Estimating the Milky Way’s Mass via Hierarchical Bayes: A Blind Test on MUGS2 Simulated Galaxies

Gwendolyn Eadie1,2,3, Benjamin Keller4, and William E. Harris5

1 Department of Astronomy, University of Washington, Seattle, WA, USA; eadieg@uw.edu
2 eScience Institute, University of Washington, Seattle, WA, USA
3 DIRAC Institute, University of Washington, Seattle, WA, USA
4 Astronomisches Rechen-Institut, Universitat Heidelberg, Heidelberg, Germany; benjamin.keller@uni-heidelberg.de
5 Department of Physics & Astronomy, McMaster University, Hamilton, Canada; harris@physics.mcmaster.ca

Received 2018 May 2; revised 2018 August 16; accepted 2018 August 18; published 2018 September 24

Abstract

In a series of three papers, Eadie et al. developed a hierarchical Bayesian method to estimate the Milky Way Galaxy’s mass given a physical model for the potential, a measurement model, and kinematic data of test particles such as globular clusters (GCs) or halo stars in the Galaxy’s halo. The Galaxy’s virial mass was found to have a 95% Bayesian credible region (c.r.) of $(0.67, 1.09) \times 10^{12} M_\odot$. In the present study, we test the hierarchical Bayesian method against simulated galaxies created in the McMaster Unbiased Galaxy Simulations 2 (MUGS2), for which the true mass is known. We estimate the masses of MUGS2 galaxies using GC analogs from the simulations as tracers. The analysis, completed as a blind test, recovers the true $M_{200}$ of the MUGS2 galaxies within 95% Bayesian c.r. in 8 out of 18 cases. Of the 10 galaxy masses that were not recovered within the 95% c.r., a large subset have posterior distributions that occupy extreme ends of the parameter space allowed by the priors. A few incorrect mass estimates are explained by the exceptional evolution history of the galaxies. We also find evidence that the model cannot describe both the galaxies’ inner and outer structure simultaneously in some cases. After removing the GC analogs associated with the galactic disks, the true masses were found more reliably (13 out of 18 were predicted within the c.r.). Finally, we discuss how representative the GC analogs are of the real GC population in the Milky Way.

Key words: clusters: general – Galaxy: fundamental parameters – globular clusters: general – Galaxy: halo – Galaxy: kinematics and dynamics – Galaxy: structure – methods: statistical

1. Introduction

The total mass of the Milky Way (MW) Galaxy is not known within a factor of two (see Figure 1 in Wang et al. 2015, for a dramatic illustration of $M_{200}$ values). This is both unfortunate and problematic, because our Galaxy’s mass is a fundamental quantity in many areas of astrophysics, astronomy, and cosmology. To complicate matters, mass estimates tend to be reported differently, making comparisons difficult. For example, some studies report a mass within a specific distance from the Galactic center, whereas others report a virial mass or $M_{200}$ based on cosmological parameters.

The wide range of mass estimates in the literature is the result of two major factors: method choice and data selection. Inference methods take a variety of approaches, such as the timing argument (e.g., Kahn & Woltjer 1959), the use of kinematic tracers to infer the gravitational potential (e.g., Little & Tremaine 1987; Wilkinson & Evans 1999; Sakamoto et al. 2003; Dehnen et al. 2006; Xue et al. 2008; Law & Majewski 2010; Watkins et al. 2010; Deason et al. 2012, 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). Different studies rely on different types of data and on different physical assumptions, making disparities between results difficult to interpret.

While each method has its own merits, the most popular approaches continue to use the kinematics of tracers (e.g., globular clusters (GCs), halo stars, and stellar streams) to constrain the MW’s gravitational potential, and thus its total mass. Since the Gaia Data Release 2 on 2018 April 25, a number of studies have already estimated the Galaxy’s mass using GC and stellar dynamics from the Gaia data (e.g., Fritz et al. 2018; Posti & Helmi 2018; Watkins et al. 2018, and G. Eadie 2018, in preparation).

Using kinematic tracers has advantages, but it also presents challenges. Different types of tracers are available, and their kinematic data suffer from incomplete velocity measurements—usually because only the line-of-sight velocity component is known. There are also differing degrees of measurement uncertainties. Incomplete velocity measurements hinder our understanding of the tracer population’s velocity anisotropy, which has been shown to influence mass estimates. Moreover, it is common practice to include or exclude data based on their (in)completeness. Which type of tracer and what components of the data researchers choose to include or exclude may contribute to the overall uncertainty in the MW’s mass.

Thankfully, the situation is improving. Data from the Gaia satellite (Perrym et al. 2001)6 and the Large Synoptic Survey Telescope (LSST)7 have and will greatly increase the number of kinematic tracers (e.g., RR Lyrae stars in the Galactic halo) and help overcome some challenges. The number of measurements for kinematic tracers will increase with these programs, especially with Gaia’s ability to measure proper motions and parallaxes of individual halo stars.

With these “big data” come a demand for reliable methods that use tracer information to estimate the mass of the Galaxy. Moreover, methods with the potential to incorporate more than one type of tracer population are needed. A hierarchical

6 http://www.cosmos.esa.int/web/gaia/release
7 https://www.lsst.org/
Bayesian approach is ideal in this scenario, as tracer populations are following the same overall gravitational potential, but may have differing spatial distributions.

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, called Galactic Mass Estimator (GME). GME has already been applied to the Galactic GC data (Eadie et al. 2015; Eadie & Harris 2016; Eadie et al. 2017a, 2017b, hereafter Papers 1, 2, and 3). In the last of these studies, GME provided a 95% Bayesian credible region (c.r.) for the MW’s virial mass: \((0.67, 1.09) \times 10^{12} M_\odot\), which is in agreement with several recent studies (e.g., Xue et al. 2008; Diaz et al. 2014; Gibbons et al. 2014; McMillan 2017; Patel et al. 2017). The median estimate for the virial mass of the MW was \(0.86 \times 10^{12} M_\odot\).

GME has at least three advantages over traditional mass estimation methods that use kinematic tracers: (1) incomplete and complete data are included simultaneously, (2) it uses a measurement model to account for observational uncertainty in position and velocity measurements, and (3) GME produces Bayesian c.r. for the cumulative mass profile of the Galaxy at any galactocentric radius, rather than point estimates of the total mass within a certain distance. As shown in Paper 3, the cumulative mass profile with a Bayesian c.r. makes it easy to compare our results with estimates from other studies that report the mass within different distances from the Galactic center.

In Papers 1–3, our method led to reasonable and encouraging results for the mass of the MW (see also Eadie 2017). As a next step, the hierarchical method could be extended for use with multiple tracer populations by adding another layer to the hierarchy. This is certainly a tempting avenue of research given the second \textit{Gaia} data release in 2018. By allowing for different spatial distributions for each tracer population, and assuming they follow a parameterized model for the total gravitational potential of the Galaxy, we can hope to better estimate the mass of the MW.

Before moving forward, however, it is important to address any uncertainty associated with the GME method thus far. Foremost, it remains unclear if the derived quantities from the posterior distribution correctly describe the \textit{true} total mass and cumulative mass profile of the Galaxy. In other words, we need to have a sense of how well our mass profile prediction represents the truth within the statistical uncertainties. We also want to better understand the limitations of the physical model (Section 3), and to be able to recognize when the model has gone awry in light of the data.

Therefore, a natural step is 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 blind tests on simulated observations of GC analogs within galaxies created by the McMaster Unbiased Galaxy Simulations 2 (MUGS2) project (see Keller et al. 2015, 2016, and Section 2). These hydrodynamical simulations incorporate the modern smoothed particle hydrodynamics code GASOLINE2 (Wadsley et al. 2004, 2017), and include comprehensive effects such as low-temperature metal cooling (Shen et al. 2010), UV background radiation, star formation, and stellar and superbubble feedback (Keller et al. 2014, 2015).

The mock galaxies provide a way to test our method’s predictive power because their stellar and dark matter profiles are more complex than the physical model assumed by GME—a similar situation when we apply our method to the real MW data.

Eighteen galaxies were created by MUGS2, and we analyze each of them individually. Mock images of the galaxies are shown in Figure 1 (reproduced from Figure 1 in Keller et al. 2016). We subject all of these galaxies to the blind test, and present detailed results for two galaxies (g15784 and g1536) as examples. Summarized results for the other galaxies are also provided, and used to make inferences about our method.

The organization of this paper is as follows. Section 2 provides a summary of the MUGS2 simulations that were carried out in previous studies (Keller et al. 2014). Next, Section 3 briefly reviews the physical model and the important points of the hierarchical Bayesian framework. Section 4 describes how mock tracers from the MUGS2 simulations were selected, and how mock heliocentric observations (with errors) were created from these tracers. In Section 5, we show the results and cumulative mass profile predictions of g15784 and g1536, and compare these to the true quantities (which were revealed only after our analysis was complete). The results from all 18 blind tests and a discussion follows in the same section. Section 6 provides a summary of our findings and avenues of future work.

2. Summary of MUGS2 Simulations

The MUGS2 (Keller et al. 2016) is a cosmological re-simulation of a sample of MW-like halos originally presented in Stinson et al. (2010) (i.e., MUGS). The initial MUGS sample was not designed to specifically create a single object like the MW, but rather to provide an unbiased sample of the kind of galaxy that lives in halos with masses \(\sim 10^{12} M_\odot\), which is where we expect \(L^*\) galaxies to reside (Moster et al. 2010). Both the MUGS and MUGS2 simulations use a Wilkinson Microwave Anisotropy Probe 3 ΛCDM cosmology with \(H_0 = 73 \text{ km} \text{s}^{-1} \text{Mpc}^{-1}\), \(\sigma_8 = 0.76\), and components \(\Omega_m = 0.24\), \(\Omega_\Lambda = 0.76\) and \(\Omega_{\text{bary}} = 0.04\) (Spergel et al. 2007).

The MUGS2 simulations differed from the original Stinson et al. (2010) simulations in two major ways. First, while Stinson et al. (2010) was simulated with the hydrodynamics code GASOLINE (Wadsley et al. 2004), MUGS2 was simulated with GASOLINE2 (Wadsley et al. 2017). The latter includes a new subgrid model for turbulent mixing of metals and energy (Shen et al. 2010), and improved hydrodynamics (see Wadsley et al. 2017, for details). Second, MUGS2 introduced a physically motivated “superbubble” model for stellar feedback; the new model captures the unresolved mixing between the cold, swept up shell and the hot interior of superbubbles driven by supernovae from star clusters (Keller et al. 2014).

MUGS2 shared the same initial conditions as the original MUGS study, in order to investigate the effects of including improved hydrodynamic methods and a more realistic model for stellar feedback. The MUGS initial conditions were drawn from a set of 18 cosmological zoom-in (Quinn & Binney 1992) galaxies, selected from a \(50 h^{-1} \text{Mpc}\) box evolved with \(256^3\) dark matter particles to \(z = 0\). Halos from the simulation were selected based on their mass and isolation alone. For halos between \(5 \times 10^{11} M_\odot\) and \(2 \times 10^{12} M_\odot\), 267 halos had no similarly sized neighbors within less than 2.7 Mpc, and of these halos, 18 were selected randomly in order to sample the spin parameter and merger history space in an unbiased way. The 18
halos were then re-simulated at higher resolution, and particles that accreted within $3r_{\text{vir}}$ at $z = 0$ were seeded with gas particles to generate a set of hydrodynamic initial conditions.

The final sample of 18 galaxies have spin parameters $\lambda' = J/\sqrt{5}/eGRM^3$ between 0.009 and 0.106. Their last major mergers occur over a range of redshifts, with $z = 7.3$ for the most quiescent of the sample (g15784), to just before redshift 0 ($z = 0.1$) for g28547. The earliest galaxy to assemble half of its mass is g15784 ($z = 1.3$), and the latest is g21647 ($z = 0.2$).

It was also found that the new feedback model in MUGS2 drives realistic, mass-loaded winds from galaxies, and becomes ineffective at regulating star formation at the peak of the stellar-mass-to-halo-mass curve (Moster et al. 2010; Keller et al. 2016) at $10^{12} M_\odot$. Thus, those galaxies with halo masses above this value contain two to three times too many stars, and significantly over-massive bulges. In nature, feedback from active galactic nuclei (AGNs) would become the dominant feedback process in these galaxies, but the MUGS2 simulations omit AGN feedback. Therefore, the galaxies with halo masses $>10^{12} M_\odot$ are referred to as “unregulated,” and the rest are referred to as “regulated.” The unregulated galaxies are distinguished by red labels in Figure 1 (from Keller et al. 2016).

During the blind tests of our method on the MUGS2 galaxies, we were not made aware of the unregulated and regulated categories. Thus, all galaxies were analyzed in the same way. Table 1 summarizes the physical characteristics of the MUGS2 galaxies, first grouped by regulated (upper-half) and unregulated (lower-half), and then listed in increasing mass.

### 3. Brief Review of the Hierarchical Bayesian Model

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

$$\Phi(r) = \frac{\Phi_0}{r^{\gamma}},$$

where $\Phi_0$ and $\gamma$ are free parameters. The radial distribution of the tracer population is also assumed to follow a power-law profile:

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

Figure 1. Mock images of the MUGS2 galaxies edge-on (top) and face-on (bottom). Reproduced from Figure 1 in Keller et al. (2016). Red labels denote the unregulated galaxies introduced in Section 2.
where $\alpha$ is a parameter. Given Equation (1), the total mass profile is given by

$$M(r) = \frac{\gamma \Phi_0}{G} \left( \frac{r}{\text{kpc}} \right)^{1-\gamma},$$

which goes to an isothermal sphere in the limit that $\gamma \to 0$, and a point mass as $\gamma \to 1$. Equations (1) and (2) are used in the Eddington formula to derive a distribution function (DF) (see Binney & Tremaine 2008),

$$f(\mathcal{E}, L) = \frac{L^{-2}\Phi(\mathcal{E})^{1/2}}{8\pi \gamma^{-2/\gamma}} \frac{1}{\beta_0} \left[ \frac{\alpha}{\gamma} - \frac{3}{2} \right] \Gamma \left( \frac{\alpha}{\gamma} - \frac{3/2}{\gamma} \right) \frac{\beta_0^{3/2}}{\gamma} \Gamma \left( \frac{\beta_0^{3/2}}{\gamma} + \frac{\alpha}{\gamma} - \frac{1}{2} \right)^{1/2},$$

where the model parameters are $\Phi_0$, $\gamma$, $\alpha$, and $\beta$. $\beta$ is the velocity anisotropy parameter (assumed constant) for the tracer population. For the curious reader, the derivation of the DF is given in the original paper by Evans et al. (1997), albeit with different notation, and in Paper 2.

The DF in Equation (4) is a probability distribution; it gives the probability of tracer having a specific energy $\mathcal{E}_i = -v_i^2/2 + \Phi(\mathcal{E})$, and specific angular momentum $L_i = r_i v_i$, given the model parameters. The data are the total speed of the tracer $v_i = \sqrt{v_r^2 + v_t^2}$, and its distance from the Galactic center $r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$. The velocity components $v_r$, $v_t$, and $v_z$ are the galactocentric radial and tangential velocity components. Assuming that the individual tracers in the population are independent, then the probability that all tracers have \{ $\mathcal{E}_i$, $L_i$ \} is

$$\prod_{i=1}^{N} f(\mathcal{E}_i, L_i).$$

The DF $f(\mathcal{E}, L)$ assumes a galactocentric reference frame, but the data are measured from our heliocentric perspective. Transforming from one frame to the other is not difficult, but properly propagating the measurement uncertainties to the galactocentric frame is. To overcome complex error propagation, we use a measurement model at the data level of the Bayesian hierarchy. That is, the measurement model is the likelihood in our hierarchical Bayesian analysis.

The measurement model (outlined more fully in Paper 3) assumes that a measurement of a quantity $x$ (e.g., line-of-sight velocity) is a random variable $X$ that is normally distributed about mean $\mu$,

$$X \sim N(\mu, \sigma^2),$$

where the variance $\sigma^2$ is set equal to the square of the known measurement uncertainty. In other words, the data (position $r$ and velocity components $v_\text{los}$, $v_r$, $v_\phi$, $\cos \delta$, and $\mu_\theta$) are assumed to be drawn from normal distributions centered on the true but unknown position and velocity components in the heliocentric frame. The true position and velocity components of each tracer are free parameters.

The prior on the likelihood is the DF (Equation (4)). The positions and velocities of the tracers are assumed to have the spatial distribution given by Equation (2) and to follow the influence of the total gravitational potential (Equation (1)). In this way, each tracer has individual parameters for their true position and velocity, but also shares with the rest of the tracers the $\Phi_0$ and $\gamma$ parameters defining the total gravitational potential.

The four model parameters $\Phi_0$, $\gamma$, $\alpha$, and $\beta$ in the DF are assigned hyperprior probability distributions, whose forms are given in Paper 3 and repeated here in Table 2. We choose truncated, uniform prior probabilities for $\Phi_0$, $\gamma$, and $\beta$, and a Gamma distribution for the prior on $\alpha$ (see Papers 2 and 3 for justification and more details). There is no direct prior on the mass ($M_{200}$) since this quantity is determined by $\Phi_0$ and $\gamma$ through Equation (3).

In summary, the posterior distribution is given by

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior} \times \text{Hyperpriors}$$

(see Eadie et al. 2016, 2017a, 2017b; Eadie 2017). In standard Bayesian inference, the posterior distribution contains all the information about the model parameters, given the data, the model, and the prior information.

The posterior distribution is most often approximated by generating a Markov chain that is a stationary distribution proportional to the posterior. We use this approach, but with
the variation of a hybrid-Gibbs within the Metropolis sampler
to increase efficiency. For information on the particulars of the
Markov Chain Monte Carlo (MCMC) sampling methods and
convergence diagnostics, the reader may refer to Paper 1, with
some updates in Papers 2 and 3.

In Paper 3, we estimated the virial mass and cumulative mass
profile of the MW using GME and the GC population. We now
apply this method to mock observations from the MUGS2
galaxies’ simulated tracer data via blind tests.

4. Creating Mock Observations

We used GCs as tracers in our analysis of the MW. In order
for our blind test to be as realistic as possible, we need tracer
data from the simulated galaxies that most closely resemble
that of the MW’s GC population. In the next few sections, we
describe the process for generating such mock data.

4.1. Finding GC Analogs in MUGS2 Galaxies

Great strides in cosmological, hydrodynamical simulations
have been made in recent years, but resolution limits have
prevented the ability to create GC populations within a fully
simulated, cosmological galactic environment.8 Instead, we
must use a selection of “star particles” from each mock galaxy
and treat them as GC analogs. The star particles represent
entire populations of stars, with each particle carrying a mass of
\( \sim 10^5 M_\odot \). Coincidentally, this mass is similar in mass to
many GCs.

The GC analogs from the MUGS2 data are star particles
with ages greater than 12 billion years and metallicities
\([\text{Fe}/\text{H}] < -1.5\). With such cuts, the stars in a GC analog
would have formed at an approximate redshift \( z \approx 3 \) or higher.
Disk-associated objects were also excluded 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.

In order to keep our analysis a true blind test, only the
galactocentric positions \((x, y, z)\) and velocities \((v_x, v_y, v_z)\) of the GC analogs were made known to us. The GC analogs may
be bound or unbound from the host galaxy, but this information
was withheld until the Bayesian solution was complete. This is
an important point, as our hierarchical Bayesian method
assumes all tracers are bound. Aside from the kinematic
information and total number of GC analogs, and the
knowledge that the MUGS2 host galaxies were “MW-type”
galaxies, we had no knowledge of their mass, mass profile, or
merger history.

The total velocities of the GC analogs as a function of their
galactocentric distance are shown in Figure 2. For comparison,
real MW GCs for which we have complete velocity
information are overplotted as solid blue squares. Because
much of the MW GC data are incomplete, the GCs shown
represent roughly half of the total GC population in the MW.

Three observations are immediately apparent in Figure 2:
(1) the GC analog velocity profiles of the MUGS2 galaxies
sometimes differ substantially from that of the MW GC’s,
(2) some velocity profiles have unique clustered features that
may represent satellites or recent mergers, and (3) there are
often many more GC analogs than GCs in the MW. The
number of GC analogs per MUGS2 galaxy ranges from
64–5106 (Table 1), and is a reflection of the different star
formation rates at high redshift. That is, the galaxies with
higher numbers of GC analogs had more star formation in
earlier times.

Because the GC analog population sizes differ from the
MW GC population, which consists of 157 known GCs
(Harris 1996, 2010), we randomly sampled GC analogs to
obtain the same sample size as the MW. Although it is possible
that the galaxies with a larger number of GC analogs are larger
galaxies, we did not use this as prior information in our
analysis.

4.2. Creating Mock Heliocentric Observations

We created mock heliocentric observations of the GC
analogs, such as would be viewed from a Sun-centered
reference frame. This involved a series of steps, including
transforming positions and velocities into a heliocentric frame,
and introducing missing data and measurement errors to
simulate real observations.

4.2.1. Transforming from Galactocentric to Heliocentric

The kinematic information of the GC analogs is in a
galactocentric, Cartesian coordinate system with positions \((x, y, z)\) and velocities \((v_x, v_y, v_z)\) (i.e., \(U, V, W\)). We first transform the
galactocentric positions into galactic coordinate \(l\) and \(b\) and heliocentric coordinates of right ascension and declination.

To create heliocentric velocities, we perform the inverse of the
transformation provided in Johnson & Soderblom (1987),
adjusting for the solar motion (we use the value from Schönrich
et al. 2010). The use of the same solar motion for every simulated
galaxy will not affect the results because this quantity is treated as
fixed and known in our analysis. That is, the same value for the
solar motion is used when the observations are transformed back
into the galactocentric frame.

The above steps were performed on the MUGS2 GC analog
data to create perfect (i.e., without error) heliocentric mock
data. To check the transformation, we converted these values
back to the galactocentric frame using the relevant code in
GME. There are discrepancies of \(\sim 2 \times 10^{-13}\) km s\(^{-1}\) in the
velocities at small \(r_{gc}\) when the transformation is performed
without scatter, due to floating point roundoff. However, the
symmetry of these discrepancies and their tiny values will not
contribute to any systematic bias in the result.

4.2.2. Introducing Measurement Uncertainties

Referring back to Equation (6), we call any difference
between a measured value \(x\) and the true value \(\mu\) the error.
In order to analyze the 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 \(\mu\) to the true value of the quantity (e.g., \(r, v_{los}\) etc.) in the MUGS2 data, deciding
on a value for \(\sigma^2\), and drawing from the normal distribution
determined by these parameters. How we choose to define \(\sigma^2\)
for each quantity \((r, v_{los}, \mu_\odot, \cos \delta, \mu_\odot)\) determines 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 velocity 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. The proper motions uncertainties range from
0.03–1.8 mas yr\(^{-1}\), and the line-of-sight velocity uncertainties

---

8 With the recent and notable exception of Pfeffer et al. (2018).
range from 0.1–5 km s⁻¹. We excluded two large measurement uncertainties in this process—those of Pal 3 and NGC 6218—to avoid assigning very large observed proper motions to the GC analogs that we deemed to be unrealistic. (In the MW analysis, we treated the proper motion of Pal 3 as unknown because the measurement uncertainty was so large.) We did investigate various distance measures to match MW GCs to GC analogs, so that GC analogs 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.

After performing reference frame transformations and introducing measurement errors for the GC analogs of each MUGS2 galaxy, we subsampled the mock data to mimic the sample size of the MW’s GC population. We randomly select 157 GC analogs from each MUGS2 galaxy, except in the case of g8893 for which there are only 64 GC analogs.

As noted previously, the velocity profiles of the GC analogs as a function of galactocentric distance are not always similar to the MW’s GC profile (Figure 2). This is reflected in the subsamples’ number density as a function of distance as well.

Figure 3 shows the empirical cumulative distribution functions (CDFs) of the galactocentric distances r of the subsampled GC analog mock data for (a) g15784 and (b) g1536, as they compare to the empirical CDF for the MW GC system. The CDFs are calculated using the function ecdf in the stats package of the R Statistical Software Environment.⁹

The black curves are the MUGS2 GC analogs, and the blue curves are the MW GCs. The points along the bottom are the distances of the individual GCs from the Galactic center (black circles for MUGS2, blue squares for MW GCs). The empirical CDFs of g15784 and g1536 are quite different, with the former being much more similar to the MW’s GC population CDF, especially at smaller r (Figure 3). The same observation is made for the other unregulated galaxies too; their CDFs appear more similar to the MW.

4.2.3. 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 20 kpc of the MW center, 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 motion measurements within 20 kpc for g1536 therefore seems unrealistic. Thus, we

⁹ https://www.r-project.org/
decide to keep all but two proper motion measurements within 20 kpc for \( g_{1536} \). All proper motions beyond 50 kpc are removed because we have only one complete data point past this distance in the MW. This procedure was needed for most regulated galaxies because they have few GC analogs within 20 kpc.

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 \( \alpha \). For the MUGS2 GC analogs, 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.

5. Results and Discussion

We now apply GME to the GC analog subsamples from each MUGS2 galaxy, and use the posterior distribution of model parameters to estimate \( r_{200} \) and \( M_{200} \) of each galaxy. The radius \( r_{200} \) is defined as the distance from the Galactic center within which the mean mass density is 200 times the critical density of the universe. We use a Hubble constant of \( 73 \pm 4 \) km s\(^{-1}\) Mpc\(^{-1}\), the same value used by Keller et al. (2016) to create the galaxies.

We begin with a detailed look at \( g_{15784} \) and \( g_{1536} \) (Section 5.1) before showing the summarized results from all MUGS2 galaxies (Section 5.2).

5.1. Detailed Cases: \( g_{15784} \) and \( g_{1536} \)

Galaxies \( g_{15784} \) and \( g_{1536} \) were chosen for the initial analysis because they have quiet merger histories and represent typical examples of the regulated and unregulated populations of galaxies from the MUGS2 simulations. Additionally, \( g_{15784} \) lacks any strange features in its velocity profile (Figure 2), has many GC analogs from which to draw samples (Table 1), and its mock image of the galaxy appears MW-like (see Figure 1). The regulated galaxy \( g_{1536} \), which is less concentrated than \( g_{15784} \), then makes for an interesting comparison.

5.1.1. Parameter Estimates and Total Mass

The mean estimates of the model parameters given by the posterior distribution for galaxy \( g_{15784} \) are presented in Table 3, where the numbers in brackets represent the bounds of the 95% marginal c.r.

Table 3

| Parameter                  | Mean   | 95% Marginal c.r. |
|----------------------------|--------|-------------------|
| \( \phi_0(10^4 \text{ km}^2 \text{s}^{-2}) \) | 47 (40, 57) |
| \( \gamma \)               | 0.41 (0.30, 0.57) |
| \( \alpha \)               | 3.04 (3.02, 3.06) |
| \( \beta \)                | 0.54 (0.41, 0.66) |

Derived Quantity

| Parameter                  | Mean   | 95% Marginal c.r. |
|----------------------------|--------|-------------------|
| \( r_{200} \) (kpc)        | 203 (173, 232) |
| \( M_{200} \) (10^{12} \text{ M}_\odot) | 1.1 (0.6, 1.6) |

\( \odot \) We also tried using the Planck Collaboration et al. (2016) result \( H_0 = 67.8 \) km s\(^{-1}\) Mpc\(^{-1}\) without significant changes to the results.
The results for $g_{1536}$ are shown in Table 4. In this case, the method did not perform as well. The true $r_{200}$ and mass are 174 kpc and $0.65 \times 10^{12} M_\odot$, whereas the predicted values were 152 (136, 170) kpc and 0.4 (0.3, 0.6) $\times 10^{12} M_\odot$.

### Table 4  
Model Parameter Estimates and Derived Quantities: $g_{1536}$

| Parameter               | Mean  | 95% Marginal Credible Region |
|-------------------------|-------|-----------------------------|
| $\phi_0 (10^4 \text{ km}^2 \text{ s}^{-2})$ | 24    | (17, 35)                    |
| $\gamma$               | 0.40  | (0.30, 0.61)                |
| $\alpha$               | 3.02  | (3.01, 3.04)                |
| $\beta$                | 0.51  | (0.28, 0.70)                |

| Derived Quantity | Value  | 95% Marginal Credible Region |
|------------------|--------|-----------------------------|
| $r_{\text{vir}}$ (kpc) | 152    | (136, 170)                  |
| $M_{200}$ ($10^{12} M_\odot$) | 0.4    | (0.3, 0.6)                  |

The cumulative mass profiles with Bayesian c.r. for each galaxy, using Equation (3) and the samples from the posterior distributions, are shown in Figure 4. The gray-shaded regions indicate the 50%, 75%, and 95% c.r., and the true cumulative mass profiles are the solid red curves. The vertical-dotted lines show the range of the subsampled data used in the analysis.

The predicted $M(<r)$ profile for both $g_{15784}$ and $g_{1536}$ falls below the true profile for most values of $r$. That is, even though the total mass is well estimated for $g_{15784}$, the predicted mass profile shows disagreement with the true mass profile in Figure 4(a). In both cases, the model fit appears to be a compromise between the inner and outer regions of the galaxy. One notable difference between the two galaxies is that the mass within 10 kpc is underpredicted for $g_{15784}$, but overpredicted for $g_{1536}$. Additionally, the true cumulative mass profile for $g_{1536}$ has a different shape than the predicted one.

### 5.1.3. Specific Energy Profiles

The hierarchical Bayesian method treats the true positions and velocities as nuisance parameters, sampling them in the MCMC hybrid-Gibbs algorithm (see Papers 1, 2, and 3). As a result, we obtain marginal posterior distributions for the galactocentric velocity and position of each GC analog.

In Paper 3, we used these distributions to estimate the specific energy $E$ of each tracer, given the mean model parameters. We compared these estimated energies to energies calculated from the actual measurements of position and velocity and the model parameters. By looking at the energies as a function of galactocentric distance $r$, we noted that GME attempts to reconcile outlier GC energies in light of the other GCs’ energies.

Here, we have the luxury of comparing the estimated energies to the true energies of the GC analogs given the actual gravitational potential calculated from the MUGS2 simulations.

Figure 5 shows the true specific energies (gray squares) and estimated specific energies (blue circles) of the GC analogs from $g_{15784}$ and $g_{1536}$. The true energies are calculated using the position and velocity data direct from the simulations, and the true gravitational potential at their distances. The estimated energies are calculated from the mean values of the nuisance parameters $(r, V_{\text{los}}, \mu_\alpha, \cos \delta, \mu_\delta)$ provided by the posterior distribution samples, and the mean estimates of the model
parameters. The estimates for the incomplete data are open circles, while the estimates for the complete data are filled.

In both galaxies, the energy estimates appear to display statistical shrinkage; the free parameters for the velocity and positions allow the estimated $E$ to move toward a common curve in $E - r$ space. Overall, this ability of our method to adjust the energy values of the GC analogs reflects the result found in Paper 3, for the real MW data.

Figure 5 also reveals a disagreement between the estimated energy profile and the true energy profile of the tracers. The differences are largest at small and large $r$ for both galaxies. The differences are more extreme for $g15784$ than $g1536$, and this likely leads to the more uncertain mass estimate in the former (Figures 4).

The estimated energy profile for galaxy $g1536$ appears to match the true energy profile more closely than that of $g15784$, but this is not an indication of a good model fit. Disagreements at small and large $r$ still exist, and the cumulative mass profile in Figure 4(b) shows a poor and underestimated fit for $g1536$’s mass.

Galaxy $g15784$ has incomplete data at all $r$, whereas almost all of the data within 20 kpc of $g1536$ are complete. A high percentage of the data for $g1536$ are also complete between 20 kpc and 50 kpc (Figure 5(b)). Consequently, the inner GC analogs of $g1536$ will have the most leverage in the model fit.

Figure 5. Energy profile of GC analogs from (a) $g15784$ and (b) $g1536$. The solid blue and open blue circles are the estimated energies, calculated using the mean values of both the model and nuisance parameters of the posterior distribution. The gray squares show GC analogs’ true energies, calculated using the true kinematic values and the true gravitational potential. The purple curve is the lower bound of the 95% c.r. for the predicted gravitational potential, and the solid black line is the true gravitational potential.
In Papers 2 and 3, we justified using a single power-law for the gravitational potential. Our argument was that at large distances a value of $\gamma = 0.5$ provides a good approximation to an NFW (Navarro et al. 1996) gravitational potential. In the case of g1536, the assumed gravitational potential does not hold at all radii, and so an abundance of complete data in the inner regions of the galaxy has probably biased the mass estimate. Indeed the complete data within 20 kpc correspond to the region in the mass profile that has the best mass prediction (Figure 4(b)).

In the next section, we review the results for the rest of the MUGS2 galaxies, and detect a similar bias in the mass estimates of the other regulated galaxies.

5.2. Analysis of all MUGS2 Galaxies
5.2.1. Parameter Estimates

The joint posterior distributions for $(\Phi_o, \gamma)$—the model parameters of the gravitational potential—are shown in Figure 6 (regulated galaxies) and 7 (unregulated galaxies). The green-dashed boxes indicate galaxies whose total mass was estimated within the 95% Bayesian c.r. (Section 5.2.2).

For many regulated galaxies, the free parameter $\gamma$ attempts to reach a location in parameter space outside of the prior distribution. Such behavior indicates the model struggled to accurately describe the gravitational potential given the data and prior assumptions.

![Figure 6. Joint distributions for $\Phi_o$ and $\gamma$, for the regulated galaxies. Points are samples from the Markov chain and contours represent (from inner to outer) the 25%, 50%, 75%, and 95% Bayesian c.r. The green-dashed outlines around the figure indicate that $M_{200}$ was correctly predicted within the 95% c.r.](image-url)
The behavior of $\gamma$ in the joint posterior distributions for the unregulated galaxies is also inconsistent. For example, the unregulated galaxies that are most underestimated and overconfident are $g22795$, $g4720$, and $g27491$. The joint distributions $p(\Phi_o, \gamma)$ in the first two cases show $\gamma \to 0.7$ (Figure 7), whereas for the distribution the latter is quite diffuse.

One interpretation of the results for $g22795$ and $g4720$ is that $\gamma$ is attempting to reach a value larger than 0.7 because of the massive bulges of these unregulated galaxies; in the limit that $\gamma \to 1$, the gravitational potential model goes to a Keplerian potential (Equation (1)). In Paper 2 when we analyzed the MW data, the parameters $\Phi_o$ and $\gamma$ appeared to be anticorrelated. Thus, a larger value of the latter leads to a smaller value of the former—which leads to a smaller mass estimate. However, in the present analysis with MUGS2 data, the parameters appear correlated. Therefore, we should be cautious about extrapolating conclusions from these blind tests to those about the MW that were arrived at using real GC data.

A posterior distribution that is truncated by a prior distribution typically indicates a poor model fit. The only joint posterior distributions that look well-behaved in Figures 6 and 7 are those of $g22437$, $g19195$, and $g25271$. Notably, the true $M_{200}$ values for these unregulated galaxies are well within the 95% Bayesian c.r. (Figure 9 and Section 5.2.2)

We also obtain estimates for the constant anisotropy parameter $\beta$ of the tracer population. Figure 8 shows the $\beta$ estimates (black circles) with 95% c.r. (errorbars), and the true constant anisotropy for (1) the GC analog subsample (blue...
diamonds) and (2) the total GC analog population (pink triangles). The unregulated galaxies are highlighted with a gray background.

The true $\beta$ value for the GC analog population is captured within the c.r. 11 out of 18 times. Although the mock data are incomplete beyond 50 kpc, making it difficult to constrain $\beta$, our estimates do not display significant bias when compared to the true constant anisotropies.

A feature of note is that galaxies g4145 and g22795 both have companions (Figure 1), and $\beta$ is very underestimated in both of these cases. Interestingly, galaxy g19195 has a noticeable feature in $v - r$ space (Figure 2) but $\beta$ is accurately estimated. Also worthy of note is the poor estimate of $\beta$ for g8893, which not only has the smallest number of GC analogs (64) but is also one of the more irregularly shaped galaxies in Figure 1. Our prior distribution on $\beta$ did not allow for values less than $-0.5$, and yet the true $\beta$ fell outside this range.

The regulated galaxies’ $\beta$ estimates perform less well than those of the unregulated galaxies. In Section 5.2.3 and Figure 11, we will see that the regulated galaxies have a higher percentage of incomplete tracer data, making it more difficult for the model to estimate the anisotropy of these tracer populations. Nevertheless, the accuracy of the $\beta$ estimate does not appear to be related to the accuracy of $M_{200}$ or $r_{200}$ shown next.

5.2.2. Mass ($M_{200}$) and $r_{200}$ Estimates

Figure 9 summarizes the median estimates of the total mass ($M_{200}$) for all MUGS2 galaxies. The estimates are shown as black circles, the true values are blue diamonds, and the 95% c.r. are shown as errorbars. The unregulated galaxies are shown with a gray background, and their $M_{200}$ estimates have notably wider marginal distributions than the other nine galaxies. Many
of the regulated galaxies of $r_{200}$ (black points), compared to the true values (blue diamonds). The errorbars represent the 95% Bayesian c.r.

The true total mass is captured within the 95% c.r. in 8 out of 18 cases, with the unregulated galaxies having better coverage than the regulated ones. Overall, the $M_{200}$ values are underestimated by the median, with the exceptions of galaxies g15807 and g21647 (which are overestimated), and g7124 and g22437 (which are estimated quite well). The estimates and true values of $r_{200}$ are shown in Figure 10, and these echo the mass results.

In the next section, we investigate how the error in the mass estimates might be related to incomplete data, model assumptions, and the evolutionary history of the MUGS2 galaxies.

### 5.2.3. Error in Mass Estimates

We now investigate why the method did not fair well in some cases, using the knowledge from the simulations. We know how many GC analogs are unbound from each galaxy, and from our mock observations we know the percentage of incomplete velocity data. Additionally, we have information about each galaxy’s merger history through the redshift of their last major merger ($z_{\text{imm}}$) and the redshift at which they acquired half of their final mass ($z_{\text{h}}$) (Keller et al. 2016).

Figure 11 shows the error in the mass estimates as a function of the following quantities: the percentage of unbound GC analogs, the percentage of incomplete data, $z_{\text{imm}}$, and $z_{\text{h}}$. The absolute percent error in the total mass is calculated by

$$\text{Percentage Error} = \left| \frac{\hat{M}_{200} - M_{200}}{M_{200}} \right| \times 100\%$$  \ (8)

where $\hat{M}_{200}$ is the estimate and $M_{200}$ is the truth.

The two galaxies with the highest absolute error (g21647 and g28547) are those that had the most recent major mergers and that have the highest percentage of unbound particles. Interestingly, these galaxies represent errors in two extremes—one was severely overestimated and the other underestimated. In general, however, there is only a slight trend for galaxies with recent major mergers to have the most inaccurate mass estimates (lower-left panel in Figure 11).

On average, the regulated galaxies have a higher proportion of incomplete data and unbound GC analogs than the unregulated galaxies. This may explain why the velocity anisotropy parameter was more poorly estimated for the regulated galaxies than that of the unregulated ones (Figure 8).

Overall, Figure 11 does not suggest any significant trends in the mass error with respect to merger history, percentage of incomplete data, or percentage of unbound particles. Thus, except in extreme cases, these quantities may not play a role in underestimating the mass. In the following section, we explore the possibility that the inner GC analogs are influencing the model fit, as suggested in Section 5.1.3.

### 5.3. Sensitivity to Inner Tracers

Our studies of the MW, and in particular the sensitivity analyses in Papers 2 and 3, found an increased mass estimate when inner GCs were not included in the analysis. We repeat this kind of analysis here by removing the inner GC analogs from the MUGS2 galaxies, and recalculating the mass estimate $M_{200}$.

Each galaxy in the MUGS2 simulation is unique in shape and spatial distribution of GC analogs. In order to remove inner GC analogs in a consistent way across all galaxies, we find the scale radius $r_s$ of each galactic disk via a simple exponential fit, and then remove GC analogs within $7.5r_s$. The MW’s scale radius is roughly 2 kpc, so in our own Galaxy this cut is akin to removing GCs within $r_{gc} = 15\text{kpc}$ of the Galactic center. It should be noted that the galaxies already had GC analogs removed within $3r_s$ in the original analysis.

The new mass estimates after removing the inner GC analogs are shown in Figure 12. In general, the estimates are better but at the cost of wider Bayesian c.r. The true $M_{200}$ values lie within the Bayesian c.r. in 13 out of 18 cases—a slight improvement from 8 out of 18 (Figure 9). Three regulated galaxy masses that were previously underestimated are now captured within the uncertainties. The mass estimate for g15807, previously a large outlier, is also notably improved.

In light of this result, and evidence presented in Section 5.1.3, we suggest that the combination of the GC analog number density profile, the percentage of complete data at small $r$, and the incomplete data at large $r$ are the culprits of the mismatch between the predicted and true mass profiles. To test this hypothesis, one could rerun the analysis in a future sensitivity test, gradually increasing the number of proper
motion measurements at large distances. We leave this to a future work.

5.4. Discussion

The underestimation of the total mass appears to be a systematic bias in our method, as it applies to these MUGS2 galaxies—especially the regulated variety. Although it may be tempting to extrapolate results from these blind tests to our previous MW results using GCs in Paper 3, it is important to make this kind of inference cautiously, for a few reasons:

1. The GC analog populations are dissimilar to the MW, and the MUGS2 galaxies, in the end, may not be very MW-like. The total number and number distribution of the GC analogs from the MUGS2 simulations differ substantially from that of the MW GC population (Table 1 and Figure 3(b)).

Additionally, the velocity profiles of the GC analogs from the regulated galaxies are dissimilar to the MW’s GC velocity profile in the inner regions (Figure 2). Instead, the velocity profiles of the unregulated galaxies appear more similar to the MW, even though these galaxies lacked appropriate feedback mechanisms in the MUGS2 simulations.

Keller et al. (2016) also note that the unregulated galaxies do not follow the standard stellar-mass-to-halo-mass relation. During the last stages of evolution in the MUGS2 simulations, feedback mechanisms were unable to effectively expel gas from the unregulated galaxies, which led to an overproduction of stars in their disks. 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). Given our basic assumption of a single power law for the gravitational

Figure 11. The absolute percentage error in $M_{200}$ vs. the percentage of unbound particles, the percentage of incomplete data, the redshift of the last major merger ($z_{lmm}$), and the redshift when the galaxy had acquired half of its final mass ($z_h$). Regulated galaxies are solid black circles, and unregulated galaxies are gray squares.
potential, it is thus not surprising that the unregulated galaxy masses are recovered more reliably than those of the regulated galaxies (Figure 9).

2. Only a single random sample of GC analogs from each MUGS2 galaxy was used in the analysis. The data were subsampled because the number of GC analogs in each MUGS2 galaxy is almost always larger than the MW GC population (Table 1). An interesting statistical test would be to repeat the analysis of Section 5 on multiple random samples from these galaxies, in order to fully understand the reliability of the mass and mass profile estimates. However, such an investigation would be a fully robustness test for each galaxy and would not improve the individual estimates for each galaxy. Rather, it would test whether the credible regions have equivalent coverage probabilities. Moreover, this is computationally expensive and would only provide insight for this particular set of simulations, which may or may not accurately represent nature.

3. Our method assumes that the galaxy is in virial equilibrium and that all GC analogs are bound—violations of these assumptions may lead to erroneous mass estimates. The MUGS2 galaxies have complex formation histories, and consequently a mixture of bound and unbound GC analogs. In particular, recent major mergers may create many unbound GC analogs. Unbound tracers will have higher total speeds than bound tracers; if the model assumes that unbound tracers are actually bound, then one would expect an overestimate of the mass. Indeed, this seems to be the case for g21647 and g15807. Galaxy g21647 had a very recent merger event (remnants of which are visible in the mock image of the galaxy, Figure 1), assembling half of its mass at \( z = 0.2 \). Unregulated galaxy g15807 had its last major merger at \( z = 2.5 \), and a feature visible in velocity space indicates that it has not fully recovered from this interaction (Figure 2). Thus, these recent interactions could explain the overestimates of the galaxies’ masses.

However, there is contradictory evidence for the other galaxies. On average, the regulated galaxies have a higher percentage of unbound GC analog tracers than the unregulated galaxies (Figure 11), and yet almost all of the regulated galaxies are underestimated (Figure 9). This is unexpected; assuming tracers are bound when they are not should lead to an overestimate of the mass. One possible explanation could be the location of complete data, discussed next.

4. The mock observations of GC analogs from the regulated galaxies and unregulated galaxies differ in their completeness. The regulated galaxies have a limited number of GC analogs within the inner regions of the galaxy (Figure 2). Thus, when mock observations were created, most of the regulated galaxies’ inner GCs analogs were given complete data within 20 kpc (Section 4.2.3 and e.g., Figure 5). Moreover, since mock observations beyond 20 kpc were made mostly incomplete, the regulated galaxies also have a higher percentage of incomplete data than the unregulated galaxies overall (Figure 11).

GME treats unknown proper motions as nuisance parameters in the model, and samples those nuisance parameters under the assumption that the tracers are bound to the galaxy. Thus, it is possible that the velocity estimates of the outermost incomplete data are indirectly influenced by the information from the inner tracers that have complete data. Since all GC analogs follow the same single power-law gravitational potential, this could lead to a lower mass estimate. The complete data in the inner regions, however limited, may have enough leverage to influence the model fit when the data at large distances are incomplete.

The evidence from our detailed investigation of regulated galaxy g1536 (Section 5.1.3), coupled with the sensitivity test (Section 5.3), supports this hypothesis. One could test this hypothesis in future work by re-running the analysis, gradually adding more complete data at larger radii.

5. Some MUGS2 galaxies are exceptional cases. Galaxies g21647 and g15807 were already mentioned as exceptional due to their merger histories. Other examples include g22795 and g4720. Both galaxies’ masses were underestimated, but both galaxies are also very compact (Figure 1), have high concentrations of tracers at their centers (Figure 2), and have the lowest gas fractions of all eighteen MUGS2 galaxies. Moreover, the Bayesian marginal posterior distributions for \( \gamma \) reached the upper limit of the prior distribution (i.e., 0.7) in both cases, indicating the model was attempting to explore values of \( \gamma > 0.7 \) and possibly a Keplerian potential (\( \gamma \rightarrow 1 \)).
This is exceptional behavior observed in the posterior brings us to the next important caveat to our results.

(6) In a Bayesian analysis, a posterior distribution that is truncated by the upper or lower limit of the prior distribution should inspire suspicion in the results.

Almost all of the regulated galaxies are inadequately matched by our model choice and prior distributions, with the $\gamma$ parameter reaching extreme ends of allowable values. In Figure 6, the mode of the joint distribution for $\Phi_0$ and $\gamma$ implies the free parameter $\gamma \rightarrow 0.3$ (the lower limit of $p(\gamma)$) for g5664, g8893, g1536, g422, g28547, and g24334. These galaxies’ masses were also very underestimated. Similar behavior is seen in the posterior distributions of some unregulated galaxies too (Figure 7). Thus, in an analysis of real data, if the posterior distribution occupies extreme parts of parameter space, then any inference should be performed with caution.

The only joint posterior distributions that look reasonable are those for g22437, g19195, and maybe g25271—and these three galaxies had masses that were estimated well within the 95% c.r.

In retrospect, using a uniform prior distribution on $\gamma$ between 0.3 and 0.7 does not necessarily reflect our prior assumptions. A value of $\gamma = 0.5$ corresponds to an Navarro–Frenk–White-like gravitational potential at large radii. A prior centered on $\gamma = 0.5$ and that drops to lower probability on both sides, for example, could instead be adopted; we leave this to a future work.

The results of the blind tests, even with these caveats, provide some insight into the behavior of the hierarchical Bayesian method as it applies to the MUGS2 data. The main results is that an abundance of complete data in the inner regions of the galaxy, and a lack of complete data in the outer regions, might bias the total mass estimate and cumulative mass profile to lower values if the gravitational potential is assumed to follow a single power law.

6. Conclusions

We have applied the hierarchical Bayesian mass estimation technique presented in Paper 3 to mock data from 18 MUGS2 hydrodynamical MW-type galaxies (Keller et al. 2015, 2016). Our method recovered the total mass within the 95% c.r. in 8 out of 18 cases, or 13 out of 18 cases, depending on which GC analogs were used in the analysis. The detailed analyses of g15784 and g1536, examples of an unregulated and a regulated galaxy, showed only moderate recovery of the cumulative mass profiles.

We can cautiously say that the hierarchical method with the current model for the gravitational potential (Equation (1)) tends to underestimates the total mass, at least for this small sample of galaxies. In particular, it is difficult for the model to predict the total mass accurately when many tracers are unbound to the galaxy and when those tracers have incomplete velocity measurements. Regardless, given the diversity of these galaxies (Figure 1) and our simple assumption for the gravitational potential, the method performs reasonably well in predicting the mass within the 95% c.r.

It is difficult to assess the reliability of our method on the MW data (i.e., the results of Paper 3) given the result of this study. Eighteen simulated galaxies is by no means a large sample size. Furthermore, nine of these galaxies (the unregulated ones) are not MW-like, and the other nine galaxies (the regulated ones) do not have MW-like GC-analog populations. The GC analogs 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.

The detailed case of g1536 and the results of the sensitivity analysis also suggest that the location of complete and incomplete data in the regulated galaxies may have played a role in underestimating the total masses. Thus, not only are simulations with GC analogs that are more similar to the MW’s GC population needed for future tests of the method, but the incomplete data at large radii are also a key piece of the puzzle.

Based on the results of this study, we suggest modifying the use of GME in future applications to the MW: it might be prudent to use only data at large distances. The trade-off may be a more uncertain result, but with less risk of bias. The Bayesian c.r. in the resulting mass estimate and cumulative mass profile will be larger, but they will be more likely to contain the truth. Additionally, we should be cautious if the posterior distribution approaches extreme ends in the allowed parameter space. In an upcoming paper, we will apply GME to only the outer tracers of the MW, whose proper motions are available from Gaia DR2 and the HSTPROMO project.

There are a variety of avenues for future work. One way forward is to compare the viability of different galaxy model assumptions within our hierarchical framework through the Bayes factor (Jeffrey 1939). However, this is complicated by the shortage of analytic DFs for galaxy models. Analytic DFs are required in the current setup of our hierarchical Bayesian framework. Non-analytic models might be possible with approximate Bayesian computation (ABC) or “forward modeling,” at the cost of substantial overhaul of the hierarchical code.

However, 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 favorable 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.

Thus, the results of this study encourage us to pursue our investigations of 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. Additionally, we plan to investigate the effects of choosing different hyperprior distributions on the model parameters, especially for $\gamma$. 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 at larger distances. The latter two are the most important next steps.

The type of blind test performed here can also be completed with mock data from other high-performance computer simulations that produce MW-type galaxies. In particular, data from the Apostle, Aquarius, Eagle, Fire, Illustris, and Latte (Springel et al. 2008; Hopkins et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Sawala et al. 2016; Wetzel et al. 2016; Sanderson et al. 2018) projects would all make interesting candidates.

Currently, we are pursuing this avenue of research with the Modeling Star cluster population Assembly In Cosmological Simulations within the EAGLE (E-MOSAICS) project (Pfeffer et al. 2018). In addition, we are investigating the importance of complete data for tracers at large radii, as well as the choice of
prior distributions for the model parameters. Our findings, including results from future tests using the E-MOSAICS data—which contain resolved GCs within a cosmological simulation—will follow in our next paper.

This research was funded in part by an Alexander Graham Bell Canada Graduate Scholarship-Doctoral (CGS D) to the first author, from the Natural Sciences and Engineering Research Council of Canada (held at McMaster University), and by the Moore, Sloan and Washington Research Foundations Data Science fellowship at the eScience Institute, University of Washington (UW). The lead author would like to thank Mario Jurić in the Department of Astronomy and DIRAC Institute and Tyler McCormick in the Department of Statistics (both at UW) for very helpful discussions regarding this research. The authors would also like to thank the anonymous referee for a thorough report that helped improve this paper.

Software: R Statistical Software Environment (R Development Core Team), including the following packages: stats (part of base R), CODA: Convergence Diagnosis and Output Analysis for MCMC (Plummer et al. 2006), emdbook: Ecological Models and Data in R (Bolker 2018), ggplot2 (Wickham 2016), MASS: Modern Applied Statistics with S Venables & Ripley (2002), moments: Moments, cumulants, skewness, kurtosis and related tests (Komsta & Novomestky 2015), praca: Practical Numerical Math Functions (Borchers 2017), RColorBrewer: ColorBrewer Palettes, and SNOW: Simple Network of Workstations (Tierney et al. 2013). (Neuwirth 2014). The MUGS2 galaxy simulations used GASOLINE2 (Wadsley et al. 2004, 2017).

ORCID ids

Gwendolyn Eadie https://orcid.org/0000-0003-3734-8177
Benjamin Keller https://orcid.org/0000-0002-9642-7193
William E. Harris https://orcid.org/0000-0001-8762-5772

References

Binney, J., & Tremaine, S. 2008, Galactic Dynamics (2nd ed.; Princeton, NJ: Princeton Univ. Press)
Bolker, B. 2018, emdbook: Ecological Models and Data in R, r Package Version 1.3.10, http://ms.mcmaster.ca/~bolker/emdbook/
Borchers, H. W. 2017, praca: Practical Numerical Math Functions, r Package version 2.0.7, https://CRAN.R-project.org/package=praca
Boylan-Kolchin, M., Besla, G., & Hernquist, L. 2011, MNRAS, 414, 1560
Deason, A. J., Belokurov, V., & Evans, N. W., et al. 2012, MNRAS, 425, 2840
Dehnen, W., McLaughlin, D. E., & Sachani, J. 2006, MNRAS, 369, 1688
Diaz, J. D., Koposov, S. E., Irwin, M., Belokurov, V., & Evans, N. W. 2014, MNRAS, 443, 1688
Eadie, G. 2017, PhD thesis, McMaster Univ.
Eadie, G., & Harris, W. 2016, ApJ, 829, 108
Eadie, G., Harris, W., & Widrow, L. 2015, ApJ, 806, 54
Eadie, G., Springfield, A., & Harris, W. 2017a, ApJ, 838, 76
Eadie, G. M., Harris, W. E., Widrow, L. M., & Springfield, A. 2016, in Proc. IAU Symp. 317, The General Assembly of Galaxy Halos: Structure, Origin and Evolution, ed. A. Bruggalio et al. (Cambridge: Cambridge Univ. Press), 296
Eadie, G. M., Springfield, A., & Harris, W. E. 2017b, ApJ, 835, 167
Evans, N. W., Hafen, R. M., & de Zeeuw, P. T. 1997, MNRAS, 286, 315
Fritz, T. K., Battaglia, G., Pawlowski, M. S., et al. 2018, A&A, in press (arXiv:1805.0090v4)
Gibbons, S. L. J., Belokurov, V., & Evans, N. W. 2014, MNRAS, 445, 3788
Harris, W. E. 1996, AJ, 112, 1487
Harris, W. E. 2010, arXiv:1012.3224
Hopkins, P. F., Kereš, D., Ohotne, J., et al. 2014, MNRAS, 445, 581
Jeffrey, H. 1939, Theory of Probability (Oxford: Oxford Univ. Press)
Johnson, D. R. H., & Soderblom, D. R. 1987, AJ, 93, 864
Kahn, F. D., & Woltjer, L. 1959, ApJ, 130, 705
Keller, B. W., Wadsley, J., Benincasa, S. M., & Couchman, H. M. P. 2014, MNRAS, 442, 3013
Keller, B. W., Wadsley, J., & Couchman, H. M. P. 2015, MNRAS, 453, 3499
Keller, B. W., Wadsley, J., & Couchman, H. M. P. 2016, MNRAS, 463, 1431
Komsta, L., & Novomestky, F. 2015, moments: Moments, cumulants, skewness, Kurtosis and Related Tests, r Package Version 0.14, https://CRAN.R-project.org/package=moments
Law, D. R., & Majewski, S. R. 2010, ApJ, 714, 229
Little, B., & Tremaine, S. 1987, ApJ, 320, 493
McMillan, P. J. 2017, MNRAS, 465, 76
Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Neuwirth, E. 2014, RColorBrewer: ColorBrewer Palettes, r Package Version 1.1-2, https://CRAN.R-project.org/package=RColorBrewer
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
Pfeffer, J., Kuijssen, J. M. D., Crain, R. A., & Bastian, N. 2018, MNRAS, 475, 4309
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, https://www.r-project.org/doc/RNews/Rnews_2006_1.pdf
Posti, L., & Helmi, A. 2018, arXiv:1805.01408
Quinn, T., & Binney, J. 1992, MNRAS, 255, 729
Sakamoto, T., Chiba, M., & Beers, T. C. 2003, A&A, 397, 899
Sanderson, R. E., Wetzl, A., Loebman, S., et al. 2018, arXiv:1806.10564
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
Shen, S., Wadsley, J., & Stinson, G. 2010, MNRAS, 407, 1581
Spitzer, D. N., Bean, R., Doré, O., et al. 2007, ApJS, 170, 377
Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685
Sörisalo, M. P., Bailin, J., Couchman, H., et al. 2010, MNRAS, 408, 812
Tierney, L., Rossini, A. J., Li, N., & Sevcikova, H. 2013, snow: Simple Network of Workstations, r Package Version 0.3-13, https://CRAN.R-project.org/package=snow
Venables, W. N., & Ripley, B. D. 2002, Modern Applied Statistics with S (4th ed.; New York: Springer)
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
Wadsley, J., Stadel, J., & Quinn, T. 2004, NewA, 9, 137
Wadsley, J. W., Keller, B. W., & Quinn, T. R. 2017, MNRAS, 471, 2357
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., Sohn, S. T., & Evans, N. W. 2018, ApJ, submitted (arXiv:1804.1134v5)
Wetzal, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, ApJL, 827, L23
Wickham, H. 2016, ggplot2: Elegant Graphics for Data Analysis (New York: Springer)
Wilkinson, M., & Evans, N. 1999, MNRAS, 310, 645
Xue, X., Rix, H., Zhao, G., et al. 2008, ApJ, 684, 1143

17