Master’s thesis, Queen’s University, Kingston, Ontario, Canada). The analysis and results in this paper, including substantial improvements to the method, are original research completed during the first year of my PhD program.

The reference for Chapter 3 is Eadie, G., & Harris, W. 2016, *ApJ*, 829, 108, DOI: 10.3847/0004 – 637X/829/2/108. The second author on this paper is my PhD supervisor Dr. William Harris.

The reference for Chapter 4 is Eadie, G. M., Springford, A., & Harris,

W. E. 2017, *ApJ*, 835, 167, DOI: 10.3847/1538 – 4357/835/2/167. The second author on this paper is PhD candidate Aaron Springford from Queen’s University, and the third author is my PhD supervisor Dr. William Harris. Aaron Springford helped develop the mathematical and statistical basis for the hierarchical Bayesian method. After publication of this paper, I found a mistake in my interpretation of the globular cluster data, fixed this mistake, re-ran the analysis, and published an erratum Eadie, G. M., Springford, A., & Harris, W. E. 2017, *ApJ*, 835, 167, DOI: 10.3847/1538 – 4357/*aa64db*, which is shown in shown in the Appendix.

Chapter 5 will have co-authors Dr. Benjamin Keller and Dr. William Harris when it is submitted as an article to *ApJ*.

Academic peer-reviewed papers require context in their introductions, and thus parts of the introductions to the articles may overlap in some content (Chapters 2, 3, and 4).

Acknowledgements

I first want to acknowledge the entire Physics & Astronomy Department—the combined effort of everyone is what makes the department such a positive, inclusive, lively, and responsive place. A sincere thanks to Tina, Cheryl, Rosemary, and Mara, whose availability, commitment, and kindness are always, deep, and true. The astronomy faculty are also an all-star team; thank you for instilling in me values of hard-work, critical thinking, kindness, humbleness, professionalism, and work-life balance. In particular, thank you Laura, James, and Allison for providing me with valuable guidance and advice these past four years.

Bill Harris, you have been an exemplary and excellent PhD supervisor. Thank you for the consistent weekly meetings that kept me on track, for teaching me how to publish papers, for taking a keen interest in my career, for sending me to conferences all over the world, for giving me the freedom to fit figure skating into my schedule, and for generally caring about my well-being. I could not have asked for a better supervisor. (And Gretchen, thank you too!)

To my academic siblings, the fellow graduate students with whom I have shared an equivalent experience in both space and time, you have been essential in creating a feeling of community. Thank you. I have been extremely lucky to share this four-year experience alongside so many cool people. I can't start listing names, lest I leave someone out. But I will single you out Gandhali, my officemate for the past four years! You have become a dear friend, and I think I'm going to miss you the most come September.

Family is not an important thing, it's everything. - Michael J. Fox

Thank you Mom and Dad, two strong, well-rounded, interesting, and loving people. Your resilience in life motivates me to be resilient too. Thank you for all the phone chats that cheered me up, and for encouraging my curious mind from a very early age. John, thanks for being an inspiring big brother. If I can be half the astronomer that you are a cellist, then I think I'll have a shot at a Nobel Prize. Colin, you are easily the smartest of the three Eadie kids, and I know a game of chess with you will do damage to my ego. You are always ready to challenge my ideas and perceptions of the physical (and metaphysical?) world, and I thank you for that. Colleen, Doug, Mark, Dianna, and Patrick, (and newest editions to the Springford clan, Stephanie, Harrison, and John), I hope you know how much your love and support has helped me (and Aaron) through graduate school.

Aaron, I can't thank you enough for all of your love, support, commitment, advice, and encouragement through graduate school. Thank you for moving to Hamilton so I could pursue a PhD too. You have always encouraged me to do what I want with my life, to choose to be happy, and to not care about what other people think (within reason, haha). Not only are you my husband, my best friend, my personal comedian, and my skating partner, but you are also a chef, a statistician, an academic collaborator, an award-winning homebrewer, and a soon-to-be Doctor of Philosophy. Your wide range of talents leaves me perplexed, and I sure cannot wait to see what you do next. But your love and companionship mean the most to me, it's true, and this thesis would not be complete without me thanking you. THANK YOU ♡

Notation and abbreviations

ABC	Approximate Bayesian Computation
AJ	The Astronomical Journal
ApJ	The Astrophysical Journal
ApJL	The Astrophysical Journal Letters
ASA	American Statistical Association
CASCA	Canadian Astronomical Society
Decl.	Declination
DF(s)	Distribution Function(s)
DG(s)	Dwarf Galaxy(Galaxies)
GC(s)	Globular Cluster(s)
GME	Galactic Mass Estimator
HVSs	High Velocity Stars
IAU	International Astronomical Union
JSM	Joint Statistical Meetings
kpc	kiloparsecs
LG	Local Group
LMC	Large Magellanic Cloud
LSR	Local Standard of Rest
LT	Little & Tremaine (1987)
MCMC	Markov Chain Monte Carlo
MNRAS	Monthly Notices of the Royal Astronomical Society
MUGS2	McMaster Unbiased Galaxy Simulations

MW	Milky Way
M31	Messier Object 31 (i.e. Andromeda Galaxy)
NFW	Navarro, Frenk, and White (1997)
NGC	New General Catalog
NSC	Nuclear Star Cluster
NSERC	National Sciences and Engineering Research Council of Canada
Pal	Palomar
RA	Right Ascension
SMC	Small Magellanic Cloud
WE99	Wilkinson & Evans (1999)
W10	Watkins et al (2010)

Contents

Abstract	v
Co-Authorship	vii
Acknowledgements	ix
Notation and abbreviations	xi
1 Introduction	1
1.1 The Milky Way Galaxy	1
1.2 Definitions of Galaxy Mass	6
1.3 Mass Estimation Techniques	8
1.3.1 Tidal Streams	8
1.3.2 The Timing Argument	11
1.3.3 Kinematics of Satellites	13
1.4 Reviewing Major Challenges	29
1.5 Contribution of this Research	32
2 Estimating the Galactic Mass Profile in the Presence of In-	
complete Data	47
Abstract	48
2.1 Introduction	49

2.2	Background	51
2.2.1	Distribution Function	51
2.2.2	Models	53
2.2.3	Bayes Theorem and Parameter Estimation	56
2.2.4	Nuisance (v_t) Parameters	59
2.2.5	The Gibbs Sampler	60
2.2.6	Prior Probabilities	62
2.2.7	MCMC Chains and Assessing Convergence	64
2.3	Simulations	66
2.4	Kinematic Data for the Milky Way	68
2.5	Results	72
2.5.1	Simulation Results	72
2.5.2	Milky Way Results	77
2.6	Discussion and Future Prospects	84
3	Bayesian Mass Estimates of the Milky Way: The Dark and Light Sides of Parameter Assumptions	95
	Abstract	96
3.1	Introduction	97
3.2	The Power Law Model	99
3.3	Kinematic Data: Globular Clusters	104
3.4	Methods	115
3.4.1	Overview and Bayesian Inference	115
3.4.2	Markov Chains	117
3.4.3	Priors	118
3.4.4	Investigations	123

3.5	Results	126
3.5.1	Group (1): Scenarios I, II, & III	127
3.5.2	Group (2): Scenarios IV, V, & VI	129
3.5.3	Group (3): Scenarios VII, VIII, & IX	133
3.5.4	Group (4): Scenarios X, XI, & XII	135
3.5.5	Summary of Groups (1) - (4)	139
3.5.6	Sensitivity Test of the GC Sample	140
3.6	Discussion and Future Work	143
3.7	Conclusion	150
4	Bayesian Mass Estimates of the Milky Way: Including Measurement Uncertainties with Heirarchical Bayes	159
	Abstract	160
4.1	Introduction	161
4.2	Method	163
4.2.1	Hierarchical Bayesian Model	164
4.2.2	Defining Priors and Hyperpriors	168
4.2.3	Transformation of Velocities	169
4.2.4	Improved Computational Methods	173
4.3	Kinematic Data	173
4.4	Analysis Overview	174
4.5	Results	174
4.6	Discussion	183
4.7	Conclusion	186
5	A Blind Test on Milky Way-Type Galaxies from the MUGS2 Hydrodynamical, ΛCDM Simulations	193

5.1	Introduction	194
5.2	Brief Review of the Physical Model	196
5.3	Creating Mock Observations of Tracers in MUGS2 Galaxies	198
5.3.1	Finding GC analogues in MUGS2 Galaxies	198
5.3.2	Selecting MUGS2 Galaxies to Investigate	199
5.3.3	Creating Mock Heliocentric Observations	200
5.3.4	Transforming from Galactocentric to Heliocentric	203
5.3.5	Assigning Measurement Uncertainties	204
5.3.6	Creating Incomplete Data	207
5.4	Preliminary Results	209
5.4.1	Detailed Test Cases: <i>g15784</i> and <i>g1536</i>	209
5.4.2	Mass Estimates of MUGS2 galaxies	214
5.5	Discussion	217
5.6	Preliminary Conclusions	222
6	Summary & Future Work	231
6.1	Future Work	234
	Appendices	241
A	Erratum: “Bayesian Mass Estimates of the Milky Way: Including Measurement Uncertainties with Hierarchical Bayes” (2017, ApJ, 835, 167)	243

List of Tables

2.1	Milky Way Kinematic Data	69
2.2	Parameter estimates for the Milky Way	79
3.1	Heliocentric kinematic data of Milky Way globular clusters . . .	108
3.2	Prior probability distributions	123
3.3	Investigations of fixed versus free parameters	124
3.4	Summary of parameter estimates	125
5.1	Model parameter estimates and derived quantities: <i>g15784</i> . . .	210
5.2	Model parameter estimates and derived quantities: <i>g1536</i> . . .	211
A.1	Estimates of model parameters and derived quantities	246

List of Figures

1.1	Milky Way mass estimates from the literature	5
2.1	Empirical distributions of M_{tot} and a	73
2.2	Predicted versus true mass profile for simulated data	73
2.3	Empirical distributions with 50 v_t values removed	75
2.4	Cumulative mass profile (incomplete data)	76
2.5	Cumulative mass profile for the Milky Way	78
2.6	Estimated energies of Milky Way satellites	82
2.7	Sensitivity Analysis	83
3.1	Example Markov chain trace plots	118
3.2	Prior probability distribution for globular cluster system profile	122
3.3	Smoothed density profile of unused globular clusters	122
3.4	$M(r < 125\text{kpc})$ estimates for all Scenarios	126
3.5	Cumulative mass profiles (Group 1)	127
3.6	Posterior distributions for Scenario III	128
3.7	Posterior distributions for Scenario IV	130
3.8	Group 2 posterior distributions	132
3.9	Cumulative mass profiles (Group 2b)	133
3.10	Group 3 posterior distributions	134
3.11	Smoothed density profile for globular cluster subsample	136
3.12	Posterior distributions for Scenario XII	137

3.13	Cumulative mass profile of the Milky Way and comparisons . . .	138
3.14	Sensitivity analysis: subselecting globular clusters	142
4.1	Prior distributions on α	169
4.2	Checking transformation of heliocentric to Galactocentric ve- locities	172
4.3	Mass profile comparison: nonhierarchical and hierarchical meth- ods	175
4.4	Mass profile comparison: hierarchical method with 89 versus 143 GCs	176
4.5	Sensitivity analysis: mass estimates	178
4.6	Sensitivity analysis: parameter estimates	180
4.7	Joint posterior distributions for model parameters	181
4.8	Specific energy profile	183
5.1	Velocity profiles of GC analogues from MUGS2 simulated galaxies	201
5.2	Checking transformations of velocities	205
5.3	Generating missing data from <i>g15784</i>	208
5.4	Predicted and true mass profiles for <i>g15784</i> and <i>g1536</i>	211
5.5	Estimated energy profile of GC analogues from <i>g15784</i>	213
5.6	Estimated energy profile of GC analogues from <i>g1536</i>	214
5.7	Estimates of mass, virial radius, and anisotropy parameter for eight Milky Way-type galaxies from MUGS2	218
5.8	MUGS2 galaxy gas fractions	219
5.9	Number density profiles for the GC analogues	221
A.1	Cumulative mass profile for the MW	245
A.2	Sensitivity analysis: M_{125} estimates	246
A.3	Specific energy profile	247

1

Introduction

Looking deep into space allows us to study the past, while observing the nearby universe allows us to study the present. In particular, the Local Group, which is dominated by the Milky Way and Andromeda galaxies, provides a local laboratory for observing large, spiral galaxies, as well as the smaller and more numerous dwarf galaxies that live amongst them. In the context of extragalactic studies from deep space observations, studies of the Local Group allow us to test cosmological theories and to make inferences about how galaxies in the universe formed and evolved into their present state. Comparing the Milky Way to other galaxies in the universe, and understanding its properties in the context of galaxy formation theory, is a key part of this endeavour.

1.1 The Milky Way Galaxy

Living inside the Milky Way Galaxy gives us a unique perspective because the Galaxy's properties can be scrutinized in more detail than any other galaxy. The fundamental properties of the Milky Way (MW), such as size, composition, shape, star formation rate, and mass, were determined by the Galaxy's formation history, which is understood to have occurred through accretion of

both baryonic and dark matter. It is widely accepted that every galaxy in the universe resides within a massive, dark matter halo. In the cold dark matter model (Λ CDM), these dark matter halos are thought to have formed from density fluctuations in the very early universe which grew over time; more dark matter fell into the gravitational potential wells as time progressed, and the baryonic matter gathered into these wells too, forming the first galaxies. Galaxies continued to grow through accretion of gas and mergers with other galaxies. These gravitationally-driven formation mechanisms are understood to have created galaxy groups, galaxy clusters, and the overall hierarchical structure of the universe that we observe today.

An early inference of dark matter and its presence was found by Zwicky (1933, 1937) (translated by Ehlers, 2008), who looked at the Coma Cluster and discovered there was not enough visible matter to support the observed velocity dispersion of the galaxies within it. Since then, there has been mounting observational evidence supporting dark matter's existence, such as the rotation curves of spiral galaxies (the seminal paper being Rubin et al., 1980), gravitational lensing by galaxy clusters (e.g. Soucail et al., 1987; Lynds & Petrosian, 1989; Bartelmann, 2010), the distribution of intergalactic gas after galaxy cluster mergers (Markevitch et al., 2004; Clowe et al., 2006), and the cosmic microwave background (Spergel et al., 2003; Bennett et al., 2013; Planck Collaboration et al., 2016) (see also the review by Roos, 2010). Although alternative theories for gravity can predict some of these phenomena (e.g. the empirical model by Milgrom (1983) predicts the rotation curves of many spiral galaxies), a new theory for gravity that explains all observations from small to large scales remains elusive. Furthermore, computer simulations

show that cold dark matter plus dark energy are required to form the large-scale visible structures that we observe in the universe. The combination of such evidence strongly suggests that every galaxy in the universe resides within a massive halo of dark matter.

The current consensus is that both the mass of the Galaxy’s initial dark matter halo and the properties of the surrounding environment governed the MW’s past and ultimately dictated how it looks today. Because the MW does not have a finite boundary, it is common to define its mass within some specified distance from the center of the Galaxy, such as the virial radius (see Section 1.2). An accurate and precise measurement of the MW’s dark matter halo mass, which dominates the total mass, could have implications for two long-standing problems in Λ CDM cosmology: (1) the “Missing Satellite Problem”, which says that the number of observed satellites around the MW is significantly fewer than the number predicted from simulations (Klypin et al., 1999), and (2) the “Too-Big-To-Fail” problem, which says that the average masses of observed satellites are less than that predicted from simulations (Boylan-Kolchin et al., 2011b). For example, a less massive MW would imply fewer satellites, easing the tension of problem (1). To overcome problem (2), the mass of the MW should be less than $1.4 \times 10^{12} M_{\odot}$, according to comparisons of the MW dwarf galaxies (DGs) to those from Millenium II simulations (Cautun et al., 2014).

Evidence also suggests a connection between the amount of dark matter in galaxies and other galaxy features. For example, star formation rates (e.g. Springel & Hernquist, 2003; Behroozi et al., 2013), the masses of central supermassive black holes (Ferrarese, 2002; Baes et al., 2003), and the total stellar masses of galaxies (e.g. Guo et al., 2010) appear to be linked to the

host galaxy's dark matter halo. Thus, knowledge of the MW's dark matter halo mass is critical for testing not only cosmological theories about galaxy evolution, but also about internal Galactic processes.

Understanding the current properties of the halo, and specifically its mass, will help determine which models of galaxy formation are most likely, and ultimately test fundamental concepts in physics. However, because the MW is made up of a bulge, a disk, a stellar halo, and a dark matter halo, all of these components contribute to the total gravitational potential of the Galaxy. The mass and shape of the dark matter halo is the most difficult of these components to measure because we cannot directly detect it. Instead, the total gravitational potential must be inferred through the motions of visible objects or via theoretical arguments.

The total mass and cumulative mass profile of the MW — which are dominated by the dark matter halo at large radii — are poorly constrained. Many techniques exist to infer the MW's mass, each having its own advantages and disadvantages, and yet as a field, astronomers and astrophysicists have yet to pin down a precise mass estimate. Results in the literature span a factor of two or more (with an average somewhere near $10^{12}M_{\odot}$), and the uncertainties in these measurements do not always overlap (see Figure 1.1 Wang et al., 2015).

Many different techniques have been developed with the goal of estimating the MW mass, and I discuss the most popular methods below (Section 1.3). As we shall see, the different techniques and use of data yield a wide variety of mass results in the literature. Moreover, many estimates are limited to the mass within the furthest data point in the tracer sample, meaning that masses are reported at a variety of distances from the Galactic center. While efforts are made to standardize the mass estimates (e.g. Wang et al., 2015), it is

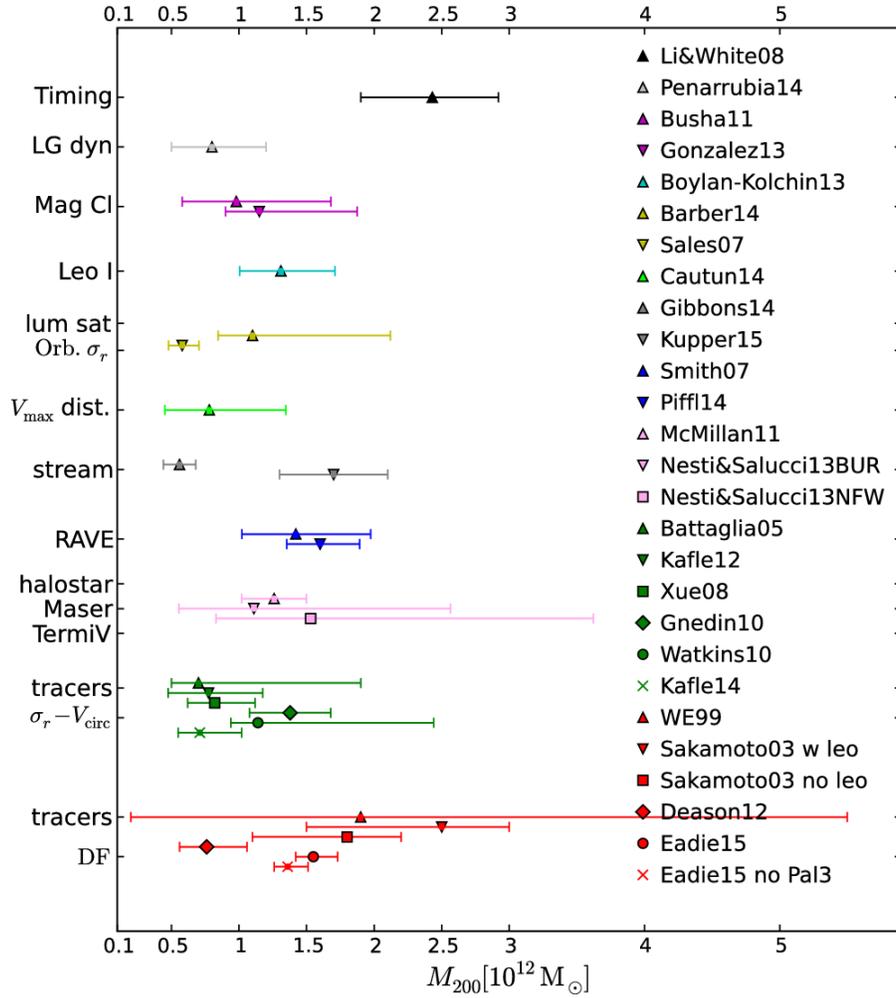


Figure 1.1: Collection of MW mass estimates (abscissa) and the techniques used to acquire them (ordinate). Reprint of Figure 1 from “Estimating the dark matter halo mass of our Milky Way using dynamical tracers” Wang et al. (2015), *Monthly Notices of the Royal Astronomical Society*, 453 (1):377-400.

nonetheless difficult to compare the mass estimates in the context of different methods.

Before outlining the most popular techniques for galaxy mass estimation, a brief review of Galaxy mass definitions is in order.

1.2 Definitions of Galaxy Mass

Total mass estimates for the MW are typically reported in terms of the *virial mass*, M_{vir} . The virial mass is defined as the mass enclosed within the virial radius, r_{vir} , which itself depends on cosmological theory and the critical density of the universe, ρ_{crit} . The following is a brief description of the cosmological theory that is often used to define the virial radius and thus the virial mass (Cole & Lacey, 1996).

The universe is assumed to be mathematically flat ($k = 0$) and have a critical density parameter of $\Omega = \rho/\rho_{crit} \equiv 1$, which implies that it will expand forever. The critical density of the universe, ρ_{crit} , is then

$$\rho_{crit} = \frac{3H^2}{8\pi G}, \quad (1.1)$$

where H is the Hubble constant and G is the gravitational constant.

In this cosmological model, the early universe contains spherically-symmetric density fluctuations that are Gaussian distributed. If one of these regions of high density has a significant overdensity compared to the background, then it undergoes idealized, spherical collapse. Linear theory predicts the overdensity at which collapse occurs, and the virial theorem and energy conservation can be used to show that the final density of such a region after virialization is

$\Delta_c = 18\pi^2 \approx 178$ (White, 2001; Cole & Lacey, 1996). The virial radius of a spherically collapsed, virialized halo is then defined such that

$$\bar{\rho}(r_{vir}) = \Delta_c \rho_{crit}, \quad (1.2)$$

where $\bar{\rho}$ is the mean density of the halo.

The mean density of a galaxy within its virial radius is defined as

$$\bar{\rho}(r_{vir}) = \frac{M_{vir}}{\frac{4}{3}\pi r_{vir}^3} \quad (1.3)$$

Thus, with Equations 1.2 and 1.3, the virial radius is found to be

$$r_{vir} = \left(\frac{3M_{vir}}{4\pi\Delta_c\rho_{crit}} \right)^{1/3}. \quad (1.4)$$

Equation 1.4 shows how the model for the mass profile $M(r)$ will play a role in the definition of the virial radius.

Often, Δ_c is set to 200, leading to the notation M_{200} (which is how we define M_{vir} in Chapter 4). In this thesis, other notations such as $M_{125} = M(r < 125 \text{ kpc})$ and $M_{300} = M(r < 300 \text{ kpc})$ are also used to define the mass *within certain radii*, and these should not be confused with the definition of M_{200} .

When a virial mass is reported in the literature, it is often but not always an extrapolation out to the virial radius. Little observational data is available beyond $\sim 150 \text{ kpc}$, and thus studies tend to report the mass contained within the Galactocentric distance of their furthest data point. With a variety of mass estimates at different Galactocentric radii, it is sometimes difficult to compare and contrast studies that use different data sets and different methods. It

would therefore be beneficial to have a method that predicts a continuous mass profile, which would enable more direct comparisons between studies that report masses within different radii.

1.3 Mass Estimation Techniques

There are numerous ways to measure the mass of the Galaxy, but ultimately all methods boil down to using the position and kinematic information of objects in and around the MW, or using the major components of the Local Group (LG) (i.e. M31 and dwarf galaxies). Traditionally, most mass estimation techniques rely on fundamental physics, but as we shall see, recently some studies have begun to obtain Galactic mass estimates by comparing observational data to simulated data from cosmologically-motivated, hydrodynamical simulations.

In the next few subsections, I outline the most popular methods for estimating the mass of the MW and highlight both advantages and disadvantages to each method.

1.3.1 Tidal Streams

Since the discovery of the Sagittarius tidal stream around the MW (Ibata et al., 2001b,a; Majewski et al., 2003), it has become popular to use stellar streams as tracers of the Galactic potential (e.g. Law & Majewski, 2010; Koposov et al., 2010; Gibbons et al., 2014; Küpper et al., 2015; Pearson et al., 2015; Bovy et al., 2016; Dierickx & Loeb, 2017). A stellar stream is assumed to be created from a satellite dwarf galaxy (DG) or globular cluster (GC) that is being torn apart by the tidal effects of gravity— leading and trailing arms of stars precede

and follow the progenitor’s orbital path.

Mass estimates derived using tidal streams are mixed. For example, the Sagittarius stream appears to imply a “lighter” MW mass ($\lesssim 10^{12} M_{\odot}$), whereas the GC stream from Pal 5 suggests a “heavier” mass. Gibbons et al. (2014) found $M(r < 100 \text{ kpc}) = 0.41 \pm 0.04 \times 10^{12} M_{\odot}$ using the stream of Sagittarius, whereas Küpper et al. (2015) found $M(r < 100 \text{ kpc}) = 0.90 \pm 0.20 \times 10^{12} M_{\odot}$ using the stream of Pal 5.

An arguable advantage to using streams over other mass estimation techniques is the possibility that they may help constrain the shape of the dark matter halo. Cosmological simulations predict different shapes and aspect ratios of dark matter halos (e.g. Vera-Ciro et al., 2011; Kuhlen et al., 2007). However, so far no shape of the MW’s dark matter halo has been ruled out, as studies have found evidence for an oblate (e.g. Loebman et al., 2014; Deg & Widrow, 2013; Law & Majewski, 2010), prolate (e.g. Bowden et al., 2016), and spherical (e.g. Smith et al., 2009; Belokurov et al., 2006) dark matter halo. It has even been suggested that the Sagittarius stream is too dynamically young to provide any information about the shape of the dark matter halo (Helmi, 2004).

Using stellar streams as probes of the Galactic potential poses some challenges. For example, the Sagittarius stream has a complex mixture of stars and wide spatial distribution on the sky (Koposov et al., 2010). The amount of kinematic information of the individual stars within stellar streams is also limited. Sophisticated techniques have been developed to use stellar streams as probes of the Galactic potential despite this shortcoming (e.g. Varghese et al., 2011; Price-Whelan et al., 2014).

For simplicity, it is often assumed that stellar streams follow the orbit of

the progenitor. However, Eyre & Binney (2011) showed that stellar streams can be made up of multiple orbits, occupying different regions of phase-space. Therefore, using streams under an assumption of a single orbit can introduce significant systematic error.

Investigations into the effectiveness of using streams have been performed. Bonaca et al. (2014) used a forward-modeling technique to investigate the use of stellar streams as probes of the Galactic dark matter halo. They found that using multiple stellar streams is favourable to using only one, because analysis of a single stream may under- or over-estimate the Galaxy's mass.

Sanderson et al. (2017) investigated how the action-space clustering of information from stellar streams can provide information about the mass of the Galactic dark matter halo. By analysing stellar streams around MW-type galaxies from the Aquarius simulations (Springel et al., 2008), they successfully recovered mass profiles that were consistent with spherically averaged dark matter profiles, even when ignoring the shape and time-dependencies. Probabilistic models using a Bayesian paradigm have also been developed, although they are currently too computationally-heavy for practical use (Sanders, 2014). The *minimum entropy method* introduced by Peñarrubia et al. (2012) provides yet another way to analyse stellar streams and constrain the Galactic potential, although the authors stress that further investigations are needed to test the method.

Clearly, using stellar streams is an active area of research, and rightly so with the new and continuing data from the *Gaia* satellite (Perryman et al., 2001; ESA, 2016), which is sure to discover more stellar streams in the Milky Way. As surveys start to detect stellar streams around other galaxies, methods to use extragalactic stellar streams to estimate masses of other galaxies will

also become important.

1.3.2 The Timing Argument

The *timing argument* uses the current distance between and relative velocities of the MW and M31, and the age of the Universe, to constrain the *total* mass $M = M_{MW} + M_{M31}$ of the two galaxies, and thus provides an approximation of the Local Group mass, M_{LG} (Kahn & Woltjer, 1959). The timing argument assumes the MW and M31:

- can be treated as point mass particles,
- had a separation of zero right after the Big Bang,
- are in bound orbits,
- obeyed conservation of energy since the big bang, and
- must have reached apocentre at least once.

The last assumption requires that the present day separation of the MW and M31 is less than that of apocentric separation, and it also puts a constraint on the orbital period of the two-body system Li & White (2008).

Early efforts using the timing argument suggested a total mass of the LG that was larger than the sum of the MW and M31 combined. In light of the high LG mass, Li & White (2008) used outputs from the *Millenium Simulation* to investigate and calibrate the timing argument. They looked at both LG-like pairs of galaxies and pairs resembling the MW-Leo I system. Li & White (2008) argued that mass estimates from the timing argument for both the LG and the MW-Leo I system are robust, and find median estimates of $M_{LG} = 5.27_{1.45}^{+1.51} \times 10^{12} M_{\odot}$, and $M_{MW} = 2.43 \times 10^{12} M_{\odot}$. Even though the lower 95% confidence limit on the MW is $0.80 \times 10^{12} M_{\odot}$, the median is high enough

to suggest the MW is larger in mass than the Andromeda Galaxy. A great deal of observational evidence brings into question this result, including M31's greater number of GCs and DGs, its higher maximum rotational velocity, and its stellar content (Diaz et al., 2014, outlines the evidence clearly).

In the last decade, it has become possible to measure the proper motions of stars within the Andromeda galaxy, enabling the use of the timing argument to estimate the total mass of the Local Group to better precision (Sohn et al., 2012; van der Marel et al., 2012b,a). The new data, in conjunction with a Bayesian analysis using prior information of the MW's mass from other studies, have led to lower estimates for the LG's virial mass, such as $3.17 \pm 0.57 \times 10^{12} M_{\odot}$ (van der Marel et al., 2012a).

A disadvantage to the timing argument is that it ignores the other massive satellites in the LG, including the Large Magellanic Clouds and the Triangulum Galaxy (M33) (Diaz et al., 2014). Another issue is that the assumption of point-masses for the two galaxies MW and M31 is often made; the halos of these two galaxies likely overlapped at earlier times.

Peñarrubia et al. (2014) modeled the local cosmic expansion with a Bayesian approach and compared their result to the timing argument. They found a LG mass of $M_{LG} = 2.3 \pm 0.7 \times 10^{12} M_{\odot}$, and ruled out models in which $M_{MW} > M_{M31}$. An advantage to their Bayesian approach meant that many parameters could be estimated simultaneously; not only did they obtain an estimate of M_{LG} , but also an estimate for the mass ratio M_{MW}/M_{M31} , the circular velocity of the MW at the solar radius, the reduced Hubble constant, and the fractional vacuum density.

Another alternative to the timing argument was suggested by Diaz et al. (2014). They estimated the total mass of the MW and M31 by assuming that

the total momentum of the LG should be zero and that the masses of these two galaxies dominate. One advantage to this line of argument, as opposed to the timing argument, is that it does not assume the two main galaxies of the LG are point masses, and instead requires that the dark matter halos of M31 and the MW do not overlap. The authors also claim that the orbits and number of passages are not needed. With this method, $M_{LG} = 2.5 \pm 0.4 \times 10^{12} M_{\odot}$, where $M_{MW} = 0.8 \pm 0.5 \times 10^{12} M_{\odot} \sim 0.43 M_{M31}$.

Research investigating the discrepancy between the timing argument mass estimates of the LG and the summed masses of the MW and M31 is ongoing, with studies of hydrodynamical and cosmological simulations hopefully providing some insight (e.g. Fattahi et al., 2016; Carlesi et al., 2017). Carlesi et al. (2017) used prior information from cosmological simulations to construct posterior distributions of the mass of the MW, M31, and the LG, finding that this method yields lower M_{LG} values than the timing argument, and masses of the MW and M31 in the ranges of $0.6 - 0.8 \times 10^{12} M_{\odot}$ and $1.0 - 2.0 \times 10^{12} M_{\odot}$ respectively. However, the authors note that differences in the tangential velocity of M31 can change the LG mass by a factor of two, and change the ratio M_{31}/M_{MW} by up to 20%.

1.3.3 Kinematics of Satellites

Satellite objects that orbit the Galaxy, such as halo stars, globular clusters (GCs), and dwarf galaxies (DGs), can be used as *tracers* of the Galactic potential, and thus probe the total mass. Using tracer kinematics is undoubtedly the most popular technique in the literature, but it is not without its challenges.

One challenge is that the heliocentric reference frame complicates position and velocity measurements of tracers in orbit around the Galaxy. Theoretical models of the Galaxy are written in mathematically simpler and tractable forms if a Galactocentric reference frame is used. Thus, it is often necessary to transform heliocentric position and velocity measurements of tracers into Galactocentric ones, which requires knowledge of the Local Standard of Rest (LSR), the solar motion, and the distance to the tracer. Although transforming positions and velocities from the heliocentric to Galactocentric frame is a solved problem, it requires careful procedures (Johnson & Soderblom, 1987).

All measurements are subject to uncertainty, but some of the measurements are more precise than others— it is important to take into account these differences carefully. Another challenge is the non-linear propagation of these measurement uncertainties from the heliocentric to Galactocentric frames, which is non-trivial without making certain assumptions about error distributions.

The velocity data for satellite objects are often *incomplete*; most line-of-sight velocities have been measured, but many proper motions have not. A datum with an incomplete velocity measurement cannot be transformed from the heliocentric frame to the Galactocentric one without making certain geometric and limiting assumptions.

Small sample size also presents a problem. For example, beyond 80kpc, the number of GCs around the MW is small, making measurements of the virial mass mostly an extrapolation. Until the last decade or so, studies relied on not only the few distant GCs but also the MW's DGs. However, use of some DGs (e.g. Leo I) as tracers of the dark matter halo have been shown to have strong influence on mass estimates (e.g. Boylan-Kolchin et al., 2013;

Sakamoto et al., 2003). The main issue is that these tracers' full velocity vectors are either unknown (i.e. only line-of-sight velocities are known) or measurements are highly uncertain. Therefore, any assumptions about such a tracer's boundedness to the Galaxy and/or the ellipticity of its orbit play a large role in the mass estimate.

All of these challenges — reference frames, incomplete data, measurement uncertainty, and small sample size — in addition to galaxy model assumptions contribute to the variation in mass estimates in the literature. Regardless, research has persisted in trying to obtain an accurate mass for the MW because of its importance. In the next few subsections, I outline the major methods by which tracer kinematics have been used to estimate the Galaxy's mass.

Mass Estimators and the Jeans Equation

Mass estimators use the kinematic information from a population of Galaxy tracers, such as halo stars, to infer the mass within the distance of the furthest tracer. Mass estimators typically use only radial velocities, and thus they depend heavily on an assumed velocity anisotropy of the system. Assumptions about the eccentricity of orbits in the system can produce very different mass estimates. However, one major advantage of mass estimators is that they rarely rely on a physical model¹.

In what follows, Hartwick & Sargent (1978) derived one of the first mass estimators, under the assumption of a spherically symmetric, collisionless system for the MW. Beginning from the first moment of the Boltzmann equation,

$$\frac{d}{dr} (n(r)\langle v_r^2 \rangle) = -\frac{GM(r)}{r^2}n(r) + \frac{(\lambda - 2)n(r)\langle v_r^2 \rangle}{r} + \frac{\bar{v}_{rot}^2 n(r)}{r}, \quad (1.5)$$

¹except, of course, that of gravity

where $n(r)$ is the radial number density of the tracers, λ determines the velocity anisotropy, v_r is the Galactocentric radial velocity, \bar{v}_{rot} is the mean streaming velocity, $M(r)$ is the mass within Galactocentric radius r , and angled brackets indicate time-averages, equation 1.5 is solved for $M(r)$ to give

$$M(r) = \frac{r\langle v_r^2 \rangle}{G} \left[-\frac{d \log n}{d \log r} - \frac{d \log \langle v_r^2 \rangle}{d \log r} + (\lambda - 2) + \frac{\bar{v}_{rot}^2}{\langle v_r^2 \rangle} \right]. \quad (1.6)$$

The term $\langle v_r^2 \rangle$ depends explicitly on the assumed anisotropy of the system.

Equation 1.6 is one way of writing the *Jeans equation*, which relates the total gravitational potential to the number density and velocity dispersion of the system's tracer population. Equation 1.6 holds for a spherical, collisionless, rotationless, and dissipation-supported system in equilibrium.

Use of the Jeans equation has been one of the most popular approaches for estimating the mass of the MW (e.g. Battaglia et al., 2005; Kafle et al., 2012), other galaxies (e.g. Łokas, 2009), and even galaxy clusters too (e.g. Côté et al., 2003). Nowadays, the Jeans' Equation is usually written in the following notation and form,

$$\frac{1}{n} \frac{d}{dr} (n \sigma_r^2) + \frac{2\beta \sigma_r^2}{r} = -\frac{GM(r)}{r^2}, \quad (1.7)$$

where σ_r is the radial velocity dispersion and where

$$\beta = 1 - \frac{\sigma_\theta^2 + \sigma_\phi^2}{2\sigma_r^2} \quad (1.8)$$

is the anisotropy parameter (Binney & Tremaine, 2008).

One drawback to using Equation 1.7 to estimate the mass of the MW, however, is that assumptions must be made about β of the tracer population,

leading to a mass-anisotropy degeneracy. The degeneracy lies in the fact that both $n(r)$ and the velocity dispersion σ_r^2 will depend on β . Some studies suggest that the mass estimate can be biased by over 70% without any knowledge of the velocity anisotropy of the satellites (Watkins et al., 2010). In addition to assuming a velocity anisotropy, use of the Jeans' equation often requires fixing the parameter for the number density profile of the satellite population (e.g. Kafle et al., 2012). Loebman et al. (2017) recently showed that Jeans' modelling is sensitive to non-constant β , and in particular could lead to over-estimates of the Galaxy's mass.

Another early mass estimator was introduced by Lynden-Bell et al. (1983) (which built upon Lynden-Bell & Frenk, 1981). They showed how the Galactocentric distances r and radial velocities v_r of satellites could be used to estimate the mass of the MW through application of the virial theorem and by making assumptions about orbital distributions of tracers. Assuming that (1) most of the Galaxy's mass resides within the solar circle and thus can be treated as a point mass, and (2) the tracers' orbits have an isotropic distribution (i.e. the time-averaged eccentricity is $\langle e^2 \rangle = 1/2$), they showed that

$$M = \frac{4\langle(v_r^2 - \epsilon^2)r\rangle}{G}, \quad (1.9)$$

where ϵ represents the velocity uncertainties. If instead the anisotropy distribution is dominated by radial orbits (i.e. $\langle e^2 \rangle = 1/8$), then the mass is given by

$$M = \frac{16\langle(v_r^2 - \epsilon^2)r\rangle}{G}, \quad (1.10)$$

which notably gives a result four times that of Equation 1.9 (Lynden-Bell et al., 1983). The isotropic mass estimator (Equation 1.9) was later used by

Olszewski et al. (1986) with 16 remote halo objects to obtain a Galaxy mass estimate of $5 \pm 2 \times 10^{11} M_{\odot}$. The authors stress, however, that the highly uncertain measurements and limited sample size impede a better estimate.

Little & Tremaine (1987) (hereafter LT) review the shortcomings of the mass estimators by Hartwick & Sargent (1978) and Lynden-Bell et al. (1983). For one, Equation 1.6 (1.7) requires the quantities such as $n(r)$ to be specified, which is arguably difficult when the tracer sample size is small. LT also call into question the confidence intervals because it is unclear how these may be calculated with a finite sample size. Finally, both mass estimators assume bound satellites, but at the same time can produce mass estimates implying that some satellites have velocities higher than the escape speed. In light of these problems, LT introduce a Bayesian approach that uses a *distribution function* (to be discussed later).

The Jeans equation has been used to estimate the mass of the Galaxy in a few different ways. For example, Battaglia et al. (2005) derived the expected radial velocity dispersion profile from the Jeans equation under both the assumption of constant β and varying $\beta(r)$, and performed a minimum- χ^2 analysis with 240 tracer objects (including GCs, halo stars, and DGs). The authors acknowledge that their analysis highlights the well-known mass-anisotropy degeneracy; multiple combinations of β and potential models can describe the observed velocity dispersion equally well. Both (1) the empirical dark halo model introduced by Navarro et al. (1996, hereafter NFW) that includes a varying anisotropy, and 2) a truncated flat rotation model with constant anisotropy, provide good fits to the data, giving mass estimates of $0.8_{-0.2}^{+1.2} \times 10^{12} M_{\odot}$ ($-0.3 \lesssim \beta \lesssim 1$) or $1.2_{-0.5}^{+1.8} \times 10^{12} M_{\odot}$ ($\beta = -0.5 \pm 0.4$) respectively. The data can be explained by either a steep dark matter halo profile

with constant anisotropy, or by a less steep profile with varying anisotropy that goes to tangentially-biased orbits at large radii. The authors thus stress the need for proper motion measurements of tracers at large radii in order to break the mass-anisotropy degeneracy. However, their method did not make use of any proper motion measurements, even for those tracers for which these measurements are available. Perhaps the mass-anisotropy degeneracy could be partially lifted (or at least some models could be ruled out) if a technique could include known proper motions measurements as well.

Xue et al. (2008) used the Jeans equation in conjunction with hydrodynamical simulations to estimate the circular velocity profile of the MW, and thus constrain the MW mass. The tracers were a collection of over 2400 Blue Horizontal Branch (BHB) stars whose positions and line-of-sight velocities were measured by the Sloan Digital Sky Survey. By binning the data into radial bins and comparing the observed line-of-sight velocity dispersions in each bin to mock observations from hydrodynamical simulations, the circular velocity as a function of radius, $V_{circ}(r)$, was scaled through a Jeans equation-based modelling approach such that the simulated data matched the observed data. With this approach, they found $M(r < 60 \text{ kpc}) = 4.0 \pm 0.7 \times 10^{11} M_{\odot}$. From here, they extrapolate the fit out to a virial radius of $\approx 250 \text{ kpc}$ to obtain a virial mass of $M_{vir} = 1.0_{-0.2}^{+0.3} \times 10^{12} M_{\odot}$ (assuming a potential model that includes a Hernquist (1990) bulge, an exponential disk, and an NFW dark matter halo). One downside to the Jeans equation-based method was that the measurement uncertainties of the BHB stars were not included directly in the analysis, and so the uncertainties in the mass estimate relied on bootstrapping the data. Matching observations to simulated observations also relies heavily on the idea that the hydrodynamical simulations are a good representation of

reality.

More recently Watkins et al. (2010) (hereafter W10) introduced two mass estimators for the MW and M31. One estimator uses only the radial velocity (v_r) and position r of the satellite data,

$$M = \frac{(\alpha + \gamma - 2\beta)r_{out}^{1-\alpha}}{G} \langle v_r^2 r^\alpha \rangle \quad (1.11)$$

where α is a parameter of the gravitational potential, γ is the power-law index for the number density of the tracer population, r_{out} is the distance of the furthest tracer, and β is the standard anisotropy parameter. The second estimator uses the complete velocity v , for when proper motion measurements are available,

$$M = \frac{1}{G} \frac{(\alpha + \gamma - 2\beta)r_{out}^{1-\alpha}}{3 - 2\beta} \langle v^2 r^\alpha \rangle. \quad (1.12)$$

The above equations require estimates of α , β , and γ , which necessitates using observational and/or simulated data. The results of this study thus depend heavily on the assumed velocity anisotropy, and W10 find the mass of the MW to be between $0.7 \times 10^{12} M_\odot$ and $3.4 \times 10^{12} M_\odot$, depending on whether β is defined from simulations or observations. Another large source of uncertainty in their mass estimate comes from assuming whether or not Leo I is bound to the Galaxy.

There are two drawbacks to using Equations 1.11 and 1.12. The first is that obtaining a standard deviation for the mass estimate requires Monte Carlo simulations, and W10 must generate mock data sets of the MW satellites to obtain uncertainties in their mass estimates. The second is that to combine all of the information from the kinematics of the satellites, a linear combination of the estimates from Equations 1.11 and 1.12 is used to arrive at a final estimate

of the mass of the MW.

Deason et al. (2012) applied the W10 radial velocity mass estimator (Equation 1.11) to new kinematic data of BHB, blue straggler, and N-type carbon stars. The data are identified out to ~ 150 kpc, but the sample size contains a modest 31 stars beyond 80kpc. Thus, binning the data at large distances leads to high levels of uncertainty, and their mass estimates within 150kpc range from 0.3×10^{12} and $1.6 \times 10^{12} M_{\odot}$. The uncertainty in their mass estimate is overwhelmed by the small sample size and incomplete data, because assumptions about velocity anisotropy and the tracer density profile play a large role in determining the mass.

Overall, one of the biggest drawbacks to mass estimators is that they typically use only line-of-sight velocities, and there is no obvious way to include both incomplete and complete data simultaneously in the analysis. Furthermore, methods to include measurement uncertainties often rely on bootstrapping the data.

Comparing Cosmological Simulations to the Magellanic Clouds

Some recent studies have taken a different approach to estimating the mass of the MW by directly comparing observations of the Large and Small Magellanic Clouds (LMC and SMC) to cosmological simulations of MW-type galaxies.

Boylan-Kolchin et al. (2011a) and Busha et al. (2011) estimated the MW mass by comparing observable features of the Magellanic Clouds to cosmologically simulated MW-type galaxies. The latter study used Millenium-II simulations (Boylan-Kolchin et al., 2009) and the former used the Bolshoi simulations (Klypin et al., 2011) as a prior in a Bayesian analysis. The two studies are in disagreement with one another, with Boylan-Kolchin et al. (2011a) suggesting

a heavy mass of at least $2 \times 10^{12} M_{\odot}$ and Busha et al. (2011) suggesting a more moderate $\sim 1.2 \times 10^{12} M_{\odot}$.

Barber et al. (2014) took a slightly different approach, and compared the ellipticity of the MW’s satellite orbits to the MW-sized dark-matter halo simulations of the Aquarius Project. They estimate that the mass range $M_{200} = 0.6 - 3.1 \times 10^{12} M_{\odot}$ for the MW.

Most recently, Patel et al. (2017) used the proper motions of the LMC and M33, along with results from the Illustris dark-matter only cosmological simulation (Vogelsberger et al., 2014a,b; Nelson et al., 2015), to determine the mass of the MW and M31. The authors follow the same Bayesian inference technique as Busha et al. (2011), and assume that the orbital energy of the massive satellites are highly correlated with the host galaxy’s halo mass. Their MW mass estimate is $M_{vir} = 1.02_{-0.55}^{+0.77} \times 10^{12} M_{\odot}$.

Using results from simulations as priors in a Bayesian analysis of observational data from the MW will depend heavily on the assumption that the simulations are a good representation of reality. Furthermore, this kind of analysis requires selection of MW-type galaxies from the simulation, and so one must question what prior information is being used in this selection.

Finally, when making inferences about the MW from aggregate simulated data, one must be wary of the *ecological fallacy* (Robinson, 1950). The ecological fallacy is a proven statistical phenomena in which “two variables at the aggregate level frequently differ markedly from the correlations between the same two variables at the individual level” (Gove & Hughes, 1980). In other words, correlations between properties at the individual level cannot be reliably inferred from correlations at the group level (Piantadosi et al., 1988). Therefore, even if the number of satellites around galaxies and the masses of

galaxies are correlated in cosmological simulations, this does not mean that the MW satellite population has to be similar in number to its counterpart in simulations.

High Velocity Stars

The velocity distribution of High Velocity Stars (HVSs) can be used to infer an escape speed for the Galaxy and to thereby infer the gravitational potential and mass of the MW. HVSs in the stellar halo of the MW are thought to be produced by strong dynamical interactions such as a binary stellar system interacting with the central massive black hole of the MW (Hills, 1988; Perets, 2009). The main idea behind using HVSs is that they will create asymmetries in the distribution of halo stars in the MW (Gnedin et al., 2005; Yu & Madau, 2007; Perets et al., 2009). Without HVSs, there would be as many stars with positive radial velocities (outgoing stars) as those with negative radial velocities (incoming). Any unbound stars then contribute to the positive side of the distribution, because by definition these stars are on their way *out* of the Galaxy. By studying this asymmetry in the distribution, identifying HVSs, and assuming a physical model for the gravitational potential, one can constrain the escape speed at different distances.

Mass estimates from recent HVS studies are mixed. Smith et al. (2007) derived a MW virial mass under the assumption of the NFW cold dark matter halo model, arriving at a mass of $1.42_{-0.54}^{+1.14} \times 10^{12} M_{\odot}$. HVSs ejected from the center regions of the MW lead to a mass estimate of $(1.2 - 1.9) \times 10^{12} M_{\odot}$ (Fragione & Loeb, 2017).

Gnedin et al. (2010) analysed HVSs from the Hypervelocity Star Survey to estimate the enclosed mass within 80kpc of the Galactic center. Assuming

a single power law for the density profile of halo stars, and restricting their analysis to objects beyond 25kpc, a Jeans modelling analysis suggested a mass of $0.69_{-1.2}^{+0.30} \times 10^{12} M_{\odot}$. The authors state that the uncertainty in this mass estimate is dominated by the lack of proper motion measurements for the HVSs and thus no information about the velocity anisotropy of the tracer population.

One potential problem with this method is contamination in the velocity distribution from stellar streams, which would have a bulk motion either towards or away from the Galactic center, although it has been suggested that the contamination is low (Piffl et al., 2011). Another issue with using HVSs to estimate the mass is heavy model dependence— the escape speed at a given distance from the Galactic center will depend on the physical model. Furthermore, estimating the mass of the Galaxy relies not only the local escape speed but also how the mass is distributed outside the solar circle (Leonard & Tremaine, 1990). Another issue is the assumption of velocity anisotropy, since proper motion measurements for HVSs are limited.

Distribution Functions

The strong Jeans Theorem implies that the positions and velocities of collisionless particles in a steady state system can be described probabilistically by a *distribution function* (DF). The DF gives the probability of a star having a particular position in 6-dimensional phase space (i.e. with velocity $\mathbf{v} = \langle \mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z \rangle$ and position $\mathbf{r} = \langle \mathbf{x}, \mathbf{y}, \mathbf{z} \rangle$) (Binney & Tremaine, 2008), and it

is a probability density function (PDF), by definition integrating to one:²

$$\int f(\mathbf{r}, \mathbf{v}) \mathbf{d}^3\mathbf{r} \mathbf{d}^3\mathbf{v} = 1. \quad (1.13)$$

In practice, the DF is usually written in terms of the relative energy per unit mass \mathcal{E} and the angular momentum per unit mass \vec{L} where

$$\mathcal{E} = -\frac{v^2}{2} + \Phi(r) \quad (1.14)$$

and

$$\vec{L} = \mathbf{r} \times \mathbf{v}. \quad (1.15)$$

Under the assumption of spherical symmetry, only the magnitude of L matters, and in this case the notation $f(\mathcal{E}, L)$ is common.

In the context of Galaxy mass estimation, analytic DFs are commonly used in Bayesian (e.g. Little & Tremaine, 1987; Kochanek, 1996; Wilkinson & Evans, 1999; McMillan, 2011) and maximum likelihood (e.g. Kulessa & Lynden-Bell, 1992; Deason et al., 2011) frameworks to analyse kinematic data from tracers. In these types of analyses, a distribution of model parameters (rather than point estimates) is obtained which can be useful for making scientific inferences about physical quantities.

Little & Tremaine (1987)(LT) used a Bayesian approach to analyse kinematic data from eight DGs and seven GCs extending to $\sim 90\text{kpc}$, and found an upper limit of the MW mass of $5.2 \times 10^{11} M_{\odot}$ (assuming a point-mass model) or $2.4_{-0.7}^{+1.3} \times 10^{11} M_{\odot}$ (assuming an infinite halo model). If the massive halo is assumed to extend out to 90kpc , then the estimate is closer to $\sim 1 \times 10^{12} M_{\odot}$.

²In practice, DFs in astronomy may be normalized such that $f(\mathbf{x}, \mathbf{r})$ integrates to a value of interest, such as the total mass or total luminosity.

The Bayesian method used in their study does not incorporate the measurement uncertainties, and so highly uncertain measurements are left out of the analysis.

The technique introduced by LT was used in a follow-up study by Zaritsky et al. (1989), in which new kinematic measurements were included for the objects Eridanus, Palomar 14, Leo I, and Leo II. One notable result was that the mass estimate changed significantly with the inclusion of the new data; it increased by a factor of two with Eridanus, Palomar 14, and Leo II, and increased by a factor of 3 when Leo I was also included. The mass estimate was also sensitive to the velocity anisotropy assumption, with a radial orbit assumption giving $9.3_{-1.2}^{+4.1} \times 10^{11} M_{\odot}$ and isotropic orbit assumption giving $12.5_{-3.2}^{+8.4} \times 10^{11} M_{\odot}$. Because the results depend heavily on Leo I and assumptions about orbits, Zaritsky et al. (1989) argue that the method introduced by LT does not accurately account for the true uncertainty in the mass estimate. A perhaps more important point, however, is that the LT method does not include measurement uncertainty, which could significantly reduce the leverage of highly uncertain measurements.

Kochanek (1996) used a Bayesian approach and all currently-available data of tracers, including line-of-sight velocities and a few proper motions to estimate the mass of the MW. This study did not incorporate the measurement uncertainties of positions and velocities directly into the Bayesian analysis. Instead, mock data sets were generated from Monte Carlo simulations based on the data and their measurement uncertainties, and the analysis was repeated to estimate the uncertainty in the mass.

Wilkinson & Evans (1999) (hereafter WE99) introduced a truncated, flat rotation curve model for the MW and derived its isotropic DF (an intractable

integral). WE99 found asymptotic approximations for the DF and used it in a Bayesian analysis of 27 MW satellites, of which 6 have proper motion measurements, to estimate the mass of the Milky Way. They investigated how different prior distributions on model parameters and the inclusion or exclusion of Leo I affect the mass estimate. Their results indicate a total mass of $1.9_{-1.7}^{+3.6} \times 10^{12} M_{\odot}$, although their mass estimate within 50kpc ($5.4_{-3.6}^{+0.2} \times 10^{11} M_{\odot}$) is the robust measurement.

WE99 derive a statistical correction factor to account for the heliocentric measurements under the assumption of constant β . However, their data are at large radii from the Galactic centre, and so this correction factor is small. The measurement uncertainties of the proper motions were accounted for by convolving the probability density function with an error function, but other measurement uncertainties were excluded from the analysis. Instead, mock data sets were used to investigate how other factors, such as uncertainties in distance and line-of-sight velocities and correlations within the data, might affect the mass estimate. WE99 conclude that the main sources of uncertainty are the measurement errors. A main result from WE99 is that including the proper motion measurement of Leo I greatly reduced its influence on the mass estimate. Their analysis highlights two strengths of using a DF within a Bayesian approach over traditional mass estimators— incomplete and complete data can be included in the analysis simultaneously, and prior assumptions must be stated explicitly.

Sakamoto et al. (2003) applied the method developed by WE99 to a mixture of Galactic tracers (satellite galaxies, GCs, and BHB stars) to estimate the mass of the MW. They find a lower limit of $1.8_{-0.7}^{+0.4} \times 10^{12} M_{\odot}$ without Leo I, and an upper limit of $2.5_{-1.0}^{+0.5} \times 10^{12}$ with Leo I— results that are in stark

contrast with WE99. The study also depends on Monte Carlo simulations to estimate the uncertainty in the mass estimate due to measurement errors in distances and velocities.

An & Evans (2006) showed that a DF approach suffers from mass-anisotropy degeneracy when only line-of-sight velocities for tracer objects are used. Depending on the assumed anisotropy, the mass of the MW derived from line-of-velocities of distant tracers (i.e. for which v_{los} is approximately the Galactocentric radial velocity) can differ up to a factor of 3. Thus, being able to include any available proper motion measurements in the analysis would be beneficial.

One drawback to using DFs is that few analytic forms are known. Two of these analytic DFs are those found by Hernquist (1990, hereafter the Hernquist model) and Jaffe (1983). These DFs are also *self-consistent*, meaning that the gravitational potential $\Phi(r)$ and the density $\rho(r)$ which determine them make up a potential-density pair which obeys Poisson's equation. A little more recently Evans et al. (1997) derived an analytic DF for a spherical system in which the gravitational potential and the number density of the tracers follow different logarithmic radial profiles, which will be discussed further in Chapter 3.

DFs for a disk, bulge, dark matter halo, and even a central black hole, are available and can be used to generate initial conditions for N-body simulations (e.g. Widrow et al., 2003; Widrow & Dubinski, 2005), however, incorporating these DFs in a maximum likelihood or Bayesian scheme to estimate the total mass of the MW would be computationally difficult. In a Bayesian analysis where Monte Carlo Markov Chains are being created, these DFs would have to be numerically integrated for each combination of model parameter values

at every step in the chain.

Although analytic DFs are difficult to come by, using DFs within a Bayesian approach has still been a fruitful endeavour and continues to hold promise. The major advantages to this approach have been shown above; mainly, the inclusion of incomplete and complete data simultaneously. Further progress can be accomplished by using a hierarchical Bayesian framework that includes a complete measurement model, as opposed to including only the measurement uncertainties of proper motions and relying on Monte Carlo simulations (i.e. the methods used by WE99 and Sakamoto et al., 2003).

1.4 Reviewing Major Challenges

As we have seen, there are many methods to measure the mass of the MW—each having its own advantages and disadvantages. The use of kinematics of satellite objects is the most popular technique, although methods using tidal streams are now becoming more prominent. Common challenges arise in all of the techniques. Here I compile and attempt to create a well-defined list of these challenges: small sample-size and incomplete data, measurement uncertainties, reference frames, and the mass-anisotropy degeneracy. I also briefly describe a challenge that will be encountered in the near future: Big Data.

Sample Size and Incomplete Data

An ideal data set for a tracer population would be large in number and complete in 6-dimensional phase-space information. Unfortunately, kinematic data from MW tracers are often incomplete, and can vary in sample size.

In the past decade, it has become possible to take observations of halo stars

further out in the halo, with BHB stars being a common choice (Kafle et al., 2012; Deason et al., 2012; Xue et al., 2008). BHB stars have several advantages: they have constant absolute magnitude, they are scattered throughout the Galaxy, and they exist as far as $\sim 150\text{kpc}$. However, even though halo stars are numerous, obtaining proper motion measurements of these objects has not been possible beyond the solar neighbourhood until very recently.

Progress is being made with programs such as Gaia and the Large Synoptic Survey Telescope (LSST-Corporation, 2016), which are performing and will perform all-sky surveys. The first release of the Gaia satellite data (DR1) has been cross referenced with the Hipparcos catalogue to estimate the proper motions of about 2 million stars using the Tycho-Gaia astrometric solution (TGAS) (Michalik et al., 2015; ESA, 2016; Lindegren et al., 2016; Gaia Collaboration et al., 2017). However, the proper motions of these halo stars will not be very reliable for distances beyond $\sim 20\text{kpc}$ from the Galactic center (Bailer-Jones, 2004). The LSST data will not be available for a number of years, but when it is released we need to be ready with reliable mass estimation methods that can include incomplete and complete data simultaneously, and efficiently.

Measurement Uncertainties

The magnitudes of measurement uncertainty in tracer data are mixed, with some velocity measurements being very uncertain and others being quite precise. How one incorporates this inhomogeneous measurement uncertainty is an important choice, and ideally all measurement uncertainties would be propagated through to the final mass estimate. Unfortunately, current methods usually require simplifying assumptions about error distributions and/or rely

on simulations to estimate statistical error.

Reference Frames

Mathematical models of the Galaxy are simplest and easiest to use in a Galactocentric reference frame, but our measurements of MW tracers are by default in a heliocentric frame. Both line-of-sight and proper motion measurements (i.e. complete data) are needed to transform a velocity vector from the heliocentric frame to the Galactocentric frame. Because tracer data tend to be incomplete, studies often limit their sample to distant tracers where $v_{los} \approx v_r$. Furthermore, mass estimators derived from the Jeans equation use only incomplete or complete data; in either case, information is being thrown away. Transforming incomplete measurements and propagating measurement uncertainties into a Galactocentric frame is particularly challenging, and propagating uncertainties into the final mass estimate is difficult and often avoided (e.g. instead, simulations are used to estimate the statistical uncertainty in the mass).

Mass Anisotropy Degeneracy

Many of the techniques outlined in Section 1.3 rely on assumptions about the tracer population's velocity dispersion, and I reviewed how assumptions about β can dramatically alter the mass estimate in Jeans equation analyses. One reason for these assumptions is the lack of proper motion data, and limited techniques for including incomplete and complete data simultaneously in the analysis. It would be better to allow β to be a free parameter in the analysis, and to estimate its value based on all available data.

Big Data

Astronomy is entering an era of *Big Data*. Data from Gaia and LSST are and will provide kinematic data of millions of halo stars, including many proper motions. However, not all of these data will be complete. Thus new computational challenges will present themselves, and the development of statistically rigorous methods for using this kinematic data will be needed.

1.5 Contribution of this Research

The main goal of my thesis research is to develop a method for estimating the mass and mass profile of the MW galaxy that overcomes most of the challenges associated with using satellite kinematics and that more easily allows for comparison to other mass estimates in the literature. I have developed a Bayesian hierarchical model that employs an analytic DF and which infers the mass and mass profile of the MW from the kinematics of tracers. The results are presented in the collection of papers in this thesis.

For the majority of this thesis work, I have been using GCs as a test data set for developing the hierarchical Bayesian method. GCs are compact star clusters which are spherical in shape, gravitationally bound, and kinematically self-supported, so treating them as point mass particles with respect to the Galaxy is not a gross approximation. GCs are some of the oldest objects in the universe, and are thought to have been either accreted by the Milky Way during our Galaxy's formation, created in-situ within our Galaxy, or both. The GCs have been in orbit around the Galaxy for billions of years, making them a nearly virialized population. Because of this property of the MW's GCs, these

objects make good kinematic tracers of the Milky Way’s total gravitational potential in the context of a DF method, and are thus a good data set to work with while developing the Galaxy mass estimation technique.

GCs in the MW are a modest data set at 157 objects, and the velocity measurements are proportionally more complete than halo stars. The GC data set is missing proper motion measurements for approximately 50% of the GCs, but this is gradually improving thanks to the HST Proper Motion Collaboration (HSTPROMO) (van der Marel et al., 2014; van der Marel, 2016) and others (Fritz et al., 2017; Watkins & van der Marel, 2017; Massari et al., 2017)³. HSTPROMO is also obtaining proper motions for a number of DGs, as other studies continue to do so as well (e.g. Piatek et al., 2016; Cioni et al., 2016, and others).

The work presented here was initially built upon the research I completed during my Masters, which involved a statistical study of Bayesian mass estimates of very simple simulated “galaxies”. These galaxies were not cosmologically motivated simulations, but instead generated in a statistical manner; they were composed of simulated point particles that followed a Hernquist spatial and kinematic distribution. I investigated the statistical properties of Bayesian mass estimators when the kinematic data was entirely incomplete (i.e. only line-of-sight measurements were known) or entirely complete, and when the assumed model had the incorrect velocity anisotropy.

During my PhD, I have made substantial and significant improvements to this method with the goal of overcoming the challenges presented in Section 1.4. The advances in this method have been incremental but important,

³*enabled by the Gaia data*

and the subsequent hierarchical Bayesian formalism has allowed us to overcome major challenges associated with using kinematic data of MW satellites. A hierarchical Bayesian method provides several advantages because:

1. ... both incomplete and complete data can be used simultaneously in the analysis. This is especially important for the use of GCs in the analysis, for which approximately 50% of the data are missing proper motion measurements.
2. ... the true positions and velocities of the satellites are treated as parameters. These parameters are constrained with the known measurement uncertainties, which are well understood, through a measurement model.
3. ... the measurement model is in the heliocentric frame, and thus negates the need for complex propagation of uncertainties to the Galactocentric frame. The varying uncertainties are naturally carried through to the posterior distribution via the hierarchical model.
4. ... the resulting posterior distribution gives probability regions for model parameters and credible intervals for physically meaningful quantities like the virial mass and the velocity anisotropy of the tracer population.
5. ... the method not only produces total mass estimates for the MW but also predicts a mass profile for the Galaxy so that comparisons may be made with other mass estimates at any radius.

The chapter contents in this thesis naturally walk through the incremental steps that led to the hierarchical Bayesian model.

First, an introduction to the method developed during my Masters, including Bayes' Theorem and sampling algorithms (e.g. Monte Carlo Markov Chain

and Gibbs Sampling) in the context of using the DF to analyse kinematic data, is provided. The method is then altered to include complete and incomplete data simultaneously and in arbitrary mixtures. Simulated incomplete data are constructed and used to test this new method in light of different velocity anisotropy assumptions. A preliminary analysis of MW GC and DG data is then performed to obtain a mass estimate for the MW under a Hernquist model assumption (Chapter 2). Not all GCs are included in this analysis because only GCs for which the line-of-sight velocity is a good approximation to the Galactocentric radial velocity could be included.

Second, a more flexible DF from the literature is incorporated that allows the number density profile of the MW satellites and the dark matter to follow different spatial distributions. This model as it applies to the MW GC data is explored through a series of investigations and sensitivity tests (Chapter 3).

Third, measurement uncertainties are incorporated via a hierarchical Bayesian framework. The hierarchical framework includes a measurement model in the heliocentric frame, which not only makes propagation of errors unnecessary, but which also allows more of the incomplete data to be included in the analysis. With the ability to include all of the available information, we subsequently arrive at a new mass estimate for the Galaxy using all of the MW's known GC population. (Chapter 4).

Finally, we are currently testing the hierarchical method from Chapter 4 on mock observations of simulated MW-type data from hydrodynamical cosmological simulations (Chapter 5). The latter is a paper in preparation, to be submitted to the *The Astrophysical Journal*. In Chapter 6, I summarize the overall findings of this thesis and discuss avenues of future work.

Bibliography

- An, J. H., & Evans, N. W. 2006, *AJ*, 131, 782
- Baes, M., Buyle, P., Hau, G. K. T., & Dejonghe, H. 2003, *MNRAS*, 341, L44
- Bailer-Jones, C. A. L. 2004, in *Proceedings of the IAU Colloquium No. 196*, ed. W. Kurtz & G. Bromage (Cambridge Journals), 429–443
- Barber, C., Starkenburg, E., Navarro, J. F., McConnachie, A. W., & Fattahi, A. 2014, *MNRAS*, 437, 959
- Bartelmann, M. 2010, *Classical and Quantum Gravity*, 27, 233001
- Battaglia, G., Helmi, A., Morrison, H., et al. 2005, *MNRAS*, 364, 433
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57
- Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006, *ApJL*, 642, L137
- Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, *ApJ Supp*, 208, 20
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics*, 2nd edn. (Princeton)
- Bonaca, A., Geha, M., Küpper, A. H. W., et al. 2014, *ApJ*, 795, 94
- Bovy, J., Bahmanyar, A., Fritz, T. K., & Kallivayalil, N. 2016, *ApJ*, 833, 31
- Bowden, A., Evans, N. W., & Williams, A. A. 2016, *MNRAS*, 460, 329

- Boylan-Kolchin, M., Besla, G., & Hernquist, L. 2011a, MNRAS, 414, 1560
- Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011b, MNRAS, 415, L40
- Boylan-Kolchin, M., Bullock, J. S., Sohn, S. T., Besla, G., & van der Marel, R. P. 2013, ApJ, 768, 140
- Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150
- Busha, M. T., Marshall, P. J., Wechsler, R. H., Klypin, A., & Primack, J. 2011, ApJ, 743, 40
- Carlesi, E., Hoffman, Y., Sorce, J. G., & Gottlöber, S. 2017, MNRAS, 465, 4886
- Cautun, M., Frenk, C. S., van de Weygaert, R., Hellwing, W. A., & Jones, B. J. T. 2014, MNRAS, 445, 2049
- Cioni, M.-R. L., Bekki, K., Girardi, L., et al. 2016, A&A, 586, A77
- Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, ApJL, 648, L109
- Cole, S., & Lacey, C. 1996, MNRAS, 281, 716
- Côté, P., McLaughlin, D. E., Cohen, J. G., & Blakeslee, J. P. 2003, ApJ, 591, 850
- Deason, A. J., Belokurov, V., & Evans, N. W. 2011, MNRAS, 411, 1480
- Deason, A. J., Belokurov, V., Evans, N. W., et al. 2012, MNRAS, 425, 2840
- Deg, N., & Widrow, L. 2013, MNRAS, 428, 912

Diaz, J. D., Koposov, S. E., Irwin, M., Belokurov, V., & Evans, N. W. 2014, MNRAS, 443, 1688

Dierickx, M. I. P., & Loeb, A. 2017, ArXiv e-prints, arXiv:1703.02137

Ehlers, J. 2008, General Relativity and Gravitation, 41, 203

ESA. 2016, Data Release Scenario, <http://www.cosmos.esa.int/web/gaia/release>, accessed: 2016-05-10

Evans, N. W., Hafner, R. M., & de Zeeuw, P. T. 1997, MNRAS, 286, 315

Eyre, A., & Binney, J. 2011, MNRAS, 413, 1852

Fattahi, A., Navarro, J. F., Sawala, T., et al. 2016, MNRAS, 457, 844

Ferrarese, L. 2002, ApJ, 578, 90

Fragione, G., & Loeb, A. 2017, New Astronomy, 55, 32

Fritz, T. K., Linden, S. T., Zivick, P., et al. 2017, ApJ, 840, 30

Gaia Collaboration, van Leeuwen, F., Vallenari, A., et al. 2017, A&A, 601, A19

Gibbons, S. L. J., Belokurov, V., & Evans, N. W. 2014, MNRAS, 445, 3788

Gnedin, O. Y., Brown, W. R., Geller, M. J., & Kenyon, S. J. 2010, ApJL, 720, L108

Gnedin, O. Y., Gould, A., Miralda-Escudé, J., & Zentner, A. R. 2005, ApJ, 634, 344

Gove, W. R., & Hughes, M. 1980, Social Forces, 58, 1157

- Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, MNRAS, 404, 1111
- Hartwick, F. D. A., & Sargent, W. L. W. 1978, ApJ, 221, 512
- Helmi, A. 2004, MNRAS, 351, 643
- Hernquist, L. 1990, ApJ, 356, 359
- Hills, J. G. 1988, Nature, 331, 687
- Ibata, R., Irwin, M., Lewis, G. F., & Stolte, A. 2001a, ApJL, 547, L133
- Ibata, R., Lewis, G. F., Irwin, M., Totten, E., & Quinn, T. 2001b, ApJ, 551, 294
- Jaffe, W. 1983, MNRAS, 202, 995
- Johnson, D. R. H., & Soderblom, D. R. 1987, The Astronomical Journal, 93, 864
- Kafle, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, ApJ, 761, 98
- Kahn, F. D., & Woltjer, L. 1959, ApJ, 130, 705
- Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102
- Kochanek, C. S. 1996, ApJ, 457, 228
- Koposov, S. E., Rix, H.-W., & Hogg, D. W. 2010, ApJ, 712, 260
- Kuhlen, M., Diemand, J., & Madau, P. 2007, ApJ, 671, 1135

- Kulesa, A. S., & Lynden-Bell, D. 1992, MNRAS, 255, 105
- Küpper, A. H. W., Balbinot, E., Bonaca, A., et al. 2015, ApJ, 803, 80
- Law, D. R., & Majewski, S. R. 2010, ApJ, 714, 229
- Leonard, P. J. T., & Tremaine, S. 1990, ApJ, 353, 486
- Li, Y.-S., & White, S. D. M. 2008, MNRAS, 384, 1459
- Lindegren, L., Lammers, U., Bastian, U., et al. 2016, A&A, 595, A4
- Little, B., & Tremaine, S. 1987, ApJ, 320, 493
- Loebman, S. R., Ivezić, Ž., Quinn, T. R., et al. 2014, ApJ, 794, 151
- Loebman, S. R., Valluri, M., Hattori, K., et al. 2017, ArXiv e-prints, arXiv:1704.06264
- Lokas, E. L. 2009, MNRAS, 394, L102
- LSST-Corporation. 2016, LSST Project Schedule, <https://www.lsst.org>, accessed: 2016-08-10
- Lynden-Bell, D., Cannon, R. D., & Godwin, P. J. 1983, MNRAS, 204
- Lynden-Bell, D., & Frenk, C. S. 1981, The Observatory, 101, 200
- Lynds, R., & Petrosian, V. 1989, ApJ, 336, 1
- Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, ApJ, 599, 1082
- Markevitch, M., Gonzalez, A. H., Clowe, D., et al. 2004, ApJ, 606, 819

- Massari, D., Posti, L., Helmi, A., Fiorentino, G., & Tolstoy, E. 2017, *A&A*, 598, L9
- McMillan, P. J. 2011, *MNRAS*, 414, 2446
- Michalik, D., Lindegren, L., & Hobbs, D. 2015, *A&A*, 574, A115
- Milgrom, M. 1983, *ApJ*, 270, 365
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Nelson, D., Pillepich, A., Genel, S., et al. 2015, *Astronomy and Computing*, 13, 12
- Olszewski, E. W., Aaronson, M., & Peterson, R. C. 1986, *ApJL*, 302, L45
- Patel, E., Besla, G., & Mandel, K. 2017, *MNRAS*, 468, 3428
- Peñarrubia, J., Koposov, S. E., & Walker, M. G. 2012, *ApJ*, 760, 2
- Peñarrubia, J., Ma, Y.-Z., Walker, M. G., & McConnachie, A. 2014, *MNRAS*, 443, 2204
- Pearson, S., Küpper, A. H. W., Johnston, K. V., & Price-Whelan, A. M. 2015, *ApJ*, 799, 28
- Perets, H. B. 2009, *ApJ*, 690, 795
- Perets, H. B., Wu, X., Zhao, H. S., et al. 2009, *ApJ*, 697, 2096
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339
- Piantadosi, S., Byar, D. P., & Green, S. B. 1988, *American journal of epidemiology*, 127, 893

- Piatek, S., Pryor, C., & Olszewski, E. W. 2016, *AJ*, 152, 166
- Piff, T., Williams, M., & Steinmetz, M. 2011, *A&A*, 535, A70
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13
- Price-Whelan, A. M., Hogg, D. W., Johnston, K. V., & Hendel, D. 2014, *ApJ*, 794, 4
- Robinson, W. S. 1950, *American Sociological Review*, 15, 351
- Roos, M. 2010, *ArXiv e-prints*, arXiv:1001.0316
- Rubin, V. C., Ford, W. K. J., & Thonnard, N. 1980, *ApJ*, 238, 471
- Sakamoto, T., Chiba, M., & Beers, T. C. 2003, *A&A*, 397, 899
- Sanders, J. L. 2014, *MNRAS*, 443, 423
- Sanderson, R. E., Hartke, J., & Helmi, A. 2017, *ApJ*, 836, 234
- Smith, M. C., Wyn Evans, N., & An, J. H. 2009, *ApJ*, 698, 1110
- Smith, M. C., Ruchti, G. R., Helmi, A., et al. 2007, *MNRAS*, 379, 755
- Sohn, S. T., Anderson, J., & van der Marel, R. P. 2012, *ApJ*, 753, 7
- Soucail, G., Fort, B., Mellier, Y., & Picat, J. P. 1987, *A&A*, 172, L14
- Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, *ApJ Supp*, 148, 175
- Springel, V., & Hernquist, L. 2003, *MNRAS*, 339, 312
- Springel, V., Wang, J., Vogelsberger, M., et al. 2008, *MNRAS*, 391, 1685

- van der Marel, R. P. 2016, Hubble Space Telescope Proper Motion Collaboration, <http://www/stsci.edu/~marel/hstpromo.html>, accessed: 2016-10-16
- van der Marel, R. P., Besla, G., Cox, T. J., Sohn, S. T., & Anderson, J. 2012a, *ApJ*, 753, 9
- van der Marel, R. P., Fardal, M., Besla, G., et al. 2012b, *ApJ*, 753, 8
- van der Marel, R. P., Anderson, J., Bellini, A., et al. 2014, in *Astronomical Society of the Pacific Conference Series*, Vol. 480, *Structure and Dynamics of Disk Galaxies*, ed. M. S. Seigar & P. Treuhardt, 43
- Varghese, A., Ibata, R., & Lewis, G. F. 2011, *MNRAS*, 417, 198
- Vera-Ciro, C. A., Sales, L. V., Helmi, A., et al. 2011, *MNRAS*, 416, 1377
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, *MNRAS*, 444, 1518
- . 2014b, *Nature*, 509, 177
- Wang, W., Han, J., Cooper, A., et al. 2015, *MNRAS*, 453, 377
- Watkins, L., Evans, N., & An, J. 2010, *MNRAS*, 406, 264
- Watkins, L. L., & van der Marel, R. P. 2017, *ApJ*, 839, 89
- White, M. 2001, *A&A*, 367, 27
- Widrow, L. M., & Dubinski, J. 2005, *ApJ*, 631, 838
- Widrow, L. M., Perrett, K. M., & Suyu, S. H. 2003, *ApJ*, 588, 311
- Wilkinson, M., & Evans, N. 1999, *MNRAS*, 310
- Xue, X., Rix, H., Zhao, G., et al. 2008, *ApJ*, 684, 1143

Yu, Q., & Madau, P. 2007, *Monthly Notices of the Royal Astronomical Society*,
379, 1293

Zaritsky, D., Olszewski, E. W., Schommer, R. A., Peterson, R. C., & Aaronson,
M. 1989, *ApJ*, 345, 759

Zwicky, F. 1933, *Helvetica Physica Acta*, 6, 110

—. 1937, *ApJ*, 86, 217

2

Estimating the Galactic Mass Profile in the Presence of Incomplete Data

Reprinted from *Eadie, G., Harris, W., & Widrow, L (2015)*, THE ASTROPHYSICAL JOURNAL, Vol. 806, No. 54. DOI: 10.1088/0004-637X/806/1/54.

Published by the American Astronomical Society ©AAS

Sections from my MSc thesis (*Eadie, G.M. 2013, Master's Thesis, QUEEN'S UNIVERSITY, Kingston, Ontario, Canada*), which had not previously been published in an academic journal, were used in the introduction and theory sections of this paper.

Abstract

A powerful method to measure the mass profile of a galaxy is through the velocities of tracer particles distributed through its halo. Transforming this kind of data accurately to a mass profile $M(r)$, however, is not a trivial problem. In particular, limited or incomplete data may substantially affect the analysis. In this paper we develop a Bayesian method to deal with incomplete data effectively; we have a hybrid-Gibbs sampler that treats the unknown velocity components of tracers as parameters in the model. We explore the effectiveness of our model using simulated data, and then apply our method to the Milky Way using velocity and position data from globular clusters and dwarf galaxies. We find that in general, missing velocity components have little effect on the total mass estimate. However, the results are quite sensitive to the outer globular cluster Pal 3. Using a basic Hernquist model with an isotropic velocity dispersion, we obtain credible regions for the cumulative mass profile $M(r)$ of the Milky Way, and provide estimates for the model parameters with 95% Bayesian credible intervals. The mass contained within 260kpc is $1.37 \times 10^{12} M_{\odot}$, with a 95% credible interval of $(1.27, 1.51) \times 10^{12} M_{\odot}$. The Hernquist parameters for the total mass and scale radius are $1.55^{+0.18}_{-0.13} \times 10^{12} M_{\odot}$ and $16.9^{+4.8}_{-4.1}$ kpc, where the uncertainties span the 95% credible intervals. The code we developed for this work, Galactic Mass Estimator (GME), will be available as an open source package in the R Project for Statistical Computing.

2.1 Introduction

Almost every galaxy in the universe is assumed to reside in a massive, dark matter halo that can extend far beyond the visible components of the galaxy. Standard methods to determine the mass distribution within visible portions of galaxies are based on rotation curves or velocity dispersion profiles. The former is applicable to spiral galaxies, while the latter method is used mainly for elliptical galaxies or disk galaxy bulges. At large radii, where the mass distribution is presumably dominated by dark matter, one can use observations of kinematic tracers to learn about a galaxy's mass profile.

The Milky Way (MW) has many distant satellites, such as globular clusters, halo stars, planetary nebulae, and dwarf galaxies (DGs). The kinematic properties of these satellites can be used to learn about the gravitational potential of the whole system, and thus the Galaxy's mass profile out to large radii.

One way to use the kinematic data of tracers to estimate the mass distribution is via a *mass estimator*, a method first suggested by Hartwick & Sargent (1978) and a method that avoids using an explicit model. The method has endured partly because it uses only the line of sight velocities and positions of the tracers. It is useful when no proper motions are available and conversion to a Galactocentric reference frame is impossible. Nowadays, however, many MW tracers have proper motion measurements, and more continue to become available. Thus, it would be beneficial to have a method which can incorporate both complete data (i.e. tracers with 3-dimensional space motions in the Galactocentric frame) and incomplete data (i.e. tracers with only line of sight velocities). In this paper we introduce such a method of mass estimation.

The method used here is based on the phase-space distribution function (DF), which is a probability distribution for a satellite in terms of its position \mathbf{r} and velocity \mathbf{v} in the Galactocentric reference frame. Using the DF for a model and a Bayesian approach to the analysis, we obtain probability distributions for the model parameters (and thus the mass). The method was first suggested by Little & Tremaine (1987), who showed how to use the DF and a Bayesian approach to estimate the mass specifically for the Milky Way. Since then, other studies have used the Bayesian approach or Maximum Likelihood methods for Galactic mass estimation (e.g. Kulessa & Lynden-Bell, 1992; Kochanek, 1996; Wilkinson & Evans, 1999; Widrow et al., 2008). Deriving analytic and physically relevant DFs has been explored by Hernquist (1990), Cuddeford (1991), Ciotti (1996), Widrow (2000), and Evans & Williams (2014), to name a few.

A DF's derivation and final form is, by default, in the Galactocentric reference frame, but previous studies have re-written DFs in terms of the line-of-sight velocity component only, in order to incorporate incomplete data (e.g. Wilkinson & Evans, 1999). This does not take full advantage of the complete data that is available, which is an issue when the method may be susceptible to large uncertainties due to small sample size (as discussed by Zaritsky et al., 1989). Furthermore, rewriting the DF in terms of the line-of-sight velocity can be mathematically difficult. In our study, we introduce a generalized approach via the Bayesian framework, whereby it is easy to incorporate complete and incomplete data simultaneously, and also unnecessary to rewrite the DF in terms of the line-of-sight velocity.

The purpose of this paper is to lay out the method fully, and set the groundwork for future studies with a range of DFs and datasets. We also test

the method with simulated data, and do some preliminary analysis with the method as it applies to the Milky Way.

The outline of this paper is as follows. Section 2.2.1 briefly describes the theoretical background of DFs, and Section 2.2.2 introduces two models that already have analytic DFs, which we use in this work. Next, in Section 2.2.3, we review how Bayes' Theorem can be used with the DFs to obtain parameter estimates of the model. Sections 2.2.4, 2.2.5, and 2.2.6 introduce the method for incorporating both complete and incomplete data via a hybrid-Gibbs Sampler, and Section 2.2.7 discusses the techniques used to assess convergence of the Markov chains.

We first apply and test our method on simulated data (described in Section 2.3) and then apply the method to some kinematic data of satellites orbiting the Milky Way (described in Section 2.4). The results of these analyses are presented and discussed in Sections 2.5 and 2.6 respectively. Many future prospects are also discussed in Section 2.6.

2.2 Background

This section provides background information and notation about distribution functions, the models we use in our analysis, and the methods of Bayesian inference as they apply to the current problem. Additional details and discussion can be found in Eadie (2013).

2.2.1 Distribution Function

The Distribution Function (DF), $f(\mathbf{r}, \mathbf{v})$, is a probability density function that gives the probability of finding a particle with a position \mathbf{r} and velocity \mathbf{v} within

a phase-space volume element $d^3\mathbf{r}d^3\mathbf{v}$ (Binney & Tremaine, 2008). Like any probability density, the DF integrates to one:

$$\int f(\mathbf{r}, \mathbf{v})d^3\mathbf{r}d^3\mathbf{v} = 1. \quad (2.1)$$

Eq. 2.1 is often renormalized so that the DF integrates to a quantity of interest, such as the total mass M_{tot} :

$$\int f(\mathbf{r}, \mathbf{v})d^3\mathbf{r}d^3\mathbf{v} = M_{\text{tot}}. \quad (2.2)$$

However, in a Bayesian framework the DF as defined in eq. 2.1 is used, and thus the left-hand-side of eq. 2.2 is divided by M_{tot} . Thus, it is important that models have a finite mass in a Bayesian analysis— if the mass is infinite, then the DF is not a proper probability distribution.

A DF can be specified by use of Jeans' Theorem, which states that any solution of the time-independent collisionless Boltzmann equation is a function of the phase-space coordinates (\mathbf{r}, \mathbf{v}) only. In a time-independent system, the Hamiltonian $H = \frac{v^2}{2} + \Phi(r)$ is always an integral of motion, and if the system is also spherical then the magnitude of the angular momentum, L , is an integral of motion too. Therefore, any non-negative function $f(H)$ or $f(H, L)$ will be a solution to the time-independent collisionless Boltzmann equation, and thus a DF for the system. Whether or not f is a function of H or both H and L determines the velocity dispersion of the system; $f(H)$ corresponds to an isotropic system, and $f(H, L)$ corresponds to an anisotropic system.

In practice, a DF corresponding to an isotropic, spherical, self-consistent system is usually written in terms of \mathcal{E} , the relative energy per unit mass,

defined as

$$\mathcal{E} = -\frac{v^2}{2} + \Psi(r) \tag{2.3}$$

where v is the speed of a particle at a distance r from the center of the system, and $\Psi(r)$ is the relative gravitational potential of the system at r , as defined in Binney & Tremaine (2008). Particles with $\mathcal{E} \leq 0$ are unbound and require $f = 0$. If the system has an anisotropic velocity dispersion, then the DF is written as a function of both \mathcal{E} and the angular momentum L .

2.2.2 Models

Models with analytic DFs are preferable to empirical distribution functions in theoretical analyses because they allow for easy sampling of the distribution, and also save computation time by avoiding numerical integration. Finding a DF that models a realistic galaxy is a difficult task, however, because galaxies are often composed of multiple subsystems such as a bulge, a stellar halo, a dark matter halo, and possibly a disk. Finding a single phase-space distribution function that is self-consistent, analytic, and that describes the intricate features of a galaxy is very challenging.

An empirical luminosity profile that has been successful in fitting the surface brightness profiles of elliptical galaxies and bulges is the de Vaucouleurs (1948) $R^{1/4}$ profile. A generalization of the $R^{1/4}$ profile is $R^{1/n}$, which was introduced by Sersic (1968). Due to the success of $R^{1/4}$, theorists have tried to develop distribution functions that can reproduce the profile. The analytic models introduced by Jaffe (1983) and Hernquist (1990) fit the $R^{1/4}$ type galaxies well for most radii.

In this work, primarily for the purpose of testing the method, we use the

Hernquist and Jaffe models because of their analytic simplicity and their ubiquity. The Hernquist model also has the benefit of having more than one analytic DF— one has an isotropic velocity dispersion, and there are a few that are anisotropic. Furthermore, both the Hernquist and Jaffe models are self-consistent and have a finite total mass, making consistent numerical computations feasible. For these reasons, we use the Hernquist-style models to lay out the methodology of our Bayesian approach and the derivation of mass profile credible regions. In future work, we will extend the method to include models with non-analytic DFs and non-finite mass distributions, such as the Navarro et al. (1996) (NFW) model.

Hernquist (1990) introduced a halo model that is a self-consistent, analytic potential-density pair. With $G \equiv 1$, the gravitational potential of the Hernquist model is

$$\Phi(r) = -\frac{M_{\text{tot}}}{r+a} \quad (2.4)$$

and the mass density profile is

$$\rho(r) = \frac{aM_{\text{tot}}}{2\pi r(r+a)^3} \quad (2.5)$$

where M_{tot} is the total mass of the system, and a is the scale radius. Integrating over a sphere, the Hernquist cumulative mass profile is then,

$$M(r) = M_{\text{tot}} \frac{r^2}{(r+a)^2} \quad (2.6)$$

Hernquist (1990) provides two DFs for their model that are written in terms of elementary functions: one for an isotropic velocity dispersion, and a second for an anisotropic velocity dispersion of the Osipkov (1979) and Merritt (1985)

type (hereafter OM-type). The OM-type allows the anisotropy to vary as a function of r , and includes a constant parameter called the anisotropy radius r_a :

$$\beta(r) = \frac{r^2}{r^2 + r_a^2} \quad (2.7)$$

The parameter r_a controls the degree of radial anisotropy in the system at different radii. As $r_a \rightarrow \infty$, $\beta(r) \rightarrow 0$ (completely isotropic).

The third Hernquist DF used in this research has a constant anisotropy $\beta = 0.5$, which was derived by Evans & An (2006). We consider this model because recent research by Deason et al. (2013) showed that blue horizontal branch (BHB) stars have a radially biased velocity anisotropy of 0.5 between 16 and 28kpc, suggesting that β may be approximately constant for most of the stellar halo.

We also consider the isotropic Jaffe (1983) model in this work, which has mass profile and potential

$$M_J(r) = M_{\text{tot}} \frac{r}{a_J(r + a_J)} \quad (2.8)$$

$$\Phi_J(r) = M_{\text{tot}} \frac{\ln(1 + a_J/r)}{a_J} \quad (2.9)$$

where a_J is the scale radius.

Overall, we fit the Milky Way data to four different models: three Hernquist models with different anisotropies (isotropic, OM-type, constant anisotropy ($\beta = 0.5$), and the isotropic Jaffe model.

2.2.3 Bayes Theorem and Parameter Estimation

Bayes' Theorem is named after Thomas Bayes (1701-1761) and was introduced posthumously by Richard Price (Bayes & Price, 1763). Using the rules of conditional probabilities, Bayes showed that the conditional probability $p(A|B)$ is,

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (2.10)$$

now known as Bayes' Theorem.

Bayesian inference involves using eq. 2.10 in data analysis to obtain probability distributions about model parameters. Bayesian inference returns a probability distribution for parameters given the data and a prior distribution on the parameters.

When Bayes' Theorem is used for Bayesian inference, it is re-written in terms of the vector of model parameters $\boldsymbol{\theta}$ and the data y , and is sometimes referred to as Bayes' rule. The Bayesian posterior probability for $\boldsymbol{\theta}$, given some data y , is then

$$p(\boldsymbol{\theta}|y) = \frac{p(y|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(y)} \quad (2.11)$$

where $p(y|\boldsymbol{\theta})$ is called the *likelihood*, $p(\boldsymbol{\theta})$ is the *prior probability* on the parameters, and $p(y)$ is the *marginal probability* of the data. Because the marginal probability does not depend on $\boldsymbol{\theta}$, and with fixed y can be considered a constant, it is common practice to sample the unnormalized posterior probability,

$$p(\boldsymbol{\theta}|y) \propto p(y|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (2.12)$$

The pioneering work by Bahcall & Tremaine (1981) showed that the DF

determines the likelihood $p(y|\boldsymbol{\theta})$. For example, the probability of the first satellite in our data set having r_1 and \mathbf{v}_1 , given the model parameters, is $f(r_1, \mathbf{v}_1|\boldsymbol{\theta})$. Assuming all n satellites in the data set are independent, the probability of their corresponding positions and velocities is the product of the DFs, and thus the likelihood is

$$p(y|\boldsymbol{\theta}) = \prod_{i=1}^n f(r_i, \mathbf{v}_i|\boldsymbol{\theta}). \quad (2.13)$$

Therefore, eq 2.12 becomes

$$p(\boldsymbol{\theta}|y) \propto \prod_{i=1}^n f(r_i, \mathbf{v}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (2.14)$$

Sampling eq. 2.14 is usually done via a Markov Chain Monte Carlo (MCMC) method, which creates a Markov chain— a sequence of random variables, $\boldsymbol{\theta}^t$, where $t = 1, 2, 3, \dots$ represents the position in the chain. Every random variable in the chain depends only on the variable before it, $\boldsymbol{\theta}^{t-1}$ (Gelman et al., 2003). When MCMC algorithms are used to sample a Bayesian posterior density, then by construction, the Markov chain is a collection of parameter vectors that have the same stationary distribution as the posterior (eq. 2.11).

We apply the Metropolis algorithm (Metropolis & Ulam, 1949; Metropolis et al., 1953) to sample eq. 2.14. The Metropolis algorithm is iterative and creates a Markov chain whose stationary distribution is proportional to the Bayesian posterior probability in question. The Markov chains in this work are constructed as follows (Gelman et al., 2003):

1. Draw a trial value $\boldsymbol{\theta}^*$ from a symmetric proposal distribution
2. Calculate $d = \frac{p(\boldsymbol{\theta}^*|y)}{p(\boldsymbol{\theta}^{t-1}|y)}$

3. If $d > 1$, then accept θ^* as θ^t
 - (a) set $\theta^t = \theta^*$
 - (b) return to step 1

4. If instead $d < 1$, then accept θ^* with probability d
 - (a) draw a random number z from the uniform distribution $U(0, 1)$
 - (b) if $d > z$ then accept θ^* as in step 3, and return to step 1
 - (c) if $d < z$ then reject θ^*
 - i. set $\theta^t = \theta^{t-1}$
 - ii. return to step 1

Accepting θ^* only when $d > z$ ensures that the θ values are accepted with probability proportional to the posterior, provided that the chain has converged to the target distribution. The above process is repeated N times, resulting in a Markov chain with N values of θ which represents samples from the posterior.

Because a Bayesian analysis leads to distributions for parameters, the results are arguably easier to interpret. The Markov chains can be used to acquire estimates, uncertainties, and probabilities pertaining to model parameters. Furthermore, the uncertainties can be carried forward to subsequent modeling and analysis.

A Bayesian analysis can also easily include nuisance parameters— parameters in the model whose values are unknown but not necessarily of interest to the researcher. This feature turns out to be useful in our current problem of galaxy mass estimation; sometimes we do not know the tangential velocity component of a satellite object, but we can treat that unknown component as a nuisance parameter in the model.

2.2.4 Nuisance (v_t) Parameters

In a Galactocentric coordinate system, the total speed of a satellite in orbit around the Galaxy can be written

$$v = \sqrt{v_r^2 + v_t^2} \quad (2.15)$$

where v_r and v_t are the radial and tangential components respectively. In turn,

$$v_t^2 = v_\phi^2 + v_\theta^2. \quad (2.16)$$

The total speed of a satellite is needed for the DF $f(r, v|\boldsymbol{\theta})$ in the likelihood of Bayes' theorem. However, proper motion measurements are not available for all satellites. In many cases, distant tracers of the MW have only line-of-sight velocity measurements with respect to our position in the Galaxy. We want to use this satellite information in our analyses, but without a proper motion, the line-of-sight velocity in the local standard of rest frame does not give us v_r or v_t in Galactocentric coordinates. For very distant objects, the line-of-sight velocity is approximately $v_{\text{los}} \approx v_r$, since the angle created by the location of the Sun, the satellite, and the center of the galaxy is quite small. However, we still have no value for v_t . If we treat these unknown v_t 's as nuisance parameters in the model, then Bayes' rule reads,

$$p(\boldsymbol{\theta}|y) \propto \prod_{i=1}^n f(r_i, v_{r,i}|\boldsymbol{\theta}, v_{t,i})p(\boldsymbol{\theta})p(v_{t,i}) \quad (2.17)$$

where $p(v_{t,i})$ is the prior probability on the tangential velocity of the i th satellite, to be discussed in section 2.2.6.

2.2.5 The Gibbs Sampler

When nuisance (v_t) parameters are present, we use a Gibbs sampler, which was first introduced by Geman & Geman (1984) in the area of image processing, and then adapted to iterative simulations in the study of statistics by Tanner & Wong (1987). Gelfand & Smith (1990) then showed how to apply it to Bayesian inference. Since then, the Gibbs sampler has been applied to many problems (Gelman et al., 2003, and references therein).

The Gibbs sampler is sometimes called alternating conditional sampling, and can be very useful in multi-dimensional problems where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ (Gelman et al., 2003). The Gibbs algorithm samples each of the parameters $(\theta_1, \dots, \theta_n)$ one at a time, based on the current value of all of the other parameters and the conditional probability given those parameters.

Consider θ_i , the i th parameter in a model with n parameters, and let θ_{-i} represent all of the other parameters. Next, let t be the t^{th} iteration of the chain—the chain that will be a sample of the posterior distribution $p(\boldsymbol{\theta}|y)$. In the Gibbs sampler, each θ_i is sampled one at a time based on its conditional probability given the current values of all of the other parameters,

$$p(\theta_i|\theta_{-i}^{t-1}, y) \tag{2.18}$$

where θ_{-i}^{t-1} represents the other parameters at their current value,

$$\theta_{-i}^{t-1} = (\theta_1^t, \dots, \theta_{i-1}^t, \theta_{i+1}^{t-1}, \dots, \theta_n^{t-1}).$$

In our work, we do not directly sample the conditional distributions (eq. 2.18)

because they are usually not available. Instead, we use a Metropolis step to update the conditional distributions. Thus, we employ a hybrid-Gibbs sampler: we use a symmetric proposal distribution, so that the accept/reject condition of the trial parameter θ_i^* follows the same algorithm as that described in section 2.2.3, except d is now

$$d = \frac{p(\theta_i^* | \theta_{-i}^{t-1}, y)}{p(\theta_i^{t-1} | \theta_{-i}^{t-1}, y)}$$

(Gelman et al., 2003).

In the problem at hand, the hybrid-Gibbs sampler is more efficient than a standard Metropolis algorithm. The latter method samples all parameters simultaneously, while the former samples parameters individually. Under the Metropolis algorithm, if even one parameter suggestion is highly improbable, then the entire vector of parameters is likely to be rejected. Therefore, a high-dimensional Markov chain may take an extremely long time to walk through parameter space and converge to the posterior distribution. By contrast, the hybrid-Gibbs sampler is much more efficient in our high-dimensional Markov chain (there are 2 model parameters, and 44 tangential velocity parameters for the Milky Way data discussed below). The parameters M_{tot} and a are sampled simultaneously, based on the current v_t parameters, and the v_t parameters are sampled individually based on the current values of all the other parameters.

Using the hybrid-Gibbs sampler for the v_t 's allows us to obtain a probability distribution for each v_t parameter efficiently. We can look at the probability distribution for each v_t and make a prediction of the most probable v_t value for each satellite. Although we find that the resulting v_t distributions are quite diffuse, and a meaningful prediction of v_t cannot be made from them, the

hybrid-Gibbs sampler method is nevertheless efficient and in general does not affect the mass estimate of the Galaxy (as will be shown below). If we assume that the satellite is bound to the galaxy, then by setting eq. 2.3 to zero we obtain an upper limit on the tangential velocity,

$$v_{t,max} = \sqrt{2\Psi(r) - v_r^2}. \quad (2.19)$$

2.2.6 Prior Probabilities

In a Bayesian analysis, the choice of a prior can be thought of as a chance for the researcher to state plainly and explicitly the prior assumptions. When little is known about the problem at hand, it is common to use a *noninformative* prior, so that the information contained in the likelihood is not overwhelmed by information contained in the prior.

In this preliminary analysis, we use uniform priors for all of the model parameters because we assume little about the mass and scale of the system. The uniform prior for each parameter θ is,

$$p(\theta) = \frac{1}{\theta_{max} - \theta_{min}} \quad (2.20)$$

where θ_{min} and θ_{max} are the lower and upper bounds of the uniform distribution. In practice, the parameters are sampled in the natural log-space to ensure that the total mass and scale radius are always positive. The Markov chain values are then exponentiated before examination of the posterior. The bounds $(\theta_{min}, \theta_{max})$ that we use for M_{tot} and a are on the order of $(10^8, 10^{15})M_{\odot}$ and $(10^{-2}, 10^6)$ kpc respectively.

When the tangential velocities are treated as nuisance parameters and sampled in the Markov chain, they too require a prior. The tangential velocity is a 2-dimensional vector on the plane of the sky, so we would like the prior on v_t to be uniform in v_t^2 . Because we are sampling v_t and not the squared tangential velocity, the uniform prior on v_t^2 needs to be transformed to one for v_t .

Here we use Jeffreys' invariance principle (Jeffrey, 1939). Suppose a parameter θ has a prior distribution $p(\theta)$, and that a one-to-one transformation is subsequently performed on θ such that $\phi = h(\theta)$. Then the Jeffrey's prior for ϕ which expresses the same belief as that of $p(\theta)$ is

$$p(\phi) = p(\theta) \left| \frac{d\theta}{d\phi} \right| = p(\theta) |h'(\theta)|^{-1} \quad (2.21)$$

(Gelman et al., 2003).

Following equation 2.21, let $\theta = v_t^2$ and $\phi = v_t$, so that $h(\theta) = \sqrt{v_t^2}$. Then the prior on v_t is given by

$$p(v_t) = \frac{2v_t}{v_{t,max}^2 - v_{t,min}^2}. \quad (2.22)$$

The minimum tangential velocity, $v_{t,min}$, is zero, and the maximum tangential velocity, $v_{t,max}$, is a large constant. Note that if a value of v_t that makes a particle unbounded is suggested in the Markov chain, it will be rejected via the likelihood.

2.2.7 MCMC Chains and Assessing Convergence

The computer code created for this research is written for the **R** Project for Statistical Computing (**R**), an open source software environment for statistical computing and graphics (R Development Core Team, 2012) with many well-developed and efficient statistical diagnostic tools. Recently, **R** has gained popularity in astronomy and the field of astrostatistics (e.g. Feigelson & Babu, 2012), and our code is being developed into an **R** package, called Galactic Mass Estimator (hereafter GME).

GME takes data of the form (r, v_r, v_t) in Galactocentric coordinates, allows the user to select one of five DFs, constructs three Markov chains in parallel, and then combines them into a final, single chain after convergence conditions have been met. The result is a single Markov chain that represents samples from the posterior distribution for the model parameters, given the data. The SNOW package (Tierney et al. (2013)) is used for parallel computing, and the CODA package (Plummer et al., 2006) is used for convergence diagnostics.

Many diagnostics to assess convergence of a Markov chain have been developed, and most of these methods use multiple chains. One advantage of running multiple chains is that the initial values can be dispersed widely in the parameter space, and then convergence can be approximated when they appear to reach a common stationary distribution. This approach allows more exploration of the parameter space and makes it less likely for a local maximum to be mistaken for the mode of the posterior. Furthermore, using multiple chains on the same data set allows estimates of convergence to be obtained in a more reliable and quantitative manner than a single chain. Gelman & Rubin (1992) suggest using the statistic \hat{R} to assess the mutual convergence

of parallel chains:

$$\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\psi|y)}{W}}. \quad (2.23)$$

In equation 2.23, $\widehat{\text{var}}^+(\psi|y)$ is the *marginal posterior variance of the estimand [parameter]*, which is essentially a weighted average of the within-chain variance W , and between-chain variance B (see Gelman et al., 2003, for more details). In practice, \widehat{R} is calculated for each parameter separately; to assume convergence, Gelman et al. (2003) recommend a value of $\widehat{R} < 1.1$ for every parameter, and this is the criterion we use. The \widehat{R} statistic is available in the **R** Software Statistical Computing Language in the CODA package (Plummer et al., 2006).

In this work, the three Markov chains' starting values are widely dispersed in the parameter space, and each chain is run for $i = 10^3$ iterations. After this initial run, \widehat{R} is calculated for each parameter. If any of the parameters have $\widehat{R} > 1.1$, then it is assumed that the chains have not converged, and the last parameter values in each chain are used as the initial parameters in three new chains. The process is repeated until all parameters across all three chains have $\widehat{R} < 1.1$, at which point convergence is assumed. The final sample of the posterior distribution is created by combining the last halves of the three chains, thus providing 1500 parameter vector samples. Prior to this final step, however, we check the effective sample sizes of the Markov chains.

In general, the draws in a Markov chain are not truly independent; some autocorrelation exists in the sequence of samples (Gelman et al., 2003). The effective sample size is the equivalent number of independent samples:

$$n_{\text{eff}} = mn \frac{\widehat{\text{var}}^+(\psi|y)}{B} \quad (2.24)$$

where m is the number of parallel Markov chains and n is the number of draws in each chain (Gelman et al., 2003). If the draws in all chains were perfectly independent, then the number of independent draws would be mn . However, the draws within a chain are general autocorrelated, and so $B > \widehat{\text{var}}^+(\psi|y)$, and $n_{eff} < mn$. An effective sample size of at least 100 is necessary to obtain reliable first-moment statistics such as the mean and median, while an effective sample size over 200 is needed for second-order moments.

In our code, we use the *effectiveSize* function in the CODA package (Plummer et al., 2006), and find that all parameters had effective sample sizes greater than 300 for all models when $i = 10^4$. An acceptance rate between 20 and 30% is required for the v_t parameters, and an acceptance rate of 30-40% is required for the model parameters. The final chain (15000 samples) for each model is visually inspected to ensure that the three chains did not reach very different maxima in the posterior distribution.

2.3 Simulations

Prior to data analysis, we explored the statistical properties of our Bayesian estimates under repeated sampling, using simulation. Unfortunately, a Bayesian analysis cannot be repeated when the data is real (i.e. for a single data set). A Bayesian analysis can be repeated, however, with simulated data sets produced from the same DF, and analysed independently using the same model. By examining the range of parameter estimates, the average behaviour of the model on this type of data can be explored. Simulations and analyses of trivial cases (i.e. when the model and data have the same distribution) are also an effective way to test code.

It is expected that when the simulated data come from the same DF as that of the model, then on average the Bayesian parameter estimates should be correct. Furthermore, quantities like uncertainties and credible intervals should be reliable (e.g. a 50% credible region should contain the true parameter values 50% of the time). In contrast, when the simulated data comes from a different DF than that of the model, biases in the estimates may occur, and the credible regions may become unreliable (e.g. overconfident). Moreover, we need to investigate whether or not treating missing v_t measurements as parameters affects other parameter estimates, regardless of whether or not the data and model share the same DF.

We simulate mock observations of 100 satellites orbiting a galaxy whose gravitational potential follows the Hernquist model. The mock tracer observations include their distances r and velocity components v_r and v_t . We explore the effects of assuming an isotropic Hernquist model in the following three scenarios:

1. isotropic data with complete velocity vectors
2. isotropic data with 50 unknown v_t values
3. constant anisotropic data from the $\beta = 0.5$ Henquist model, with 50 unknown v_t 's

For each scenario, we create 500 data sets with 100 particles each. The Bayesian analysis is performed on each set of data, as described in Sections 2.2.3 - 2.2.7, yielding 500 Markov chains for each scenario. From these chains, statistics such as the mean parameter estimate and credible intervals are calculated.

For all the simulations, we use $M_{\text{tot}} = 10^{12} M_{\odot}$ and $a = 15 \text{ kpc}$ to generate the data. Our choice for the total mass is based on many other studies that

have shown the Milky Way's mass to be close to this order of magnitude. For numerical simplicity, the following units are used in our code: the gravitational constant $G \equiv 1$, r is measured in kiloparsecs (kpc), velocity components are measured in 100km s^{-1} , and mass is measured in $2.325 \times 10^9 M_{\odot}$.

2.4 Kinematic Data for the Milky Way

In principle, any well-defined object orbiting the Galaxy with a measured distance from the Galactic center and at least one velocity measurement may be used to estimate the mass of the Milky Way. In this work, as a first run, we use only globular clusters (GCs) and dwarf galaxies (DGs). It is possible to measure the proper motions and line-of-sight velocities of these tracers in the Galaxy's halo, and to convert these measurements into Galactocentric coordinates. Indeed, many of the kinematic measurements and conversions have already been made (e.g. Dinescu et al., 1999; Casetti-Dinescu et al., 2010, 2013; Boylan-Kolchin et al., 2013). However, the proper motions of many GCs and DGs have yet to be measured, and so the conversion from our frame of reference to a Galactocentric one cannot be performed. Nevertheless, the line-of-sight velocities of these objects are available, and could contain useful information about the Galaxy's mass profile. Thus, we incorporate some of this incomplete data into our analysis.

The data used in this research are in Galactocentric coordinates (see Table 2.1). The first 59 objects are GCs, and the last 29 are DGs. Note that 26 GCs and 18 DGs listed do not have tangential velocities, because they have no proper motion measurements. The Galactocentric radial velocities for these data must be approximated, for which we assume $v_r \approx v_{\text{los}}$. We use

this approximation only for objects with $|\cos \gamma| \geq 0.95$ (where γ is the angle subtended by the line connecting the Sun and the Galactic Centre, from the object), guaranteeing that any further adjustment to v_{los} will be small. We also exclude the following clusters, even though they have $|\cos \gamma| \geq 0.95$, because they are either associated with the Sagittarius dwarf galaxy or their measurements suffer from high extinction: Djorg 1, NGC 6401, NGC 6715, NGC 6544, NGC 6715, Pal 6, Terzan 1, Terzan 6, Terzan 7, and Terzan 8.

Table 2.1: Milky Way Kinematic Data

Object	r (kpc)	v_r (km s ⁻¹)	Δv_r	v_t (km s ⁻¹)	Δv_t	$\cos \gamma$
NGC 104	7	17.0	0.2	171.0	22.0	0.15
NGC 288	11	16.0	0.4	59.0	18.0	0.75
NGC 362	9	55.0	0.5	85.0	31.2	0.61
NGC 1851	16	186.0	0.6	170.0	42.4	0.89
NGC 1904	18	93.0	0.5	83.0	44.7	0.94
NGC 2298	14	-58.0	1.3	100.0	52.7	0.89
Pal 3	85	-247.0	8.4	242.0	121.5	1.00
NGC 4147	19	57.0	1.0	161.0	65.7	0.93
NGC 4590	9	-99.0	0.6	300.0	35.6	0.69
NGC 5024	18	-106.0	4.1	250.0	86.5	0.90
NGC 5139	6	-31.0	0.7	65.0	14.1	0.05
NGC 5272	12	2.0	0.4	164.0	24.5	0.75
NGC 5466	16	254.0	0.3	216.0	66.8	0.88
Pal 5	16	-11.0	16.0	62.0	38.0	0.95
NGC 5897	7	49.0	1.0	138.0	59.4	0.79
NGC 5904	6	-313.0	0.5	234.0	39.6	0.33
NGC 6093	3	60.0	4.1	85.0	28.2	0.67
NGC 6121	6	-58.0	0.4	25.0	22.6	-0.91
NGC 6144	3	109.0	1.1	137.0	33.3	0.47
NGC 6171	4	20.0	0.3	156.0	36.9	-0.29
NGC 6205	8	279.0	0.9	129.0	35.0	0.48
NGC 6218	5	-21.0	0.6	168.0	22.0	-0.46
NGC 6254	5	-53.0	1.1	178.0	28.3	-0.59
NGC 6341	9	70.0	1.7	46.0	26.2	0.61
NGC 6362	5	-40.0	0.6	134.0	20.5	0.25

Object	r (kpc)	v_r (km s ⁻¹)	Δv_r	v_t (km s ⁻¹)	Δv_t	$\cos \gamma$
NGC 6397	6	18.0	0.1	166.0	16.3	-0.83
NGC 6584	6	150.0	15.0	185.0	55.9	0.88
NGC 6626	3	8.0	1.0	172.0	26.4	-0.87
NGC 6656	5	172.0	0.6	214.0	31.9	-0.94
NGC 6712	4	208.0	0.6	132.0	21.5	-0.08
NGC 6752	5	-19.0	1.5	200.0	11.4	-0.50
NGC 6779	9	172.0	0.9	39.0	58.1	0.63
NGC 6809	4	-181.0	0.4	119.0	30.4	-0.49
NGC 6838	7	3.0	0.2	180.0	17.8	-0.05
NGC 6934	12	-305.0	1.6	124.0	93.0	0.86
NGC 7078	10	-74.0	0.6	141.0	34.7	0.70
NGC 7089	10	46.0	0.9	331.0	63.9	0.74
NGC 7099	7	14.0	1.0	120.0	30.8	0.46
NGC 5634	21	-0.8	6.6	–	–	0.95
NGC 6284	8	0.3	1.7	–	–	0.98
NGC 6356	7	0.6	4.3	–	–	0.97
NGC 6426	14	-0.5	23.0	–	–	0.96
NGC 6441	4	-0.0	1.0	–	–	0.95
NGC 6453	4	-0.9	8.3	–	–	0.98
NGC 6540	3	0.0	1.4	–	–	-0.97
NGC 6569	3	-0.2	5.6	–	–	0.95
NGC 6864	15	-1.1	3.6	–	–	0.96
IC 1257	18	-66.5	2.1	–	–	0.99
Arp 2	21	153.0	10.0	–	–	0.99
NGC 7492	25	-97.4	0.6	–	–	0.95
NGC 5824	26	-117.7	1.5	–	–	0.98
Pal 13	27	192.4	0.3	–	–	0.95
NGC 5694	29	-228.1	0.8	–	–	0.98
NGC 6229	30	22.6	7.6	–	–	0.96
Whiting 1	35	-103.5	1.8	–	–	0.98
Pal 2	35	-104.4	57.0	–	–	1.00
Pal 15	38	147.8	1.1	–	–	0.99
NGC 7006	39	-185.2	0.4	–	–	0.98
Pyxis	41	-195.2	1.9	–	–	0.98
Pal 14	72	165.4	0.2	–	–	1.00
NGC 2419	90	-26.4	0.5	–	–	1.00
Eridanus	95	-141.0	2.1	–	–	1.00
Pal 4	111	50.5	2.1	–	–	1.00
AM 1	125	-41.6	20.0	–	–	1.00
Fornax	140	-31.8	1.7	196.0	29.0	1.00

Object	r (kpc)	v_r (km s ⁻¹)	Δv_r	v_t (km s ⁻¹)	Δv_t	$\cos \gamma$
LeoI	261	167.9	2.8	101.0	34.4	1.00
LMC	49	93.2	3.7	346.0	8.5	0.99
SMC	60	6.8	2.4	259.0	17.0	0.99
Sculptor	87	79.0	6.0	198.0	50.0	1.00
Draco	92	-98.5	2.6	210.0	25.0	1.00
BootesI	57	106.6	1.0	–	–	0.99
BootesII	43	-115.6	5.0	–	–	0.98
CanesVenaticiI	219	76.8	1.0	–	–	1.00
CanesVenaticiII	150	-96.1	1.0	–	–	1.00
Carina	102	14.3	1.0	–	–	1.00
ComaBernices	45	82.6	5.0	–	–	0.98
Hercules	141	142.9	1.0	–	–	1.00
LeoII	235	26.5	8.0	–	–	1.00
LeoIV	154	13.9	1.0	–	–	1.00
LeoV	175	62.3	3.0	–	–	1.00
Sagittarius	16	166.3	60.0	–	–	0.93
Segue1	28	113.5	1.0	–	–	0.97
Segue2	41	39.7	1.0	–	–	0.99
Sextans	89	78.2	1.0	–	–	1.00
UrsaMajorI	101	-8.8	1.0	–	–	1.00
UrsaMajorII	36	-36.5	2.0	–	–	0.99
UrsaMinor	77	-89.8	8.0	–	–	1.00
Willman1	42	33.7	2.0	–	–	0.98

NOTE: Columns from left to right: objects' names, Galactocentric distance, radial velocity, uncertainty in radial velocity, tangential velocity, uncertainty in tangential velocity, and $\cos \gamma$. All data are in Galactocentric coordinates (r , v_r , v_t) as described in Section 2.2.4, with the exception of GCs and DGs that lack tangential velocities (see text). Conversions from line-of-sight and proper motion measurements to Galactocentric measurements were completed by the studies mentioned in Section 2.4.

The GC data are taken from Dinescu et al. (1999), Casetti-Dinescu et al. (2010), Casetti-Dinescu et al. (2013), and Harris (1996), while data for six of the DGs are taken from Sohn et al. (2013) (Leo I), Pryor et al. (2014) (Draco),

and Boylan-Kolchin et al. (2013) (Fornax, LMC, SMC, and Sculptor). The rest of the dwarf galaxy data, which do not have tangential velocities, are from the compilation given in Watkins et al. (2010) and references therein. Uncertainties in the Watkins et al. (2010) dwarfs' v_r values are taken from the HyperLeda Catalogue (Paturel et al., 2003), with the exception of those for Coma Berenices, Sagittarius, and Sextans, which are taken from Simon & Geha (2007), Ibata et al. (1997), and Walker et al. (2006) respectively. The r -values in Table 2.1 are based on mean magnitudes of RR Lyrae and horizontal branch stars, and are uncertain to typically 5% (see Harris, 1996). The uncertainties associated with r and v_r , and the differences in the LSR assumed motion used among the different studies are $\lesssim 15$ km/s, and thus unimportant compared to the uncertainties associated with the v_t values.

In the following analysis, we specifically assume (a) a spherical Hernquist-like or Jaffe-like halo potential, (b) equal weights for all data points, (c) no net rotation of the halo, and (d) that all tracers are bound to the Galaxy.

2.5 Results

2.5.1 Simulation Results

Figure 2.1 shows the distribution of the mean parameter estimates from scenario 1. Black dots are the mean of the estimates, and red dashed lines are the true parameter values. On average, the estimates are unbiased within one standard deviation (sd), and the sd of the chains is roughly equal to the sd of the estimates.

Because the Markov chain represents the posterior distribution, we can also

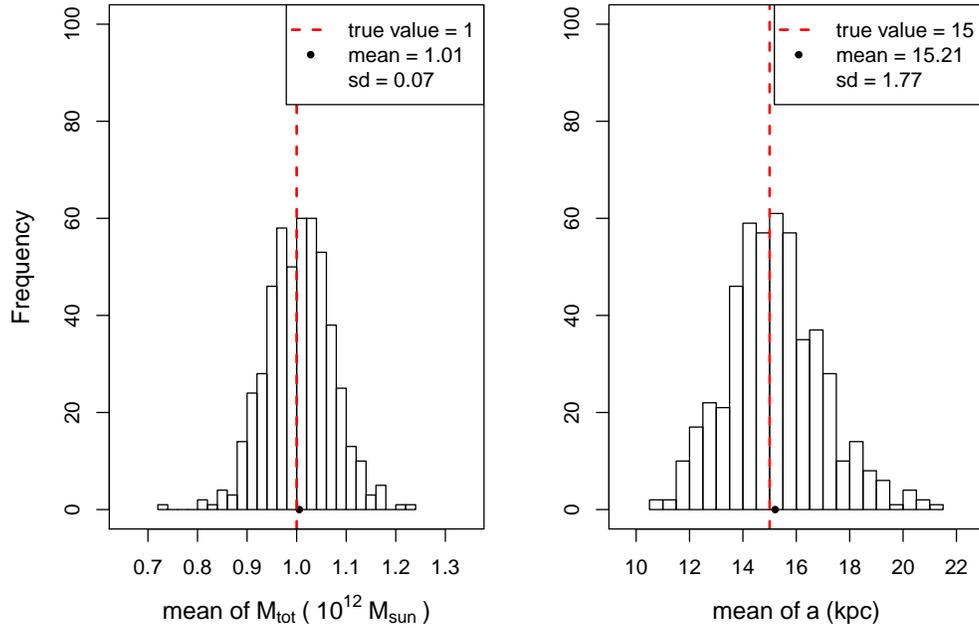


Figure 2.1: Empirical distribution of M_{tot} and a estimates from simulated data analysis. Black points and red dashed lines show the mean of the estimates and the true value of the parameter respectively. The standard deviation of the estimates is 0.07 and 1.77 for M_{tot} and a respectively.

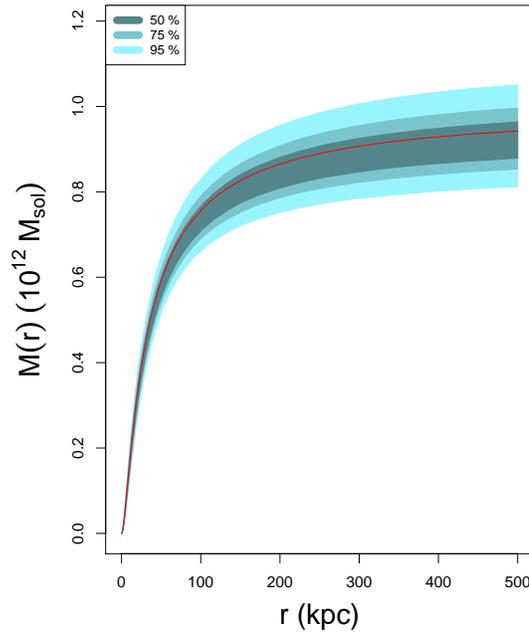


Figure 2.2: Example of predicted and true mass profile from analysis of simulated data. The true $M(r)$ profile is shown in red, and the 50, 75, and 96% credible regions are shown as shades of teal. Note: this is the result from one analysis (i.e. one data set).

calculate credible regions — Bayesian analogues of confidence intervals — for the $M(r)$ profile. An example of the mass profile credible regions for one data set is shown in Figure 2.2, where shades of teal show the 50, 75, and 95% credible intervals as a function of r . Credible regions are found by calculating $M(r)$ at several different r values, for every set of parameters in the Markov chain. The true $M(r)$ profile is the solid red line, calculated from eq. 2.6 and the true M_{tot} and a . We find that the credible regions are reliable when the DF of the assumed model and the DF of the data are the same. For example, the true $M(r)$ curve fell within the 75% credible region seventy-five percent of the time over the course of the 500 analyses for scenario 1.

In scenario 2, we randomly remove 50 v_t 's from each data set, and treat them as parameters in the analysis. We find a very small positive bias in both the M_{tot} and a estimates, as shown in the top two panels in Fig. 2.3. The bias is insignificant, as the means of the estimates ($1.01 \times 10^{12} M_\odot$ and 15.2kpc) are within one standard deviation of the distribution ($0.08 \times 10^{12} M_\odot$ and 1.9kpc, respectively). The slight although insignificant positive bias suggests that the median may be a better estimate of the mass than the mean, but the median is almost identical to the mean in all cases.

In scenario 3 recall that an isotropic model is assumed, but the data sets in scenario 3 have constant anisotropic velocity dispersions $\beta = 0.5$. Despite the data and model having different DFs, the estimates show only a slight positive bias; the true M_{tot} and a are still within one standard deviation of the distribution (see Fig. 2.3). The mean of the estimates for M_{tot} and a are $1.01 \pm 0.08 \times 10^{12} M_\odot$ and $15.3 \pm 1.8\text{kpc}$, respectively.

Examples of mass profile credible regions from scenarios 2 and 3 are presented in Fig. 2.4. Note that the introduction of v_t parameters tends to increase

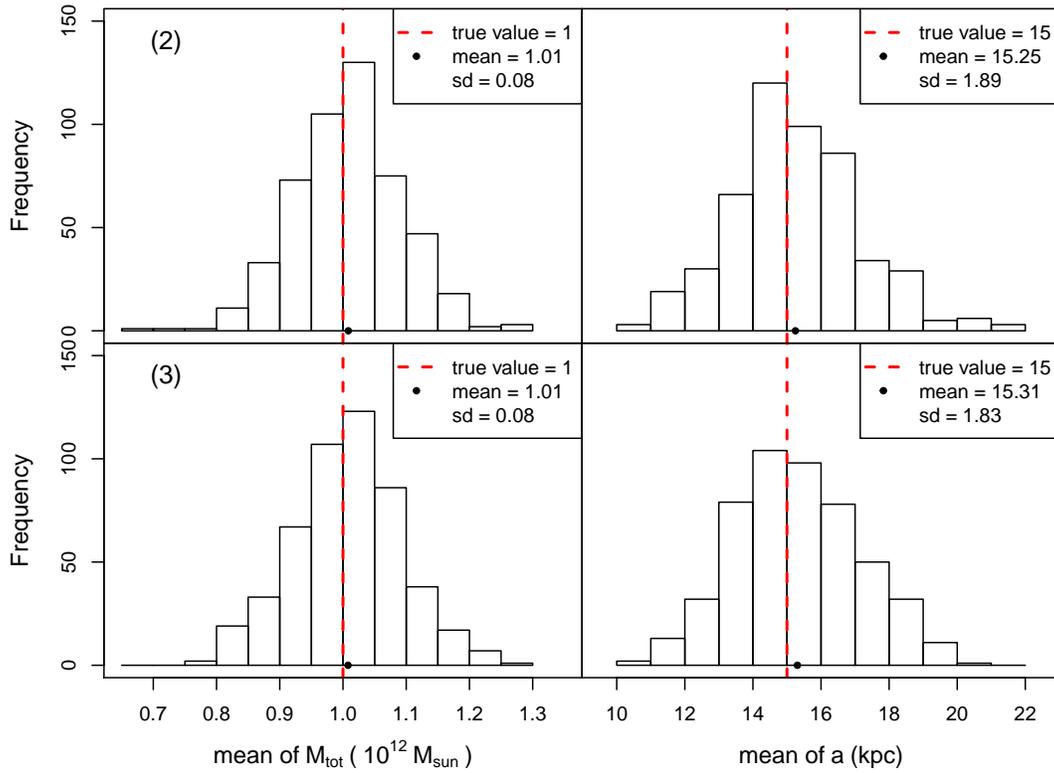


Figure 2.3: Empirical distribution of M_{tot} and a estimates from simulated data analysis, with 50 tangential velocities removed. The top panels are for scenario (2) and the bottom, (3). The black points and red dashed lines show the mean of the estimates and the true value of the parameters respectively. The standard deviations of the estimates in scenario (2) are $0.08 \times 10^{12} M_{\odot}$ and 1.9kpc, while the standard deviations are $0.08 \times 10^{12} M_{\odot}$ and 1.8kpc in (3).

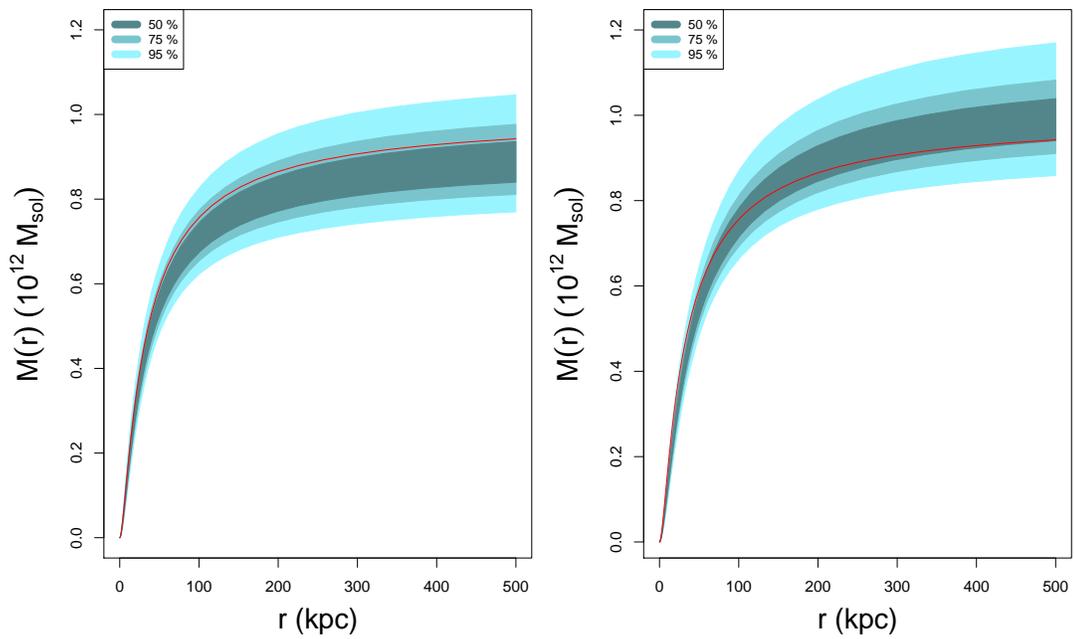


Figure 2.4: Example cumulative mass profile when 50 v_t 's are unknown, and an isotropic Hernquist model is assumed. The true profile is shown as a solid red line, and the credible regions are shown as shades of teal. The left profile is scenario (2) and the right is scenario (3).

the width of the credible regions at all r values compared to Fig. 2.2. In scenario 2, we find the credible regions to be slightly over confident for values of $17 < r < 35$ kpc, with the true $M(r)$ falling within the 50, 75, and 95% regions 48, 73, and 93% of the time over the 500 analyses. At all other r values, however, the credible regions are reliable. In scenario 2, the credible regions are slightly lower than the true cumulative mass profile; the opposite is the case for scenario 3, but for both the true curve still lies in the 75% credible region for most r (see Fig. 2.4). We reiterate, however, that Fig. 2.4 are examples of $M(r)$ profiles from a single data set and analysis. Over 500 analyses we find that the 50, 75, and 95% credible regions do contain the true $M(r)$ 50, 75, and 95% of the time, in both scenarios 2 and 3, for almost all r .

2.5.2 Milky Way Results

Assuming an isotropic Hernquist model, and using all the kinematic data from Table 2.1, we find a mean M_{tot} of $1.55 \pm 0.08 \times 10^{12} M_{\odot}$ and a scale radius of 16.9 ± 2.3 kpc, where the uncertainties are the standard deviations of the parameters in the Markov chain. The 95% credible regions for M_{tot} and a are $(1.42, 1.73) \times 10^{12} M_{\odot}$ and $(12.8, 21.7)$ kpc respectively. We also report the mean M_{tot} and scale radius, with uncertainties of one standard deviation, when the other models are assumed (Table 2.2). The mass estimates and scale radii vary only slightly between Hernquist models. The Jaffe model's mass is similar, but the Jaffe scale radius cannot be compared directly to that of Hernquist because they have physically different definitions.

The mass profile credible regions are shown in Fig. 2.5. The innermost dark region corresponds to the 50% credible region. The vertical dashed lines show

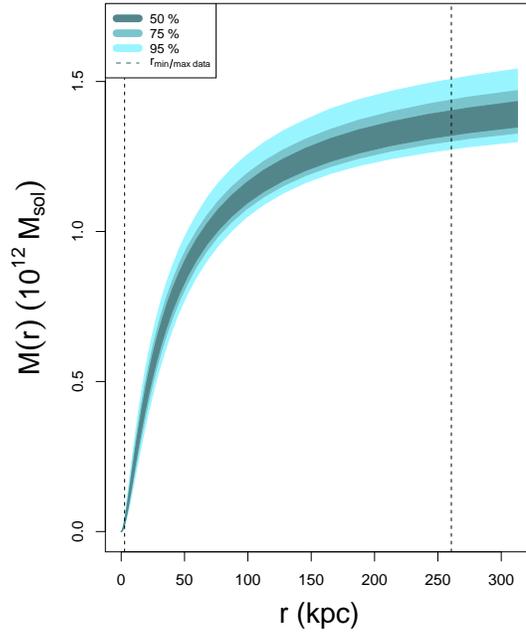


Figure 2.5: Mass profile credible regions assuming a Hernquist model with isotropic velocity dispersion. The dashed lines indicate the location of NGC 6540 and Leo I (the closest and furthest objects from the Galactic center respectively in our dataset).

the extent of the data, with NGC 6540 and Leo I being the closest and furthest objects from the Galactic center respectively. The mass contained within the distance of Leo I is $1.37^{+0.14}_{-0.10} \times 10^{12} M_{\odot}$, where the uncertainties correspond to the 95% credible interval.

Table 2.2: Parameter Estimates for the Milky Way.

	All Data		Without Pal 3		Without Draco	
Model - σ^2	M_{tot} ($10^{12}M_{\odot}$)	Scale Radius (kpc)	M_{tot} ($10^{12}M_{\odot}$)	Scale Radius (kpc)	M_{tot} ($10^{12}M_{\odot}$)	Scale Radius (kpc)
H - iso	1.55 ± 0.08	16.9 ± 2.3	1.36 ± 0.07	16.7 ± 2.2	1.55 ± 0.08	16.8 ± 2.3
H - OM	1.52 ± 0.08	16.7 ± 2.2	1.34 ± 0.06	16.4 ± 2.1	1.52 ± 0.08	16.6 ± 2.2
H - $\beta = 0.5$	1.47 ± 0.07	12.1 ± 1.9	1.31 ± 0.06	12.3 ± 1.9	1.46 ± 0.07	11.9 ± 1.8
Jaffe - iso	1.61 ± 0.09	47.7 ± 8.5	1.38 ± 0.06	48.9 ± 8.7	1.57 ± 0.08	45.2 ± 8.2

NOTE: In the first column, the first three models are of the Hernquist type, with isotropic, OM-type anisotropy, and constant anisotropy. The last row shows the results of assuming an isotropic Jaffe model. Uncertainties are one standard deviation of the posterior distribution.

Some satellites may have a large effect on the mass estimate of the Galaxy. Leo I, for example, remained a contentious object for many years, because it is at a large distance from the Galactic center and it was unclear whether or not it is bound to the MW. Recently, however, Boylan-Kolchin et al. (2013) showed that Leo I is likely bound to the MW. Furthermore, when Leo I's proper motion is taken into account, the object has little effect on the mass estimate of the Galaxy (Wilkinson & Evans, 1999). We run our analysis assuming the isotropic Hernquist model both with and without Leo I, and we also find that it has no effect within error on M_{tot} . When Leo I is removed from the analysis, $M_{tot} = 1.52 \pm 0.07 \times 10^{12} M_{\odot}$ and $a = 16.2 \pm 2.2$ kpc, very similar to the values obtained when Leo I is present.

The five other DGs in our data set that have measured proper motions are Draco, Fornax, Sculptor, and the Large and Small Magellanic Clouds (hereafter LMC and SMC). We obtain parameter estimates assuming an isotropic Hernquist model with each of these dwarfs removed, and find that the mass and scale radii do not change within error in any case.

Another object to consider is Sagittarius. In Section 2.4, we argued that $v_{los} \approx v_r$ for tracers with $|\cos \gamma| \geq 0.95$, but Sagittarius is relatively close-by at 16 kpc and has $|\cos \gamma| = 0.93$, so one may question the inclusion of this object. However, once again we find little change in the mass estimate without it, for all models. For example, the isotropic Hernquist model returned $M_{tot} = 1.55 \pm 0.08 \times 10^{12} M_{\odot}$, which is almost identical to the result obtained using all the data (see Table 2.2). The scale radius is also unchanged within error (17.0 ± 2.2 kpc).

We also investigate the effects on the mass estimate when tangential velocities are treated as parameters. To do this, we first obtain a mass estimate

using only the Dinescu data (i.e. using only objects with complete velocity vectors), and find a slightly lower mass of $1.47 \pm 0.08 \times 10^{12} M_{\odot}$. Next, we remove five tangential velocities from the data, and repeat the analysis treating those missing v_t 's as parameters. Repeating this process and removing five different v_t 's each time, we find that the v_t 's cannot be well estimated. However, treating v_t 's as parameters has little to no effect on the mass estimate, within error. There is one exception to the latter statement: when Pal 3's tangential velocity was removed, the mass estimate was reduced significantly.

To investigate the influence of Pal 3 further, we performed an analysis using only the Dinescu data, but without Pal 3's v_t value. Treating Pal 3's v_t as a parameter, the mass estimate of the Milky Way fell by more than 50% ($M_{tot} = 0.76 \pm 0.06 \times 10^{12} M_{\odot}$). We also ran the analysis using only the Dinescu data, but without *any* v_t 's. In this case, $M_{tot} = 0.8 \pm 0.1 \times 10^{12} M_{\odot}$, which is similar to the estimate obtained in the former analysis. We note, however, that Pal 3 has the most uncertain v_t measurement in the list (Table 2.1). It is evident that including measurement uncertainties in the analysis would reduce its leverage considerably.

Using all kinematic data, but removing Pal 3 from the analysis, also resulted in reduced mass estimates. Furthermore, the effect is observed regardless of the selected model (Table 2.2). Thus, Pal 3's proper motion, and indeed Pal 3 in general, has significant influence on the mass estimate of the Galaxy. This issue regarding Pal 3 confirms the finding of Sakamoto et al. (2003), who noted that high-velocity objects such as Pal 3 and Draco can have a significant effect on the mass estimate of the Galaxy. As mentioned previously, we also test the effect of Draco on the mass estimate and find that it has little effect on the mass estimate (Table 2.2).

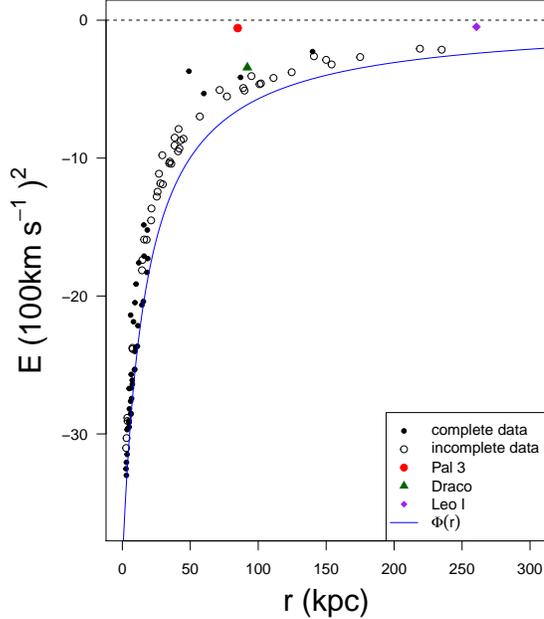


Figure 2.6: Satellite energies given the model parameters from the isotropic Hernquist model fit. Satellites without tangential velocities are shown as open circles, and are plotted using the estimated v_t value from the model fit. Unbound (escaping) objects would lie above the dotted line.

To demonstrate the effectiveness of using v_t 's as parameters, consider Figure 2.6. Using eq. 2.3 and the mean parameter values from the isotropic Hernquist model fit, we plot the negative of the tracers' specific energies as a function of r . Filled points are satellite data with complete velocity vectors, and hollow points are data with unknown v_t 's (plotted using the mean v_t estimates from the Markov chain). As demonstrated by our simulations and tests with the Dinescu data, the v_t values cannot be well estimated. However, the v_t parameters appear to converge to an average v_t for a given r value, and this is reflected in the positions of the hollow points in Fig. 2.6.

The results demonstrate that in a small sample of data, some objects carry greater influence on the mass estimate than others. Furthermore, the variation in these results implies that it would be fruitful to weight the data by their

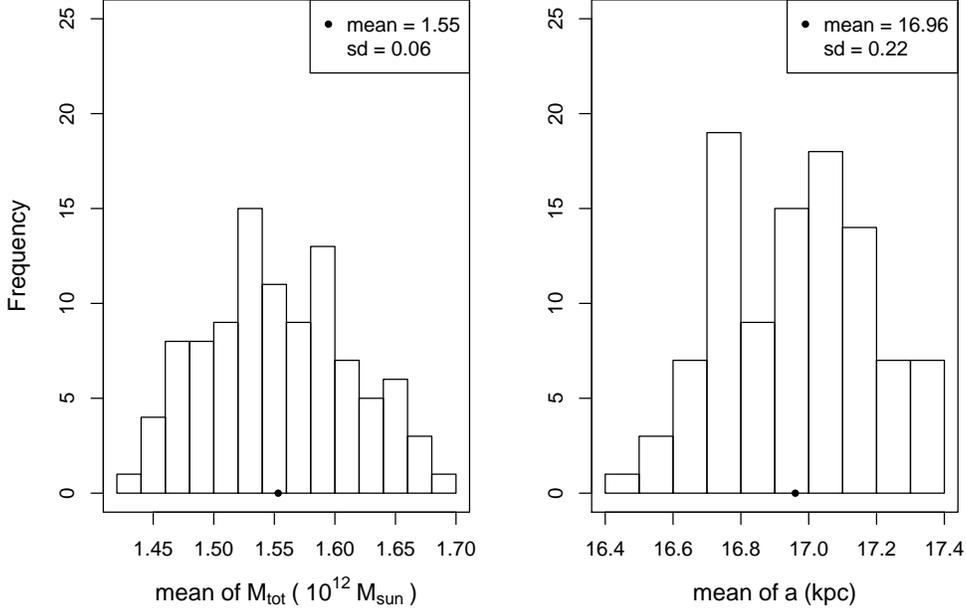


Figure 2.7: Distribution of M_{tot} and a estimates from the third sensitivity analysis. The black dots show the mean of the estimates. The value of the mean and the standard deviation of the empirical distribution are shown in the legend.

measurement uncertainties. In a Bayesian analysis, however, a fully hierarchical approach is necessary to properly include the measurement uncertainties of the data, and a probability distribution for the errors must be assumed. We leave this analysis to a future paper, and instead perform three more approximate sensitivity analyses.

The first two sensitivity analyses are extreme cases: (1) all the tangential velocities are increased by $2\Delta v_t$, and (2) all the tangential velocities are decreased by $2\Delta v_t$. In the third and more realistic sensitivity analysis, we randomly change each v_t into a new tangential velocity $v_{t,new}$ via,

$$v_{t,new} = v_t + N(0, \Delta v_t) \quad (2.25)$$

where $N(0, \Delta v_t)$ represents a random number drawn from a normal distribution with mean zero and variance Δv_t . Using eq. 2.25, we generate 100 data sets with new v_t values, and then analyze these data sets assuming the isotropic Hernquist model. The estimates of M_{tot} and a from the 100 analyses have a distribution that confirms the results of the original analysis (Fig. 2.7); the mean of the estimates is nearly identical to the result in Table 2.2.

The results of the sensitivity analyses show that a proper treatment of the measurement uncertainties is worth pursuing. In future analyses, we plan to incorporate the measurement uncertainties of the data via a hierarchical model.

2.6 Discussion and Future Prospects

The results of this study are promising. Not only does the Bayesian analysis provide an effective way of incorporating complete and incomplete data, but it also enables easy calculation of probabilistic credible regions for the cumulative mass profile. Furthermore, even though this is a preliminary analysis, and mostly meant to lay the groundwork for future studies, our total mass estimates are similar to other studies that use different methods.

Because our method returns a sample of parameter values representing the posterior distribution, it is easy to compare our results with mass estimates, at any radii, obtained in other studies. We can compute $M(r)$ credible regions from our Markov chain at any r value, and obtain a mass estimate at that radius, with uncertainties.

A collection of total mass estimates within $r = 100\text{kpc}$, from 10 different studies, has been compiled by Courteau et al. (2014). Our mass estimate at this

radius assuming the isotropic Hernquist model is $M_{100} = 1.14 \times 10^{12} M_{\odot}$ with a 95% credible region of $(1.05, 1.26) \times 10^{12} M_{\odot}$, which is within the range of values listed in the review.

Watkins et al. (2010) find the mass within 300 kpc to be $0.9 \pm 0.3 \times 10^{12} M_{\odot}$ for an isotropic model. Our estimate M_{300} assuming an isotropic model is $1.39 \times 10^{12} M_{\odot}$ with a 95% credible region of $(1.29, 1.53) \times 10^{12} M_{\odot}$. When they consider an anisotropic velocity distribution with β derived from the observational data, however, they find $3.4 \pm 0.9 \times 10^{12} M_{\odot}$, in contrast to our $\beta = 0.5$ constant anisotropic model that gives $M_{300} = 1.35 \times 10^{12} M_{\odot}$, with a 95% credible interval of $(1.27, 1.51) \times 10^{12} M_{\odot}$.

Deason et al. (2013) used BHB stars to trace the MW's mass, and found $M(r = 50\text{kpc})$ to be approximately $4 \times 10^{11} M_{\odot}$, assuming a model of constant anisotropy with $\beta = 0.5$. However, our mass estimate for the MW at 50kpc, using the Hernquist constant anisotropic model, is substantially higher at $9.5 \times 10^{11} M_{\odot}$ with a 95% credible interval of $(8.5, 11.0) \times 10^{11} M_{\odot}$. Even removing Pal 3 from the data set does not lower this estimate significantly, reducing it to $8.5 \times 10^{11} M_{\odot}$ with a 95% credible interval of $(7.5, 9.6) \times 10^{11} M_{\odot}$.

Using a truncated, flat rotation curve model, Battaglia et al. (2005) found the mass of the MW dark matter halo to be $1.2_{-0.5}^{+1.8} \times 10^{12} M_{\odot}$, and with an NFW model found a virial radius of $0.8_{-0.2}^{+1.2} \times 10^{12} M_{\odot}$. Boylan-Kolchin et al. (2013) estimate the MW virial mass at $1.6 \times 10^{12} M_{\odot}$, with a 90% confidence interval of 1.0 to $2.4 \times 10^{12} M_{\odot}$. Sohn et al. (2013) use the timing argument of Leo I to arrive at a virial mass estimate $M_{vir} = 3.15_{-1.36}^{+1.58} \times 10^{12} M_{\odot}$. Li & White (2008) found the virial mass to be $2.4 \times 10^{12} M_{\odot}$, with a lower 95% confidence level of $0.8 \times 10^{12} M_{\odot}$. Thus, our preliminary results are on par with many other studies that use different methodologies.

To our knowledge, no other studies besides Sakamoto et al. (2003) have found Pal 3 to carry so much weight in the analysis. Pal 3's proper motion is already known (though with large uncertainties) and the satellite does not lie as far from the Galactic center as Leo I and other satellite dwarfs, which may have allowed its effect to go unnoticed. Removing Pal 3's true v_t from the analysis and treating it as a parameter lowered the mass significantly, suggesting that the tangential velocity is in the tail of the v_t distribution at $r = 84\text{kpc}$. The DGs in our data set, on the other hand, seem to have little individual influence on the mass estimate of the Galaxy even though some have high velocities and large distances from the galactic center.

Many improvements, challenges, and exploratory analyses remain:

1. One way to substantially improve the analysis is to incorporate measurement uncertainties via a hierarchical model, rather than the preliminary sensitivity analysis performed here. In the Bayesian paradigm, a probability density function of the measurement errors must be assumed. For example, for each data point the known measurement uncertainty may be used to define the variance of a Gaussian distribution centered on the measurement value. Objects with high influence when measurement errors are ignored (e.g. Pal 3) might have reduced influence when measurement errors are included.
2. An immediate challenge is finding models with distribution functions that are tractable and that describe the Milky Way in a more sophisticated manner. Using the NFW model is of particular interest because it is known to fit the dark matter halos of galaxies on many different scales, as well as groups and clusters of galaxies. The NFW DF, however, is not

analytic. Although numerical solutions for the NFW DF have been derived, applying them in the Bayesian framework is more difficult because of the model's infinite mass. When a model's DF does not integrate to a finite mass (eq. 2.2) then it is not a proper probability distribution— a requirement when applying Bayes' theorem. We are currently exploring the problem through an Approximate Bayesian Computation (ABC) algorithm, which allows for calculations of posterior distributions without explicit calculation of the likelihood.

3. The DFs for all of the models employed here are analytic. Furthermore, the models are self-consistent— i.e. we implicitly assume that the dark matter and the satellites follow the same distribution. However, it is possible that the tracers (GCs and DGs) have a different distribution than the dark matter halo particles. For example, the tracers may have a Hernquist-type density profile $\rho_H(r)$ given by eq. 2.5, but may reside in an NFW gravitational potential given by

$$\Phi_{\text{NFW}}(r) = -4\pi\rho_o r_s^2 \frac{\ln(1+r/r_s)}{r/r_s}. \quad (2.26)$$

In this situation, there are two extra parameters, r_s and ρ_o , which correspond to the scale length and density parameter of the dark matter halo. We can derive the DF for such a model via the Eddington formula:

$$f(\mathcal{E}) = \frac{1}{\sqrt{8\pi^2}} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E} - \Psi}} \left(\frac{d^2\rho}{d\Psi^2} \right) d\Psi + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho}{d\Psi} \right)_{\Psi=0} \quad (2.27)$$

where $\Psi = \Phi - \Phi_o$ is the relative potential (see Binney & Tremaine, 2008). For the Hernquist model, ρ can be written as an analytic function

of Ψ , and the integral can be evaluated in closed form. For the case at hand, however, the relation between ρ_H and Ψ_{NFW} is a transcendental equation. Nevertheless, the integral required in the Eddington formula can be evaluated numerically. Widrow et al. (2008), for example, used this method to derive DFs for their self-consistent disk-bulge-halo galaxy models. A numerically derived DF may still be used with our method, as long as it is a normalized probability distribution. We plan to implement models of this type in future studies of tracer populations.

4. Different velocity anisotropy formalisms are also of interest. For example, other Hernquist model DFs of different anisotropies are discussed by Cuddeford (1991) and Baes & Dejonghe (2002). The former derived a velocity anisotropy that is a generalization of the OM-type anisotropy, where another parameter β_0 is introduced such that

$$\beta(r) = \frac{r^2 + \beta_0 r_a^2}{r^2 + r_a^2}. \quad (2.28)$$

When $\beta_0 = 0$, eq. 2.28 reduces to OM-type anisotropy. As $r_a \rightarrow \infty$, $\beta(r) \rightarrow \beta_0$, in contrast to eq. 2.7. Baes & Dejonghe (2002) derive a Hernquist DF using this formalism, and show that only four values of β_0 lead to DFs that can be expressed in terms of elementary functions. The simplest of these DFs occurs when $\beta_0 = 0.5$, while the other DFs are "...somewhat more elaborate" and not provided (Baes & Dejonghe, 2002).

5. We will explore biases that may occur due to selection effects. In the Hernquist simulations used here, some tracers are unrealistically far from

the Galactic center (e.g. more than 500kpc away), while our kinematic data set has a range from $r = 3\text{kpc}$ to 261 kpc. Imposing a more realistic range on simulated data may or may not introduce biases in parameter estimates.

6. Further along the line, it will be exciting to apply the method presented here to large datasets of field halo stars, leading up to the GAIA data. Sakamoto et al. (2003) showed that including many field horizontal branch stars greatly reduced the effect of high-velocity objects (such as Draco and Pal 3) on the mass estimate of the Milky Way. Therefore, it can be expected that the accurate and abundant kinematic data from GAIA will also improve our mass estimates in a major way and decrease the effect of outliers.
7. The method outlined in this paper could also be extended to obtain mass estimates of other galaxies for which tracer objects will have only the projected positions and line of sight velocities.

Summary

We have introduced a method to estimate the mass of the Milky Way that incorporates both complete and incomplete data for positions and velocities of tracers. The method treats unknown tangential velocities as parameters in the model. Simulations showed that although the tangential velocities cannot be well constrained, treating v_t 's as parameters has little effect on the mass estimate when other complete velocity vectors are available. An exception does occur, however, when a tracer has an unusually extreme tangential velocity

(e.g. Pal 3).

Under simple assumptions of a Hernquist-like halo potential and modest anisotropy, we find $M_{tot} \approx 1.3 - 1.5 \times 10^{12} M_{\odot}$, in good agreement with other recent work. For an isotropic model we find $M_{tot} = 1.55 \times 10^{12} M_{\odot}$ with a 95% credible interval of $(1.42, 1.73) \times 10^{12} M_{\odot}$, and a scale radius of $a = 16.9 \text{kpc}$. We also report the mass contained within 260kpc: $1.37 \times 10^{12} M_{\odot}$, with a 95% credible interval of $(1.27, 1.51) \times 10^{12} M_{\odot}$.

In future research, we will be incorporating measurement uncertainties into the analysis and will test more extensively for parameter biases. We also plan to use the NFW model and find other DF's to use in the GME code. The method outlined here will eventually be applied to extragalactic studies, where the complete velocity vectors of the tracers are unknown.

Acknowledgements

WEH and LMW acknowledge the financial support of NSERC. GME thanks the reviewer for very useful comments and suggestions, and would also like to thank Aaron Springford for useful discussions about sensitivity analyses.

Bibliography

- Baes, M., & Dejonghe, H. 2002, *Astronomy and Astrophysics*, 393, 485
- Bahcall, J. N., & Tremaine, S. 1981, *The Astrophysical Journal*, 244, 805
- Battaglia, G., Helmi, A., Morrison, H., et al. 2005, *MNRAS*, 364, 433
- Bayes, M., & Price, M. 1763, *Philosophical Transactions (1683-1775)*, 370
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics*, 2nd edn. (Princeton)
- Boylan-Kolchin, M., Bullock, J. S., Sohn, S. T., Besla, G., & van der Marel, R. P. 2013, *ApJ*, 768, 140
- Casetti-Dinescu, D. I., Girard, T. M., Jílková, L., et al. 2013, *AJ*, 146, 33
- Casetti-Dinescu, D. I., Girard, T. M., Korchagin, V. I., van Altena, W. F., & López, C. E. 2010, *AJ*, 140, 1282
- Ciotti, L. 1996, *The Astrophysical Journal*, 471, 68
- Courteau, S., Cappellari, M., de Jong, R., et al. 2014, *Rev. Mod. Phys.*, 86, 47
- Cuddeford, P. 1991, *MNRAS*, 253, 414
- de Vaucouleurs, G. 1948, *Annales d'Astrophysique*, 11, 247
- Deason, A. J., Belokurov, V., Evans, N. W., & Johnston, K. V. 2013, *ApJ*, 763, 113

- Dinescu, D. I., Girard, T. M., & van Altena, W. F. 1999, *AJ*, 117, 1792
- Eadie, G. M. 2013, Master's thesis, Queen's University, Kingston, Ontario, Canada
- Evans, N. W., & An, J. H. 2006, *Physical Review D*, 73
- Evans, N. W., & Williams, A. A. 2014, *MNRAS*, 443, 791
- Feigelson, E., & Babu, G. 2012, *Modern Statistical Methods for Astronomy: With R Applications* (Cambridge University Press)
- Gelfand, A. E., & Smith, A. F. 1990, *Journal of the American statistical association*, 85, 398
- Gelman, A., Carlin, J., Stern, H., & Rubin, D. 2003, *Bayesian data analysis* (Chapman & Hall/CRC)
- Gelman, A., & Rubin, D. 1992, *Statistical science*, 7, 457
- Geman, S., & Geman, D. 1984, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 6, 721
- Harris, W. E. 1996, *AJ*, 112, 1487
- Hartwick, F. D. A., & Sargent, W. L. W. 1978, *The Astrophysical Journal*, 221, 512
- Hernquist, L. 1990, *The Astrophysical Journal*, 356, 359
- Ibata, R. A., Wyse, R. F. G., Gilmore, G., Irwin, M. J., & Suntzeff, N. B. 1997, *AJ*, 113, 634
- Jaffe, W. 1983, *MNRAS*, 202, 995

- Jeffrey, H. 1939, *Theory of Probability* (Oxford University Press)
- Kochanek, C. S. 1996, *The Astrophysical Journal*, 457, 228
- Kulesa, A. S., & Lynden-Bell, D. 1992, *MNRAS*, 255, 105
- Li, Y.-S., & White, S. D. M. 2008, *MNRAS*, 384, 1459
- Little, B., & Tremaine, S. 1987, *The Astrophysical Journal*, 320, 493
- Merritt, D. 1985, *Astronomical Journal*, 90, 1027
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *The journal of chemical physics*, 21, 1087
- Metropolis, N., & Ulam, S. 1949, *Journal of the American Statistical Association*, 44, 335
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *The Astrophysical Journal*, 462, 563
- Osipkov, L. P. 1979, *Soviet Astron. Letters*, 5, 42
- Paturel, G., Petit, C., Prugniel, P., et al. 2003, *A&A*, 412, 45
- Plummer, M., Best, N., Cowles, K., & Vines, K. 2006, *R News*, 6, 7
- Pryor, C., Piatek, S., & Olszewski, E. W. 2014, *ArXiv e-prints*, arXiv:1407.3509
- R Development Core Team. 2012, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0

- Sakamoto, T., Chiba, M., & Beers, T. C. 2003, *A&A*, 397, 899
- Sersic, J. L. 1968, Cordoba, Argentina: Observatorio Astronomico, 1968, 1
- Simon, J. D., & Geha, M. 2007, *ApJ*, 670, 313
- Sohn, S. T., Besla, G., van der Marel, R. P., et al. 2013, *ApJ*, 768, 139
- Tanner, M. A., & Wong, W. H. 1987, *Journal of the American statistical Association*, 82, 528
- Tierney, L., Rossini, A. J., Li, N., & Sevcikova, H. 2013, snow: Simple Network of Workstations, r package version 0.3-13
- Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2006, *ApJL*, 642, L41
- Watkins, L., Evans, N., & An, J. 2010, *MNRAS*, 406, 264
- Widrow, L. M. 2000, *ApJ Supp*, 131, 39
- Widrow, L. M., Pym, B., & Dubinski, J. 2008, *ApJ*, 679, 1239
- Wilkinson, M., & Evans, N. 1999, *MNRAS*, 310
- Zaritsky, D., Olszewski, E. W., Schommer, R. A., Peterson, R. C., & Aaronson, M. 1989, *ApJ*, 345, 759

3

Bayesian Mass Estimates of the Milky Way: The Dark and Light Sides of Parameter Assumptions

Reprinted from *Eadie, G. and Harris, W.*, THE ASTROPHYSICAL JOURNAL, Vol. 829, No. 108. DOI: 10.3847/0004 – 637X/829/2/108. Published by the American Astronomical Society ©AAS

Abstract

We present mass and mass profile estimates for the Milky Way (MW) Galaxy using the Bayesian analysis developed by Eadie et al. (2015b) and using globular clusters (GCs) as tracers of the Galactic potential. The dark matter and GCs are assumed to follow different spatial distributions; we assume power-law model profiles and use the model distribution functions described in Evans et al. (1997) and Deason et al. (2011, 2012a). We explore the relationships between assumptions about model parameters and how these assumptions affect mass profile estimates. We also explore how using subsamples of the GC population beyond certain radii affect mass estimates. After exploring the posterior distributions of different parameter assumption scenarios, we conclude that a conservative estimate of the Galaxy’s mass within 125kpc is $5.22 \times 10^{11} M_{\odot}$, with a 50% probability region of $(4.79, 5.63) \times 10^{11} M_{\odot}$. Extrapolating out to the virial radius, we obtain a virial mass for the MW of $6.82 \times 10^{11} M_{\odot}$ with 50% credible region of $(6.06, 7.53) \times 10^{11} M_{\odot}$ ($r_{vir} = 185_{-7}^{+7}$ kpc). If we consider only the GCs beyond 10 kpc, then the virial mass is $9.02 (5.69, 10.86) \times 10^{11} M_{\odot}$ ($r_{vir} = 198_{-24}^{+19}$ kpc). We also arrive at an estimate of the velocity anisotropy parameter β of the GC population, which is $\beta = 0.28$ with a 50% credible region (0.21, 0.35). Interestingly, the mass estimates are sensitive to both the dark matter halo potential and visible matter tracer parameters, but are not very sensitive to the anisotropy parameter.

3.1 Introduction

The Milky Way's (MW's) composition, structure, dynamical properties, and formation history are heavily influenced by two important properties: its total mass and mass profile. However, inferring the mass profile of the Milky Way (MW) is a task fraught with uncertainty. Direct observations of dark matter, the most important component of the Galaxy's mass, still elude us. Therefore, astronomers must rely on kinematic information of tracer objects such as globular clusters (GCs), dwarf galaxies (DGs), stellar streams, and halo stars whose orbits are influenced by the Galaxy's gravitational potential.

Attempts to use these objects are made more difficult because many of the three-dimensional velocity measurements are incomplete (i.e., proper motions are unknown). Although some tracers' velocity measurements are complete, popular mass estimators do not make use of both incomplete and complete data at the same time. For example, projected mass estimators rely on line-of-sight velocities (e.g., the mass estimators introduced by Bahcall & Tremaine, 1981; Evans et al., 2003; Watkins et al., 2010). On the other hand, mass estimators that use proper motions only use a subset of the data, because they require complete velocity vectors. So overall, we seem to have a dilemma; either we (1) throw away proper motion measurements, or (2) throw away some of the line-of-sight velocity measurements.

Watkins et al. (2010) (hereafter W10) developed two different mass estimators: one that uses line-of-sight velocities only, and one that uses full three-dimensional space motions. In an attempt to use all the data available, W10 employed each mass estimator separately, and combined the two estimates by a weighted average. This approach, however, requires a decision about how to

weight each estimator, and also relies on Monte Carlo simulations to determine uncertainties. Furthermore, W10 find the mass estimators are quite sensitive to the velocity anisotropy.

Cosmological simulations have also shown that the spatial distributions of the dark matter and the tracers are probably quite different, and thus the density profiles of the tracers and the gravitational potential of the dark matter are not self-consistent. This inconsistency makes it difficult to model the phase-space distribution function (DF) of the tracer particles.

Because of the aforementioned issues, the MW's mass and mass profile estimates remain very uncertain, with values routinely varying between 10^{11} and $10^{12}M_{\odot}$ (see Wang et al., 2015, for a graphic of mass estimates from studies using different methods).

For quite some time now, in an effort to obtain better Galactic mass estimates, both maximum likelihood and Bayesian methods have been adopted, with the pioneering paper being Little & Tremaine (1987) (see also Kulessa & Lynden-Bell, 1992; Kochanek, 1996; Wilkinson & Evans, 1999; McMillan, 2011; Kafle et al., 2012; Eadie et al., 2015b; Williams & Evans, 2015; Küpper et al., 2015).

Eadie et al. (2015b, hereafter EHW) introduced a Bayesian analysis which estimates the cumulative mass profile of the Galaxy with both complete and incomplete kinematic data used simultaneously. The method uses the phase-space DF of the tracers, $f(\mathcal{E}, L)$ as determined by the physical model (Binney & Tremaine, 2008). Thorough testing of the method was performed with simulated data, and a preliminary analysis was done using real data (GCs and local DGs). The simulations showed that the Bayesian method is a powerful way to include complete and incomplete kinematic data simultaneously, and

the preliminary analysis gave a total mass for the Galaxy that was in agreement with many other studies (although the range of values in the literature is wide). Furthermore, we found that estimating the mass in this way was relatively insensitive to the velocity anisotropy assumption.

In the discussion section of EHW, we listed ways to improve the analysis in future work. One major step-forward is to implement a model in which the spatial distribution of the dark matter halo is different from the spatial distribution of the satellites. Deriving a DF for such a model, in terms of the energies and momenta of tracers, can be quite difficult; the Eddington-equation method described in Binney & Tremaine (2008) (and also used below) requires the density profile of the tracers ρ to be written as a function of the gravitational potential Φ , which may not be possible if ρ and Φ do not obey Poissons equation. Still, there are realistic cases in which ρ can be written as a function of Φ , even when they are not self-consistent, and for which a non-negative DF can be found.

3.2 The Power Law Model

We assume a galaxy model first proposed by Evans et al. (1997), and also used by Deason et al. (2011); Watkins et al. (2010), in which the gravitational potential and the density profile of the satellites follow different power-law profiles. For such power-law profiles, the DF is analytic (see below), and thus provides a major advantage for assessing the effects of the important parameters of the model.

The first step in setting up the model is to derive the DF in terms of the parameters. Evans et al. (1997) derived a family of DFs for their generalized

power-law model in terms of the specific energy and angular momentum of galactic satellites. This DF was later adopted by Deason et al. (2011, 2012a), with a slightly different parametrization of the gravitational potential, and used in a maximum likelihood analysis to obtain an estimate for the mass profile of the MW, given the kinematic information of blue horizontal branch (BHB) stars.

Various pieces of the DF derivation are given in Evans et al. (1997) and Deason et al. (2011, 2012a), but their notations differ substantially and the exact form of the normalization constant is unclear. In our experience, this can lead to confusion. Therefore, for completeness and clarity, we provide a short derivation of the DF using the parameterization and notation introduced by Deason et al. (2011).

The number density profile for the satellites (tracers) is given by

$$\rho_t \propto \frac{1}{r^\alpha} \quad (3.1)$$

where α may be a free parameter (we drop the constant for Equation 3.1 as this factor is only related to the number of satellites). The gravitational potential of the dark matter halo (assumed to be spherical) is

$$\Phi = \frac{\Phi_o}{r^\gamma} \quad (3.2)$$

where both γ and Φ_o may be free parameters. The values of γ and Φ_o will determine the cumulative mass profile of the dark matter halo through the equation:

$$M(r) = \frac{\gamma\Phi_o}{G} \left(\frac{r}{\text{kpc}} \right)^{1-\gamma} \quad (3.3)$$

(Deason et al., 2012b). For $\gamma \rightarrow 0$, $M(r)$ approaches an isothermal sphere. The opposite extreme, $\gamma \rightarrow 1$, corresponds to the Keplerian case of a central point mass. Mathematically, the values of Φ_o and γ in Equation 3.3 may be any pair of real numbers. However, physically there are restrictions on their allowed values: Φ_o plays a large role in determining the mass and must be positive, and $0 < \gamma < 1$ if the cumulative mass profile is to be a constant or increasing function of radius. Although the parameters α and β do not appear in Equation 3.3, their values will determine the shape of the posterior distribution, and may affect the estimates for Φ_o and γ .

We use the Eddington-equation method described in Binney & Tremaine (2008) to solve for the isotropic DF $f(\mathcal{E})$ in terms of the binding energy $\mathcal{E} = -v^2/2 + \Phi(r)$, where

$$f(\mathcal{E}) = \frac{1}{\sqrt{8\pi^2}} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E} - \Phi}} \left(\frac{d^2 \rho_t}{d\Phi^2} \right) d\Phi + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{d\rho_t}{d\Phi} \right)_{\Phi=0} \quad (3.4)$$

To solve Equation 3.4, Equation 3.1 must be written as a function of Equation 3.2. Solving Equation 3.2 for r , and substituting into Equation 3.1 gives

$$\rho_t(\Phi) \propto \left(\frac{\Phi}{\Phi_o} \right)^{\alpha/\gamma}. \quad (3.5)$$

The derivatives of the above equation are,

$$\frac{d\rho_t}{d\Phi} \propto \frac{\alpha}{\gamma} \left(\frac{\Phi}{\Phi_o} \right)^{\frac{\alpha}{\gamma}-1} \frac{1}{\Phi_o} \quad (3.6)$$

and

$$\frac{d^2 \rho_t}{d\Phi^2} \propto \frac{\alpha}{\gamma} \left(\frac{\alpha}{\gamma} - 1 \right) \left(\frac{\Phi}{\Phi_o} \right)^{\frac{\alpha}{\gamma}-2} \frac{1}{\Phi_o^2}. \quad (3.7)$$

We assume and require that $\frac{\alpha}{\gamma} > 1$ (i.e. the satellite profile is steeper than the dark matter profile). With this restriction in place, Equation 3.6 goes to zero as $\Phi \rightarrow 0$, making the second term in Equation 3.4 vanish. The DF for the isotropic case is then,

$$f(\mathcal{E}) = \frac{1}{\sqrt{8\pi^2}} \frac{\alpha}{\gamma} \left(\frac{\alpha}{\gamma} - 1 \right) \frac{1}{\Phi_o^2} \int_0^{\mathcal{E}} \frac{1}{\sqrt{\mathcal{E} - \Phi}} \left(\frac{\Phi}{\Phi_o} \right)^{\frac{\alpha}{\gamma} - 2} d\Phi. \quad (3.8)$$

The solution to this integral is analytic, albeit tedious. Use the substitution $u = \Phi/\mathcal{E}$ to solve the integral, and then apply the recursion relation $x\Gamma(x) = \Gamma(x+1)$, twice, to simplify the final expression to

$$f(\mathcal{E}) = \frac{\mathcal{E}^{\frac{\alpha}{\gamma} - \frac{3}{2}} \Gamma\left(\frac{\alpha}{\gamma} + 1\right)}{\sqrt{8\pi^3} \Phi_o^{\frac{\alpha}{\gamma}} \Gamma\left(\frac{\alpha}{\gamma} - \frac{1}{2}\right)}. \quad (3.9)$$

Equation 3.9 is the probability distribution of the specific energies of tracers in the potential Φ , *assuming an isotropic velocity dispersion*.

The velocity dispersion of Milky Way satellites is likely to be at least mildly anisotropic. Cuddeford (1991) showed that a way to incorporate velocity anisotropy is to multiply the DF by the specific angular momentum $L = rv_t$,

$$f(\mathcal{E}, L) \propto L^{-2\beta} f(\mathcal{E}). \quad (3.10)$$

Here, β is the velocity anisotropy parameter,

$$\beta = 1 - \frac{\sigma_\theta^2 + \sigma_\phi^2}{2\sigma_r^2} \quad (3.11)$$

and σ_θ^2 , σ_ϕ^2 , and σ_r^2 are the velocity variances in spherical coordinates (Binney & Tremaine, 2008). A system with completely radial orbits or completely

tangential orbits will have $\beta = 1$ or $\beta \rightarrow -\infty$ respectively, while an isotropic velocity dispersion will have $\beta = 0$.

Evans et al. (1997) derived the normalization for Equation 3.10, and this factor depends on the parameters of interest. We reproduce their result here, the complete anisotropic DF, slightly re-organized and in Deason's notation:

$$f(\mathcal{E}, L) = \frac{L^{-2\beta} \mathcal{E}^{\frac{\beta(\gamma-2)}{\gamma} + \frac{\alpha}{\gamma} - \frac{3}{2}}}{\sqrt{8\pi^3} 2^{-2\beta} \Phi_o^{-\frac{2\beta}{\gamma} + \frac{\alpha}{\gamma}}} \frac{\Gamma\left(\frac{\alpha}{\gamma} - \frac{2\beta}{\gamma} + 1\right)}{\Gamma\left(\frac{\beta(\gamma-2)}{\gamma} + \frac{\alpha}{\gamma} - \frac{1}{2}\right)}. \quad (3.12)$$

Note that as $\beta \rightarrow 0$, Equation 3.12 reduces to Equation 3.9. In summary, this model has four parameters:

Φ_o the scale factor for the gravitational potential

γ the power-law slope of the gravitational potential

α the power-law slope of the satellite population

β the velocity anisotropy parameter

The parameters γ , α , and β are restricted by Equation 3.12 and the requirement that the DF be non-negative,

$$\alpha > \beta(2 - \gamma) + \frac{\gamma}{2} \quad (3.13)$$

(Evans et al., 1997, but be aware of notational differences).

In practice, the dark matter halo profile is often assumed to follow an NFW-type or Sérsic/Einasto-type function (e.g. Merritt et al., 2006). We experimented with such models but found that converting these to a DF becomes intractable analytically. There are numerical approximations to such models

that are themselves quite complex (e.g. Widrow, 2000), but for the present purpose we stick to analytic models for simplicity. While a power-law potential for the dark matter is simplistic, it is also a common assumption in methodologies that use DFs (Deason et al., 2012a; Williams & Evans, 2015). Furthermore, a power-law potential (Equation 3.2) has the benefit of approximating an NFW profile at large radii when $\gamma = 0.5$ (Watkins et al. 2010; Deason et al. 2011), with recent analyses suggesting a transition radius around 10kpc (Huang et al., 2016; Harris, 2001).

In this study, we use the Deason power-law model shown in Equation 3.12, and assume that all tracers are bound to the spherically symmetric, non-rotating system, i.e., $\mathcal{E} > 0$. We explore the parameter space and mass profiles predicted by the model when it is confronted with real data.

3.3 Kinematic Data: Globular Clusters

In this study, and in contrast to EHW, we use only GCs to trace the Galactic potential. In principle, we could include kinematic data for both DGs and GCs to help extend the estimated $M(r)$ profile to larger distances. However, the model parameter α is meant to describe the power-law slope distribution of a single population, and the GCs' and DGs' spatial distributions may be quite different. Thus, we only use GCs, but will return to this point in a later paper. Although metal-rich and metal-poor GCs may also have different distributions, we treat them as a *single* population: the metal-poor ones ($[\text{Fe}/\text{H}] < -1$) dominate the numbers, particularly at large Galactocentric radius.

Table 3.1 lists all the kinematic data available for 157 MW GCs, using the

catalog of Harris (1996, 2010 edition) as a starting point. The Heliocentric line-of-sight velocities v_{los} and the Galactocentric distances r are from the Harris catalog, while the proper motions are taken from a variety of studies (see the “ μ Reference” column in the table). The Galactocentric distances were calculated assuming the Sun’s position with respect to the Galactic center as $(X_{\odot}, Y_{\odot}, Z_{\odot}) = (8.0, 0, 0.02)$ kpc (the height above the midplane is from Humphreys & Larsen, 1995).

Almost half of the GCs in Table 3.1 have measured proper motions, many of which are from the series of papers by Casetti-Dinescu, referenced collectively as “Casetti” in the table (Dinescu et al., 1999, 2004, 2005; Casetti-Dinescu et al., 2010, 2013)¹. The GC M79 (NGC 1904) has two proper motion measurements we are aware of: the first was calculated by Dinescu et al. (1999) and is included in the Casetti online catalog, and the second was calculated by Wang et al. (2005) using 29 years of photographic plates from the Shanghai Observatory. We use the result from Wang et al. (2005) because it is more recent, and agrees well with Dinescu et al. (1999).

Many Galactic GC proper motions are still unknown, although there are observational programs such as HSTPROMO (Sohn, S. et al 2016, in progress) which are beginning to remedy this issue. For example, the proper motions of inner bulge GCs NGC 6522, NGC 6558, NGC 6540, NGC 6652, AL 3, ESO 456-SC38, Palomar 6, Terzan 2, Terzan 4, and Terzan 9 were recently measured by Rossi et al. (2015). A proper motion measurement for NGC 6681 was made for the first time by Massari et al. (2013), and an updated measurement for 47 Tuc (NGC 104) was recently completed by Cioni et al. (2016).

Two GCs found in the bulge, NGC 6528 and NGC 6553, have proper

¹Updated catalog: www.astro.yale.edu/dana/gc.html

motion measurements too (Feltzing & Johnson, 2002; Zoccali et al., 2001, respectively), which can also be found in the Caseti online catalog. Pal 5's proper motion was measured by Fritz & Kallivayalil (2015) and Küpper et al. (2015), and their uncertainties overlap, so we use the average μ of these two studies.

We exclude some GCs in Table 3.1 from our analysis:

1. A few objects simply do not have any velocity measurements, such as Ko 1, Ko 2, and AM 4.
2. Some GCs are known to be associated with the Sagittarius DG, and their motions may be highly correlated with one another; these objects include Arp 2, NGC 6715, NGC 5634, Terzan 7, Terzan 8, and Whiting 1 (Law & Majewski, 2010a).
3. The following GCs are subject to very high reddening: NGC 6401, NGC 6544, Terzan 1, Pal 6, Djorg 1, and Terzan 6. However, because of the careful new measurements by Rossi et al. (2015), we include Terzan 1 and Pal 6 in the analysis.
4. Another GC of issue is outer-halo object Pal 3; EHW showed that Pal 3's proper motion substantially affects the total mass estimate for the Milky Way. However, the proper motion measurement for Pal 3 is highly uncertain (Majewski & Cudworth, 1993). We are unaware of an updated proper motion measurement for Pal 3, and thus decide to treat the proper motion for this GC as unknown in the analysis.
5. We exclude some GCs that do not have proper motions, and explain why below.

To incorporate the incomplete data (GCs without proper motions) via the methodology of EHW, we must make the approximation that the Heliocentric line-of-sight velocity $|v_{\text{los}}|$ is approximately equal to the Galactocentric radial velocity $|v_r|$. This approximation is only valid when the GC is either (1) very far away, or (2) in line with the Sun and the Galactic center. To quantify this argument, we define ξ as the angle subtended by the Sun, the GC, and the center of the Galaxy, and as in EHW require that GCs without a proper motion measurement have $|\cos\xi| > 0.95$ in order to be included. In these cases we can safely make the approximation that $|v_{\text{los}}| \approx |v_r|$.

After culling the data for the above reasons, only 89 GCs remain in the sample. We note that our main motivation for using the EHW Bayesian method was to incorporate all the available data, but we have just tossed aside over a third of the data, mostly due to our imposed geometric requirement that $|\cos\xi| > 0.95$. However, this is a temporary problem; in our next paper (G. M. Eadie et al. 2016, in preparation), we will introduce a hierarchical version of the EHW method that negates the need for this geometric requirement. The GCs included in our current sample have a ✓ in the “Included” column in Table 3.1; 18 of the 89 GCs included in our sample have incomplete velocity measurements.

Table 3.1: Kinematic Data of Galactic Globular Clusters

Object	r kpc	$\mu_\alpha \cos \delta$ mas/year	μ_δ mas/year	v_{los} km s ⁻¹	$\cos \xi$	μ Reference	Included
NGC 104	7.4	7.26 ± 0.03	-1.25 ± 0.03	-18 ± 0.1	0.17	Cioni	✓
NGC 288	12.0	4.675 ± 0.219	-5.6 ± 0.35	-45.4 ± 0.2	0.75	Cassetti	✓
NGC 362	9.4	4.873 ± 0.514	-2.727 ± 0.824	223.5 ± 0.5	0.61	Cassetti	✓
Whiting 1	34.5	—	—	-130.6 ± 1.8	0.98	—	—
NGC 1261	18.1	—	—	68.2 ± 4.6	0.90	—	—
Pal 1	17.2	—	—	-82.8 ± 3.3	0.93	—	—
AM 1	124.6	—	—	116 ± 20	1.00	—	✓
Eridanus	95.0	—	—	-23.6 ± 2.1	1.00	—	✓
Pal 2	35.0	—	—	-133 ± 57	1.00	—	✓
NGC 1851	16.6	1.28 ± 0.68	2.39 ± 0.65	320.5 ± 0.6	0.89	Cassetti	✓
NGC 1904	18.8	2.34 ± 0.69	-0.5 ± 0.75	205.8 ± 0.4	0.94	Wang	✓
NGC 2298	15.8	4.05 ± 1	-1.72 ± 0.98	148.9 ± 1.2	0.89	Cassetti	✓
NGC 2419	89.9	—	—	-20.2 ± 0.5	1.00	—	✓
Ko 2	41.9	—	—	—	1.00	—	—
Pyxis	41.4	—	—	34.3 ± 1.9	0.98	—	✓
NGC 2808	11.1	0.58 ± 0.45	2.06 ± 0.46	101.6 ± 0.7	0.71	Cassetti	✓
E 3	9.1	—	—	—	0.57	—	—
Pal 3	95.7	0.33 ± 0.23	0.3 ± 0.31	83.4 ± 8.4	1.00	Majewski & Cudworth	✓ (μ not included)
NGC 3201	8.8	5.28 ± 0.32	-0.98 ± 0.33	494 ± 0.2	0.43	Cassetti	✓
Pal 4	111.2	—	—	74.5 ± 2.1	1.00	—	✓
Ko 1	49.3	—	—	—	0.99	—	—
NGC 4147	21.4	-1.54 ± 0.54	-3.285 ± 0.516	183.2 ± 0.7	0.93	Cassetti	✓

Object	r kpc	$\mu_\alpha \cos \delta$ mas/year	μ_δ mas/year	v_{los} km s ⁻¹	$\cos \xi$	μ Reference	Included
NGC 4372	7.1	-6.49 ± 0.33	3.71 ± 0.32	72.3 ± 1.2	0.24	Casseti	✓
Rup 106	18.5	—	—	-44 ± 3	0.93	—	—
NGC 4590	10.2	-3.76 ± 0.66	1.79 ± 0.62	-94.7 ± 0.2	0.70	Casseti	✓
NGC 4833	7.0	-8.11 ± 0.35	-0.96 ± 0.34	200.2 ± 1.2	0.31	Casseti	✓
NGC 5024	18.4	0.5 ± 1	-0.1 ± 1	-62.9 ± 0.3	0.90	Casseti	✓
NGC 5053	17.8	—	—	44 ± 0.4	0.90	—	—
NGC 5139	6.4	-5.08 ± 0.35	-3.57 ± 0.34	232.1 ± 0.1	0.06	Casseti	✓
NGC 5272	12.0	-0.12 ± 0.607	-2.667 ± 0.404	-147.6 ± 0.2	0.75	Casseti	✓
NGC 5286	8.9	—	—	57.4 ± 1.5	0.73	—	—
AM 4	27.8	—	—	—	0.98	—	—
NGC 5466	16.3	-3.9 ± 1	1 ± 1	110.7 ± 0.2	0.88	Casseti	✓
NGC 5634	21.2	—	—	-45.1 ± 6.6	0.96	—	—
NGC 5694	29.4	—	—	-140.3 ± 0.8	0.98	—	✓
IC 4499	15.7	—	—	31.5 ± 0.2	0.91	—	—
NGC 5824	25.9	—	—	-27.5 ± 1.5	0.98	—	✓
Pal 5	18.6	-2.343 ± 0.356	-2.3085 ± 0.331	-58.7 ± 0.2	0.95	Fritz; Kupper	✓
NGC 5897	7.4	-4.93 ± 0.86	-2.33 ± 0.84	101.5 ± 1	0.79	Casseti	✓
NGC 5904	6.2	4.267 ± 0.597	-11.3 ± 1.457	53.2 ± 0.4	0.33	Casseti	✓
NGC 5927	4.6	-5.72 ± 0.39	-2.61 ± 0.4	-107.5 ± 1	0.23	Casseti	✓
NGC 5946	5.8	—	—	128.4 ± 1.8	0.67	—	—
BH 176	12.9	—	—	—	0.94	—	—
NGC 5986	4.8	-3.81 ± 0.45	-2.99 ± 0.37	88.9 ± 3.7	0.67	Casseti	✓
Lynga 7	4.3	—	—	8 ± 5	0.27	—	—
Pal 14	71.6	—	—	72.3 ± 0.2	1.00	—	✓
NGC 6093	3.8	-3.31 ± 0.58	-7.2 ± 0.67	8.2 ± 1.5	0.66	Casseti	✓

Object	r kpc	$\mu_\alpha \cos \delta$ mas/year	μ_δ mas/year	v_{los} km s ⁻¹	$\cos \xi$	μ Reference	Included
NGC 6121	5.9	-12.657 ± 0.285	-19.387 ± 0.288	70.7 ± 0.2	-0.94	Casseti	✓
NGC 6101	11.2	—	—	361.4 ± 1.7	0.87	—	—
NGC 6144	2.7	-3.06 ± 0.64	-5.11 ± 0.72	193.8 ± 0.6	0.47	Casseti	✓
NGC 6139	3.6	—	—	6.7 ± 6	0.70	—	—
Terzan 3	2.5	—	—	-136.3 ± 0.7	0.23	—	—
NGC 6171	3.3	-0.7 ± 0.9	-3.1 ± 1	-34.1 ± 0.3	-0.29	Casseti	✓
1636-283	2.1	—	—	—	0.27	—	—
NGC 6205	8.4	-0.103 ± 0.797	4.687 ± 0.813	-244.2 ± 0.2	0.48	Casseti	✓
NGC 6229	29.8	—	—	-154.2 ± 7.6	0.97	—	✓
NGC 6218	4.5	1.15 ± 1.95	-7.75 ± 1.672	-41.4 ± 0.2	-0.48	Casseti	✓
FSR 1735	3.7	—	—	—	0.63	—	—
NGC 6235	4.2	—	—	87.3 ± 3.4	0.89	—	—
NGC 6254	4.6	-5.75 ± 0.778	-4.75 ± 1.45	75.2 ± 0.7	-0.58	Casseti	✓
NGC 6256	3.0	—	—	-101.4 ± 1.9	0.83	—	—
Pal 15	38.4	—	—	68.9 ± 1.1	0.99	—	✓
NGC 6266	1.7	-3.5 ± 0.37	-0.82 ± 0.37	-70 ± 1.4	-0.64	Casseti	✓
NGC 6273	1.7	-2.86 ± 0.49	-0.45 ± 0.51	135 ± 4	0.55	Casseti	✓
NGC 6284	7.5	-3.66 ± 0.64	-5.39 ± 0.83	27.6 ± 1.7	0.99	Casseti	✓
NGC 6287	2.1	-3.68 ± 0.88	-3.54 ± 0.69	-288.7 ± 3.5	0.73	Casseti	✓
NGC 6293	1.9	0.26 ± 0.85	-5.14 ± 0.71	-146.2 ± 1.7	0.83	Casseti	✓
NGC 6304	2.3	-2.59 ± 0.29	-1.56 ± 0.29	-107.3 ± 3.6	-0.88	Casseti	✓
NGC 6316	2.6	-2.42 ± 0.63	-2.65 ± 0.56	71.5 ± 8.9	0.94	Casseti	✓
NGC 6341	9.6	-3.575 ± 0.893	-0.6 ± 0.601	-120 ± 0.1	0.61	Casseti	✓
NGC 6325	1.1	—	—	29.8 ± 1.8	-0.11	—	—
NGC 6333	1.7	-0.57 ± 0.57	-3.7 ± 0.5	229.1 ± 7	0.05	Casseti	✓

Object	r kpc	$\mu_\alpha \cos \delta$ mas/year	μ_δ mas/year	v_{los} km s ⁻¹	$\cos \xi$	μ Reference	Included
NGC 6342	1.7	-2.77 ± 0.71	-5.84 ± 0.65	115.7 ± 1.4	0.39	Cassetti	✓
NGC 6356	7.5	-3.14 ± 0.68	-3.65 ± 0.53	27 ± 4.3	0.97	Cassetti	✓
NGC 6355	1.4	—	—	-176.9 ± 7.1	0.88	—	—
NGC 6352	3.3	—	—	-137 ± 1.1	-0.59	—	—
IC 1257	17.9	—	—	-140.2 ± 2.1	0.98	—	✓
Terzan 2	0.8	-0.94 ± 0.3	0.15 ± 0.42	109 ± 15	-0.59	Rossi	✓
NGC 6366	5.0	—	—	-122.2 ± 0.5	-0.76	—	—
Terzan 4	1.0	3.5 ± 0.69	0.35 ± 0.58	-50 ± 2.9	-0.78	Rossi	✓
HP 1	0.5	—	—	45.8 ± 0.7	0.43	—	—
NGC 6362	5.1	-3.09 ± 0.46	-3.83 ± 0.46	-13.1 ± 0.6	0.26	Cassetti	✓
Liller 1	0.8	—	—	52 ± 15	0.30	—	—
NGC 6380	3.3	—	—	-3.6 ± 2.5	0.91	—	—
Terzan 1	1.3	0.51 ± 0.31	-0.93 ± 0.29	114 ± 14	-1.00	Rossi	✓
Ton 2	1.4	—	—	-184.4 ± 2.2	0.23	—	—
NGC 6388	3.1	-1.9 ± 0.45	-3.83 ± 0.51	80.1 ± 0.8	0.71	Cassetti	✓
NGC 6402	4.0	—	—	-66.1 ± 1.8	0.52	—	—
NGC 6401	2.7	—	—	-65 ± 8.6	0.97	—	—
NGC 6397	6.0	3.69 ± 0.29	-14.88 ± 0.26	18.8 ± 0.1	-0.82	Cassetti	✓
Pal 6	2.2	2.95 ± 0.41	1.24 ± 0.19	181 ± 2.8	-1.00	Rossi	✓
NGC 6426	14.4	—	—	-162 ± 23	0.96	—	✓
Djorg 1	5.7	—	—	-362.4 ± 3.6	1.00	—	—
Terzan 5	1.2	—	—	-93 ± 2	-0.90	—	—
NGC 6440	1.3	—	—	-76.6 ± 2.7	0.45	—	—
NGC 6441	3.9	-2.86 ± 0.45	-3.45 ± 0.76	16.5 ± 1	0.95	Cassetti	✓
Terzan 6	1.3	—	—	126 ± 15	-0.91	—	—

Object	r kpc	$\mu_\alpha \cos \delta$ mas/year	μ_δ mas/year	v_{los} km s $^{-1}$	$\cos \xi$	μ Reference	Included
NGC 6453	3.7	—	—	-83.7 ± 8.3	0.98	—	✓
UKS 1	0.7	—	—	57 ± 6	-0.24	—	—
NGC 6496	4.2	—	—	-112.7 ± 5.7	0.86	—	—
Terzan 9	1.1	0 ± 0.38	-3.07 ± 0.49	59 ± 10	-0.79	Rossi	✓
Djorg 2	1.8	—	—	—	-0.93	—	—
NGC 6517	4.2	—	—	-39.6 ± 8	0.74	—	—
Terzan 10	2.3	—	—	—	-0.94	—	—
NGC 6522	0.6	3.35 ± 0.6	-1.19 ± 0.34	-21.1 ± 3.4	-0.47	Rossi	✓
NGC 6535	3.9	—	—	-215.1 ± 0.5	-0.05	—	—
NGC 6528	0.6	-0.35 ± 0.23	0.27 ± 0.26	206.6 ± 1.4	-0.13	Feltzing & Johnson	✓
NGC 6539	3.0	—	—	31 ± 1.7	0.12	—	—
NGC 6540	2.8	0.07 ± 0.4	1.9 ± 0.57	-17.7 ± 1.4	-0.95	Rossi	✓
NGC 6544	5.1	—	—	-27.3 ± 3.9	-0.95	—	—
NGC 6541	2.1	—	—	-158.7 ± 2.4	-0.11	—	—
2MS-GC01	4.5	—	—	—	-0.95	—	—
ESO-SC06	14.0	—	—	—	0.98	—	—
NGC 6553	2.2	2.5 ± 0.065	5.35 ± 0.076	-3.2 ± 1.6	-0.88	Zoccali	✓
2MS-GC02	3.2	—	—	-238 ± 36	-0.95	—	—
NGC 6558	1.0	-0.12 ± 0.55	0.47 ± 0.6	-197.2 ± 1.5	-0.56	Rossi	✓
IC 1276	3.7	—	—	155.7 ± 1.3	-0.53	—	—
Terzan 12	3.4	—	—	94.1 ± 1.5	-0.90	—	—
NGC 6569	3.1	—	—	-28.1 ± 5.6	0.95	—	✓
BH 261	1.7	—	—	—	-0.85	—	—
GLIMPSE02	3.0	—	—	—	-0.75	—	—

Object	r kpc	$\mu_\alpha \cos \delta$ mas/year	μ_δ mas/year	v_{los} km s ⁻¹	$\cos \xi$	μ Reference	Included
NGC 6584	7.0	-0.22 ± 0.62	-5.79 ± 0.67	222.9 ± 15	0.88	Casseti	✓
NGC 6624	1.2	—	—	53.9 ± 0.6	-0.01	—	—
NGC 6626	2.7	0.63 ± 0.67	-8.46 ± 0.67	17 ± 1	-0.89	Casseti	✓
NGC 6638	2.2	—	—	18.1 ± 3.9	0.71	—	—
NGC 6637	1.7	—	—	39.9 ± 2.8	0.55	—	—
NGC 6642	1.7	—	—	-57.1 ± 5.4	0.16	—	—
NGC 6652	2.7	4.75 ± 0.07	-4.45 ± 0.1	-111.7 ± 5.8	0.80	Rossi	✓
NGC 6656	4.9	7.37 ± 0.5	-3.95 ± 0.42	-146.3 ± 0.2	-0.95	Casseti	✓
Pal 8	5.5	—	—	-43 ± 15	0.92	—	—
NGC 6681	2.2	1.58 ± 0.18	-4.57 ± 0.16	220.3 ± 0.9	0.55	Massari	✓
GLIMPSE01	4.9	—	—	—	-0.54	—	—
NGC 6712	3.5	4.2 ± 0.4	-2 ± 0.4	-107.6 ± 0.5	-0.09	Casseti	✓
NGC 6715	18.9	—	—	141.3 ± 0.3	0.99	—	—
NGC 6717	2.4	—	—	22.8 ± 3.4	-0.23	—	—
NGC 6723	2.6	-0.17 ± 0.45	-2.16 ± 0.5	-94.5 ± 3.6	0.41	Casseti	✓
NGC 6749	5.0	—	—	-61.7 ± 2.9	0.30	—	—
NGC 6752	5.2	-0.69 ± 0.42	-2.85 ± 0.45	-26.7 ± 0.2	-0.50	Casseti	✓
NGC 6760	4.8	—	—	-27.5 ± 6.3	0.19	—	—
NGC 6779	9.2	0.3 ± 1	1.4 ± 1	-135.7 ± 0.8	0.63	Casseti	✓
Terzan 7	15.6	—	—	166 ± 4	0.98	—	—
Pal 10	6.4	—	—	-31.7 ± 0.4	0.16	—	—
Arp 2	21.4	—	—	115 ± 10	0.99	—	—
NGC 6809	3.9	-3.31 ± 0.945	-9.695 ± 0.554	174.7 ± 0.3	-0.47	Casseti	✓
Terzan 8	19.4	—	—	130 ± 8	0.98	—	—
Pal 11	8.2	—	—	-68 ± 10	0.83	—	—

Object	r kpc	$\mu_\alpha \cos \delta$ mas/year	μ_δ mas/year	v_{los} km s ⁻¹	$\cos \xi$	μ Reference	Included
NGC 6838	6.7	-2.3 ± 0.8	-5.1 ± 0.8	-22.8 ± 0.2	-0.06	Casseti	✓
NGC 6864	14.7	—	—	-189.3 ± 3.6	0.96	—	✓
NGC 6934	12.8	1.2 ± 1	-5.1 ± 1	-411.4 ± 1.6	0.86	Casseti	✓
NGC 6981	12.9	—	—	-345.1 ± 3.7	0.89	—	—
NGC 7006	38.5	-0.96 ± 0.35	-1.14 ± 0.4	-384.1 ± 0.4	0.98	Casseti	✓
NGC 7078	10.4	-1.233 ± 0.617	-7.567 ± 1.77	-107 ± 0.2	0.70	Casseti	✓
NGC 7089	10.4	5.9 ± 0.849	-4.95 ± 0.849	-5.3 ± 2	0.74	Casseti	✓
NGC 7099	7.1	1.42 ± 0.69	-7.71 ± 0.65	-184.2 ± 0.2	0.45	Casseti	✓
Pal 12	15.8	-1.2 ± 0.3	-4.21 ± 0.29	27.8 ± 1.5	0.91	Casseti	✓
Pal 13	26.9	2.3 ± 0.26	0.27 ± 0.25	25.2 ± 0.3	0.95	Casseti	✓
NGC 7492	25.3	—	—	-177.5 ± 0.6	0.95	—	✓

Note: This table is available in machine-readable format

3.4 Methods

3.4.1 Overview and Bayesian Inference

We estimate the mass profile of the Milky Way's dark matter halo by assuming the power-law model described in Section 3.2, using the Bayesian method outlined in EHW, and confronting this model with the GC data described in Section 3.3. For numerical purposes, we use $G \equiv 1$ units. The cumulative mass profile in $10^{12}M_{\odot}$ units is then given by

$$M(< r) = 2.325 \times 10^{-3} \gamma \Phi_o \left(\frac{r}{\text{kpc}} \right)^{1-\gamma}. \quad (3.14)$$

where Φ_o has units $10^4 \text{km}^2 \text{s}^{-2}$ and r is in kpc.

The DFs in Eq. 3.9 and 3.10 require Galactocentric velocities in a spherical coordinate system, rather than the Heliocentric proper motion and line-of-sight measurements presented in Table 3.1. In the Galactocentric spherical coordinate system, the binding energy \mathcal{E} is given by

$$\mathcal{E} = -\frac{1}{2}(v_r^2 + v_t^2) + \Phi(r) \quad (3.15)$$

where v_r and $v_t = \sqrt{v_{\phi}^2 + v_{\theta}^2}$ are the radial and tangential velocities respectively.

Heliocentric velocities $(v_{\text{los}}, \mu_{\alpha} \cos \delta, \mu_{\delta})$ are transformed to Galactocentric velocities (U, V, W) in a right-handed cylindrical coordinate system, following the method outlined in Johnson & Soderblom (1987), but using J2000 epoch values for the North Galactic Pole. We assume the velocity of the Sun with respect to the local standard of rest is $(U_{\odot}, V_{\odot}, W_{\odot}) = (11.1, 12.24, 7.25)$

km s⁻¹ (Schönrich et al., 2010), and take the local standard of rest velocity to be 22 km s⁻¹. After transforming to (U, V, W) , the velocities are transformed to spherical coordinates (v_r, v_ϕ, v_θ) .

When a tracer does not have a proper motion measurement, then the transformations described above cannot be computed. For these objects, v_t in Equation 3.15 is still unknown. This is where using the Bayesian paradigm comes in handy: the unknown v_t 's can be treated as parameters in the model.

Bayes' theorem states that the posterior probability distribution $p(\boldsymbol{\theta}|\mathbf{y})$ is the probability of model parameters $(\boldsymbol{\theta})$, conditional on a set of data \mathbf{y} :

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \quad (3.16)$$

where $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood, and $p(\boldsymbol{\theta})$ is the prior probability on $\boldsymbol{\theta}$ (Bayes & Price, 1763). The denominator is a normalization constant whose value is not of interest— we may instead sample a distribution that is proportional to the posterior distribution,

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (3.17)$$

to obtain probabilities of model parameters given the data.

If there are n tracers, each with data (r, v_r, v_t) , and assumed to be independent, then the posterior probability is proportional to the product

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &\propto \prod_i^n p(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &\propto \prod_i^n p((r_i, v_{r,i}, v_{t,i})|\boldsymbol{\theta})p(\boldsymbol{\theta}). \end{aligned}$$

In the case that v_t is unknown, it becomes a nuisance parameter in the model,

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \prod_i^n p((r_i, v_{r,i})|\boldsymbol{\theta}, v_{t,i}) p(v_{t,i}) p(\boldsymbol{\theta}). \quad (3.18)$$

We define $p(v_{t,i})$, the prior probability on $v_{t,i}$, as a uniform distribution in v_t^2 (this accounts for spherical geometry). The nuisance parameters are sampled via a hybrid-Gibbs sampler, which is a mixture of a standard Metropolis algorithm (Metropolis & Ulam, 1949; Metropolis et al., 1953) and a Gibbs sampler (Geman & Geman, 1984). This method is an efficient way to treat the unknown tangential velocities as parameters; see Eadie (2013) and EHW for more details. For a comprehensive description of Gibbs sampling, see Gelman et al. (2003).

3.4.2 Markov Chains

Samples of the posterior distribution are drawn via the hybrid-Gibbs sampler described in EHW. We run three independent Markov chains in parallel: the chains are initialized in different parts of parameter space and run until they reach a common stationary distribution (Figure 3.1). The mutual convergence of the chains is assessed by inspecting the trace plots of the chains and by calculating the \widehat{R} statistic (Gelman & Rubin, 1992).

Figure 3.1 is an example of a trace plot for three Markov chains that were initialized at different Φ_o values, but which have reached a common posterior distribution. Within the first few hundred iterations (the burn-in), these chains appear to have reached a common location in parameter space. The burn-in is discarded, and the Markov chains are run for at least 10^4 more iterations, after which we confirm the effective sample size (n_{eff}) of the chains is at least 1000

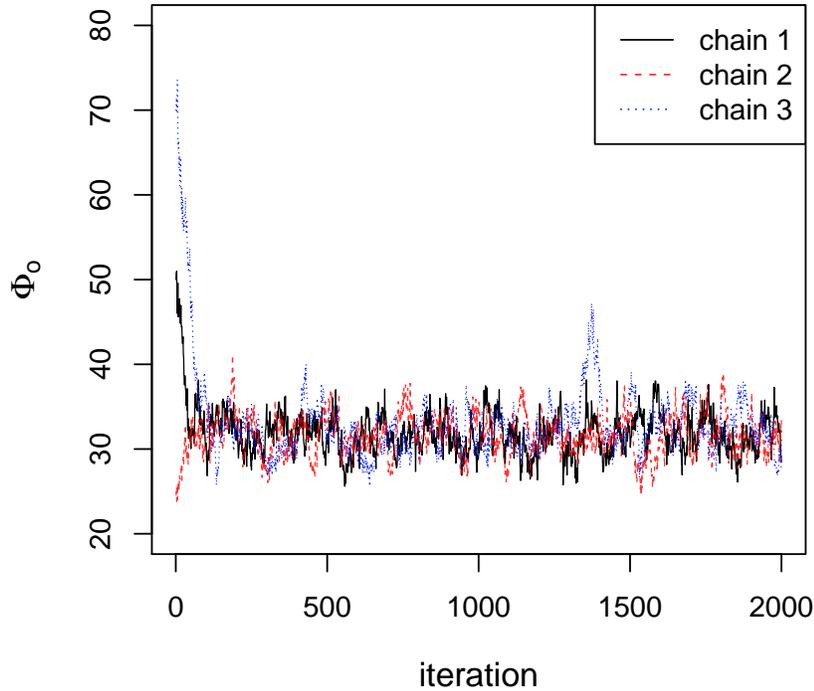


Figure 3.1: Example trace plot of three parallel Markov chains that are sampling the parameter space of Φ_0 .

(see EHW for a brief description of n_{eff}). After all requirements have been met (visual convergence, a 30%-45% acceptance rate, $\hat{R} < 1.1$, and $n_{\text{eff}} > 1000$), we accept that the final Markov chain samples have a distribution that is proportional to the posterior distribution (Equation 3.18). At this point, we calculate statistics, estimates, and probabilities of model parameters.

3.4.3 Priors

Bayesian inference requires choosing prior probability distributions for model parameters. We use flat, uniform prior probability distributions for three out of four model parameters, with the lower and upper bounds listed in Table 3.2. When a parameter is held fixed, then the prior probability is a δ -function

centered on the chosen parameter value.

The prior on Φ_o is quite wide, representing little prior knowledge of the Galaxy’s mass. In the cases of γ and β , there are clear mathematical reasons for the prior bounds; γ must be positive for the halo potential to decrease with distance (Equation 3.2), and β is the conventional anisotropy parameter, which has the limits described in Section 3.2.

Unfortunately, using $\beta \rightarrow -\infty$ as a lower-bound to a uniform distribution is ill-defined. On the other hand, there is strong evidence to expect $\beta \gtrsim 0$ for the Galactic stellar halo (Kafle et al., 2014), and previous studies have shown that the velocity distribution of GCs in the MW halo is mildly radial (Deason et al., 2011). Furthermore, values of $\beta < -1$ are known to be unrealistic velocity anisotropies for distant halo stars (e.g. Cunningham et al., 2016; Deason et al., 2013). Taking all of this information into account, we set a conservative lower limit of $\beta_{\text{lower}} = -0.5$.

The parameter α determines the spherical density distribution of the GCs, and has been shown to follow a power-law profile with index ~ 3.5 (Harris, 1976; Djorgovski & Meylan, 1994). Given our knowledge of the GC spatial distribution around the MW, it seems reasonable to define a slightly more informative prior distribution for this parameter. Although it may be tempting to look at all the GC r values, fit a power-law profile, and then use the best-fit parameter value of α as a way to define a prior distribution, Bayesian inference can only use the data once; we cannot use the data to define a prior and then also use the same data to calculate the posterior distribution. This is especially true here, where the r value of the GCs will help constrain not only α , but also the potential through Equation 3.15. So, how do we define a prior for α ?

We use the 68 MW GCs that are *excluded* from our analysis (see Section 3.3) to determine the prior distribution for α . The procedure to define the prior probability distribution $p(\alpha)$ is as follows.

First we hypothesize that the true density profile of the GCs is a power-law profile $\rho(r) \propto r^{-\alpha}$, implicitly assuming spherical symmetry. We can re-write the power law as,

$$\rho(r) \propto \frac{1}{4\pi r^2} \times \frac{1}{r^{\alpha-2}}. \quad (3.19)$$

where the first term is one over the area of a sphere with radius r . Thus, in 1-dimension, the GCs follow a power-law with index $\alpha - 2$. This kind of distribution can be described by a Pareto Distribution of the first kind.

The Pareto Distribution is a pdf defined as

$$f(x|\eta, b) = \eta b^\eta \frac{1}{x^{(\eta+1)}}, \quad 0 < b < x < \infty, \quad (3.20)$$

where b is a threshold parameter—the x value beyond which the data follow a power-law with index $\eta + 1$ (Howlander et al., 2007; Feigelson & Babu, 2012). The term ηb^η is the normalization constant. Note the power-law slope in Equation 3.20 is $\eta + 1$, but in Equation 3.19 is $\alpha - 2$. Thus, for the GCs, $\alpha = \eta + 3$.

Following Howlander et al. (2007), the posterior pdf for η given data vector \mathbf{x} of length n is,

$$p(\eta|b, \mathbf{x}) = \frac{(nc_o + p)^{c+n}}{\Gamma(c+n)} \eta^{c+n-1} e^{-\eta(nc_o+p)}, \eta > 0 \quad (3.21)$$

where c and p are parameters, and where

$$nc_o = \sum_i^n \log\left(\frac{x_i}{b}\right). \quad (3.22)$$

Equation 3.21 is a Gamma distribution with shape and scale parameters $(n+c)$ and $(nc_o + p)$.

We use Equation 3.21 to calculate the probability distribution of η given the extra data ($n = 68$), and then reparameterize to obtain a prior probability distribution for α (Figure 3.2). We let the extra GC data determine the shape and scale of the Gamma distribution as much as possible by defining $b = 0.4\text{kpc}$, $c = 0.001$, and $p = 0.001$. The most probable value in Figure 3.2 is $\alpha \approx 3.4$. Interestingly, this value is in excellent agreement with the power-law best-fit obtained by Wilkinson & Evans (1999), who used a mixture of both globular clusters and dwarf galaxies beyond 20kpc.

As a check, we plot the smoothed density estimate of the extra GC data and a power-law profile with index 3.4 in Figure 3.3. The smoothed density estimate is made using the *density* function in the **R** software environment, with a Gaussian kernel of bandwidth 1kpc. The power-law profile with slope 3.4 approximates the smoothed density quite well beyond 2kpc. We will return to this point in Section 3.5.3 below.

The spatial distribution of the 68 excluded GCs may be different from that of the 89 included GCs, due to selection effects (e.g., closer GCs may be more likely to have proper motion measurements). However, in a Bayesian analysis this is not a problem. By including the r information of the excluded GCs via the prior, we are actually taking into account the spatial distribution of the entire GC population. The estimate for α after using Bayes' theorem is

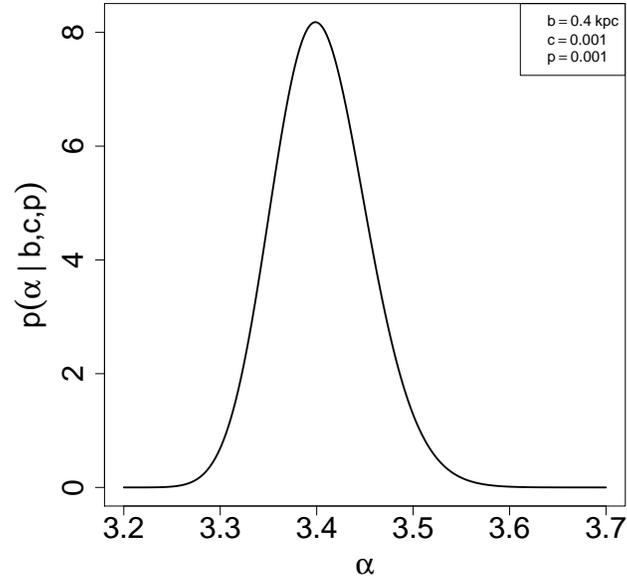


Figure 3.2: Prior probability distribution (a Gamma distribution) for α , as determined by the extra GC data ($n = 68$).

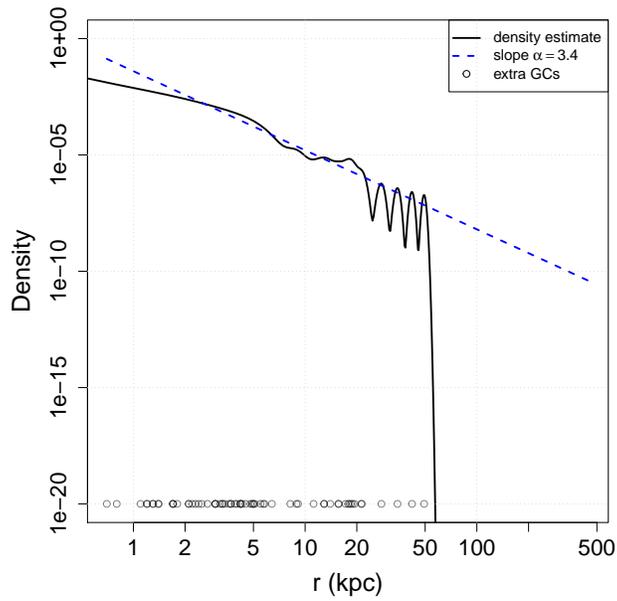


Figure 3.3: Density profile estimate of Milky Way Globular Clusters not used in this study. The dashed blue line has $\alpha = 3.4$.

Table 3.2: Prior Probability Distributions for Parameters

Parameter	Prior	Prior Parameters
Φ_o	Uniform	$\Phi_{o,\min} = 1, \Phi_{o,\max} = 200$
γ	Uniform	$\gamma_{\min} = 0, \gamma_{\max} = 1$
α	Gamma	$b = 0.4\text{kpc}, c = 0.001, p = 0.001$
β	Uniform	$\beta_{\min} = -0.5, \beta_{\max} = 1$

a compilation of the information from the prior and the information from the data.

3.4.4 Investigations

The Φ_o and γ parameters directly determine the MW’s mass profile (Equation 3.3), while parameters α and β may indirectly affect it. Thus, a simple analysis would let only Φ_o and γ be free parameters, while a more general analysis would let all four model parameters ($\Phi_o, \gamma, \alpha, \beta$) be free. To cover the range of possibilities, we instead perform an entire suite of investigations categorized first into Groups and then into Scenarios, as listed in Table 3.3. In every investigation, the parameter Φ_o is free. Group (1) holds $\gamma = 0.5$ and $\alpha = 3.5$ fixed, Group (2) holds $\alpha = 3.5$ fixed, and Group (3) holds $\gamma = 0.5$ fixed. Within each Group, we test Scenarios with different velocity anisotropies. For example, in Scenario IV, Φ_o and γ are free, α is fixed at 3.5, and we assume an isotropic velocity dispersion ($\beta = 0$). In Scenario VI, however, the anisotropy β is a free parameter. Scenario XII is the most general analysis; ($\Phi_o, \gamma, \alpha, \beta$) are all free.

Table 3.3: Investigations: fixed *vs.* free parameters

Group	Scenario	Potential	Tracers	Anisotropy
		γ	α	β
(1)	I			0
	II	0.5	3.5	0.5
	III			free
(2)	IV			0
	V	free	3.5	0.5
	VI			free
(3)	VII			0
	VIII	0.5	free	0.5
	IX			free
(4)	X			0
	XI	free	free	0.5
	XII			free

Table 3.4: Summary of Parameter Estimates and 50% marginal credible regions

Group	Scen.	Φ_o ($10^4 \text{km}^2 \text{s}^{-2}$)		γ	α		β		M_{125} (10^{11}) M_\odot		
(1)	I	32.9	(28.4-36.4)	0.50	—	3.50	—	0.00	—	4.28	(3.69-4.73)
	II	33.7	(28.8-37.4)	0.50	—	3.50	—	0.50	—	4.38	(3.75-4.86)
	III	33.4	(28.6-37.0)	0.50	—	3.50	—	0.35	(0.29-0.42)	4.34	(3.72-4.81)
(2)	IVb	29.5	(25.2-33.1)	0.32	(0.31-0.33)	3.50	—	0.00	—	5.88	(5.00-6.61)
	Vb	30.4	(25.7-34.3)	0.33	(0.31-0.34)	3.50	—	0.50	—	5.93	(4.97-6.72)
	VIb	29.5	(27.2-31.5)	0.33	(0.31-0.35)	3.50	—	0.27	(0.21-0.35)	5.70	(5.23-6.17)
(3)	VII	31.4	(27.4-34.5)	0.50	—	3.20	(3.18-3.23)	0.00	—	4.09	(3.56-4.48)
	VIII	31.6	(27.3-34.6)	0.50	—	3.20	(3.17-3.23)	0.50	—	4.10	(3.55-4.50)
	IX	31.5	(27.4-34.6)	0.50	—	3.20	(3.17-3.23)	0.36	(0.30-0.43)	4.10	(3.56-4.50)
(4)	X	27.4	(23.4-30.7)	0.32	(0.31-0.32)	3.20	(3.17-3.22)	0.00	—	5.47	(4.66-6.13)
	XI	27.4	(23.1-30.9)	0.33	(0.31-0.34)	3.20	(3.17-3.22)	0.50	—	5.36	(4.50-6.06)
	XII	26.9	(24.9-28.8)	0.33	(0.31-0.35)	3.20	(3.18-3.22)	0.28	(0.21-0.35)	5.22	(4.79-5.63)

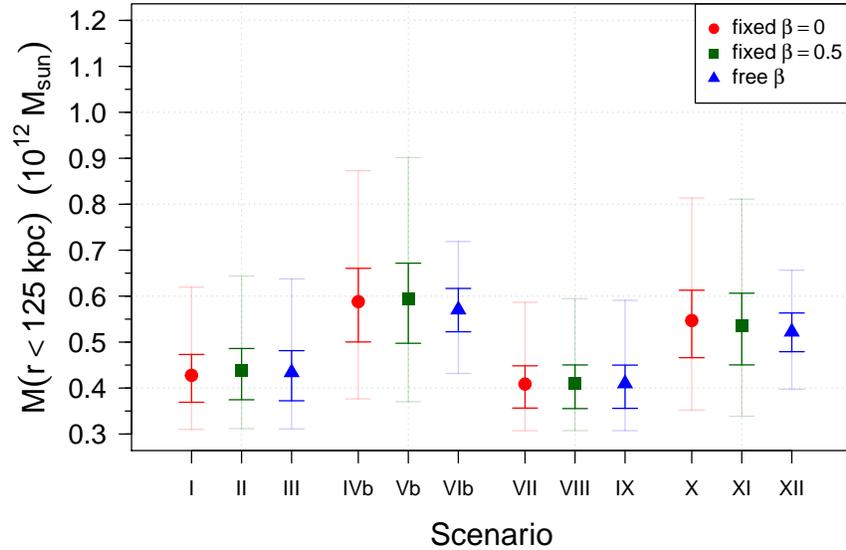


Figure 3.4: $M(r < 125)$ estimates for all Scenarios, with 50% and 95% Bayesian credible intervals shown as bright and faint error bars respectively.

3.5 Results

Table 3.4 summarizes the results of our study, showing the estimates of each parameter and their 50% marginal credible intervals in brackets. Figure 3.4 shows the mass estimates within 125kpc, with error bars indicating the 50% and 95% Bayesian credible intervals. The mass estimates are grouped as in Tables 3.3 and 3.4 in order to highlight both the differences between groups, and differences between the anisotropy assumptions within a group.

In the following four sections, the results are presented in more detail. Each section pertains to one of the Groups in Table 3.3. In this way, we intend to highlight the differences in mass estimates due to anisotropy assumptions versus other parameter assumptions, and also describe why these differences occur.

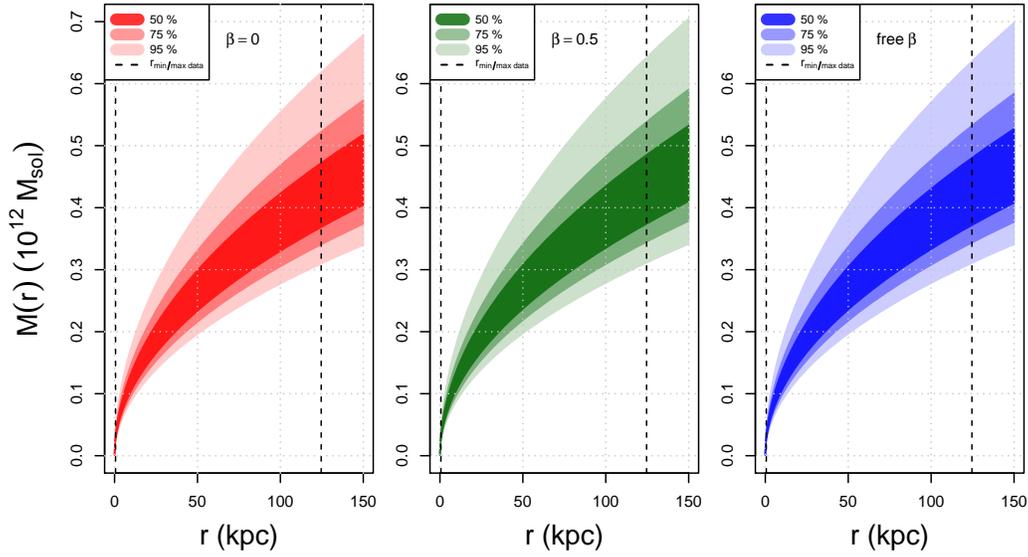


Figure 3.5: Credible regions for the cumulative mass profile in Scenarios I, II, and III. The percentages in the legend correspond to the three Bayesian credible regions shown, and the dotted lines show the extent of the GC data.

3.5.1 Group (1): Scenarios I, II, & III

Group (1) is the most rudimentary analysis, because it assumes that the dark matter profile is NFW-like in the outer halo, and that the power-law profile of the tracers is known with certainty.

The cumulative mass profile estimates for Scenarios I, II, & III are presented in Figure 3.5, with the darkest regions representing the 50% Bayesian credible regions. The velocity anisotropy assumption varies from left to right in the figure: the first assumes an isotropic velocity dispersion, the second a constant anisotropy of 0.5, and the third has β as a free parameter.

Despite the variation in the anisotropy assumption between Scenarios I, II, and III, the mass profiles appear quite similar. The three estimates for the mass within 125kpc (M_{125}) are 4.28 , 4.38 , and $4.34 \times 10^{11} M_{\odot}$ and the 50% probability credible regions are $(3.69, 4.73)$, $(3.75, 4.86)$, and $(3.72, 4.81) \times 10^{11} M_{\odot}$ respectively. The mass estimate in Scenario II ($\beta = 0.5$) is only slightly higher

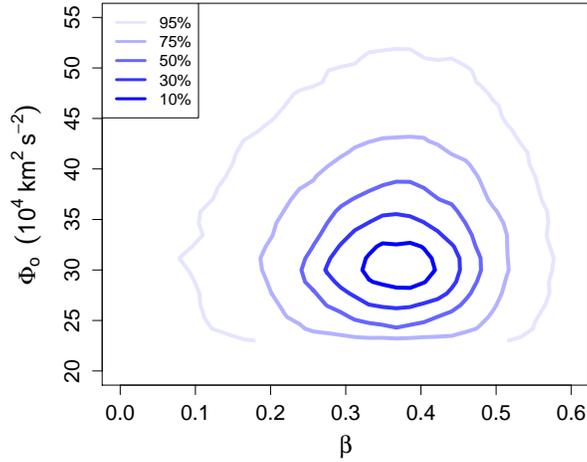


Figure 3.6: Posterior distribution for Scenario III. The contours show the 10, 30, 50, 75, and 95% joint credible regions for Φ_o and β when the parameters $alpha$ and γ are fixed.

than Scenarios in I ($\beta = 0$) and III (β free), but all estimates are in agreement within the 50% credible regions.

For comparison, we also use W10s mass estimator (their Equation (24)) with our complete data to compute a mass estimate, and then compare this result with our method. Using our data, $\alpha = 3.5$, $\gamma = 0.5$, and assuming an isotropic velocity assumption, W10s mass estimator returns $1.79 \times 10^{11} M_{\odot}$ for the mass within 38.5kpc (the position of the outermost GC with a proper motion measurement). Under an anisotropic assumption of $\beta = 0.5$, the W10 mass estimator gives $2.01 \times 10^{11} M_{\odot}$. In contrast, our cumulative mass profile and 50% credible regions for Scenarios I and II are $M(r < 38.5kpc) = 2.37 (2.05, 2.63) \times 10^{11} M_{\odot}$, and $2.43 (2.08, 2.70) \times 10^{11} M_{\odot}$. Therefore, the W10 mass estimator gives a slightly lower value using the complete data than our Bayesian method does using the complete and incomplete data together.

Figure 3.6 shows Scenario IIIs 10, 30, 50, 75, and 95% joint credible regions

for Φ_o and β ². There is no evidence for strong correlation between the two parameters, and the estimate of β suggests a slightly anisotropic velocity distribution. The 50% marginal credible interval for β is between 0.29 and 0.42, and the mean estimate for β is 0.35. Note that the 50% joint credible interval for β and Φ_o has a slightly wider range than the marginal credible interval for β alone.

3.5.2 Group (2): Scenarios IV, V, & VI

Scenarios IV - VI investigate the case in which γ is a free parameter, constrained to the lower and upper bounds given in Table 3.2. In principle, this Group is the one we should pay most attention to, because γ is the least constrained by observations.

Figure 3.7 shows the joint credible regions for Φ_o and γ in Scenario IV, after assessing for convergence. Even though γ was allowed to vary between 0 and 1 (a flat, uniform prior), very few samples are drawn from the region $\gamma > 0.3$. The parameters Φ_o and γ have a strong, highly non-linear correlation at low γ values, which results in larger estimates for the mass. The shape of the distribution is also reminiscent of that seen in Deason et al. (2012b). The mean value of Φ_o from the posterior distribution is 102.7, with a 50% credible region of $(64.2, 138.1) \times 10^4 \text{km}^2 \text{s}^{-2}$, and the mean estimate for γ is 0.06 (the median is 0.05), with 50% marginal distribution samples between 0.03 and 0.08. The latter estimate is quite surprising, considering that Deason et al. (2012a) found $\gamma \sim 0.35$ using the same model and BHB stars as tracers.

When the mean Φ_o and γ from the posterior distribution in Figure 3.7 are naively used as the best estimates for these parameters, then the cumulative

²Contours are drawn with the *emdbook* package in **R** (see Bolker, 2008, 2016)

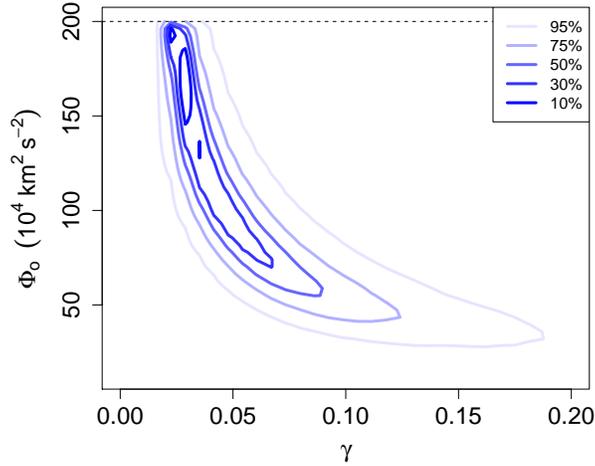


Figure 3.7: Scenario IV posterior distribution for Φ_o and γ . The contours show the Bayesian joint credible region. The dotted horizontal line is the upper bound of the uniform prior on Φ_o .

mass profile increases almost linearly with distance from the Galactic center, as it should for $\gamma \rightarrow 0$ (Equation 3.3). The mean mass estimate within 125kpc is also significantly larger than in Group (1), at a value of $1.09 \times 10^{12} M_\odot$, with a 50% credible region of $(0.92, 1.24) \times 10^{12} M_\odot$.

A numerical issue is that the posterior distribution of Φ_o and γ in Scenario IV may be multi-modal, and that the Markov chains may be stuck in a local maximum. Figure 3.7 shows more than one peak in the posterior distribution which may be real modes as opposed to numerical artifacts, (10^6 pairs of Φ_o and γ were drawn in Scenario IV). One could use a more complicated sampling method to try to explore other parts of the parameters space (e.g. the affine invariant sampler introduced by Goodman & Weare, 2010), but we view this as unnecessary since we have good reasons to put a narrower prior distribution on γ .

An isothermal profile for the dark matter halo ($\gamma \rightarrow 0$ in Equation 3.3) has been ruled out in the case of constant anisotropy (Battaglia et al., 2005). At

the other extreme, $\gamma \rightarrow 1$, $M(r)$ goes to a point mass which is unrealistic for a dark matter halo. As stated previously, $\gamma = 0.5$ is a good approximation to an outer NFW-type dark matter halo, for galaxies like the MW (Watkins et al., 2010; Deason et al., 2011). This is why we choose $\gamma = 0.5$ in Scenarios where γ is fixed. In the scenarios where γ is a free parameter, we relax the NFW approximation slightly and apply a uniform prior with lower and upper bounds of 0.3 and 0.7 respectively, calling these Scenarios IVb, Vb, and VIb. The range $\gamma = 0.3$ to 0.7 by itself covers a large range in dark matter halo central concentrations, while staying within arguably realistic bounds.

The above adjustment to the prior $p(\gamma)$ allows for slightly non-NFW type potentials while notably excluding some parameter ranges. Because of the relationship seen in Figure 3.7, the new prior on γ *will* change our mass estimates. However, choosing this slightly informative prior is important in the Bayesian paradigm—we apply prior distributions based on our assumptions and current knowledge about the situation, including external information not contained in the GC data themselves. At the same time, we present our results with the reminder that they are influenced by our assumptions, and thus are open to interpretation and criticism.

With the new prior on γ , the mean estimates for Scenario IVb are of $\Phi_o = 29.5$ (25.2, 33.1) $\times 10^4 \text{km}^2 \text{s}^{-2}$ and $\gamma = 0.318$ (0.305, 0.325). These estimates are in better agreement with Deason et al. (2012a), where the same model was applied to BHB stars using a maximum likelihood method rather than a Bayesian analysis. The mean estimates of Φ_o and γ in Scenarios Vb and VIb are very similar, but the shapes of the 95% joint posterior probability contours are rather different, shown in Figure 3.8. Notice that when β is a free parameter, the range of Φ_o is considerably smaller. Although it may

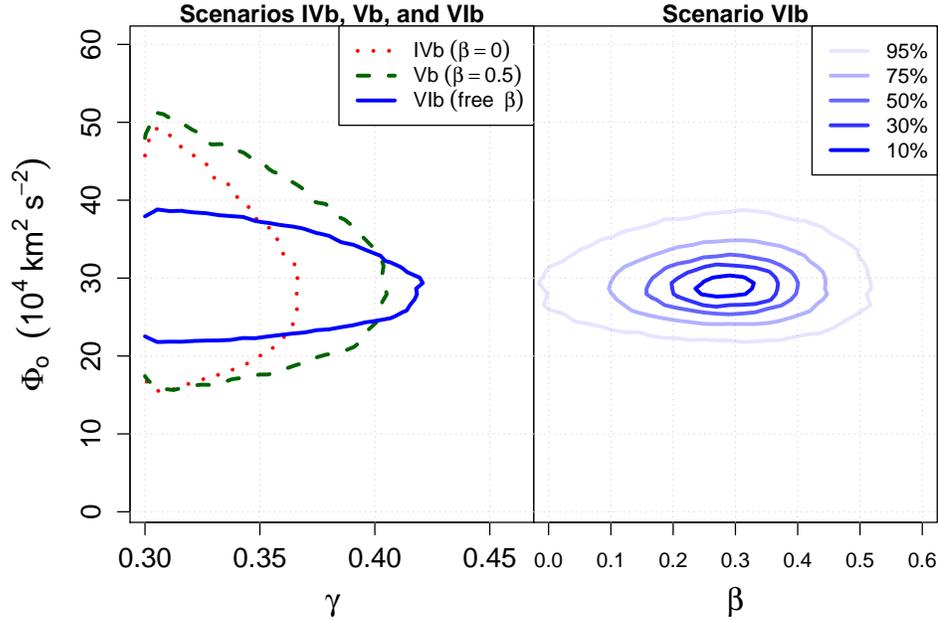


Figure 3.8: LEFT: The 95% joint credible regions for Group (2). The regions for the isotropic ($\beta = 0$), constant anisotropic ($\beta = 0.5$), and constant anisotropic (free parameter) are shown in dotted (red), dashed (green), and solid (blue) lines respectively. RIGHT: The posterior distribution and joint credible regions for Φ_o and γ in Scenario VIb.

initially seem strange that an extra free parameter would cause the posterior distribution to narrow, one explanation is that the posterior distribution of Φ_o and β is narrow as well (the right panel of Figure 3.8).

The M_{125} estimates for Scenarios IVb, Vb, and VIb are considerably higher than those in Scenarios I, II, and III, but the Bayesian credible intervals are also substantially larger (Figure 3.4 and Table 3.4). Likewise, the cumulative mass profile credible regions are also substantially wider (Figure 3.9). In the isotropic and anisotropic cases (Scenario IVb and Vb), the 50% credible regions for M_{125} are $5.00 - 6.61 \times 10^{11} M_{\odot}$ and $4.97 - 6.72 \times 10^{11} M_{\odot}$ respectively. The M_{125} estimate is most constrained in Scenario VIb (free β) with a 50% credible interval of $5.23 - 6.17 \times 10^{11} M_{\odot}$. This is attributed to the narrowed marginal distribution of Φ_o when β is a free parameter (Figure 3.8).

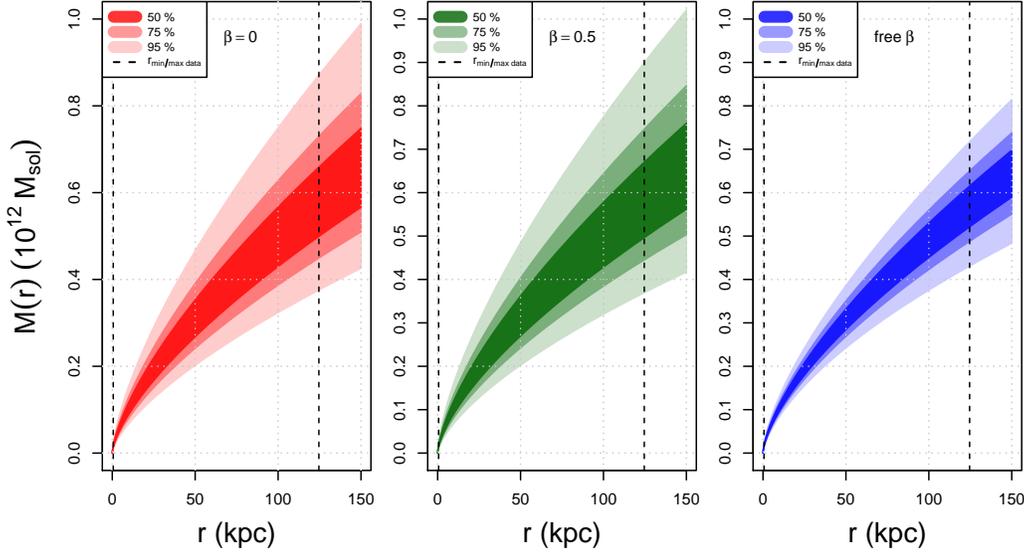


Figure 3.9: Credible regions for the cumulative mass profile in Scenarios IVb, Vb, and VIb. The percentages in the legend correspond to the three Bayesian credible regions shown, and the dotted lines show the extent of the GC data.

The differences in mass estimates between Group (1) and Group (2) are significant in terms of the 50% credible intervals. The 95% credible intervals for Scenarios IVb, Vb, and VIb contain the mass estimates obtained in Group (1), although VIb is a close call. The differences between Groups (1) and (2) must arise because γ is a free parameter. The estimate of γ is near the lower bound of the prior distribution $p(\gamma)$, which might indicate that the prior is too strong of an assumption. On the other hand, the uncertainty in the mass is very large when γ is free, suggesting that there is insufficient information in the GC data alone to pin down the shape of the dark matter halo.

3.5.3 Group (3): Scenarios VII, VIII, & IX

The cumulative mass profile credible regions from Group (3) (α free) are almost identical to those of Group (1) and so we do not bother showing them. Similar

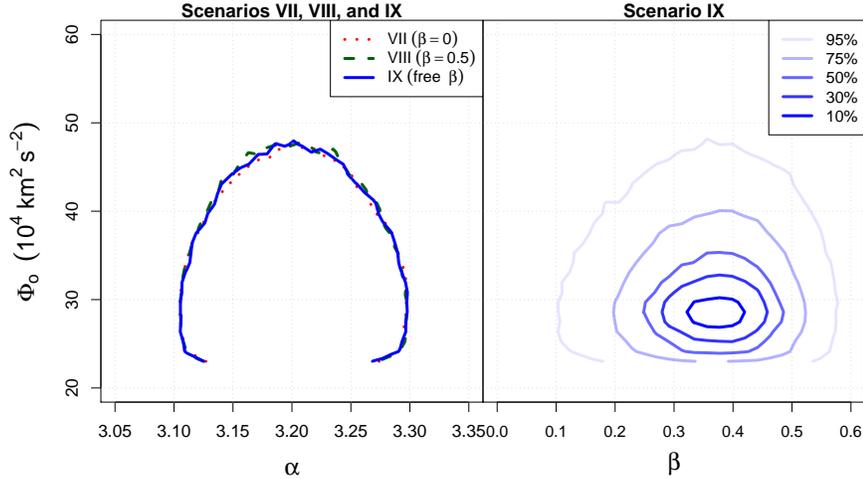


Figure 3.10: **Left:** The 95% joint credible region contours for Scenarios VII, VIII, and IX. The regions for the isotropic ($\beta = 0$), constant anisotropic ($\beta = 0.5$), and constant anisotropic (free parameter) are shown in dotted (red), dashed (green), and solid (blue) lines respectively. **Right:** The posterior distribution and 95% credible region for parameters Φ_o and γ in Scenario VIIb.

to the results in Section 5.1, the cumulative mass profiles between Scenarios VII, VIII, and IX are strikingly similar, regardless of the velocity anisotropy assumption. Overall, the estimate of M_{125} is approximately $4.1 \times 10^{11} M_\odot$, with 50% credible regions of about $(3.6, 4.5) \times 10^{11} M_\odot$. The M_{125} estimates are significantly lower than the estimates in Group (2), and only slightly lower than the estimates in Group (1).

The α estimates in Scenarios VII, VIII, and IX were all ~ 3.2 , with very narrow 50% marginal credible intervals (Table 3.4). This implies a slightly shallower tracer profile than we expected. It is interesting that the mass estimates are also lower in these Scenarios—Deason et al. (2012a) noticed the same behavior between α and the mass estimate.

One possible explanation for the lower α estimate is that the GCs in our subsample of Table 3.1 do not follow a power-law profile with index $\alpha \approx 3.5$. As a check, we calculate the spherical density profile for both the entire set

of GCs from Table 3.1 and the subset we used in our analysis (Figure 3.11). The density profile of the subsample appears to follow a power-law slope of 3.6 beyond $\sim 7\text{kpc}$ just as well as the entire data set. However, we can also see that a power-law slope of 1.9 follows the density profile quite well within $\sim 6\text{kpc}$. In our subsample, 42 of the 89 GCs fall within this inner region, so when α is free, the best model fit of a single power-law is a compromise of these two slopes. A broken power-law might be a better description of the GC profile in the MW. Alternatively, we could exclude data within 6.64kpc (the point of intersection of the two straight lines shown in Figure 11) at the expense of a smaller sample size of tracers.

To test the latter hypothesis, we run Scenario IX again but only use data for which $r > 6.64\text{kpc}$. We also re-define the prior distribution on α for consistency, using $b = 6.64\text{kpc}$ and only using the extra data beyond 6.64kpc . While the resulting estimate of M_{125} is larger ($5.06 \times 10^{11} M_{\odot}$, with 50% credible region $4.23\text{-}5.71 \times 10^{11} M_{\odot}$), the estimates for α and β are relatively unchanged (3.14 and 0.33 respectively). Therefore, excluding inner region objects leads to a slightly higher mass estimate despite an unchanged α estimate. We return to this point in more detail in Section 3.5.6.

3.5.4 Group (4): Scenarios X, XI, & XII

The mean M_{125} estimates for Group (4) are lower than those in Group (2), but higher than those in Groups (1) and (3) (Table 3.4 and Figure 3.4). However, Scenario X and XI's 95% credible regions for the mass estimate overlap with the mass estimates from all Groups (Figure 3.4).

Scenario XII deserves some attention, as all four model parameters are

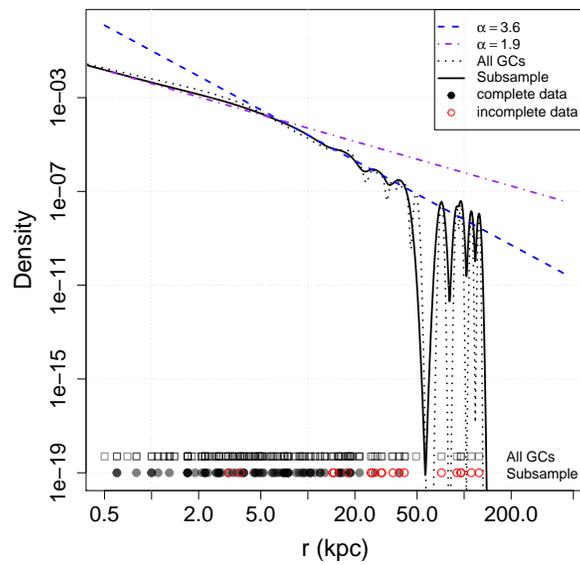


Figure 3.11: Smoothed radial density distribution of the GC subsample (solid line). Lines with power-law slopes 3.6 and 1.9 follow the outer and inner regions of GCs respectively. Points along the bottom of the graph indicate the r values of individual GCs; the top row shows the entire GC population from Table 3.1, and the bottom row is the subsample used in our analysis. The lack of GCs around $r = 50\text{kpc}$ results in the large drop in the density estimate in this region.

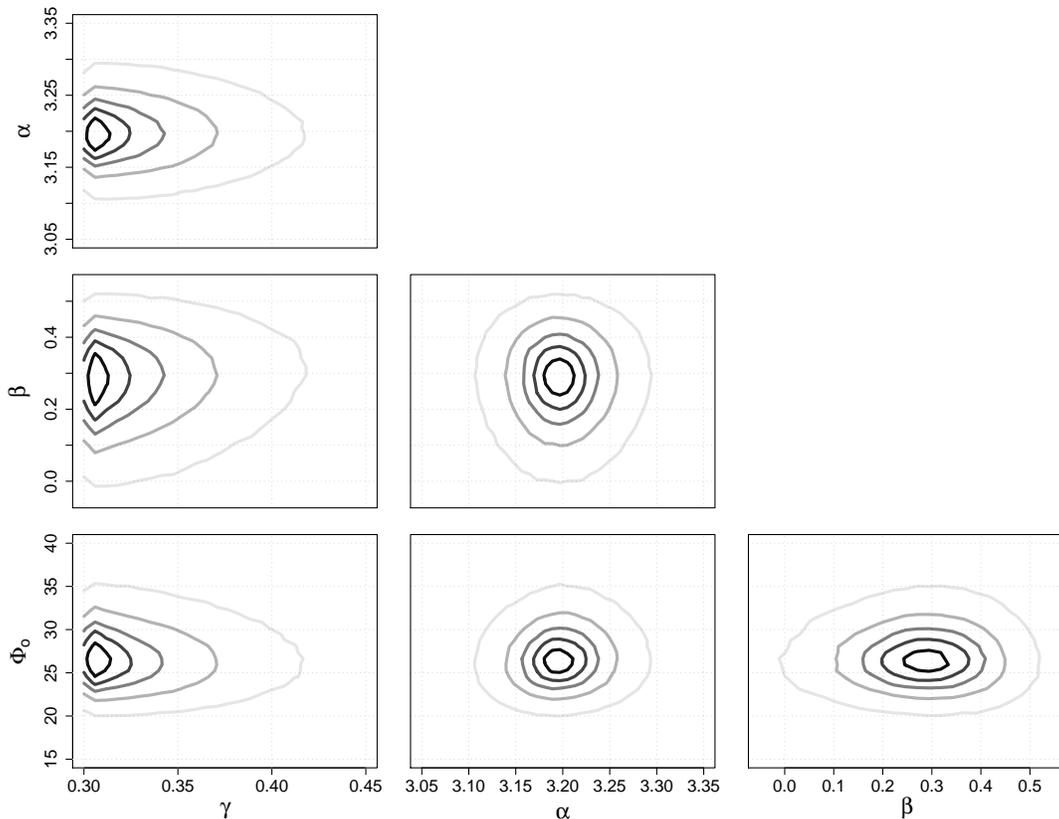


Figure 3.12: Posterior distributions for Scenario XII. The 10, 30, 50, 75, and 95% joint credible regions are shown as contours.

free. The joint posterior credible regions for all parameter combinations are shown in Figure 3.12, the parameter estimates are shown in Table 3.4, and the cumulative mass profile is on the left-hand side of Figure 3.13.

One notable feature in Figure 3.12 is in the marginal posterior distributions for γ (the leftmost column in the figure). The mode for γ is very close to the lower bound defined in the prior distribution, similar to the situation seen in Group (2) (Figure 3.8). The posterior distribution for γ also has an asymmetric shape, reminiscent of the shape seen in Figure 3.7. The parameter γ was poorly constrained when we used a less informative prior (Figure 3.7), and appears to be constrained in Group (2) and (4) mainly because of the more informative

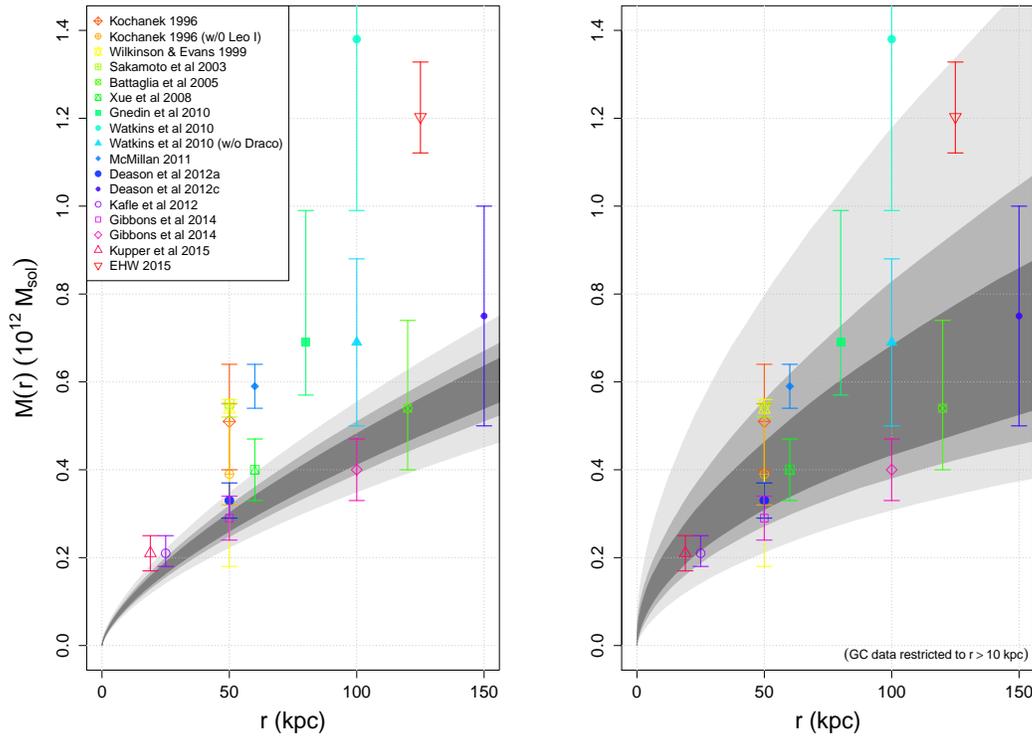


Figure 3.13: Comparisons of Scenario XII $M(R < r)$ profile credible regions to mass estimates from other studies (those not mentioned in the text are Kochanek, 1996; McMillan, 2011; Kafle et al., 2012; Eadie et al., 2015b; Küpper et al., 2015). The shades of grey (dark to light) correspond to the Bayesian credible regions (95, 75, and 50%). *Left*: Mass profile when all GC data are used in the analysis. *Right*: Mass profile when GC data at Galactocentric positions $r < 10\text{kpc}$ are excluded.

prior. Therefore, the present GC sample may not provide enough information about the dark matter halo to constrain its shape, without making relatively strong prior assumptions. The most we can say is that perhaps the potential is shallower than NFW.

The other notable feature about Scenario XII is that the 95% credible regions for M_{125} overlap the 50% credible regions from all the other Scenarios.

The points in Figure 3.13 are results from other studies, which are discussed below (Section 3.6). In general, however, it is clear that the results of Scenario XII are in agreement with studies that favor a “lighter” dark matter halo.

3.5.5 Summary of Groups (1) - (4)

Figure 3.4 shows the effect that parameter assumptions can have on mass estimates. In Group (1) — Scenarios I, II, and III — the only parameter that was allowed to vary in every case was Φ_o . The mass estimates are in very good agreement with one another, despite the different assumptions of velocity anisotropy.

In Group (2) (γ free) the mass estimates are consistently higher than Groups (1) and (3), but the uncertainties in these estimates are also substantially larger. Likewise, the credible regions for the $M(r)$ profiles are much larger (Figure 3.9). This may imply that it is difficult to constrain the steepness of the dark matter profile using the kinematic information of the tracers, most of which lie within 30kpc. When a strong assumption is made about the dark matter potential shape ($\gamma = 0.5$), the mass profile is more constrained but may be biased. When we relax the restriction on γ , the mass becomes more uncertain and relies more heavily on the prior distribution $p(\gamma)$, again implying that the current sample of GCs cannot constrain γ well on their own.

In Group (3) (α free), the mass estimates are lower than in any other group, but the results of Scenarios VII, VIII, and IX are in agreement with each other. We also observe that the tracer number density parameter α has a minor effect on the M_{125} estimate, despite its lack of appearance in Equation 3.3.

The mass estimates in Group (4) are slightly lower than those in Group (2), but higher than Groups (1) and (3). The uncertainties in Group (4) are similar to those seen in Group (2), where the β -free case has smaller credible regions. This is attributed to allowing β to be a free parameter, which narrows the allowed values of Φ_o via the likelihood, as shown in Section 3.5.2. In Groups

(3) and (4) (α free), the mass estimates are lower than in Groups (1) and (2) respectively. The lower estimates of α in comparison to a fixed $\alpha = 3.5$ (Table 3.4) may be causing the lower mass estimate, a relationship also noted by Deason et al. (2012a).

Table 3.3 shows the anisotropy parameter estimates for the Scenarios in which β was a free parameter (Scenarios III, VIb, IX, and XII). All estimates of β are in agreement with one another within the 95% credible intervals despite the mass estimates for these Scenarios differing in a large way. For a direct comparison of the anisotropy estimates with observations, we also estimate β directly from the complete data (71 GCs have both radial and tangential velocity measurements), and obtain a mean value $\langle\beta\rangle = 0.209$. This value is slightly but not dramatically smaller than the values in Table 3.3, and falls within our 95% credible regions of the β estimates.

Taking all considerations into account, we choose Scenario XII as our estimate for the MW dark matter halo. Using the posterior distribution from Scenario XII, and assuming a Hubble constant of $67.8\text{km s}^{-1}\text{Mpc}^{-1}$ (Planck Collaboration et al., 2015), we extrapolate out to the virial radius defined by $\rho_{200} = 200\rho_{crit}$. We find $r_{vir} = 185_{-7}^{+7}\text{kpc}$, and obtain a virial mass for the MW of $6.82 \times 10^{11} M_{\odot}$ with 50% credible region of $(6.06, 7.53) \times 10^{11} M_{\odot}$.

3.5.6 Sensitivity Test of the GC Sample

In this section, we run a simple sensitivity test of the relative importance of the inner versus outer GCs. Although the GCs cover a range of $0.6 < r < 125\text{kpc}$, many of the GCs in our sample are within $r = 30\text{kpc}$. Setting $\gamma = 0.5$ is akin to assuming an NFW potential beyond $\simeq 10\text{kpc}$. However, a single power-law

will not account for the inner DM halo which is presumably flatter than the outskirts. Furthermore, the inner GCs ($r < 10\text{kpc}$) are in a region where the bulge and disk are important contributors to the gravitational potential.

To investigate possible biases in our results, we perform a simple empirical test that has little reliance on a particular form (such as NFW) for the shape of the potential. We repeat the analysis for the whole suite of investigations (Table 3.3) using four different r cut-off values, r_{cut} , for the data. Mass estimates within 125kpc are obtained when GCs within 5, 10, 15, and 20kpc are excluded from the analysis respectively. We expect that as r_{cut} becomes larger, the uncertainties in the mass estimate will increase simply due to the lower number of data points. The question remains whether or not there is a significant trend toward a lower or higher mass as r_{cut} increases.

The number of data points decreases quickly as r_{cut} increases. For $r_{cut} = 5$ kpc, only 53 GCs remain in the sample (down from 89). For r_{cut} values of 10, 15 and 20 kpc, the number of GCs drops to 33, 24, and 16 respectively. Furthermore, as r_{cut} increases from 5 to 20kpc, the percentage of incomplete data increases from 30.2% to 81.2%, making the estimates of β in Scenarios III, VI, IX, and XII more uncertain.

Figure 3.14 shows the median mass estimates within 125kpc for all Scenarios listed in Table 3.3. Note that in all Groups, the Bayesian credible regions increase significantly as more data is excluded, as we expected. In Groups (1) and (3), when $\gamma = 0.5$, there is a noticeable dependence on the inner GC data that causes lower mass estimates; the mass estimates trend to higher values as more inner GCs are excluded from the analysis. This dependence, however, is not as strong when the 95% Bayesian credible regions are taken into account (the faint error bars in Figure 3.14).

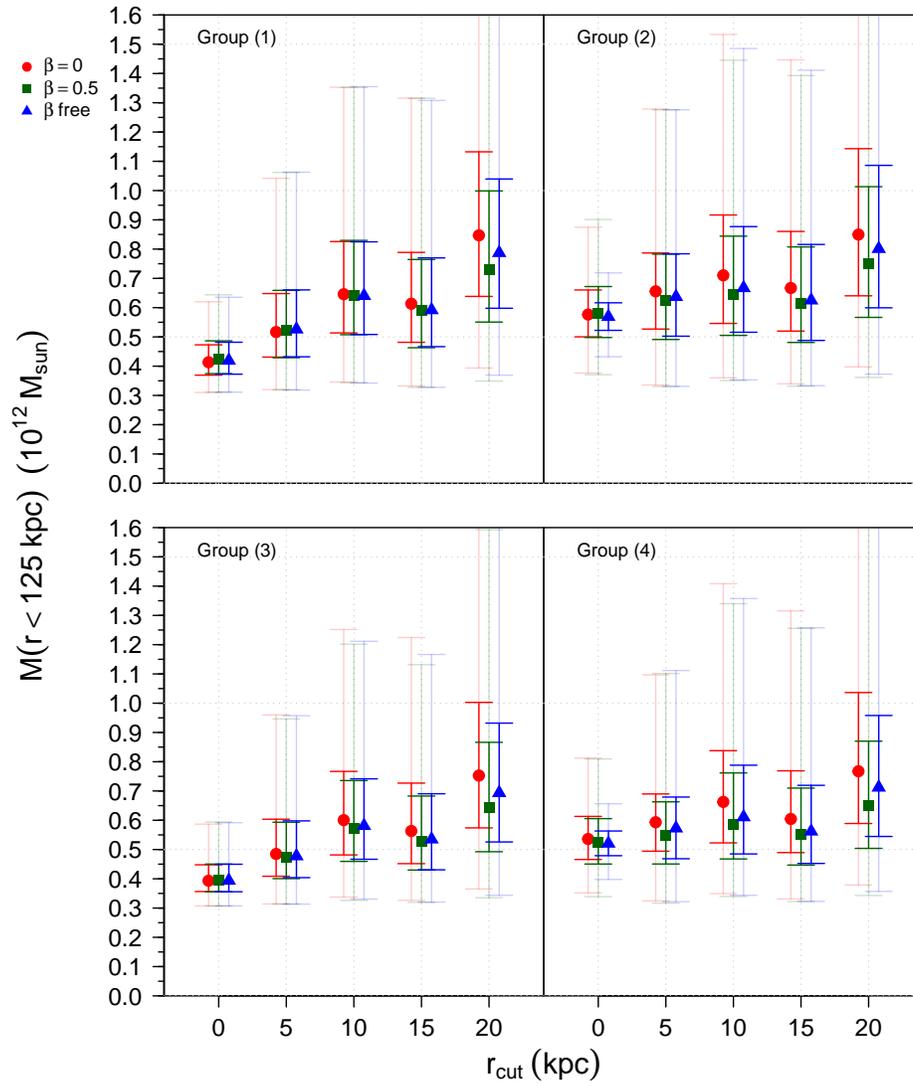


Figure 3.14: Results from sensitivity tests; dependency of the M_{125} estimates on the GC data sample. Each mass estimate was determined using GCs beyond r_{cut} . Bright and faint error bars correspond to the 50% and 95% credible regions respectively.

In contrast, for Groups (2) and (4), where γ is a free parameter, the trend of increasing mass estimates as r_{cut} increases is lessened significantly. Closer examination of Figure 3.14 reveals that the biggest mass differences between Groups occurs when all of the data is used in the analysis. On the other hand, there is little variation between the Groups' mass estimates when $r_{cut} > 10\text{kpc}$. Thus, the best choice may be to treat γ as a free parameter, and exclude GCs within 10kpc.

The estimates of β when it is a free parameter (i.e., Scenarios III, VI, IX, and XII) do not depend heavily on the r_{cut} values, despite the percentage of incomplete data increasing as r_{cut} increases. For example, the 50% credible regions for β in Scenario III, in increasing order of r_{cut} , are (0.28, 0.44), (0.21, 0.44), (0.34, 0.58), and (0.01, 0.49).

Another interesting feature of Figure 3.14 is that as r_{cut} increases, the difference in mass estimates between anisotropy scenarios ($\beta = 0, 0.5$ or free) becomes more pronounced. This can be attributed to the lowered percentage of complete data in the analysis as inner GCs are excluded. As the percentage of incomplete data becomes higher, any assumptions about β will have a larger impact on the mass estimate.

3.6 Discussion and Future Work

The mass estimate for the Galaxy in this work is significantly lower than the estimate from EHWs analysis (Figure 3.13). However, EHW used two populations of satellites (DGs and GCs) and assumed a self-consistent Hernquist model. When we apply the isotropic Hernquist model to the data used in this paper (GCs only), the results are in closest agreement with

Group (3); the isotropic Hernquist model gives a mass interior to 125kpc of $3.74 (3.64, 3.81) \times 10^{11} M_{\odot}$. Thus, the biggest difference results from dropping the DGs, which populate $r \gtrsim 50\text{kpc}$.

In general, our results are closely consistent with a number of other studies which suggest a “light” MW total mass. We take Scenario XII as our best estimate for the MW mass profile (Figure 3.13) and compare it to other studies. We consider the results when all GC data are used, and when only GCs at $r > 10\text{kpc}$ are included. It is clear from Figure 3.13 that the latter case results in better agreement with more studies, simply because of the increased range of the Bayesian credible regions at each radius.

We now discuss and compare our results from Scenario XII with a selection of the mass estimates shown in Figure 3.13. We also discuss some other studies whose results cannot be displayed in the figure.

Wilkinson & Evans (1999) used kinematic data of GCs and DGs, and a truncated flat-rotation curve (TF) model, to estimate the mass of the MW. Their result for $M(r < 50\text{kpc})$ and the uncertainties are shown in Figure 3.13. The lower bound of their uncertainty completely overlaps our mass estimate at 50kpc, despite the difference in model assumption. The point estimate is not in good agreement with ours, but this may be attributed to the different model choice.

Our results are in good agreement with Sakamoto et al. (2003), who used satellite galaxies, GCs, and field horizontal branch stars as kinematic tracers and found $M(r < 50\text{kpc}) = 5.5^{+0.0}_{-0.2} \times 10^{11} M_{\odot}$.

Battaglia et al. (2005) studied the radial velocity dispersion profile of 240 halo objects (satellite galaxies, GCs, and halo stars), and found that an NFW model predicts a virial mass of $0.8^{+1.2}_{-0.2} \times 10^{12} M_{\odot}$ with a concentration of $c = 18$.

The virial mass is an extrapolation beyond their furthest data point (120kpc), so they also quote their best mass estimate $M(r < 120\text{kpc}) = 5.4_{-1.4}^{+2.0} \times 10^{11} M_{\odot}$. Figure 3.13 shows that their result is in excellent agreement with our estimate at 120kpc.

Xue et al. (2008) compared the line-of-sight velocity distribution of 2401 BHB stars to cosmological galaxy simulations to constrain the mass of the MW. They found an enclosed mass within 60kpc of $4.0 \pm 0.7 \times 10^{11} M_{\odot}$, in agreement with our Scenario XII results (see Figure 3.13). However, when they assume an NFW halo profile the virial mass estimate is $1.0_{-0.2}^{+0.3} \times 10^{12} M_{\odot}$, which is not in agreement with M_{vir} presented here (see Section 3.5.4). This disagreement cannot be attributed to the dark matter profile parameter γ , since we found that a lower γ estimate leads to a higher mass estimate (Section 3.5.2), and we found $\gamma \approx 0.3$, rather than the NFW approximation of $\gamma = 0.5$.

Li & White (2008) report a virial mass of $2.43 \times 10^{12} M_{\odot}$, with a lower limit of $0.8 \times 10^{12} M_{\odot}$ at the 95% confidence level. This is at the higher end of the results in the literature and disagrees with our results.

As mentioned in Section 3.1, W10 calculated the MWs mass with two different mass estimators, assuming an NFW-type halo and using kinematic data from 26 DGs. Their mass estimates depended significantly on both the velocity anisotropy assumption for the tracers and the inclusion (or not) of proper motion measurements; they reported $M(r < 300\text{kpc})$ (M_{300}) values ranging from 7.0 to $14.0 \times 10^{11} M_{\odot}$. To compare, we extrapolate the Group (3) analyses (i.e., those which assume an NFW-type halo) out to M_{300} and find a range of 4.8 to $9.2 \times 10^{11} M_{\odot}$ for the 95% credible intervals, independent of the velocity anisotropy. Although our estimates were made using GCs rather than DGs, the 95% credible regions do overlap with the lower end of the mass

estimates from W10. If instead we compare our Scenario XII results extrapolated out to M_{300} , we obtain a 95% credible range of $(6.84, 11.93) \times 10^{11} M_{\odot}$, which is in much better agreement with W10's estimates. However, it should be noted that in Scenario XII, the posterior distribution for γ does not suggest an NFW-type potential, which *was* assumed by W10.

Using distant halo stars and a high-velocity star sample, and the spherical Jeans equation, Gnedin et al. (2010) found an enclosed mass for the Milky Way of $M(80\text{kpc}) = 6.9_{-1.2}^{+3.0} \times 10^{11} M_{\odot}$. This is in disagreement with Scenario XII's estimate of $M(80\text{kpc})$ (see Figure 3.13), but in better agreement with Scenarios IVb, Vb, and VIb (refer to Figure 3.8).

Busha et al. (2011) also used Bayesian inference to estimate the mass of the Milky Way, but instead incorporated Λ CDM-based simulations and the phase-space information of the Small and Large Magellanic Clouds. They arrived at a virial mass estimate of $1.2_{-0.4}^{+0.7} \times 10^{12} M_{\odot}$, where we quote only the statistical errors. Their virial radius was $250_{-30}^{+60} \text{kpc}$. Our Scenario XII estimates within 250kpc and 310kpc are in close agreement: $0.83 (0.61, 1.05) \times 10^{12} M_{\odot}$ and $0.96 (0.70, 1.22) \times 10^{12} M_{\odot}$.

Deason et al. (2012a) used the power-law model employed here in a maximum likelihood analysis of BHB star kinematics. When they assume spherical symmetry and set $\alpha = 3.5$, their γ , Φ_o , and β values were $0.35_{-0.17}^{+0.08}$, $30 \pm 5 \times 10^4 \text{km}^2 \text{s}^{-2}$, and $0.4_{+0.1}^{-0.2}$ respectively, which leads to mass estimate within 50kpc of $3.3 \pm 0.4 \times 10^{11} M_{\odot}$. In our equivalent set-up (Scenario VIb), we estimated $\gamma \approx 0.33$, $\Phi_o \approx 30$, $\beta \approx 0.27$, and a mass within 50kpc of $3.09 (2.43, 3.86) \times 10^{11} M_{\odot}$, which is in very good agreement despite the use of different tracer objects and a different method.

Using the TME from W10, Deason et al. (2012c) concluded that the mass

within 150kpc is $(5-10)\times 10^{11}M_{\odot}$. When we extrapolate the Scenario VIb mass profile out to 150kpc, we find the mass and 95% credible regions to be $6.44 \times 10^{11}M_{\odot}$, $(4.84, 8.14)\times 10^{11}M_{\odot}$, again in good agreement. In Scenario XII the estimate with 95% credible regions is $5.90 (4.45, 7.43) \times 10^{11}M_{\odot}$, which is also in good agreement with the TME. Again, this suggests a “light” Milky Way total mass.

The virial mass of the Milky Way was estimated by Boylan-Kolchin et al. (2013) to be $1.6_{-0.6}^{+1.0} \times 10^{12}M_{\odot}$, where the uncertainties represent 90% confidence intervals and the virial radius was ~ 300 kpc. The 95% credible regions for $M(r < 300$ kpc) from Scenario XII are in agreement at $(0.68, 1.19) \times 10^{12}M_{\odot}$.

Lastly, Gibbons et al. (2014) used a Bayesian method and kinematic data from the Sagittarius stream to measure the Galactic mass distribution, and reported the mass within 50 and 100 kpc: $M(50$ kpc) = $2.9 \pm 0.5 \times 10^{11}M_{\odot}$ and $M(100$ kpc) = $4.0 \pm 0.7 \times 10^{11}M_{\odot}$. The former is in agreement with our Scenario XII estimate for $M(50$ kpc) mentioned above, and is also in reasonable agreement with $M(100$ kpc) (see Figure 3.13). Their result is also in line with the aforementioned papers that support a lighter Milky Way Galaxy. Gibbons et al. (2014) also point out that their leaner mass estimate of the Milky Way helps to solve the “Too big to fail” problem.

In Scenario XII, where the GC subsample is limited to $r > 10$ kpc, the mass profile credible regions widen dramatically (right-hand side of Figure 3.13). Under this circumstance, our results agree with almost every value we have quoted from the literature.

There are some issues with the analysis presented here:

1. We assumed a spherically symmetric DM halo. However, the geometry

of the Milky Way's DM halo is not well known. Some studies favour a triaxial shape or oblate shape (Loebman et al., 2014; Deg & Widrow, 2013; Law & Majewski, 2010b), others a prolate shape (Bowden et al., 2016), and still others show that a spherical shape is not ruled out (Smith et al., 2009). The ideal way to allow for a non-spherical halo under the methodology of EHW would be to use a DF that includes an angular-dependent dark matter potential through extra model parameters. These parameters could then be estimated via the posterior distribution, and used to make inference about the geometry of the dark matter halo.

2. The model we used here assumes a constant anisotropy of the tracers, not an anisotropy that can vary with distance.
3. The assumption that tracers are randomly distributed about the Galaxy is probably incorrect at large r . Substructure in the distribution of Galactic satellites and halo stars arises in many hierarchical formation simulations of Milky Way-type galaxies, is becoming increasingly obvious in M31 (e.g. McConnachie et al., 2009; Ibata et al., 2007), and is without a doubt present in our own Galaxy (e.g. Yanny et al., 2000; Belokurov et al., 2006). However, recently Yencho et al. (2006) showed that the errors introduced by assuming a randomly distributed tracer population are actually quite small (at the 20% level) in comparison to the errors introduced by measurement uncertainties. This brings us to the next point.
4. The mass estimate under the power-law model presented here depends on the range of GCs used in the analysis. To fully understand this dependence, a better test of this simple power-law form ($\gamma = \text{constant}$) will

be to apply this analysis to simulated galaxies built from Λ CDM hydrodynamic simulations. This will be the subject of an upcoming paper (G. M. Eadie et al 2016, in preparation).

5. Measurement uncertainties have not been included in this analysis, but as EHW showed they are extremely important. In the third paper of this series, we will discuss how uncertainties can be included in the Bayesian paradigm; the interested reader may look at some preliminary tests of this method in Eadie et al. (2015a,c).
6. Finally, although our method does use incomplete and complete data at the same time, it relies on geometric arguments to incorporate the incomplete data (that is, the requirement that $|\cos\xi| > 0.95$). There are 157 GCs in Table 3.1, and we used only 89 of them ($\sim 56\%$). In our next paper, we will show how this problem is remedied through the use of a hierarchical model that includes the measurements uncertainties.

The Bayesian analysis performed here highlights the important influence of parameter assumptions and selecting prior probabilities. The results of Scenarios IV, V, and VI, as well as IVb, Vb, and VIb in Section 3.5.2 showed that the data cannot constrain the dark matter halo profile parameter γ very well, without prior information from other studies and knowledge gained from simulations. However, this highlights a strength of the Bayesian approach—we must mathematically and explicitly state our prior knowledge and assumptions. Many of these assumptions are incorporated into studies that use maximum likelihood and standard frequentist techniques too, but the assumptions may be more implicit.

3.7 Conclusion

We have performed a Bayesian analysis to determine the mass and cumulative mass profile of the MW out to 125kpc using GCs as tracers of the Galactic halo. The model and method we used is sensitive to the assumptions about model parameters, and most notably assumptions about the power-law profile of the dark matter halo potential. There also appears to be a dependence on the positions of the GCs, as excluding inner GCs leads to slightly higher mass estimates, especially when strong assumptions have been made about the dark matter profile.

One advantage of the EHW method is that we can easily obtain an estimate for the mass enclosed at any radius, and immediately obtain uncertainties in that estimate. This feature of the method makes comparing our results to other studies relatively straightforward (e.g., Figure 3.13). Another advantage is that using both complete and incomplete data simultaneously seems to remove any mass-anisotropy degeneracy. Furthermore, independent of parameter assumptions, the results suggest that the GC population has a mild radially anisotropic velocity distribution.

The first data release from the Gaia mission, including astrometric and photometric measurements, will occur in summer 2016 (ESA, 2016). Tycho-Gaia astrometric solution, which uses data from the Hipparcos catalog and the new measurements from GAIA, could yield proper motion, parallax, and position measurements for 2.5 million Tycho-2 stars (Michalik et al., 2015). The analytical approach described here will be well suited to this data.

Lastly, our analysis and comparison to other studies, and in particular Figure 3.13, strongly emphasizes the need for remote, virialized tracers ($r \gtrsim$

30kpc) with proper motion measurements to place stronger constraints on the slope of the dark matter halo profile, and ultimately the total mass of the MW. As more complete data at large Galactic distances become available, it will also be easier to rule out possible dark matter halo models via Bayesian model comparison tests.

Acknowledgements

GME and WEH thank the anonymous referee for thoughtful criticisms and constructive feedback on this paper. GME also thanks Aaron Springford for extremely helpful input about conjugate prior distributions and Gandhali Joshi for some helpful mathematical conversations. The authors also want to thank Alis Deason for the detailed and timely answers to our inquiries, and Dana Casetti for sharing the updated list of GC proper motions.

WEH and GME acknowledge the financial support of NSERC.

Bibliography

- Bahcall, J. N., & Tremaine, S. 1981, *The Astrophysical Journal*, 244, 805
- Battaglia, G., Helmi, A., Morrison, H., et al. 2005, *MNRAS*, 364, 433
- Bayes, M., & Price, M. 1763, *Philosophical Transactions (1683-1775)*, 370
- Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006, *ApJL*, 642, L137
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics*, 2nd edn. (Princeton)
- Bolker, B. 2016, *emdbook: Ecological Models and Data in R*, r package version 1.3.9
- Bolker, B. M. 2008, *Ecological Models and Data in R* (Princeton University Press)
- Bowden, A., Evans, N. W., & Williams, A. A. 2016, *MNRAS*, arXiv:1604.06885
- Boylan-Kolchin, M., Bullock, J. S., Sohn, S. T., Besla, G., & van der Marel, R. P. 2013, *ApJ*, 768, 140
- Busha, M. T., Marshall, P. J., Wechsler, R. H., Klypin, A., & Primack, J. 2011, *ApJ*, 743, 40
- Casetti-Dinescu, D. I., Girard, T. M., Jílková, L., et al. 2013, *AJ*, 146, 33

Casetti-Dinescu, D. I., Girard, T. M., Korchagin, V. I., van Altena, W. F., & López, C. E. 2010, *AJ*, 140, 1282

Cioni, M.-R. L., Bekki, K., Girardi, L., et al. 2016, *A&A*, 586, A77

Cuddeford, P. 1991, *MNRAS*, 253, 414

Cunningham, E. C., Deason, A. J., Guhathakurta, P., et al. 2016, *ApJ*, 820, 18

Deason, A. J., Belokurov, V., & Evans, N. W. 2011, *MNRAS*, 411, 1480

Deason, A. J., Belokurov, V., Evans, N. W., & An, J. 2012a, *MNRAS*, 424, L44

Deason, A. J., Belokurov, V., Evans, N. W., & McCarthy, I. G. 2012b, *ApJ*, 748, 2

Deason, A. J., Van der Marel, R. P., Guhathakurta, P., Sohn, S. T., & Brown, T. M. 2013, *ApJ*, 766, 24

Deason, A. J., Belokurov, V., Evans, N. W., et al. 2012c, *MNRAS*, 425, 2840

Deg, N., & Widrow, L. 2013, *MNRAS*, 428, 912

Dinescu, D. I., Girard, T. M., & van Altena, W. F. 1999, *AJ*, 117, 1792

Dinescu, D. I., Keeney, B. A., Majewski, S. R., & Girard, T. M. 2004, *AJ*, 128, 687

Dinescu, D. I., Martínez-Delgado, D., Girard, T. M., et al. 2005, *ApJL*, 631, L49

Djorgovski, S., & Meylan, G. 1994, *AJ*, 108, 1292

Eadie, G., Harris, W., & Springford, A. 2015a, in Joint Statistical Meetings Proceedings (American Statistical Association)

Eadie, G. M. 2013, Master's thesis, Queen's University, Kingston, Ontario, Canada

Eadie, G. M., Harris, W. E., & Widrow, L. M. 2015b, *ApJ*, 806, 54

Eadie, G. M., Harris, W. E., Widrow, L. M., & Springford, A. 2015c, in Proceedings of the International Astronomical Union, ed. A. Bragaglia, M. Arnaboldi, M. Rejkuba, & D. Romano (Cambridge Journals)

ESA. 2016, Data Release Scenario, <http://www.cosmos.esa.int/web/gaia/release>, accessed: 2016-05-10

Evans, N. W., Hafner, R. M., & de Zeeuw, P. T. 1997, *MNRAS*, 286, 315

Evans, N. W., Wilkinson, M. I., Perrett, K. M., & Bridges, T. M. 2003, *The Astrophysical Journal*, 583, 752

Feigelson, E., & Babu, G. 2012, *Modern Statistical Methods for Astronomy: With R Applications* (Cambridge University Press)

Feltzing, S., & Johnson, R. A. 2002, *A&A*, 385, 67

Fritz, T. K., & Kallivayalil, N. 2015, *ApJ*, 811, 123

Gelman, A., Carlin, J., Stern, H., & Rubin, D. 2003, *Bayesian data analysis* (Chapman & Hall/CRC)

Gelman, A., & Rubin, D. 1992, *Statistical science*, 7, 457

- Geman, S., & Geman, D. 1984, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 6, 721
- Gibbons, S. L. J., Belokurov, V., & Evans, N. W. 2014, MNRAS, 445, 3788
- Gnedin, O. Y., Brown, W. R., Geller, M. J., & Kenyon, S. J. 2010, ApJL, 720, L108
- Goodman, J., & Weare, J. 2010, Communications in applied mathematics and computational science, 5, 65
- Harris, W. E. 1976, AJ, 81, 1095
- . 1996, AJ, 112, 1487
- . 2001, Globular Cluster Systems, Vol. 28 (Springer-Verlag Berlin Heidelberg), 223
- . 2010 edition, ArXiv e-prints, arXiv:1012.3224
- Howlander, H. A., Hossain, A. M., & Makhmin, O. 2007, International Journal of Statistical Sciences, 6, 137
- Huang, Y., Liu, X., Yuan, H., et al. 2016, ArXiv e-prints, arXiv:1604.01216
- Humphreys, R. M., & Larsen, J. A. 1995, AJ, 110, 2183
- Ibata, R., Martin, N. F., Irwin, M., et al. 2007, ApJ, 671, 1591
- Johnson, D. R. H., & Soderblom, D. R. 1987, The Astronomical Journal, 93, 864
- Kaffe, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, ApJ, 761, 98

—. 2014, *ApJ*, 794, 59

Kochanek, C. S. 1996, *The Astrophysical Journal*, 457, 228

Kulesa, A. S., & Lynden-Bell, D. 1992, *MNRAS*, 255, 105

Küpper, A. H. W., Balbinot, E., Bonaca, A., et al. 2015, *ApJ*, 803, 80

Law, D. R., & Majewski, S. R. 2010a, *ApJ*, 718, 1128

—. 2010b, *ApJ*, 714, 229

Li, Y.-S., & White, S. D. M. 2008, *MNRAS*, 384, 1459

Little, B., & Tremaine, S. 1987, *The Astrophysical Journal*, 320, 493

Loebman, S. R., Ivezić, Ž., Quinn, T. R., et al. 2014, *ApJ*, 794, 151

Majewski, S. R., & Cudworth, K. M. 1993, *PASP*, 105, 987

Massari, D., Bellini, A., Ferraro, F. R., et al. 2013, *ApJ*, 779, 81

McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2009, *Nature*, 461, 66

McMillan, P. J. 2011, *MNRAS*, 414, 2446

Merritt, D., Graham, A. W., Moore, B., Diemand, J., & Terzić, B. 2006, *The Astronomical Journal*, 132, 2685

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *The journal of chemical physics*, 21, 1087

Metropolis, N., & Ulam, S. 1949, *Journal of the American Statistical Association*, 44, 335

- Michalik, D., Lindegren, L., & Hobbs, D. 2015, *A&A*, 574, A115
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, ArXiv e-prints, arXiv:1502.01589
- Rossi, L. J., Ortolani, S., Barbuy, B., Bica, E., & Bonfanti, A. 2015, *MNRAS*, 450, 3270
- Sakamoto, T., Chiba, M., & Beers, T. C. 2003, *A&A*, 397, 899
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, 403, 1829
- Smith, M. C., Wyn Evans, N., & An, J. H. 2009, *ApJ*, 698, 1110
- Wang, J., Chen, L., & Chen, D. 2005, *Chinese Astronomy and Astrophysics*, 29, 386
- Wang, W., Han, J., Cooper, A., et al. 2015, *MNRAS*, 453, 377
- Watkins, L., Evans, N., & An, J. 2010, *MNRAS*, 406, 264
- Widrow, L. M. 2000, *ApJ Supp*, 131, 39
- Wilkinson, M., & Evans, N. 1999, *MNRAS*, 310
- Williams, A. A., & Evans, N. W. 2015, *MNRAS*, 454, 698
- Xue, X., Rix, H., Zhao, G., et al. 2008, *ApJ*, 684, 1143
- Yanny, B., Newberg, H. J., Kent, S., et al. 2000, *ApJ*, 540, 825
- Yencho, B. M., Johnston, K. V., Bullock, J. S., & Rhode, K. L. 2006, *ApJ*, 643, 154
- Zoccali, M., Renzini, A., Ortolani, S., Bica, E., & Barbuy, B. 2001, *AJ*, 121, 2638

4

Bayesian Mass Estimates of the Milky Way: Including Measurement Uncertainties with Heirarchical Bayes

Reprinted from *Eadie, G., Springford, A., & Harris, W. (2017), THE ASTROPHYSICAL JOURNAL, Vol. 835, No. 167.* DOI: 10.3847/1538-4357/835/2/167.

An erratum associated with this paper is included in Appendix A. *Published by the American Astronomical Society ©AAS*

Abstract

We present a hierarchical Bayesian method for estimating the total mass and mass profile of the Milky Way Galaxy. The new hierarchical Bayesian approach further improves the framework presented by Eadie et al. (2015b); Eadie & Harris (2016) and builds upon the preliminary reports by Eadie et al. (2015a,c). The method uses a distribution function $f(\mathcal{E}, L)$ to model the galaxy and kinematic data from satellite objects such as globular clusters (GCs) to trace the Galaxy's gravitational potential. A major advantage of the method is that it not only includes complete and incomplete data simultaneously in the analysis, but also incorporates measurement uncertainties in a coherent and meaningful way. We first test the hierarchical Bayesian framework, which includes measurement uncertainties, using the same data and power-law model assumed in Eadie & Harris (2016), and find the results are similar but more strongly constrained. Next, we take advantage of the new statistical framework and incorporate all possible GC data, finding a cumulative mass profile with Bayesian credible regions. This profile implies a mass within 125kpc of $4.8 \times 10^{11} M_{\odot}$ with a 95% Bayesian credible region of $(4.0, 5.8) \times 10^{11} M_{\odot}$. Our results also provide estimates of the true specific energies of all the GCs. By comparing these estimated energies to the measured energies of GCs with complete velocity measurements, we observe that (the few) remote tracers with complete measurements may play a large role in determining a total mass estimate of the Galaxy. Thus, our study stresses the need for more remote tracers with complete velocity measurements.

4.1 Introduction

In our two previous papers, Eadie et al. (2015b, hereafter Paper I) and Eadie & Harris (2016, hereafter Paper II), we estimated the Galaxy's mass and mass profile using a new Bayesian method and the kinematic data of Milky Way globular clusters (GCs) and dwarf galaxies (DGs). Paper I laid the groundwork: we tested the method on simulated data and then applied the method to Milky Way satellite data in a preliminary analysis. A main advantage of the new Bayesian method was that both complete and incomplete velocity vectors were included in the analysis simultaneously. Furthermore, the tests on simulated data showed that our Galactic mass estimates were insensitive to incorrect velocity anisotropy assumptions. Paper I incorporated an analytic Hernquist model (for simplicity and testing of the method), and used GCs and DGs as tracers of the Milky Way's potential. The satellites were assumed to follow the same spatial distribution as the dark matter. Despite the simplicity of the model, the results were in agreement with many other studies (see Wang et al., 2015, for a comparison figure).

The promising results of Paper I led us to implement an arguably more realistic model for the Milky Way in Paper II, in which the distributions of the dark matter and the Galactic tracers are allowed to differ. The Paper II model uses power-law profiles with different parameters for the dark matter and tracers, and also includes velocity anisotropy as a parameter. This model is explained in detail by Evans et al. (1997), and previous applications of the model to the Milky Way and other galaxies were completed by Deason et al. (2011, 2012a,b, note that the notations vary between Evans' and Deason's papers).

Because the model includes a spatial profile for only a single population of tracers, we used GC kinematic data alone instead of a mixture of DGs and GCs. The results in Paper II suggested a mass estimate for the Milky Way that was significantly lower than the mass found in Paper I under the Hernquist model, but closer in agreement to recent studies which suggest a “light” Milky Way (e.g. Gibbons et al., 2014).

An issue that is not fully addressed in Paper I or II is the inclusion of measurement uncertainty. Measurement uncertainties can differ substantially from object to object, with some tracers having very precise radial velocities or proper motions and others having very imprecise ones.

Using a sensitivity analysis, we found in Paper I that measurement uncertainties can play a significant role in the mass estimate of the Galaxy, contributing up to 50% of the uncertainty in the estimate. In addition, we found that certain individual objects had very high leverage. For example, when the single GC Palomar 3 was removed from the analysis, the mass estimate of the Galaxy decreased by more than 12%. Thus, it seems prudent to include measurement uncertainties in a rigorous and consistent way when estimating the mass and mass profile of the Galaxy.

Here, we substantially improve upon Paper II by introducing a *hierarchical* Bayesian method that includes the measurement uncertainties of proper motions and line-of-sight velocities in a measurement model. Preliminary tests of this method have been reported by Eadie, Harris, Widrow, & Springford (2015c) and Eadie, Harris, & Springford (2015a) using the Hernquist model and data from GCs and DGs in Paper I, but here we apply the arguably more realistic tracer model from Paper II, and also use all of the available GC data.

4.2 Method

In Papers I and II, we defined the posterior distribution from Bayes' theorem as $p(\boldsymbol{\theta}|\mathbf{y})$, where $\boldsymbol{\theta}$ is the vector of model parameters, and \mathbf{y} is the vector of data. In practice, the posterior distribution is difficult to calculate directly, and Markov Chain Monte Carlo (MCMC) methods are used to sample a distribution that is proportional to the posterior distribution. We write this distribution as

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \prod_i^n p(y_i|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (4.1)$$

$$= \prod_i^n p((r_i, v_{r,i}, v_{t,i})|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (4.2)$$

Above, r_i , $v_{r,i}$, and $v_{t,i}$ represent the Galactocentric distance, radial velocity, and tangential velocity of the i^{th} tracer (GC). We assume that the GC positions and velocities are independent of one another, conditional on the value of $\boldsymbol{\theta}$.

In Paper II, we defined $p((r_i, v_{r,i}, v_{t,i})|\boldsymbol{\theta})$ by the *distribution function* (DF). The model for the dark matter halo's gravitational potential follows a power-law profile of $\Phi(r) = \Phi_o r^{-\gamma}$, and the spatial number density profile of the tracers follows $\rho(r) \propto r^{-\alpha}$. Using the Eddington formula as described in Binney & Tremaine (2008), the DF is found to be

$$f(\mathcal{E}, L) = \frac{L^{-2\beta} \mathcal{E}^{\frac{\beta(\gamma-2)}{\gamma} + \frac{\alpha}{\gamma} - \frac{3}{2}} \Gamma\left(\frac{\alpha}{\gamma} - \frac{2\beta}{\gamma} + 1\right)}{\sqrt{8\pi^3} 2^{-2\beta} \Phi_o^{-\frac{2\beta}{\gamma} + \frac{\alpha}{\gamma}} \Gamma\left(\frac{\beta(\gamma-2)}{\gamma} + \frac{\alpha}{\gamma} - \frac{1}{2}\right)} \quad (4.3)$$

where $\mathcal{E} = -v^2/2 + \Phi(r)$, $L = rv_t$, and the model parameters are $\boldsymbol{\theta} = (\Phi_o, \gamma, \alpha, \beta)$ (beware of notational differences between Evans et al. (1997) and

Deason et al. (2011, 2012a,b)). The parameter β is the standard anisotropy parameter, where the limits $\beta = 1$ and $\beta \rightarrow -\infty$ correspond to completely radial or completely tangential orbital distributions for the tracers (Binney & Tremaine, 2008).

The DF in Equation 4.3 assumes a spherical and non-rotating system, and also requires that the relative energy \mathcal{E} is greater than zero (i.e. that tracers are bound to the Galaxy). Under this model, the mass profile of the dark matter halo is,

$$M(r) = \frac{\gamma\Phi_o}{G} \left(\frac{r}{\text{kpc}} \right)^{1-\gamma} \quad (4.4)$$

(Deason et al., 2012b), which has the physical limits of an isothermal sphere ($\gamma \rightarrow 0$) and a central point mass ($\gamma \rightarrow 1$).

Equation 4.3 is written in the Galactocentric reference frame—the frame in which the geometry of the model is the most straightforward. The GC kinematic data and their uncertainties, on the other hand, are measured in the Heliocentric reference frame. Although the mathematical transformation of velocity and position vectors from a Heliocentric frame to a Galactocentric frame is relatively straightforward, transforming *uncertainties* from one frame to the other requires complex error propagation which is non-linear, and that likely results in non-Gaussian errors. Therefore, we employ a different approach to incorporating the measurement uncertainties using a hierarchical Bayesian model.

4.2.1 Hierarchical Bayesian Model

In Paper II, all stochasticity in $\{r_i, v_{r,i}, v_{t,i}\}$ was due to Equation 4.3, and none was due to measurement uncertainty. The measured values of $\{r_i, v_{r,i}, v_{t,i}\}$

were assumed to be the true values, which we conditioned upon to obtain the posterior distribution for θ , the model parameters.

Now, we include a model for measurement uncertainty. The approach starts with a slight change in perspective: instead of treating the measurements of the position r , line-of-sight velocity v_{los} , and proper motions in right ascension (R.A.) ($\mu_\alpha \cos \delta$) and declination (decl.) (μ_δ) as the true values, we treat these measurements as *samples* drawn from a distribution which depends on the true (but unknown) values. That is, the true values are now included as parameters in the model. These parameters, the true Galactocentric position and Heliocentric velocity components, are denoted with a \dagger symbol:

$$\boldsymbol{\vartheta} = \left(r^\dagger, v_{los}^\dagger, \mu_\delta^\dagger, \mu_\alpha \cos \delta^\dagger \right), \quad (4.5)$$

For a given GC, the *measurements* are denoted as

$$\boldsymbol{y} = (r, v_{los}, \mu_\delta, \mu_\alpha \cos \delta), \quad (4.6)$$

and the measurement *uncertainties* are denoted:

$$\boldsymbol{\Delta} = (\Delta r, \Delta v_{los}, \Delta \mu_\delta, \Delta \mu_\alpha \cos \delta). \quad (4.7)$$

We assume that the measurements are samples drawn from Gaussian (normal) distributions centered on $\boldsymbol{\vartheta}$, and the measurement uncertainties $\boldsymbol{\Delta}$ are taken to be standard deviations. For example, the measurement of the line-of-sight velocity is drawn from a normal distribution centered on the true line-of-sight velocity, with a standard deviation equal to the measurement uncertainty. In statistical terms, this is akin to saying that V_{los} is a *random*

variable normally distributed with mean v_{los}^\dagger and variance Δv_{los}^2 :

$$V_{los} \sim \mathcal{N}(v_{los}^\dagger, \Delta v_{los}^2) \quad (4.8)$$

(where $\mathcal{N}(\mu, \sigma^2)$ represents the Normal distribution). With this assumption, the probability of obtaining a *measurement* v_{los} is

$$p(V_{los} = v_{los} | v_{los}^\dagger, \Delta v_{los}^2) = \frac{1}{\sqrt{2\pi\Delta v_{los}^2}} e^{-\frac{(v_{los} - v_{los}^\dagger)^2}{2\Delta v_{los}^2}}. \quad (4.9)$$

The same Gaussian assumption is made for the probabilities of the other measurements $p(\mu_\delta | \mu_\delta^\dagger, \Delta\mu_\delta)$, $p(\mu_\alpha \cos \delta | \mu_\alpha \cos \delta^\dagger, \Delta\mu_\alpha \cos \delta)$, and $p(r | r^\dagger, \Delta r)$. We assume that measurement errors are independent given the true values, so that the probability of measuring all components of a GC's kinematic quantities is simply the product of the probabilities defined above. Thus, the total likelihood is

$$\begin{aligned} \mathcal{L}(\mathbf{y} | \mathbf{\Delta}, \vartheta) &= p(r | r^\dagger, \Delta r) p(v_{los} | v_{los}^\dagger, \Delta v_{los}^2) \times \\ &\quad p(\mu_\delta | \mu_\delta^\dagger, \Delta\mu_\delta) p(\mu_\alpha \cos \delta | \mu_\alpha \cos \delta^\dagger, \Delta\mu_\alpha \cos \delta) \end{aligned} \quad (4.10)$$

(Eadie et al., 2015a,c). This defines our measurement model. We acknowledge that the two components of the proper motion measurements are not actually independent. Their correlation could be incorporated using a multivariate normal, but because these correlations are not usually reported, we do not pursue it here.

Equipped with an expression for the likelihood (Eq. 4.10), we next define prior distributions on the parameters. The prior distributions on ϑ link the

measurement model to the tracer/galactic mass model. The parameters $\boldsymbol{\vartheta}$ represent the true positions and velocities, and we assume that these parameters have a prior distribution defined by Equation 4.3, i.e. the DF. Thus, the DF is the prior distribution on $\boldsymbol{\vartheta}$, and is denoted in shorthand as $p(h(\boldsymbol{\vartheta})|\boldsymbol{\theta})$, where h is the transformation from Heliocentric to Galactocentric coordinates (Section 4.2.3).

Because the DF (the prior distribution on $\boldsymbol{\vartheta}$) has its own parameters $\boldsymbol{\theta}$, then hyperpriors $p(\boldsymbol{\theta})$ must also be specified. Thus, for a single GC or tracer, Bayes's rule is written as

$$p(\boldsymbol{\theta}|\mathbf{y}_i, \boldsymbol{\Delta}_i) \propto \mathcal{L}(\mathbf{y}_i|\boldsymbol{\Delta}_i, \boldsymbol{\vartheta}_i) \times p(h(\boldsymbol{\vartheta}_i)|\boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \quad (4.11)$$

$$\propto \text{Likelihood} \times \text{Prior} \times \text{Hyperprior} \quad (4.12)$$

Assuming that the GCs are conditionally independent, the posterior distribution is proportional to

$$p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\Delta}) \propto \prod_{i=1}^N \mathcal{L}(\mathbf{y}_i|\boldsymbol{\vartheta}_i, \boldsymbol{\Delta}_i) p(h(\boldsymbol{\vartheta}_i)|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (4.13)$$

The hierarchical Bayesian specification above provides a couple of improvements to Papers I and II (Equation 4.1). First and foremost, measurement uncertainties are included in the analysis in a meaningful way. Second, whereas before only 89 of 157 GCs could be included¹, we can now include 143 GCs. The remaining 14 GCs are objects for which no measurements of velocity are available (see Table 4 in Paper II).

¹Mainly due to the GCs' locations, see Papers I & II

4.2.2 Defining Priors and Hyperpriors

Defining priors in the Bayesian paradigm is an opportunity for the researcher to state prior knowledge, gained from previous studies, and prior assumptions about model parameters. For this study, we use the same prior distributions for the model parameters Φ_o , γ , and β that were used previously: uniform distributions with bounds given and justified in Paper II. These lower and upper bounds for Φ_o , γ , and β are (1, 200), (0.3, 0.7), and (-0.5, 1.0) respectively.

The prior on the GC spatial distribution parameter, $p(\alpha)$, is a gamma distribution. This choice was established and justified in Paper II and was defined using the extra 68 GCs not otherwise included in the analysis. In this study, however, most of these previously excluded GCs can now be included because we do not have to depend on geometric assumptions to approximate v_{los} . There remain 14 GCs that are excluded in the data sample because they have only position measurements (see Section 4.2.1). We use these 14 GC positions to estimate and define a prior distribution for the parameter α , in the same way as in Paper II. Figure 4.1 compares the new $p(\alpha)$ to that used in Paper II. Note that the new prior is wider than the one used in Paper II, because fewer GCs were used to estimate and define it. Including the extra GC data in the prior is akin to including the positions of all GCs in the analysis.

In summary, there are two sets of parameters in the hierarchical model: (1) the position and velocity parameters $\boldsymbol{\vartheta}$ and (2) the DF parameters $\boldsymbol{\theta}$. Bayes' theorem and the rules of conditional probability require a prior probability for both sets of parameters. The prior probability distribution for $\boldsymbol{\vartheta}$ is Equation 4.3 (the DF), because we assume that the positions and velocities come from the set of models determined by $\boldsymbol{\theta}$. The prior distributions on $\boldsymbol{\theta}$ are the

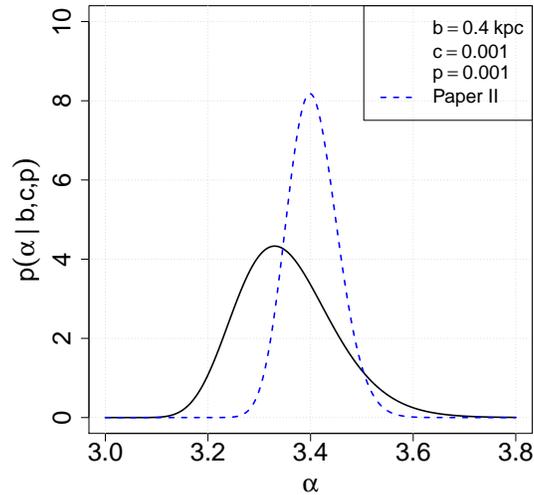


Figure 4.1: The solid black line is the prior distribution used in this paper, and the blue dashed line was the prior distribution used in Paper II. The solid-line prior probability distribution for α is determined using the extra GC data ($n = 14$) that is not used in the rest of the analysis. Thus, the prior used in this study is less informative than that used in Paper II.

hyperprior distributions $p(\boldsymbol{\theta})$ described above and in Paper II.

4.2.3 Transformation of Velocities

In this section we discuss the function $h(\boldsymbol{\vartheta})$ first mentioned in Section 4.2.1. The $h(\boldsymbol{\vartheta})$ notation symbolizes the transformation of velocity parameters in Eq. 4.5 from a Heliocentric parameterization $(v_{los}^\dagger, \mu_\delta^\dagger, \mu_\alpha \cos \delta^\dagger)$ to a Galactocentric parameterization $(v_r^\dagger, v_t^\dagger)$, following the method presented in Johnson & Soderblom (1987). We review the Johnson & Soderblom (1987) method here for completeness and in order to highlight some important points.

The first step is to transform the Heliocentric velocities into Galactic space-velocities (U, V, W) in a right-handed coordinate system:

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \mathbf{T} \cdot \mathbf{A} \begin{bmatrix} v_{los} \\ k\mu_{\alpha}\cos(\delta)/\lambda \\ k\mu_{\delta}/\lambda \end{bmatrix} + \begin{bmatrix} U_{\odot} \\ V_{\odot} \\ W_{\odot} \end{bmatrix} \quad (4.14)$$

where U is positive toward the Galactic center, V is positive in the direction of Galactic rotation, and W is positive above the Galactic plane. The solar motion is set to $(U_{\odot}, V_{\odot}, W_{\odot}) = (11.1, 12.24, 7.25)$ (Schönrich et al., 2010), $k=4.74057$ (the equivalent in km s^{-1} of 1 AU in one tropical year), and λ is the parallax (in arcsec) of the GC (Johnson & Soderblom, 1987). The matrices T and A depend on the R.A. and decl. of the North Galactic Pole (as determined by the Hipparcos catalog) and GCs, respectively, where

$$\mathbf{T} = \begin{bmatrix} -0.0548755604 & -0.8734370902 & -0.4838350155 \\ +0.4941094279 & -0.4448296300 & +0.7469822445 \\ -0.8676661490 & -0.1980763734 & +0.4559837762 \end{bmatrix} \quad (4.15)$$

(ESA, 1997) and where \mathbf{A} for a single GC is

$$\mathbf{A} = \begin{bmatrix} +\cos \alpha \cos \delta & -\sin \alpha & -\cos \alpha \sin \delta \\ +\sin \alpha \cos \delta & +\cos \alpha & -\sin \alpha \sin \delta \\ +\sin \delta & 0 & +\cos \delta \end{bmatrix}. \quad (4.16)$$

Above, α and δ are the R.A. and decl., respectively, in decimal degrees (this α is of course different from the one used to parameterize the GC distribution above). We take the GCs' parallax and R.A. and decl. positions as fixed but

treat the Galactocentric distance r as a parameter in the model and assign an uncertainty of 5% to the measured r value (Harris, 1996, 2010 edition).

The next step is to transform the Cartesian, rotating Galactic frame velocity components (U, V, W) into components in a cylindrical, non-rotating Galactocentric reference frame (Π, Θ, Z) . First the adjustment for the rotation of the Galaxy at $R_\odot = 8.0\text{kpc}$ is taken to be 220km s^{-1} , to obtain (U_{gc}, V_{gc}, W_{gc}) , and then this vector is transformed to a non-rotating, right-handed cylindrical system via

$$\begin{bmatrix} \Pi \\ \Theta \\ Z \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U_{gc} \\ V_{gc} \\ W_{gc} \end{bmatrix}. \quad (4.17)$$

As a test of the entire transformation, we compare our derived (Π, Θ, W) to the Casseti online catalog of GC velocity measurements (Figure 4.2) (Dinescu et al., 1999, 2004, 2005; Casseti-Dinescu et al., 2010, 2013)².

Finally, the velocity components in Eq. 4.17 are transformed to the spherical coordinate system

$$\begin{bmatrix} v_r \\ v_\theta \\ v_\phi \end{bmatrix} = \begin{bmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{bmatrix} \begin{bmatrix} \Pi \\ \Theta \\ W \end{bmatrix} \quad (4.18)$$

where $v_t^2 = v_\theta^2 + v_\phi^2$. To reiterate, the complete transformation from the Galactocentric parameterization to the Heliocentric parameterization described above is represented by $h(\boldsymbol{\vartheta})$ in Equation 4.3.

²Updated catalog: www.astro.yale.edu/dana/gc.html

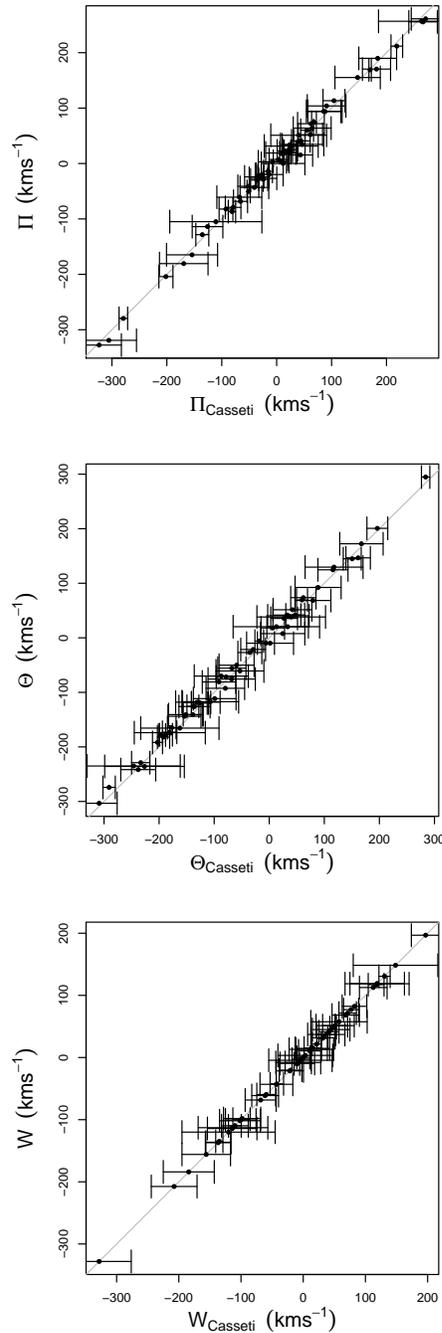


Figure 4.2: Our transformation from Heliocentric velocities to Galactocentric velocities (Θ, Π, W) for GCs with proper motions, compared to the Casseti catalog. The abscissa are from the Casseti online catalog, and the ordinates are our transformation values. The error bars are the uncertainties given in the Casseti catalog, and the grey line has a slope of one.

4.2.4 Improved Computational Methods

The posterior distribution is sampled using the same general MCMC method and hybrid-Gibbs sampler that was utilized in Papers I and II. One improvement is that the proposal distributions for ϑ and θ are determined using the adaptive MCMC method described by Roberts & Rosenthal (2009). A multivariate covariance matrix is determined for each GC's ϑ parameters and for θ , via an iterative burn-in process. The advantage of the adaptive MCMC method is that the target posterior distribution is sampled much more efficiently by taking into account correlations between parameters. A second minor change is that we now run seven independent parallel chains instead of three, thereby obtaining the same number of samples in less than half the time.

4.3 Kinematic Data

The kinematic data used in this study are presented in Table 1 of Paper II. In Paper II, only 89 out of 157 GCs were used in the analysis, mainly because the approximation $|v_{los}| \approx |v_r|$ did not hold for most GCs without proper motions. Other GCs were excluded in the analysis of Paper II, due to high reddening association with the Sagittarius dwarf galaxy, or to no velocity measurements.

As described in Sections 4.2 and 4.3, using the hierarchical Bayesian framework allows all of the incomplete data to be included without having to make any geometric arguments like those used in Papers I and II, because the likelihood \mathcal{L} is written in the Heliocentric frame. Now that we are accounting for uncertainties, we also include the GCs subject to high reddening. In the

present analysis, the GCs associated with the Sagittarius dwarf do not change the result significantly, and therefore we choose to include them. Altogether, this increases the size of the data set significantly, from 89 to 143 GCs.

4.4 Analysis Overview

To make a fair comparison between the nonhierarchical method of Paper II and the hierarchical method presented here and to thereby directly test the influence of measurement uncertainties, we first apply the hierarchical Bayesian method to the same kinematic data analyzed in Paper II (i.e., only 89 GCs). In this case, we use the prior distribution $p(\alpha)$ for the tracer spatial parameter that was used in Paper II (the dashed blue line in Figure 4.1).

Next, we use the hierarchical Bayesian method with 143 GCs, using prior distribution $p(\alpha)$ defined by the extra 14 GCs without velocity measurements (solid line in Figure 4.1).

4.5 Results

Figure 4.3 compares the 95% Bayesian credible regions for the mass profiles of the Milky Way from Paper II (the dashed black lines) to the 50%, 75%, and 95% regions from the present paper (the blue shaded regions). Both results rely on the same 89 GC sample used in Paper II; the only difference between the two analyses is that measurement uncertainties are now included. The main result of including measurement uncertainties via the hierarchical method is a stronger constraint on the mass profile and mass estimate compared to that from the method used in Paper II.

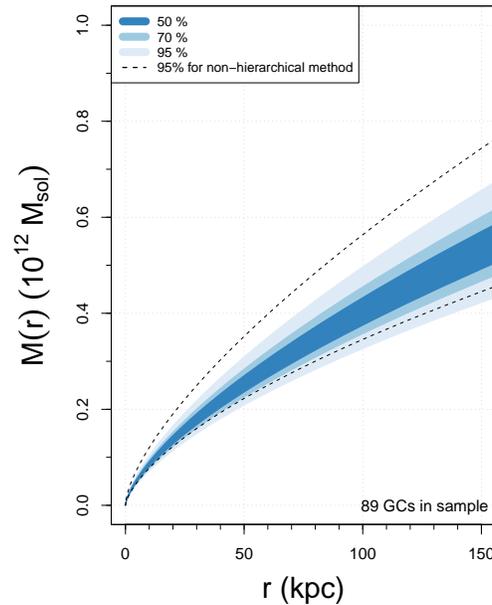


Figure 4.3: The blue shaded areas are the Bayesian credible regions for the cumulative mass profile of the Milky Way, using the hierarchical method and 89 GCs. The black dashed lines show the 95% credible regions for the non-hierarchical method and 89 GCs (i.e. the results from Paper II).

As mentioned in Section 4.2.1 and 4.3, one advantage of the hierarchical model is that the GC sample size is increased from 89 to 143. Figure 4.4 compares the estimated mass profile using 89 GCs to the profile using 143 GCs. The dashed blue lines indicate the 95% Bayesian credible regions from Figure 4.3, and the black shaded regions indicate the credible regions when 143 GCs are included. The increase in total sample size likely accounts for the slight narrowing of the Bayesian credible regions. However, the difference between the hierarchical results from the 89 GC sample and those from the 143 GC sample is not as large as might be expected.

We conjecture that a more constrained estimate using the present method will require a higher proportion of complete data. Although the sample size increases by more than 60%, the GC data that are added to the sample are

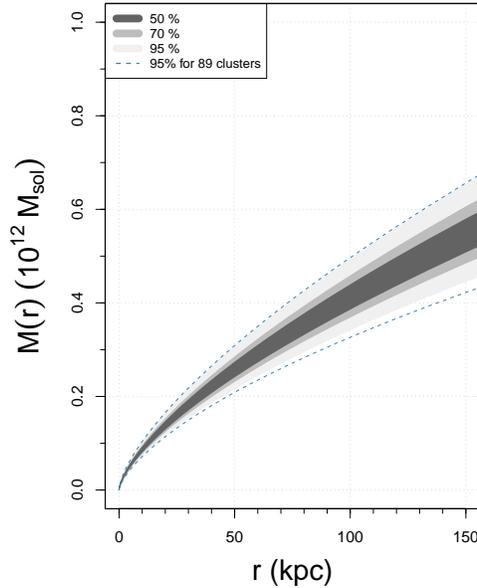


Figure 4.4: The grey shaded areas are the Bayesian credible regions for the cumulative mass profile of the Milky Way using the hierarchical method and 143 GCs. The blue dashed lines show the 95% credible regions for the hierarchical method and 89 GCs (i.e. the outermost blue region in Figure 4.3).

all incomplete. Of the 89 GCs used in Paper II, 71 had complete velocity data. Thus, with 143 GCs the proportion of GCs with complete velocity data decreased to about 50%.

GCs are subject to the total gravitational potential within their orbits and thus trace the Galaxy’s total mass out to 125kpc (the distance of the farthest GC in our sample). Using the hierarchical Bayesian method presented here, the power-law models, and the priors, and confronting this coherent model with data from 143 GCs returns a total mass within 125kpc of $4.8 \times 10^{11} M_{\odot}$, with a 95% credible region of $(4.0, 5.8) \times 10^{11} M_{\odot}$.

Extrapolating our mass profile in Figure 4.4 to a virial radius that corresponds to 200 times the critical density of the universe, assuming $H_o = 67.8 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al., 2016), we find that $r_{200} =$

179(164, 194) kpc and $M(r_{200} = 6.2 (4.7, 7.8) \times 10^{11} M_{\odot}$ (the numbers in brackets correspond to the 95% Bayesian credible regions). Extrapolating further, we find that the mass within 300 kpc is $M(300 \text{ kpc}) = 0.9 (0.7, 1.1) \times 10^{12} M_{\odot}$.

In Paper II, we performed a sensitivity analysis to determine how the spatial sample of GCs might affect the mass estimate of the Milky Way under our assumed power-law model. The sensitivity analysis involved obtaining mass estimates after removing GCs with positions within five different r_{cut} values: 0, 5, 10, 15, and 20 kpc. Here we repeat the sensitivity analysis using the same set of r_{cut} values but using the full sample of 143 GCs. The sensitivity analysis implicitly examines how disk- and bulge-associated GCs might affect the mass estimate, because when $r_{cut} = 10$ kpc, almost all (93/97) of the excluded GCs have $|z| < 5$ kpc. Figures 4.5 and 4.6 display how the mass and individual model parameters Φ_o , γ , α , and β vary in the sensitivity analysis.

In contrast to Paper II, we find that the mass estimate within 125 kpc is robust to the systematic exclusion of inner GCs, except perhaps when only GCs beyond 20 kpc are used in the analysis (Figure 4.5). We note, however, that the sample size beyond 20 kpc is small (19 GCs), and only 4 of these GCs have proper motion measurements. Accordingly, the uncertainty in the mass increases significantly in this case, and the 95% credible regions overlap with mass estimates under smaller r_{cut} values.

The results suggest that the current model is adequate for describing the profile of GCs, at least with regard to estimating the total mass within 125 kpc. If the current tracer model was not able to describe the data, then we might expect changes in the α estimate and the mass estimate, as inner tracers were systematically removed. However, we see no evidence of this occurring within the 95% credible regions of the posterior distributions for α (Figure 4.6), and

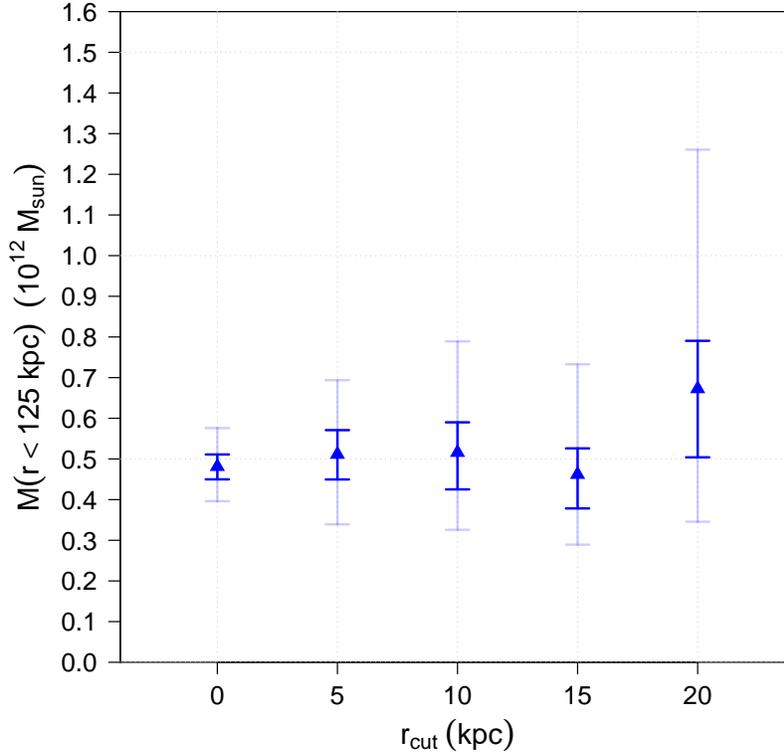


Figure 4.5: Sensitivity analysis: the estimate of the total mass within 125 kpc (M_{125}) for different GC samples. GCs within the r_{cut} value are removed from the sample (for example, M_{125} at $r_{cut} = 10$ corresponds to the mass estimate when GCs within 10 kpc are removed from the sample). Inner bars are 50% credible regions, and outer bars are 95% credible regions.

there is little evidence that changes in α affect the mass estimate (Figure 4.5). The power-law slope of the GC population is highly constrained in the analysis, regardless of the GC sample that is used. One thing to note is that the prior $p(\alpha)$ becomes less and less informative for each r_{cut} , as the extra data available to define a prior change from 14 GCs to 6, 5, 3, and 3 GCs.

A positive correlation in the estimates of Φ_o and γ is immediately obvious in the upper two panels of Figure 4.6, and as more inner GCs are excluded (i.e., as r_{cut} increases), $\gamma \rightarrow 0.5$. This value of γ corresponds to an approximate Navarro et al. (1996) profile at large radii (Deason et al., 2011), albeit with

very large uncertainty. The significant change in γ and in its uncertainty in the sensitivity analysis suggests that the shape of the dark matter profile cannot be well constrained using only the outermost GCs. To constrain the shape with more confidence, all of the data must be used. The single power-law profile for the gravitational potential does not take into account the Galaxy's bulge and disk components. However, despite the relatively simplistic model for the gravitational potential and the changes in γ during the sensitivity analysis, the mass estimate is robust.

The β estimates in the sensitivity analysis are in good agreement with one another, despite the percentage of complete data decreasing as r_{cut} increases. We can therefore conclude that the GC population has a mildly radial constant anisotropy under this model assumption. However, when the GC sample is limited to clusters outside 20 kpc, the uncertainty in β becomes quite large.

To summarize the entire posterior distribution for the full sample of 143 GCs, we also show the joint credible regions for all four model parameters (Figure 4.7).

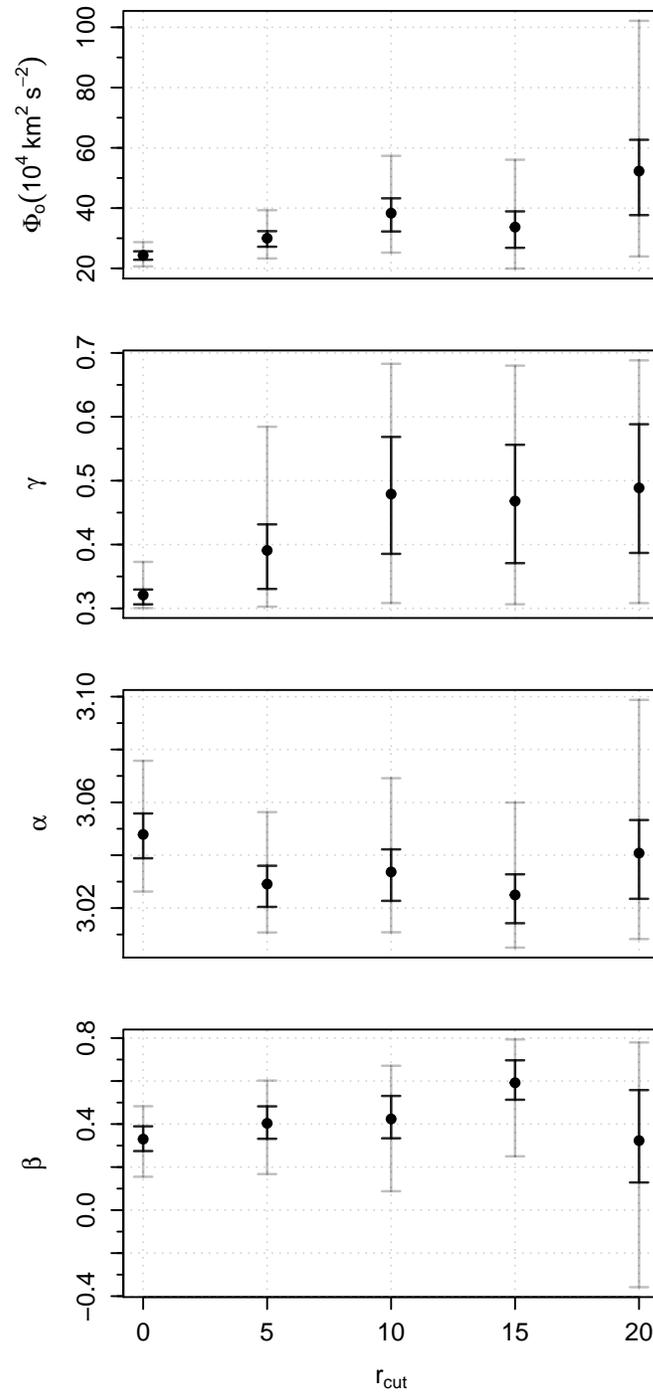


Figure 4.6: Sensitivity analysis: Parameter estimates when GCs within the r_{cut} value are removed from the sample. Inner bars are 50% credible regions and outer bars are 95% credible regions.

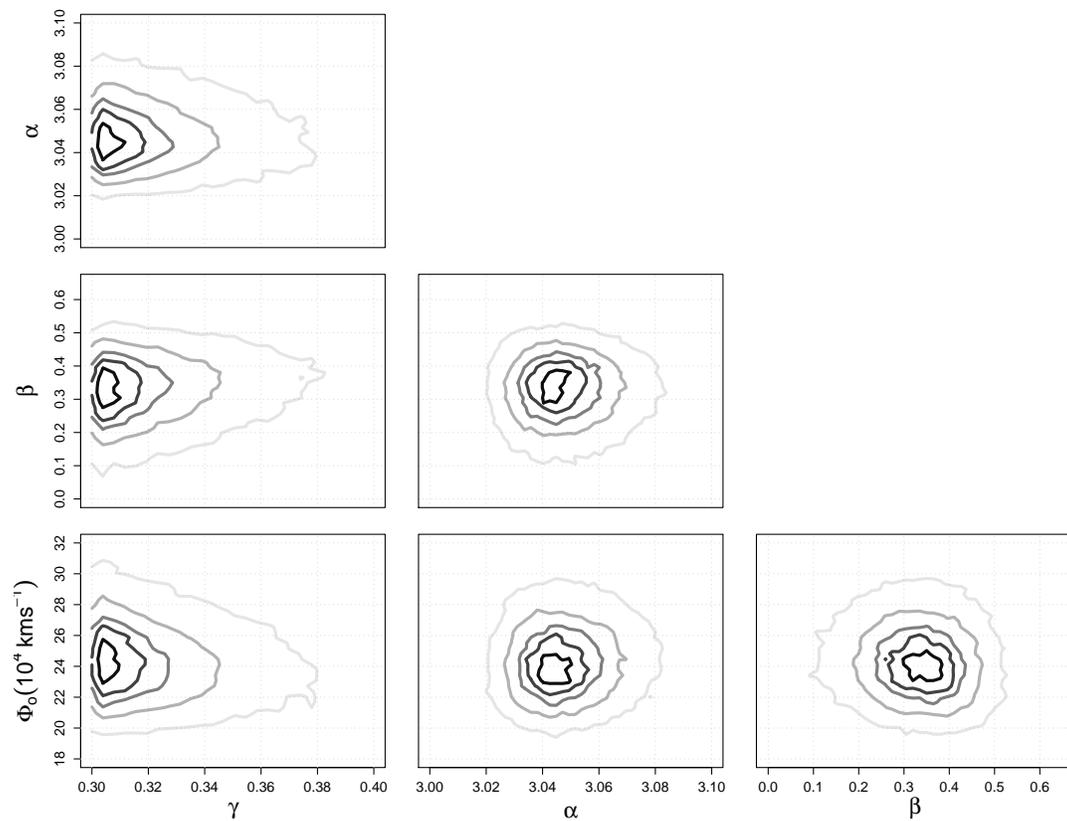


Figure 4.7: Joint posterior distributions for the model parameters. The curves show the 10, 20, 30, 50, 75, and 95% Bayesian credible regions.

Because the hierarchical method treats the Heliocentric distances and velocities as parameters in the model, the final posterior distribution provides estimates and credible regions for the parameters $\boldsymbol{\vartheta} = (r^\dagger, v_{los}^\dagger, \mu_\delta^\dagger, \mu_\alpha \cos \delta^\dagger)$ (i.e., there are 572 parameters in the GC measurement model alone). Using the posterior distributions for these parameters, we derive an estimate of the specific energy E for each GC, with credible regions. Figure 4.8 shows these energy estimates as a function of Galactocentric position: the hollow and solid blue circles are the mean energy estimates of the incomplete and complete data parameters, respectively. The solid green diamonds are the energies derived from the measurements of the complete data (there are no hollow green diamonds because energies cannot be derived without proper motions). Arrows from the solid green diamonds to the solid blue points connect the same GC. For legibility, we do not show the 95% credible regions for the energies, but we have checked that they are reasonable. The purple shaded curves enclose the 50% and 95% credible regions for the gravitational potential, determined pointwise as a function of r .

Figure 4.8 provides a consistency check of the hierarchical method in three ways: (1) the distribution of points is consistent with our initial assumptions that all GCs are bound to the Galaxy, (2) the incomplete and complete data energy distributions populate the region between the gravitational potential and the zero line, and (3) the incomplete and complete data do not appear to have different energy distributions. Another feature of note is the tendency for the estimated energies based on positions and velocities to shrink toward a curve similar in shape to the $\Phi(r)$ profile. The reason is that the posterior distributions for each tracer’s energy are in some sense a compromise between the prior implied by the tracer model (Equation 4.3) and the measured value.

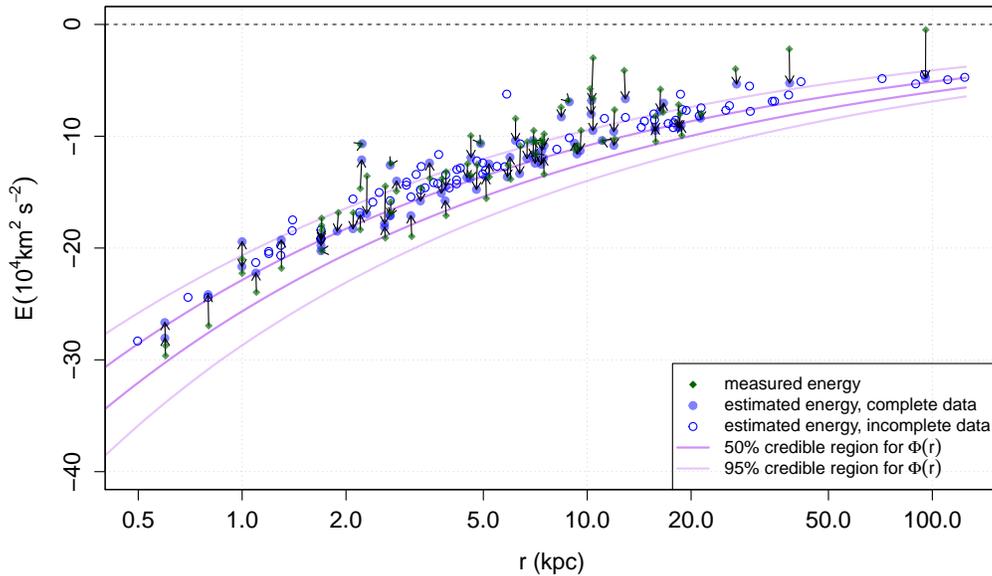


Figure 4.8: The posterior means of the GCs' specific energies as a function of Galactocentric position. The blue circles are the mean energy estimates for each GC; solid points are complete data and hollow points are incomplete data. The solid green diamonds are the measured energies for complete data. The arrows connect GC measured energies to GC posterior mean energies, and the purple shaded curves represent the 50 and 95% credible regions for the potential.

Whether the posterior distribution is closer to the measured value or to the value implied by the tracer model is a function of the width of the prior compared to the measurement uncertainty of the tracer.

4.6 Discussion

The Bayesian method presented here has an advantage over traditional point mass estimation techniques in the literature because it uses complete and incomplete data simultaneously in the analysis, whereas other techniques use either complete or incomplete data only (e.g. the mass estimators introduced by Bahcall & Tremaine, 1981; Evans et al., 2003; Watkins et al., 2010). Furthermore, although other studies have used a Bayesian analysis to infer the

mass of the Milky Way (e.g. Little & Tremaine, 1987; Kulessa & Lynden-Bell, 1992; Kochanek, 1996; Wilkinson & Evans, 1999; McMillan, 2011; Kafle et al., 2012; Williams & Evans, 2015; Küpper et al., 2015), to our knowledge none of these studies has included measurement uncertainties using a coherent measurement model as we have done here.

Including the measurement uncertainties in a measurement model introduced four additional parameters for every GC, which increased the computational cost of the analysis. Nonetheless, even with 576 parameters (572 measurement model parameters $\boldsymbol{\vartheta}$ and 4 tracer model parameters $\boldsymbol{\theta}$), we were able to run these analyses overnight on a personal computer with four cores after sufficient Markov chain burn-in.

We found that including uncertainties in the analysis resulted in a tighter constraint on the cumulative mass profile of the Milky Way compared to ignoring measurement uncertainties (Figure 4.3). This somewhat paradoxical result might be explained by attributing some of the variation in GC kinematics to the measurement process, as described in Figure 4.8. Without allowing for measurement error, the tracer model is made to explain all of the variation, which apparently results in increased overall uncertainty in the mass profile.

When the sample size of GCs went from 89 to 143, neither the value nor the spread of the mass profile changed substantially (Figure 4.4). Introducing additional data might be expected to decrease the width of the Bayesian credible regions, but this was not observed. We suspect that the credible region width did not change because including additional incomplete data decreased the proportion of complete measurements. When 143 GCs were included in the analysis, nearly 50% of the data were incomplete, in contrast to almost 80% of the data being complete when 89 GCs were used. We therefore stress the

importance of having complete and accurate motion data for tracer objects. In particular, there is a need for remote tracers with complete measurements. This point is highlighted by both the sensitivity analysis (Figures 4.5 and 4.6 and 6) and the energy estimates of the GCs (Figure 4.8).

Figures 4.5 and 4.6 display how the uncertainty in the mass and parameter estimates changes as inner GCs are removed from the sample; as the percentage of incomplete data increases, the results become much less constrained.

Figure 4.8 shows that the outermost GCs with complete data have estimated energies that are lower than their measured counterparts. The complete data E estimates (solid blue circles) appear to move away from the measurement values (solid green diamonds) and toward the E estimates of the incomplete data (hollow blue circles). However, there is very little information beyond 20 kpc, because the proportion of GCs at large distances without proper motion measurements is high. If complete velocity measurements of these remote GCs suggest that they have high energies, then the mass estimate obtained with this model will increase. If they do not, then the mass estimate of the Galaxy may be closer to the value we found in this study. Ultimately, this question cannot be answered without measuring the proper motions of the other remote GCs.

The results of the sensitivity analysis, the estimated energy profile of the GCs, and the relatively unchanged result between 89 and 143 clusters lead us to conclude that it is absolutely critical to have proper motions for distant tracers. Obtaining proper motions of GCs at large radii is critical to understanding the distribution of energies at large radii and thus the mass of the Milky Way.

An illuminating follow-up investigation to this study is to analyze simulations of Milky Way-type galaxies and their satellites using our hierarchical

method. We are currently performing such analyses of realistic galaxy simulations (Keller et al., 2015, 2016) to determine how much proper motion data is necessary to constrain the mass profile further, and to study what biases may occur under the Galaxy model that we have employed here when the distribution of the tracers does not follow a single power-law spatial distribution (G. M. Eadie et al. 2016, in preparation).

The mass profile result we have obtained in this study is at the lower end of most mass estimates in the literature but is also in agreement with some more recent measurements (e.g. Deason et al., 2012a; Battaglia et al., 2005; Gibbons et al., 2014). Because the result obtained in this study is so similar to the mass profile of Paper II, we refer the reader to that paper for further comparison to other studies. We end by noting, however, that our results could change substantially with the inclusion of proper motion data from remote tracers. The number of complete velocity measurements for GCs at large distances will soon increase thanks to projects such as the HST Proper Motion Collaboration (HSTPROMO³; van der Marel et al., 2014; Sohn et al., 2016), and with these data, a better estimate of the Galaxy’s total mass will be possible.

4.7 Conclusion

We have described a coherent, hierarchical Bayesian method for estimating the mass profile of the Milky Way Galaxy and applied this method to the Galaxy using GC data. This statistical framework allows us to take full advantage of all available GC kinematic data and also provides a meaningful and coherent

³HSTPROMO Project: <http://www.stsci.edu/marel/hstpromo.html>

probabilistic way to incorporate measurement uncertainties.

Using the assumptions of the power-law model (Section 4.2.1), the hierarchical framework for including uncertainties (Section 4.2), and the prior distributions (Section 4.2.2) and confronting this model with data from 143 GCs around the Milky Way, we arrive at a cumulative mass profile for the Galaxy with uncertainties (Figure 4.4) and a mass estimate within 125 kpc of $4.8 \times 10^{11} M_{\odot}$ (the 95% Bayesian credible regions are $(4.0 - 5.8) \times 10^{11} M_{\odot}$). When we extrapolate the mass profile to the virial radius (≈ 179 kpc), we find $M_{vir} = 6.2 \times 10^{11} M_{\odot}$ with a 95% Bayesian credible region of $(4.7 - 7.8) \times 10^{11} M_{\odot}$. This mass estimate is notably lower than those in other studies.

The statistical framework presented here will be highly useful and appropriate for other tracer objects around the Milky Way, such as halo stars and DGs. Using our approach with data sets from large programs, such as Gaia, could yield a well-constrained mass estimate for the Galaxy. Incorporating large data sets in this analysis will present some computational challenges, but given the effectiveness of our MCMC sampler we are confident that this will be a tractable problem through parallelization.

The first order of business, however, is to better understand what tracer populations will provide the most information about the Milky Way's gravitational potential. Thus, in our next paper (G. M. Eadie et al. 2016, in preparation), we perform a series of blind tests of simulated data of Milky Way-type galaxies that were created through hydrodynamical simulations (Keller et al., 2015, 2016) and investigate which tracer information is necessary for constraining the mass of the Milky Way.

Bibliography

- Bahcall, J. N., & Tremaine, S. 1981, *ApJ*, 244, 805
- Battaglia, G., Helmi, A., Morrison, H., et al. 2005, *MNRAS*, 364, 433
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics*, 2nd edn. (Princeton)
- Casetti-Dinescu, D. I., Girard, T. M., Jílková, L., et al. 2013, *AJ*, 146, 33
- Casetti-Dinescu, D. I., Girard, T. M., Korchagin, V. I., van Altena, W. F., & López, C. E. 2010, *AJ*, 140, 1282
- Deason, A. J., Belokurov, V., & Evans, N. W. 2011, *MNRAS*, 411, 1480
- Deason, A. J., Belokurov, V., Evans, N. W., & An, J. 2012a, *MNRAS*, 424, L44
- Deason, A. J., Belokurov, V., Evans, N. W., & McCarthy, I. G. 2012b, *ApJ*, 748, 2
- Dinescu, D. I., Girard, T. M., & van Altena, W. F. 1999, *AJ*, 117, 1792
- Dinescu, D. I., Keeney, B. A., Majewski, S. R., & Girard, T. M. 2004, *AJ*, 128, 687
- Dinescu, D. I., Martínez-Delgado, D., Girard, T. M., et al. 2005, *ApJL*, 631, L49

Eadie, G., & Harris, W. 2016, ApJ, 829, 108

Eadie, G., Harris, W., & Springford, A. 2015a, in JSM Proceedings (ASA)

Eadie, G., Harris, W., & Widrow, L. 2015b, ApJ, 806, 54

Eadie, G., Harris, W., Widrow, L., & Springford, A. 2015c, in Proceedings of the IAU, ed. A. Bragaglia, M. Arnaboldi, M. Rejkuba, & D. Romano (Cambridge Journals)

ESA. 1997, The Hipparcos and Tycho Catalogues

Evans, N. W., Hafner, R. M., & de Zeeuw, P. T. 1997, MNRAS, 286, 315

Evans, N. W., Wilkinson, M. I., Perrett, K. M., & Bridges, T. M. 2003, ApJ, 583, 752

Gibbons, S. L. J., Belokurov, V., & Evans, N. W. 2014, MNRAS, 445, 3788

Harris, W. E. 1996, AJ, 112, 1487

—. 2010 edition, ArXiv e-prints, arXiv:1012.3224, updated Globular Cluster catalogue

Johnson, D. R. H., & Soderblom, D. R. 1987, The Astronomical Journal, 93, 864

Kaffe, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, ApJ, 761, 98

Keller, B. W., Wadsley, J., & Couchman, H. M. P. 2015, MNRAS, 453, 3499

—. 2016, MNRAS, 463, 1431

- Kochanek, C. S. 1996, *ApJ*, 457, 228
- Kulesa, A. S., & Lynden-Bell, D. 1992, *MNRAS*, 255, 105
- Küpper, A. H. W., Balbinot, E., Bonaca, A., et al. 2015, *ApJ*, 803, 80
- Little, B., & Tremaine, S. 1987, *ApJ*, 320, 493
- McMillan, P. J. 2011, *MNRAS*, 414, 2446
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13
- Roberts, G. O., & Rosenthal, J. S. 2009, *Journal of Computational and Graphical Statistics*, 18, 349
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, 403, 1829
- Sohn, S. T., Van Der Marel, R. P., Deason, A. J., et al. 2016, in *American Astronomical Society Meeting Abstracts*, Vol. 228, American Astronomical Society Meeting Abstracts, 319.06
- van der Marel, R. P., Anderson, J., Bellini, A., et al. 2014, in *Astronomical Society of the Pacific Conference Series*, Vol. 480, *Structure and Dynamics of Disk Galaxies*, ed. M. S. Seigar & P. Treuhardt, 43
- Wang, W., Han, J., Cooper, A., et al. 2015, *MNRAS*, 453, 377
- Watkins, L., Evans, N., & An, J. 2010, *MNRAS*, 406, 264
- Wilkinson, M., & Evans, N. 1999, *MNRAS*, 310
- Williams, A. A., & Evans, N. W. 2015, *MNRAS*, 454, 698

5

A Blind Test on Milky Way-Type Galaxies from the MUGS2 Hydrodynamical, Λ CDM Simulations

In the chapter, I present preliminary results of original research that is in preparation for *The Astrophysical Journal*. The first author is myself, Gwendolyn Eadie, the second author is Dr. Benjamin Keller, and the third is Dr. William Harris. Dr. Keller supplied the mock data. In what follows, I performed the analysis and wrote the text.

5.1 Introduction

Multiple techniques have been created to measure the total mass and mass profile of the Milky Way (MW) Galaxy, including the timing argument (Kahn & Woltjer, 1959), the use of kinematics of globular clusters (GCs), halo stars, and stellar streams to trace the gravitational potential (e.g. Little & Tremaine, 1987; Wilkinson & Evans, 1999; Sakamoto et al., 2003; Dehnen et al., 2006; Xue et al., 2008; Deason et al., 2012; Law & Majewski, 2010, and many others), and more recently the direct comparison to cosmological simulations (e.g. Boylan-Kolchin et al., 2011; Busha et al., 2011; Patel et al., 2017). While each method has its own merits, the most popular approach continues to be using the kinematics of tracers to estimate the MW’s gravitational potential, and thus its total mass. The Gaia satellite (Perryman et al., 2001; ESA, 2016) and the Large Synoptic Survey Telescope (LSST-Corporation, 2016, hereafter LSST) data will greatly increase the number of kinematic tracers, and with this big data comes a demand for reliable methods that use tracer information to estimate the mass of the Galaxy.

In this vein, we have been developing a hierarchical Bayesian method to measure the mass and mass profile of the MW that uses kinematic tracers. The method is tentatively called *Galactic Mass Estimator* (GME), and it has been applied to the Galactic globular cluster (GC) data in a series of papers (Eadie et al., 2015; Eadie & Harris, 2016; Eadie et al., 2017b,a, hereafter Papers 1, 2, and 3¹ respectively.). GME uses the distribution function (DF) from Evans et al. (1997), which was also used by Deason et al. (2011) in a maximum likelihood analysis of halo stars to measure the mass of the MW. Our results provide

¹this paper includes an erratum, see Appendix A

a 95% Bayesian credible region for the virial mass of $(0.67, 1.09) \times 10^{12} M_{\odot}$, which is in agreement with several recent studies (e.g. Xue et al., 2008; Diaz et al., 2014; Gibbons et al., 2014; Patel et al., 2017; McMillan, 2017). Our method has at least two advantages over traditional mass estimation methods, one being the inclusion of incomplete data (tracers which only have line-of-sight velocity measurements) and two being a measurement model that accounts for observational error. Our method produces Bayesian credible regions for the cumulative mass profile of the Galaxy at any Galacocentric radii, rather than point estimates of the total mass within a certain distance.

What remains unclear in our previous analyses, however, is whether the mass profile correctly describes the shape and true total mass of the Galaxy. It is impossible to know the *true* Galactic mass profile, but it would be useful to have some sense of how well our mass profile prediction encloses the true mass profile within the uncertainties. We also want to obtain a better understanding of the model’s limitations. Thus, it seems a natural step from Papers 1, 2, and 3 to test the hierarchical Bayesian method on mock observations derived from hydrodynamical simulations of MW-type galaxies, in order to obtain insight into the predictive properties of our choice of model.

In this study, we perform such tests on simulated observations of GC analogues from the McMaster Unbiased Galaxy Simulations 2 (MUGS2) hydrodynamical simulations (see Keller et al., 2015, 2016). The simulations incorporate the modern smoothed particle hydrodynamics code GASOLINE2 (Wadsley et al., 2004; Wadsley et al., 2017) and include such effects as low-temperature metal cooling, UV background radiation, star formation, and stellar and super-bubble feedback (Keller et al., 2015; Stinson et al., 2010). These mock galaxies provide a way to test our method’s predictive power because their stellar and

dark matter profiles are probably more complicated than our model— a situation that is likely similar with the real data of the MW.

We perform blind tests on the mock data (i.e. without prior knowledge of the true mass of the simulated galaxies) from eight MUGS2 simulated galaxies. Two of these galaxies, *g15784* and *g1536*, are first analysed in detail, and then six more are analysed to obtain a greater sample. Section 5.2 briefly reviews the physical model that is used in our hierarchical Bayesian framework. Section 5.3 describes how mock tracers from the MUGS2 simulations were selected, and Section 5.3.3 describes how mock heliocentric observations (with errors) were created from these tracers. In Section 5.4, we show the mass profile and virial mass predictions and compare these to the true quantities (which were revealed to us from Keller after our analysis was complete). A brief discussion follows in Section 5.5, and Section 5.6 provides a summary of our findings.

5.2 Brief Review of the Physical Model

We adopt the same physical model that was used in Papers 2 and 3 (Evans et al., 1997; Deason et al., 2011). The model assumes a total gravitational potential given by

$$\Phi(r) = \frac{\Phi_o}{r^\gamma} \quad (5.1)$$

where Φ_o and γ are parameters. The radial distribution of the tracer population is also assumed to follow a power-law profile:

$$\rho(r) \propto \frac{1}{r^\alpha} \quad (5.2)$$

where α is a parameter. Given Equation 5.1, the total mass profile is given by

$$M(r) = \frac{\gamma\Phi_o}{G} \left(\frac{r}{\text{kpc}} \right)^{1-\gamma}, \quad (5.3)$$

which goes to an isothermal sphere in the limit that $\gamma \rightarrow 0$, and a point mass as $\gamma \rightarrow 1$. Equations 5.1 and 5.2 are used to derive a distribution function (DF) through the Eddington formula (see Binney & Tremaine, 2008). For the curious reader, the derivation of the DF is given in both the original paper by Evans et al. (1997) and in Paper 2 (Chapter 4, albeit with different notation).

The DF is a probability distribution; it gives the probability of a single tracer having a particular specific energy \mathcal{E} and specific angular momentum L , given the model parameters $(\Phi_o, \gamma, \alpha, \beta)$, where β is the constant velocity anisotropy parameter for the tracer population (Binney & Tremaine, 2008). In our Bayesian set-up, all four model parameters are free. The prior probabilities on the model parameters are uniform distributions for (Φ_o, γ, β) , and the prior on α is a Gamma distribution.

The measurement model, which incorporates the measurement uncertainties in the data, is described fully in Paper 3. In short, the measurement model treats the true position and velocity components of the tracers as nuisance parameters, and constrains these parameters using both the data and the measurement uncertainties. For information on the particulars of the MCMC methods for sampling the posterior distribution, the reader may refer to Paper 1, with some updates in Papers 2 and 3.

5.3 Creating Mock Observations of Tracers in MUGS2 Galaxies

In order for our blind test to be as realistic as possible, we need to create a set of tracer data from the simulated galaxies that most closely resembles that of the MW. In the next few sections, we describe this process.

5.3.1 Finding GC analogues in MUGS2 Galaxies

Although cosmological, hydrodynamical simulations have made great strides in recent years, resolution limits have prevented the ability to create GC populations within a fully simulated, cosmological galactic environment. Therefore, MUGS2 does not have GCs with which we can use to test our method. Instead, we must use a selection of star particles from each mock galaxy and treat them as *GC analogues*. These *star particles* represent entire populations of stars, with each particle carrying a mass of $\sim 10^5 M_\odot$. Coincidentally, this mass is similar to the mass of many GCs found in the MW.

In order to keep our analysis a true blind test, we asked Keller to select the GC analogues independently and send us only their kinematic information. Keller chose GC analogues from the MUGS2 data by filtering out star particles with ages greater than 12 billion years and metallicities $[Fe/H] < -1.5$. With such cuts, the stars in a GC analogue would have formed at an approximate redshift $z \approx 3$. Keller also removed any disk-associated objects by removing star particles within a galaxy-centered cylinder with radius $3r_e$ and height r_e , where r_e is the half-light radius of the Galaxy.

Keller then provided us with the Galactocentric positions (x, y, z) and velocities (v_x, v_y, v_z) of the GC analogues, and told us that the MUGS2 host galaxies were “Milky Way-type”. We had no knowledge of their mass, mass profile, or merger history.

The number of GC analogues in each MUGS2 galaxy varies significantly, ranging from populations of 64 to 5106— a reflection of the different star formation rates at high redshift. Galaxies with more GC analogues had more star formation in earlier times. Some of these population sizes are not realistic when compared to the MW GC population, which consists of 157 known GCs (Harris, 1996, 2010). Although the number of GC analogues does not match the number of MW GCs, we can randomly sample GC analogues from the population to mimic the MW.

For this preliminary work, we randomly select 157 GC analogues from each of g15784 and g1536. Having thousands of GC analogues at our disposal, while unrealistic, will make for some interesting statistical studies in the future. For example, we can repeat the sampling and perform the blind test more than once on the same galaxy to uncover and understand possible biases, although this is left to future work.

5.3.2 Selecting MUGS2 Galaxies to Investigate

Eighteen MW-type galaxies were created in the MUGS2 simulations. The black points in (Figure 5.1) show the GC analogues’ total velocity v as a function of galactocentric distance r . For comparison, the MW GC data is also plotted in blue (only GCs with complete velocity measurements are shown).

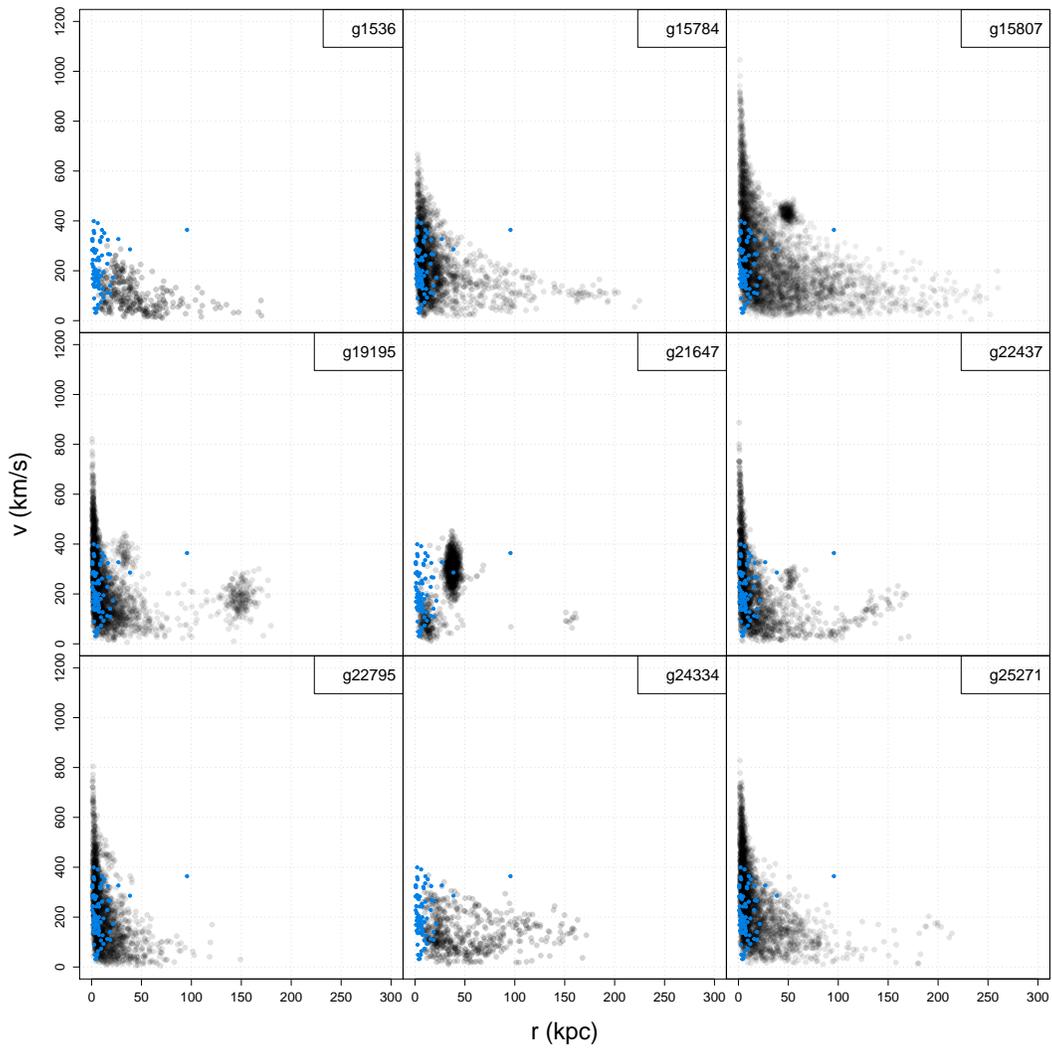
Most of the galaxies’ GC analogues display the same overall shape to their

velocity profile—the average speed tends to decrease with distance r from the galactic centre. However, there are also clumpy features in some of the GC analogue profiles which could indicate recent merger events and/or the presence of massive satellites. The GC system of the MW is thought to be an approximately virialized population, and indeed these kinds of features are not present in the MW GC data. Therefore, it seems prudent to avoid MUGS2 galaxies with these features for this preliminary analysis).

As a first step, we chose to work with $g15784$ because it lacked any strange features in its velocity profile, had many GC analogues from which I could draw samples, and frankly, because its velocity profile and mock image of the galaxy “looked nice”. It should be noted that we did not look at the *comparison* of the MW GC data to the MUGS2 data (Figure 5.1) until after the analysis was complete, and only after-the-fact did we realize the MW GC data overlaps quite well with the GC analogues for $g15784$. After analysing $g15784$, we studied $g1536$ at the suggestion of Keller, who relayed the information that $g1536$ was a less concentrated galaxy than 15784 and thus would make an interesting comparison. In the two galaxies, $g15784$ and $g1536$, the GC analogue populations are 2381 and 311 respectively. We follow-up this detailed examination by applying our method to six more MUGS2 galaxies: $g19195$, $g22437$, $g22795$, $g25271$, $g4145$, and $g4720$.

5.3.3 Creating Mock Heliocentric Observations

After selecting MUGS2 galaxies to analyse, we created mock heliocentric observations of the GC analogues, such as might be viewed from a Sun-centered reference frame. This involved a series of steps, including not only transforming



(a)

Figure 5.1: Velocity profiles of GC analogues from MUGS2 simulated galaxies, with MW GC velocities (for GCs with complete data) overplotted in blue. Galaxies *g1536* and *g15784* are the first two frames in the top row.

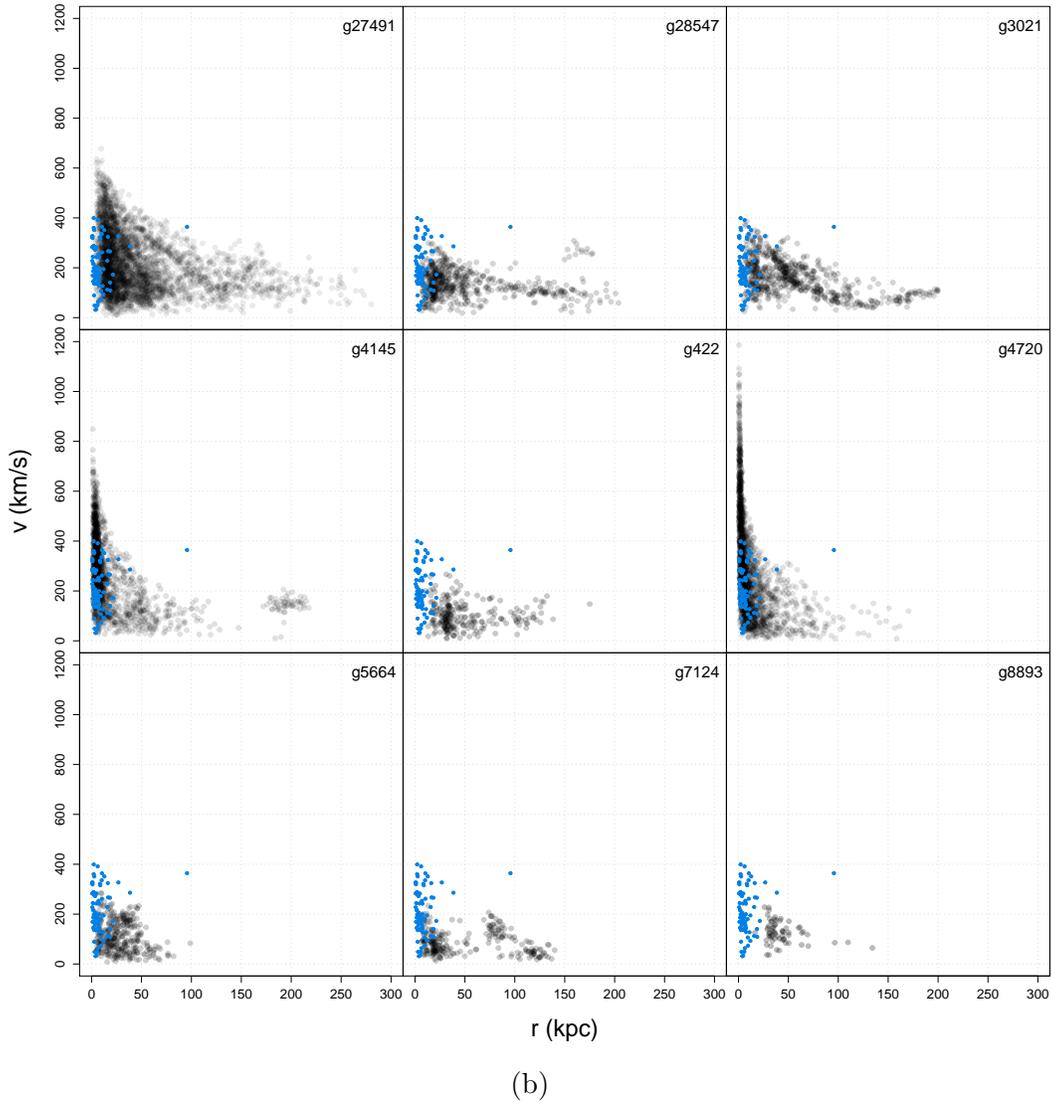


Figure 5.1: (cont'd) Velocity profiles of GC analogues from MUGS2 simulated galaxies, with MW GC velocities (for GCs with complete data) over-plotted in blue.

positions and velocities into a heliocentric frame, but also introducing missing data and measurement errors to simulate real observations.

5.3.4 Transforming from Galactocentric to Heliocentric

The kinematic information of the GC analogues directly from the MUGS2 simulations is in a galactocentric, Cartesian coordinate system with positions (x, y, z) and velocities (v_x, v_y, v_z) . To create heliocentric positions and velocities, we perform a series of steps:

1. Transform the galactocentric Cartesian coordinates into heliocentric Cartesian coordinates by adjusting for the solar position.
2. Calculate the distance d from the Sun to the GC analogue, which is then used to calculate the parallax λ .
3. Transform the heliocentric positions (Cartesian coordinates) into galactic coordinates l and b , and subsequently into right ascension (RA, α) and declination (decl., δ).
4. Using the GC analogues' λ , α , δ and Galactocentric velocities, we perform the inverse of the transformation described in Johnson & Soderblom (1987) to obtain the line-of-sight velocities (v_{los}) and proper motions ($\mu_\alpha \cos \delta$, μ_δ):

$$\begin{bmatrix} v_{los} \\ k\mu_\alpha \cos(\delta)/\lambda \\ k\mu_\delta/\lambda \end{bmatrix} = (\mathbf{T} \cdot \mathbf{A})^{-1} \left(\begin{bmatrix} U \\ V \\ W \end{bmatrix} - \begin{bmatrix} U_\odot \\ V_\odot \\ W_\odot \end{bmatrix} \right), \quad (5.4)$$

where $k=4.74057$ (the equivalent in km s^{-1} of 1 AU in one tropical year), the vector $(U_{\odot}, V_{\odot}, W_{\odot})$ is the solar motion (we use the value from Schönrich et al., 2010)², the matrix \mathbf{T} is determined by the RA and decl. of the North Galactic Pole, and the matrix \mathbf{A} is determined by α and δ of the GC analogue.

The above steps were performed for the MUGS2 GC analogue data, and then checked by transforming the heliocentric data back to Galactocentric with the relevant code in GME. This was done without introducing scatter, to confirm that we could obtain the correct Galactocentric values when transforming back to the Galactocentric frame.

We notice discrepancies of $\simeq 2 \times 10^{-13} \text{km s}^{-1}$ in the velocities at small r_{gc} when the transformation is performed *without* scatter, which are due to numerical issues (Figure 5.2). The symmetry of these discrepancies and their small values will not contribute to any systematic bias in the result.

5.3.5 Assigning Measurement Uncertainties

The measurement model in the hierarchical Bayesian method (Paper 3) assumes that a measurement of some quantity x is a random variable X that is normally distributed,

$$X \sim \mathcal{N}(\mu, \sigma^2) \tag{5.5}$$

where μ , the mean, is the true value of the quantity and where σ^2 , the variance, is equal to the square of the measurement uncertainty. Any difference between the measured x and the true value μ is the *error*. In order to analyse the

² U is positive toward the Galactic center, V is positive in the direction of Galactic rotation, and W is positive above the Galactic plane.

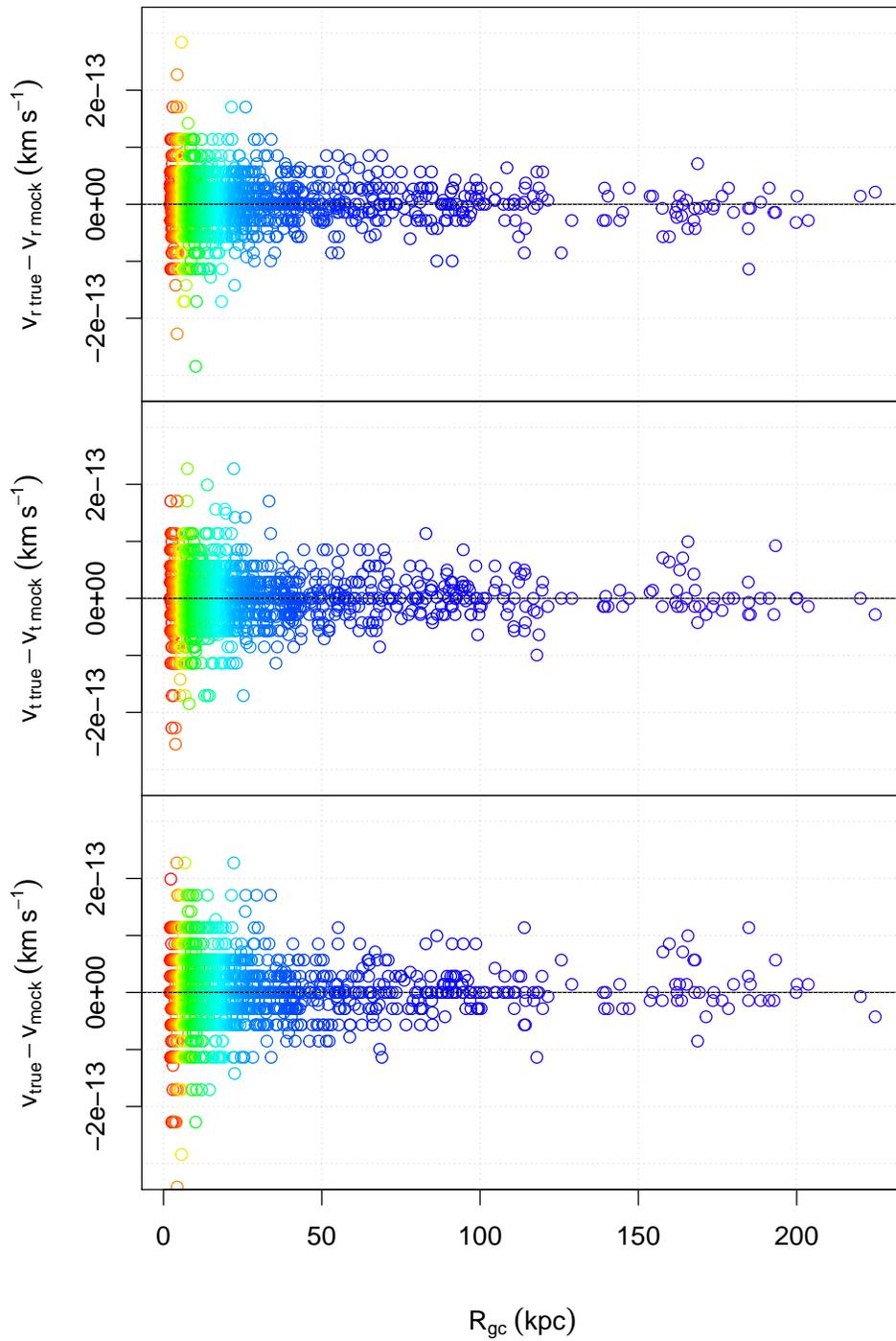


Figure 5.2: Difference between the true radial, tangential, and total velocities and those calculated from the heliocentric velocities, as a function of Galactocentric distance (indicated by colours).

MUGS2 data in a way that is most similar to the MW analysis, we must create realistic, observational errors. We achieve this by setting the mean μ to the true value of the quantity (e.g. r , v_{los} , etc) in the MUGS2 data, deciding on a value for σ^2 , and drawing from the normal distribution determined by these parameters. How we choose to define σ^2 for each quantity (r , v_{los} , $\mu_\alpha \cos \delta$, μ_δ) is important because it may determine how much leverage a data point has on the final analysis.

The galactocentric distances r were assigned a measurement uncertainty of 5% (see Harris, 1996, 2010). The proper motion and line-of-sight measurement uncertainties were drawn with replacement from the real data uncertainties by randomly selecting a row from the MW GC list given in Paper 2. We excluded two large measurement uncertainties in this process — that of Pal 3 and NGC 6218 — to avoid assigning very large observed proper motions to the GC analogues that we deemed to be unrealistic³. We also investigated various distance measures to match MW GCs to GC analogues, so that GC analogues could be assigned uncertainties that were similar to their MW counterparts, but found these procedures gave final error distributions that were indistinguishable from the simple random sampling.

Once the transformations of the MUGS2 data were complete and measurement errors were created for the GC analogues of each MUGS2 galaxy, a subsample of 157 was taken at random from each galaxy to mimic the sample size of the MW’s GC population.

³in the MW analysis, we treated the proper motion of Pal 3 as unknown, because the measurement uncertainty was so large.

5.3.6 Creating Incomplete Data

Of the 157 GCs in the MW listed in Paper 2, 85 do not have proper motion measurements and 14 of this subset also lack line-of-sight velocity measurements. Within 20kpc of the MW centre, approximately 50% (67/135) of the data are missing proper motion measurements, and beyond this distance approximately 87% are missing proper motions. We mimic this distribution of incomplete data in the subsamples by removing 50% of the proper motions within 20kpc, and removing 87% outside of this distance. After doing this, there are many GC analogues with proper motion measurements beyond 50kpc, which to us seems unrealistic, since the real MW has only one complete data point beyond this distance. Thus, we decided to remove all proper motions beyond 50kpc.

In Papers 2 and 3, we used the 14 GCs that lacked line-of-sight measurements to define the prior distribution in the number density profile parameter α (see Chapter 3). For the MUGS2 GC analogues, we randomly remove 14 line-of-sight velocities and use the positions of these objects in the same way as we did for the real data. To check that the distribution of incomplete data is similar to that of the MW, we plot the smoothed distributions of the GC analogue mock data and compare it to that of the MW GC data. The subsample for *g15784* is shown as an example in Figure 5.3.

In the case of *g1536*, any given subsample of 157 GC analogues of the simulated data has ~ 30 GC analogues residing within 20kpc of the galactic centre. This is in stark contrast to the MW, which has 135 GCs within 20kpc. In the MW, proper motion measurements are available for at least 67 GCs within 20kpc. Removing 50% of the proper motion measurements within 20kpc for *g1536* (as was done for *g15784*) seems unrealistic— if we lived in a galaxy with

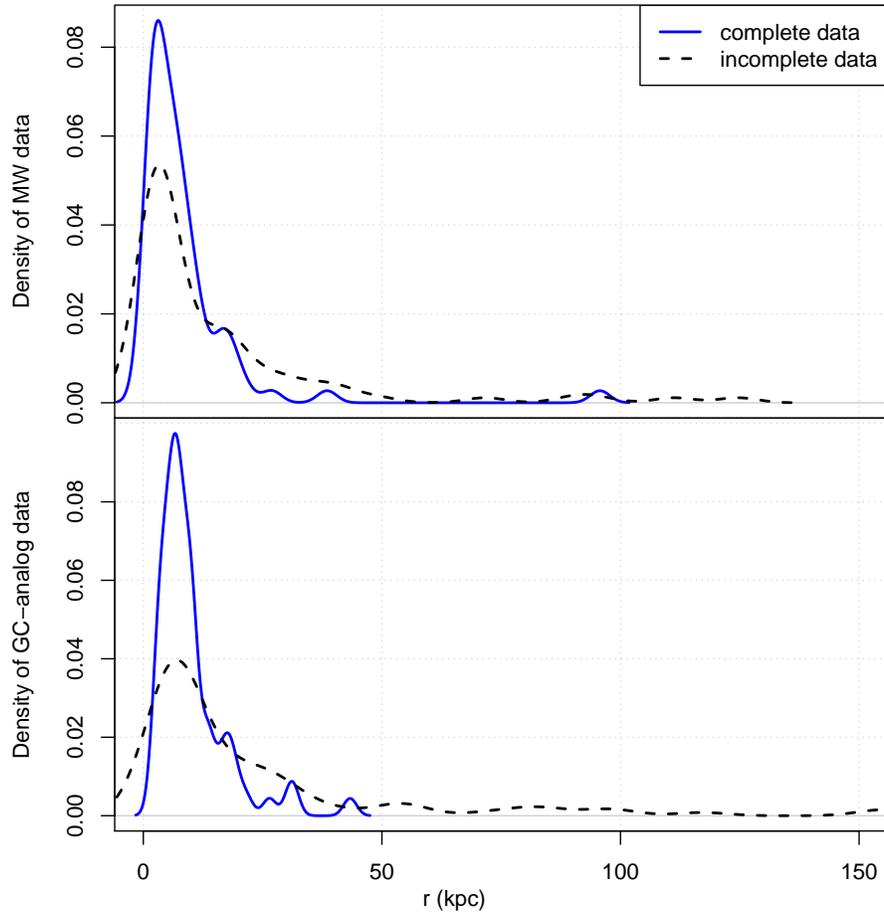


Figure 5.3: Top: The distribution of the MW complete and incomplete data, as a function of Galactocentric distance (solid blue and dashed black lines respectively). Bottom: The distribution of complete and incomplete mock data generated from a subsample of $g15784$ GC analogues.

only 30 GCs within 20kpc of the Galactic centre, then surely we would have proper motion measurements for almost all of them! Thus, we decide to keep all but two proper motion measurements within 20kpc. Beyond 50kpc, all proper motions are removed, based on the fact that measurements of the most distant GCs require a certain amount of time, technology, and commitment on the part of us humans.

5.4 Preliminary Results

We now apply the GME code to the GC analogue subsamples from eight MUGS2 galaxies, and present the results below (Section 5.4.1). We begin with a detailed look at $g15784$ and $g1536$ (Section 5.4.1). Next, we repeat the analysis on six other MW-type MUGS2 galaxies ($g19195$, $g22437$, $g22795$, $g25271$, $g4145$, and $g4720$) that have similar GC-analogue spatial distributions to that of the MW's real GC system (Section 5.4.2).

5.4.1 Detailed Test Cases: $g15784$ and $g1536$

The mean estimates of the model parameters given by the posterior distribution for galaxy $g15784$ are presented in Table 5.1, where the numbers in brackets represent the bounds of the 95% marginal credible regions.

We also calculate the virial radius and virial mass of galaxy $g15784$ from the posterior distribution of model parameters. We define the virial radius as the distance from the galactic centre within which the mean mass density of the galaxy is 200 times the critical density of the Universe, and use a Hubble constant of $67.8 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al., 2016). The mean virial radius and virial mass of galaxy $g15784$ as predicted from

the hierarchical Bayesian analysis are $r_{vir} = 217$ (184, 249) kpc and $M_{vir} = 1.1$ (0.7, 1.7) $\times 10^{12} M_{\odot}$, where numbers in brackets are 95% Bayesian credible regions (Table 5.1). These values are strikingly accurate—the true values from the simulations are 219 kpc and $1.3 \times 10^{12} M_{\odot}$.

After completing the analysis of $g15784$, galaxy $g1536$ was recommended by Keller for a second blind test, because its structure is somewhat different from $g15784$. The results are shown in Table 5.2. In this case, the method did not perform as well; the true virial radius and mass are 174kpc and $0.65 \times 10^{12} M_{\odot}$ respectively, whereas the predicted values were $r_{vir} = 162$ (143, 182)kpc and $M_{vir} = 0.5$ (0.3, 0.6) $\times 10^{12} M_{\odot}$.

We also calculate the cumulative mass profile Bayesian credible regions, using Equation 5.3 (Figure 5.4a). Although the virial mass is well estimated using the GC analogues of $g15784$, the predicted mass profile shows disagreement with the true mass profile. The predicted $M(r)$ profile falls below the true profile for many values of r for both $g15784$ and $g1536$.

The hierarchical Bayesian method treats the true positions and velocities as nuisance parameters, sampling them in the Monte Carlo Markov Chain hybrid-Gibbs algorithm (see Paper 1 and 3). As a result, we obtain marginal

Table 5.1: Model Parameter Estimates: $g15784$

Parameter	Mean	95% Marginal Credible Region
$\Phi_o(10^4 \text{km}^2 \text{s}^{-2})$	47	(40, 57)
γ	0.41	(0.31, 0.57)
α	3.04	(3.02, 3.06)
β	0.54	(0.41, 0.66)
Derived Quantity		
r_{vir} (kpc)	216	(183, 248)
M_{vir} ($10^{12} M_{\odot}$)	1.1	(0.7, 1.6)

Table 5.2: Model Parameter Estimates: $g1536$

Parameter	Mean	95% Marginal Credible Region
$\Phi_o(10^4 \text{km}^2 \text{s}^{-2})$	25	(19, 36)
γ	0.41	(0.30, 0.62)
α	3.03	(3.01, 3.04)
β	0.46	(0.22, 0.64)
Derived Quantity		
r_{vir} (kpc)	162	(143, 182)
M_{vir} ($10^{12} M_\odot$)	0.5	(0.3, 0.6)

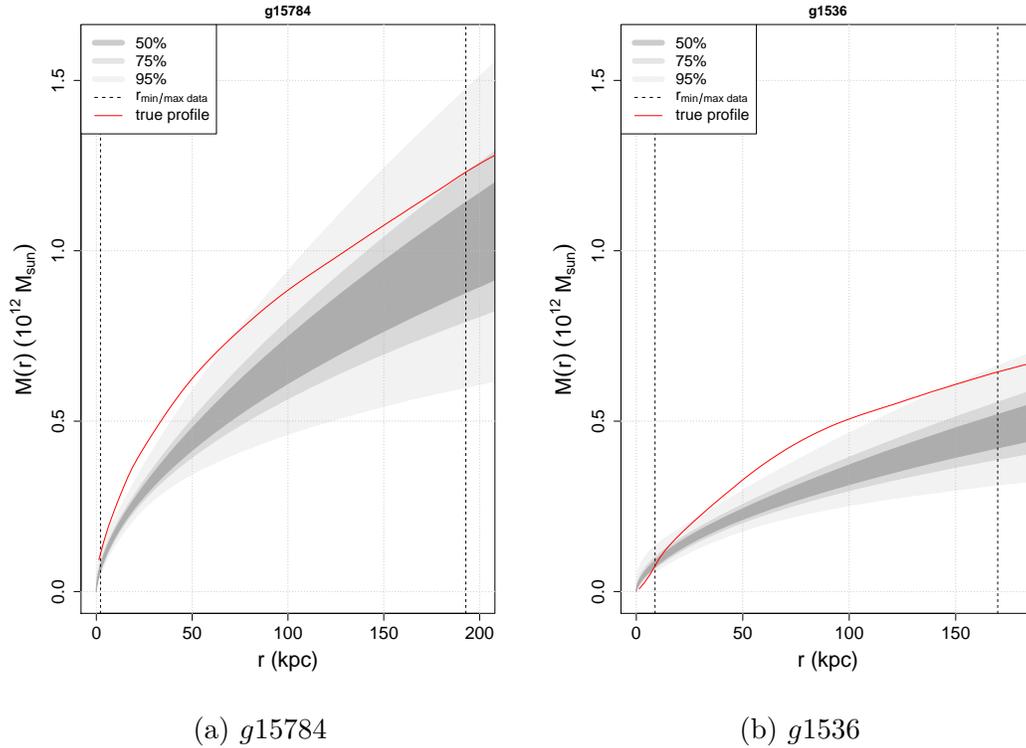


Figure 5.4: The predicted (grey shaded regions) and true (red line) cumulative mass profiles for galaxy (a) $g15784$ and (b) $g1536$. The dashed vertical lines indicate the range of the mock observations.

posterior distributions for the galactocentric velocity and position of each GC analogue, which can be used to estimate the specific energy \mathcal{E} . In Paper 3, we used these marginal posterior distributions to create an energy profile as a function of radius, and compared the predicted energies to the measured energies. We noted that the estimated energies were altered when the measurement uncertainties were large, in perhaps the model’s attempt to describe the GC’s kinematic energy in light of the other GCs. Here, we repeat the same procedure of comparing the mock “measured” energies of the GC analogs to the estimated energies, but also have the luxury of comparing the latter to the true energies of the GC analogues (given the estimates for the model parameters).

The estimated, measured, and true specific energies of the GC analogues from *g15784* are shown in Figure 5.5. All energies are calculated using the mean estimates of the model ($\Phi_o, \gamma, \alpha, \beta$) and nuisance ($r, v_{los}, \mu_\alpha \cos \delta, \mu_\delta$) parameters. The measured complete data (solid green diamonds) are connected to the estimated energies (solid blue circles) via black arrows. The estimated energies from the posterior distribution for the incomplete data are shown as open blue circles. The true energies, given the mean model parameters, are shown as grey squares. There is neither an over- nor under-estimated systematic bias in the energy estimates. However, we can see that the estimates are pulled toward the profile created by the incomplete data measurements (Figure 5.5). Overall, the ability of GME to adjust the energy values of the GC analogues reflects the result found in Paper 3, for the real MW data.

The estimated energy profile of the GC analogues from *g1536* is shown in Figure 5.6. Again, the mock measurements of the complete data and mean model parameters are used to calculate the measured energy (solid green diamonds), and the estimated energies (solid blue = complete, hollow blue =

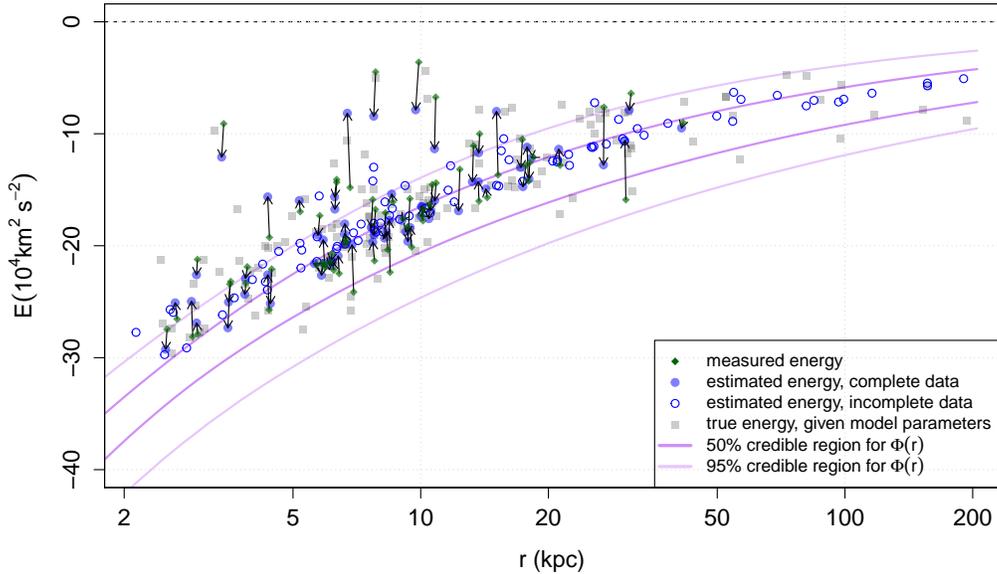


Figure 5.5: Energy profile of GC analogues from $g15784$. The solid green diamonds show the specific energy of the GC analogues which have complete velocity measurements. The solid blue and hollow blue points are the estimated energies, calculated using the mean values of both the model and nuisance parameters of the posterior distribution. Arrows connect the measured energies to the estimated energies for the GC analogues with complete data. The grey squares show the true energies, based on the best estimates of the model parameters. The purple curves are the 50 and 95% regions for the gravitational potential. Any object above the dotted line ($\mathcal{E} = 0$) would be unbound.

incomplete) are calculated using the means of the nuisance and model parameters. The true values of the energies, given the model parameters, are shown as grey solid squares. Immediately obvious in Figure 5.6 are ~ 6 GC analogues whose measured energies differ greatly from the predicted ones. However, the true values (grey squares) are quite different as well, and the estimated energies are in better agreement with the true energies. The model has taken into account the large uncertainties of these objects and adjusted their position and velocity parameters such that the GC analogue energies match the average distribution of the other energies at that distance.

For galaxy $g1536$, about 50% of the GC analogue data are complete between

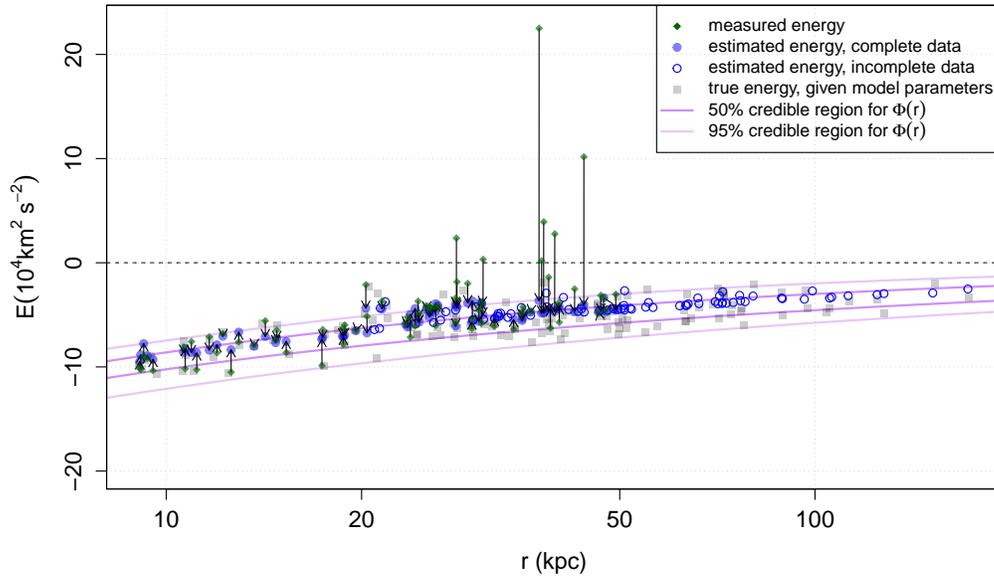


Figure 5.6: Estimated energy profile of GC analogues from $g1536$. The symbol notation is the same as Figure 5.5, except that grey squares are added to indicate the true energies of the GC analogues.

20 and 50kpc, and the estimated energies for a significant portion of these GC analogues are not very different from the true values. There were six GC analogues with large measured energies, but the measurement model accounted for their large errors and estimated their energies to be much lower—which was the correct assumption in this case. However, despite accurately predicting the approximate energy of the six GC analogues that had large errors, the total mass and mass profile for $g1536$ was underestimated.

5.4.2 Mass Estimates of MUGS2 galaxies

We now apply GME to six more MUGS2 galaxies, to bring the total to eight, and present the results. Galaxies $g19195$, $g22437$, $g22795$, $g25271$, $g4145$, and $g4720$ were chosen for analysis because their GC analogue populations have

radial density profiles similar to the GC population of the MW. Figure 5.7 summarizes the mean estimates of the virial mass (top), virial radius (middle), and constant velocity anisotropy (bottom) for all eight MUGS2 galaxies analysed thus far. The estimates are shown as black circles, the true values are blue diamonds, and the 95% credible regions are shown as error bars.

The virial masses are underestimated for seven out of eight of the MUGS2 galaxies, and the 95% credible regions contain the true virial mass for all but $g22795$ and $g4720$ (top panel, Figure 5.7, top row). This could hint at a systematic bias in our method, but it is important to make this kind of inference cautiously, for a few reasons.

First, recall that the six additional galaxies were chosen because their GC analogue populations had similar spatial distributions to the real GCs of the MW. After the analysis was complete and the true masses and characteristics of the galaxies were revealed to us, we learned we had chosen six “unregulated” MUGS2 galaxies (Keller et al., 2016), discussed next.

Half of the galaxies in the MUGS2 simulations were categorized as *unregulated* (Keller et al., 2016). During the last stages of evolution in the simulations, feedback mechanisms were unable to effectively expel gas from these galaxies, which led to an overproduction of stars in their discs. Ultimately, each of these unregulated galaxies formed a massive bulge at its center which eventually depleted its gas reservoir, and created a strong central Keplerian potential (see Keller et al., 2016, Figure 4). When the gas mass is normalized by the virial mass, it becomes clear that $g22795$ and $g4720$ are two of the most extreme unregulated galaxies (Figure 5.8). Furthermore, Keller et al. (2016) note that the unregulated galaxies do not follow the standard stellar mass to halo mass relation, and therefore are not quite MW-type galaxies. Of the eight MUGS2

galaxies we have analysed thus far, seven out of eight were unregulated, with the exception being *g1536*.

Second, the two most poorly estimated virial masses, *g22795* and *g4720*, have marginal posterior distributions for γ that reached the upper limit of the prior (i.e. 0.7). In a Bayesian analysis, a marginal distribution for a parameter that is truncated by the upper or lower limit of the prior distribution should inspire suspicion in the results, because this behaviour indicates that the model and prior assumptions are not capable of describing the data. In the cases of *g22795* and *g4720*, it seems that γ was attempting to reach a value larger than 0.7 because of the extremely compact discs; in the limit that $\gamma \rightarrow 1$, the gravitational potential model goes to a Keplerian potential (Equation 5.1). As we saw in Paper 2, the parameters Φ_o and γ are anti-correlated, and thus a larger value of the latter will lead to a smaller value of the former— which consequently means a smaller mass estimate.

Third, the sample size of galaxies is only eight, and each galaxy was analysed with one sample of GC analogues. Thus, the probability that seven out of eight galaxies would be underestimated by chance is not insignificant.

Returning to Figure 5.7, the virial radius estimates (middle row) are in reasonably good agreement with the true values and within the 95% credible regions for all but the two special cases mentioned above. The β parameter estimates, on the other hand, are more poorly estimated, and tend to indicate a less radially-biased anisotropy than the true velocity anisotropy calculated from the GC analogue subsample (5.7, bottom row). Nevertheless, the accuracy of the β estimate does not appear to be related to the accuracy of the virial mass or virial radius. It is unclear why the velocity anisotropy is underestimated, although it is important to remember that the mock data was

incomplete beyond 50kpc. It will be interesting to test how the β estimates change as more proper motions measurements are included in future tests.

5.5 Discussion

The results of this preliminary study are intriguing. While the measurement model seems to account for the measurement uncertainties very well, the physical model poorly describes the shape of the galaxy's mass profile at all radii. On the other hand, the virial radius and virial mass estimates are quite good for most of the galaxies and within the 95% credible regions.

The virial mass estimates of the eight MUGS2 galaxies analysed so far do show a tendency to be underestimated (top panel of Figure 5.7). However, all of the galaxies we analysed so far (with the exception of *g1536*) were the unregulated type from the MUGS2 sample. We selected out these galaxies for analysis by looking at the radial density profile of their GC analogues and without considering their star formation histories or final gas fractions. The galaxies *g22795* and *g4720* in particular are two extreme case, and indeed these are the two galaxies which are the most poorly estimated in our study.

One of the most important next steps is to apply the hierarchical Bayesian method to the *regulated* MUGS2 galaxies. The regulated galaxies are probably better representations of the MW and therefore will provide a second test of our method.

Although we analysed mostly unregulated galaxies in this study, we nevertheless have gained insight into the features of our method and the behaviour of the physical model. Our method did recover the virial mass within the 95% credible regions in most cases, but the physical model could not describe

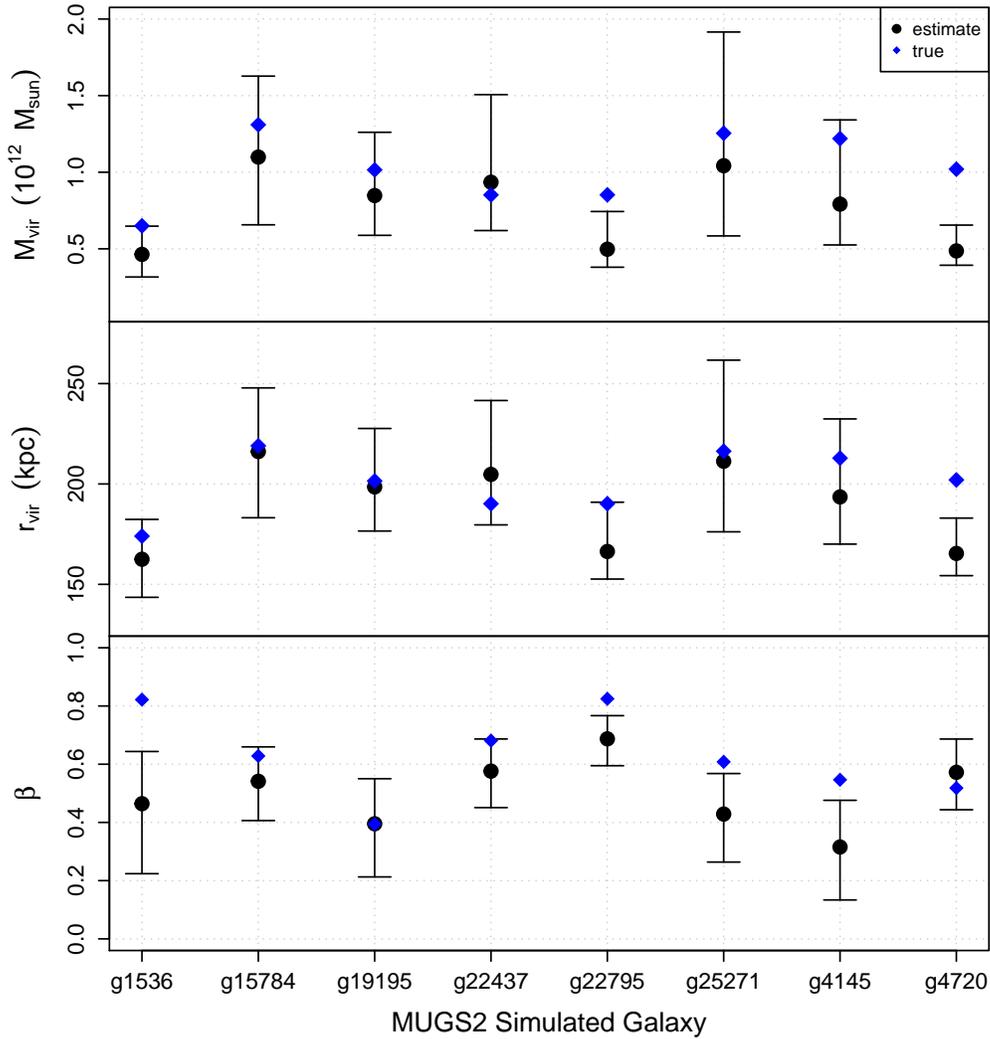


Figure 5.7: True values (blue diamonds), estimates (black circles), and 95% credible regions (error bars) for each MUGS2 galaxy’s virial mass (top), virial radius (middle), and velocity anisotropy parameter (bottom). The estimates were calculated after analysing mock observations of 157 GC analogues for each of the eight MUGS2 simulated galaxies.

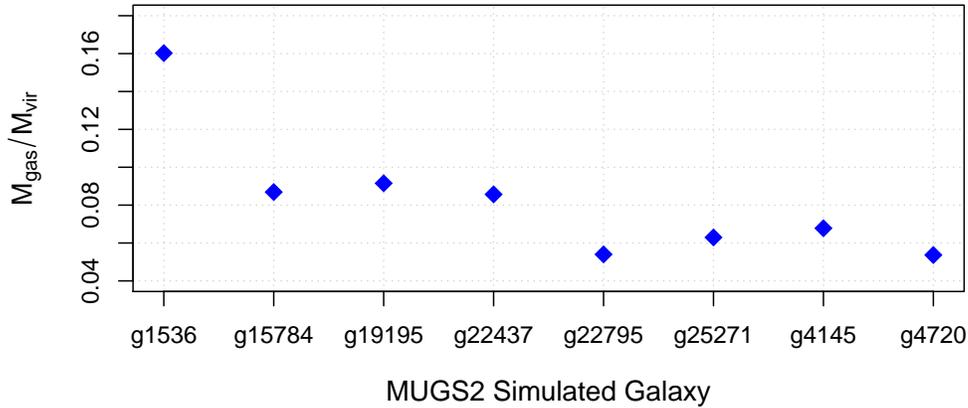


Figure 5.8: The gas mass fraction at $z = 0$ of the eight MUGS2 galaxies analysed in this study (created using data from Keller et al. (2016), Table 1).

the inner regions of the cumulative mass profiles of the galaxies. If our interpretation of the analysis thus far is correct, then an important next step is introducing a better model for the gravitational potential.

Treating the positions and velocities as nuisance parameters enables the estimated energy to change significantly when the measurements are highly uncertain. For objects whose measurement uncertainties are large, including these uncertainties can have a profound affect at the individual level (i.e. the energy estimate for that GC analogue) and at the mass measurement level. If we had not included the measurement uncertainties, then the six outliers in *g1536* would have carried significantly more weight and led to a larger virial mass estimate.

Including measurement uncertainties may lead to narrower credible regions in the mass profile, but if the model describes poorly the underlying mass distribution then these credible regions may be overconfident. It is therefore

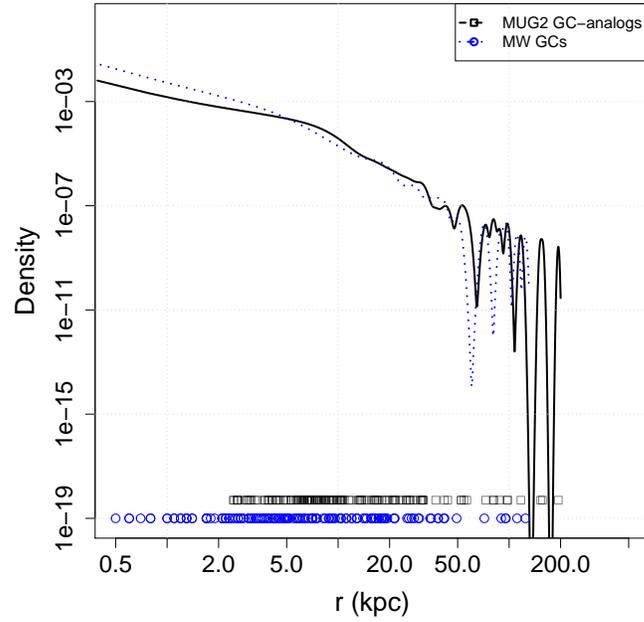
important that we investigate using only the outer tracers from the simulations, to see if they provide a better estimate of the MW mass. However, it is interesting that the less concentrated simulated galaxy (*g1536*) was more poorly fit than the more concentrated one, even though there were fewer GC analogues at inner radii.

The MUGS2 galaxies that have GC analogue populations most similar in size to the MW are *g8893* (64), *g7124* (247), and *g422* (251), so a next step might be to analyse the GC analogues of these galaxies. Revisiting Figure 5.1, however, shows us that the velocity profiles of these GC analogues are rather dissimilar to the GC analogues in the MW, populating regions much further out in the their host galaxies (again, due to the original cuts that removed disk-associated GC analogues).

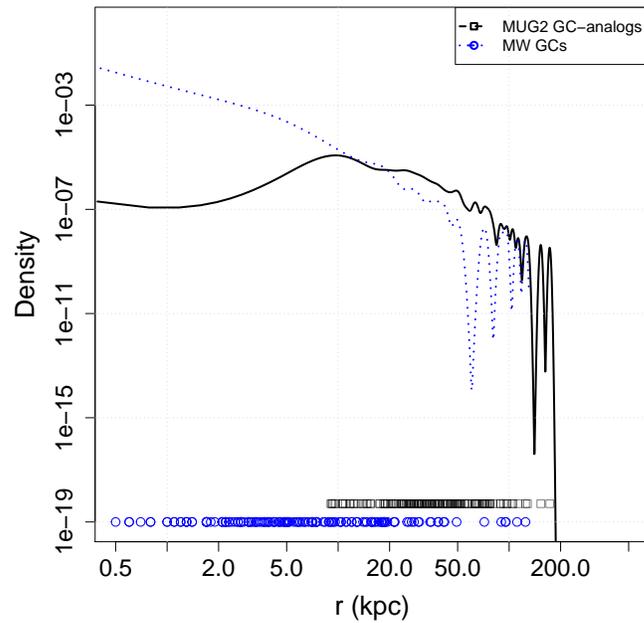
One way to test whether the GC analogues are from a distribution similar to that of the MW GCs would be to perform an empirical statistical test. For this preliminary work, we simply compare the normalized number density profile of the GC analogs for both simulated galaxies to the GC number density profile in the MW (Figure 5.9). We can see that the cylindrical cuts to remove disk-associated GC analogues significantly changes the spatial distribution at small radii. However, the distributions at large radii look similar.

One hypothesis is that the combination of the GC analogue number density profile and the incomplete data at large radii are culprits for the mismatch between the predicted and true mass profiles (see Figures 5.5 and 5.6).

Our studies of the MW, and in particular the sensitivity analyses in Papers 2 and 3, suggest that the mass estimate might increase slightly if the inner GC analogues were not included in the analysis. We could perform a sensitivity analysis on the mock data, and iteratively remove the inner GC analogues



(a) *g15784*



(b) *g1536*

Figure 5.9: Solid lines show the GC analogue number density profiles for the subsampled data from MUGS2 galaxies (a) *g15784* and (b) *g1536*. Dotted lines show the GC number density profile for the MW GC data. The points along the bottom show the positions of the individual GC analogues (top) and MW GCs (bottom).

from the sample to find out how the mass estimate and mass profile prediction would change. However, as we have seen, the number density profile of the GC analogues does seem to follow a single power law at large radii. Therefore, perhaps the incomplete data at large radii are the most important piece of the puzzle.

To investigate this idea, we plan to re-run the analysis, gradually increasing the number of proper motion measurements at large distances. Perhaps we can answer the question: *Will the predictive power of the mass profile improve with the inclusion of complete data at large distances?*

5.6 Preliminary Conclusions

We have applied the hierarchical Bayesian mass estimation technique presented in Eadie et al. (2017b,a) to mock data from eight MUGS2 hydrodynamical Milky Way-type galaxies (Keller et al., 2015, 2016). First we analysed galaxies *g15784* and *g1536* in detail; while the virial radii and virial masses of these two galaxies were well estimated, the mass profiles were underestimated at small radii. The hierarchical method applied to the other six MUGS2 galaxies showed similar results, in that the mass tended to be underestimated. These results warrant further investigation and a more comprehensive statistical analysis. Because we plan to extend GME to other applications (e.g. other galaxies) in the long term, it is imperative to investigate how our choice of tracer and physical model might bias results. At this point no strong conclusions can be made, but we can cautiously say the method may tend underestimate the virial mass, at least for this small sample of unregulated MUGS2 galaxies.

An important caveat to the results in Section 5.4, is that only a single random sample of GC analogues from each MUGS2 galaxy was investigated. The number of GC analogues in each MUGS2 galaxy is usually larger; for example, the total number of GC analogues in *g15784* is 2381, and in *g1536* is 311. An interesting statistical test would be to repeat the analysis of Section 5.4 on multiple random samples of these galaxies, in order to fully understand the reliability of the mass and mass profile estimates. Another caveat to our results is that the GC analogues may not be representative of a GC population in a MW-type galaxy, insofar as the MW is typical for one of its shape, size, and mass. Finally, the most important caveat is that seven out of eight galaxies we chose to analyse were *unregulated* ones. Why did the unregulated galaxies have a spatial distribution of GC analogues most similar to the MW's GC population, when these galaxies are supposedly least like the MW? The answer is unclear, and this question also brings into doubt the reliability of treating star tracers in the MUGS2 simulation as GC analogues.

Nevertheless, the results of this preliminary study encourage us to pursue our investigations of these simulated galaxies. A more thorough analysis involving repeated sampling of the MUGS2 data will provide us with a better understanding of both the model choice and the method. Furthermore, by increasing the number of complete measurements at larger radii, we will be able to investigate how well the model predicts the mass profile in the presence of more complete data. And of course, analysing the *regulated* galaxies from MUGS2 will also enable a more well-rounded interpretation of the present results.

Another area to investigate is that of tracer choice. In our studies thus far, we have been using the MW GC population as a testbed for the hierarchical

Bayesian method. One way to investigate how the tracer choice (GCs) affects the mass estimate is to apply our hierarchical Bayesian method to other tracer populations in the MW halo, such as Blue Horizontal Branch (BHB) stars or dwarf galaxies (DGs). While this would allow for comparison between mass estimates from different tracers, at the same time it would not tell us which mass estimate is best (if they were different) nor if the physical model we assume is a good predictor of true Galactic mass. Interpreting and comparing results from such a study would be quite model dependent, and rely on some conjecture. Therefore, using different tracers from the MUGS2 simulations might be a fruitful path to pursue, first.

There are many high-performance computer simulations that produce MW-type galaxies, and it would be interesting to test our method with the power-law models on more than the MUGS2 simulated galaxies. In particular, data from the *Apostle*, *Aquarius*, *Eagle*, *Fire*, *Illustris*, and *Latte* (Sawala et al., 2016; Springel et al., 2008; Schaye et al., 2015; Hopkins et al., 2014; Vogelsberger et al., 2014; Wetzel et al., 2016) projects would all make interesting candidates.

To compare the viability of different galaxy model assumptions within our hierarchical framework, we could perform Bayesian model comparison tests (e.g. the Bayes' factor Jeffrey, 1939). However, this is complicated by the shortage of analytic distribution functions (DFs) for galaxy models. Analytic DFs are required in the current set-up of our hierarchical Bayesian framework. Non-analytic models might be possible with Approximate Bayesian Computation (ABC) or "Forward Modeling", but this would involve substantial overhaul of the hierarchical code.

We should not immediately discount the idea that the galaxy model employed here, although simple, may still be a good predictor of the Galaxy's mass and mass profile if we can understand how best to use it. If this is the case, then it would be a favourable alternative to computationally heavy methods like ABC for computing the mass of the MW (and in the future, other galaxies), especially with the deluge of data coming from Gaia and LSST in the near future.

Bibliography

- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics*, 2nd edn. (Princeton)
- Boylan-Kolchin, M., Besla, G., & Hernquist, L. 2011, *MNRAS*, 414, 1560
- Busha, M. T., Marshall, P. J., Wechsler, R. H., Klypin, A., & Primack, J. 2011, *ApJ*, 743, 40
- Deason, A. J., Belokurov, V., & Evans, N. W. 2011, *MNRAS*, 411, 1480
- Deason, A. J., Belokurov, V., Evans, N. W., et al. 2012, *MNRAS*, 425, 2840
- Dehnen, W., McLaughlin, D. E., & Sachania, J. 2006, *MNRAS*, 369, 1688
- Diaz, J. D., Koposov, S. E., Irwin, M., Belokurov, V., & Evans, N. W. 2014, *MNRAS*, 443, 1688
- Eadie, G., & Harris, W. 2016, *ApJ*, 829, 108
- Eadie, G., Harris, W., & Widrow, L. 2015, *ApJ*, 806, 54
- Eadie, G., Springford, A., & Harris, W. 2017a, *ApJ*, 838, 76
- Eadie, G. M., Springford, A., & Harris, W. E. 2017b, *ApJ*, 835, 167
- ESA. 2016, Data Release Scenario, <http://www.cosmos.esa.int/web/gaia/release>, accessed: 2016-05-10

- Evans, N. W., Hafner, R. M., & de Zeeuw, P. T. 1997, MNRAS, 286, 315
- Gibbons, S. L. J., Belokurov, V., & Evans, N. W. 2014, MNRAS, 445, 3788
- Harris, W. E. 1996, AJ, 112, 1487
- . 2010, ArXiv e-prints, arXiv:1012.3224
- Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581
- Jeffrey, H. 1939, Theory of Probability (Oxford University Press)
- Johnson, D. R. H., & Soderblom, D. R. 1987, The Astronomical Journal, 93, 864
- Kahn, F. D., & Woltjer, L. 1959, ApJ, 130, 705
- Keller, B. W., Wadsley, J., & Couchman, H. M. P. 2015, MNRAS, 453, 3499
- . 2016, MNRAS, 463, 1431
- Law, D. R., & Majewski, S. R. 2010, ApJ, 714, 229
- Little, B., & Tremaine, S. 1987, ApJ, 320, 493
- LSST-Corporation. 2016, LSST Project Schedule, <https://www.lsst.org>, accessed: 2016-08-10
- McMillan, P. J. 2017, MNRAS, 465, 76
- Patel, E., Besla, G., & Mandel, K. 2017, MNRAS, 468, 3428
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

- Sakamoto, T., Chiba, M., & Beers, T. C. 2003, *A&A*, 397, 899
- Sawala, T., Frenk, C. S., Fattahi, A., et al. 2016, *MNRAS*, 457, 1931
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, 403, 1829
- Springel, V., Wang, J., Vogelsberger, M., et al. 2008, *MNRAS*, 391, 1685
- Stinson, G. S., Bailin, J., Couchman, H., et al. 2010, *MNRAS*, 408, 812
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *MNRAS*, 444, 1518
- Wadsley, J., Stadel, J., & Quinn, T. 2004, *New Astronomy*, 9, 137
- Wadsley, J. W., Keller, B. W., & Quinn, T. R. 2017, *Monthly Notices of the Royal Astronomical Society*, 471, 2357
- Wetzell, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, *ApJL*, 827, L23
- Wilkinson, M., & Evans, N. 1999, *MNRAS*, 310
- Xue, X., Rix, H., Zhao, G., et al. 2008, *ApJ*, 684, 1143

6

Summary & Future Work

I have presented the development of a hierarchical Bayesian model for measuring the mass and cumulative mass profile of the Milky Way Galaxy, as well as its application to simple simulated data, real data, and mock data from hydrodynamic, cosmological simulations. The immediate goal of this work has been to develop a thorough statistical framework for measuring the mass of the MW using the kinematic information of Galactic satellites. In particular, I have sought to overcome some of the major challenges associated with using tracer data that were discussed in Chapter 1. Another important goal has been to make the method accessible to those who are not familiar with Bayesian analysis, and I hope that I have succeeded in this task, especially with the introduction to Bayesian inference in Chapter 2.

The series of investigations performed in Chapter 3 highlighted some features of the physical model when applied to MW GC data. For example, it was seen that the mass estimate increased slightly as inner GCs were removed from the analysis, but that the uncertainties also increased significantly. The mass estimates, despite this change, agree at the level of the 95% credible regions even when numerous inner GCs are removed. We also noted that the inclusion of complete and incomplete data simultaneously seemed to remove

any mass-anisotropy degeneracy; mass estimates were in agreement with each other within the 50% credible regions, regardless of whether $\beta = 0$, $\beta = 0.5$, or β was a free parameter. Also, any slight disagreement in the mass due to β appears to increase as the proportion of proper motions decreased, as expected. This result exemplifies why obtaining proper motions for tracers at large Galactic radii is so important.

Chapter 4 brings the biggest improvement to our method thus far— including measurement uncertainties in a hierarchical Bayesian paradigm. Treating the positions and velocities of the GCs as parameters in the model not only reduces the influence of highly uncertain data, but also allows us to analyse the specific energy profile of the GCs, which can provide useful insight. Our best estimate of the mass and mass profile of the MW using the hierarchical Bayesian method are shown in Appendix A. I repeat the virial mass and virial radius with the 95% credible intervals here, for convenience: $M_{vir} = 0.87 (0.67, 1.09) \times 10^{12} M_{\odot}$ and $r_{vir} = 201 (184, 217)$ kpc. This mass estimate is in agreement with many of the studies discussed in Chapter 1, albeit the range of mass estimates in the literature is wide. It also is in agreement with two of the most recent studies of the MW’s virial mass, which found $M_{vir} = 1.02^{+0.77}_{-0.55} \times 10^{12} M_{\odot}$ (Patel et al., 2017) and $M_{vir} = 1.30(\pm 0.30) \times 10^{12} M_{\odot}$ (McMillan, 2017).

The results from blind tests performed on mock data from the hydrodynamical and cosmologically motivated MUGS2 simulations are compelling and warrant further investigation (Chapter 5). The 95% Bayesian credible regions for the virial radius contained the true value in most cases, and the true virial mass was contained within the same regions for all but three galaxies. The mass profiles, on the other hand, were underestimated at small radii for the

unregulated galaxies, and overestimated at small radii for $g1536$. Because all but $g1536$ were unregulated galaxies from MUGS2, we cannot make strong conclusions about our method's predictive power for the MW.

An important next step is to understand whether the physical model poorly described the mass profiles because of random or systematic effects. If the underestimation in the mass profile is a true systematic bias, then we can hope to understand it and develop a way to correct for it. Alternatively, the underestimation could be a chance occurrence based on the subsample of GC analogues that were taken from the simulated galaxies. To fully comprehend the behaviour of the model as it is confronted with GC analogues, independent analyses of multiple subsamples of the MUGS2 data are being performed. Regardless of the mass profile discrepancies, the possibility that the simple power-law profile model for the Galactic potential can consistently recover the *virial mass* for these simulated galaxies is encouraging.

In addition to the work presented in Chapter 5 and discussed above, I will also use the MUGS2 data to study how the mass estimate changes as more proper motions are included at larger distances. I may also compare the distribution of the current GC analogues to MW metal-poor halo star populations in a quantitative manner, and investigate the possibility of using different definitions to find GC analogues and/or halo star analogues in the MUGS2 data.

The blind tests in Chapter 5 highlight the importance of high-resolution simulations for testing new Galactic mass estimation methods, insofar as they can describe nature. Ideally, actual GCs within galactic simulations would make better mock data than GC analogues (not just for this study, but for many other areas of astronomy as well). However, creating realistic GCs within

a cosmological simulations requires much higher resolution that is not yet achievable (although some simulations are getting close, e.g. Kim et al., 2017).

Overall, the hierarchical Bayesian method provides a probabilistic framework to estimate the mass and mass profile of the MW, includes all available tracer kinematic data, and incorporates measurement uncertainties for all known positions and velocities. From the posterior distribution, estimates and probability regions for meaningful physical quantities can be derived, and a mass profile with probability regions can be calculated to compare mass estimates with other studies.

6.1 Future Work

The method I have developed for measuring the mass of the MW could be improved in a variety of ways. For example, one could

- derive a new DF that includes a gravitational potential for the disk and bulge component of the Galaxy,
- include a shape parameter that allows for non-spherical dark matter halos (e.g. Deason et al., 2012),
- treat the solar motion and the LSR as parameters in the model (stellar stream studies have done this with success, e.g. Küpper et al., 2015),
- design a way to allow satellites to be unbound in the model, with some probability, so that statements may be made about whether or not a tracer is bound to the MW,

- incorporate a broken power-law for stellar tracers as done by Kafle et al. (2012), and
- allow β to vary with radius (instead of assuming a constant velocity anisotropy).

Determining which of these improvements is the most immediately important may depend on the results of our current study (Chapter 5). However, at this point it appears that the first and last items are perhaps the most important improvements to pursue. Using a DF that includes the gravitational potential for the disk and bulge components of the Galaxy will provide more model flexibility, and evidence suggests this improvement is needed to better describe the data. The sensitivity tests in Chapters 3 and 4 showed a slight increase in the mass estimate as inner GCs were removed from the analysis (although the uncertainties increased substantially), and in Chapter 5 the innermost regions of the mass profiles of the simulated galaxies were poorly recovered (although here the number of blind tests was small). The last item, allowing β to vary with radius, would provide us with a $\beta(r)$ profile for the MW, which is potentially very useful. Studies of both N-body and hydrodynamical simulations of MW-type galaxies suggest that changes in β with radius provide information about the Galaxy's accretion history (Loebman et al., 2017).

Beyond improving the method itself, one could also apply the method to other tracers and new data. The recent work by the HSTPROMO team and others is increasing the number of proper motion measurements of GCs in the MW. Some of these GCs are at large distances and are being measured relatively precisely. Re-running our analysis to include this new data may provide significant information about the dark matter halo at larger radii.

The mass estimate for the MW when dwarf galaxies were included in our analysis (Chapter 2) was larger than our more recent results (Appendix A). However, some of the DG data are highly uncertain. Now that we have improved the method to incorporate measurement uncertainty and are using a slightly more realistic model for the tracer profile, it would be useful to re-run the analysis on the dwarf galaxy data too.

Kinematic data used in other studies (e.g. Xue et al., 2008, used BHB stars) could also be analysed with our method, and the results might make for interesting comparisons. RR Lyrae, a subset of BHB stars, will also be useful tracers when their kinematic data becomes available from observational programs such as Gaia and LSST. With these new instruments, a mixture of both complete and incomplete velocity measurements will be available, and methods which can deal with both kinds of data simultaneously will be needed.

The method developed in this thesis could be adapted and applied to other mass estimation problems in astronomy. For example, one could use it to estimate masses of objects at a range of scales, including globular clusters, the Nuclear Star Cluster (NSC) of the MW, spherical dwarf galaxies, other galaxies, and possibly galaxy clusters.

Currently, I have started to investigate using GME to estimate the mass profile of the MW's NSC. By looking at the central region of the NSC, one could include the gravitational potential of the central supermassive black hole in addition to the background potential. Such a model would return estimates for not only the parameters of the NSC, but also the mass of the central supermassive black hole. Furthermore, the mass profile generated from GME could be used with the integrated light profile to constrain a mass-to-light ratio profile with Bayesian credible regions. A natural progression from application

to the NSC is to apply the method to GCs themselves— we could determine the probability that a GC contains an intermediate mass black hole at its centre.

Future work using the hierarchical Bayesian approach I have developed in this thesis is promising, both for continued studies of the MW and for other systems. Astronomy is entering an era of Big Data, and kinematic information from large-scale surveys such as Gaia and LSST are and will become available. Despite the massive data sets, data will remain incomplete and subject to measurement uncertainty. Modern statistical methods, such as the hierarchical Bayesian approach presented in this thesis, will be needed to properly interpret the data, test theoretical models, and help us decide which research paths to take next.

Bibliography

Deason, A. J., Belokurov, V., Evans, N. W., & An, J. 2012, MNRAS, 424, L44

Kaffe, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, ApJ, 761,
98

Kim, J.-h., Ma, X., Grudić, M. Y., et al. 2017, ArXiv e-prints,
arXiv:1704.02988

Küpper, A. H. W., Balbinot, E., Bonaca, A., et al. 2015, ApJ, 803, 80

Loebman, S. R., Valluri, M., Hattori, K., et al. 2017, ArXiv e-prints,
arXiv:1704.06264

McMillan, P. J. 2017, MNRAS, 465, 76

Patel, E., Besla, G., & Mandel, K. 2017, MNRAS, 468, 3428

Xue, X., Rix, H., Zhao, G., et al. 2008, ApJ, 684, 1143

Appendices

A

Erratum: “Bayesian Mass Estimates of the Milky Way: Including Measurement Uncertainties with Hierarchical Bayes” (2017, ApJ, 835, 167)

Reprinted from *Eadie, G., Springford, A., & Harris, W. (2017), THE ASTROPHYSICAL JOURNAL, Vol. 835, No. 167. DOI: 10.3847/1538 – 4357/aa64db.*

Published by the American Astronomical Society ©AAS

This paper is an erratum to the paper presented in Chapter 4.

In the analysis for the original manuscript Eadie et al. (2017), the Cartesian Galactocentric velocities (U_{gc}, V_{gc}, W_{gc}) of the globular clusters (GCs) were incorrectly transformed to the right-hand cylindrical velocities (Π, Θ, Z) (Equation (17)). The angle θ in Equation (17) is measured in the xy -plane from the positive x -axis, such that $\cos \theta = \frac{X_{gc}}{r}$ and $\sin \theta = \frac{Y_{gc}}{r}$ where X_{gc} and Y_{gc} are the Galactocentric x and y coordinates of a GC in a right-hand Cartesian system, and where r is the projected distance of the GC onto the xy -plane. In the analysis, the incorrect r values were used to calculate $\cos \theta$ and $\sin \theta$. This mistake was partly due to the differences in notation between the Caseti and Harris online GC catalogs (in the former, R_{gc} is defined as the projected distance r , while in the latter R_{gc} is the three-dimensional distance). The Harris values of R_{gc} were used as the projected distance, and thus many of the GC spatial velocities were underestimated.

We have corrected this mistake, recalculated all posterior distributions, and performed the sensitivity analysis again. A summary of the new model parameter estimates is presented in the top half of Table A.1. The estimates for the derived parameters $M_{125} = M(r < 125 \text{ kpc})$, $M_{vir} = M(r < r_{vir})$, r_{vir} , and $M_{300} = M(r < 300 \text{ kpc})$ are also calculated from the new posterior distributions of the model parameters and shown in the lower half of the table. Of the four model parameters, only the Φ_o and β estimates changed significantly; Φ_o and the derived mass estimates are higher because the velocities were previously underestimated, and β is lower because the tangential components of the velocities were the most affected. The shapes of the joint posterior distributions (Figure 7 in the original paper) did not change. The cumulative mass profile $M(r)$ is now in better agreement with previous studies (Figure A.1). The estimated energy profile is similar in shape to the original Figure 8, but

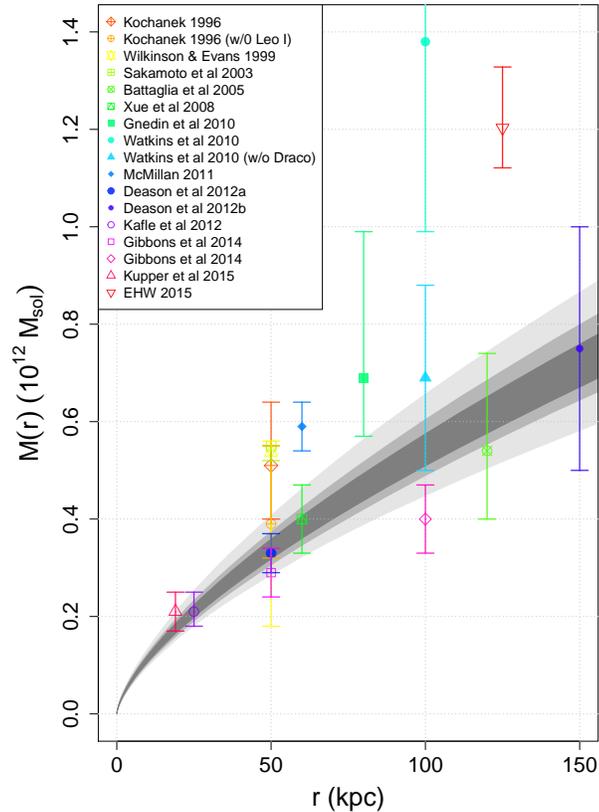


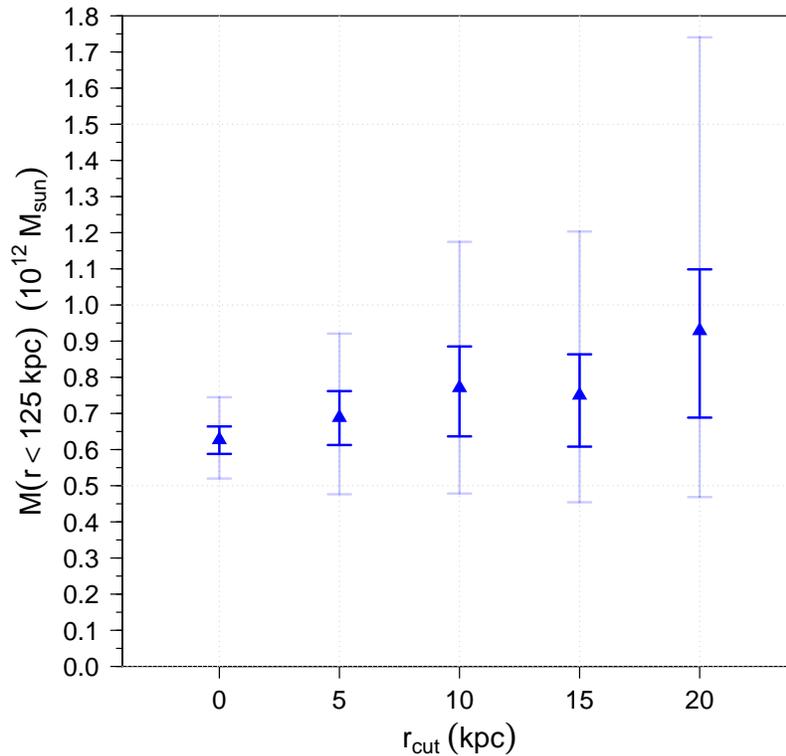
Figure A.1: Cumulative mass profile for the Milky Way as estimated with the kinematic data of 143 GCs. The grey shaded areas show the 50%, 75%, and 95% Bayesian credible regions, and the points with error bars are results from other studies. (Replaces Figures 3 and 4 in the original paper.)

is shifted toward more negative energies (Figure A.3). Our interpretation of the energy profile remains the same.

The parameter trends found in the sensitivity analysis (Figure 6 in the paper) are unchanged, despite the different values of Φ_o and β . The new estimates for β for different r_{cut} values are between 0.1 and 0.3, which still indicates a mild radial velocity anisotropy for the GC population. The trend in the M_{125} estimate is slightly stronger than that shown in Figure 5 of the original paper, albeit with large uncertainties (Figure A.2).

Table A.1: Parameter and Derived Parameter Estimates with Bayesian Marginal Credible Regions

Parameter	Units	Mean	Median	50% Cred. Region	95% Cred. Region
Φ_o	$(10^4 \text{km}^2 \text{s}^{-2})$	31.6	31.4	(29.7, 33.2)	(26.9, 37)
γ	—	0.32	0.31	(0.31, 0.33)	(0.30, 0.37)
α	—	3.05	3.05	(3.04, 3.06)	(3.03, 3.08)
β	—	0.14	0.14	(0.06, 0.21)	(-0.09, 0.34)
M_{125}	$10^{12} M_\odot$	0.63	0.63	(0.59, 0.66)	(0.52, 0.74)
M_{vir}	$10^{12} M_\odot$	0.87	0.86	(0.80, 0.94)	(0.67, 1.09)
r_{vir}	kpc	201	201	(195, 206)	(184, 217)
M_{300}	$10^{12} M_\odot$	1.14	1.14	(1.06, 1.21)	(0.92, 1.36)


 Figure A.2: The new M_{125} estimates from the sensitivity analysis. Bright and faint error bars correspond to 50% and 95% credible regions. (Replaces Figure 5 in the original paper.)

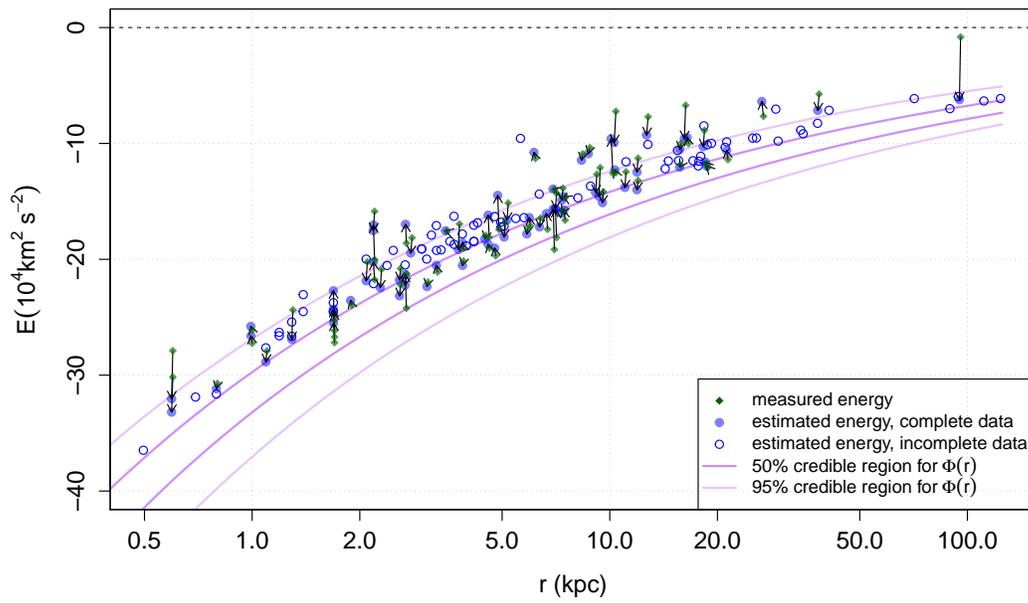


Figure A.3: Specific energy profile for GCs. (Replaces Figure 8 in the original paper.)

Bibliography

Eadie, G. M., Springford, A., & Harris, W. E. 2017, ApJ, 835, 167

Bibliography

- An, J. H., & Evans, N. W. 2006, *AJ*, 131, 782
- Baes, M., Buyle, P., Hau, G. K. T., & Dejonghe, H. 2003, *MNRAS*, 341, L44
- Baes, M., & Dejonghe, H. 2002, *Astronomy and Astrophysics*, 393, 485
- Bahcall, J. N., & Tremaine, S. 1981, *ApJ*, 244, 805
- Bailer-Jones, C. A. L. 2004, in *Proceedings of the IAU Colloquium No. 196*, ed. W. Kurtz & G. Bromage (Cambridge Journals), 429–443
- Barber, C., Starckenburg, E., Navarro, J. F., McConnachie, A. W., & Fattahi, A. 2014, *MNRAS*, 437, 959
- Bartelmann, M. 2010, *Classical and Quantum Gravity*, 27, 233001
- Battaglia, G., Helmi, A., Morrison, H., et al. 2005, *MNRAS*, 364, 433
- Bayes, M., & Price, M. 1763, *Philosophical Transactions (1683-1775)*, 370
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57
- Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006, *ApJL*, 642, L137
- Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, *ApJ Supp*, 208, 20
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics*, 2nd edn. (Princeton)

- Bolker, B. 2016, emdbook: Ecological Models and Data in R, r package version 1.3.9
- Bolker, B. M. 2008, Ecological Models and Data in R (Princeton University Press)
- Bonaca, A., Geha, M., Küpper, A. H. W., et al. 2014, ApJ, 795, 94
- Bovy, J., Bahmanyar, A., Fritz, T. K., & Kallivayalil, N. 2016, ApJ, 833, 31
- Bowden, A., Evans, N. W., & Williams, A. A. 2016, MNRAS, 460, 329
- Boylan-Kolchin, M., Besla, G., & Hernquist, L. 2011a, MNRAS, 414, 1560
- Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011b, MNRAS, 415, L40
- Boylan-Kolchin, M., Bullock, J. S., Sohn, S. T., Besla, G., & van der Marel, R. P. 2013, ApJ, 768, 140
- Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150
- Busha, M. T., Marshall, P. J., Wechsler, R. H., Klypin, A., & Primack, J. 2011, ApJ, 743, 40
- Carlesi, E., Hoffman, Y., Sorce, J. G., & Gottlöber, S. 2017, MNRAS, 465, 4886
- Casetti-Dinescu, D. I., Girard, T. M., Jílková, L., et al. 2013, AJ, 146, 33
- Casetti-Dinescu, D. I., Girard, T. M., Korchagin, V. I., van Altena, W. F., & López, C. E. 2010, AJ, 140, 1282

Cautun, M., Frenk, C. S., van de Weygaert, R., Hellwing, W. A., & Jones, B. J. T. 2014, MNRAS, 445, 2049

Cioni, M.-R. L., Bekki, K., Girardi, L., et al. 2016, A&A, 586, A77

Ciotti, L. 1996, ApJ, 471, 68

Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, ApJL, 648, L109

Cole, S., & Lacey, C. 1996, MNRAS, 281, 716

Côté, P., McLaughlin, D. E., Cohen, J. G., & Blakeslee, J. P. 2003, ApJ, 591, 850

Courteau, S., Cappellari, M., de Jong, R., et al. 2014, Rev. Mod. Phys., 86, 47

Cuddeford, P. 1991, MNRAS, 253, 414

Cunningham, E. C., Deason, A. J., Guhathakurta, P., et al. 2016, ApJ, 820, 18

de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247

Deason, A. J., Belokurov, V., & Evans, N. W. 2011, MNRAS, 411, 1480

Deason, A. J., Belokurov, V., Evans, N. W., & An, J. 2012a, MNRAS, 424, L44

Deason, A. J., Belokurov, V., Evans, N. W., & Johnston, K. V. 2013a, ApJ, 763, 113

Deason, A. J., Belokurov, V., Evans, N. W., & McCarthy, I. G. 2012b, ApJ, 748, 2

- Deason, A. J., Van der Marel, R. P., Guhathakurta, P., Sohn, S. T., & Brown, T. M. 2013b, *ApJ*, 766, 24
- Deason, A. J., Belokurov, V., Evans, N. W., et al. 2012c, *MNRAS*, 425, 2840
- Deg, N., & Widrow, L. 2013, *MNRAS*, 428, 912
- Dehnen, W., McLaughlin, D. E., & Sachania, J. 2006, *MNRAS*, 369, 1688
- Diaz, J. D., Koposov, S. E., Irwin, M., Belokurov, V., & Evans, N. W. 2014, *MNRAS*, 443, 1688
- Dierickx, M. I. P., & Loeb, A. 2017, *ArXiv e-prints*, arXiv:1703.02137
- Dinescu, D. I., Girard, T. M., & van Altena, W. F. 1999, *AJ*, 117, 1792
- Dinescu, D. I., Keeney, B. A., Majewski, S. R., & Girard, T. M. 2004, *AJ*, 128, 687
- Dinescu, D. I., Martínez-Delgado, D., Girard, T. M., et al. 2005, *ApJL*, 631, L49
- Djorgovski, S., & Meylan, G. 1994, *AJ*, 108, 1292
- Eadie, G., & Harris, W. 2016, *ApJ*, 829, 108
- Eadie, G., Harris, W., & Springford, A. 2015a, in *Joint Statistical Meetings Proceedings (American Statistical Association)*
- Eadie, G., Harris, W., & Widrow, L. 2015b, *ApJ*, 806, 54
- Eadie, G., Springford, A., & Harris, W. 2017a, *ApJ*, 838, 76
- Eadie, G. M. 2013, Master's thesis, Queen's University, Kingston, Ontario, Canada

- Eadie, G. M., Harris, W. E., Widrow, L. M., & Springford, A. 2015c, in Proceedings of the International Astronomical Union, ed. A. Bragaglia, M. Arnaboldi, M. Rejkuba, & D. Romano (Cambridge Journals)
- Eadie, G. M., Springford, A., & Harris, W. E. 2017b, ApJ, 835, 167
- Ehlers, J. 2008, General Relativity and Gravitation, 41, 203
- ESA. 2016, Data Release Scenario, <http://www.cosmos.esa.int/web/gaia/release>, accessed: 2016-05-10
- Evans, N. W., & An, J. H. 2006, Physical Review D, 73
- Evans, N. W., Hafner, R. M., & de Zeeuw, P. T. 1997, MNRAS, 286, 315
- Evans, N. W., & Williams, A. A. 2014, MNRAS, 443, 791
- Eyre, A., & Binney, J. 2011, MNRAS, 413, 1852
- Fattahi, A., Navarro, J. F., Sawala, T., et al. 2016, MNRAS, 457, 844
- Feigelson, E., & Babu, G. 2012, Modern Statistical Methods for Astronomy: With R Applications (Cambridge University Press)
- Feltzing, S., & Johnson, R. A. 2002, A&A, 385, 67
- Ferrarese, L. 2002, ApJ, 578, 90
- Fragione, G., & Loeb, A. 2017, New Astronomy, 55, 32
- Fritz, T. K., & Kallivayalil, N. 2015, ApJ, 811, 123
- Fritz, T. K., Linden, S. T., Zivick, P., et al. 2017, ApJ, 840, 30

Gaia Collaboration, van Leeuwen, F., Vallenari, A., et al. 2017, *A&A*, 601, A19

Gelfand, A. E., & Smith, A. F. 1990, *Journal of the American statistical association*, 85, 398

Gelman, A., Carlin, J., Stern, H., & Rubin, D. 2003, *Bayesian data analysis* (Chapman & Hall/CRC)

Gelman, A., & Rubin, D. 1992, *Statistical science*, 7, 457

Geman, S., & Geman, D. 1984, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 6, 721

Gibbons, S. L. J., Belokurov, V., & Evans, N. W. 2014, *MNRAS*, 445, 3788

Gnedin, O. Y., Brown, W. R., Geller, M. J., & Kenyon, S. J. 2010, *ApJL*, 720, L108

Gnedin, O. Y., Gould, A., Miralda-Escudé, J., & Zentner, A. R. 2005, *ApJ*, 634, 344

Goodman, J., & Weare, J. 2010, *Communications in applied mathematics and computational science*, 5, 65

Gove, W. R., & Hughes, M. 1980, *Social Forces*, 58, 1157

Guo, Q., White, S., Li, C., & Boylan-Kolchin, M. 2010, *MNRAS*, 404, 1111

Harris, W. E. 1976, *AJ*, 81, 1095

—. 1996a, *AJ*, 112, 1487

—. 1996b, *VizieR Online Data Catalog*, 7195, 0

- . 2001, *Globular Cluster Systems*, Vol. 28 (Springer-Verlag Berlin Heidelberg), 223
- . 2010, ArXiv e-prints, arXiv:1012.3224
- Hartwick, F. D. A., & Sargent, W. L. W. 1978, *ApJ*, 221, 512
- Helmi, A. 2004, *MNRAS*, 351, 643
- Hernquist, L. 1990, *ApJ*, 356, 359
- Hills, J. G. 1988, *Nature*, 331, 687
- Howlander, H. A., Hossain, A. M., & Makhnin, O. 2007, *International Journal of Statistical Sciences*, 6, 137
- Huang, Y., Liu, X.-W., Yuan, H.-B., et al. 2016, *MNRAS*, 463, 2623
- Humphreys, R. M., & Larsen, J. A. 1995, *AJ*, 110, 2183
- Ibata, R., Irwin, M., Lewis, G. F., & Stolte, A. 2001a, *ApJL*, 547, L133
- Ibata, R., Lewis, G. F., Irwin, M., Totten, E., & Quinn, T. 2001b, *ApJ*, 551, 294
- Ibata, R., Martin, N. F., Irwin, M., et al. 2007, *ApJ*, 671, 1591
- Ibata, R. A., Wyse, R. F. G., Gilmore, G., Irwin, M. J., & Suntzeff, N. B. 1997, *AJ*, 113, 634
- Jaffe, W. 1983, *MNRAS*, 202, 995
- Jeffrey, H. 1939, *Theory of Probability* (Oxford University Press)

- Johnson, D. R. H., & Soderblom, D. R. 1987, *The Astronomical Journal*, 93, 864
- Kafle, P. R., Sharma, S., Lewis, G. F., & Bland-Hawthorn, J. 2012, *ApJ*, 761, 98
- . 2014, *ApJ*, 794, 59
- Kahn, F. D., & Woltjer, L. 1959, *ApJ*, 130, 705
- Keller, B. W., Wadsley, J., & Couchman, H. M. P. 2015, *MNRAS*, 453, 3499
- . 2016, *MNRAS*, 463, 1431
- Kim, J.-h., Ma, X., Grudić, M. Y., et al. 2017, *ArXiv e-prints*, arXiv:1704.02988
- Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, *ApJ*, 522, 82
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, 740, 102
- Kochanek, C. S. 1996, *ApJ*, 457, 228
- Koposov, S. E., Rix, H.-W., & Hogg, D. W. 2010, *ApJ*, 712, 260
- Kuhlen, M., Diemand, J., & Madau, P. 2007, *ApJ*, 671, 1135
- Kulessa, A. S., & Lynden-Bell, D. 1992, *MNRAS*, 255, 105
- Küpper, A. H. W., Balbinot, E., Bonaca, A., et al. 2015, *ApJ*, 803, 80
- Law, D. R., & Majewski, S. R. 2010a, *ApJ*, 718, 1128
- . 2010b, *ApJ*, 714, 229

- Leonard, P. J. T., & Tremaine, S. 1990, *ApJ*, 353, 486
- Li, Y.-S., & White, S. D. M. 2008, *MNRAS*, 384, 1459
- Lindegren, L., Lammers, U., Bastian, U., et al. 2016, *A&A*, 595, A4
- Little, B., & Tremaine, S. 1987, *ApJ*, 320, 493
- Loebman, S. R., Ivezić, Ž., Quinn, T. R., et al. 2014, *ApJ*, 794, 151
- Loebman, S. R., Valluri, M., Hattori, K., et al. 2017, *ArXiv e-prints*,
arXiv:1704.06264
- Lokas, E. L. 2009, *MNRAS*, 394, L102
- LSST-Corporation. 2016, LSST Project Schedule, <https://www.lsst.org>, accessed: 2016-08-10
- Lynden-Bell, D., Cannon, R. D., & Godwin, P. J. 1983, *MNRAS*, 204
- Lynden-Bell, D., & Frenk, C. S. 1981, *The Observatory*, 101, 200
- Lynds, R., & Petrosian, V. 1989, *ApJ*, 336, 1
- Majewski, S. R., & Cudworth, K. M. 1993, *PASP*, 105, 987
- Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003,
ApJ, 599, 1082
- Markevitch, M., Gonzalez, A. H., Clowe, D., et al. 2004, *ApJ*, 606, 819
- Massari, D., Bellini, A., Ferraro, F. R., et al. 2013, *ApJ*, 779, 81
- Massari, D., Posti, L., Helmi, A., Fiorentino, G., & Tolstoy, E. 2017, *A&A*,
598, L9

- McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2009, *Nature*, 461, 66
- McMillan, P. J. 2011, *MNRAS*, 414, 2446
- McMillan, P. J. 2017, *MNRAS*, 465, 76
- Merritt, D. 1985, *Astronomical Journal*, 90, 1027
- Merritt, D., Graham, A. W., Moore, B., Diemand, J., & Terzić, B. 2006, *AJ*, 132, 2685
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *The journal of chemical physics*, 21, 1087
- Metropolis, N., & Ulam, S. 1949, *Journal of the American Statistical Association*, 44, 335
- Michalik, D., Lindegren, L., & Hobbs, D. 2015, *A&A*, 574, A115
- Milgrom, M. 1983, *ApJ*, 270, 365
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Nelson, D., Pillepich, A., Genel, S., et al. 2015, *Astronomy and Computing*, 13, 12
- Olszewski, E. W., Aaronson, M., & Peterson, R. C. 1986, *ApJL*, 302, L45
- Osipkov, L. P. 1979, *Soviet Astron. Letters*, 5, 42
- Patel, E., Besla, G., & Mandel, K. 2017, *MNRAS*, 468, 3428
- Paturel, G., Petit, C., Prugniel, P., et al. 2003, *A&A*, 412, 45
- Peñarrubia, J., Koposov, S. E., & Walker, M. G. 2012, *ApJ*, 760, 2

- Peñarrubia, J., Ma, Y.-Z., Walker, M. G., & McConnachie, A. 2014, MNRAS, 443, 2204
- Pearson, S., Küpper, A. H. W., Johnston, K. V., & Price-Whelan, A. M. 2015, ApJ, 799, 28
- Perets, H. B. 2009, ApJ, 690, 795
- Perets, H. B., Wu, X., Zhao, H. S., et al. 2009, ApJ, 697, 2096
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
- Piantadosi, S., Byar, D. P., & Green, S. B. 1988, American journal of epidemiology, 127, 893
- Piatek, S., Pryor, C., & Olszewski, E. W. 2016, AJ, 152, 166
- Piffl, T., Williams, M., & Steinmetz, M. 2011, A&A, 535, A70
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
- Plummer, M., Best, N., Cowles, K., & Vines, K. 2006, R News, 6, 7
- Price-Whelan, A. M., Hogg, D. W., Johnston, K. V., & Hendel, D. 2014, ApJ, 794, 4
- R Development Core Team. 2012, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0
- Robinson, W. S. 1950, American Sociological Review, 15, 351
- Roos, M. 2010, ArXiv e-prints, arXiv:1001.0316

- Rossi, L. J., Ortolani, S., Barbuy, B., Bica, E., & Bonfanti, A. 2015, MNRAS, 450, 3270
- Rubin, V. C., Ford, W. K. J., & Thonnard, N. 1980, ApJ, 238, 471
- Sakamoto, T., Chiba, M., & Beers, T. C. 2003, A&A, 397, 899
- Sanders, J. L. 2014, MNRAS, 443, 423
- Sanderson, R. E., Hartke, J., & Helmi, A. 2017, ApJ, 836, 234
- Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829
- Sersic, J. L. 1968, Cordoba, Argentina: Observatorio Astronomico, 1968, 1
- Simon, J. D., & Geha, M. 2007, ApJ, 670, 313
- Smith, M. C., Wyn Evans, N., & An, J. H. 2009, ApJ, 698, 1110
- Smith, M. C., Ruchti, G. R., Helmi, A., et al. 2007, MNRAS, 379, 755
- Sohn, S. T., Anderson, J., & van der Marel, R. P. 2012, ApJ, 753, 7
- Sohn, S. T., Besla, G., van der Marel, R. P., et al. 2013, ApJ, 768, 139
- Soucail, G., Fort, B., Mellier, Y., & Picat, J. P. 1987, A&A, 172, L14
- Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJ Supp, 148, 175
- Springel, V., & Hernquist, L. 2003, MNRAS, 339, 312
- Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685
- Tanner, M. A., & Wong, W. H. 1987, Journal of the American statistical Association, 82, 528

- Tierney, L., Rossini, A. J., Li, N., & Sevcikova, H. 2013, snow: Simple Network of Workstations, r package version 0.3-13
- van der Marel, R. P. 2016, Hubble Space Telescope Proper Motion Collaboration, <http://www/stsci.edu/~marel/hstpromo.html>, accessed: 2016-10-16
- van der Marel, R. P., Besla, G., Cox, T. J., Sohn, S. T., & Anderson, J. 2012a, *ApJ*, 753, 9
- van der Marel, R. P., Fardal, M., Besla, G., et al. 2012b, *ApJ*, 753, 8
- van der Marel, R. P., Anderson, J., Bellini, A., et al. 2014, in *Astronomical Society of the Pacific Conference Series*, Vol. 480, *Structure and Dynamics of Disk Galaxies*, ed. M. S. Seigar & P. Treuhardt, 43
- Varghese, A., Ibata, R., & Lewis, G. F. 2011, *MNRAS*, 417, 198
- Vera-Ciro, C. A., Sales, L. V., Helmi, A., et al. 2011, *MNRAS*, 416, 1377
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, *MNRAS*, 444, 1518
- . 2014b, *Nature*, 509, 177
- Wadsley, J., Stadel, J., & Quinn, T. 2004, *New Astronomy*, 9, 137
- Wadsley, J. W., Keller, B. W., & Quinn, T. R. 2017, *Monthly Notices of the Royal Astronomical Society*, 471, 2357
- Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2006, *ApJL*, 642, L41
- Wang, J., Chen, L., & Chen, D. 2005, *Chinese Astronomy and Astrophysics*, 29, 386
- Wang, W., Han, J., Cooper, A., et al. 2015, *MNRAS*, 453, 377

- Watkins, L., Evans, N., & An, J. 2010, MNRAS, 406, 264
- Watkins, L. L., & van der Marel, R. P. 2017, ApJ, 839, 89
- Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, ApJL, 827, L23
- White, M. 2001, A&A, 367, 27
- Widrow, L. M. 2000, ApJ Supp, 131, 39
- Widrow, L. M., & Dubinski, J. 2005, ApJ, 631, 838
- Widrow, L. M., Perrett, K. M., & Suyu, S. H. 2003, ApJ, 588, 311
- Widrow, L. M., Pym, B., & Dubinski, J. 2008, ApJ, 679, 1239
- Wilkinson, M., & Evans, N. 1999, MNRAS, 310
- Wilkinson, M. I., Evans, N. W., Perrett, K. M., & Bridges, T. M. 2003, ApJ, 583, 752
- Williams, A. A., & Evans, N. W. 2015, MNRAS, 454, 698
- Xue, X., Rix, H., Zhao, G., et al. 2008, ApJ, 684, 1143
- Yanny, B., Newberg, H. J., Kent, S., et al. 2000, ApJ, 540, 825
- Yencho, B. M., Johnston, K. V., Bullock, J. S., & Rhode, K. L. 2006, ApJ, 643, 154
- Yu, Q., & Madau, P. 2007, Monthly Notices of the Royal Astronomical Society, 379, 1293
- Zaritsky, D., Olszewski, E. W., Schommer, R. A., Peterson, R. C., & Aaronson, M. 1989, ApJ, 345, 759

Zoccali, M., Renzini, A., Ortolani, S., Bica, E., & Barbuy, B. 2001, *AJ*, 121,
2638

Zwicky, F. 1933, *Helvetica Physica Acta*, 6, 110

—. 1937, *ApJ*, 86, 217