

Sizing from the smallest scales: the mass of the Milky Way

M. K. Rodriguez Wimberly^{1,2,★}, M. C. Cooper^{1,2}, D. C. Baxter^{1,2}, M. Boylan-Kolchin^{1,3},
J. S. Bullock^{1,2}, S. P. Fillingham^{1,4}, A. P. Ji⁵, L. V. Sales¹ and J. D. Simon⁶

¹Department of Physics and Astronomy, University of California Riverside, 900 University Avenue, CA 92507, USA

²Center for Cosmology, Department of Physics & Astronomy, University of California, 4129 Reines Hall, Irvine, CA 92697, USA

³Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712, USA

⁴Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA

⁵Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637, USA

⁶Observatories of the Carnegie Institution for Science, Pasadena, CA 91101, USA

Accepted 2022 May 3. Received 2022 May 3; in original form 2021 August 16

ABSTRACT

As the Milky Way and its satellite system become more entrenched in near field cosmology efforts, the need for an accurate mass estimate of the Milky Way’s dark matter halo is increasingly critical. With the second and early third data releases of stellar proper motions from *Gaia*, several groups calculated full 6D phase-space information for the population of Milky Way satellite galaxies. Utilizing these data in comparison to subhalo properties drawn from the Phat ELVIS simulations, we constrain the Milky Way dark matter halo mass to be $\sim 1\text{--}1.2 \times 10^{12} M_{\odot}$. We find that the kinematics of subhaloes drawn from more- or less-massive hosts (i.e. $>1.2 \times 10^{12} M_{\odot}$ or $<10^{12} M_{\odot}$) are inconsistent, at the 3σ confidence level, with the observed velocities of the Milky Way satellites. The preferred host halo mass for the Milky Way is largely insensitive to the exclusion of systems associated with the Large Magellanic Cloud, changes in galaxy formation thresholds, and variations in observational completeness. As more Milky Way satellites are discovered, their velocities (radial, tangential, and total) plus Galactocentric distances will provide further insight into the mass of the Milky Way dark matter halo.

Key words: Galaxy: fundamental parameters – galaxies: dwarf – galaxies: evolution – galaxies: general – galaxies: kinematics and dynamics – Local Group.

1 INTRODUCTION

Several of the most pressing cosmological problems challenging the Lambda cold dark matter (Λ CDM) paradigm, namely the Too Big to Fail (TBTf; Boylan-Kolchin, Bullock & Kaplinghat 2011, 2012) and the Missing Satellites problems (Klypin et al. 1999; Moore et al. 1999), depend heavily on the Milky Way’s dark matter halo mass. One way to resolve the TBTf problem within Λ CDM is through the assumption of a less massive Milky Way, for which fewer massive satellites with high central densities are expected (Wang et al. 2012; Cautun et al. 2014). Similarly, in conjunction with suppression of galaxy formation on the very smallest scales (e.g. Efstathiou 1992; Thoul & Weinberg 1996), the Missing Satellites problem can also be largely eliminated by lowering the assumed Milky Way dark matter halo mass (and thus the predicted number of satellite systems). As such, the Milky Way’s dark matter halo mass is a critical parameter in testing Λ CDM and models of galaxy formation on small scales (see discussion in Bullock & Boylan-Kolchin 2017).

Alternative resolutions to both the TBTf and Missing Satellites problems lie in the possibility that the Milky Way may be an outlier relative to the cosmic norm. For example, only ~ 10 per cent of Milky Way-like systems are estimated to have satellites as massive as

the Large and Small Magellanic Clouds (Boylan-Kolchin et al. 2010; Busha et al. 2011; Tollerud et al. 2011; Santos-Santos et al. 2021). The Milky Way’s satellite population is also remarkable in another characteristic – its Vast Polar Structure (VPOS; e.g. Lynden-Bell 1976; Kroupa, Theis & Boily 2005; Pawłowski, Pflamm-Altenburg & Kroupa 2012; Fritz et al. 2018; Pawłowski & Kroupa 2020). While our ability to observe such structures in systems beyond our very local Universe is still relatively new (Conn et al. 2013; Ibata et al. 2013; Collins et al. 2015; Müller et al. 2018), our observed flattened polar distribution of satellites, the VPOS, seems to be uncommon (Metz, Kroupa & Libeskind 2008; Ibata et al. 2014; Pawłowski et al. 2014; Pawłowski & McGaugh 2014; Cautun et al. 2015; Buck, Dutton & Macciò 2016; Ahmed, Brooks & Christensen 2017; Shao et al. 2018; Shao, Cautun & Frenk 2019). Another unusual feature of our local system may be the high fraction of quenched (or passive) satellite galaxies. Extragalactic surveys, such as SAGA, have found that the majority of satellites around Milky Way-like systems are actively star forming – SAGA in fact finds 85 per cent of low-mass satellites ($M_{*} \sim 10^{7\text{--}8.5} M_{\odot}$) are star-forming across 36 Milky Way-like systems (Geha et al. 2017; Mao et al. 2021). The radial distributions of the Milky Way satellites versus observed and simulated Milky Way analogues is also a contentious point which may place the Milky Way out of the cosmic norm. Some recent work highlights discrepancies in the 3D radial distributions of Milky Way satellites and various cosmological simulations, in particular that Milky Way satellites are more radially concentrated than their simulated counterparts (Moore

* E-mail: mariaro@ucr.edu

† NSF MPS-Ascend Postdoctoral Fellow.

2001; Willman et al. 2004; Yniguez et al. 2014; see also Carlsten et al. 2020, but see also Macciò et al. 2010; Bose et al. 2020; Font, McCarthy & Belokurov 2021; Samuel et al. 2021). The dark matter halo mass of the Milky Way has strong implications for its ability to quench satellite galaxies and for the radial distribution of its satellite population. More broadly, our reliance upon the Milky Way as a Cosmic Rosetta Stone (Boylan-Kolchin et al. 2016) requires a strong constraint on its dark matter halo mass.

Gaia has opened a new opportunity to study the distribution and dynamics of the Milky Way satellite population and to constrain the Milky Way’s dark matter halo mass. Prior to the second data release (DR2) of proper motions from *Gaia*, the Milky Way’s dark matter halo mass limits were $0.8\text{--}4.5 \times 10^{12} M_{\odot}$ (e.g. Boylan-Kolchin et al. 2013; Phelps, Nusser & Desjacques 2013; Kafle et al. 2014, and in particular see fig. 1 in Wang et al. 2020 for a recent literature summary). Since *Gaia* DR2, this mass has been inferred in various ways — from calculating the escape speed from counter-rotating stars in the Galaxy’s outer halo (Monari et al. 2018); using a scale-free mass estimator involving the density, potential, and anisotropy (β) of the satellites — galaxies or globular clusters — surrounding the Milky Way (Watkins et al. 2019; Fritz et al. 2020); comparing phase-space distributions in simulations and semi-analytical models to observed distributions to then infer the mass (Patel et al. 2018; Callingham et al. 2019; Eadie & Jurić 2019; Li et al. 2020); fitting physically motivated models to the *Gaia* DR2 stellar rotation curve (Cautun et al. 2020); and calculating the mass within 100 kpc via a distribution function method then extrapolating total mass (Deason et al. 2020). The results of these recent studies range from $M_{200} = 0.7^{+0.11}_{-0.08} - 1.55^{+0.64}_{-0.51} \times 10^{12} M_{\odot}$ (respectively, Monari et al. 2018; Eadie & Jurić 2019). It is important to note that even within confidence intervals, many of these results do not agree with one another.

As an alternate approach to these direct dynamical methods, in this work we constrain the Milky Way’s dark matter halo mass through comparison of subhalo kinematics in a suite of high-resolution N -body simulations to corresponding observational measures of the Milky Way satellite population using a statistical test to measure the ‘*antigoodness of fit*’ between these data sets. Herein, we utilize orbital parameters for the Milky Way satellites, derived primarily from proper motion measurements contained in the early third *Gaia* Data Release (EDR3, *Gaia* Collaboration 2016, 2021), subhalo kinematics from the Phat ELVIS suite of Milky Way-like simulations (Kelley et al. 2019) and the Mann–Whitney U test (Mann & Whitney 1947). In Section 2, we discuss our observed tracers of the Milky Way host potential along with the comparison suite of cosmological simulations. Section 3 details our primary analysis techniques, while our results are presented in Section 4. In Section 5, we examine various sources of potential systematic errors and points of further discussion, including the impact of satellites associated with the Large Magellanic Cloud, our adopted lower limit for peak subhalo velocity, orbital characteristics, limitations of our approach, comparisons to previous studies and observational completeness. Additionally, in this section we make some predictions for how future observations might impact our results. Finally, we summarize in Section 6.

2 DATA

2.1 Gaia

Gaia has spurred a dramatic improvement in our understanding of the orbital parameters for nearby stars, including those within the satellites of the Milky Way (MW) (*Gaia* Collaboration 2016). In

the second and early third data releases (DR2 and EDR3, *Gaia* Collaboration 2018, 2021; Lindegren et al. 2018), *Gaia* provides precise parallaxes and/or proper motions for over one billion sources, in an absolute reference frame defined entirely by *Gaia* observations. From this vast data set, several groups calculated full phase-space information, including tangential velocities, for a majority of the MW satellites (e.g. using DR2: Helmi et al. 2018; Simon 2018; Fritz et al. 2018; Massari & Helmi 2018; Pace & Li 2019; Kallivayalil et al. 2018 plus McConnachie & Venn 2020a; Li et al. 2021; Battaglia et al. 2022, which utilize EDR3). Herein, we utilize the Galactocentric tangential velocities from McConnachie & Venn (2020a, hereafter referred to as MCV20a). Heliocentric radial velocities are taken from McConnachie & Venn (2020b) and converted to the Galactocentric reference frame using *astropy* (*Astropy* Collaboration 2013, 2018). Heliocentric distances and associated errors are taken from Simon (2019)¹, Karachentsev & Kashibadze (2006), Weisz et al. (2016), and Torrealba et al. (2019). These distances are also converted to the Galactocentric reference frame using *astropy*. These quantities, converted to the Galactocentric reference frame, along with other properties of the MW satellites used in this work can be found in Table 1.

We limit our sample of Milky Way satellites to those systems within a Galactocentric distance of 300 kpc and exclude unconfirmed systems that are likely not galaxies (e.g. Indus I and DESJ0225+0304). In addition, we exclude the Sagittarius (Sgr) dwarf from our sample. Sgr is currently being disrupted via tidal interactions with the Milky Way (e.g. Ibata, Gilmore & Irwin 1994; Law, Majewski & Johnston 2009; Koposov et al. 2012), such that it is poorly reproduced in our comparison simulation data set (see Section 2.2). Our primary sample includes 44 satellite galaxies. Fig. 1 shows the distribution of these systems as a function of Galactocentric distance and total velocity. Finally, we identify a subsample of 34 systems with higher precision tangential velocities, such that $|V_{\text{tan, err}}/V_{\text{tan}}| \leq 0.30$. While this subset of systems is biased towards smaller Galactocentric distance, it does span a broad range of velocities (see Fig. 1). The particular selection limit used to define this subsample was adopted to exclude those systems with exceptionally uncertain tangential velocities while maintaining a statistically significant sample size.

2.2 Phat ELVIS

As a comparison data set, we utilize the Phat ELVIS (phELVIS) suite of 12 high-resolution, dissipationless simulations of MW-like haloes (Kelley et al. 2019). Building upon the ELVIS (Exploring the Local Volume In Simulations) suite of Local Group and MW-like simulations (Garrison-Kimmel et al. 2014), phELVIS incorporates the effects of tidal disruption due to an artificial disc potential (e.g. Garrison-Kimmel et al. 2017; Sawala et al. 2017). This new suite includes a total of 24 MW-like simulations, encompassing 12 high-resolution cosmological dark matter-only (DMO) simulations

¹Data presented by Simon (2019) are compiled from Torrealba et al. (2016a, b, 2018), Dall’Ora et al. (2006, 2012), Walsh et al. (2008), Kuehn et al. (2007), Greco et al. (2007), Musella et al. (2009, 2012), Kinemuchi et al. (2008), Longeard et al. (2018), Crnojević et al. (2016), Rizzi et al. (2007), Koposov et al. (2015, 2018), Bechtol et al. (2015), Vivas et al. (2016), Bellazzini et al. (2004), Bellazzini, Gennari & Ferraro (2005), Moretti et al. (2009), Medina et al. (2018), Sand et al. (2012), Mutlu-Pakdil et al. (2018), Pietrzyński et al. (2008), Belokurov et al. (2007), Boettcher et al. (2013), Lee, Grebel & Hodge (2003), Hamaonowicz et al. (2016), Carlin et al. (2017), Drlica-Wagner et al. (2015), Garofalo et al. (2013), and Willman et al. (2005a).

Table 1. Properties of the MW satellite galaxies used in this work. Column (1) Status of whether the system is a spectroscopically confirmed galaxy or not (i.e. Y = confirmed galaxy and P = not confirmed but probably a galaxy). Columns (2) and (3) Proper motions derived by McConnachie & Venn (2020a) from *Gaia* EDR3 in mas yr⁻¹. Column (4) Galactocentric distance with errors in kpc. Columns (5)–(7) radial, tangential, and total velocities, respectively, in the Galactocentric frame of reference, all in km s⁻¹. Galactocentric distance and radial velocity were converted from the heliocentric frame of reference using *astropy* and quantities from MCV20a, MCV20b and the heliocentric distances referenced in Section 2.1. Galactocentric tangential velocity was converted from the Galactocentric tangential velocity components provided in MCV20a. Total velocity was then calculated from its two components. Column (8) indicates the various subgroups a galaxy belongs to which are used throughout this work. 1 indicates belonging to the group of systems with proportionally low tangential velocity errors. 2 indicates the system is a long-term satellite of the LMC as determined by Patel et al. (2020). 3 indicates the system is a short-term satellite or recent interactor with the LMC as determined by Patel et al. (2020).

Galaxy	Confirmed?	$\mu_{\alpha} \cos \delta$ (mas yr ⁻¹)	μ_{δ} (mas yr ⁻¹)	D_{MW} (kpc)	V_{rad} (km s ⁻¹)	V_{tan} (km s ⁻¹)	$V_{3\text{D}}$ (km s ⁻¹)	Notes
Antlia II	Y	-0.09 ± 0.01	0.12 ± 0.01	133.0 ± 6.0	70.4 ± 0.5	125.0 ± 6.1	143.5 ± 5.3	1
Aquarius II	Y	-0.17 ± 0.1	-0.43 ± 0.08	105.3 ± 3.3	30.4 ± 7.2	157.0 ± 51.6	159.9 ± 50.7	3
Boötes I	Y	-0.39 ± 0.01	-1.06 ± 0.01	63.6 ± 2.0	91.2 ± 2.1	156.0 ± 3.0	180.7 ± 2.8	1
Boötes II	Y	$-2.33^{+0.09}_{-0.08}$	-0.41 ± 0.06	39.8 ± 1.0	-54.3 ± 3.9	$319.0^{+17.9}_{-16.0}$	$323.6^{+17.6}_{-15.8}$	1
Canes Venatici I	Y	-0.11 ± 0.02	-0.12 ± 0.02	210.8 ± 6.0	78.2 ± 0.5	69.0 ± 22.5	104.3 ± 14.9	
Canes Venatici II	Y	-0.15 ± 0.07	-0.27 ± 0.06	160.6 ± 4.0	-96.7 ± 0.2	31.0 ± 62.9	101.5 ± 19.2	
Carina	Y	0.53 ± 0.01	0.12 ± 0.01	107.6 ± 5.0	8.5 ± 0.3	187.0 ± 7.6	187.2 ± 7.6	1
Carina II	Y	1.88 ± 0.01	0.13 ± 0.02	37.1 ± 0.6	219.5 ± 1.9	268.0 ± 3.8	346.4 ± 3.2	1,2
Carina III	Y	3.12 ± 0.05	$1.54^{+0.06}_{-0.07}$	29.0 ± 0.6	58.7 ± 4.7	$395.0^{+12.4}_{-12.5}$	$399.3^{+12.3}_{-12.4}$	1,2
Columba I	P	0.19 ± 0.06	-0.36 ± 0.06	187.6 ± 10.0	-22.8 ± 6.8	205.0 ± 52.4	206.3 ± 52.1	
Coma Berenices	Y	0.41 ± 0.02	-1.71 ± 0.02	43.2 ± 1.5	31.9 ± 0.7	264.0 ± 4.2	265.9 ± 4.2	1
Crater II	Y	-0.07 ± 0.02	-0.11 ± 0.01	116.4 ± 1.1	-76.0 ± 0.1	102.0 ± 9.4	127.2 ± 7.5	1
Draco	Y	0.042 ± 0.005	-0.19 ± 0.01	82.0 ± 6.0	-103.0 ± 0.4	156.0 ± 1.4	186.9 ± 1.2	1
Draco II	P	1.08 ± 0.07	0.91 ± 0.08	23.9 ± 3.8	-156.6 ± 1.6	299.0 ± 15.0	337.5 ± 13.3	1
Fornax	Y	0.382 ± 0.001	-0.359 ± 0.002	141.1 ± 3.0	-34.5 ± 0.2	142.0 ± 0.0	$146.1^{+5.3}_{-0.0}$	1
Grus I	P	0.07 ± 0.05	$-0.29^{+0.06}_{-0.07}$	116.2 ± 11.5	-187.3 ± 4.3	$71.0^{+38.0}_{-41.3}$	$200.3^{+14.1}_{-15.2}$	
Grus II	P	0.38 ± 0.03	-1.46 ± 0.04	48.4 ± 5.0	-124.8 ± 1.7	139.0 ± 10.2	186.8 ± 7.7	1
Hercules	Y	-0.03 ± 0.04	-0.36 ± 0.03	126.3 ± 6.0	141.1 ± 1.3	146.0 ± 24.4	203.0 ± 17.6	1
Horologium I	Y	0.82 ± 0.03	-0.61 ± 0.03	87.3 ± 12.0	-28.4 ± 4.3	193.0 ± 15.9	195.1 ± 15.8	1,2
Horologium II	P	$0.76^{+0.2}_{-0.29}$	$-0.41^{+0.23}_{-0.21}$	79.1 ± 7.5	29.1 ± 25.0	$128.0^{+111.9}_{-157.7}$	$131.3^{+109.3}_{-153.8}$	
Hydra II	P	-0.34 ± 0.1	$-0.09^{+0.08}_{-0.09}$	148.1 ± 7.5	136.6 ± 0.7	$97.0^{+71.8}_{-72.6}$	$167.5^{+41.6}_{-42.0}$	
Hydrus I	Y	3.79 ± 0.01	-1.5 ± 0.01	25.7 ± 0.5	-40.1 ± 1.4	363.0 ± 5.0	365.2 ± 4.9	1,2
Leo I	Y	-0.05 ± 0.01	-0.11 ± 0.01	257.9 ± 15.5	174.7 ± 0.1	75.0 ± 12.1	190.1 ± 4.8	1
Leo II	Y	-0.14 ± 0.02	-0.12 ± 0.02	235.6 ± 14.0	26.5 ± 0.3	103.0 ± 22.6	106.3 ± 21.9	1
Leo IV	Y	-0.08 ± 0.09	-0.21 ± 0.08	154.6 ± 5.0	8.6 ± 0.2	52.0 ± 63.8	52.7 ± 62.9	
Leo V	Y	-0.06 ± 0.09	$-0.25^{+0.09}_{-0.08}$	169.8 ± 4.0	55.8 ± 1.6	$69.0^{+83.5}_{-78.6}$	$88.8^{+64.9}_{-61.1}$	
Phoenix II	Y	0.48 ± 0.04	-1.17 ± 0.05	81.3 ± 4.0	-38.1 ± 5.9	271.0 ± 20.7	273.7 ± 20.6	1,3
Pisces II	Y	0.11 ± 0.11	$-0.24^{+0.12}_{-0.11}$	182.1 ± 15.0	-75.7 ± 8.7	$47.0^{+102.1}_{-93.7}$	$89.1^{+54.3}_{-49.9}$	
Reticulum II	Y	2.39 ± 0.01	-1.36 ± 0.02	33.0 ± 1.4	-92.2 ± 1.7	214.0 ± 2.2	233.0 ± 2.1	1,3
Reticulum III	P	0.36 ± 0.14	$0.05^{+0.19}_{-0.25}$	92.0 ± 13.0	113.4 ± 17.9	$78.0^{+83.1}_{-109.1}$	$137.6^{+49.4}_{-63.6}$	
Sagittarius II	Y	-0.77 ± 0.03	-0.89 ± 0.02	63.0 ± 2.3	-115.7 ± 1.8	239.0 ± 10.4	265.5 ± 9.4	1
Sculptor	Y	0.099 ± 0.002	-0.16 ± 0.002	86.1 ± 5.0	76.3 ± 0.3	163.0 ± 0.0	180.0 ± 0.1	1,3
Segue 1	Y	-2.21 ± 0.06	-3.34 ± 0.05	28.0 ± 1.9	136.9 ± 0.0	240.0 ± 6.5	276.3 ± 5.7	1,3
Segue 2	Y	1.47 ± 0.04	-0.31 ± 0.04	43.1 ± 3.0	59.9 ± 5.0	134.0 ± 7.2	146.8 ± 6.9	1
Sextans I	Y	-0.41 ± 0.01	0.04 ± 0.01	98.1 ± 3.0	88.1 ± 0.1	220.0 ± 4.0	237.0 ± 3.7	1
Triangulum II	P	0.56 ± 0.05	0.07 ± 0.06	34.8 ± 1.6	-265.6 ± 3.2	159.0 ± 8.9	309.5 ± 5.3	1
Tucana II	Y	0.9 ± 0.02	-1.26 ± 0.02	54.2 ± 7.9	-182.0 ± 4.6	210.0 ± 8.1	277.9 ± 6.9	1
Tucana III	P	-0.08 ± 0.01	-1.62 ± 0.02	23.0 ± 1.9	-223.4 ± 2.6	126.0 ± 2.0	256.5 ± 2.5	1,3
Tucana IV	Y	0.54 ± 0.06	-1.67 ± 0.07	45.4 ± 3.9	-90.6 ± 5.3	197.0 ± 16.0	216.9 ± 14.7	1
Tucana V	P	$-0.14^{+0.06}_{-0.05}$	$-1.15^{+0.08}_{-0.06}$	51.8 ± 8.9	-157.8 ± 5.9	$181.0^{+30.9}_{-24.9}$	$240.1^{+23.6}_{-19.1}$	1
Ursa Major I	Y	-0.39 ± 0.03	-0.63 ± 0.03	102.1 ± 5.8	-0.8 ± 0.7	126.0 ± 20.8	126.0 ± 20.8	1
Ursa Major II	Y	1.72 ± 0.02	-1.89 ± 0.03	41.0 ± 1.9	-64.5 ± 1.6	262.0 ± 6.5	269.8 ± 6.4	1
Ursa Minor	Y	-0.124 ± 0.004	0.078 ± 0.04	77.9 ± 4.0	-83.4 ± 1.4	148.0 ± 0.0	169.9 ± 0.7	1
Willman 1	Y	0.21 ± 0.06	-1.08 ± 0.09	49.7 ± 9.9	16.7 ± 1.9	120.0 ± 17.5	121.2 ± 17.3	1

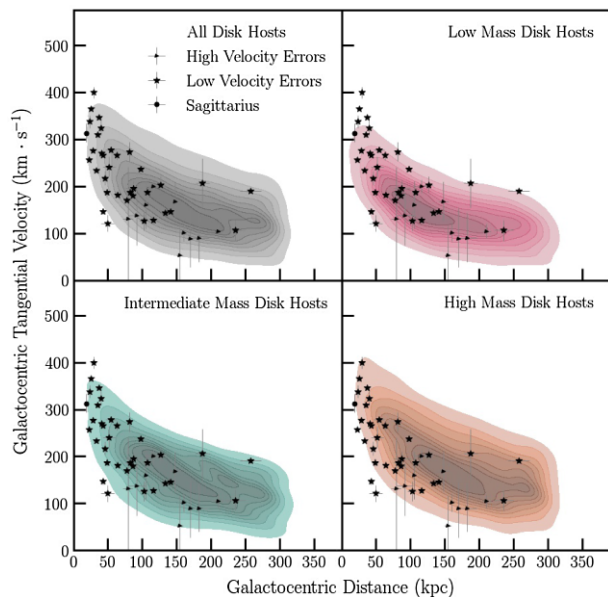


Figure 1. Comparison of phase-space distributions between Phat ELVIS subhaloes (density contours) and satellites in McConnachie & Venn (2020a) (MCV20a, black markers). In each plot, the stars denote satellites with low tangential velocity errors ($V_{\text{tan, err}} \leq 0.30 V_{\text{tan}}$). The triangles denote all other satellites excluding Sagittarius, which is represented as the circle. Sagittarius is poorly reproduced by the simulations and thus omitted from this analysis. The total velocity information for Sagittarius comes from Fritz et al. (2018), which uses *Gaia* Data Release 2 proper motions. The total velocity errors are taken from MCV20a while the distance errors are taken from the literature (mainly Simon 2019 and references therein). The top left panel shows the phase-space density contours for subhaloes across all 12 Phat ELVIS hosts. Meanwhile, in the three remaining panels, we display the corresponding contours with subhaloes divided according to host mass – $0.7\text{--}1 \times 10^{12} M_{\odot}$ (low mass, burgundy shading), $1\text{--}1.2 \times 10^{12} M_{\odot}$ (intermediate mass, aqua shading), and $1.4\text{--}2 \times 10^{12} M_{\odot}$ (high mass, sienna shading). Each mass bin includes 4 hosts, with the adopted colour scheme for the host mass sets carried throughout this paper.

of isolated MW-like haloes and 12 re-runs of those DMO simulations with an embedded galaxy potential matching the observed MW disc and bulge (from here on referred to as the disc runs). The 12 disc runs begin as identical duplicates to the 12 DMO suites. At $z = 3$, a galaxy potential, including a stellar disc, gaseous disc, and Hernquist bulge component, is inserted into each of the disc hosts.² While the potentials temporally evolve, each disc host ends up at $z = 0$ as an observationally constrained MW (Kelley et al. 2019).

Each simulation occurs within a global cosmological box of length 74.06 Mpc ($50 h^{-1}$ Mpc) with a dark matter particle mass of $3 \times 10^4 M_{\odot}$ and a Plummer-equivalent force softening length of $\epsilon = 37$ parsecs. These parameters allow for the subhalo catalogues to be complete down to a maximum circular velocity of $V_{\text{max}} > 4.5 \text{ km s}^{-1}$, i.e. a total bound mass of $\gtrsim 5 \times 10^6 M_{\odot}$. Refer to fig. 2 in Kelley et al. (2019) for a visualization of completeness limits – in V_{peak} cumulative distributions, roll-off of the functions begin at low V_{peak} (at approximately $V_{\text{peak}} \approx 6 \text{ km s}^{-1}$ within 50 kpc), yet throughout the radii explored, there is no roll-off in V_{max} going as far down as $V_{\text{max}} = 4.5 \text{ km s}^{-1}$. This serves as a measure of incompleteness. The

phELVIS halo catalogues are constructed of 152 snapshots, evenly spaced in scale factor with a time resolution of roughly 100 Myr. We spline interpolate the subhalo positions and velocities to achieve a time resolution of ~ 10 Myr and use the interpolated data to calculate only pericentric passage and infall time. Phat ELVIS adopts the cosmology of Planck Collaboration XIII (2016) with the following parameters: $\Omega_m = 0.3121$, $\Omega_{\Lambda} = 0.6879$, and $h = 0.6751$.

In order to more directly compare to the MW satellite population, we select subhaloes from the phELVIS suite with $V_{\text{peak}} > 6 \text{ km s}^{-1}$. While suppression of galaxy formation due to reionization is often predicted to occur below a mass limit of $V_{\text{peak}} \sim 20\text{--}25 \text{ km s}^{-1}$ (e.g. Gnedin 2000; Hoeft et al. 2006; Ocvirk et al. 2016), the observed abundance of ultrafaint satellites of the MW are better matched via a lower mass limit (Graus et al. 2019). Adopting a more inclusive mass selection yields a considerably larger subhalo population for comparison, better sampling the host potential and allowing control of systematics associated with observational completeness. Radial profiles and infall times are both potentially biased by V_{peak} limits. Recent work has shown that subhalo radial profiles are largely independent of V_{peak} limits (Newton et al. 2018) and investigating the infall- V_{peak} relation in phELVIS illustrated that there is a tight correlation amongst distributions of subhalo infall times throughout our V_{peak} range, where the median difference between distributions is 0.6 Gyr.

While the phELVIS subhaloes catalogues at $z = 0$ provide thousands of subhaloes for comparison, they are limited to a single snapshot of each subhalo orbit. To better sample the host potential, we expand our subhalo population to include subhaloes at two earlier snapshots. These two other snapshots were selected based on the average growth histories of the phELVIS hosts to minimize variation in the host mass. A majority of the hosts have minor ($\lesssim 2$ per cent) to no growth after $z = 0.05$, which corresponds to the original 8th time-step prior to $z = 0$. The two snapshots chosen to examine here are evenly spread – specifically the 8th and 4th (corresponding to $z = 0.05, 0.02$, respectively). The vast majority (~ 90 per cent) of the subhalo population is present at all 3 time-steps, with a small number of subhaloes missing (or added) at earlier time-steps due to recent accretion, backslashing, and/or tidal destruction.

Throughout this work, we focus on the 12 MW-like hosts in the disc runs. The hosts with embedded disc potentials are chosen for their ability to better represent the observations relative to the dark matter-only (DMO) hosts. In the DMO runs, the greatest subhalo Galactocentric total velocities, which are all found at small Galactocentric distances, are systematically lower than those found in the disc runs. This trend is seen in all three of the host halo mass ranges displayed in Fig. 1 and further strengthens the argument initially made in Kelley et al. (2019) – central-galaxy dynamics must be included to match observations of the satellite population. The disc host haloes range in virial mass³ from $0.71\text{--}1.95 \times 10^{12} M_{\odot}$. We split the hosts evenly into 3 groups based on mass – least massive, intermediate mass, and most massive. Specifically, the mass ranges of the 3 bins are: $0.71\text{--}0.96 \times 10^{12} M_{\odot}$ (low mass), $1.04\text{--}1.20 \times 10^{12} M_{\odot}$ (intermediate mass), and $1.40\text{--}1.95 \times 10^{12} M_{\odot}$ (high mass). Excluding subhaloes with $V_{\text{peak}} < 6 \text{ km s}^{-1}$, there are an average of 1200 subhaloes ($< R_{\text{vir}}$) associated with each of the 4 low-mass hosts, in comparison to an average of 1400 (2100) subhaloes for each of the 4 intermediate-mass (high-mass) hosts. Within each

²At $z = 0$ the inserted masses are $M_{\text{Stellar disc}} = 4.1 \times 10^{10} M_{\odot}$, $M_{\text{Gas disc}} = 1.9 \times 10^{10} M_{\odot}$, and $M_{\text{Bulge}} = 0.9 \times 10^{10} M_{\odot}$, see table 1 of Kelley et al. (2019) for more details.

³In the phELVIS simulations, virial mass, M_{vir} , follows the Bryan & Norman (1998) definition.

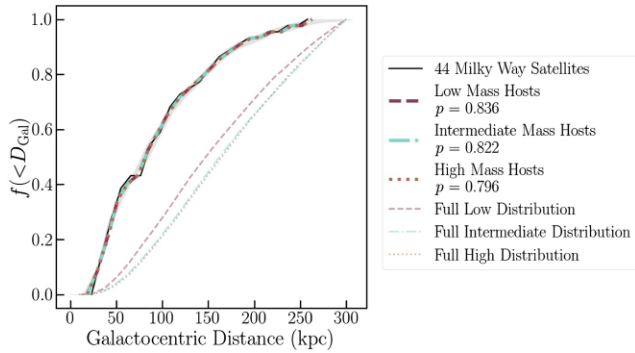


Figure 2. The cumulative distributions of MW-centric and host halocentric physical distances are engineered to be nearly identical via our distance-matching scheme. The solid black line is the cumulative distribution of Galactocentric distances for our primary sample of 44 satellite galaxies. The dashed burgundy line, dashed-dotted aqua line, and the dotted sienna line are the corresponding cumulative distributions for subhaloes drawn from the three host halo sets, where Galactocentric distance is measured with respect to the corresponding host halo. The thinner dashed, dashed-dotted and dotted lines are the cumulative distributions of all the subhaloes in the respective host mass bins. The (thin) grey shaded region is the range of reported errors in the literature (see Simon 2019 and references therein). The Mann–Whitney U Test p -value statistic is calculated using these errors. The legend reports the harmonic mean of MWU p -values from 500 randomly selected, distance-matched subhalo distributions. This tight comparison (as we cannot reject the null hypothesis that the distributions are drawn from the same parent population due to the two-sided p -values all being well above the statistical significance level of 0.05), reduces the potential biases associated with the incompleteness of the MW satellite population.

of the 3 host mass bins, the halo-to-halo scatter in subhalo count is not great. The normalized, cumulative distribution of each bin is approximately the average of the 4 individual host distributions that comprise that particular bin. As shown in Fig. 1, due to the tidal disruption of subhaloes in the disc runs, phELVIS includes exceedingly few analogues to the Sagittarius dwarf. As discussed in Section 2.1, for this reason Sgr is excluded from the sample of MW satellites studied.

3 ANALYSIS

To study how the dynamics of the MW satellites depend on host halo mass, we select subhaloes from the phELVIS simulations from each of the host mass divisions discussed in Section 2.2. We then compare the subhalo samples to the observational data set via Galactocentric velocities – namely, radial, tangential, and total. To mitigate selection effects driven by incompleteness in the sample of MW satellites, we match our sample of MW satellites to phELVIS haloes via Galactocentric distance. For each of the three sets of host haloes, we randomly select (with replacement) 10 subhaloes for each satellite, selecting the subhaloes from distance bins of width 10 kpc centred on the Galactocentric distance of the satellite, where there is an average of 600 subhaloes in each satellite’s distance bin. The size of the distance bins encompasses the observational radial distance errors for ~ 70 per cent of the MW satellites. These distance-matched subhaloes are randomly selected from the parent catalogue that combines the subhalo populations from all three time-steps. For our primary sample of 44 satellites, this produces three comparison samples of 440 haloes each, associated with the low-mass, intermediate-mass, and high-mass hosts. As shown in Fig. 2, this distance-based matching enforces a very

close correlation between the Galactocentric distances of the MW satellite sample and our comparison subhalo samples. One caveat to this method of matching – it does not guarantee each distance-matched halo is located within R_{vir} of the host. For example, Leo I’s Galactocentric distance is 258 kpc while roughly only 25 per cent of the explored phELVIS host haloes have virial radii greater than this distance. Thus, some of the subhaloes, drawn from the lower mass hosts and distanced matched to Leo I, may reside beyond R_{vir} . We choose not to match subhaloes to observed systems on normalized distance (i.e. Galactocentric distance that has been normalized to the MW/host’s virial radius) to avoid introducing biases associated with the boundedness of a system at or near the virial radius or possible tidal disruption for systems near the host halo’s centre.

To quantify the observation-to-simulation comparisons (i.e. to measure if the distanced-matched haloes do not represent the MW satellites, an ‘*antigoodness of fit*’), we employ the Mann–Whitney U (MWU) test (Mann & Whitney 1947; Jones et al. 2001). This is a non-parametric statistical ranked summation test that examines two independent samples. This test does not require any knowledge of the underlying distribution in either of the independent samples. To avoid possible underlying biases in the ranked summation (Fong & Huang 2019), we increase our observational sample size by randomly sampling the observational quantities’ errors to match the size of the distance matched halo population (i.e. 440 observational values are compared to 440 simulated values). For satellite characteristics that have asymmetric error distributions (e.g. for V_{tan} and V_{tot}), the errors are drawn equally from the positive and negative sides. No galaxy in our observational set has plus-minus errors that are extremely different from one another. Fortunately, the MWU is attuned to only median changes, compared to say the Kolmogorov–Smirnov (KS) test which is sensitive to the shape of the underlying distributions as well as the medians. The null hypothesis for the MWU is that the two-independent samples are, in fact, drawn from the same parent distribution. We report the MWU test results as the associated two-sided p -values. Our statistical significance level to reject the null hypothesis is set at $p \leq 0.05$.

In an additional step towards bias avoidance, we conduct Monte Carlo sampling, by randomly selecting input values for the MWU test from the observational errors, 500 times per parameter. This is to reduce the possibility of sampling a randomly skewed distribution, and incorporate the observational errors into our modelling. In the two-sided MWU test, we use 500 sets of 10N distance-matched subhaloes, which are randomly selected from 10 kpc distance bins (which encompass the radial distance errors for most satellites), and 10N measures of corresponding satellite properties, as drawn from the observed error distributions. A p -value is calculated for each MWU test. We then take the harmonic mean of these 500 p -values and use this as our statistical result.

To create a baseline to our distance-matched analysis, we compare the unmatched subhalo distributions in the same way as comparing the matched distributions. Here, we compare 10N subhalo properties randomly chosen from any subhalo in the host halo set (independent of distance) to a set of 10N values randomly sampled from the error distributions of the observational quantities. We do this twice – once where $N = 44$ for the full set of satellites and $N = 34$ for the set of systems with proportionally low tangential velocity errors. Finally, we create the random samples 500 times, comparing the unmatched set to the observational set each time and then take the harmonic mean of the resulting p -values. As detailed in Table 3, the comparison of Galactocentric distances for both observational sets to each of the 3 unmatched host halo sets are rejected at greater than 5σ , where 5σ maps to $p = 0.00001$ as determined by our choice of significance

Table 2. Orbital and infall properties of the MW satellite galaxies used in this work. Columns (1)–(3) Pericentric passage distances, in kpc, for each system in a low, intermediate, and high MW potential, with an NFW profile and concentration of 15.3, via GALPY. The corresponding MW dark matter halo masses used for the 3 potentials are 0.835 , 1.120 , $1.675 \times 10^{12} M_{\odot}$, respectively. These are the average masses of the hosts in each of our fiducial phELVIS host sets. Columns (4)–(6) Orbital eccentricities for each system in the three GALPY MW potentials, with smaller eccentricities corresponding to more circular orbits and larger eccentricities corresponding to more radial/plunging orbits. Column (7) Satellite infall times, in Gyr, as derived by Fillingham et al. (2019).

Galaxy	$D_{\text{peri, l}}$ (kpc)	$D_{\text{peri, i}}$ (kpc)	$D_{\text{peri, h}}$ (kpc)	e_l	e_i	e_h	t_{infall} (Gyr)
Antlia II	69.8 ± 13.1	54.9 ± 10.8	43.4 ± 8.3	0.4 ± 0.0	0.5 ± 0.0	0.5 ± 0.0	–
Aquarius II	92.7 ± 45.2	72.5 ± 35.8	54.6 ± 25.5	0.2 ± 0.3	0.2 ± 0.3	0.3 ± 0.3	$1.6^{+5.4}_{-3.5}$
Boötes I	43.9 ± 3.1	37.4 ± 3.2	30.8 ± 2.9	0.4 ± 0.0	0.4 ± 0.0	0.4 ± 0.0	$10.7^{+0.6}_{-1.9}$
Boötes II	39.1 ± 1.0	39.0 ± 1.0	38.8 ± 1.1	0.9 ± 0.0	0.7 ± 0.0	0.4 ± 0.0	1.1 ± 0.6
Canes Venatici I	47.4 ± 33.4	37.8 ± 25.7	30.7 ± 19.8	0.7 ± 0.1	0.7 ± 0.1	0.8 ± 0.1	$9.4^{+0.9}_{-2.3}$
Canes Venatici II	9.8 ± 32.0	8.3 ± 25.5	7.2 ± 20.4	0.9 ± 0.2	0.9 ± 0.2	0.9 ± 0.2	$9.0^{+1.0}_{-2.8}$
Carina	107.3 ± 5.1	106.5 ± 5.7	83.6 ± 25.3	0.3 ± 0.2	0.1 ± 0.1	0.1 ± 0.1	$9.9^{+0.6}_{-2.7}$
Carina II	37.1 ± 0.6	27.0 ± 0.7	25.9 ± 0.7	0.9 ± 0.0	0.8 ± 0.0	0.7 ± 0.0	$7.9^{+2.5}_{-2.4}$
Carina III	28.9 ± 0.6	28.5 ± 0.6	28.4 ± 0.6	1.0 ± 0.0	0.8 ± 0.1	0.6 ± 0.1	$7.6^{+2.4}_{-2.7}$
Columba I	185.7 ± 10.7	185.0 ± 10.8	182.8 ± 10.7	0.7 ± 0.0	0.5 ± 0.1	0.2 ± 0.1	–
Coma Berenices	42.5 ± 1.6	42.3 ± 1.6	41.4 ± 1.9	0.5 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	$10.2^{+2.6}_{-3.3}$
Crater II	39.9 ± 8.4	31.9 ± 6.5	25.8 ± 5.0	0.6 ± 0.1	0.6 ± 0.1	0.7 ± 0.0	$7.8^{+2.7}_{-3.0}$
Draco	49.0 ± 5.9	41.2 ± 5.3	33.8 ± 4.3	0.5 ± 0.0	0.5 ± 0.0	0.5 ± 0.0	$10.4^{+2.4}_{-3.1}$
Draco II	19.9 ± 4.3	19.4 ± 4.4	18.7 ± 4.6	0.8 ± 0.1	0.6 ± 0.1	0.5 ± 0.1	$10.2^{+1.8}_{-2.4}$
Fornax	98.7 ± 12.6	73.4 ± 10.1	55.9 ± 7.2	0.2 ± 0.0	0.3 ± 0.0	0.4 ± 0.0	$10.7^{+0.8}_{-3.1}$
Grus I	17.1 ± 9.2	14.7 ± 7.8	12.7 ± 6.5	0.9 ± 0.0	0.9 ± 0.0	0.9 ± 0.0	$1.1^{+1.0}_{-0.9}$
Grus II	26.0 ± 9.4	22.4 ± 9.0	18.9 ± 7.8	0.5 ± 0.0	0.5 ± 0.1	0.5 ± 0.1	–
Hercules	70.2 ± 62.3	61.4 ± 13.2	52.0 ± 12.5	0.7 ± 0.1	0.6 ± 0.0	0.6 ± 0.0	$6.6^{+2.3}_{-0.7}$
Horologium I	86.1 ± 12.7	85.6 ± 13.1	83.3 ± 15.2	0.6 ± 0.3	0.3 ± 0.4	0.1 ± 0.3	$8.8^{+1.8}_{-2.0}$
Horologium II	49.3 ± 37.2	37.4 ± 44.9	29.3 ± 50.4	0.3 ± 0.5	0.4 ± 0.1	0.5 ± 0.2	–
Hydra II	62.6 ± 7.5	52.8 ± 6.8	43.8 ± 5.7	0.7 ± 0.0	0.7 ± 0.0	0.7 ± 0.0	$9.4^{+1.7}_{-1.8}$
Hydrus I	25.5 ± 0.5	25.5 ± 0.5	25.5 ± 0.5	0.9 ± 0.0	0.7 ± 0.0	0.5 ± 0.0	$10.7^{+1.3}_{-1.4}$
Leo I	257.7 ± 15.5	257.7 ± 15.5	35.6 ± 6.4	0.7 ± 0.0	0.5 ± 0.0	0.9 ± 0.0	$10.5^{+1.5}_{-2.4}$
Leo II	120.3 ± 24.0	88.1 ± 16.4	67.2 ± 11.6	0.3 ± 0.1	0.5 ± 0.0	0.6 ± 0.0	$2.3^{+0.6}_{-0.5}$
Leo IV	20.2 ± 107.7	16.5 ± 72.9	13.7 ± 52.7	0.8 ± 0.7	0.8 ± 0.5	0.8 ± 0.4	$7.8^{+3.3}_{-2.0}$
Leo V	32.5 ± 89.2	26.1 ± 70.1	21.3 ± 52.8	0.7 ± 0.4	0.7 ± 0.4	0.8 ± 0.3	10.4 ± 1.4
Phoenix II	80.3 ± 4.3	80.1 ± 4.4	79.6 ± 4.6	0.9 ± 0.0	0.6 ± 0.1	0.4 ± 0.1	–
Pisces II	28.7 ± 108.7	23.3 ± 88.5	19.2 ± 68.0	0.8 ± 0.4	0.8 ± 0.4	0.8 ± 0.4	–
Reticulum II	28.7 ± 2.3	27.3 ± 2.7	24.8 ± 3.3	0.5 ± 0.1	0.3 ± 0.0	0.3 ± 0.0	8.3 ± 1.8
Reticulum III	17.2 ± 87.7	14.4 ± 58.5	12.1 ± 52.9	0.8 ± 0.1	0.8 ± 0.2	0.8 ± 0.3	–
Sagittarius II	51.1 ± 2.1	48.0 ± 2.0	43.1 ± 1.8	0.6 ± 0.0	0.5 ± 0.0	0.4 ± 0.0	–
Sculptor	58.9 ± 1.6	48.4 ± 1.1	38.8 ± 0.8	0.4 ± 0.0	0.4 ± 0.0	0.4 ± 0.0	$10.6^{+1.6}_{-1.9}$
Segue 1	21.4 ± 3.1	20.2 ± 3.5	18.5 ± 3.8	0.6 ± 0.1	0.5 ± 0.1	0.4 ± 0.0	$9.9^{+1.7}_{-2.9}$
Segue 2	22.1 ± 8.3	17.9 ± 6.8	14.7 ± 5.4	0.4 ± 0.1	0.4 ± 0.1	0.5 ± 0.1	$10.8^{+1.3}_{-1.4}$
Sextans I	97.9 ± 3.0	82.5 ± 3.9	75.4 ± 4.5	0.6 ± 0.1	0.4 ± 0.0	0.3 ± 0.0	$10.8^{+1.6}_{-1.9}$
Triangulum II	11.7 ± 1.3	10.7 ± 1.2	9.7 ± 1.1	0.9 ± 0.0	0.8 ± 0.0	0.8 ± 0.0	$8.4^{+2.7}_{-0.9}$
Tucana II	38.6 ± 11.9	36.6 ± 12.6	33.8 ± 13.4	0.9 ± 0.1	0.7 ± 0.2	0.6 ± 0.1	$9.5^{+1.5}_{-2.1}$
Tucana III	3.0 ± 0.6	2.8 ± 0.4	2.5 ± 0.5	0.9 ± 0.0	0.9 ± 0.0	0.9 ± 0.0	–
Tucana IV	36.1 ± 6.3	32.5 ± 7.1	27.6 ± 7.0	0.4 ± 0.1	0.3 ± 0.0	0.3 ± 0.0	–
Tucana V	27.9 ± 9.4	24.7 ± 8.7	21.4 ± 7.7	0.6 ± 0.0	0.6 ± 0.0	0.6 ± 0.0	–
Ursa Major I	70.6 ± 3.7	51.4 ± 2.6	39.3 ± 1.9	0.2 ± 0.0	0.3 ± 0.0	0.4 ± 0.0	$9.5^{+2.4}_{-2.8}$
Ursa Major II	39.3 ± 2.2	38.9 ± 2.4	38.1 ± 2.7	0.7 ± 0.1	0.5 ± 0.1	0.3 ± 0.1	$1.5^{+5.1}_{-1.6}$
Ursa Minor	44.8 ± 8.1	36.4 ± 7.0	29.5 ± 5.6	0.4 ± 0.0	0.4 ± 0.0	0.5 ± 0.0	$10.7^{+1.4}_{-2.3}$
Willman 1	24.6 ± 21.4	19.4 ± 15.2	15.7 ± 11.2	0.3 ± 0.2	0.4 ± 0.2	0.5 ± 0.1	$10.7^{+1.7}_{-2.0}$

Table 3. Mann–Whitney U Test p -values for the observation to distance-matched subhalo comparisons. The reported p -values are the harmonic means of 500 MWU tests conducted on sets of 10N distance-matched subhalo properties and 10N MW satellite properties drawn from the properties' errors. These p -values are two-sided, i.e. $p_{\max} = 1$. Our statistical significance level is set at $p = 0.05$. Values below this level reject the null hypothesis that the two compared distributions are drawn from the same parent population. Values above our chosen significance level signify that the null hypothesis cannot be rejected and are highlighted in yellow. Values are reported for our fiducial phELVIS host sets for all the various groups of satellites discussed throughout the paper.

	D_{MW}	V_{Rad}	V_{Tan}	$V_{3\text{D}}$
All satellites (44) to unmatched subhaloes				
Low-mass hosts	<0.001	<0.001	<0.001	<0.001
Intermediate-mass hosts	<0.001	0.003	<0.001	<0.001
High-mass hosts	<0.001	0.367	<0.001	<0.001
Satellites with proportionally low				
Tangential velocity errors (LTVE) (34) to unmatched subhaloes				
Low-mass hosts	<0.001	<0.001	<0.001	<0.001
Intermediate-mass hosts	<0.001	0.001	<0.001	<0.001
High-mass hosts	<0.001	0.384	<0.001	<0.001
All satellites (44)				
Low-mass hosts	0.836	<0.001	0.142	0.006
Intermediate-mass hosts	0.822	0.735	0.005	0.013
High-mass hosts	0.796	<0.001	<0.001	<0.001
Satellites with proportionally low				
Tangential velocity errors (LTVE) (34)				
Low-mass hosts	0.767	<0.001	0.003	<0.001
Intermediate-mass hosts	0.744	0.896	0.098	0.141
High-mass hosts	0.702	<0.001	<0.001	<0.001
LTVE satellites excluding				
Satellites associated with the LMC (25)				
Low-mass hosts	0.823	0.003	0.019	0.003
Intermediate-mass hosts	0.759	0.722	0.081	0.057
High-mass hosts	0.778	0.001	<0.001	<0.001
LTVE satellites with $D_{\text{MW}} < 100$ kpc (25)				
Low-mass hosts	0.644	<0.001	0.012	<0.001
Intermediate-mass hosts	0.641	0.061	0.005	0.044
High-mass hosts	0.565	0.032	<0.001	<0.001
LTVE satellites excluding satellites associated				
with the LMC and using limit of $V_{\text{peak}} > 10 \text{ km s}^{-1}$ (25)				
Low-mass hosts	0.769	0.060	0.015	0.003
Intermediate-mass hosts	0.798	0.055	0.049	0.008
High-mass hosts	0.730	<0.001	<0.001	<0.001

level ($p = 0.05$) and the fact that we calculate a two-sided p -value. The tangential and total velocity comparisons are also rejected at greater than 5σ for all three sets. The radial velocity p -values are a bit different in these comparisons. For the low- and intermediate-mass host haloes sets, the p -values are rejected at greater than or near 3σ , with 3σ mapping to $p = 0.0027$. The comparison for the high-mass host halo set in radial velocity cannot be rejected ($p > 0.05$) – this is the only non-rejectable null hypothesis between the unmatched subhalo distributions and the observations.

4 RESULTS

We refine halo mass constraints for the MW by comparing the MW satellites' Galactocentric velocities to distance-matched distributions of Phat ELVIS subhaloes split into 3 groups based on host halo virial mass. The three host halo mass bins range from $<10^{12}$ to $\sim 2 \times 10^{12} M_{\odot}$, with $\sim 10^{12} M_{\odot}$ being the intermediate bin. We

focus on two sets of satellites drawn from the [MCV20a](#) sample – all MW satellites and satellites with proportionally small tangential velocity errors (see Section 2.1). These two subsets included 44 and 34 satellites, respectively.

The subhalo distribution from each of the three host halo sets is well matched in distance to each of the two main satellite sets, by design (see fig. 2 and Section 3). Given these subhalo samples that are well matched on Galactocentric distance to the observed MW satellite population, the velocity distributions of each satellite set are then compared to the corresponding measure for the distance-matched subhaloes from the 4 highest mass host haloes, the 4 intermediate-mass host haloes, and the 4 lowest mass host haloes. This results in three harmonic mean p -values for the each of the velocity components. When $p \leq 0.05$, the MWU test's null hypothesis can be rejected, which equates to the galaxy sample being poorly represented by the distance-matched subhaloes in a specific host halo mass set (based on that particular measure of velocity).

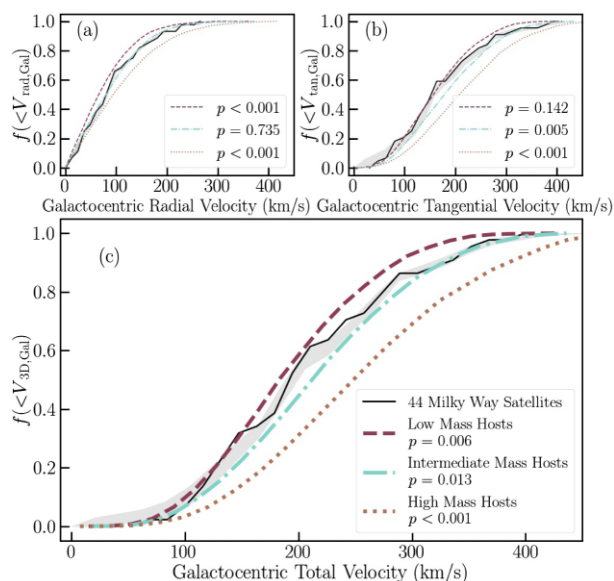


Figure 3. Cumulative distributions of Galactocentric velocities – namely in panel (a) V_{rad} , in panel (b) V_{tan} , and in panel (c) V_{tot} – for the MW satellites in comparison to that of the simulated subhaloes. The solid black line is the distribution for the 44 satellite galaxies in the MCV20a sample. The dashed burgundy line, dashed-dotted aqua line, and the dotted sienna line are the distributions for subhaloes drawn from the three bins in host mass. The grey shaded regions are the cumulative distribution of the range of 500 randomly sampled values from each systems reported errors. In panel (a), the comparison in the well constrained parameter of Galactocentric radial velocity results in the preference towards only an intermediate mass MW dark matter halo. Panels (b) and (c) tell a different story. These subplots display how the inclusion of the further phase-space information from *Gaia* shifts the preference towards a less massive MW in Galactocentric tangential velocity, in that the low-mass hosts yield the only non-rejected p -value, while no host halo mass range is consistent with the observed Galactocentric total velocities for this sample.

With tight distance-matched populations, we first examine the most well-constrained kinematic property – Galactocentric radial velocity (V_{rad}). As illustrated in panel (a) of Fig. 3, radial velocities for the main set of 44 satellites are in good agreement with the corresponding velocities for subhaloes in the intermediate-mass host halo set, while the subhaloes drawn from the low-mass and high-mass hosts are inconsistent with the observations at $\gtrsim 5\sigma$ ($p < 0.0001$).

Since the strength of *Gaia*’s data is the ability to calculate full 6D phase-space, we take the analysis a step further by incorporating the not as richly studied Galactocentric total and tangential velocities. For the main set of 44 satellites, the preference for an intermediate-mass host halo is *not* evident when examining either of these two velocities – i.e. the associated p -values reject the null hypothesis that the two samples are drawn from the same parent distribution. As shown in panels (b) and (c) of Fig. 3, the subhaloes of the intermediate-mass hosts are inconsistent with the observational set. In panel (b), the tangential velocity comparison prefers the low-mass hosts, i.e. this is the only non-rejected p -value, while in panel (c) the total velocity comparison rules out all host mass ranges – i.e. all samples yield $p < 0.05$.

A caveat to the *Gaia*-derived velocities is that a significant group of the observed systems have proportionally large errors associated with their proper motions which translates to proportionally large errors associated with the system’s tangential and total velocities. These larger uncertainties allow for the possibility of extreme velocities

that are not well represented in the phELVIS simulations. Inclusion of these systems in the analysis potentially creates a bias primarily towards lower mass hosts. The systems in the MCV20a sample with proportionally high tangential velocity errors have the lowest tangential velocities of all 44 satellites and are at distances further than roughly half the sample. These kinematically cool systems become more rare with increasing host halo mass – in higher mass hosts, hotter systems are the norm. For example, Leo IV is one such system with proportionally large tangential velocity errors (i.e. $|V_{\text{tan, err}}/V_{\text{tan}}| \geq 0.30$). Of the 1108 haloes within ± 5 kpc of Leo IV’s distance (154.59 ± 4.99 kpc) in the 4 highest mass host haloes, there are exactly 0 haloes with a tangential velocity in the bottom range of Leo IV’s 1σ tangential velocity error ($V_{\text{tan}} < 14 \text{ km s}^{-1}$).

To address this potential bias, we rerun our distance-matching analysis using the subsample of 34 satellites with low fractional uncertainty in V_{tan} – specifically, we define proportionally low-error systems to have $|V_{\text{tan, err}}/V_{\text{tan}}| \leq 0.30$. We then compare the resulting velocity distributions for this pared-down set. As illustrated in Fig. 4, the radial velocity comparison, panel (a), is essentially unaffected by the removal of systems with proportionally high tangential velocity errors. However, the comparison in tangential velocity space, panel (b), now prefers the intermediate-mass hosts. The preferred halo mass in the total velocity comparison also changes from Fig. 3 to now prefer the intermediate-mass hosts. The low- and high-mass hosts in all three velocity component comparisons are rejected ($p < 0.05$) at or near 3σ . All p -values discussed here can be found in Table 3.

To explore the limits of this preferred intermediate-mass range, $1.04\text{--}1.20 \times 10^{12} M_{\odot}$, we rerun the analysis with thinner and wider intermediate-mass ranges. More specifically, we ran the analysis with a thinner intermediate mass range of $\sim 1.04\text{--}1.10 \times 10^{12} M_{\odot}$ (3 host haloes) and a wider intermediate mass range of $\sim 0.96\text{--}1.40 \times 10^{12} M_{\odot}$ (6 intermediate-mass host haloes instead of 4). We compare these varying host halo mass ranges to the observational set of systems with proportionally low tangential velocity errors. In the velocity comparisons for the thinner intermediate-mass host halo set, this new intermediate-mass host halo set of 3 haloes is preferred across all 3 velocity components. The 3 p -values for this host halo set are all greater than the significance limit of 0.05, while the p -values for the low- and high-mass hosts in all 3 velocity components are rejected at or near 3σ . In the velocity comparisons for the wider host mass set, there is not one host halo set preferred across all 3 velocity components. Though the intermediate-mass sample is nearly preferred across the components – the radial and total velocity p -values are greater than the significance limit while the tangential velocity p -value is just below this limit ($p = 0.045$). Overall, the results based on comparing the set of MW satellites with proportionally low tangential velocity errors to the phELVIS simulations strongly indicate that the *Gaia*-based distances and velocities are consistent with an MW dark matter halo mass of $\sim 1\text{--}1.20 \times 10^{12} M_{\odot}$.

5 DISCUSSION

5.1 Large Magellanic Cloud (LMC) Satellites

Of the 12 MW-like systems in the phELVIS suite with an embedded disc potential, there is only 1 host with a Large Magellanic Cloud-like subhalo (i.e. with $M_{\text{vir}} \geq 8 \times 10^{10} M_{\odot}$). If this restriction is lowered to $M_{\text{vir}} \geq 3 \times 10^{10} M_{\odot}$, then there are 2 Large Magellanic Cloud-like subhaloes throughout the 12 disc hosts. It can be argued that including MW satellites that were originally LMC satellites in

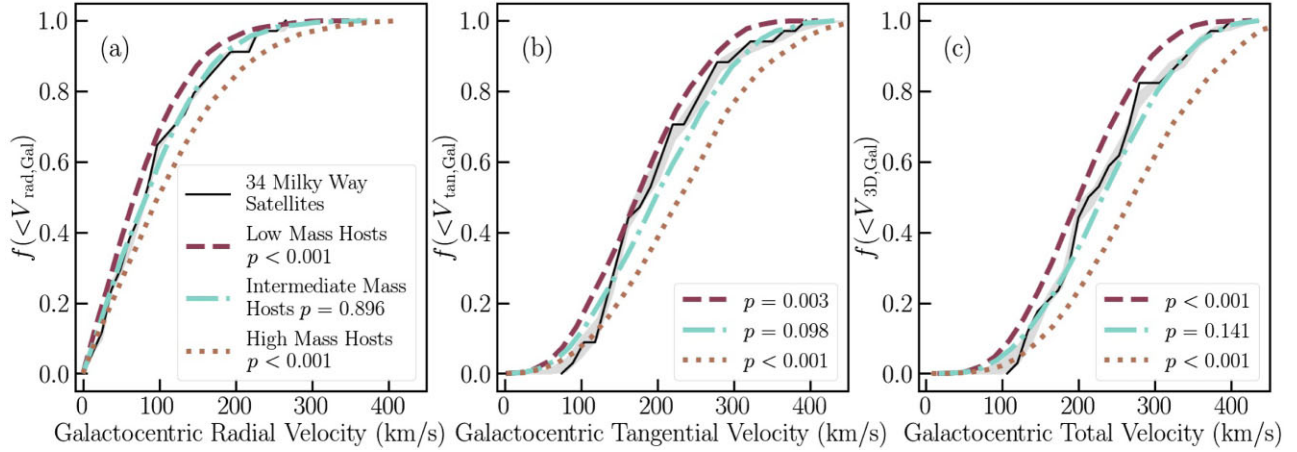


Figure 4. Cumulative distributions of all velocity components for the pared-down set of 34 satellites all with proportionally low tangential velocity errors. The colour coding, line style, and legend conventions are identical to Fig. 3. These three plots display how excluding satellites with their average plus-minus error ≥ 30 per cent of their tangential velocity settles the host mass preference on an intermediate mass MW. Specifically, compared to Fig. 3, the resulting p -values for the intermediate mass host haloes in the Galactocentric tangential and total velocities are now above the statistical significance level of 0.05 while the p -values for the low- and high-mass host haloes are rejected and therefore the intermediate mass hosts haloes is the preferred host mass range.

our analysis may bias our results, as these systems may not be well-represented in the pHELVIS simulation suite. Here, we explore the impact of removing LMC satellites from our observational sample of galaxies with low tangential velocity errors.

While there is some contention over which galaxies are satellites of the LMC, the derived proper motions from *Gaia* DR2 have allowed for more direct investigation into potential associations with the LMC. In addition to Horologium I, which has been found to be a likely LMC satellite in multiple studies (Sales et al. 2017; Erkal & Belokurov 2019; Patel et al. 2020; Santos-Santos et al. 2021), Carina II, Carina III, and Hydrus I have also been classified as long-term satellites of the LMC via their *Gaia* DR2 proper motions (Kallivayalil et al. 2018; Patel et al. 2020), where long term is defined by being bound to the LMC for at least 2 consecutive orbits. Furthermore, Patel et al. (2020) found another 5 galaxies to be recently captured LMC satellites (Reticulum II and Phoenix II) or have had prior interactions with the LMC (Sculptor, Segue 1, and Tucana III).

To explore how our results may be biased by the dynamical influence of the LMC, we fully rerun our distance-matching analysis on the set of MW satellites with low tangential velocity error fractions excluding the 9 LMC-associated satellites, which includes long-term satellites plus recent satellites and LMC interactors. When excluding these systems associated with the LMC, across all 3 velocity components, the null hypothesis is rejected when comparing to subhaloes drawn from the low- and high-mass hosts (i.e. $p < 0.05$), with the distribution of observed velocities for the Milky Way satellites intermediate between these two subhalo samples (i.e. again favouring an intermediate-mass Milky Way). However, while the radial velocity distribution for subhaloes drawn from the intermediate-mass sample is visually consistent with that of the Milky Way satellite population when excluding the LMC-associated satellites, we find that the tangential and total velocities are less consistent (as compared to the distributions in Fig. 4). As seen in Fig. 5 and Table 3, the intermediate-mass host samples yield non-rejected p -values, $p \geq 0.05$, when comparing to the observed radial, tangential and total velocities. Overall, when satellites associated with the LMC are removed from our analysis, the distance-matched subhaloes continue to show a preference for an intermediate-mass host halo ($\sim 1\text{--}1.2 \times 10^{12} M_{\odot}$).

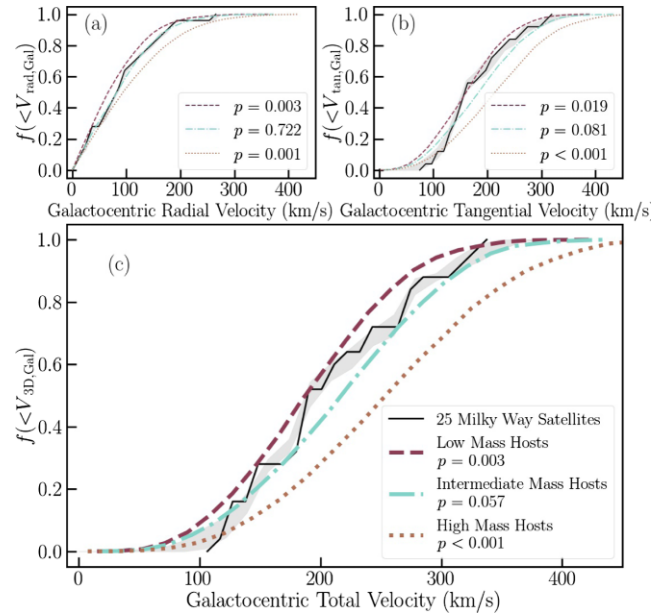


Figure 5. Cumulative distributions of individual Galactocentric velocity components for the set of (25) satellite galaxies excluding all 9 LMC-associated systems (and excluding satellites with proportionally high tangential velocity errors). The colour coding, line style, and legend conventions are identical to Fig. 3. The exclusion of the 9 LMC-associated systems does not change the intermediate host mass preference. Across all 3 velocity component comparisons, the p -values for the low- and high-mass host haloes are rejected ($p < 0.05$) while the intermediate-mass host halo p -values are above the significance limit.

So far in this work, all distance-matched subhaloes are selected according to a peak maximum circular velocity limit of $V_{\text{peak}} > 6 \text{ km s}^{-1}$. As shown by Graus et al. (2019), the abundance of Milky Way satellites can be reproduced with subhaloes down to $V_{\text{peak}} > 10 \text{ km s}^{-1}$ when excluding those systems associated with the LMC. While this more restrictive subhalo selection roughly quarters the subhalo populations, only Draco II does not have at least 10 subhaloes

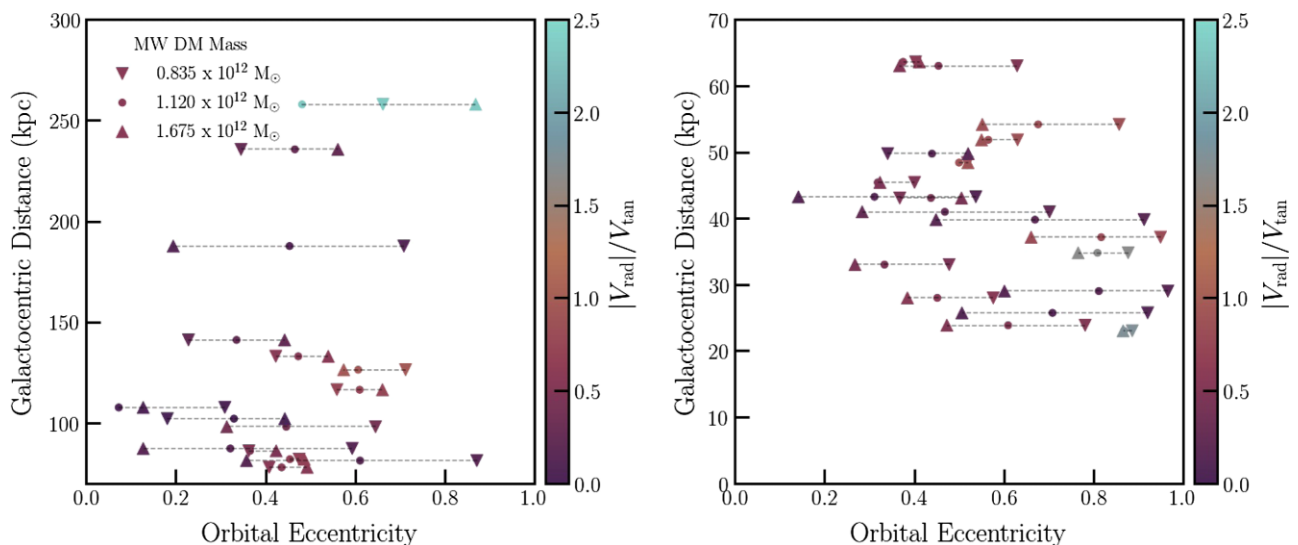


Figure 6. Galactocentric distance against orbital eccentricity for the MW satellites with proportionally low tangential velocity errors colour coded by the ratio of Galactocentric radial velocity versus tangential velocity. The satellites are separated into two groups: (*left-hand panel*) systems outside 70 kpc and (*right-hand panel*) systems inside 70 kpc. These figures display trends in eccentricity – decreasing eccentricity with increasing MW dark matter halo mass for those systems inside 70 kpc (*right-hand panel*) and the lack of a strong trend in eccentricity for those systems outside 70 kpc (*left-hand panel*). As seen by the colour coding, $|V_{\text{rad}}|/V_{\text{tan}} > 1$ indicates more radial orbits while those < 1 point towards more circular orbits. Eccentricity may serve well as host mass diagnostics.

in its distance bin across all host mass sets. As shown in Table 3, when limiting subhaloes to $V_{\text{peak}} > 10 \text{ km s}^{-1}$, there is no consistent host mass preference. The high-mass host samples are inconsistent with the observed Milky Way satellites across all 3 velocity measures. Meanwhile, the subhaloes drawn from the low- and intermediate-mass hosts are consistent with the observed radial velocities, but unable to reproduce the tangential and total velocity distributions of the Milky Way satellites.

5.2 Pericentric passage, eccentricity, and satellite infall

To explore the orbits of the observed MW satellites in various dark matter halo potentials, we calculate pericentres and orbital eccentricities, defined as $\text{eccentricity} = (r_{\text{apocentre}} - r_{\text{pericentre}})/(r_{\text{apocentre}} + r_{\text{pericentre}})$, for a low-, intermediate-, and high-mass MW potential using GALPY (Bovy 2015). We employ GALPY’s standard `MWPotential2014` potential model, which contains a spherical buldge, a Miyamoto Nagai Potential disc and an NFW dark matter halo with concentration 15.3, along with the EDR3 proper motions and other satellite properties from the literature (i.e. RA, Dec., heliocentric distance, and line-of-sight velocity). We modify the `MWPotential2014` by adopting 3 host potentials based on the average halo masses from our 3 host sets – i.e. average masses of 0.835, 1.120, $1.675 \times 10^{12} M_{\odot}$ – 1.04, 1.4, and, $2.09 \times \text{MWPotential2014}$, respectively. The resulting inferred orbital properties (3 pericentres and 3 eccentricities) for the MW satellites can be found in Table 2.

Comparing the resulting 3 pericentric passages derived using GALPY, there is a mild preference for smaller pericentres in increasing host potentials – e.g. for satellites 30–60 kpc from the centre of the Milky Way, the median pericentre in the largest host potential is ~ 30 kpc, while the median in the smallest potential is $\gtrsim 35$ kpc. This host mass–pericentre correlation is not as strong amongst the distance–matched phELVIS subhaloes, where the median pericentres in this same distance bin (30–60 kpc) is

roughly half the spread of that seen in the GALPY pericentres.⁴ While these predicted pericentres for all 44 satellites decrease with increasing host potential, the correlation between orbital eccentricity and host halo mass is more complicated. As illustrated in Fig. 6, for satellites currently in the outer MW halo ($70 < D_{\text{MW}}/\text{kpc} < 300$), there is not a strong correlation between host potential and orbital eccentricity. In the inner MW halo ($D_{\text{MW}} < 70$ kpc), however, there is a clear negative correlation, such that eccentricity decreases with increasing host potential for ~ 75 per cent of the satellites – i.e. orbits become more circular in greater potentials. The preferential circularization of orbits with increasing host mass is likely the result of satellite disruption associated with tidal forces. At a given host-centric distance within the inner halo, surviving satellites of more massive hosts tend to populate circular orbits, as tidal destruction has preferentially destroyed systems on more plunging orbits.

Beyond pericentre and eccentricity, another critical orbital parameter is the infall time on to the Milky Way (or host halo). Within the simulations, we are able to directly trace the infall of subhaloes, such that infall time is defined as the lookback time when a subhalo *first* crossed the host halo’s virial radius. For observed satellites of the MW, on the other hand, constraining the infall time is more challenging (given that we lack a DeLorean and a flux capacitor). Using *Gaia* proper motions from Fritz et al. (2018) to estimate the binding energy of each Milky Way satellite, Fillingham et al. (2019) estimate the infall time according to a correlation between infall time and binding energy derived for subhaloes in the phELVIS simulations (see also Rocha, Peter & Bullock 2012). For each satellite’s infall time, Fillingham et al. (2019) adopt the peak value in that satellite’s kernel density estimation (KDE) from binding energy-matched subhalo infall times. When computing the binding

⁴The subhalo pericentres are likely overestimated, as they were calculated after spline interpolating distances and velocities for each subhalo (Richings et al. 2020), which may affect the varying trends between data sets.

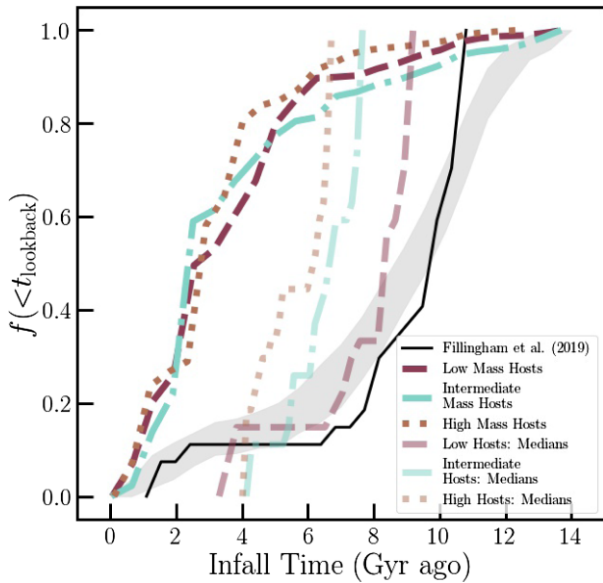


Figure 7. Cumulative distribution of satellite/subhalo infall times. The solid black line shows the cumulative distribution of first infall times for 26 MW satellites with low tangential velocity errors, as inferred by Fillingham et al. (2019). The grey shaded region is the cumulative distribution of the range of 500 randomly sampled values from each of the 26 systems' reported errors. The opaque dashed burgundy line, opaque dashed-dotted aqua line, and the opaque dotted sienna line are the corresponding distributions for all the distance-matched subhaloes belonging to our adopted host-mass bins. The half-transparent versions of these colorful lines are cumulative distributions of the median infall time for each set of subhaloes distance-matched to each galaxy in the Fillingham et al. sample. These median distributions mimic methods employed by Fillingham et al. showing better agreement to their distribution. In contrast, the opaque lines account for the full range of possible infall times which allows for later infall times to become more common, skewing those distributions to earlier cosmic times. As a whole, the distribution of infall times, as inferred from *Gaia* proper motions, does not match that found in the simulations, with the simulations favouring later (i.e. more recent) infall times.

energy of the Milky Way satellites, they assume a host halo mass of $1.3 \times 10^{12} M_{\odot}$, which is directly between that of our intermediate- and high-mass host halo samples. As illustrated in Fig. 7, however, the distribution of infall times within phELVIS is largely independent of host mass.

For the 26 MW satellites with proportionally low tangential velocity errors and infall times estimated by Fillingham et al. (2019), we draw distance-matched subhalo samples as described in Section 3. Fig. 7 shows cumulative distributions of infall times for all distance-matched subhaloes in this subsample (opaque colourful lines). We note that the distribution of infall times obtained in this way is different than using only a single value, such as the median of the distribution, for each set of subhaloes distance-matched to each galaxy, which is shown by the half-transparent lines. Accounting for the full range of possible infall times (opaque lines) instead of just the median value per galaxy (half-transparent lines) allows for later infall times to become more common, independent of the host mass. Interestingly our distribution of median infall times (half-transparent lines) are in reasonable agreement with estimates from Fillingham et al. (2019), which are based on energy-matching subhaloes instead of distance matching as in this work.

The inferred median infall times for the MW satellite population in phELVIS (half-transparent lines) are skewed to earlier cosmic times

relative to the distribution of most likely (KDE peak) infall times calculated for the MW (black solid line), an effect even stronger when allowing for the whole distribution of infall times for the distance matched subhaloes to be included (opaque lines). This suggests that it is possible that the MW may be an outlier with regard to its accretion history, such that a larger fraction of its satellites were accreted at early cosmic time (Elias et al. 2018), potentially via correlated accretion of substructures (D'Souza & Bell 2021). While the likely recent infall of the LMC (and associated satellites) would counter this potential bias in accretion history to some degree (Besla et al. 2007; Kallivayalil et al. 2013), the *possibly* anomalous satellite quenched fraction for the Milky Way might serve as further evidence of a bias towards early accretion – and excess quenching – relative to other nearby Milky Way-like systems (Wheeler et al. 2014; Fillingham et al. 2015; Geha et al. 2017).

5.3 Limitations and comparison to previous studies

One obvious limitation of our method is the underlying assumption that the formation history of the MW is represented in our sample of phELVIS simulations. Encouragingly, as discussed in Section 5.1, the low rate of large satellites with their own satellite systems (> 20 per cent phELVIS host haloes host an MC system) does not affect our results. A limitation that does impact our results is the coarse mass determination due to small host halo sample size – i.e. phELVIS has only 12 MW-like host haloes. Due to time and computational costs of running new, larger simulations, doing so to obtain a more precise result is beyond the scope of this paper.

Historically there has been a wide range of estimates of the MW's mass from $0.56 \pm 0.12 \times 10^{12} M_{\odot}$, via stellar stream modeling (using *N*-body realizations of the Sagittarius stream Gibbons, Belokurov & Evans 2014), to $2.65^{+1.58}_{-1.36} \times 10^{12} M_{\odot}$, via timing mass argument [using *Hubble Space Telescope* (*HST*) derived proper motions for Leo I Sohn et al. 2013]. Within the uncertainties these results do not agree with one another or with our results. Recent studies using stellar streams, such as Craig et al. (2021), who modelled the Magellanic Stream motivated by *HST* proper motions, found the mass to be $1.5 \pm 0.3 \times 10^{12} M_{\odot}$, which just overlaps our determination at 1σ . The timing mass argument constrains the Local Group mass, and MW mass within it, through present-day kinematics, the impending major merger between the MW and M31 as well as mass restrictions to have overcome universal expansion [see Bland-Hawthorn & Gerhard (2016), Benisty & Guendelman (2020), Benisty (2021) for thorough discussions on the method, its implications and limitations]. Recent results using the timing argument, and data from the H3 Spectroscopic Survey, find a range from 0.9 to $1.5 \times 10^{12} M_{\odot}$ (Zaritsky et al. 2020), where our results fit snugly within.

There are many more approaches to calculating the virialized dark matter mass of the MW. Analysing kinematics of various objects within the MW dark matter halo is one of the most common methods as there are many tracer object options (e.g. halo stars, globular clusters, hypervelocity stars, streams, and satellite galaxies). The advent of *Gaia* has dramatically increased the precision of such efforts with its proper motion prowess. Wang et al. (2020) found that of the studies basing their observational measurements off *Gaia* DR2 data, the result of Callingham et al. (2019), who compared satellite dynamics to model satellites in the EAGLE cosmological hydrodynamics simulations, is the median value at $1.17^{+0.21}_{-0.15} \times 10^{12} M_{\odot}$ – our results fully agree with theirs. The extremes of the *Gaia*-based studies are $M_{200} = 0.7^{+0.11}_{-0.08} \times 10^{12} M_{\odot}$ (via cumulative mass profiles derived from globular cluster kine-

matics Eadie & Jurić 2019), and $M_{200} = 1.31^{+0.45}_{-0.40} \times 10^{12} M_{\odot}$ (via satellite galaxy proper motions in a scale-free mass estimator Fritz et al. 2020). Our results are in agreement with the more massive result but not with the least massive. As a caveat to kinematics-based mass estimations, Erkal, Belokurov & Parkin (2020) recently illustrated that not including the LMC in such methods can result in overestimating the MW’s mass by up to 50 per cent due to the LMC pushing the MW out of equilibrium. This issue is bypassed in methods directly comparing N -body simulations, where simulated systems are not in equilibrium, to the observed kinematics, as is done in this work.

As described in this section, the field is starting to converge on a well constrained value of the mass of the Milky Way where our results, $1\text{--}1.2 \times 10^{12} M_{\odot}$, fit within the range. Furthermore, the method presented here achieves a metric most other studies do not – we fully quantify which model masses deviate from the observed satellite population at $\gtrsim 3\sigma$ significance. This ‘antigoodness of fit’ measure is a great advantage of null hypothesis tests, such as the MWU.

5.4 Observational completeness

With many new satellite galaxy discoveries within the past 2 decades (York et al. 2000; Willman et al. 2005a, b; Zucker et al. 2006a, b; Belokurov et al. 2010; Bechtol et al. 2015; Drlica-Wagner et al. 2015), the debate related to the observational completeness of the MW satellite population has been revived (Tollerud et al. 2008; Walsh, Willman & Jerjen 2009; Wang, Frenk & Cooper 2013; Hargis, Willman & Peter 2014; Garrison-Kimmel et al. 2017; Jethwa, Erkal & Belokurov 2018; Newton et al. 2018; Carlsten et al. 2020; Samuel et al. 2020). By matching our subhalo subsamples to the observed MW satellite population based on host-centric distance, our analysis effectively minimizes any systematic bias associated with incompleteness. To more fully explore the potential impact of observational completeness on our results, however, we limit the Milky Way satellite population (and corresponding subhalo samples from phELVIS) to systems within 100 kpc. At these Galactocentric distances (<100 kpc), the MW satellite population is relatively complete, especially in the Southern Hemisphere thanks to surveys such as the Dark Energy Survey (DES) and other imaging campaigns using the Dark Energy Camera (DECam) on the Victor M. Blanco Telescope at the Cerro Tololo Inter-American Observatory (Drlica-Wagner et al. 2020).

In our observational data set of systems with proportionally low tangential velocity errors, there are 25 satellite galaxies within 100 kpc. For this restricted – yet largely complete – sample, the observed velocity distributions are again inconsistent with the kinematics of subhaloes drawn from the low- and high-mass host samples (see Table 3). Meanwhile, while the distributions of observed radial velocities for the nearby Milky Way satellites and for the distance-matched phELVIS subhaloes in the intermediate-mass hosts are consistent, the tangential (and total) velocity distribution for the nearby MW satellites is inconsistent ($p < 0.05$) with that of the intermediate-mass subhalo distribution. This slight disagreement between the velocity distributions is likely due to a preference for circular orbits at small host-centric distance in phELVIS, such that the distribution of V_{tan} is biased towards higher velocities relative to that of the observed MW satellites. This perhaps indicates that the tidal disruption of subhaloes within phELVIS may be slightly overestimated or otherwise incomplete in its characterization of orbits within the inner part of the host halo.

5.5 Observational predictions

As more MW satellites are discovered through deep and wide imaging surveys, such as the Legacy Survey of Space and Time (LSST) at the Vera Rubin Observatory (Ivezić et al. 2019) or the Nancy Grace Roman Space Telescope (Akeson et al. 2019), or via future data releases from *Gaia*, the virial mass of the Milky Way might be further refined. Fig. 8 attempts to illustrate the potential future refinement based on different observable quantities. The cumulative kernel density estimates plotted trace the unmatched distribution of all haloes within 350 kpc of the respective host halo from hosts split into the three mass bins used throughout this work, with the median number of haloes per host mass bin being 1650. The low-mass (burgundy dashed lines) and high-mass (sienna dotted lines) sets are compared to the intermediate-mass (black dashed-dotted lines) host halo sets across 9 subhalo (or satellite) characteristics. MWU p -values were calculated for comparisons between the subhalo distributions drawn from the low-mass and high-mass hosts relative to those in the intermediate-mass hosts. Any characteristic with rejected p -values ($p < 0.05$), i.e. large differences between the three host mass binned subhalo populations, stand to be good metrics to test the preferred host mass range as new data becomes available.

Galactocentric total velocity, tangential velocity, physical Galactocentric distance, and infall have the most discernible differences between the host mass sets within phELVIS. Since infall must be inferred from simulations (Fillingham et al. 2019) or modelling the orbital history of the satellite (e.g. Patel et al. 2020) and thus has greater measurement uncertainty, it is likely to be of less help in discriminating between different host mass regimes. Distance and total velocity have the largest differences between subhalo distributions. Subhaloes in the more massive hosts are kinematically hotter and at further distances from the centre of their host halo than those in the less massive hosts. As new satellites are discovered, the characterization of the radial selection function of observed MW satellites will improve, enabling distance to be used as a mass estimator. It will be particularly interesting to explore their total velocity-based phase space to further refine the halo mass of the Milky Way.

6 SUMMARY

Using the Phat ELVIS suite of N -body Milky Way-like cosmological simulations with embedded disc potentials along with the full phase-space information for Milky Way satellites from *Gaia* EDR3, we constrain the dark matter halo mass of the Milky Way and find a preferred mass range of $\sim 1\text{--}1.2 \times 10^{12} M_{\odot}$. A more complete summary of our main results are as follows:

- (i) As illustrated in Fig. 4, when limiting the observed sample of Milky Way satellites to those systems with well-measured kinematics, we find that the observed distribution of satellite velocities (V_{rad} , V_{tan} , and V_{tot}) are consistent with a host halo mass of $\sim 1\text{--}1.2 \times 10^{12} M_{\odot}$.
- (ii) Across all samples probed, the distribution of satellite velocities inferred from *Gaia* observations of the Milky Way satellites are inconsistent, at the 3σ confidence level, with that of subhaloes populating host haloes with masses $<10^{12}$ or $>1.2 \times 10^{12} M_{\odot}$. Our use of the MWU test allows us to quantify the inconsistency or ‘antigoodness of fit’.
- (iii) Excluding systems associated with the LMC does not significantly change our results, with the observed kinematics of the Milky Way satellites favouring a host halo mass of $\sim 1\text{--}1.2 \times 10^{12} M_{\odot}$ when compared to distance-matched subhalo populations in phELVIS.

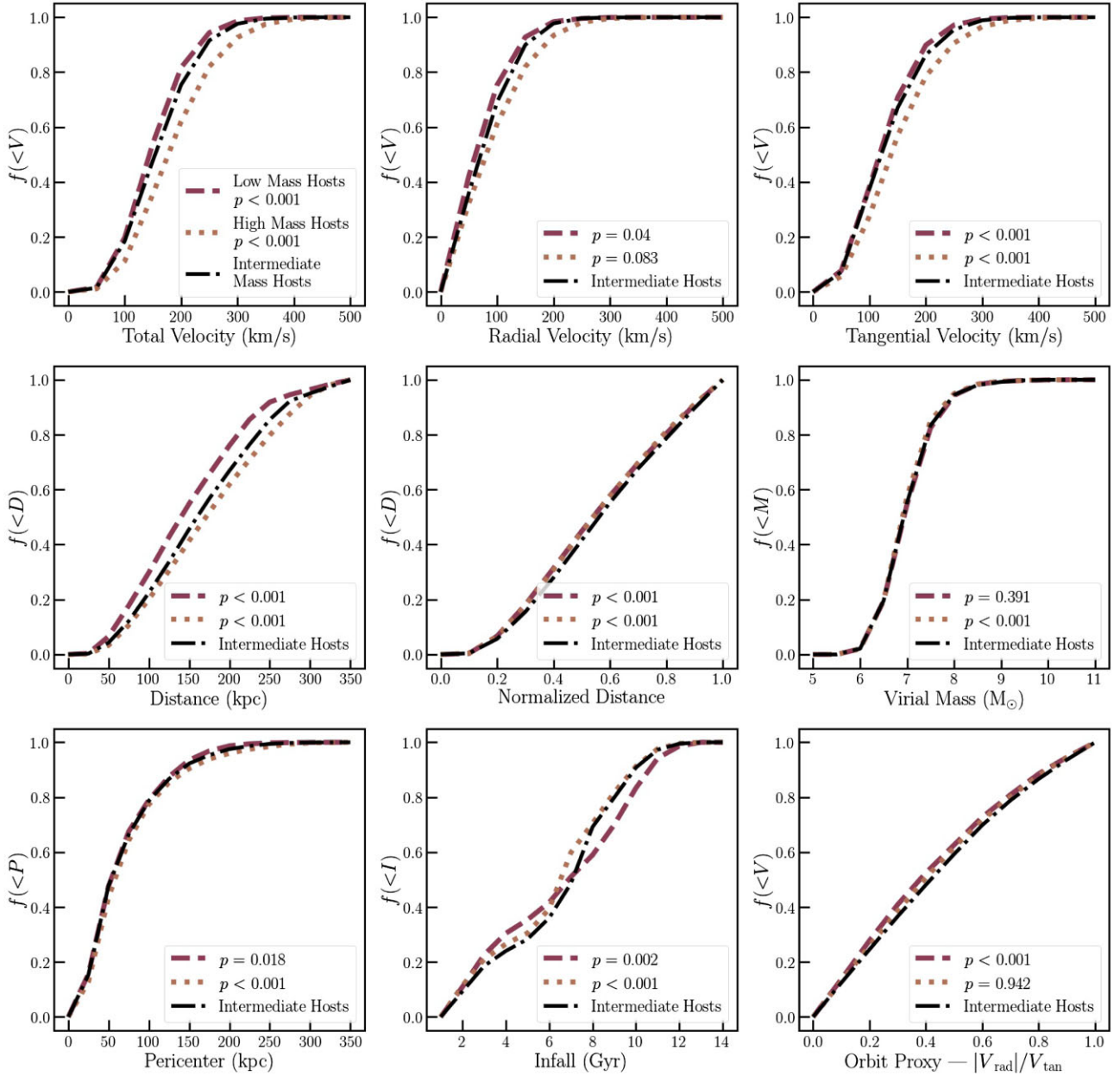


Figure 8. Cumulative distributions of all haloes within 350 kpc of their respective host halo split into three host halo mass sets. The low-mass (burgundy dashed lines) and high-mass (sienna dotted lines) sets are compared to the intermediate mass (black dashed–dotted lines) host halo sets across 9 subhalo properties. These plots illustrate which physical parameters of newly discovered satellites will assist us in further refining the dark matter mass content of the MW – namely two of the most straightforward to obtain properties – distance and total velocity.

(iv) In the inner halo ($D_{\text{MW}} < 100$ kpc), we find a correlation between host mass and the eccentricity of satellite orbits (as predicted by GALPY), such that at a given Galactocentric distance increasingly circular orbits are found in higher mass hosts. This is likely a consequence of subhalo destruction preferentially removing satellites on more radial orbits in more massive hosts.

(v) The distribution of infall times inferred from *Gaia* phase-space measures (Fillingham et al. 2019) are systematically skewed towards early cosmic times (i.e. early accretion) relative to that of distance-matched subhaloes drawn from the pHELVIS simulation suite.

(vi) The distribution of pericentric distances for subhaloes in pHELVIS show little dependence on host mass, in contrast to the expectations from GALPY that favour smaller pericentric distances for satellites in more massive host haloes.

(vii) Looking towards the discovery of future Milky Way satellites by next-generation observational facilities, we show that the observed distribution of Galactocentric total velocity and Galactocentric distance stand to be good metrics to test the preferred host mass range for the Milky Way.

ACKNOWLEDGEMENTS

We thank Tyler Kelley and Marcel Pawlowski for helpful discussions regarding this project. This work was supported in part by National Science Foundation (NSF) grants AST-1815475 and AST-1518257. MKRW acknowledges support from the National Science Foundation Graduate Research Fellowship and Math and Physical Sciences

(MPS) Ascend Postdoctoral Research Fellowship. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1321846. This material is based upon work supported by the National Science Foundation MPS-Ascend Postdoctoral Research Fellowship under Grant No. AST-2138144. MBK acknowledges support from NSF CAREER award AST-1752913, NSF grant AST-1910346, National Aeronautics and Space Administration (NASA) grant NNX17AG29G, and *HST*-AR-15006, *HST*-AR-15809, *HST*-GO-15658, *HST*-GO-15901, *HST*-GO-15902, *HST*-AR-16159, and *HST*-GO-16226 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555. LVS acknowledges support from the NASA ATP 80NSSC20K0566 and NSF CAREER 1945310 grants. DCB thanks the Large Synoptic Survey Telescope Corporation (LSSTC) Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant #1829740, the Brinson Foundation, and the Moore Foundation; participation in the program has greatly benefited this work.

This research has made use of NASA's Astrophysics Data System Bibliographic Services. This research also utilized *astropy*, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013, 2018). Additionally, the PYTHON packages NUMPY (Harris et al. 2020), iPython (Pérez & Granger 2007), SCIPY (Jones et al. 2001), MATPLOTLIB (Hunter 2007), and scikit-learn (Pedregosa et al. 2012) were utilized for our data analysis and presentation.

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

DATA AVAILABILITY

The data used within this work is available from the corresponding author upon request.

REFERENCES

- Ahmed S. H., Brooks A. M., Christensen C. R., 2017, *MNRAS*, 466, 3119
 Akeson R. et al., 2019, preprint ([arXiv:1902.05569](https://arxiv.org/abs/1902.05569))
 Astropy Collaboration, 2013, *A&A*, 558, A33
 Astropy Collaboration, 2018, *AJ*, 156, 123
 Battaglia G., Taibi S., Thomas G. F., Fritz T. K., 2022, *A&A*, 657, A54
 Bechtol K. et al., 2015, *ApJ*, 807, 50
 Bellazzini M., Gennari N., Ferraro F. R., Sollima A., 2004, *MNRAS*, 354, 708
 Bellazzini M., Gennari N., Ferraro F. R., 2005, *MNRAS*, 360, 185
 Belokurov V. et al., 2007, *ApJ*, 654, 897
 Belokurov V. et al., 2010, *ApJ*, 712, L103
 Benisty D., 2021, *A&A*, 656, A129
 Benisty D., Guendelman E. I., 2020, *Phys. Dark Univ.*, 30, 100708
 Besla G., Kallivayalil N., Hernquist L., Robertson B., Cox T. J., van der Marel R. P., Alcock C., 2007, *ApJ*, 668, 949
 Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
 Boettcher E. et al., 2013, *AJ*, 146, 94
 Bose S., Deason A. J., Belokurov V., Frenk C. S., 2020, *MNRAS*, 495, 743
 Bovy J., 2015, *ApJS*, 216, 29
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, *MNRAS*, 406, 896
 Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, *MNRAS*, 415, L40
 Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2012, *MNRAS*, 422, 1203
 Boylan-Kolchin M., Bullock J. S., Sohn S. T., Besla G., van der Marel R. P., 2013, *ApJ*, 768, 140
 Boylan-Kolchin M., Weisz D. R., Bullock J. S., Cooper M. C., 2016, *MNRAS*, 462, L51
 Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
 Buck T., Dutton A. A., Macciò A. V., 2016, *MNRAS*, 460, 4348
 Bullock J. S., Boylan-Kolchin M., 2017, *ARA&A*, 55, 343
 Busha M. T., Wechsler R. H., Behroozi P. S., Gerke B. F., Klypin A. A., Primack J. R., 2011, *ApJ*, 743, 117
 Callingham T. M. et al., 2019, *MNRAS*, 484, 5453
 Carlin J. L. et al., 2017, *AJ*, 154, 267
 Carlsten S. G., Greene J. E., Peter A. H. G., Greco J. P., Beaton R. L., 2020, *ApJ*, 902, 124
 Cautun M., Frenk C. S., van de Weygaert R., Hellwing W. A., Jones B. J. T., 2014, *MNRAS*, 445, 2049
 Cautun M., Bose S., Frenk C. S., Guo Q., Han J., Hellwing W. A., Sawala T., Wang W., 2015, *MNRAS*, 452, 3838
 Cautun M. et al., 2020, *MNRAS*, 494, 4291
 Collins M. L. M. et al., 2015, *ApJ*, 799, L13
 Conn A. R. et al., 2013, *ApJ*, 766, 120
 Craig P., Chakrabarti S., Baum S., Lewis B. T., 2021, preprint ([arXiv:2107.09791](https://arxiv.org/abs/2107.09791))
 Crnojević D., Sand D. J., Zaritsky D., Spekkens K., Willman B., Hargis J. R., 2016, *ApJ*, 824, L14
 D'Souza R., Bell E. F., 2021, *MNRAS*, 504, 5270
 Dall'Ora M. et al., 2006, *ApJ*, 653, L109
 Dall'Ora M. et al., 2012, *ApJ*, 752, 42
 Deason A. J., Fattahi A., Frenk C. S., Grand R. J. J., Oman K. A., Garrison-Kimmel S., Simpson C. M., Navarro J. F., 2020, *MNRAS*, 496, 3929
 Drlica-Wagner A. et al., 2015, *ApJ*, 813, 109
 Drlica-Wagner A. et al., 2020, *ApJ*, 893, 47
 Eadie G., Jurić M., 2019, *ApJ*, 875, 159
 Efsthathiou G., 1992, *MNRAS*, 256, 43P
 Elias L. M., Sales L. V., Creasey P., Cooper M. C., Bullock J. S., Rich R. M., Hernquist L., 2018, *MNRAS*, 479, 4004
 Erkal D. et al., 2019, *MNRAS*, 487, 2685
 Erkal D., Belokurov V. A., Parkin D. L., 2020, *MNRAS*, 498, 5574
 Fillingham S. P., Cooper M. C., Wheeler C., Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., 2015, *MNRAS*, 454, 2039
 Fillingham S. P. et al., 2019, preprint ([arXiv:1906.04180](https://arxiv.org/abs/1906.04180))
 Fong Y., Huang Y., 2019, *Am. Stat.*, 73, 43
 Font A. S., McCarthy I. G., Belokurov V., 2021, *MNRAS*, 505, 783
 Fritz T. K., Battaglia G., Pawlowski M. S., Kallivayalil N., van der Marel R., Sohn S. T., Brook C., Besla G., 2018, *A&A*, 619, A103
 Fritz T. K., Di Cintio A., Battaglia G., Brook C., Taibi S., 2020, *MNRAS*, 494, 5178
 Gaia Collaboration, 2016, *A&A*, 595, A1
 Gaia Collaboration, 2018, *A&A*, 616, A1
 Gaia Collaboration, 2021, *A&A*, 649, A1
 Garofalo A. et al., 2013, *ApJ*, 767, 62
 Garrison-Kimmel S. et al., 2017, *MNRAS*, 471, 1709
 Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, *MNRAS*, 438, 2578
 Geha M. et al., 2017, *ApJ*, 847, 4
 Gibbons S. L. J., Belokurov V., Evans N. W., 2014, *MNRAS*, 445, 3788
 Gnedin N. Y., 2000, *ApJ*, 542, 535
 Graus A. S., Bullock J. S., Kelley T., Boylan-Kolchin M., Garrison-Kimmel S., Qi Y., 2019, *MNRAS*, 488, 4585
 Greco C. et al., 2007, *ApJ*, 675, L73
 Hamanowicz A. et al., 2016, *AcA*, 66, 197
 Hargis J. R., Willman B., Peter A. H. G., 2014, *ApJ*, 795, L13
 Harris C. R. et al., 2020, *Nature*, 585, 357
 Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, *Nature*, 563, 85
 Hoeft M., Yepes G., Gottlöber S., Springel V., 2006, *MNRAS*, 371, 401
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
 Ibata R. A., Gilmore G., Irwin M. J., 1994, *Nature*, 370, 194
 Ibata R. A. et al., 2013, *Nature*, 493, 62

- Ibata R. A., Ibata N. G., Lewis G. F., Martin N. F., Conn A., Elahi P., Arias V., Fernando N., 2014, *ApJ*, 784, L6
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jethwa P., Erkal D., Belokurov V., 2018, *MNRAS*, 473, 2060
- Jones E., et al., 2001, SciPy: Open Source Scientific Tools for Python, <http://www.scipy.org/>
- Kafle P. R., Sharma S., Lewis G. F., Bland-Hawthorn J., 2014, *ApJ*, 794, 59
- Kallivayalil N., van der Marel R. P., Besla G., Anderson J., Alcock C., 2013, *ApJ*, 764, 161
- Kallivayalil N. et al., 2018, *ApJ*, 867, 19
- Karachentsev I. D., Kashibadze O. G., 2006, *Astrophysics*, 49, 3
- Kelley T., Bullock J. S., Garrison-Kimmel S., Boylan-Kolchin M., Pawlowski M. S., Graus A. S., 2019, *MNRAS*, 487, 4409
- Kinemuchi K., Harris H. C., Smith H. A., Silberman N. A., Snyder L. A., Lacluyzé A. P., Clark C. L., 2008, *AJ*, 136, 1921
- Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, *ApJ*, 522, 82
- Koposov S. E. et al., 2012, *ApJ*, 750, 80
- Koposov S. E., Belokurov V., Torrealba G., Evans N. W., 2015, *ApJ*, 805, 130
- Koposov S. E. et al., 2018, *MNRAS*, 479, 5343
- Kroupa P., Theis C., Boily C. M., 2005, *A&A*, 431, 517
- Kuehn C. et al., 2007, *ApJ*, 674, L81
- Law D. R., Majewski S. R., Johnston K. V., 2009, *ApJ*, 703, L67
- Lee H., Grebel E. K., Hodge P. W., 2003, *A&A*, 401, 141
- Li H., Hammer F., Babusiaux C., Pawlowski M. S., Yang Y., Arenou F., Du C., Wang J., 2021, *ApJ*, 916, 8
- Li Z.-Z., Qian Y.-Z., Han J., Li T. S., Wang W., Jing Y. P., 2020, *ApJ*, 894, 10
- Lindegren L. et al., 2018, *A&A*, 616, A2
- Longeard N. et al., 2018, *MNRAS*, 480, 2609
- Lynden-Bell D., 1976, *MNRAS*, 174, 695
- Macciò A. V., Kang X., Fontanot F., Somerville R. S., Koposov S., Monaco P., 2010, *MNRAS*, 402, 1995
- Mann H. B., Whitney D. R., 1947, *Ann. Math. Statist.*, 18, 50
- Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, *ApJ*, 907, 85
- Massari D., Helmi A., 2018, *A&A*, 620, A155
- McConnachie A. W., Venn K. A., 2020a, *Res. Notes Am. Astron. Soc.*, 4, 229 (MCV20a)
- McConnachie A. W., Venn K. A., 2020b, *AJ*, 160, 124
- Medina G. E. et al., 2018, *ApJ*, 855, 43
- Metz M., Kroupa P., Libeskind N. I., 2008, *ApJ*, 680, 287
- Monari G. et al., 2018, *A&A*, 616, L9
- Moore B., 2001, in Wheeler J. C., Martel H., eds, AIP Conf. Proc. Vol. 586, 20th Texas Symposium on Relativistic Astrophysics. Am. Inst. Phys., New York, p. 73
- Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, *ApJ*, 524, L19
- Moretti M. I. et al., 2009, *ApJ*, 699, L125
- Müller O., Pawlowski M. S., Jerjen H., Lelli F., 2018, *Science*, 359, 534
- Musella I. et al., 2009, *ApJ*, 695, L83
- Musella I. et al., 2012, *ApJ*, 756, 121
- Mutlu-Pakdil B. et al., 2018, *ApJ*, 863, 25
- Newton O., Cautun M., Jenkins A., Frenk C. S., Helly J. C., 2018, *MNRAS*, 479, 2853
- Ocvirk P. et al., 2016, *MNRAS*, 463, 1462
- Pace A. B., Li T. S., 2019, *ApJ*, 875, 77
- Patel E., Besla G., Mandel K., Sohn S. T., 2018, *ApJ*, 857, 78
- Patel E. et al., 2020, *ApJ*, 893, 121
- Pawlowski M. S., Kroupa P., 2020, *MNRAS*, 491, 3042
- Pawlowski M. S., McGaugh S. S., 2014, *ApJ*, 789, L24
- Pawlowski M. S., Pflamm-Altenburg J., Kroupa P., 2012, *MNRAS*, 423, 1109
- Pawlowski M. S. et al., 2014, *MNRAS*, 442, 2362
- Pedregosa F. et al., 2012, preprint ([arXiv:1201.0490](https://arxiv.org/abs/1201.0490))
- Pérez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Phelps S., Nusser A., Desjacques V., 2013, *ApJ*, 775, 102
- Pietrzyński G. et al., 2008, *AJ*, 135, 1993
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Richings J. et al., 2020, *MNRAS*, 492, 5780
- Rizzi L., Held E. V., Saviane I., Tully R. B., Gullieuszk M., 2007, *MNRAS*, 380, 1255
- Rocha M., Peter A. H. G., Bullock J., 2012, *MNRAS*, 425, 231
- Sales L. V., Navarro J. F., Kallivayalil N., Frenk C. S., 2017, *MNRAS*, 465, 1879
- Samuel J., Wetzel A., Chapman S., Tollerud E., Hopkins P. F., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., 2021, *MNRAS*, 504, 1379
- Sand D. J., Strader J., Willman B., Zaritsky D., McLeod B., Caldwell N., Seth A., Olszewski E., 2012, *ApJ*, 756, 79
- Santos-Santos I. M. E., Fattahi A., Sales L. V., Navarro J. F., 2021, *MNRAS*, 504, 4551
- Sawala T., Pihajoki P., Johansson P. H., Frenk C. S., Navarro J. F., Oman K. A., White S. D. M., 2017, *MNRAS*, 467, 4383
- Shao S., Cautun M., Deason A. J., Frenk C. S., Theuns T., 2018, *MNRAS*, 479, 284
- Shao S., Cautun M., Frenk C. S., 2019, *MNRAS*, 488, 1166
- Simon J. D., 2018, *ApJ*, 863, 89
- Simon J. D., 2019, *ARA&A*, 57, 375
- Sohn S. T., Besla G., van der Marel R. P., Boylan-Kolchin M., Majewski S. R., Bullock J. S., 2013, *ApJ*, 768, 139
- Thoul A. A., Weinberg D. H., 1996, *ApJ*, 465, 608
- Tollerud E. J., Bullock J. S., Strigari L. E., Willman B., 2008, *ApJ*, 688, 277
- Tollerud E. J., Boylan-Kolchin M., Barton E. J., Bullock J. S., Trinh C. Q., 2011, *ApJ*, 738, 102
- Torrealba G., Koposov S. E., Belokurov V., Irwin M., 2016a, *MNRAS*, 459, 2370
- Torrealba G. et al., 2016b, *MNRAS*, 463, 712
- Torrealba G. et al., 2018, *MNRAS*, 475, 5085
- Torrealba G. et al., 2019, *MNRAS*, 488, 2743
- Vivas A. K. et al., 2016, *AJ*, 151, 118
- Walsh S. M., Willman B., Sand D., Harris J., Seth A., Zaritsky D., Jerjen H., 2008, *ApJ*, 688, 245
- Walsh S. M., Willman B., Jerjen H., 2009, *AJ*, 137, 450
- Wang J., Frenk C. S., Navarro J. F., Gao L., Sawala T., 2012, *MNRAS*, 424, 2715
- Wang J., Frenk C. S., Cooper A. P., 2013, *MNRAS*, 429, 1502
- Wang W., Han J., Cautun M., Li Z., Ishigaki M. N., 2020, *Sci. China Phys. Mech. Astron.*, 63, 109801
- Watkins L. L., van der Marel R. P., Sohn S. T., Evans N. W., 2019, *ApJ*, 873, 118
- Weisz D. R. et al., 2016, *ApJ*, 822, 32
- Wheeler C., Phillips J. I., Cooper M. C., Boylan-Kolchin M., Bullock J. S., 2014, *MNRAS*, 442, 1396
- Willman B., Governato F., Dalcanton J. J., Reed D., Quinn T., 2004, *MNRAS*, 353, 639
- Willman B. et al., 2005a, *AJ*, 129, 2692
- Willman B. et al., 2005b, *ApJ*, 626, L85
- Yniguez B., Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., 2014, *MNRAS*, 439, 73
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zaritsky D., Conroy C., Zhang H., Naidu R. P., Bonaca A., Caldwell N., Cargile P. A., Johnson B. D., 2020, *ApJ*, 888, 114
- Zucker D. B. et al., 2006a, *ApJ*, 643, L103
- Zucker D. B. et al., 2006b, *ApJ*, 650, L41

This paper has been typeset from a \LaTeX file prepared by the author.